FACULTY OF ENGINEERING
AND ARCHITECTURE

# Markovian Imprecise Jump Processes: Foundations, Algorithms and Applications

**Alexander Erreygers**

Doctoral dissertation submitted to obtain the academic degree of
Doctor of Mathematical Engineering

**Supervisors**
Prof. Jasper De Bock, PhD* - Prof. Gert De Cooman, PhD* - Prof. Em. Herwig Bruneel, PhD**

* Department of Electronics and Information Systems
Faculty of Engineering and Architecture, Ghent University

** Department of Telecommunications and Information Processing
Faculty of Engineering and Architecture, Ghent University

September 2021

GHENT
UNIVERSITY

ISBN 978-94-6355-515-9
NUR 919
Wettelijk depot: D/2021/10.500/63

## Members of the Examination Board

### Chair

Prof. Hennie De Schepper, PhD, Ghent University

### Other members entitled to vote

Prof. Alessio Benavoli, PhD, Trinity College Dublin, Ireland
Prof. Stijn De Vuyst, PhD, Ghent University
Prof. Matthias Troffaes, PhD, Durham University, United Kingdom
Prof. Joris Walraevens, PhD, Ghent University
Prof. Cassio de Campos, PhD, Technische Universiteit Eindhoven, the Netherlands

### Supervisors

Prof. Jasper De Bock, PhD, Ghent University
Prof. Gert De Cooman, PhD, Ghent University
Prof. Em. Herwig Bruneel, PhD, Ghent University

# *Table of contents*

# *Preface*

Six years of doctoral research, and what an incredible ride it has been. The pages of this dissertation are filled with the fruits of my labour, and it is my absolute pleasure that you are on the cusp of browsing through them – I *do* hope that you will read more than just this preface, because this book would be a terrible waste of paper otherwise.

I have had a lot of help and support along the way, and I would like to use this opportunity to express my gratitude to all those lovely people who have been there for me. I do apologise, my dear reader, for the ton of clichés that I am about to throw your way; I am afraid that after writing nearly 500 pages over the course of a bit more than a year, I have exhausted my ability to write creatively.

First and foremost, I would like to thank my three promotors, professors Jasper De Bock, Gert de Cooman and Herwig Bruneel. Obviously, this dissertation would not have made it into existence without them.

Herwig, who gave me the opportunity to start my academic career at the SMACS research group. I was not an obvious candidate with my background in mechanical engineering, so thanks for believing in me – and kudos to professor Stijn De Vuyst, who, after having been part of the jury for my Master's dissertation, recommended me to you.

Gert, who lured me into the wonderful world of imprecise probabilities – the classes for the linear systems course served as the perfect breadcrumb trail to put me on the right path. You are a true inspiration, and not only on a scientific level. Furthermore, I appreciate tremendously the friendly atmosphere that you have created in our research group. Most of my fondest memories of conferences – road trips, lakes, golf carts and a boat spring to mind – are from those occasions when you took us in tow with your childlike enthusiasm.

Jasper, who taught me all the facets of doing (mathematical) research. I cannot thank you enough for your guidance in general, and in particular for the – at times seemingly endless – stream of painstaking remarks on drafts of papers or this dissertation. Your continuous suggestions for things to look into were always intriguing, and our blackboard sessions – always interesting,

usually rather intense, occasionally heated and sparingly frustrating – were instrumental to the results in this dissertation of which I am most proud of.

More generally, I tip my hat to all my past and present colleagues at Ghent University. For all the fun times during lunches, dinners, drinks and conferences, I am indebted to – in alphabetical order – Alain, Arne, Arthur, Floris, Keano, Meizhu, Michiel, Natan, Simon, Stavros and Thomas. I should thank Thomas in particular for the interesting discussions, and for our fruitful collaboration – I do not even regret doing research whilst visiting a brewery in Compiegne.

Right before I started writing this dissertation, I visited the University of Oviedo for three monts. Enrique was the best host one can imagine, and I owe him many thanks for making me feel welcome, for pulling me out of my comfortable stochastic process bubble, for introducing me to Spanish cuisine and for taking me out on the most wonderful walks in the beautiful Asturian mountains. Our daily coffee breaks were something to look forward to, and I thank Ignacio, Javier and many others for making them as pleasant as they were.

The members of the examination board also deserve my gratitude. Carefully reading and evaluating a 500-page dissertation is a thankless job, so I appreciate their efforts tremendously. I thank them for their valuable comments on my manuscript, and for their thought-provoking questions during both defences.

Ook buiten de academische wereld zijn er velen aan wie ik mijn dankbaarheid zou willen uiten. Eerst en vooral dank ik mijn vrienden – wiens namen ik hier niet zal opnoemen om zeker niemand te vergeten – om er voor me te zijn wanneer nodig, en om te zorgen voor de af en toe zo broodnodige verstrooiing. Bedankt voor de avonden thuis, op restaurant of op café, al dan niet eindigend op de toog van Den Draak; voor het lachen, zwanzen en afzien op menig klimmuur en rotswand; voor het wieltjeszuigen als de wind weer eens tegen zat; en voor zo veel meer. In het bijzonder wil ik graag Sylke bedanken voor haar luisterend oor, bemoedigende woorden en goede raad.

Ten tweede dank ik mijn familie – papa, mama, Tessa & Sara – voor hun onvoorwaardelijke steun. Wat een luxe om – misschien niet frequent genoeg – terug te kunnen keren naar het warme nest dat jullie bieden.

Ten slotte rest er me nog één persoon om te bedanken: mijn steun en toeverlaat, Milena. Dat ik het afgelopen jaar zowel mentaal als fysiek heb overleefd, heb ik grotendeels aan jouw zorg en liefde te danken. Ik ben nog steeds verwonderd over hoe jij op vanzelfsprekende wijze en zonder ooit te morren – tegen mij althans – de afgelopen anderhalf jaar bijna in je eentje ons huishouden hebt doen draaien. ¡Muchisimas gracias por todos, buddy!

Alexander Erreygers
September 2021

# *Summary*

This dissertation covers several theoretical and practical aspects of Markovian imprecise jump processes. A Markovian imprecise jump process is a particular type of stochastic process, meaning that it is a mathematical model for a dynamical system whose state evolves over time in an uncertain manner. In particular, it is a specific type of stochastic process for a system that evolves in continuous time and whose state assumes values in a finite state space.

Because a Markovian imprecise jump process models uncertainty, we begin this dissertation with a brief overview of some of the mathematical tools that can be used to model uncertainty. We adhere to the coherence framework for modelling uncertainty as conceived by de Finetti (1970), P. M. Williams (1975), and Walley (1991), and not to the more conventional measure-theoretical framework advanced by Kolmogorov (1933). In particular, we use coherent conditional probabilities as elementary uncertainty models. Unlike with the probability measures that are used in the measure-theoretical framework, this allows us to condition on events with probability zero without any issue or ambiguity.

Next, we present our take on the framework of (im)precise jump processes, which was originally put forward by Krak et al. (2017). We define a (precise) jump process as a coherent conditional probability on a specific domain: we consider finitary events – events that depend on the state of the system at a finite number of future time points – in combination with conditioning events that fix the state of the system at a finite number of (past) time points. Under the classical Markovianity and (time-)homogeneity assumptions – and a mild continuity assumption – a jump process is uniquely defined by two parameters: its probability mass function and its rate operator. Thus, given these two parameters, we can determine the (conditional) probability of any finitary event, and therefore the (conditional) expectation of any simple variable – a simple variable depends on the state of the system at a finite number of time points. Note that (conditional) probabilities are a special case of (conditional) expectations, so we can focus on the latter.

Specifying precise values for the parameters of a (homogeneous) Marko-

vian jump process can be infeasible if not impossible, especially if these are learned from data and/or are elicited from an expert. This is where Markovian imprecise jump processes come in, as they generalise Markovian jump processes to allow for partial parameter specification. Whereas a Markovian jump process is fully determined by an initial probability mass function and a rate operator, a Markovian imprecise jump process is determined by a set of initial probability mass functions and a bounded set of rate operators. However, there is not one Markovian imprecise jump process, but there are three that we consider. All three of them are defined as sets of jump processes that are consistent with the set of initial probability mass functions and the set of rate operators, but they differ in the type of processes that are considered: the first contains all consistent homogeneous and Markovian jump process, the second contains all consistent Markovian homogeneous jump processes – so it includes the homogeneous ones – and the third simply contains all consistent jump process – so it includes the Markovian ones.

Because we work with sets of jump processes, there is not a single value for the (conditional) expectation of a simple variable, but a set of values; our aim is not to determine this set, but to determine tight lower and upper bounds, which we call lower and upper expectations. Whether or not we can compute these lower and upper bounds in a tractable manner depends on the structure of the set of rate operators. If this set is infinite, which usually is the case, then computing lower and upper expectations of simple variables is computationally intractable for the set of consistent homogeneous and Markovian jump processes. Quite remarkably, it turns out that if the set of rate operators is separately specified (and convex), then we can nevertheless tractably compute tight lower and upper bounds for the other two sets of consistent jump processes. Krak et al. (2017) identify two such cases: (i) separately specified is sufficient for simple variables that depend on the state of the system at a single future time point; and (ii) separately specified rows and convexity are sufficient for general simple variables, although only for the set of all consistent jump process. We identify a third case that sits somewhere between the previous two extreme ones: if the set of rate operators is separately specified, then we can tractably compute lower and upper expectations for simple variables that have a so-called sum-product representation.

In many applications, the variables of interest depend on the state of the system at all time points in a (bounded) time period, so on the state of the system at more than a finite number of time points. One of the more important contributions of this dissertation is that we extend the domain of Markovian imprecise jump processes to deal with such variables. Crucial to our extension is that we consider càdlàg sample paths from the start. Many important variables are then the point-wise limit of a sequence of simple variables, and we call these idealised variables; examples are temporal averages, hitting times and indicators of until events. We show that for any (bounded) set of rate operators, the (conditional) expectation corresponding to any consistent jump process satisfies monotone convergence, so we can extend the domain of this expectation through Daniell's (1918) method of integration. Even more, we show that the tight lower and upper bounds on these extended expectations satisfy (imprecise generalisations of) the Monotone Convergence Theorem and Lebesgue's Dominated Convergence theorem. In general, this convergence might be to conservative bounds, but we show that the convergence is tight for three important types of idealised variables – indicators of time-bounded until events, truncated hitting times and temporal averages. Even more, these idealised variables are the point-wise limits of sequences of simple variables that have a sum-product representation, and we can therefore tractably compute their lower and upper expectations, at least for the set of all consistent (Markovian) jump processes, whenever the set of rate operators is separately specified.

After this theoretical part, we turn to a – still rather theoretical – setting where parameter indeterminacy arises naturally. In many applications with homogeneous and Markovian jump process models, the state space is so large that computing expectations becomes intractable. Lumping the states – sometimes also called grouping or aggregating states – can then significantly reduce the number of states. Unfortunately, characterising the resulting lumped jump process exactly is not possible due to loss of information – at least not in general. We show that this lumped jump process is consistent with a set of initial probability mass functions on the lumped state space that follows naturally from the original initial probability mass functions, and that this lumped jump process is consistent with a set of rate operators on the lumped state space that is induced by the original rate operator. Consequently, we can use the corresponding Markovian imprecise jump process to tractably compute lower and upper bounds on expectations that we could not tractably compute otherwise.

Finally, we show that all of the aforementioned theory serves a purpose. To this end, we consider the problem of spectrum fragmentation in a single optical link, which is an example where the state space of the exact homogeneous Markovian jump process model is too large. Kim, Yan, et al. (2015) only consider the random allocation policy, and they reduce the number of states by lumping; they deal with the resulting parameter indeterminacy through an approximate homogeneous and Markovian model, and they use this model to approximate the blocking ratios. With our method, we obtain guaranteed lower and upper bounds on the blocking ratios instead of approximations, and we do so for the random allocation policy but also for two other policies. Even more, we can determine lower and upper bounds on the blocking ratios that hold for any allocation policy.

# *Samenvatting*

Dit proefschrift behandelt verscheidene theoretische en praktische aspecten van Markoviaanse imprecieze sprongprocessen. Een Markoviaans imprecies sprongproces is een type stochastisch proces, wat betekent dat het een wiskundig model is voor een dynamisch systeem waarvan de toestand op een onzekere manier evolueert in de tijd. Meer specifiek is het een type stochastisch proces voor een systeem dat evolueert in continue tijd en waarvan de toestand waarden aanneemt in een eindige toestandsruimte.

Omdat een Markoviaans imprecies sprongproces onzekerheid modelleert, beginnen we dit proefschrift met een bondig overzicht van enkele wiskundige manieren om onzekerheid te modelleren. We houden ons aan de aanpak om onzekerheid te modelleren van de Finetti (1970), Walley (1991) en P. M. Williams (1975) die op coherentie is gebaseerd, en dus niet aan de meer conventionele maattheoretische aanpak van Kolmogorov (1933). In het bijzonder maken we gebruik van coherente conditionele waarschijnlijkheden als elementaire onzekerheidsmodellen. In tegenstelling tot de waarschijnlijkheidsmaten die gebruikt worden in de maattheoretische waarschijnlijkheidsleer, maakt dat het mogelijk om te conditioneren op gebeurtenissen met waarschijnlijkheid nul, en dit zonder enig probleem of enige ambiguïteit.

Vervolgens presenteren we het kader van (im)precieze sprongprocessen zoals voorgesteld door Krak e.a. (2017), maar dan vanuit een eigen insteek. We definiëren een (precies) sprongproces als een coherente conditionele waarschijnlijkheid met een specifiek domein: we beschouwen zogenoemde eindige gebeurtenissen – gebeurtenissen die afhangen van de toestand van het systeem op een eindig aantal tijdspunten – in combinatie met conditionerende gebeurtenissen die de toestand van het systeem vastleggen op een eindig aantal tijdspunten (in het verleden). Onder de klassieke aannames van Markovianiteit en (tijds-)homogeniteit – en onder een milde continuïteitsaanname – is een sprongproces uniek gedefinieerd door twee parameters: zijn initiële massafunctie en zijn transitietempo-operator. Bijgevolg kunnen we met deze twee parameters de (conditionele) waarschijnlijkheid van elke eindige gebeurtenis bepalen, en daarom ook de (conditionele) verwachtingswaarde van elke eindige toevallige veranderlijke – een eindige toevallige veranderlijke is een toevallige veranderlijke die afhangt van de toestand van het systeem op een eindig aantal tijdspunten. Merk op dat (conditionele) waarschijnlijkheden een speciaal geval zijn van (conditionele) verwachtingswaarden, dus kunnen we ons beperken tot die laatste.

Het is vaak onmogelijk om exact de waarden voor de parameters van een (homogeen) Markoviaans sprongproces op te geven, zeker als ze geschat worden uit data en/of als een expert ze aanlevert. Dit is waar Markoviaanse imprecieze sprongprocessen nut hebben, aangezien ze homogene Markoviaanse sprongprocessen veralgemenen op een manier die onbepaaldheid van de parameters toelaat. In tegenstelling tot een homogeen Markoviaans sprongproces, dat volledig bepaald wordt door één initiële massafunctie en één transitietempo-operator, wordt een Markoviaans imprecies sprongproces bepaald door een verzameling van initiële massafuncties en een (begrensde) verzameling van transitietempo-operatoren. Er zijn echter meerdere Markoviaanse imprecieze sprongprocessen die gekarakteriseerd worden door deze parameters, en we beschouwen er hier drie. Alle drie zijn ze gedefinieerd als verzamelingen van sprongprocessen die consistent zijn met de verzameling initiële massafuncties en de verzameling transitietempo-operatoren, maar ze verschillen in het type processen dat in aanmerking komt: het eerste bevat alle consistente homogene Markoviaanse sprongprocessen, het tweede bevat alle consistente (niet noodzakelijk homogene) Markoviaanse sprongprocessen en het derde bevat alle consistente (niet noodzakelijk Markoviaanse) sprongprocessen.

Aangezien we niet langer uitgaan van één sprongproces maar van een verzameling, is er ook niet langer één verwachtingswaarde van een eindige toevallige veranderlijke maar een verzameling van zulke verwachtingswaarden. Het is niet ons doel om deze verzamelingen te bepalen, maar wel hun boven- en ondergrenzen. Of we al dan niet deze boven- en ondergrenzen kunnen berekenen hangt af van de structuur van de verzameling transitietempo-operatoren. Als deze verzameling oneindig is, wat meestal het geval is, dan is het praktisch onmogelijk om onder- en bovenverwachtingswaarden van eindige toevallige veranderlijken te berekenen voor de verzameling van consistente homogene Markoviaanse sprongprocesses. Merkwaardig genoeg blijkt dat we de boven- en ondergrenzen wel kunnen berekenen voor de twee andere verzamelingen van consistente sprongprocessen, of toch indien de verzameling transitietempo-operatoren componentsgewijs beschreven (en convex) is. Krak e.a. (2017) tonen dit aan voor twee 'extreme' gevallen: (i) componentsgewijs beschreven is voldoende voor eindige toevallige veranderlijken die afhangen van de toestand van het systeem op één toekomstig tijdspunt; en (ii) componentsgewijs beschreven en convex is voldoende voor algemene eindige toevallige veranderlijken, maar wel enkel voor de verzameling van *alle* consistente sprongprocessen. Wij tonen dit aan voor een derde geval: als de verzameling transitietempo-operatoren componentsgewijs beschreven is, dan kunnen we de boven- en onderverwachtingswaarde berekenen voor eindige toevallige veranderlijken die een som-productvorm hebben.

In veel toepassingen hangen de toevallige veranderlijken waarin we geïnteresseerd zijn af van de toestand van het systeem in alle tijdspunten in een (begrensde) tijdsperiode, en dus niet van de toestand van het systeem in een eindig aantal tijdspunten. Een van de belangrijkere bijdragen van dit proefschrift is dat we het domein van Markoviaanse imprecieze sprongprocessen uitbreiden naar zulke variabelen. Het is essentieel voor deze uitbreiding dat we van bij het begin aannemen dat de paden in de mogelijkhedenruimte càdlàg zijn. Veel relevante toevallige veranderlijken zijn dan de puntsgewijze limiet van een rij eindige toevallige veranderlijken, en we noemen ze geïdealiseerde variabelen; voorbeelden zijn tijdsgemiddelden, de tijd tot bereiken en de indicator van de gebeurtenis van bereiken. We tonen aan dat voor elke (begrensde) verzameling transitietempo-operatoren en voor elk consistent proces, de bijbehorende (conditionele) verwachtingswaardeoperator monotoon convergent is, waardoor we deze kunnen uitbreiden aan de hand van Daniells (1918) integratiemethode. Meer nog, we tonen voor de uitgebreide boven- en onderverwachtingswaardeoperatoren aan dat ze voldoen aan varianten van monotone convergentie en gedomineerde convergentie. In het algemeen is deze convergentie conservatief, maar we tonen aan dat deze convergentie exact is voor drie belangrijke types van geïdealiseerde toevallige veranderlijken – indicatoren van de gebeurtenis van bereiken, afgeknotte tijd tot bereiken en tijdsgemiddelden. Meer nog, deze geïdealiseerde toevallige veranderlijken zijn de puntsgewijze limiet van een rij van eindige toevallige veranderlijken die een som-productvorm hebben, en daarom kunnen we hun boven- en onderverwachtingswaarden berekenen, tenminste als de verzameling transitietempo-operatoren componentsgewijs beschreven is.

Na deze brok theorie is het tijd voor een eerste – nog steeds redelijk theoretische – 'praktische' toepassing waar de parameters op natuurlijke wijze onbepaald zijn. Het komt vaak voor dat de toestandsruimte van een homogeen Markoviaans sprongproces zo groot is, dat het uitrekenen van verwachtingswaarden praktisch onmogelijk wordt. We kunnen dan de toestandsruimte verkleinen door toestanden op te hopen, of anders gezegd, door toestanden samen te nemen. In het algemeen gaat dit ophopen gepaard met een verlies aan informatie, waardoor we het 'opgehoopte' sprongproces niet meer exact kunnen bepalen. Wij tonen aan dat dit opgehoopte sprongproces consistent is met een verzameling initiële massafuncties op de opgehoopte toestandsruimte die op natuurlijke wijze volgt uit de oorspronkelijke initiële massafunctie, en ook dat dit opgehoopte sprongproces consistent is met een verzameling transitietempo-operatoren die afgeleid is van de oorspronkelijke transitietempo-operator. Hierdoor kunnen we het overeenkomstige Markoviaans imprecies sprongproces gebruiken om onder- en bovengrenzen te berekenen op verwachtingswaarden die we anders niet hadden kunnen berekenen.

Tenslotte tonen we aan dat al deze theorie wel degelijk een – praktisch – nut heeft. Hiervoor kijken we naar de spectrumversplintering in een optische kabel. Dit kan bestudeerd worden met een homogeen Markoviaans sprongproces, maar de toestandsruimte van dit proces is te groot om er berekeningen mee te kunnen maken. Kim, Yan e.a. (2015) beschouwen enkel de willekeurige toewijzingsprocedure, en zij maken de toestandsruimte kleiner door toestanden op te hopen. Ze omzeilen het verlies aan informatie door met een benaderend homogeen Markoviaans sprongproces te werken, en met dit proces benaderen ze de blokkeringsverhoudingen. Met onze methode berekenen we boven- en ondergrenzen op de blokkeringsverhoudingen in plaats van benaderingen, en dit voor de willekeurige toewijzingsprocedure en nog twee andere toewijzingsprocedures. Meer nog, we kunnen zelfs boven- en ondergrenzen op de blokkeringsverhoudingen berekenen die gelden voor elke mogelijke toewijzingsprocedure.

# 1 Introduction

The theory of Markovian imprecise jump processes has various fascinating facets, and this dissertation aims to come to grips with some of the more elementary ones. Not only does this dissertation extend the existing theory, but it also aims to show that Markovian imprecise jump processes are a useful (computational) tool in applications.

This introductory chapter starts off with some context on and motivation for Markovian imprecise jump processes. It also contains some information on the internal and external references in this work, a short overview of its contents, a list of publications that led to this dissertation and some preliminaries about (sequences of) natural, integer and (extended) real numbers.

## 1.1 Context and motivation

A Markovian imprecise jump process is a type of stochastic process, meaning that it is a mathematical model for a dynamical system whose state evolves over time in a non-deterministic manner. This concept of a 'dynamical system' is perhaps a bit abstract, so let us consider an example that we are all familiar with: a shop that is open 24 hours a day, 7 days a week. Suppose we are interested in the evolution of the number of customers inside such a shop. In this case, the state space of the system – the set of possible values for the state – is discrete and the state evolves in continuous time. It is fair to say that we do not know with certainty when a new customer will enter the shop nor how long they will stay, so the temporal evolution of the state of this system – the number of customers in the shop – is uncertain. Nevertheless, we may still be interested in some derived quantities: straightforward examples are the number of customers that are in the shop at noon tomorrow, or the average number of customers in the shop over the following 24 hour period. Then the idea is that a stochastic process models our uncertainty about (the temporal evolution of) the number of customers, and that we can use this mathematical model to make inferences about the system, that is, that we can use this model to determine the probability of relevant events – for example, the event that at noon tomorrow there will be 10 customers – and the expectation of relevant variables – for example, the temporal average of the number of customers over the following 24 hours.

The number of customers is of course only one example of a 'state' for our system. Alternatively, we could be interested in the temperature in the shop or in the number of clients that visit the shop per day; the former is an example of a state with a continuous state space, and the latter is an example of a state that evolves in discrete time. This illustrates that we can categorize systems – and therefore also the corresponding stochastic processes – according to the nature of their state space and time axis. In this dissertation, we consider a (finite) discrete state space and a continuous time axis; in that case, a stochastic process is called a jump process.

Let us start with the basic case of a classical – or 'precise' – jump process. To specify such a jump process, we need to specify two types of probabilities. The first type of probabilities are the initial probabilities; these model the uncertainty regarding the initial state of the system. The second type of probabilities are the transition probabilities; these model the uncertainty regarding the evolution of the state. In general, a transition probability gives the probability that the system is in some state $y$ at a future time point $r$, given that the state of the system is $x$ at the current time point $t$ and that the system's state was $z_1, \ldots, z_n$ at the past time points $s_1, \ldots, s_n$. For example, the probability that there are 7 customers at 10:10 today given that there are 10 customers now at 10:00 and that there were 8 customers at 9:15 and 12 customers at 8:12. Specifying these transition probabilities is a nuisance, because we have to specify a transition probability for *all* combinations of a future time point and future state, current time point and current state and past time points and past states; for this reason, it is customary to make the following two simplifying assumptions.

First and foremost, it is customary to assume that the transition probabilities are Markovian,[1] in the sense that they only depend on the state of the system at the current time point $t$ and *not* on the system's state at the past time points $s_1, \ldots, s_n$. For our example from before, this means that the probability that there are 7 customers at 10:10 today given that there are 10 customers now at 10:00 and that there were 8 customers at 9:15 and 12 customers at 8:12, is equal to the probability that there are 7 customers at 10:10 today given that there are 10 customers at 10:00. Second, it is convenient to assume that the transition probabilities are (time-)homogeneous, meaning that it is only the length $r - t$ of the period between the current time point $t$ and the future time point $r$ that matters, and not the start (or end) point of this time period. In our example, this means that the probability that there

[1] After Markov (1906), who popularised a similar assumption in his work on the Weak Law of Large Numbers (Seneta, 1996, 2006).

are 7 customers at 10:10 today given that there are 10 customers at 10:00 is equal to the probability that there are 7 customers at 23:23 given that there are 10 customers at 23:13.

We call a jump chain that satisfies these two assumptions a homogeneous Markovian jump chain. These were first studied by Kolmogorov (1931), Doeblin (1938)[2] and Fréchet (1938), and their theory was further developed by Doob (1953), Chung (1960), Feller (1968) and many others. For a more contemporary account of this theory, see (Iosifescu, 1980; Norris, 1997).

Due to the assumptions of Markovianity and homogeneity, we now only need to specify the transition probabilities for the initial time point $t = 0$ and any initial state $x$, and for any future time point $r$ and any future state $y$. In fact, it has been shown that – under a mild continuity assumption – these transition probabilities are fully determined by their initial rate of change, so by their derivatives at $r = 0$. Thus, a homogeneous Markovian jump process is fully determined by its initial probabilities and its transition rates. Notwithstanding this succinct parameterisation, a homogeneous Markovian jump process can be used to make various non-trivial inferences: from the probability of an event that depends on the state of the system at a finite number of future time points – for example, the probability that there is no customer in the shop at noon tomorrow – to the expectation of a variable that depends on the state of the system at *all* future time points over some (possibly unbounded) time period – for example, the expected temporal average of the number of customers in the shop over the next 24 hours.

Homogeneous Markovian jump process have been used successfully in an assortment of scientific fields, including biochemistry (Ganguly et al., 2014), epidemiology (Kay, 1986), reliability analysis (Besnard et al., 2010; Troffaes et al., 2015) and telecommunications engineering (Kim, Yan, et al., 2015), mainly because of their simplicity. However, that is not to say that there are no issues with using homogeneous Markovian jump process in these applications. One issue that all of these applications have in common, is that the defining parameters of the model – and in particular the transition rates – are learned from data and/or elicited from experts. Usually, this results in a single 'exact' estimate for each of the parameter values. However, the inference of interest often depends on the parameters in a non-trivial way, so it might occur that slightly different parameter values lead to widely different conclusions. Thus, there is a clear need for a type of jump process that can deal with partially specified parameters instead of exact parameter values. A second issue is that the homogeneity and/or Markovianity assumptions may not be justified. For instance, in our example of the shop, the homogeneity is difficult to justify; to give just one possible issue, the number of customers will evolve differently on a working day than on a bank holiday. Thus, there is

[2]Doeblin's tragic life story is well worth the read; see for example Lévy's (1955) account, or the more recent one by Bru et al. (2002)

a also a clear need for a type of jump process that does not need to assume homogeneity and/or Markovianity.

Hartfiel (1985) was the first to investigate the dependency of the inferences on the (rate) parameters of a (not necessarily homogeneous) Markovian jump process – at least in a general setting – but he only considers inferences that depend on a single future time point. Quite some time later, Škulj (2015) did something similar; specifically, he provides an algorithm to compute tight lower and upper bounds on inferences that depend on the state at a single future time point. Krak et al. (2017) put Škulj's (2015) work on a sound theoretical basis, and in doing so they defined Markovian imprecise jump processes – although they call them imprecise continuous-time Markov chains. They define a Markovian imprecise jump process through a notion of 'consistency' with sets of parameters, and focus on three specific sets: (i) the set of all consistent homogeneous and Markovian jump processes; (ii) the set of all consistent (not necessarily homogeneous) Markovian jump processes; and (iii) the set of all (not necessarily Markovian) jump processes. Thus, they provide a way to deal with parameter indeterminacy, and at the same time they are also able to let go of the homogeneity assumption and/or the Markovianity assumption. That said, their theory can only deal with events and variables that depend on the state of the system at a finite number of future time points, and this significantly limits the applicability of Markovian imprecise jump processes.

This brings us to the first main contribution of this dissertation, which is to enlarge the scope of the theory of Markovian imprecise jump processes by extending it to events and variables that depend on the state of the system at *all* time points in some (unbounded) time period. Besides the necessary theoretical framework, we also provide algorithms to determine (tight lower and upper bounds on) the probability of such events or the expectation of such variables.

The second main contribution of this dissertation addresses yet another issue with homogeneous Markovian jump processes: the computational methods to determine probabilities and expectations for such processes become intractable if the state space is too large. Burke et al. (1958) propose to solve this issue by lumping together states. Unfortunately, this lumping is coupled with a loss of information, and this means that the resulting 'lumped' jump process is hard to characterise. This dissertation shows that lumping essentially causes parameter indeterminacy and a loss of Markovianity and homogeneity, and establishes that a 'lumped' Markovian imprecise jump process can be used to capture this. This enables us to tractably compute bounds on probabilities and expectations that we could not tractably compute otherwise.

## 1.2 References

As you have undoubtedly already noticed by now, we refer to other works with the name of the (first) author and the year of publication. The full bibliographical details of these references are available in the Bibliography, which starts on page 485.

This dissertation also contains countless internal references – figuratively speaking, that is – to chapters, (sub)sections, equations, figures, tables, theorems, propositions, lemmas, corollaries and (running) examples. Locating the referenced material can be a rather cumbersome affair, especially in a work of this size. For this reason, we resort to Quaeghebeur's (2009) system of locational clues, as has become tradition for doctoral dissertations and monographs written by members of FLip[3] (Quaeghebeur, 2009; Troffaes et al., 2014; De Bock, 2015; Lopatatzidis, 2017; Van Camp, 2017; Krak, 2021). Whenever an internal reference is not located on the same double-page spread as the material it references, we add a clue to the location of this material as a subscript: ↶ or ↷ whenever the material is on the recto page or the verso page, respectively, and a page number otherwise. For example: the proof of Theorem 5.19$_{230}$ – quite an important result in Section 5.2.1$_{228}$ of Chapter 5$_{215}$ – has been relegated to Appendix 5.C$_{252}$ and uses the terminology and tools of measure-theoretic probability theory as introduced in Appendix C$_{461}$

## 1.3 Brief overview

Besides this introductory chapter and the conclusions in Chapter 9$_{439}$, this dissertation consists of seven chapters and three appendices. This section gives a *very* brief overview of these chapters and appendices.

Chapter 2$_{11}$ introduces the basic mathematical tools that we will use to model uncertainty. Especially important is Section 2.4$_{40}$ where we introduce coherent conditional probabilities, as these are our elementary uncertainty models for jump processes.

In Chapter 3$_{53}$, we introduce the framework of (im)precise jump process as conceived by (Krak et al., 2017). First we define jump processes as coherent conditional probabilities with a specific domain, second we consider Markovian (and homogeneous) jump processes as a special case, and third we define imprecise jump processes as sets of (consistent) jump processes.

Chapter 4$_{157}$ is all about computing lower and upper expectations for imprecise jump processes. Most important are the two algorithms we give to compute the lower and upper expectation of simple variables. We also investigate how we can evaluate the operator exponential of a lower rate operator, and we have our first encounter with ergodicity.

[3] The Foundations Lab for imprecise probabilities; a research group at Ghent University formerly known as SYSTeMS.

Next, we move on from simple variables to idealised variables in Chapter $5_{215}$. This chapter has two parts. In the first part, we discover how Daniell's (1918) integration method extends the expectation corresponding to a generic countably additive probability charge. In the second part, we use this extension method to extend the domain of a generic countably additive jump process. We subsequently show that any (consistent) jump process in an imprecise jump process is countably additive, and finally use this result to extend the lower and upper expectations for an imprecise jump process.

Chapter $6_{273}$ is to Chapter $5_{215}$ what Chapter $4_{157}$ was to Chapter $3_{53}$: a second chapter about computing lower and upper expectations, but in this one we focus on four types of idealised variables.

Moving on, we consider lumping in Chapter $7_{337}$. First, we investigate the issues that arise when lumping a single jump process. Second, we consider the particular case of lumping a jump process that is consistent with a set of rate operators, and we discover how in this case we can use a lumped imprecise jump process to describe the original jump process. Third, we propose two lumping-based methods to compute limit expectations for a homogeneous Markovian jump process.

We put (our methods for) imprecise jump processes to the test in Chapter $8_{403}$, where we model the problem of spectrum allocation in an optical link.

Chapter $9_{439}$ concludes the main text of this dissertation. What remains are three appendices, a list of symbols and the bibliography. Appendix $A_{443}$ establishes compactness of the set of lower transition operators and any bounded set of (lower) rate operators, Appendix $B_{451}$ contains some extra material – including relegated proofs – on Daniell's (1918) extension method and Appendix $C_{461}$ gives a brief introduction to those parts of measure-theoretic probability theory that we need.

## 1.4 List of publications

This dissertation gathers my research on the topic of Markovian imprecise jump processes in one coherent – and largely self-contained – narrative. Hence, many of the results that it presents have previously appeared elsewhere. For example, many of the results in Chapters $4_{157}$, $7_{337}$ and $8_{403}$ appeared in, or are preceded by work in, the following publications.

(i) Cristina Rottondi, Alexander Erreygers, Giacomo Verticale & Jasper De Bock (2017). Modelling spectrum assignment in a two-service flexi-grid optical link with imprecise continuous-time Markov chains. In: *Proceedings of the 13th International Conference on the Design of Reliable Communication Networks (DRCN 2017)* [best paper award]

(ii) Alexander Erreygers & Jasper De Bock (2017a). Imprecise continuous-time Markov chains: Efficient computational methods with guaranteed error bounds. In: *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2017)*

(iii) Alexander Erreygers & Jasper De Bock (2018a). Computing inferences for large-scale continuous-time Markov chains by combining lumping with imprecision. In: *Uncertainty Modelling in Data Science (Proceedings of SMPS 2018)* [best paper award]

(iv) Alexander Erreygers, Cristina Rottondi, Giacomo Verticale & Jasper De Bock (2018b). Imprecise Markov models for scalable and robust performance evaluation of flexi-grid spectrum allocation policies. In: *IEEE Transactions on Communications*

(v) Alexander Erreygers & Jasper De Bock (2019a). Bounding inferences for large-scale continuous-time Markov chains: A new approach based on lumping and imprecise Markov chains. In: *International Journal of Approximate Reasoning*

The results in Chapters $5_{215}$ and $6_{273}$ are entirely new, but some of them appear (without proof) in the proceedings of ISIPTA 2021.

(vi) Alexander Erreygers & Jasper De Bock (2021). Extending the domain of imprecise jump processes from simple variables to measurable ones. In: *Proceedings of the Twelfth International Symposium on Imprecise Probabilities: Theories and Applications (ISIPTA 2021)*

Furthermore, I have also done research on the topic of learning Markovian imprecise jump processes. While these results are relevant to the material in this dissertation, I have chosen not to include them because statistical estimation is not directly related to the main goals of this dissertation.

(vii) Thomas Krak, Alexander Erreygers & Jasper De Bock (2018). An imprecise probabilistic estimator for the transition rate matrix of a continuous-time Markov chain. In: *Uncertainty Modelling in Data Science (Proceedings of SMPS 2018)*

During my time as a doctoral candidate, I have also endeavoured research on topics other than Markovian imprecise jump processes. This has led to the following publications, some of which are tangentially related to the material in this dissertation.

(viii) Alexander Erreygers, Jasper De Bock, Gert de Cooman & Arthur Van Camp (2019). Optimal control of a linear system subject to partially specified input noise. In: *International Journal of Robust and Nonlinear Control*

(ix) Alexander Erreygers & Jasper De Bock (2019b). First steps towards an imprecise Poisson process. In: *Proceedings of the Eleventh Interna-*

*tional Symposium on Imprecise Probabilities: Theories and Applications (ISIPTA 2019)*

(x) Alexander Erreygers & Enrique Miranda (2020). A study of the set of probability measures compatible with comparative judgements. In: *Information Processing and Management of Uncertainty in Knowledge-Based Systems (Proceedings of IPMU 2020)*

(xi) Jasper De Bock, Alexander Erreygers & Thomas Krak (2021). Sum-product laws and efficient algorithms for imprecise Markov chains. In: *Proceedings of the 37th Conference on Uncertainty in Artificial Intelligence (UAI 2021)*

(xii) Alexander Erreygers & Enrique Miranda (2021). A graphical study of comparative probabilities. In: *Journal of Mathematical Psychology*

## 1.5 Mathematical preliminaries

We denote the set of integers by $\mathbb{Z}$; additionally, $\mathbb{Z}_{\geq 0}$ denotes the set of non-negative integers and $\mathbb{N} := \mathbb{Z}_{>0}$ that of the natural numbers, that is, the positive integers. Similarly, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}_{\geq 0}$ that of the non-negative real numbers and $\mathbb{R}_{>0}$ that of the positive real numbers.

We will also need the set of extended real numbers $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$, and we extend the binary relations '=', '≤' and '<' and the arithmetic operations of addition '+' and multiplication '×' to the extended real numbers in the usual way – see also (Taylor, 1985, Sections 1-7 and 4-1) or (Troffaes et al., 2014, Appendix D). That is, we say that $+\infty = +\infty$, $-\infty = -\infty$ and $-\infty < +\infty$, that $a \leq +\infty$ and $-\infty \leq a$ for all $a$ in $\overline{\mathbb{R}}$ and that $a < +\infty$ and $-\infty < a$ for all $a$ in $\mathbb{R}$. As far as the arithmetic operations are concerned, we extend the addition '+' from $\mathbb{R}$ to $\overline{\mathbb{R}}$ as follows:

$$(+\infty) + (+\infty) = +\infty, \qquad (-\infty) + (-\infty) = -\infty,$$
$$a + (+\infty) = (+\infty) + a = +\infty, \qquad a + (-\infty) = (-\infty) + a = -\infty,$$

where $a$ is any real number in $\mathbb{R}$. This way, the addition '+' on $\overline{\mathbb{R}}$ is commutative and associative. Note that we do not define $+\infty + (-\infty)$ or $-\infty + (+\infty)$; whenever a sum of extended real numbers cannot be reduced to one of these two cases, we call it *well-defined*. Furthermore, we extend the binary relation '×' from $\mathbb{R}$ to $\overline{\mathbb{R}}$ as follows:

$$(+\infty) \times (+\infty) = (-\infty) \times (-\infty) = +\infty, \quad (+\infty) \times (-\infty) = (-\infty) \times (+\infty) = -\infty,$$
$$0 \times (+\infty) = (+\infty) \times 0 = 0, \qquad 0 \times (-\infty) = (-\infty) \times 0 = 0;$$

for any positive real number $a$ in $\mathbb{R}_{>0}$, we let

$$a \times (+\infty) = (+\infty) \times a = +\infty, \qquad a \times (-\infty) = (-\infty) \times a = -\infty,$$

and for any negative real number $a$ in $\mathbb{R}_{<0}$, we let

$$a \times (+\infty) = (+\infty) \times a = -\infty, \qquad a \times (-\infty) = (-\infty) \times a = +\infty.$$

Note that the multiplication '×' on $\overline{\mathbb{R}}$ is commutative and associative.

If $(a_n)_{n\in\mathbb{N}}$ is a sequence of (extended) real numbers, then we say that it is *non-decreasing* whenever $a_n \leq a_{n+1}$ for all $n$ in $\mathbb{N}$, and *increasing* whenever these inequalities are strict; for *non-increasing* and *decreasing* sequences, the relation is $a_n \geq a_{n+1}$ and $a_n > a_{n+1}$.

# 2 Modelling uncertainty

At its very core, this dissertation is about jump processes, a type of mathematical models for systems that evolve over time in a non-deterministic or, better, uncertain fashion. In other words, jump processes are a specific model for uncertainty. Thus, it is certainly worthwhile to first take a look at the bigger picture of modelling uncertainty in general, and this is precisely what we set out to do in this chapter.

We start this chapter in Section 2.1 with some general remarks concerning mathematical models of uncertainty. We subsequently introduce the coherence-centred approach to modelling uncertainty in Section $2.2_{16}$. In Section $2.3_{31}$, we will see how classical probability theory uses probability charges to model uncertainty, how this is related to coherence and how conditioning works for probability charges. Finally, we turn to coherent conditional probabilities in Section $2.4_{40}$.

This chapter is rather technical, and is for a large part made up of – more or less – standard concepts, definitions and results in the mathematical theory of modelling uncertainty. In order to make the process of reading this rather technical material as pleasant as possible, we will illustrate most of the material with a running example. In choosing this running example, we have drawn inspiration from Doob (1953), Feller (1968), and Billingsley (1995) – and virtually every other treatise on probability theory.

**Bruno's Example 2.1.** Bruno, a brilliant scientist, comes into possession of a very precious golden coin – a florin, to be more precise – on what he will later describe as the best day of his life. The first thing he does after acquiring his golden treasure, is to flip it. ₰

## 2.1 Modelling uncertainty in general

The aim of probability theory – or, more generally, uncertainty theory – is, of course, to model *uncertainty*. More specifically, with uncertainty theory we mean the entire collection of mathematical tools that aim to model the uncertainty of 'someone' about the outcome of some 'experiment' that he or she is interested in. Common examples of experiments are flipping a coin – as in Bruno's Example 2.1↶ – rolling a die, or the weather tomorrow; but experiments can be way more involved, as we will see in Bruno's Example 2.31$_{31}$ or Chapter 3$_{53}$ further on.

### 2.1.1 Outcomes and the possibility space

The first step in obtaining a mathematical model of uncertainty is to set up an abstract framework. In this dissertation, we will make exclusive use of the well-known abstract framework that deals with a fixed possibility space. For alternative abstract frameworks that do not require a (fixed) possibility space, we refer to (de Finetti, 2017) or (P. M. Williams, 1975, 2007).

Central to our abstract framework is the uncertain outcome of an experiment, denoted by $X$; an example is the tossing of a coin, as in Bruno's Example 2.1↶. We denote a generic realisation or outcome by $x$, and we collect the possible outcomes of our experiment in the *possibility space* $\mathcal{X}$. A possibility space $\mathcal{X}$ is supposed to be *exhaustive* – that is, at least one of the outcomes occurs – and *mutually exclusive* – that is, at most one of the outcomes occurs. Thus, a possibility space is a non-empty set.

**Bruno's Example 2.2.** For a single coin toss, the possibility space is $\mathcal{X} = \{\mathtt{H}, \mathtt{T}\}$, where H and T denote the outcome 'the coin lands with heads facing up' and 'the coin lands with tails facing up', respectively.

Note that even in the most basic case of a single coin toss, we make an idealisation. For example, we exclude the possibility that the coin lands precisely on its side and stays upright. That being said, a single coin flip is not the most interesting experiment, and it certainly does not allow us to illustrate all the intricacies involved in the remainder. We therefore slightly alter the setting of our running example.

**Bruno's Example 2.3.** Bruno is so excited by his discovery, that he cannot keep himself to flipping the coin only once. Therefore, we will consider $n$ consecutive coin flips, with $n$ a natural number. The outcome of this experiment is the $n$-tuple $X_{1:n} := (X_1, \ldots, X_n)$, where $X_k$ is the (uncertain) outcome of the $k$-th coin flip. Consequently, the possibility space $\mathcal{X}$ consists of all $n$-tuples in $\{\mathtt{H}, \mathtt{T}\}$:

$$\mathcal{X} := \{\mathtt{H}, \mathtt{T}\}^n.$$

We will denote a generic outcome in $\mathcal{X}$ by $x_{1:n} = (x_1, \ldots, x_n)$. Note that our use of $X_{1:n}$ and $x_{1:n}$ is in line with the general convention of using $X$ for the uncertain outcome of the experiment and $x$ for a generic outcome.

### 2.1.2 Events

In practice, an event is a statement about the outcome of the experiment that is either true or false, depending on the realisation $x$ of the uncertain outcome $X$. For this reason, we identify an event with the subset of outcomes for which this statement is true. In our theory, we simply call any subset $A$ of the possibility space $\mathcal{X}$ an *event*. Thus, the *set of all events* $\mathcal{P}(\mathcal{X})$ is simply the power set of $\mathcal{X}$.

Two obvious examples of events are the degenerate ones $\emptyset$ and $\mathcal{X}$, often called the *impossible event* and the *sure event*, respectively. When it is necessary to exclude the impossible event from the set of all events, we will write $\mathcal{P}(\mathcal{X})_{\supset\emptyset} := \mathcal{P}(\mathcal{X}) \setminus \{\emptyset\}$.

**Bruno's Example 2.4.** Bruno is bursting with excitement after finding his florin. Among other things, he is very curious whether the $k$-th of the $n$ coin flips will result in heads, with $k$ a natural number such that $k \leq n$. In the formalism that we have set up, he is interested in the event

$$\{X_k = \mathtt{H}\} := \{x_{1:n} \in \mathcal{X} \colon x_k = \mathtt{H}\}.$$

Furthermore, he wonders if all of the first $k$ flips will produce heads; in our formalism, this corresponds to the event

$$H_k := \big\{x_{1:n} \in \mathcal{X} \colon (\forall \ell \in \{1, \dots, k\})\ x_\ell = \mathtt{H}\big\}.$$

Observe that

$$H_k = \bigcap_{\ell=1}^{k} \{X_\ell = \mathtt{H}\}.$$

### 2.1.3 Variables

In practice, a variable is a quantity that is determined by the outcome of the experiment. In our theory, a *variable* corresponds to a function on the possibility space $\mathcal{X}$. Such a variable is said to be an *extended real variable*, a *real variable* or a *non-negative real variable* if its codomain is the extended set of real numbers $\overline{\mathbb{R}}$, the set of real numbers $\mathbb{R}$ or the set of non-negative real numbers $\mathbb{R}_{\geq 0}$, respectively. We collect all extended real variables in $\overline{\mathbb{V}}(\mathcal{X})$ and all real variables in $\mathbb{V}(\mathcal{X})$.

We extend all binary operations on the extended real numbers to the extended real variables – and therefore also to the real variables – in a pointwise manner. For any extended real variables $f$ and $g$, and with $*$ denoting one of the binary operations on the extended real numbers – that is, addition, subtraction, multiplication or division – we let $f * g$ be the extended real variable defined by

$$[f * g](x) := f(x) * g(x) \quad \text{for all } x \in \mathcal{X},$$

whenever the right-hand side of this inequality is well-defined for all outcomes $x$ in the possibility space $\mathcal{X}$. We will often use more than these binary operations to construct extended real variables. One prevalent way is to take the absolute value; more concretely, we let $|f|$ be the non-negative extended real variable that maps every outcome $x$ to the absolute value $|f(x)|$ of its image under $f$. Another way is to take point-wise minima and maxima. If $f$ and $g$ are extended real variables on $\mathcal{X}$, then their point-wise minimum is the extended real variable

$$f \wedge g \colon \mathcal{X} \to \overline{\mathbb{R}} \colon x \mapsto [f \wedge g](x) := \min\{f(x), g(x)\} \tag{2.1}$$

and similarly, their point-wise maximum is

$$f \vee g \colon \mathcal{X} \to \overline{\mathbb{R}} \colon x \mapsto [f \vee g](x) := \max\{f(x), g(x)\}. \tag{2.2}$$

Finally, we will often use real variables that take on the value $\mu$ everywhere, with $\mu$ any real number. For the sake of brevity, we denote this constant variable by $\mu$ as well whenever it is clear from the context whether $\mu$ is a real number or the corresponding constant real variable.

In a similar point-wise manner, we also extend the binary relations '=', '≤', '≥' to extended real variables. A notable exception are the binary relations '<' and '>'; we denote their point-wise extensions by '⋖' and '⋗', and reserve '<' and '>' for the standard strict order corresponding to '≤' and '≥'. More precisely, we write $f < g$ whenever $f \leq g$ and $f \neq g$, and $f > g$ if $g < f$.

To shorten our notation, we will write $\sup f$ and $\inf f$ instead of $\sup\{f(x) \colon x \in \mathcal{X}\}$ and $\inf\{f(x) \colon x \in \mathcal{X}\}$, respectively, and the same holds for $\max f$ and $\min f$, if applicable. An extended real variable $f$ is *bounded above* whenever $\sup f < +\infty$, and *bounded below* whenever $\inf f > -\infty$. If the extended real variable $g$ is both bounded above and below it is simply called *bounded*; we often refer to bounded real variables as *gambles*. Because gambles will play an import role in Section 2.2$_{16}$, we denote the real vector space of all gambles by $\mathbb{G}(\mathcal{X})$. Note that any constant real variable is bounded, so it is a gamble.

An important subclass of gambles are the indicator variables. With any event $A$ in $\mathcal{P}(\mathcal{X})$, we associate the *indicator* $\mathbb{I}_A$ of $A$, which takes the value 1 on $A$ and 0 elsewhere:

$$\mathbb{I}_A \colon \mathcal{X} \to \{0,1\} \colon x \mapsto \mathbb{I}_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Besides providing a way to go from events to variables, indicators also provide a mnemonic for the notation of taking point-wise minima and maxima: for any two events $A$ and $B$, $\mathbb{I}_A \wedge \mathbb{I}_B = \mathbb{I}_{A\cap B}$ and $\mathbb{I}_A \vee \mathbb{I}_B = \mathbb{I}_{A\cup B}$.

While indicators allow us to go from events to variables, level sets allow us to go the other way around. For any extended real variable $f$ and any real number $\alpha$, we define the *level set*

$$\{f \gtrdot \alpha\} := \{x \in \mathcal{X} \colon f(x) > \alpha\}. \tag{2.4}$$

Note that in this definition, we use some of the notational conventions discussed in the previous: on the left-hand side, we identify the real number $\alpha$ with the corresponding constant real variable (or gamble) and the relation '$\gtrdot$' is the point-wise extension of '$>$'.

**Bruno's Example 2.5.** Alicia, who shares an office with Bruno, is interested in the number of times Bruno will flip heads in the first $k$ flips, with $k$ a natural number such that $k \leq n$. This corresponds to the variable

$$h_k \colon \{\mathtt{H}, \mathtt{T}\}^n \to \{0, \ldots, k\} \colon x_{1:n} \mapsto \left|\left\{\ell \in \{1, \ldots, k\} \colon x_\ell = \mathtt{H}\right\}\right|.$$

Note that $h_k$ is a non-negative gamble. Furthermore, $h_k$ can be written as a sum of indicators:

$$h_k = \sum_{\ell=1}^{k} \mathbb{I}_{\{X_\ell = \mathtt{H}\}}.$$

Observe that for any real number $\alpha$ such that $0 \leq \alpha < 1$, the level set $\{h_k \gtrdot \alpha\}$ corresponds to the union of the events $\{X_1 = \mathtt{H}\}, \ldots, \{X_k = \mathtt{H}\}$. Similarly, if $\alpha$ is a real number such that $k-1 \leq \alpha < k$, then the level set $\{h_k \gtrdot \alpha\}$ corresponds to the intersection of the events $\{X_1 = \mathtt{H}\}, \ldots, \{X_k = \mathtt{H}\}$. Finally, if $\alpha \geq k$, then the level set $\{h_k \gtrdot \alpha\}$ is clearly empty. ∯

### 2.1.4 Interpretation

A model, and consequently also the inferences that we can make with it, only make sense if we decide on an interpretation. Interpretations of mathematical models of uncertainty are available in abundance. According to Walley (1991, Section 1.3) and Hájek (2019, Section 3), there are two main types of interpretations: those that are focused on aleatory uncertainty and those for which uncertainty is epistemic.

In the *aleatoric* interpretation, uncertainty is something that occurs empirically. One example is the frequentist interpretation, in which probability is interpreted as the limit of the relative frequency. Therefore, we can only use frequentism to interpret models of uncertainty for experiments that can be repeated infinitely – or at least often enough – under the same exact circumstances.

This is in contrast with the *epistemic* interpretation of uncertainty, in which uncertainty is subjective, representing someone's personal beliefs. An important example is the class of behavioural interpretations, in which someone's behavioural dispositions – for example, her or his betting behaviour – are used to quantify their uncertainty.

Important to emphasise here is that the results in this dissertation hold regardless of the adopted interpretation. We will, however, avail ourselves of the subjective and behavioural interpretation to motivate the coherence framework in this section and, more importantly, in order to motivate specific choices when modelling jump processes in Chapter 353 further on.

## 2.2 Modelling uncertainty using coherence

The approach of modelling uncertainty using coherence was popularised by de Finetti (1970) in his seminal work, in which he lays out a subjective and behavioural theory for modelling uncertainty by means of fair prices in a betting situation, giving them the name *coherent previsions*. As a convenience to the reader – who (hopefully) understands English but might not have sufficiently mastered the Italian language just yet – we will refer to the recently republished version (de Finetti, 2017) of the English translation (de Finetti, 1974, 1975) instead of de Finetti's (1970) original book in Italian.

Many authors later expanded and generalised de Finetti's initial ideas. For example, Regazzini (1985), Berti et al. (1991), Berti et al. (2002) and many others extended his work to conditional previsions. Other authors have generalised this theory to allow for imprecision, by relaxing the concepts of fair price and prevision to supremum buying price and lower prevision – or their conjugates, infimum selling price and upper prevision. This was essentially initiated by P. M. Williams (1975, 2007) in his pioneering work, which in turn inspired the formative works of Walley (1991) and Troffaes et al. (2014); see also (Augustin et al., 2014) for a good introductory-level overview.

Our treatment of modelling uncertainty using coherence more or less follows that of Troffaes et al. (2014). One key difference – at least when it comes down to language – is that we prefer to use 'expectation' instead of 'prevision', because this will allow us to link back more easily when discussing the classical approach in Sections $2.3_{31}$ and $5.1_{216}$ further on. In this section, we first discuss acceptable gambles, the foundation of the theory, in Section 2.2.1. In Section $2.2.2_{19}$, we then use this theory to introduce coherent expectations. Section $2.2.3_{24}$ concerns extending coherent expectations, and we conclude this section in Section $2.2.4_{27}$ with a discussion of coherent lower expectations.

### 2.2.1 Acceptable gambles

Essential to the motivation of modelling uncertainty using coherence is that a gamble $g$ in $\mathbb{G}(\mathcal{X})$ is interpreted as an uncertain reward. This uncertain reward is expressed in some predetermined linear utility scale, and we interpret a negative reward as a loss of utility. Examples of linear utility scales are 'sufficiently small' amounts of money (de Finetti, 2017, Section 3.2) or lottery tickets of a lottery with a single prize (Walley, 1991, Section 2.2.2).

Also fundamental is the idea that the beliefs that 'someone' has about the experiment are reflected in their betting behaviour. We will refer to this 'someone' as *the subject*, and from here on we are interested in those gambles on (the outcome of) the experiment that she[1] is disposed to accept. Following

[1] Throughout this dissertation, we will consistently use the pronouns 'she', 'her' and 'hers' when referring to the subject.

Troffaes et al. (2014, Section 3.3), we say that the subject *accepts* the gamble $g$ – or, alternatively, that $g$ is an *acceptable gamble* to her – if she is disposed to agree to the transaction where first the outcome $x$ in $\mathcal{X}$ of our experiment is determined, and she is subsequently rewarded $g(x)$.

In this light, it is important to realise that not all gambles – that is, bounded real variables – actually allow this gambling interpretation. More precisely, the gambling interpretation only makes sense for those gambles that are *determinable*,[2] which, in the words of de Finetti (2017, Section 2.3.2), means that they 'should be specified in such a way that a possible bet based upon it can be decided without question'. That being said, we will introduce the mathematical theory of coherence without paying much attention to this distinction, mainly because the mathematics are valid regardless of the betting interpretation. We will, however, bring this distinction back to the fore occasionally, especially all the way at the end of this chapter in Section $2.4.4_{51}$, and also in Sections $3.1_{55}$ and $5.1.1_{217}$ further on.

### Rationality criteria

We assume that the subject's behaviour is rational, and characterise this assumption through the four formal rationality criteria as posited by Troffaes et al. (2014, Axiom 3.1). First and foremost, we require that she *avoids partial loss*, that is,

A1. she should not be disposed to accept any gamble that never increases but can decrease her utility,

and that she *accepts partial gain*, that is,

A2. she should be disposed to accept any gamble that cannot decrease her utility.

Second, we have the following two consequences of the linearity of the utility scale.

A3. If she accepts the gamble $g$, then she should also be disposed to accept the gamble $\lambda g$ for any non-negative real number $\lambda$.

A4. If she accepts the gambles $g$ and $h$, then she should also be disposed to accept $g + h$.

These four rationality criteria are not the only rationality criteria that we could use, and in fact many alternatives have been proposed and/or used. The main difference is that not all authors consider the zero gamble 0 to be acceptable; this is reflected in their slightly different versions of (A1), (A2) and/or (A3). For an overview and discussion of possible alternatives, we refer to (Quaeghebeur, 2014) and references therein. That being said, Troffaes et al. (2014, Section 3.4.1) argue that these differences are only essential to

[2] The translators use the term 'well-determined' instead of determinable.

the study of acceptable gambles themselves, and that they do not influence the study of coherent (lower) expectations in any significant way.

#### Avoiding partial loss

It is clearly infeasible if not impossible to ask the subject to specify for every gamble $g$ on $\mathcal{X}$ whether or not she deems it acceptable. Therefore, we always start from a *partial* assessment of acceptability; for example, an assessment that is obtained through elicitation will only consist of a finite number of gambles for practical reasons. More concretely, we assume that we have some set $\mathcal{A}$ of gambles on $\mathcal{X}$ that the subject is disposed to accept; for obvious reasons, we will call such a subset $\mathcal{A}$ of $\mathbb{G}(\mathcal{X})$ a *set of acceptable gambles*. Because it is only a partial assessment, the set $\mathcal{A}$ of acceptable gambles need not be – and, usually, will not be – exhaustive, that is, not every gamble $g$ that is not included in $\mathcal{A}$ is necessarily unacceptable to the subject.

Any partial assessment $\mathcal{A}$ can be enlarged with gambles that are implied by any of the four rationality criteria listed above. Note that the criterion of avoiding partial loss is not constructive, but we will come back to this in a minute. First, we observe that the other three criteria imply that the subject should be disposed to accept the gambles in

$$\mathcal{E}(\mathcal{A}) := \left\{ h + \sum_{k=1}^{n} \lambda_k g_k \colon h \in \mathbb{G}(\mathcal{X})_{\geq 0}, n \in \mathbb{N}, \lambda_k \in \mathbb{R}_{\geq 0}, g_k \in \mathcal{A} \right\}. \tag{2.5}$$

We still have to check if the subject can be forced to accept a partial loss though. Observe that a gamble $g$ incurs partial loss whenever $g \leq 0$ and $g \neq 0$; this can be written more concisely as $g < 0$, using the notational conventions introduced in Section $2.1.3_{13}$.

**Definition 2.6.** A set $\mathcal{A}$ of acceptable gambles on some possibility space $\mathcal{X}$ *avoids partial loss* if for all natural numbers $n$, all non-negative real numbers $\lambda_1, \ldots, \lambda_n$ and all $g_1, \ldots, g_n$ in $\mathcal{A}$,

$$\sum_{k=1}^{n} \lambda_k g_k \nless 0.$$

The following result follows almost immediately from the previous definition and Eq. (2.5).

**Proposition 2.7.** *A set $\mathcal{A}$ of acceptable gambles on some possibility space $\mathcal{X}$ avoids partial loss if and only if $\mathcal{E}(\mathcal{A})$ avoids partial loss.*

One – trivial – example of a set of acceptable gambles that avoids partial loss is the *vacuous* one.

**Vacuous Example 2.8.** Consider any possibility space $\mathcal{X}$. A trivial example of a set of acceptable gambles that avoids partial loss is $\mathbb{G}(\mathcal{X})_{\geq 0}$, the set of non-negative gambles. Clearly, for all natural numbers $n$, all non-negative real numbers $\lambda_1, \dots, \lambda_n$ and all $g_1, \dots, g_n$ in $\mathbb{G}(\mathcal{X})_{\geq 0}$,

$$\sum_{k=1}^{n} \lambda_k g_k \geq 0.$$ ¤

For a more involved example, we return to the familiar setting of our running example.

**Bruno's Example 2.9.** When dealing with acceptable gambles, the case of a single coin toss, corresponding to $n = 1$, *is* interesting. Therefore, we will restrict us to this case in this running example until further notice.

One reason that the case of a binary possibility space is interesting, is that gambles and sets of acceptable gambles then have a clarifying graphical representation. Quaeghebeur (2014) explains that to obtain this well-known graphical representation, it suffices to identify the gamble $g$ on $\{\mathtt{H}, \mathtt{T}\}$ with the point $(g(\mathtt{H}), g(\mathtt{T}))$ in the $(\mathtt{H}, \mathtt{T})$-plane.

Consider a set of acceptable gambles $\mathcal{A}$. Then with its graphical representation, we can determine $\mathcal{E}$ as the convex hull of $\mathcal{A}$ and the first quadrant $\mathbb{G}(\{\mathtt{H}, \mathtt{T}\})_{\geq 0}$. Furthermore, it follows from Definition 2.6 and Proposition 2.7 that with this graphical representation, the set $\mathcal{A}$ avoids partial loss if and only if $\mathcal{E}(\mathcal{A})$ excludes the third quadrant $-\mathbb{G}(\{\mathtt{H}, \mathtt{T}\})_{<0}$, to be more precise. We can use these observations to graphically check whether $\mathcal{A}$ avoids partial loss.

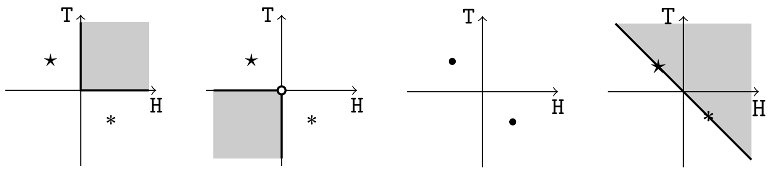
To see how this works, we consider the set

$$\mathcal{A} := \{g^{\star}, g^{*}\},$$

where we let $g^{\star}$ be the gamble on $\{\mathtt{H}, \mathtt{T}\}$ that maps $\mathtt{H}$ to $-1$ and $\mathtt{T}$ to $1$, and we let $g^{*} := -g^{\star}$. It is quite easy to formally check that $\mathcal{A}$ avoids partial loss, but verifying this becomes trivial when looking at the graphical representation of $\mathcal{E}(\mathcal{A})$, as depicted in Fig. 2.1. ∲

### 2.2.2 Coherent expectations

We now move from the realm of acceptable gambles to that of real-valued maps on a subset of the set of gambles $\mathbb{G}(\mathcal{X})$. Let us consider a real-valued map $E$ on a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. Following de Finetti (2017), we think of $E(g)$ as the subject's fair price for the gamble $g$. This means that she accepts to pay any price $\alpha < E(g)$ in exchange for the reward $g$, and that she accepts to receive any price $\beta > E(g)$ in exchange for the reward $-g$. Because she is disposed to buying $g$ for any price $\alpha < E(g)$ and selling $g$ for any price $\beta >$




(a) $\mathbb{G}(\{\mathtt{H},\mathtt{T}\})_{\geq 0}$ (b) $\mathbb{G}(\{\mathtt{H},\mathtt{T}\})_{<0}$ (c) $\mathcal{A}$ (d) $\mathcal{E}(\mathcal{A})$

We use shading to indicate the relevant set of gambles and thick lines to indicate their border. A filled circle indicates an included gamble and an open circle one that is excluded. The gambles $g^\star = (-1, 1)$ and $g^* = (1, -1)$ are indicated by $\star$ and $*$, respectively.

**Figure 2.1** Graphical representation of the sets of (acceptable) gambles in Bruno's Example 2.9.

$E(g)$, we associate with the real-valued map $E$ the set of acceptable gambles

$$\mathcal{A}_E := \big\{(g-\alpha)\colon g \in \mathcal{G}, \alpha \in \mathbb{R}, \alpha < E(g)\big\} \\ \cup \big\{(\beta - g)\colon g \in \mathcal{G}, \beta \in \mathbb{R}, \beta > E(g)\big\}. \quad (2.6)$$

**Bruno's Example 2.10.** As in Bruno's Example 2.9, we consider the case of a singe coin flip. Recall from Bruno's Example 2.9 that $g^\star$ is the gamble that maps $\mathtt{H}$ to $-1$ and $\mathtt{T}$ to 1, and that $g^* = -g^\star$. Here, we consider the real-valued map $E_1$ on $\mathcal{G}_1 := \{g^\star\}$ defined by $E_1(g^\star) = 0$. Additionally, we also define the expectation $E_2$ on $\mathcal{G}_2 := \{g^\star, g^*\}$ as $E_2(g^\star) = 0$ and $E_2(g^*) = 0$. By Eq. (2.6), the set of acceptable gambles associated with $E_1$ is

$$\mathcal{A}_{E_1} = \big\{(g^\star - \alpha)\colon \alpha \in \mathbb{R}, \alpha < E_1(g^\star) = 0\big\} \cup \big\{(\beta - g^\star)\colon \beta \in \mathbb{R}, \beta > E_1(g^\star) = 0\big\}.$$

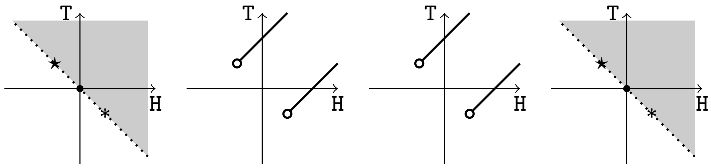
Similarly, we find that

$$\mathcal{A}_{E_2} = \big\{(g^\star - \alpha)\colon \alpha \in \mathbb{R}, \alpha < 0\big\} \cup \big\{(g^* - \alpha)\colon \alpha \in \mathbb{R}, \alpha < 0\big\} \\ \cup \big\{(\beta - g^\star)\colon \beta \in \mathbb{R}, \beta > 0\big\} \cup \big\{(\beta - g^*)\colon \beta \in \mathbb{R}, \beta > 0\big\}.$$

Because $g^* = -g^\star$, it is obvious that $\mathcal{A}_{E_1} = \mathcal{A}_{E_2}$. ∲

### Avoiding partial loss for expectations

Now that we have laid down how a real-valued map $E$ on $\mathcal{G}$ corresponds to a set of acceptable gambles $\mathcal{A}_E$, the natural question to ask is whether this set $\mathcal{A}_E$ avoids partial loss. It is not difficult to see that by combining Definition 2.6$_{18}$ with Eq. (2.6), we end up with the following necessary and sufficient condition – see (de Finetti, 2017, Section 3.3.5) or (Walley, 1991, Definition 2.8.1).




(a) $\mathcal{E}(\mathcal{A}_{E_1})$ (b) $\mathcal{A}_{E_1}$ (c) $\mathcal{A}_{E_2}$ (d) $\mathcal{E}(\mathcal{A}_{E_2})$

Besides the graphical conventions of Fig. 2.1, we use thick dotted lines to indicate that the border is excluded.

**Figure 2.2** Graphical representation of the sets of acceptable gambles in Bruno's Example 2.10.

**Theorem 2.11.** *Consider a possibility space $\mathcal{X}$ and a real-valued map $E$ on a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. Then $\mathcal{A}_E$ avoids partial loss if and only if for all natural numbers $n$, all real numbers $\mu_1, \dots, \mu_n$ and all $g_1, \dots, g_n$ in $\mathcal{G}$,*

$$\sup\left(\sum_{k=1}^{n} \mu_k\big(g_k - E(g_k)\big)\right) \geq 0. \tag{2.7}$$

Whenever a real-valued map $E$ satisfies any of the two equivalent conditions in Theorem 2.11, we call it a (coherent) expectation. As mentioned before at the beginning of Section 2.2$_{16}$, our use of the term expectation is somewhat atypical; de Finetti (2017, Section 3.3.5) prefers the term prevision, while Walley (1991, Definition 2.8.1) and Troffaes et al. (2014, Definition 4.11) use the term linear prevision.

**Definition 2.12.** Consider a possibility space $\mathcal{X}$ and a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. A *(coherent) expectation on $\mathcal{G}$* is a real-valued map on $\mathcal{G}$ such that $\mathcal{A}_E$ avoids partial loss. We denote the set of all coherent expectations on $\mathcal{G}$ by $\mathbb{E}_{\mathcal{G}}$.

For any coherent expectation $E$ and any gamble $g$ in its domain, we call $E(g)$ the *expectation of $g$*. Furthermore, for any event $A$ in $\mathcal{P}(\mathcal{X})$ whose indicator $\mathbb{I}_A$ belongs to the domain of $E$, we call $P(A) := E(\mathbb{I}_A)$ the *probability of $A$*. Let us look at an example of a coherent expectation in the setting of our running example.

**Bruno's Example 2.13.** Let $E_1$ and $E_2$ be the real-valued maps as defined in Bruno's Example 2.10. From the graphical representation depicted in Fig. 2.2, we conclude immediately that $\mathcal{A}_{E_1}$ and $\mathcal{A}_{E_2}$ avoid partial loss, so $E_1$ and $E_2$ are coherent expectations.

### Properties of coherent expectations

Coherent expectations have some properties that are intuitive when $E(g)$ is interpreted as a subject's fair price for $g$. We summarise the most essential properties in the following result; we refer to (Troffaes et al., 2014, Corollary 4.14) for a proof and much more exhaustive lists of properties.

**Proposition 2.14.** *Consider a possibility space $\mathcal{X}$ and a coherent expectation $E$ on a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. Then*

E1. $E(g) \geq \inf g$ *for all* $g \in \mathcal{G}$*;*

E2. $E(\mu g) = \mu E(g)$ *for all* $\mu \in \mathbb{R}$ *and* $g \in \mathcal{G}$ *such that* $\mu g \in \mathcal{G}$*;*

E3. $E(g+h) = E(g) + E(h)$ *for all* $g, h \in \mathcal{G}$ *such that* $g + h \in \mathcal{G}$.

*Furthermore, if we let g and h be gambles in $\mathcal{G}$ and μ a real number, then the following properties hold whenever every term is well-defined:*

E4. $\inf g \leq E(g) \leq \sup g$*;*

E5. $E(g+\mu) = E(g) + \mu$*;*

E6. $E(g) \leq E(h)$ *whenever* $g \leq h$*;*

E7. $|E(g) - E(h)| \leq E(|g-h|)$.

Observe that (E1) is a straightforward consequence of accepting partial gains $(A2)_{17}$: if $\alpha < \inf g$, then $g - \alpha \geq 0$ is a partial gain so the subject should be disposed to accept it. Similarly, (E2) is justified through $(A3)_{17}$. Finally, (E3) is justified by $(A4)_{17}$. To see this, we fix any $\alpha_g < E(g)$, $\alpha_h < E(h)$, $\beta_g > E(g)$ and $\beta_h > E(h)$. Because the subject is disposed to accept $(g - \alpha_g)$ and $(h - \alpha_h)$, $(A4)_{17}$ implies that she should also be disposed to accept $(g + h - \alpha_g - \alpha_h)$. Similarly, the subject is disposed to accept $(\beta_g - g)$ and $(\beta_h - h)$, so $(A4)_{17}$ implies that she should also be disposed to accept $(\beta_g + \beta_h - g - h)$. Consequently, her fair price for $g + h$ is $E(g) + E(h)$.

Properties (E2) and (E3) ensure that the coherent expectation $E$ is a linear functional. The three properties (E1)–(E3) are necessary for a real-valued map $E$ on $\mathcal{G}$ to be a coherent expectation, but they are not sufficient in general. That said, in case the domain $\mathcal{G}$ of the map $E$ is a linear space, two of the three turn out to suffice in order for $E$ to be a coherent expectation; this was essentially proven by de Finetti (2017), but also occurs in (Walley, 1991, Theorem 2.8.4) and (Troffaes et al., 2014, Theorem 4.16).

**Proposition 2.15.** *Consider a possibility space $\mathcal{X}$ and a real-valued map $E$ on a linear subspace $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. Then $E$ is a coherent expectation if and only if it satisfies (E1) and (E3) of Proposition 2.14.*

### Probability mass functions

One straightforward consequence of Proposition 2.15 is that it allows us to characterise coherent expectations in terms of probability mass functions, at least for a *finite* possibility space.

**Definition 2.16.** Consider a finite possibility space $\mathcal{X}$. A *probability mass function p on* $\mathcal{X}$ is a real-valued map on $\mathcal{X}$ that is

M1. non-negative, in the sense that $p(x) \geq 0$ for all $x$ in $\mathcal{X}$; and

M2. normalised, in the sense that $\sum_{x\in\mathcal{X}} p(x) = 1$.

We denote the set of all probability mass functions on $\mathcal{X}$ by $\Sigma_{\mathcal{X}}$.

With any probability mass function $p$ on some finite possibility space $\mathcal{X}$, we can associate the expectation

$$E_p \colon \mathbb{G}(\mathcal{X}) \to \mathbb{R} \colon g \mapsto E_p(g) := \sum_{x\in\mathcal{X}} p(x)g(x). \tag{2.8}$$

Note that evaluating $E_p(g)$ corresponds to taking the normed weighted average of the values of $g$ with the weights – or masses – given by $p$, or more generally, that $E_p(g)$ is a convex mixture of the values of $g$. For this reason, $E_p$ satisfies (E1) and (E3). Because furthermore $\mathbb{G}(\mathcal{X})$ is a linear space, we deduce with the help of Proposition 2.15 that $E_p$ is coherent.

**Corollary 2.17.** *Consider a finite possibility space $\mathcal{X}$. Then for any probability mass function p on $\mathcal{X}$, the real-valued map $E_p$ as defined by Eq. (2.8) is a coherent expectation.*

*Proof.* Follows immediately from Proposition 2.15. □

Conversely, one can show that for every coherent expectation $E$ on $\mathcal{G}$, there is at least one probability mass function $p$ on $\mathcal{X}$ such that $E(g) = E_p(g)$ for all $g$ in $\mathcal{G}$. In case $\mathcal{G} = \mathbb{G}(\mathcal{X})$,[3] this probability mass function is unique and given by $p(x) = E(\mathbb{I}_{\{x\}})$ for all $x$ in $\mathcal{X}$.

**Proposition 2.18.** *Consider a finite possibility space $\mathcal{X}$. Then for any coherent expectation E on $\mathbb{G}(\mathcal{X})$,*

$$p_E \colon \mathcal{X} \to \mathbb{R} \colon x \mapsto p_E(x) := E(\mathbb{I}_{\{x\}})$$

*is a probability mass function, and this is the unique probability mass function p on $\mathcal{X}$ such that $E = E_p$.*

*Proof.* See (Troffaes et al., 2014, Section 5.1). □

[3] Actually, all we need for there to be a unique probability mass function is that $\mathcal{G}$ contains a basis for $\mathbb{G}(\mathcal{X})$; this is for example the case if the indicator $\mathbb{I}_{\{x\}}$ of every outcome $x$ in $\mathcal{X}$ belongs to $\mathcal{G}$.

### 2.2.3 Extending coherent expectations

Consider a coherent expectation $E$ on some subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. Let us investigate whether we can use the subject's fair prices for the gambles in $\mathcal{G}$ to say something about some other gamble $g$ in $\mathbb{G}(\mathcal{X}) \setminus \mathcal{G}$. Because we have interpreted the corresponding set $\mathcal{A}_E$ as a partial acceptability assessment from the subject, it follows from the constructive rationality requirements that she should be disposed to accept the gambles in $\mathcal{E}(\mathcal{A}_E)$, as explained in Section 2.2.1$_{16}$. Thus, for any[4] gamble $g$ in $\mathbb{G}(\mathcal{X})$, she should be disposed to accept to buy the gamble $g$ for any price strictly lower than

$$\underline{\mathcal{E}}_E(g) := \sup\{\alpha \in \mathbb{R} \colon (g - \alpha) \in \mathcal{E}(\mathcal{A}_E)\}, \tag{2.9}$$

so we call this her (inferred) supremum acceptable buying price for $g$. Similarly, she should be disposed to accept to sell the gamble $g$ for any price strictly greater than

$$\overline{\mathcal{E}}_E(g) := \inf\{\beta \in \mathbb{R} \colon (\beta - g) \in \mathcal{E}(\mathcal{A}_E)\}, \tag{2.10}$$

and we therefore call this her (inferred) infimum acceptable selling price for $g$. Note that, by definition, $\underline{\mathcal{E}}_E$ and $\overline{\mathcal{E}}$ are *conjugate*, in the sense that for all $g$ in $\mathbb{G}(\mathcal{X})$,

$$\begin{aligned}
\underline{\mathcal{E}}_E(g) &= \sup\{\alpha \in \mathbb{R} \colon (g - \alpha) \in \mathcal{E}(\mathcal{A}_E)\} \\
&= \sup\{\alpha \in \mathbb{R} \colon ((-\alpha) - (-g)) \in \mathcal{E}(\mathcal{A}_E)\} \\
&= -\inf\{-\alpha \in \mathbb{R} \colon ((-\alpha) - (-g)) \in \mathcal{E}(\mathcal{A}_E)\} \\
&= -\overline{\mathcal{E}}(-g).
\end{aligned}$$

Walley (1991, Definition 3.1.1 and Section 3.1.2) calls $\underline{\mathcal{E}}_E$ the *natural extension* of $E$ – and so do Troffaes et al. (2014, Definition 4.8) – and he shows that $\underline{\mathcal{E}}_E$ is real-valued – and therefore, by conjugacy, so is $\overline{\mathcal{E}}_E$.

Our theory would be contradictory were the subject's fair price for $g$ not to coincide with her (inferred) supremum acceptable buying price and infimum acceptable selling price; fortunately, this is not the case.

**Proposition 2.19.** *Consider a possibility space $\mathcal{X}$ and a coherent expectation $E$ on a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. Then for all $g$ in $\mathcal{G}$,*

$$E(g) = \sup\{\alpha \in \mathbb{R} \colon (g - \alpha) \in \mathcal{E}(\mathcal{A}_E)\} = \inf\{\beta \in \mathbb{R} \colon (\beta - g) \in \mathcal{E}(\mathcal{A}_E)\}.$$

*Proof.* Follows immediately from (Troffaes et al., 2014, Definition 4.8 and Theorem 4.42). □

[4] Let us forget for a moment that we should really only use the implications of the rationality requirements for determinable gambles.

For all other gambles $g$ in $\mathbb{G}(\mathcal{X}) \setminus \mathcal{G}$, Walley (1991, Section 3.1.2) proves that

$$\underline{\mathcal{E}}_E(g) \leq \overline{\mathcal{E}}_E(g).$$

If this inequality holds with equality, then the common value of $\underline{\mathcal{E}}_E(g)$ and $\overline{\mathcal{E}}_E(g)$ is the subject's (inferred) fair price for $g$. Thus, on the basis of our rationality requirements, we could extend $E$ to the (possibly) larger domain

$$\left\{g \in \mathbb{G}(\mathcal{X}) : \underline{\mathcal{E}}_E(g) = \overline{\mathcal{E}}_E(g)\right\}$$

by setting $E(g) = \underline{\mathcal{E}}_E(g) = \overline{\mathcal{E}}_E(g)$. Clearly, we would then still have the interpretation of $E(g)$ as being the subject's fair price for $g$. Furthermore, it is not all too difficult to show using Proposition 2.718 that the set of acceptable gambles corresponding to this extension would avoid partial loss, thus making this extension a coherent expectation. Here, we will follow a slightly different route to prove this.

One convenient property of coherent expectations is that they can always be extended to a coherent expectation on a larger domain. Troffaes et al. (2014, Theorem 4.42) establish this, and we repeat their result here.

**Proposition 2.20.** *Consider a possibility space $\mathcal{X}$, a real-valued map E on a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$ and a superset $\mathcal{K}$ of $\mathcal{G}$. Then E is a coherent expectation if and only if it can be extended to a coherent expectation $E^\star$ on $\mathcal{K}$.*

**Bruno's Example 2.21.** Let us again consider the two coherent expectations $E_1$ and $E_2$ as defined in Bruno's Example 2.1020. First, let us determine a coherent expectation $E_1^\star$ on $\mathbb{G}(\mathcal{X})$ that extends $E_1$. Then by definition, we need that $E_1^\star(g^\star) = E_1(g^\star) = 0$. Furthermore, because the possibility space $\mathcal{X} = \{\mathtt{H}, \mathtt{T}\}$ is finite, we know from Proposition 2.1823 that for any such coherent expectation $E_1^\star$, $E_1^\star = E_p$ with $p$ a probability mass function on $\{\mathtt{H}, \mathtt{T}\}$. Clearly, there is only one probability mass function $p$ on $\{\mathtt{H}, \mathtt{T}\}$ such that

$$0 = E_1(g^\star) = E_1^\star(g^\star) = E_p(g^\star) = p(\mathtt{H})g^\star(\mathtt{H}) + p(\mathtt{T})g^\star(\mathtt{T}) = -p(\mathtt{H}) + p(\mathtt{T}),$$

and that is the one that assigns mass 1/2 to H and to T. Thus, in this case $E_p$ is the only coherent expectation on $\mathbb{G}(\{\mathtt{H}, \mathtt{T}\})$ that extends $E_1$.

A similar argument shows that $E_p$ also is the unique coherent expectation on $\mathbb{G}(\{\mathtt{H}, \mathtt{T}\})$ that extends $E_2$. ∲

We should emphasise that in Proposition 2.20, the coherent expectation $E^\star$ that extends the coherent expectation $E$ to the superset $\mathcal{K}$ of the domain $\mathcal{G}$ of $E$ need not be unique. Let us take a closer look at the set

$$\mathcal{M}_E^{\mathcal{K}} := \left\{E^\star \in \mathbb{E}_{\mathcal{K}} : (\forall g \in \mathcal{G})\; E^\star(g) = E(g)\right\} \tag{2.11}$$

of coherent expectations on $\mathcal{K}$ that extend $E$. Fix any coherent extension $E^\star$ of $E$ in $\mathcal{M}_E^{\mathcal{K}}$. Because $E^\star$ is a coherent expectation, we can interpret $E^\star(g)$

as a fair price for the gamble $g$ in $\mathcal{K}$. De Finetti (2017, Section 3.10) proves that this fair price is always greater than the subject's (inferred) supremum buying price for $g$ and lower than her (inferred) infimum selling price for $g$. Even more, he proves that these bounds are tight, and that any value in the corresponding interval is attained by at least one coherent extension $E^\star$ of $E$. Here, we repeat Walley's (1991, Corollary 3.4.3) statement of this result.

**Proposition 2.22.** *Consider a possibility space $\mathcal{X}$, a coherent expectation $E$ on a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$ and a superset $\mathcal{K}$ of $\mathcal{G}$. Then for all $g$ in $\mathcal{K}$,*

$$\{E^\star(g)\colon E^\star \in \mathcal{M}_E^{\mathcal{K}}\} = [\underline{\mathcal{E}}_E(g), \overline{\mathcal{E}}_E(g)],$$

*and therefore*

$$\underline{\mathcal{E}}_E(g) = \min\{E^\star(g)\colon E^\star \in \mathcal{M}_E^{\mathcal{K}}\} \quad \textit{and} \quad \overline{\mathcal{E}}_E(g) = \max\{E^\star(g)\colon E^\star \in \mathcal{M}_E^{\mathcal{K}}\}.$$

**Bruno's Example 2.23.** With Proposition 2.22, we can obtain the results in Bruno's Example 2.21 in a different manner. Here, we exclusively focus on $E_1$.

Our intention is to invoke Proposition 2.22, so we should determine $\underline{\mathcal{E}}_{E_1}$ and $\overline{\mathcal{E}}_{E_1}$. It is clear from Fig. 2.2(a)$_{21}$ – or equivalently, it follows from Eqs. (2.6)$_{20}$ and (2.5)$_{18}$ – that

$$\mathcal{E}(\mathcal{A}_{E_1}) = \left\{g \in \mathbb{G}(\{\mathtt{H},\mathtt{T}\})\colon \frac{g(\mathtt{H}) + g(\mathtt{T})}{2} > 0 \text{ or } g = 0\right\}.$$

For this reason, for all $g$ in $\mathbb{G}(\{\mathtt{H},\mathtt{T}\})$,

$$\begin{aligned}
\underline{\mathcal{E}}_{E_1}(g) &= \sup\{\alpha \in \mathbb{R}\colon (g-\alpha) \in \mathcal{E}(\mathcal{A}_{E_1})\} \\
&= \sup\left\{\alpha \in \mathbb{R}\colon \frac{[g-\alpha](\mathtt{H}) + [g-\alpha](\mathtt{T})}{2} > 0\right\} \cup \{\alpha \in \mathbb{R}\colon g-\alpha = 0\} \\
&= \sup\left\{\alpha \in \mathbb{R}\colon \frac{g(\mathtt{H}) + g(\mathtt{T})}{2} > \alpha\right\} \cup \{\alpha \in \mathbb{R}\colon g = \alpha\} \\
&= \frac{g(\mathtt{H}) + g(\mathtt{T})}{2} \\
&= E_p(g),
\end{aligned}$$

where, as in Bruno's Example 2.21, $p$ is the probability mass function on $\{\mathtt{H},\mathtt{T}\}$ that assigns mass $1/2$ to $\mathtt{H}$ and to $\mathtt{T}$. Similarly, we see that for all $g$ in $\mathbb{G}(\{\mathtt{H},\mathtt{T}\})$,

$$\overline{\mathcal{E}}_{E_1}(g) = \inf\left\{\beta \in \mathbb{R}\colon \frac{[\beta-g](\mathtt{H}) + [\beta-g](\mathtt{T})}{2} > 0\right\} = \frac{g(\mathtt{H}) + g(\mathtt{T})}{2} = E_p(g).$$

Hence, it follows from Proposition 2.22 that $E_p$ is the unique coherent expectation on $\mathbb{G}(\{\mathtt{H},\mathtt{T}\})$ that extends $E_1$. Of course, this confirms our conclusion in Bruno's Example 2.21. ♯

Let us again consider a coherent expectation $E$ on $\mathcal{G}$, and the (possibly larger) domain

$$\{g \in \mathbb{G}(\mathcal{X}) : \underline{\mathcal{E}}_E(g) = \overline{\mathcal{E}}_E(g)\}. \tag{2.12}$$

Then it follows immediately from Proposition 2.22 that there is a *unique* coherent expectation $E^\star$ that extends $E$ to this domain, and this extension is given by $E^\star(g) = \underline{\mathcal{E}}_E(g) = \overline{\mathcal{E}}_E(g)$ for all $g$ in $\mathcal{K}$. In other words, this answers our question whether it makes sense to extend the domain of $E$ to include those gambles $g$ for which the (inferred) supremum buying price $\underline{\mathcal{E}}_E(g)$ is equal to the (inferred) infimum selling price $\overline{\mathcal{E}}_E(g)$. We will come back to this particular extension in Section 5.1.1217 further on. There, we also use the following more convenient expression for the natural extension $\underline{\mathcal{E}}_E$ of $E$; it is due to Walley (1991, Theorem 3.1.4), but also occurs in (Troffaes et al., 2014, Theorem 4.34).

**Proposition 2.24.** *Consider a possibility space $\mathcal{X}$, a coherent expectation $E$ on a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$ and a superset $\mathcal{K}$ of $\mathcal{G}$. If $\mathcal{G}$ is a linear space that contains all constant gambles, then*

$$\underline{\mathcal{E}}_E(g) = \sup\{E(h) : h \in \mathcal{G}, h \leq g\} \quad \textit{for all } g \in \mathbb{G}(\mathcal{X})$$

*and, by conjugacy,*

$$\overline{\mathcal{E}}_E(g) = \inf\{E(h) : h \in \mathcal{G}, h \geq g\} \quad \textit{for all } g \in \mathbb{G}(\mathcal{X}).$$

Thus, if the domain $\mathcal{G}$ of the coherent expectation $E$ is a linear space that contains all constant gambles, then the natural extension corresponds to *inner and outer approximations.* That is, determining the natural extension $\underline{\mathcal{E}}_E(g)$ then corresponds to approximating $g$ from below with gambles in the domain $\mathcal{G}$ of $E$, or from above for $\overline{\mathcal{E}}_E(g)$. This is an important general concept in modelling uncertainty, and we will re-encounter it further down the line, more specifically in Section 5.1216.

### 2.2.4 Lower envelopes of expectations

Next, let us consider a (non-empty) *set* $\mathcal{M}$ of coherent expectations, all with the same domain $\mathcal{G}$. We have just seen one setting where this is natural, but there is a second one: the setting of a *sensitivity analysis.* In this setting, we regard the set $\mathcal{M}$ of coherent expectations as a partial specification of the 'true' coherent expectation $E$. More specifically, we take on the view that the subject believes there is some 'true' coherent expectation $E$, but her beliefs only permit her to say that it belongs to the *credal set* $\mathcal{M}$ – a term coined by Levi (1983). For a more extensive discussion of the ideas behind the sensitivity analysis setting, we refer to (Walley, 1991, Section 5.9).

As in Proposition 2.22, we are interested in the minimum and the maximum of the coherent expectations in the credal set $\mathcal{M}$. That is, we are interested in the *lower envelope*

$$\underline{E}_{\mathcal{M}}: \mathcal{G} \to \mathbb{R}: g \mapsto \underline{E}_{\mathcal{M}}(g) := \inf\{E(g): E \in \mathcal{M}\} \tag{2.13}$$

and the *upper envelope*

$$\overline{E}_{\mathcal{M}}: \mathcal{G} \to \mathbb{R}: g \mapsto \overline{E}_{\mathcal{M}}(g) := \sup\{E(g): E \in \mathcal{M}\}. \tag{2.14}$$

Observe that $\underline{E}_{\mathcal{M}}$ and $\overline{E}_{\mathcal{M}}$ are real-valued functions on $\mathcal{G}$ due to (E1)$_{22}$. Furthermore, it is obvious that $\underline{E}_{\mathcal{M}}$ and $\overline{E}_{\mathcal{M}}$ are conjugate, in the sense that for all $g$ in $\mathcal{G}$ such that $-g$ belongs to $\mathcal{G}$,

$$\overline{E}_{\mathcal{M}}(g) = \sup\{E(g): E \in \mathcal{M}\} = \sup\{-E(-g): E \in \mathcal{M}\} = -\inf\{E(-g): E \in \mathcal{M}\} = -\underline{E}_{\mathcal{M}}(-g),$$

where for the second equality we used (E2)$_{22}$. In most practical cases – including the ones in this dissertation – the domain $\mathcal{G}$ is negation invariant, in the sense that $\mathcal{G} = -\mathcal{G} := \{-g: g \in \mathcal{G}\}$. In that case, the upper envelope $\overline{E}_{\mathcal{M}}$ is completely determined by the lower envelope $\underline{E}_{\mathcal{M}}$ through the conjugacy relation, so we can focus on one of the two; in this dissertation, we typically focus on the lower envelope $\underline{E}_{\mathcal{M}}$.

For all $g$ in $\mathcal{G}$, we call $\underline{E}_{\mathcal{M}}(g)$ the *lower expectation of* $g$ and $\overline{E}_{\mathcal{M}}(g)$ the *upper expectation of* $g$. Moreover, for any event $A$ in $\mathcal{P}(\mathcal{X})$ such that its indicator $\mathbb{I}_A$ belongs to the domain $\mathcal{G}$, we call $\underline{P}_{\mathcal{M}}(A) := \underline{E}_{\mathcal{M}}(\mathbb{I}_A)$ the *lower probability of* $A$ and $\overline{P}_{\mathcal{M}}(A) := \overline{E}_{\mathcal{M}}(\mathbb{I}_A)$ the *upper probability of* $A$, respectively.

In general, we call every real-valued function on $\mathcal{G}$ that is a lower envelope of coherent expectations a *coherent lower expectation*, after Walley (1991, Theorem 3.3.3) (see also Troffaes et al., 2014, Proposition 4.20 and Theorem 4.38).

**Definition 2.25.** Consider a possibility space $\mathcal{X}$ and a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. A real-valued map $\underline{E}$ on $\mathcal{G}$ is called a *coherent lower expectation on* $\mathcal{G}$ if there is a non-empty set $\mathcal{M}$ of coherent expectations on $\mathcal{G}$ such that $\underline{E} = \underline{E}_{\mathcal{M}}$.

For any coherent lower expectation $\underline{E}$ on $\mathcal{G}$, the corresponding *conjugate upper expectation* $\overline{E}$ is the real-valued map on $-\mathcal{G}$ defined by

$$\overline{E}(g) := -\underline{E}(-g) \quad \text{for all } g \in -\mathcal{G}.$$

We have already encountered one example of a coherent lower expectation: by Proposition 2.22$_{26}$, the natural extension $\underline{\mathcal{E}}_E$ of a coherent expectation $E$ is a coherent lower expectation on $\mathbb{G}(\mathcal{X})$. For another example, we turn to our running example.

**Bruno's Example 2.26.** Bruno has had it with us restricting him to a single coin flip, so we now again allow him to make two consecutive coin flips.

Hence, we let $n = 2$ and $\mathcal{X} = \{\mathtt{H},\mathtt{T}\}^2 = \{(\mathtt{H},\mathtt{H}),(\mathtt{H},\mathtt{T}),(\mathtt{T},\mathtt{H}),(\mathtt{T},\mathtt{T})\}$. Following Walley (1991, Section 2.7.3), we will define two coherent lower expectations through a lower envelope.

First, we define three coherent expectations on $\mathbb{G}(\mathcal{X})$. These are defined through the probability mass functions $p_1$, $p_2$ and $p_3$ on $\mathcal{X} = \{\mathtt{H},\mathtt{T}\}^2$, whose values are slightly adapted from (Walley, 1991, Section 2.7.3) and are summarised in the following table:

| $x_{1:2}$ | $(\mathtt{H},\mathtt{H})$ | $(\mathtt{H},\mathtt{T})$ | $(\mathtt{T},\mathtt{H})$ | $(\mathtt{T},\mathtt{T})$ |
|---|---|---|---|---|
| $p_1(x_{1:2})$ | 3/7 | 1/7 | 1/7 | 2/7 |
| $p_2(x_{1:2})$ | 2/7 | 3/7 | 1/7 | 1/7 |
| $p_3(x_{1:2})$ | 3/7 | 2/7 | 1/7 | 1/7 |

For any $k$ in $\{1,2,3\}$, we let $E_{p_k}$ be the coherent expectation associated with $p_k$ according to Eq. (2.8)23. Let $\mathcal{M} := \{E_{p_1}, E_{p_2}\}$ and $\mathcal{M}' := \{E_{p_1}, E_{p_2}, E_{p_3}\}$, and denote their lower envelopes by $\underline{E} := \underline{E}_{\mathcal{M}}$ and $\underline{E}' := \underline{E}_{\mathcal{M}'}$. By definition, both $\underline{E}$ and $\underline{E}'$ are coherent lower expectations.

We let $g$ denote the gamble on $\mathcal{X} = \{\mathtt{H},\mathtt{T}\}^2$ that represents the number of tails in the two coin tosses; so $g(\mathtt{H},\mathtt{H}) = 0$, $g(\mathtt{H},\mathtt{T}) = 1 = g(\mathtt{T},\mathtt{H})$ and $g(\mathtt{T},\mathtt{T}) = 2$. Then

$$\underline{E}(g) = \min\{6/7, 6/7\} = 6/7 \quad \text{and} \quad \underline{E}'(g) = \min\{6/7, 6/7, 5/7\} = 5/7.$$

One crucial – but perhaps a bit trivial – observation is that different sets of non-empty coherent expectations can have the same lower envelope $\underline{E}$. Our running example elucidates this.

**Bruno's Example 2.27.** We return to the setting of Bruno's Example 2.26↶. Additionally, we consider the set

$$\mathcal{M}^\star := \{E_p : \alpha \in [0,1], p = \alpha p_1 + (1-\alpha) p_2\} = \{\alpha E_{p_1} + (1-\alpha) E_{p_2} : \alpha \in [0,1]\}$$

of linear expectations on $\mathbb{G}(\mathcal{X})$. It is not difficult to see that $\underline{E}_{\mathcal{M}} = \underline{E}_{\mathcal{M}^\star}$, even though $\mathcal{M} \neq \mathcal{M}^\star$.

So we have seen that for a given coherent lower expectation $\underline{E}$ on $\mathcal{G}$, there may be several sets of coherent expectations that have $\underline{E}$ as lower envelope; all these sets have one thing in common, though: every coherent expectation $E$ in such a set must dominate $\underline{E}$, in the sense that for all $g$ in $\mathcal{G}$, $E(g) \geq \underline{E}(g)$. For this reason, for any real-valued map $\underline{E}$ on $\mathcal{G}$, it makes sense to consider the set

$$\mathcal{M}_{\underline{E}} := \{E \in \mathbb{E}_{\mathcal{G}} : (\forall g \in \mathcal{G})\ E(g) \geq \underline{E}(g)\} \tag{2.15}$$

of all coherent expectations on $\mathcal{G}$ that dominate $\underline{E}$. Clearly, if $\underline{E}$ is a coherent lower expectation, then $\mathcal{M}_{\underline{E}}$ is the largest set of coherent expectations on $\mathcal{G}$ that has $\underline{E}$ as lower envelope. In fact, one can prove that for this particular credal set, the infimum in Eq. $(2.13)_{28}$ is a minimum. For a proof of this result, we refer to (Walley, 1991, Theorems 3.3.3 and 3.4.1) or (Troffaes et al., 2014, Proposition 4.20 and Theorem 4.38).

**Theorem 2.28.** *Consider a possibility space $\mathcal{X}$ and a real-valued map $\underline{E}$ on a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. Then $\underline{E}$ is a coherent lower expectation if and only if the corresponding set $\mathcal{M}_{\underline{E}}$ of dominating coherent expectations is non-empty and*

$$\underline{E}(g) = \min\{E(g) \colon E \in \mathcal{M}_{\underline{E}}\} \quad \textit{for all } g \in \mathcal{G}.$$

### Properties of coherent lower expectations

Because coherent lower expectations are lower envelopes of coherent expectations, they have properties that mirror those of coherent expectations listed in Proposition $2.14_{22}$. For a proof and a much more exhaustive list of properties, we refer to (Troffaes et al., 2014, Theorem 4.13) or (Walley, 1991, Section 2.6.1).

**Proposition 2.29.** *Consider a possibility space $\mathcal{X}$ and a coherent lower expectation $\underline{E}$ on a subset $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. Then*

LE1. $\underline{E}(g) \geq \inf g$ *for all* $g \in \mathcal{G}$*;*

LE2. $\underline{E}(\lambda g) = \lambda \underline{E}(g)$ *for all* $\lambda \in \mathbb{R}_{\geq 0}$ *and all* $g \in \mathcal{G}$ *such that* $\lambda g \in \mathcal{G}$*;*

LE3. $\underline{E}(g+h) \geq \underline{E}(g) + \underline{E}(h)$ *for all* $g, h \in \mathcal{G}$ *such that* $g+h \in \mathcal{G}$.

*Furthermore, if we let $g$ and $h$ be gambles in $\mathcal{G}$ and $\mu$ a real number, then the following properties hold whenever every term is well-defined:*

LE4. $\inf g \leq \underline{E}(g) \leq \overline{E}(g) \leq \sup g$*;*

LE5. $\underline{E}(g+\mu) = \underline{E}(g) + \mu$*;*

LE6. $\underline{E}(g) \leq \underline{E}(h)$ *whenever* $g \leq h$*;*

LE7. $|\underline{E}(g) - \underline{E}(h)| \leq \overline{E}(|g-h|)$.

The properties (LE1)–(LE3) are known as the *coherence properties*. The reason for this is that like coherent expectations, coherent lower expectations have an intuitive betting interpretation. We do not need this interpretation in the remainder, so we will not go into the nitty gritty. It suffices to understand that like the natural extension $\underline{\mathcal{E}}_{E}$, a general coherent lower expectation $\underline{E}$ can be interpreted as a subject's supremum acceptable buying price. Under this interpretation, the coherence properties (LE1)–(LE3) are more or less immediate consequences of the rationality requirements $(A3)_{17}$–$(A4)_{17}$. The interested reader can find the details regarding this betting interpretation in (Walley, 1991, Chapter 2) or (Troffaes et al., 2014, Chapter 4).

Because the three coherence properties (LE1)–(LE3) can be more easily checked than the condition in Definition $2.25_{28}$, one might wonder if they are also sufficient for coherence. As was the case for coherent expectations, this is *not* the case in general. However, the following result establishes that the coherence properties (LE1)$_{\curvearrowleft}$–(LE3)$_{\curvearrowleft}$ suffice in case the domain $\mathcal{G}$ is a linear space. For a proof, we refer to (Walley, 1991, Theorems 2.5.5 and 3.3.3) or (Troffaes et al., 2014, Theorem 4.15 and Propostion 4.20); note that it actually suffices to impose (LE2)$_{\curvearrowleft}$ for positive real numbers only.

**Proposition 2.30.** *Consider a possibility space $\mathcal{X}$ and a linear subspace $\mathcal{G}$ of $\mathbb{G}(\mathcal{X})$. A real-valued map $\underline{E}$ on $\mathcal{G}$ is a coherent lower expectation if and only if it satisfies (LE1)$_{\curvearrowleft}$–(LE3)$_{\curvearrowleft}$.*

## 2.3 Modelling uncertainty using a probability charge

Probability mass functions are not the only link between coherent expectations – and, through lower envelopes, coherent lower expectations – and the classical approach towards modelling uncertainty. In this section, we will discover a more general way in which this classical approach is intertwined with the theory of coherence. We continue to use our running example to illustrate the necessary concepts, but this time around we move from the finite possibility space corresponding to $n$ consecutive coin flips to a more involved possibility space.

**Bruno's Example 2.31.** Bruno is so obsessed with flipping his coin that he hardly gets any work done. Some time after Bruno's life-changing discovery, Alicia decides that she cannot let Bruno wither away because of his strange obsession. She realises that the only way to get Bruno back to his research, is to get some kind of machine to do the coin flipping for him. Knowing that Bruno is an avid tinkerer, she cunningly plants into his mind the idea to build this magical coin flipping contraption. Many sleepless nights and frustrating days later, Bruno finishes his most incredible invention to date. He has realised a magnificent feat of engineering: a machine that can flip his precious golden coin until eternity.

From here on, we take the experiment to be the following: Bruno switches on his coin flipping machine, and this machine registers the results of the consecutive coin flips. Because the incredible machine keeps flipping Bruno's coin until eternity, an (idealised) outcome of this experiment is an infinite sequence of heads and tails. Therefore, the possibility space $\mathcal{X}$ is made up of all infinite sequences $x = (x_n)_{n\in\mathbb{N}}$ in $\{\mathtt{H},\mathtt{T}\}$, with $x_n$ the outcome of the $n$-th coin toss. In line with this, we denote the uncertain outcome of our experiment by $X = (X_n)_{n\in\mathbb{N}}$, where $X_n$ is the uncertain outcome of the $n$-th coin toss. ∅

### 2.3.1 Fields of events

Probability charges are the fundamental tool of the classical approach towards probability theory. They essentially model uncertainty by assigning a probability, a real number between 0 and 1, to all events in some set of events. Probability charges have to satisfy three essential properties – the laws of probability – and we can only impose these properties if the set of events has sufficient structure. The minimal structure for a set of events is that of a field, which will return time and time again in this work. A field of events is a Boolean algebra[5] with conjunction $\cap$, disjunction $\cup$ and complementation (or negation) $\bullet^{\mathrm{c}}$.

**Definition 2.32.** Consider a possibility space $\mathcal{X}$. A *field of events* $\mathcal{F}$ over $\mathcal{X}$ is a set of events – that is, a subset of $\mathcal{P}(\mathcal{X})$ – such that

F1. $\mathcal{X} \in \mathcal{F}$;

F2. $A^{\mathrm{c}} \in \mathcal{F}$ for any event $A \in \mathcal{F}$;

F3. $A \cup B \in \mathcal{F}$ for all events $A, B \in \mathcal{F}$.

Note that a field of events $\mathcal{F}$ is also closed under intersection. Indeed, for any two events $A$ and $B$ in $\mathcal{F}$, it follows from De Morgan's laws that $A \cap B = (A^{\mathrm{c}} \cup B^{\mathrm{c}})^{\mathrm{c}}$; now $A^{\mathrm{c}}$ and $B^{\mathrm{c}}$ belong to $\mathcal{F}$ by (F2), therefore $A^{\mathrm{c}} \cup B^{\mathrm{c}}$ belongs to $\mathcal{F}$ by (F3) and therefore $(A^{\mathrm{c}} \cup B^{\mathrm{c}})^{\mathrm{c}} = A \cap B$ belongs to $\mathcal{F}$ by (F2). Additionally, it is easy to show using induction that a field of events $\mathcal{F}$ is also closed under finite unions and intersections as well.

There are three straightforward examples of families of events that are fields: (i) the smallest one $\{\emptyset, \mathcal{X}\}$ that only consists of the impossible and the sure event; (ii) a trivial one $\{\emptyset, A, A^{\mathrm{c}}, \mathcal{X}\}$ that consists of the impossible and the sure event and an event $A$ in $\mathcal{P}(\mathcal{X})$ and its complement $A^{\mathrm{c}}$; and (iii) the greatest one $\mathcal{P}(\mathcal{X})$ that consists of all events. For a more involved example, we return to our running example.

**Bruno's Example 2.33.** A non-trivial field of events over our possibility space $\mathcal{X}$ is the set of those events that concern the result of a finite number of coin flips. To formally construct this set, we need to introduce some notation. For any natural number $n$ and any $n$-tuple $y_{1:n} = (y_1, \ldots, y_n)$ in $\{\mathtt{H}, \mathtt{T}\}^n$, we denote the event 'the first $n$ flips are $y_1, \ldots, y_n$' as

$$\{X_{1:n} = y_{1:n}\} := \{x \in \mathcal{X} \colon x_{1:n} = y_{1:n}\},$$

where for any $x = (x_k)_{k\in\mathbb{N}}$ in $\mathcal{X}$, we let $x_{1:n}$ denote $(x_1, \ldots, x_n) = (x_n)_{n\in\{1,\ldots,n\}}$. Similarly, for any subset $A$ of $\{\mathtt{H}, \mathtt{T}\}^n$, we define the event

$$\{X_{1:n} \in A\} := \{x \in \mathcal{X} \colon x_{1:n} \in A\} = \bigcup_{y_{1:n}\in A} \{X_{1:n} = y_{1:n}\},$$

where we follow the convention that an empty union is equal to the empty set $\emptyset$.

[5] Despite its name, a field of events is actually not a field in the strict mathematical sense, because it lacks additive and multiplicative inverses (see D. Williams, 1991, p. 16). Be that as it may, we will frequently use 'a field $\mathcal{F}$' as a shorthand for 'a field of events $\mathcal{F}$'.

It is now relatively straightforward to verify – see for instance (Billingsley, 1995, pp. 27 and 28) or (Lopatatzidis, 2017, Lemma 6) – that the set

$$\mathcal{F} \coloneqq \left\{\{X_{1:n} \in A\}\colon\ n \in \mathbb{N}, A \subseteq \{\mathtt{H},\mathtt{T}\}^n\right\}$$

is a field of events.

To illustrate what kind of events belongs to $\mathcal{F}$, we extend the events as defined in Bruno's Example 2.4$_{13}$ to our idealised setting. The event that the $n$-th coin flip of the machine is heads is now

$$\{X_n = \mathtt{H}\} \coloneqq \{x \in \mathcal{X}\colon\ x_n = \mathtt{H}\}$$

and the event that the first $n$ flips of the machine are all heads is

$$H_n \coloneqq \left\{x \in \mathcal{X}\colon\ \big(\forall k \in \{1,\dots,n\}\big)\ x_k = \mathtt{H}\right\}.$$

These two events only depend on the result of a finite number of coin flips, so they should belong to the field $\mathcal{F}$. To see that this is the case, we observe that $\{X_n = \mathtt{H}\} = \{X_{1:n} \in A^n_{1:n}\}$, with

$$A^n_{1:n} = \left\{y_{1:n} \in \{\mathtt{H},\mathtt{T}\}^n\colon\ y_n = \mathtt{H}\right\}.$$

To see that $H_n$ belongs to the field $\mathcal{F}$ as well, we observe that

$$H_n = \bigcap_{k=1}^{n} \{X_k = \mathtt{H}\}.$$

Because the field $\mathcal{F}$ is closed under taking finite intersections, we infer from this that $H_n$ also belongs to $\mathcal{F}$. ∲

It is not always desirable or feasible to immediately and explicitly specify a field of events. Fortunately, this is not necessary, as *any* set of events $\mathcal{F}$ over some possibility space $\mathcal{X}$ can be enlarged to a field, in the sense that there are fields of events that include $\mathcal{F}$. For example, the set of all events $\mathcal{P}(\mathcal{X})$ is a field and trivially includes the set $\mathcal{F}$. Even more, there always is a smallest field of events that includes the set $\mathcal{F}$; the following constructive result can be found in (Bhaskara Rao et al., 1983, Theorem 1.1.11).

**Lemma 2.34.** *Consider any set of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. We close the set under complements to yield*

$$\mathcal{F}_1 \coloneqq \{\emptyset,\mathcal{X}\} \cup \mathcal{F} \cup \{A^{\mathrm{c}}\colon\ A \in \mathcal{F}\},$$

*subsequently close the set under finite intersections to yield*

$$\mathcal{F}_2 \coloneqq \left\{\bigcap_{k=1}^{n} A_k\colon\ n \in \mathbb{N}, A_k \in \mathcal{F}_1\right\},$$

*and finally close the set under finite unions of disjoint events to yield*

$$\langle \mathcal{F} \rangle := \left\{ \bigcup_{k=1}^{n} A_k \colon n \in \mathbb{N}, A_k \in \mathcal{F}_2, \big(\forall \ell \in \{1,\dots,n\}, k \neq \ell\big)\ A_k \cap A_\ell = \emptyset \right\}.$$

*Then $\langle \mathcal{F} \rangle$ is the smallest field of events over $\mathcal{X}$ containing $\mathcal{F}$, in the sense that any other field $\mathcal{F}'$ that includes $\mathcal{F}$ also includes $\langle \mathcal{F} \rangle$.*

**Bruno's Example 2.35.** Recall from Bruno's Example 2.33$_{32}$ that for any natural number $n$ and any $y_{1:n} = (y_1, \dots, y_n)$ in $\{\mathtt{H},\mathtt{T}\}^n$, we have defined the event

$$\{X_{1:n} = y_{1:n}\} = \{x \in \mathcal{X} \colon x_{1:n} = y_{1:n}\}.$$

The set of events

$$\mathcal{F}' := \big\{\{X_{1:n} = y_{1:n}\} \colon n \in \mathbb{N}, y_{1:n} \in \{\mathtt{H},\mathtt{T}\}^n\big\}$$

is not a field. However, Lopatatzidis (2017, Lemma 6) shows that the smallest field of events that includes it is $\mathcal{F}$; that is, $\langle \mathcal{F}' \rangle = \mathcal{F}$. ∲

### 2.3.2 Probability charges

A probability charge is a normalised, non-negative and finitely additive function on a field of events. Some authors choose to adopt the term finitely additive probability measure; we prefer to reserve the term probability measure for countably additive probability charges on a σ-field of events – see Appendix C$_{461}$. Formally, we adhere to the following definition.

**Definition 2.36.** Consider a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. A *probability charge* $P$ (on $\mathcal{F}$) is a real-valued map on $\mathcal{F}$ that satisfies the following three properties:

P1. $P(\mathcal{X}) = 1$;

P2. $P(A) \geq 0$ for all events $A$ in $\mathcal{F}$;

P3. $P(A \cup B) = P(A) + P(B)$ for any two disjoint events $A$ and $B$ in $\mathcal{F}$.

Note that conditions (F1) and (F3) of Definition 2.32$_{32}$ ensure that $\mathcal{X}$ and $A \cup B$ belong to the domain of $P$.

In case the possibility space $\mathcal{X}$ is finite, there is a straightforward one-to-one correspondence between probability mass functions and probability charges. With any probability mass function $p$ on $\mathcal{X}$, we can associate the map

$$P_p \colon \mathcal{P}(\mathcal{X}) \to \mathbb{R} \colon A \mapsto P_p(A) := \sum_{x \in A} p(x),$$

where we follow the convention that the empty sum is equal to zero. It then follows almost immediately from Definitions 2.16$_{23}$ and 2.36 that $P_p$ is a probability charge. On the other hand, given a probability charge $P$ on $\mathcal{P}(\mathcal{X})$, the map

$$p_P\colon \mathcal{X} \to \mathbb{R}\colon x \mapsto p_P(x) := P(\{x\})$$

is a probability mass function such that $P(A) = \sum_{x \in A} p_P(x)$. For a more involved example of a probability charge, we return to our running example.

**Bruno's Example 2.37.** The uncertainty regarding a single coin toss is usually modelled as 'the probability of heads is $q$', where $q$ is equal to $1/2$ for a fair coin but can be any real number in the unit interval $[0,1]$ in general. This information is captured in the probability mass function

$$p\colon \{\mathtt{H},\mathtt{T}\} \to [0,1]\colon x \mapsto p(x) := \begin{cases} q & \text{if } x = \mathtt{H}, \\ 1-q & \text{if } x = \mathtt{T}. \end{cases}$$

We can use this mass function $p$ to define a probability charge $P$ on the field $\mathcal{F}$ of events that depend on the results of a finite number of coin flips by the machine. First, we define the probability of the events in $\mathcal{F}$ of the form $\{X_{1:n} = y_{1:n}\}$ as

$$P(X_{1:n} = y_{1:n}) := \prod_{k=1}^{n} p(y_k), \tag{2.16}$$

where – in classical parlance – we have assumed that the consecutive coin flips are independent and identically distributed. Note that we write $P(X_{1:n} = y_{1:n})$ instead of $P(\{X_{1:n} = y_{1:n}\})$ in an effort not to unnecessarily complicate our notation; we will implicitly carry through the same notational convention in other cases as well. Subsequently, for any general event $\{X_{1:n} \in A\}$ in $\mathcal{F}$, with $A$ a subset of $\{\mathtt{H},\mathtt{T}\}^n$, we set

$$P(X_{1:n} \in A) := \sum_{y_{1:n} \in A} P(X_{1:n} = y_{1:n}) = \sum_{y_{1:n} \in A} \prod_{k=1}^{n} q(y_k), \tag{2.17}$$

where we again follow the convention that an empty sum is equal to 0. At first sight, this definition depends on the representation of the event $\{X_{1:n} \in A\}$, but one can check that $P(X_{1:m} \in B) = P(X_{1:n} \in A)$ for every other representation $\{X_{1:m} \in B\}$ – with $B$ a subset of $\{\mathtt{H},\mathtt{T}\}^m$ – of the event $\{X_{1:n} \in A\}$ (see Billingsley, 1995, pp. 28 and 29). Now that we know that $P$ is properly defined, we should verify that the map $P$ satisfies (P1)–(P3); we will not explicitly do this here, but refer to (Billingsley, 1995, pp. 28 and 29) for a proof. ∯

### 2.3.3 Expectations of simple variables

To see how classical probability charges fit into the coherence-centred approach, we need to establish how they naturally induce expectation operators. For the time being, we restrict ourselves to the expectation of simple variables; we will consider the expectation of more general variables in Section 5.1216 further on (see also Troffaes et al., 2014, Chapter 8).

### Simple variables

Basically, a simple variable is a linear combination of indicators; we adhere to the formal definition of Troffaes et al. (2014, Definition 1.16), similar definitions are used by D. Williams (1991, Section 5.1) and Shiryaev (2016, Chapter 2, Section 6.1).

**Definition 2.38.** Consider a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. An *$\mathcal{F}$-simple variable $f$* is a real variable that has a representation of the form

$$f = \sum_{k=1}^{n} a_k \mathbb{I}_{A_k}, \tag{2.18}$$

where $n$ is a natural number, $a_1, \ldots, a_n$ are real numbers and $A_1, \ldots, A_n$ are events in $\mathcal{F}$. We collect all $\mathcal{F}$-simple real variables in $\mathbb{S}(\mathcal{F})$.

It follows immediately from Definition 2.38 – more specifically, from Eq. (2.18) – that an $\mathcal{F}$-simple variable $f$ has a finite range. Hence, an $\mathcal{F}$-simple variable $f$ is bounded – and therefore a gamble – and it furthermore holds that $\sup f = \max f$ and $\inf f = \min f$. It is moreover easy to verify that the $\mathcal{F}$-simple real variables constitute a real vector space – they are closed under scalar multiplication and point-wise addition – that includes the constant gambles.

**Lemma 2.39.** *Consider a field of events $\mathcal{F}$ on a possibility space $\mathcal{X}$. Then $\mathbb{S}(\mathcal{F})$ is a real vector space that includes all constant gambles, because*

(i) $\mu f \in \mathbb{S}(\mathcal{F})$ *for all* $\mu \in \mathbb{R}$ *and* $f \in \mathbb{S}(\mathcal{F})$*;*

(ii) $f + g \in \mathbb{S}(\mathcal{F})$ *for all* $f, g \in \mathbb{S}(\mathcal{F})$*; and*

(iii) $\mu \in \mathbb{S}(\mathcal{F})$ *for all* $\mu \in \mathbb{R}$.

**Bruno's Example 2.40.** Like in Bruno's Example 2.5[15], we consider the number of times Bruno's machine will flip heads in its first $n$ flips, with $n$ a natural number. In our formalism, this corresponds to the (bounded) real variable

$$h_n := \sum_{k=1}^{n} \mathbb{I}_{\{X_k = \mathtt{H}\}}.$$

Recall from Bruno's Example 2.33[32] that all the events $\{X_k = \mathtt{H}\}$ belong to the field $\mathcal{F}$, so we conclude that $h_n$ is an $\mathcal{F}$-simple variable. ∯

### Dunford integration

Let $P$ be a probability charge over a field $\mathcal{F}$ of events over some possibility space $\mathcal{X}$. Then for any $\mathcal{F}$-simple variable $f$, we define its expectation with respect to $P$ as its Dunford integral

$$E_P(f) := \sum_{k=1}^{n} a_k P(A_k), \tag{2.19}$$

where $\sum_{k=1}^{n} a_k \mathbb{I}_{A_k}$ is any representation of the $\mathcal{F}$-simple variable $f$, as in Eq. (2.18)↶. This is a proper definition because the value of $E_P(f)$ is independent of the representation of $f$; see (Troffaes et al., 2014, Definition 8.13) for a proof.

**Bruno's Example 2.41.** Recall that $h_n$, the number of heads in the first $n$ flips, is an $\mathcal{F}$-simple random variable with representation

$$h_n = \sum_{k=1}^{n} \mathbb{I}_{\{X_k=\mathtt{H}\}} = \sum_{k=1}^{n} \mathbb{I}_{\{X_{1:k}\in A_{1:k}^k\}},$$

with $A_{1:k}^k$ as defined in Bruno's Example 2.33$_{32}$. Consequently, the expected number of heads in the first $n$ flips is

$$\begin{aligned} E_P(h_n) &= \sum_{k=1}^{n} P(X_k = \mathtt{H}) = \sum_{k=1}^{n} P(X_{1:k} \in A_{1:k}^k) = \sum_{k=1}^{n} \sum_{y_{1:k}\in A_{1:k}^k} \prod_{\ell=1}^{k} p(y_\ell) \\ &= \sum_{k=1}^{n} q = nq, \end{aligned}$$

where for the third equality we have used Eq. (2.17)$_{35}$ and the penultimate equality follows from the definition of $A_{1:k}^k$ and some straightforward calculations.

The operator $E_P$ on $\mathbb{S}(\mathcal{F})$ has some well-known properties. Because these properties all follow more or less immediately from Eq. (2.19)↶, we have chosen to omit a formal proof.

**Proposition 2.42.** *Consider a probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. For all $\mathcal{F}$-simple variables $f$ and $g$ and any real number $\mu$,*

ES1. $\min f \le E_P(f) \le \max f$;

ES2. $E_P(\mu f) = \mu E_P(f)$;

ES3. $E_P(f+g) = E_P(f) + E_P(g)$.

*Furthermore,*

ES4. $E_P(f) \le E_P(g)$ *whenever* $f \le g$.

### Link with coherence

Recall that we have already encountered the three properties (ES1)–(ES3) of Proposition 2.42 before, in Section 2.2.2$_{19}$ to be precise. Because $\mathbb{S}(\mathcal{F})$ is a real vector space by Lemma 2.39↶, it follows from Proposition 2.15$_{22}$ that the expectation $E_P$ with respect to the probability charge $P$ is coherent. We have thus established the following result, which also occurs in (Troffaes et al., 2014, Lemma 8.14).

**Proposition 2.43.** *Consider a probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then $E_P$, the expectation with respect to the probability charge $P$ as defined in Eq.* (2.19)*36*, *is a coherent expectation on* $\mathbb{S}(\mathcal{F})$.

Even stronger, Troffaes et al. (2014, Theorem 8.22) prove that given a field $\mathcal{F}$ of events, there is a one-to-one correspondence between coherent expectations on the set of all '$\mathcal{F}$-measurable variables' – which includes the $\mathcal{F}$-simple variables (see Troffaes et al., 2014, Definition 1.17 and Proposition 1.18) – and probability charges on $\mathcal{F}$; more precisely, they show that any coherent expectation is essentially determined by the values it assumes on the indicators. Here, we limit ourselves to the following weaker result, which essentially generalises Proposition 2.18₂₃.

**Proposition 2.44.** *Consider a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$, and a coherent expectation $E$ on $\mathbb{S}(\mathcal{F})$. Then*

$$P_E\colon \mathcal{F} \to [0,1]\colon A \mapsto P_E(A) := E(\mathbb{I}_A)$$

*is a probability charge on $\mathcal{F}$. Moreover, $E = E_{P_E}$, where $E_{P_E}$ is the Dunford integral with respect to the probability charge $P_E$ on $\mathcal{F}$.*

Whereas coherent expectations are essentially determined by the values that they assume on the indicators, this is *not* the case for coherent lower expectations. This is an important and well-known difference between coherent expectations and coherent lower expectations, which is excellently argued by Walley (1991, Section 2.7.3) and Troffaes et al. (2014, Section 5.2). For the sake of completeness, we also give an example of two coherent lower expectations that agree on all indicators, but that do not agree on all gambles.

**Bruno's Example 2.45.** We briefly return to the setting of Bruno's Example 2.26₂₈. Observe that for any event $A$ in $\mathcal{P}(\mathcal{X})$, $\underline{E}(\mathbb{I}_A) = \underline{E}'(\mathbb{I}_A)$ and $\overline{E}(\mathbb{I}_A) = \overline{E}'(\mathbb{I}_A)$. In other words, both $\underline{E}$ and $\underline{E}'$ yield the same lower and upper probabilities. However, as we have already seen in Bruno's Example 2.26₂₈, $\underline{E}$ and $\underline{E}'$ do not agree on the gamble $g$ that represents the number of tails in the two coin tosses. ∲

### 2.3.4 Conditioning

We end our introduction to probability charges with a brief look at how to do conditioning with them. To this end, we fix a possibility space $\mathcal{X}$, a field of events $\mathcal{F}$ over $\mathcal{X}$ and a probability charge $P$ on $\mathcal{F}$. Furthermore, we fix some *conditioning event* $C$ in $\mathcal{F}$, with $P(C) > 0$. For any event $A$, we define the conditional probability of $A$ conditional on – or, alternatively, under the condition – $C$ as

$$P(A\,|\,C) := \frac{P(A \cap C)}{P(C)}.$$

The equality above is know as *Bayes's rule*; the rationale behind this rule is to only take into account those outcomes in the event $A$ that agree with the condition $C$, and then renormalise.

It is easy to see that the resulting real-valued map

$$P(\bullet\,|\,C)\colon \mathcal{F} \to \mathbb{R}\colon A \mapsto P(A\,|\,C) \tag{2.20}$$

is again a probability charge on $\mathcal{F}$. For this reason, we call $P(\bullet\,|\,C)$ a *conditional probability charge*. Because $P(\bullet\,|\,C)$ is a probability charge, it induces the *conditional expectation* operator

$$E_P(\bullet\,|\,C)\colon \mathbb{S}(\mathcal{F}) \to \mathbb{R}\colon f \mapsto E_P(f\,|\,C) := E_{P(\bullet|C)}(f), \tag{2.21}$$

where $E_{P(\bullet|C)}$ is the expectation operator with respect to the probability charge $P(\bullet\,|\,C)$ according to Eq. (2.19)$_{36}$. It is easy to verify that, for all $\mathcal{F}$-simple variables $f$,

$$E_P(f\,|\,C) = \frac{E_P(f\mathbb{I}_C)}{P(C)}. \tag{2.22}$$

Conditional expectations are especially convenient when we consider conditioning events that form a partition of the state space. To see why, we let $\mathcal{C} := \{C_1, \dots, C_n\}$ be a partition of $\mathcal{X}$ in events in $\mathcal{F}$ with $P(C_k) > 0$ for all $k$ in $\{1, \dots, n\}$. Observe that

$$E_P(f\,|\,\mathcal{C}) := \sum_{k=1}^{n} E_P(f\,|\,C_k)\mathbb{I}_{C_k}$$

is an $\mathcal{F}$-simple variable. Consequently, it follows from Eqs. (2.19)$_{36}$ and (2.22) and (ES3)$_{37}$ that

$$E_P(E_P(f\,|\,\mathcal{C})) = \sum_{k=1}^{n} E_P(f\,|\,C_k)P(C_k) = \sum_{k=1}^{n} E_P(f\mathbb{I}_{C_k}) = E_P(f). \tag{2.23}$$

This is a special case of an important law known as the *law of iterated expectations*, sometimes also called the law of total expectation. For now, it suffices to know that this law is a convenient tool when computing expectations; we will discuss this law in more detail in Chapter 3$_{53}$ further on.

**Bruno's Example 2.46.** Recall from Bruno's Example 2.37$_{35}$ that $P(X_1 = \mathtt{H}) = q$ and $P(X_1 = \mathtt{T}) = 1 - q$. Hence, whenever $0 < q < 1$, both the events $\{X_1 = \mathtt{H}\}$ and $\{X_1 = \mathtt{T}\}$ have non-zero probability. We now assume that this is the case, and therefore $P(\bullet\,|\,X_1 = \mathtt{H})$ and $P(\bullet\,|\,X_1 = \mathtt{T})$ are well-defined. After some easy calculations, we see that

$$P(X_1 = \mathtt{H}\,|\,X_1 = \mathtt{H}) = 1, \qquad P(X_2 = \mathtt{H}\,|\,X_1 = \mathtt{H}) = q,$$
$$P(X_1 = \mathtt{H}\,|\,X_1 = \mathtt{T}) = 0, \qquad P(X_2 = \mathtt{H}\,|\,X_1 = \mathtt{T}) = q.$$

It follows from this and Eq. (2.21)↶ that

$$E_P(h_2 \mid X_1 = \mathtt{H}) = E_{P(\bullet \mid X_1=\mathtt{H})}(h_2) = P(X_1 = \mathtt{H} \mid X_1 = \mathtt{H}) + P(X_2 = \mathtt{H} \mid X_1 = \mathtt{H})$$
$$= 1 + q,$$

where we have used that $h_2 = \mathbb{I}_{\{X_1=\mathtt{H}\}} + \mathbb{I}_{\{X_2=\mathtt{H}\}}$. Similarly,

$$E_P(h_2 \mid X_1 = \mathtt{T}) = P(X_1 = \mathtt{H} \mid X_1 = \mathtt{T}) + P(X_2 = \mathtt{H} \mid X_1 = \mathtt{T}) = q.$$

Observe that $\{X_1 = \mathtt{H}\}$ and $\{X_1 = \mathtt{T}\}$ form a partition of the possibility space $\mathcal{X}$. Therefore, we may use the law of iterated expectations to find that

$$E_P(h_2) = E_P(h_2 \mid X_1 = \mathtt{H})P(X_1 = \mathtt{H}) + E_P(h_2 \mid X_1 = \mathtt{T})P(X_1 = \mathtt{T}) = 2q,$$

which agrees with what we previously obtained in Bruno's Example 2.41₃₇. ∲

One clear drawback of using a probability charge to model uncertainty is that the conditional probability (or expectation) is undefined whenever the condition $C$ has probability zero. There is a way to mitigate this – through the Radon-Nikodym theorem, see for example (Billingsley, 1995, Section 33), (Fristedt et al., 1997, Chapter 21) or (Shiryaev, 2016, Section 7) – but this alternative approach essentially suffers from the same downside, because the resulting conditional probabilities (and expectations) are not uniquely defined. Additionally, it sometimes makes more sense to construct a probability charge starting from uniquely specified conditional probabilities – even for conditioning events that turn out to have probability zero – instead of the other way around.

## 2.4 Coherence and conditioning

An elegant way to deal with conditioning on events that have probability zero, is to use full conditional probabilities. It is for this reason that we introduce them, and the related concept of coherent conditional probabilities, in Section 2.4.1. We subsequently explain in Section 2.4.2₄₆ how coherent conditional probabilities induce conditional expectations. Next, we rid ourselves of the requirement of precision in Section 2.4.3₄₉, where we will allow for imprecision by considering sets of coherent conditional probabilities and taking lower envelopes of the associated conditional expectations, inspired by Section 2.2.4₂₇. We end this chapter in Section 2.4.4₅₁, where we will see how determinable gambles and simple variables are connected.

### 2.4.1 Coherent conditional probabilities

In the classical framework, conditional probabilities are only unequivocally defined for conditioning events that do not have zero probability. Dubins (1975), Regazzini (1985) and Berti et al. (2002) – and many others – get rid of this ambiguity by taking conditional probabilities as elementary instead of as being derived through Bayes's rule. More concretely, Dubins (1975, Section 3) and (Regazzini, 1985, Definition 2) generalise (conditional) probability charges to conditional probabilities as follows; in this definition, we use $\mathcal{C}_{\supset\emptyset} := \mathcal{C} \setminus \{\emptyset\}$ to exclude the empty conditioning event from the field $\mathcal{C}$ of conditioning events.[6]

**Definition 2.47.** Consider a possibility space $\mathcal{X}$, and two fields of events $\mathcal{F}$ and $\mathcal{C}$ over $\mathcal{X}$ such that $\mathcal{F}$ includes $\mathcal{C}$. A *full conditional probability $P$ on $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$* is a real-valued map on $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$ that satisfies the following four axioms. For any $A, B$ in $\mathcal{F}$ and $C, D$ in $\mathcal{C}_{\supset\emptyset}$,

CP1. $P(A|C) = 1$ whenever $A$ includes $C$;

CP2. $P(A|C) \geq 0$;

CP3. $P(A \cup B|C) = P(A|C) + P(B|C)$ whenever $A$ and $B$ are disjoint;

CP4. $P(A \cap D|C) = P(A|D \cap C)P(D|C)$.

If the field of events $\mathcal{F}$ and the field of conditioning events $\mathcal{C}$ are equal, then we call $P$ a full conditional probability on $\mathcal{F}$ following Dubins (1975, Section 3). For any event $A$ in $\mathcal{F}$ and any conditioning event $C$ in $\mathcal{C}_{\supset\emptyset}$, we call $P(A|C)$ *the probability of $A$ conditional on $C$*. The sole exception to this occurs if the conditioning event $C$ is the sure event $\mathcal{X}$; then we call $P(A|\mathcal{X})$ the *probability of $A$*, and usually resort to the shorthand $P(A) := P(A|\mathcal{X})$ whenever there can be no confusion about its meaning.

Observe that (CP1)–(CP3) are simply the conditional versions of $(P1)_{34}$–$(P3)_{34}$. For this reason, for any $C$ in $\mathcal{C}_{\supset\emptyset}$,

$$P(\bullet|C)\colon \mathcal{F} \to \mathbb{R}\colon A \mapsto P(A|C)$$

is a probability charge on $\mathcal{F}$. To make sense of (CP4), we set $C = \mathcal{X}$, to yield

$$P(A \cap D) = P(A|D)P(D),$$

where we have used the notational convention $P(\bullet) = P(\bullet|\mathcal{X})$. Clearly, this is the multiplicative version of Eq. $(2.20)_{39}$ – that is, Bayes's rule. This being said, the crucial difference with (conditional) probability charges is that the conditional probabilities are fundamental this time around; more precisely, we specify the conditional probability $P(\bullet|C)$ for *every* conditioning event $C$ in $\mathcal{C}_{\supset\emptyset}$, even if the conditioning event $C$ has (unconditional) probability $P(C) = P(C|\mathcal{X})$ zero. That said, given a probability charge $P$ on $\mathcal{F}$ and a field $\mathcal{C}$ of conditioning events that is included in $\mathcal{F}$, we can always derive a

[6] To see why we need to exclude the empty event $\emptyset$ as a conditioning event, we fix any non-empty event $A$ in $\mathcal{F}$. Were we to allow $\emptyset$ as a conditioning event, then (CP1) would imply that $P(A|\emptyset) = 1$ and $P(A^c|\emptyset) = 1$, but $P(A|\emptyset) = P(A|\emptyset) + P(A^c|\emptyset)$ due to (CP3), which is a contradiction.

full conditional probability through Bayes's rule whenever every conditioning event $C$ in $\mathcal{C}_{\supset\emptyset}$ has non-zero probability $P(C)$.

**Bruno's Example 2.48.** In Bruno's Example 2.37<sub>35</sub>, we defined a probability charge $P$ on the field $\mathcal{F}$ of events that only depend on a finite number of flips of Bruno's machine. We briefly repeat this here, with some changes in notation. For any probability $q$ of heads in $[0,1]$, we let $p_q$ be the probability mass function on $\{\mathtt{H},\mathtt{T}\}$ defined by $p_q(\mathtt{H}) = q$ and $p_q(\mathtt{T}) = 1-q$. This probability mass function $p_q$ induces the corresponding probability charge $P_q$ through Eqs. (2.16)<sub>35</sub> and (2.17)<sub>35</sub>:

$$P_q\{X_{1:n} \in A\} := \sum_{y_{1:n}\in A} P_q\{X_{1:n} = y_{1:n}\} = \sum_{y_{1:n}\in A} \prod_{k=1}^{n} p_q(y_k),$$

where $\{X_{1:n} \in A\}$ is a generic event in the field $\mathcal{F}$.

One can verify that $P_q\{X_{1:n} \in A\} > 0$ for *every* non-empty event $\{X_{1:n} \in A\}$ in $\mathcal{F}$ if and only if $0 < q < 1$. In this case, we can use Eq. (2.20)<sub>39</sub> to define a conditional probability charge $P_q(\bullet|C)$ on $\mathcal{F}$ for every non-empty event $C$ in $\mathcal{F}$. That is, whenever $0 < q < 1$, we can define the real-valued map $P_q$ on $\mathcal{F} \times \mathcal{F}_{\supset\emptyset}$ as

$$P_q(A|C) := \frac{P_q(A \cap C)}{P_q(C)} \quad \text{for all } (A,C) \in \mathcal{F} \times \mathcal{F}_{\supset\emptyset}.$$

Note that $P_q(A|\mathcal{X}) = P_q(A\cap\mathcal{X})/P_q(\mathcal{X}) = P_q(A)$, so we can unambiguously use the shorter notation $P_q(\bullet)$ for $P_q(\bullet|\mathcal{X})$.

By definition, $P_q$ satisfies (CP4)↶. Additionally – and as previously mentioned right after Eq. (2.20)<sub>39</sub> – $P_q(\bullet|C)$ is a probability charge on $\mathcal{F}$ and therefore satisfies (P1)<sub>34</sub>–(P3)<sub>34</sub>, so $P_q$ satisfies (CP1)↶–(CP3)↶. In conclusion, if $0 < q < 1$, then $P_q$ is a full conditional probability on $\mathcal{F}$. ∯

The following result lists some well-known properties of full conditional probabilities.

**Proposition 2.49.** *Consider a full conditional probability $P$ on $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$, where $\mathcal{F}$ and $\mathcal{C}$ are two fields of events over some possibility space $\mathcal{X}$ such that $\mathcal{F}$ includes $\mathcal{C}$. Then for any $A, B$ in $\mathcal{F}$ and $C$ in $\mathcal{C}_{\supset\emptyset}$,*

CP5. $P(\emptyset|C) = 0$;

CP6. $P(A|C) \leq 1$;

CP7. $P(A^{c}|C) = 1 - P(A|C)$;

CP8. $P(A|C) \leq P(B|C)$ *whenever $A$ is included in $B$;*

CP9. $P(A|C) = P(A\cap B|C)$ *whenever $P(B|C) = 1$.*

*Proof.* CP5. Because $\emptyset \cap \emptyset = \emptyset$, it follows from (CP3)↶ that $P(\emptyset|C) = 2P(\emptyset|C)$. Because $P(\emptyset|C)$ is a real number, this can only be true if $P(\emptyset|A) = 0$.

CP6. Observe that the events $A$ and $C \setminus A$ are disjoint, and their union $A \cup (C \setminus A) = A \cup C$ includes $C$. The property now follows immediately from this and (CP3)$_{41}$, (CP2)$_{41}$ and (CP1)$_{41}$.

CP7. Observe that $A$ and $A^{\mathrm{c}}$ are disjoint, and that their union $A \cup A^{\mathrm{c}} = \mathcal{X}$ includes $C$. The property now follows from (CP3)$_{41}$ and (CP1)$_{41}$.

CP8. Observe that $A$ and $B \setminus A$ are disjoint. The property now follows from this, (CP3)$_{41}$ and (CP2)$_{41}$.

CP9. Observe that the events $A \cap B$ and $A \cap B^{\mathrm{c}}$ are disjoint, and that their union is $A$. Therefore, it follows from (CP3)$_{41}$ that

$$P(A \,|\, C) = P(A \cap B \,|\, C) + P(A \cap B^{\mathrm{c}} \,|\, C).$$

The property now follows if $P(A \cap B^{\mathrm{c}} \,|\, C) = 0$. To see that this is the case, we observe that $A \cap B^{\mathrm{c}}$ is included in $B^{\mathrm{c}}$, and use (CP2)$_{41}$, (CP8)$_{\curvearrowleft}$ and (CP7)$_{\curvearrowleft}$, to yield

$$0 \leq P(A \cap B^{\mathrm{c}} \,|\, C) \leq P(B^{\mathrm{c}} \,|\, C) = 1 - P(B \,|\, C) = 0. \qquad \square$$

The structure that is required of the domain $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$ of a full conditional probability is rather strong. For example, due to the requirement that $\mathcal{C}$ should be a field, it might be that we need to specify probabilities conditional on events $C$ that we are not actually interested in. For this reason, we would like to define a 'conditional probability' on a subset $\mathcal{D}$ of $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})_{\supset\emptyset}$ that is *not* a Cartesian product of two nested fields. One could again require that this map $P$ should satisfy the laws of conditional probability – that is, (CP1)$_{41}$–(CP4)$_{41}$ – whenever all relevant terms are well-defined, but this would yield some undesirable behaviour. The following example – essentially taken from (Krak et al., 2017, Example 4.1) – illustrates that the laws of probability alone do not guarantee that we can extend $P$ to a full conditional probability $P^{\star}$ on a domain $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$ that includes the domain $\mathcal{D}$ of $P$.

**Bruno's Example 2.50.** Let

$$\mathcal{D}' := \{(H_1, \mathcal{X}), (H_1^{\mathrm{c}}, \mathcal{X})\},$$

where $H_1 = \{X_1 = \mathtt{H}\}$ and $H_1^{\mathrm{c}} = \{X_1 = \mathtt{T}\}$ are as defined in Bruno's Example 2.33$_{32}$. Consider the real-valued map $P'$ on $\mathcal{D}'$ defined by $P'(H_1 \,|\, \mathcal{X}) := 1/3$ and $P'(H_1^{\mathrm{c}} \,|\, \mathcal{X}) := 1/3$. It is almost trivial to verify that $P'$ satisfies (CP1)$_{41}$–(CP4)$_{41}$ for those events $A$, $B$, $C$ and $D$ in its domain $\mathcal{D}'$ for which all terms in these properties are well-defined. Despite this, there is *no* full conditional probability $P^{\star}$ on $\mathcal{F} \times \mathcal{F}_{\supset\emptyset}$ – or even on $\{\emptyset, H_1, H_1^{\mathrm{c}}, \mathcal{X}\} \times \{\mathcal{X}\}$ – that extends $P'$. To verify this, we assume ex-absurdo that there is one, and denote it by $P^{\star}$. On the one hand,

$$P^{\star}(H_1 \,|\, \mathcal{X}) + P^{\star}(H_1^{\mathrm{c}} \,|\, \mathcal{X}) = P'(H_1 \,|\, \mathcal{X}) + P'(H_1^{\mathrm{c}} \,|\, \mathcal{X}) = \frac{2}{3}$$

because $P^{\star}$ extends $P'$. On the other hand, it also follows from (CP3)$_{41}$ and (CP1)$_{41}$ that

$$P^{\star}(H_1 \,|\, \mathcal{X}) + P^{\star}(H_1^{\mathrm{c}} \,|\, \mathcal{X}) = P^{\star}(H_1 \cup H_1^{\mathrm{c}} \,|\, \mathcal{X}) = P^{\star}(\mathcal{X} \,|\, \mathcal{X}) = 1.$$

Clearly, these two equalities contradict each other. ⦰

To ensure that a conditional probability $P$ on $\mathcal{D}$ can always be extended to a full conditional probability on a domain that includes $\mathcal{D}$, we return to the notion of coherence, this time for conditional probabilities. P. M. Williams (1975, 2007, Definition 2) was the first to characterise this, although he establishes coherence in the more general setting of conditional (lower) expectations. Regazzini (1985, Definition 1) focuses more explicitly on coherent conditional probabilities, which is why we prefer to follow him here. One exception is our definition, though, because we opt to repeat the simplified formulation of Lopatatzidis (2017, Definition 5) and Krak et al. (2017, Definition 4.2).

**Definition 2.51.** Consider a possibility space $\mathcal{X}$ and a subset $\mathcal{D}$ of $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})_{\supset\emptyset}$. A *coherent conditional probability* $P$ on $\mathcal{D}$ is a real-valued map on $\mathcal{D}$ such that for all natural numbers $n$, all real numbers $\mu_1, \dots, \mu_n$ and all $(A_1, C_1)$, …, $(A_n, C_n)$ in $\mathcal{D}$,

$$\max\left\{\sum_{k=1}^{n} \mu_k \mathbb{I}_{C_k}(x)\big(\mathbb{I}_{A_k}(x) - P(A_k \,|\, C_k)\big) \colon x \in \bigcup_{k=1}^{n} C_k\right\} \geq 0. \tag{2.24}$$

One motivation for Eq. (2.24) – the coherence condition – is through a behavioural interpretation in terms of betting behaviour, as in Section $2.2.2_{19}$. More precisely, a justification that is reminiscent of the behavioural interpretation of coherent expectations as fair prices. For this behavioural justification, we interpret the conditional probability $P$ on some subset $\mathcal{D}$ of $\mathcal{P}(\mathcal{F}) \times \mathcal{P}(\mathcal{F})_{\supset\emptyset}$ as a partial acceptability assessment of the subject, just like we did for expectations $E$ in Section $2.2.2_{19}$. More specifically, for any $(A, C)$ in $\mathcal{D}$, we interpret $P(A\,|\,C)$ as the subject's *called-off fair price for the gamble* $\mathbb{I}_A$ *contingent on* $C$; that is, she is imposed to accept a transaction where she buys the uncertain reward $\mathbb{I}_A$ for any buying price $\alpha$ lower than $P(A\,|\,C)$ or sells $\mathbb{I}_A$ for any selling price $\beta$ greater than $P(A\,|\,C)$, under the condition that the transaction is called off if $C$ does not occur. Formally, she is imposed to accept the gambles

$$\mathbb{I}_C(\mathbb{I}_A - \alpha) \quad \text{and} \quad \mathbb{I}_C(\beta - \mathbb{I}_A)$$

for any buying price $\alpha < P(A\,|\,C)$ and any selling price $\beta > P(A\,|\,C)$. This way, the map $P$ on $\mathcal{D}$ induces the set of acceptable gambles

$$\mathcal{A}_P \coloneqq \left\{\mathbb{I}_C(\mathbb{I}_A - \alpha) \colon (A, C) \in \mathcal{D}, \alpha \in \mathbb{R}, \alpha < P(A\,|\,C)\right\} \\ \cup \left\{\mathbb{I}_C(\beta - \mathbb{I}_A) \colon (A, C) \in \mathcal{D}, \beta \in \mathbb{R}, \beta > P(A\,|\,C)\right\}.$$

Under this betting interpretation, the condition of Definition 2.51 ensures that the subject's set of acceptable gambles $\mathcal{A}_P$ avoids partial loss, similarly to what we saw in Theorem $2.11_{21}$ (see Troffaes et al., 2014, Definition 13.18).

**Theorem 2.52.** *Consider a possibility space $\mathcal{X}$ and a real-valued map $P$ on a subset $\mathcal{D}$ of $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})_{\supset\emptyset}$. Then $\mathcal{A}_P$ avoids sure loss if and only if $P$ is a coherent conditional probability.*

This betting interpretation also provides an intuitive reason for why in Eq. (2.24) we only take into account the outcomes $x$ in $\bigcup_{k=1}^{n} C_k$: with all other outcomes, all transactions are called off so the subject gains neither loses any utility. In other words, in case $\bigcup_{k=1}^{n} C_k \neq \mathcal{X}$, the condition would otherwise be trivially true.

A second, more pragmatic motivation for the condition of Definition 2.51 is that – as promised – it allows us to first specify a coherent conditional probability on a domain $\mathcal{D}$ that need not be of the form $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$, and subsequently extend this to a full conditional probability with domain $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$ that includes $\mathcal{D}$. Before we get to this result, we first establish that full conditional probabilities and coherent conditional probabilities are essentially equivalent (see Regazzini, 1985, Theorem 3).

**Theorem 2.53.** *Consider a possibility space $\mathcal{X}$ and a real-valued map $P$ on $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$, where $\mathcal{F}$ and $\mathcal{C}$ are fields of events over $\mathcal{X}$ and $\mathcal{F}$ includes $\mathcal{C}$. Then $P$ is a full conditional probability if and only if it is a coherent conditional probability.*

Next, we establish what is probably the single most important result regarding (coherent) conditional probabilities; we refer to (Regazzini, 1985, Theorem 4) for a proof.

**Theorem 2.54.** *Consider a possibility space $\mathcal{X}$ and a coherent conditional probability $P$ on a subset $\mathcal{D}$ of $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})_{\supset\emptyset}$. Then for any superset $\mathcal{D}^\star$ of $\mathcal{D}$, there is a coherent conditional probability $P^\star$ on $\mathcal{D}^\star$ that extends $P$ to $\mathcal{D}^\star$, or equivalently, that coincides with $P$ on $\mathcal{D}$.*

Theorem 2.54 is a highly useful tool in proofs, because it allows us to extend a coherent conditional probability that is defined on some small and manageable domain $\mathcal{D}$ to some larger domain $\mathcal{D}^\star$ that we are actually interested in. This is especially relevant because checking the coherence condition of Definition 2.51 can simplify considerably – or, more precisely, can become feasible – when the domain $\mathcal{D}$ is sufficiently small. That being said, we should emphasise that this extension is not necessarily unique. Fortunately, the laws of (conditional) probability usually impose convenient restrictions on the probabilities in the extension, and they determine them uniquely often enough. For a good example of this, we refer to Section 3.1.573 further on. That coherent conditional probabilities indeed satisfy the laws of probability is established by the following result.

**Corollary 2.55.** *Consider a possibility space $\mathcal{X}$ and a coherent conditional probability $P$ on a subset $\mathcal{D}$ of $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})_{\supset\emptyset}$. Then $P$ satisfies (CP1)41–(CP9)42 on its domain $\mathcal{D}$.*

*Proof.* Follows immediately from Definition 2.47$_{41}$, Proposition 2.49$_{42}$ and Theorems 2.53$_{\curvearrowleft}$ and 2.54$_{\curvearrowleft}$. □

To end our introduction of (coherent) conditional probabilities, we return to one of our motivations for introducing them: we establish that a coherent conditional probability can always be extended to a full conditional probability.

**Corollary 2.56.** *Consider a possibility space $\mathcal{X}$ and a subset $\mathcal{D}$ of $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})_{\supset\emptyset}$. Then a real-valued map $P$ on $\mathcal{D}$ is a coherent conditional probability if and only if it can be extended to a full conditional probability on $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$, with $\mathcal{F}$ and $\mathcal{C}$ any two fields over $\mathcal{X}$ such that $\mathcal{F}$ includes $\mathcal{C}$ and $\mathcal{F} \times \mathcal{C}_{\supset\emptyset} \supseteq \mathcal{D}$.*

*Proof.* Follows immediately from Theorems 2.53$_{\curvearrowleft}$ and 2.54$_{\curvearrowleft}$. □

### 2.4.2 From conditional probability to conditional expectation

The domain $\mathcal{D}$ of a coherent conditional probability $P$ is not required to have any structure whatsoever. Nonetheless, most of the coherent conditional probabilities that we will encounter throughout this dissertation do have a specifically structured domain, and it will be precisely this structure that allows us to incorporate classical probability theory. More precisely, we will often restrict ourselves to domains $\mathcal{D}$ that have the following structure.

**Definition 2.57.** Consider a possibility space $\mathcal{X}$. A subset $\mathcal{D}$ of $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})_{\supset\emptyset}$ is a *structure of fields* if there is a non-empty set $\mathcal{C}$ of non-empty events in $\mathcal{P}(\mathcal{X})_{\supset\emptyset}$ and, for every conditioning event $C$ in $\mathcal{C}$, a field $\mathcal{F}_C$ of events over $\mathcal{X}$ such that

$$\mathcal{D} = \big\{(A, C)\colon C \in \mathcal{C}, A \in \mathcal{F}_C\big\}. \tag{2.25}$$

Observe that the domain $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$ of a full conditional probability is a structure of fields, but not every structure of fields is of the form $\mathcal{F} \times \mathcal{C}_{\supset\emptyset}$, with $\mathcal{C}$ and $\mathcal{F}$ two nested fields of events. This structure is interesting because it allows us to relate coherent conditional probabilities to probability charges.

**Corollary 2.58.** *Consider a possibility space $\mathcal{X}$ and a structure of fields $\mathcal{D} = \{(A, C)\colon C \in \mathcal{C}, A \in \mathcal{F}_C\}$. If $P$ is a coherent conditional probability on $\mathcal{D}$, then for every conditioning event $C$ in $\mathcal{C}$,*

$$P(\bullet\,|\,C)\colon \mathcal{F}_C \to \mathbb{R}\colon A \mapsto P(A|C)$$

*is a probability charge on $\mathcal{F}_C$.*

*Proof.* Recall from Corollary 2.55$_{\curvearrowleft}$ that $P$ satisfies (CP1)$_{41}$–(CP9)$_{42}$ on its domain $\mathcal{D}$. Consequently, $P(\bullet\,|\,C)$ satisfies (P1)$_{34}$–(P3)$_{34}$. □

It is important to realise that the converse of Corollary 2.58 is not necessarily true, that is, we cannot arbitrarily define a probability charge $P(\bullet\,|\,C)$ for every conditioning event $C$ and expect to end up with a coherent conditional probability. The reason for this is, essentially, that the conditional probabilities should be linked through (CP4)$_{41}$, the multiplicative version of Bayes's rule.

### Expectation of simple variables

Let $P$ be a coherent conditional probability on a structure of fields $\mathcal{D} = \{(A,C)\colon C \in \mathcal{C}, A \in \mathcal{F}_C\}$. Due to Corollary 2.58, for any conditioning event $C$ in $\mathcal{C}$, $P(\bullet\,|\,C)$ is a probability charge on $\mathcal{F}_C$. As we have seen in Section 2.3.3$_{35}$, this (conditional) probability charge $P(\bullet\,|\,C)$ defines a (conditional) expectation operator $E_P(\bullet\,|\,C)$ on $\mathbb{S}(\mathcal{F}_C)$ as

$$E_P(f\,|\,C) := E_{P(\bullet|C)}(f) \quad \text{for all } f \in \mathbb{S}(\mathcal{F}_C), \tag{2.26}$$

where we use $E_P(f\,|\,C)$ to denote the conditional expectation to emphasize that it is actually conditional, and where $E_{P(\bullet|C)}$ is the Dunford integral with respect to the probability charge $P(\bullet\,|\,C)$ as defined by Eq. (2.19)$_{36}$. This way, $P$ induces the conditional expectation operator $E_P$ with domain

$$\mathbb{C}\mathbb{S}(\mathcal{D}) := \big\{(f,C)\colon C \in \mathcal{C}, f \in \mathbb{S}(\mathcal{F}_C)\big\}. \tag{2.27}$$

Whenever there can be no confusion, we will leave out the conditioning event $C$ if it is the sure event $\mathcal{X}$:

$$E_P(f) := E_P(f\,|\,\mathcal{X}) = E_{P(\bullet|\mathcal{X})}(f) \quad \text{for all } f \in \mathbb{S}(\mathcal{F}_{\mathcal{X}});$$

in short, we follow the same convention as with conditional probabilities.

### Extension through coherence

An interesting side note is that we can also arrive at the conditional expectation $E_P$ through the notion of coherence for conditional (lower) expectations, and the natural extension that goes along with it. We will not provide definitions for these concepts here, but invite the interested reader to take a look at (P. M. Williams, 1975, 2007), (Regazzini, 1985) and (Troffaes et al., 2014). There, you will find out that P. M. Williams (1975) introduces the notion of coherent conditional lower expectations on gambles, that Regazzini (1985) more thoroughly studies the special cases of coherent conditional expectations – also on gambles – and probabilities, and that Troffaes et al. (2014, Chapter 13) generalise Williams' work from gambles to (unbounded) real variables. As we are about to see, the concepts of coherence and natural extension for conditional expectations are similar to their unconditional counterparts of Section 2.2$_{16}$.

Let $P$ be a coherent conditional probability on some structure of fields $\mathcal{D} = \{(A, C) \colon C \in \mathcal{C}, A \in \mathcal{F}_C\}$. Clearly, this coherent conditional probability $P$ contains the same information as the restriction of $E_P$ to $\mathcal{G}_{\mathcal{D}} \coloneqq \{(\mathbb{I}_A, C) \colon C \in \mathcal{C}, A \in \mathcal{F}_C\}$, which is a coherent conditional expectation in the sense of (Regazzini, 1985, Definition 2) and (Troffaes et al., 2014, Definition 13.28). Troffaes et al. (2014, Definition 13.23 and Theorem 13.47) explain how the most conservative coherent conditional lower expectation that extends the coherent conditional expectation $E_P$ from $\mathcal{G}_{\mathcal{D}}$ to $\mathbb{CS}(\mathcal{D})$ is the natural extension $\underline{\mathcal{E}}_P$ on $\mathbb{CS}(\mathcal{D})$, defined for any $(f, C)$ in $\mathbb{CS}(\mathcal{D})$ by

$$\underline{\mathcal{E}}_P(f \mid C) \coloneqq \sup\bigl\{\alpha \in \mathbb{R} \colon \mathbb{I}_C(f - \alpha) \in \mathcal{E}(\mathcal{A}_P)\bigr\}.$$

Observe that this is a straightforward adaptation to the conditional case of the unconditional natural extension as defined in Eq. $(2.9)_{24}$. Because $E_P$ is coherent, it coincides with its natural extension $\underline{\mathcal{E}}_P$ on its domain; formally,

$$\underline{\mathcal{E}}_P(\mathbb{I}_A \mid C) = E_P(\mathbb{I}_A \mid C) = P(A \mid C) = \overline{\mathcal{E}}_P(\mathbb{I}_A \mid C) \quad \text{for all } (A, C) \in \mathcal{D}, \tag{2.28}$$

where $\overline{\mathcal{E}}_P \coloneqq -\underline{\mathcal{E}}_P(-\bullet \mid \bullet)$ is the conjugate coherent conditional upper expectation of the natural extension $\underline{\mathcal{E}}_P$.

To confirm that we could also have obtained $E_P$ through coherence, we fix any conditioning event $C$ in $\mathcal{C}$ and any $\mathcal{F}_C$-simple variable $f$ with representation $f = \sum_{k=1}^{n} a_k \mathbb{I}_{A_k}$. First, it follows from (Troffaes et al., 2014, Theorem 3.31(i)), the conditional counterpart of $(\text{LE4})_{30}$, that

$$\underline{\mathcal{E}}_P(f \mid C) \le \overline{\mathcal{E}}_P(f \mid C)$$

Next, it follows from the properties of coherent conditional lower expectations – more precisely from their super-additivity and non-negative homogeneity (see Troffaes et al., 2014, Theorem 13.31) – that

$$\underline{\mathcal{E}}_P(f \mid C) \ge \sum_{k=1}^{n} \underline{\mathcal{E}}_P(a_k \mathbb{I}_{A_k} \mid C) = \sum_{k=1}^{n} a_k \underline{\mathcal{E}}_P(\mathbb{I}_{A_k} \mid C) = \sum_{k=1}^{n} a_k P(A_k \mid C) = E_P(f \mid C),$$

where for the last two equalities we have used Eq. (2.28) and Eq. $(2.19)_{36}$. Similarly, for the conjugate of the natural extension we obtain that

$$\overline{\mathcal{E}}_P(f \mid C) \le \sum_{k=1}^{n} \overline{\mathcal{E}}_P(a_k \mathbb{I}_{A_k} \mid C) = \sum_{k=1}^{n} a_k P(A_k \mid C) = E_P(f \mid C).$$

We combine the preceding three inequalities, to yield

$$\underline{\mathcal{E}}_P(f \mid C) = E_P(f \mid C) = \overline{\mathcal{E}}_P(f \mid C) \quad \text{for all } (f, C) \in \mathbb{CS}(\mathcal{D}).$$

In other words, the natural extension $\underline{\mathcal{E}}_P$ is self-conjugate on $\mathbb{CS}(\mathcal{D})$.[7] For this reason, $E_P$ is a coherent conditional expectation in the sense of (Troffaes et al., 2014, Definition 13.28).

[7] Troffaes et al. (2014, Proposition 8.17) make this point as well, but they only explicitly do so for probability charges and coherent lower expectations.

### 2.4.3 Sets of coherent conditional probabilities

By introducing coherent conditional probabilities, we have successfully mitigated what we perceive as one of the main drawbacks of classical probability theory: the issues with conditioning on events that have probability zero. That being said, coherent conditional probabilities still require us to specify conditional probabilities *precisely* or *exactly*, and this is often infeasible if not impossible. For this reason, we free ourselves of this requirement of precision here, and instead consider sets of coherent conditional probabilities.

We fix some structure of fields $\mathcal{D} = \{(A,C) : C \in \mathcal{C}, A \in \mathcal{F}_C\}$, and let $\mathcal{P}$ be a non-empty set of coherent conditional probabilities on $\mathcal{D}$. Formally, there are no additional restrictions whatsoever on the coherent conditional probabilities in the set $\mathcal{P}$, but within the context of this dissertation, we will typically interpret it from a sensitivity analysis point of view – which we have previously mentioned in Section $2.2.4_{27}$. We leave a more detailed discussion for Chapter $3_{53}$ though; for now, it can be helpful to think of the following example.

**Bruno's Example 2.59.** Alicia is convinced that the infinite sequence $(X_n)_{n\in\mathbb{N}}$ of coin flips of Bruno's machine can be adequately modelled by what is commonly known as a sequence of 'independent and identically distributed random variables'. In short, this means that she believes that the coherent conditional probability $P_q$ of Bruno's Example $2.48_{42}$, with $q$ in $]0,1[$ the probability of heads, is an adequate model. Alicia is a bit of a cautious woman though, and this stops her from fixing a single value for the parameter $q$; she only feels comfortable to say that $q$ can have any value in the interval $[3/7, 4/7]$. In short, her beliefs are captured by the set

$$\mathcal{P} := \{P_q : q \in [3/7, 4/7]\}.$$

#### Lower and upper envelopes

As in Section $2.2.4_{27}$, we look at lower and upper envelopes with respect to the non-empty set $\mathcal{P}$ of coherent conditional probabilities. For starters, the conditional lower and upper probabilities $\underline{P}_{\mathcal{P}}$ and $\overline{P}_{\mathcal{P}}$ are the real-valued maps on the structure of fields $\mathcal{D}$ defined as

$$\underline{P}_{\mathcal{P}}(A\,|\,C) := \inf\{P(A\,|\,C) : P \in \mathcal{P}\} \qquad \text{for all } (A,C) \in \mathcal{D}$$

and

$$\overline{P}_{\mathcal{P}}(A\,|\,C) := \sup\{P(A\,|\,C) : P \in \mathcal{P}\} \qquad \text{for all } (A,C) \in \mathcal{D}.$$

Note that, as an immediate consequence of $(\text{CP7})_{42}$, these lower and upper conditional probabilities are conjugate, in the sense that $\overline{P}_{\mathcal{P}}(A\,|\,C) = 1 - \underline{P}_{\mathcal{P}}(A^{\mathrm{c}}\,|\,C)$ for all $(A,C)$ in $\mathcal{D}$. Thus, it suffices to study either the lower or the upper conditional probability; in this dissertation, we choose the former.

Recall from Section 2.4.2$_{46}$ that because every coherent conditional probability $P$ in $\mathcal{P}$ has the structure of fields $\mathcal{D}$ as domain, it induces a conditional expectation $E_P$ on $\mathbb{CS}(\mathcal{D})$. Thus, we can also take the lower and upper envelope of the set $\{E_P \colon P \in \mathcal{P}\}$ of induced conditional expectations. Formally, these conditional lower and upper expectations $\underline{E}_{\mathcal{P}}$ and $\overline{E}_{\mathcal{P}}$ are the real-valued maps on $\mathbb{CS}(\mathcal{D})$ defined by

$$\underline{E}_{\mathcal{P}}(f \mid C) := \inf\{E_P(f \mid C) \colon P \in \mathcal{P}\} \qquad \text{for all } (f, C) \in \mathbb{CS}(\mathcal{D})$$

and

$$\overline{E}_{\mathcal{P}}(f \mid C) := \sup\{E_P(f \mid C) \colon P \in \mathcal{P}\} \qquad \text{for all } (f, C) \in \mathbb{CS}(\mathcal{D}).$$

Here too, it suffices to only study the lower envelope because lower and upper envelopes are conjugate; more precisely, it follows from (ES2)$_{37}$ that $\overline{E}_{\mathcal{P}}(f \mid C) = -\underline{E}_{\mathcal{P}}(-f \mid C)$ for all $(f, C)$ in $\mathbb{CS}(\mathcal{D})$. Furthermore, the lower conditional expectation includes lower conditional probabilities – and thus by conjugacy also upper conditional probabilities – as a special case, because

$$\underline{P}_{\mathcal{P}}(A \mid C) = \underline{E}_{\mathcal{P}}(\mathbb{I}_A \mid C) \quad \text{for all } (A, C) \in \mathcal{D}.$$

We again follow the convention that the conditioning event is left out of the notation whenever it is the sure event $\mathcal{X}$, so $\underline{E}_{\mathcal{P}}(\bullet) := \underline{E}_{\mathcal{P}}(\bullet \mid \mathcal{X})$, and similarly for the conjugate upper conditional expectation and the conditional lower and upper probabilities.

**Bruno's Example 2.60.** In Bruno's Example 2.41$_{37}$, we defined the $\mathcal{F}$-simple variable $h_n$ as the number of heads in the first $n$ flips. With the notation of Bruno's Example 2.48$_{42}$, we now find that

$$E_{P_q}(h_n \mid \mathcal{X}) = E_{P_q}(h_n) = nq,$$

where the final equality is established in Bruno's Example 2.41$_{37}$. With $\mathcal{P}$ as defined in Bruno's Example 2.59$_{\frown}$, we therefore find that

$$\begin{aligned}\underline{E}_{\mathcal{P}}(h_n) = \underline{E}_{\mathcal{P}}(h_n \mid \mathcal{X}) &= \inf\{E_P(h_n \mid \mathcal{X}) \colon P \in \mathcal{P}\}\\ &= \inf\{E_{P_q}(h_n \mid \mathcal{X}) \colon q \in [3/7, 4/7]\} = \inf\{nq \colon q \in [3/7, 4/7]\} = n\frac{3}{7}.\end{aligned}$$

Similarly, $\overline{E}_{\mathcal{P}}(h_n) = n4/7$.

### Coherent extension of the lower envelope

Recall that for any coherent conditional probability $P$ in the set $\mathcal{P}$, $E_P$ is a coherent conditional expectation in the sense of (Troffaes et al., 2014, Definition 13.28). For this reason, it follows from (Troffaes et al., 2014, Proposition 13.42) – the conditional counterpart of Definition 2.25$_{28}$ – that $\underline{E}_{\mathcal{P}}$ is a coherent conditional lower expectation in the sense of (Troffaes et al., 2014, Definition 13.25). We could therefore use the 'natural extension for conditional lower expectations' (see Troffaes et al., 2014, Section 13.7) to extend $\underline{E}_{\mathcal{P}}$ to any subset $\mathcal{K}$ of $\mathbb{G}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})_{\supset\emptyset}$ – or, to be more precise, $\mathbb{V}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})_{\supset\emptyset}$ – that includes $\mathbb{CS}(\mathcal{D})$. We will not detail how this works here, because this extension is ill-suited for our purposes. We will return to this point in Chapter $5_{215}$ further on; for now, it suffices to understand that this has to do with whether or not we should use the natural extension for gambles (or variables) that are not determinable.

### 2.4.4 Connecting determinable gambles and simple variables

Recall from Section $2.2.1_{16}$ that if we want to adhere to a gambling interpretation, then we should only consider those gambles that are determinable. As promised, we end this chapter with a closer look at what type of gambles are determinable. Let us first use our running example to give an example of a gamble that is not determinable.

**Bruno's Example 2.61.** Consider the event that 'all coin flips are heads', which in our notation, is defined as

$$H_{\mathrm{lim}} := \bigcap_{n\in\mathbb{N}} \{X_n = \mathtt{H}\}.$$

The indicator $\mathbb{I}_{H_{\mathrm{lim}}}$ of this event is a gamble, as this is the case for any indicator.

However, we can never decide without question whether the event $H_{\mathrm{lim}}$ occurs or not, because even after a million consecutive coin flips that result in heads, we can never be sure that the next coin flip will not turn out to be tails. This means that any bet based upon this event, and therefore any bet that involves the gamble $\mathbb{I}_{H_{\mathrm{lim}}}$, cannot be decided without question. For this reason, the gamble $\mathbb{I}_{H_{\mathrm{lim}}}$ is *not* determinable. ∯

Next, we try to determine the type of gambles that are determinable in the setting of our running example.

**Bruno's Example 2.62.** Recall that a gamble $g$ on $\mathcal{X}$ is determinable whenever its value can be decided without question. Clearly, we can only decide on the value of those gambles that only require a finite number of consecutive coin flips. In other words, for any determinable gamble $g$, there is a natural number $n$ such that $g$ only depends on the first $n$ coin tosses; that is, there is some real-valued function $h$ on $\{\mathtt{H},\mathtt{T}\}^n$ such that

$$g(x) = h(x_{1:n}) \quad \text{for all } x = (x_k)_{k\in\mathbb{N}} \in \mathcal{X}.$$

A gamble that satisfies this condition is called *n-measurable* by (Lopatatzidis, 2017, Section 3.5.1). Observe that any n-measurable gamble $g$ is $\mathcal{F}$-simple, because

$$g = \sum_{y_{1:n}\in\{\mathtt{H},\mathtt{T}\}^n} h(y_{1:n})\mathbb{I}_{\{X_{1:n}=y_{1:n}\}}.$$

Note that it need not be that every $\mathcal{F}$-simple variable is determinable. For example, one might argue that after some physical horizon we can no longer observe the outcome of the coin flip; in that case, not every $\mathcal{F}$-simple variable is determinable. The issue is then of course that we need to specify this horizon. One convenient way to circumvent this arbitrary choice, is to make the mathematical idealization that there is no such physical horizon.

The crucial observation is that, in the setting of our running example, every gamble that is determinable is a simple variable, and that it is convenient to use the idealization that every simple variable is determinable as well. This is by no means exclusive to the specific setting of our running example; in fact, we will also encounter this when dealing with jump processes in Chapter $3_{\curvearrowright}$ further on. In any case, because coherence and the natural extension are motivated through a betting interpretation, it is arguably not sensible to use the (conditional version of) natural extension to extend the domain of $\underline{E}_{\mathcal{P}}$ to gambles (or variables) that are not determinable – or, when the equivalence holds, simple. That is as far as we will go with this for now, but we will return to this in great length in Chapter $5_{215}$ further on.

# 3 *Modelling jump processes*

In this chapter, we use the basic mathematical tools for modelling uncertainty to introduce the main topic of this dissertation: jump processes. A jump process is a mathematical model for a subject's uncertainty regarding a system she is interested in. More precisely, she is interested in the state of this system, and how this state changes – or 'jumps' – over time, but she cannot specify the initial state nor how the state evolves over time with certainty.

We speak in general terms to make our framework as widely applicable as possible, but we will illustrate this wide applicability with some concrete examples. For a first example, we return to Bruno's Example $2.31_{31}$, where we were interested in the outcomes of the consecutive coin flips of Bruno's machine. In the terminology of the previous paragraph, Bruno's machine is the system, and its state is either heads or tails. It is natural to think of 'the outcome of the $n$-th coin flip' as 'the state at time $n$', so this is an example of a system that evolves over discrete time.

In this dissertation, however, we will mainly be concerned with systems that are more naturally studied in continuous time. Examples of such systems are abundant. One example that everyone has to endure once in a while is the queue at the checkout of a shop; in this case, the 'state at time $t$' is 'the number of customers in the queue at time $t$'. More involved queueing situations – with people but also with more general 'customers' – often also fit in this abstract framework, as we will see in Queuing Network Example $7.3_{340}$ in Chapter $7_{337}$ and also in Chapter $8_{403}$ further on. For a different example that is not related to queueing, we give our running example a little shake-up.

**Joseph's Example 3.1.** Bruno is getting a lot of attention with his coin flipping machine, to the envy of some of colleagues, including Joseph. In an attempt to steal away some of Bruno's limelight, Joseph comes up with a plan to build his own flipping coin flipping machine in his spare time. After a couple of design iterations, his contraption is ready to be revealed to his colleagues.

Joseph has no doubt that his machine will draw a crowd, because it works in a substantially different manner than Bruno's. While Bruno's machine mechanically flips an actual coin, Joseph's machine makes use of radioactive decay. More precisely, Joseph's machine contains some amount of radioactive isotope and a detector; whenever the machine detects a decay, it changes its display from heads to tails or vice versa.

In the example above, the 'system' is Joseph's machine, and 'the state at time $t$' is 'the output of the display at time $t$'. It is natural to use continuous time in this example, because the machine can flip – that is, change state – at any point in time. Throughout the remainder of this dissertation, we will use the system of Joseph's Example 3.1$_{\curvearrowleft}$ as a running example. It will help to keep in mind that the setting of Joseph's Example 3.1$_{\curvearrowleft}$ is simply the continuous time version of that in Bruno's Example 2.31$_{31}$.

The rest of this chapter is structured as follows. In Section 3.1$_{\curvearrowright}$, we apply the concepts of Chapter 2$_{11}$ to model the subject's uncertainty regarding a system that evolves over continuous time with a 'jump process', which is a coherent conditional probability on a specific domain. Next, we study how a jump process simplifies when it is Markovian and time-homogeneous in Section 3.2$_{74}$. Crucially, we will establish how under these two properties, the jump process is uniquely characterised by two parameters: the initial mass function and the rate operator. Finally, we come to imprecise jump processes in Section 3.3$_{88}$, which we simply define as sets of jump process from the point of view of a sensitivity analysis. Concretely, instead of a single initial mass function and rate operator, we assume we have a set of initial mass functions and a set of rate operators, and then consider the set of all jump processes that are 'consistent' with these sets. Remarkably, it turns out that the lower envelope with respect to this set of consistent jump processes is also Markovian and time-homogeneous, so we end up with a proper imprecise version of (time-homogeneous) Markovian jump processes.

This chapter is largely based on (Krak et al., 2017) and – to a lesser extent – also on (De Bock, 2017b). There are a couple of differences between our exposition and that of Krak et al. (2017), though. One key difference is that we impose additional requirements on the outcomes in the possibility space, which will prove indispensable in Chapter 5$_{215}$ further on. Another key difference is that we propose a novel definition of the notion of 'consistency with a set of rate operators' – see Definitions 3.46$_{97}$ and 3.50$_{99}$ and Proposition 3.57$_{104}$ further on. Furthermore, we also present some novel results, the most notable of these being Theorems 3.75$_{114}$ and 3.89$_{120}$. From a historical perspective, it is important to note that Škulj (2015) was the first to consider Markovian imprecise jump processes. Furthermore, Škulj (2015), De Bock (2017b) and Krak et al. (2017) draw inspiration from earlier work on imprecise Markov chains – including but not limited to (De Cooman et al., 2008), (De Cooman et al., 2009), (Hermans et al., 2012) and (Škulj et al., 2013).

## 3.1 Jump processes

We want to model the subject's uncertainty regarding (the evolution of) the state of some system over time. This kind of uncertainty model is generally called a *stochastic process* – sometimes also random process – and it is common practice to call the set of all possible states of the system the *state space*.

Traditionally, stochastic processes are further distinguished based on the nature of time and the state space, both of which are either continuous or discrete. Most authors, like Doob (1953), Chung (1960) Anderson (1991), Norris (1997), Tornambè (1995), Stewart (2009) or Krak et al. (2017), use the term *chain* to refer to any stochastic process with a finite – or sometimes also countable – state space, adding the adjective 'discrete-time' or 'continuous-time' to distinguish the nature of the time axis. A minority of authors, including Feller (1968), Kendall et al. (1971), Iosifescu (1980), Billingsley (1995) or Borovkov (2013), opt to reserve the term chain for a stochastic process with a finite or countable state space that evolves in discrete time. Because we prefer not to use the adjectives 'continuous-time' and 'discrete-time' all the time, we follow the second group and exclusively associate the term chain with discrete time.

In this dissertation, we consider a system with a finite state space that evolves in continuous time. If such a system changes state we say that it *jumps* from one state to another. Similarly, we refer to a stochastic process for such a system as a *jump process*, following Gikhman et al. (1969) and Le Gall (2016).

The aim of this section is to formally define a jump process as a coherent conditional probability on a specific domain. To this end, we basically follow the steps as laid out in Chapter $2_{11}$. First, we fix the possibility space in Section 3.1.1. In Section $3.1.2_{58}$, we subsequently introduce some essential events and variables on this possibility space. Finally, we give the formal definition of a jump process in Section $3.1.3_{65}$, and establish some useful properties in Sections $3.1.4_{69}$ and $3.1.5_{73}$.

### 3.1.1 Càdlàg paths

We denote the state space of the system by $\mathscr{X}$,[1] and always assume that it is non-empty and *finite*. This might seem like an overly restrictive assumption from a mathematical point of view, but it makes sense from a practical point of view. A plethora of applications – including those we will encounter in Chapters $7_{337}$ and $8_{403}$ further on – can be analysed with a finite state space.

[1] Note that we denote both the generic state space $\mathscr{X}$ and the generic possibility space $\mathcal{X}$ with a 'calligraphic' capital X, although they are from a different font. Here and in the remainder, we consistently use the calligraphic font of the Times font family for 'general' uncertainty models and the calligraphic font of the STIX font family for jump processes.

This being said, all of the material in this section naturally generalises to countable state spaces; this has already been done in (Erreygers & De Bock, 2019b) for the special case of counting processes.

Because the system evolves in continuous time, it makes sense to think of 'the state of the system at every time point' as an outcome. Thus, an outcome is a function that maps every time point to the state of the system at this time point. We take $\mathbb{R}_{\geq 0}$ as time axis, so an outcome is a function $\omega$ from $\mathbb{R}_{\geq 0}$ to $\mathscr{X}$. We will refer to such outcomes as paths; that is, we call any function from $\mathbb{R}_{\geq 0}$ to the state space $\mathscr{X}$ a *path*. We collect all paths in the *set of paths* $\tilde{\Omega}_{\mathscr{X}}$, which we denote by $\tilde{\Omega}$ whenever there cannot be any confusion regarding the state space.

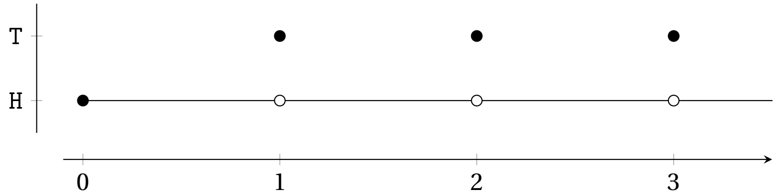
**Joseph's Example 3.2.** Clearly, the state of Joseph's machine is either heads or tails, so the state space $\mathscr{X}$ is equal to $\{\mathtt{H},\mathtt{T}\}$. Joseph's machine is the continuous-time counterpart of Bruno's machine, and this is reflected in the outcomes that we use: a path $\omega\colon \mathbb{R}_{\geq 0} \to \{\mathtt{H},\mathtt{T}\}$ is the natural continuous-time counterpart of a sequence $(x_n)_{n\in\mathbb{N}}$ in $\{\mathtt{H},\mathtt{T}\}$. ♫

In general, a path $\omega$ can have pretty erratic behaviour. Take, for example, the path $\omega_1$ that assumes $\mathtt{T}$ whenever the time point $t$ is natural and $\mathtt{H}$ otherwise, as depicted in Fig. 3.1(a)↷; or even more extreme, the path that assumes $\mathtt{T}$ whenever the time point $t$ is rational and the state $\mathtt{H}$ whenever the time point $t$ is irrational. This is not the type of behaviour that we expect from our system, so it makes sense to a priori exclude these erratic paths from the possibility space. Instead of this erratic behaviour, we expect that after the system changes states, it stays in the new state for some time. Mathematically, this translates to the requirement that a path $\omega\colon \mathbb{R}_{\geq 0} \to \mathscr{X}$ should be continuous from the right at all time points $t$ in $\mathbb{R}_{\geq 0}$. Even with this condition, a path can still display some volatile behaviour; we illustrate this with our running example.

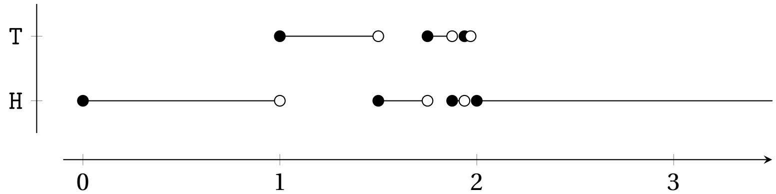
**Joseph's Example 3.3.** An example of a path that is right-continuous but nevertheless still behaves in a strange manner is the path $\omega_2$ defined by

$$\omega_2(t) := \begin{cases} \mathtt{H} & \text{if } t \in \bigcup_{n\in\mathbb{N}} [2-2^{-2n+3}, 2-2^{-2n+2}[, \\ \mathtt{T} & \text{if } t \in \bigcup_{n\in\mathbb{N}} [2-2^{-2n+2}, 2-2^{-2n+1}[, \\ \mathtt{H} & \text{if } t \geq 2, \end{cases} \quad \text{for all } t \in \mathbb{R}_{\geq 0},.$$

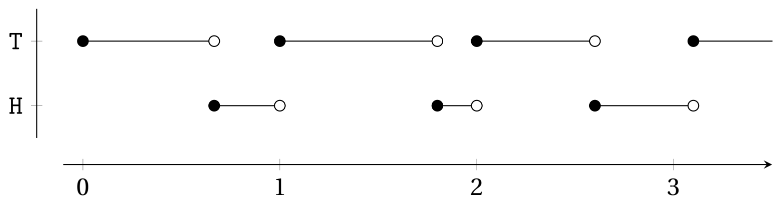
In Fig. 3.1(b)↷, we see that the path $\omega_2$ is right-continuous everywhere by construction. It starts off in the state $\mathtt{H}$, and then alternates between $\mathtt{T}$ and $\mathtt{H}$, every time changing states after a shorter time interval. This continues ad infinitum until the time point 2, after which the state is always $\mathtt{H}$. In other words, Joseph's machine switches between heads and tails infinitely often *before* the time point 2. Observe that the limit from the left exists for every positive time point $t$, *except* for $t = 2$. In other words, we can say that




(a) Graph of the path $\omega_1 : \mathbb{R}_{\geq 0} \to \{\mathtt{H}, \mathtt{T}\}$ that assumes T whenever the time point $t$ is natural and H otherwise.




(b) Graph of the path $\omega_2 : \mathbb{R}_{\geq 0} \to \{\mathtt{H}, \mathtt{T}\}$ as defined in Joseph's Example 3.3. Note that the graph is inaccurate to the left of $t = 2$ for obvious reasons.




(c) Graph of a càdlàg path $\omega_3 : \mathbb{R}_{\geq 0} \to \{\mathtt{H}, \mathtt{T}\}$.

**Figure 3.1** Examples of paths for the state space $\mathscr{X} = \{\mathtt{H}, \mathtt{T}\}$.

the state changes from heads to tails at every time point in the sequence $(2 - 2^{-2n+2})_{n \in \mathbb{N}}$ and from tails to heads at every time point in the sequence $(2 - 2^{-2n+1})_{n \in \mathbb{N}}$. Both of these sequences converge to 2, but it is impossible to say whether the state right before the time point 2 is heads or tails.

In order to get rid of this second type of strange behaviour, it suffices to additionally require that for every path $\omega$ in the possibility space, the left limit exists at all time points. The following standard definition – see, for example, (Billingsley, 1999, Section 12) or (Le Gall, 2016, p. 54) – combines these two requirements.

**Definition 3.4.** A path $\omega \colon \mathbb{R}_{\geq 0} \to \mathscr{X}$ is *càdlàg*[2] if it is continuous from the right and has limits from the left with respect to the discrete topology on $\mathscr{X}$.

[2] An acronym for the French phrase '[une fonction qui est] continue à droite [et admet une] limite à gauche'.

More formally, we require that

$$\big(\forall t \in \mathbb{R}_{\geq 0}\big)\big(\exists \delta \in \mathbb{R}_{>0}\big)\big(\forall r \in \,]t, t+\delta[\big)\ \omega(r) = \omega(t), \tag{3.1}$$

and

$$\big(\forall t \in \mathbb{R}_{>0}\big)\big(\exists \delta \in \,]0, t]\big)\big(\forall s, s' \in \,]t-\delta, t[\big)\ \omega(s) = \omega(s'). \tag{3.2}$$

We collect all càdlàg paths in the set $\Omega_{\mathscr{X}}$, which we will denote by $\Omega$ whenever the state space $\mathscr{X}$ is clear from the context.

A typical example of a càdlàg path is depicted in Fig. 3.1(c)$_{\curvearrowleft}$; note that the limit from the left indeed prevents the second type of strange behaviour. More precisely put, the càdlàg assumption ensures that any càdlàg path has at most a finite number of state changes or 'jumps' over any finite period of time – see Lemma 5.20$_{232}$ further on for the precise statement.

We take the set $\Omega$ as possibility space because we want to exclude the erratic behaviour of non-càdlàg paths a priori. Important to stress here is that this restriction to càdlàg paths is not really necessary for the definitions and results in this chapter; it will become vital in Chapter 5$_{215}$ further on, though. More precisely, as far as this and the next chapter are concerned, it would suffice to take as possibility space instead of $\Omega$ any set of paths $\tilde{\Omega}'$ that satisfies the condition of the following result (see Krak et al., 2017, Eq. (12)).

**Lemma 3.5.** *For every natural number $n$, any increasing sequence of time points $(t_1, \dots, t_n)$ in $\mathbb{R}_{\geq 0}$ and any states $x_1, \dots, x_n$ in $\mathscr{X}$, there is at least one path $\omega$ in $\Omega$ such that $\omega(t_k) = x_k$ for all $k$ in $\{1, \dots, n\}$.*

*Proof.* By construction, the path

$$\omega\colon \mathbb{R}_{\geq 0} \to \mathscr{X}\colon t \mapsto \omega(t) := \begin{cases} x_1 & \text{if } t \in [0, t_1), \\ x_k & \text{if } (\exists k \in \{1, \dots, n-1\})\ t_k \leq t < t_{k+1}, \\ x_n & \text{if } t \geq t_n, \end{cases}$$

satisfies the requirements of the statement: $\omega$ is càdlàg and for any $k$ in $\{1, \dots, n\}$, $\omega(t_k) = x_k$. □

### 3.1.2 Finitary events

Because we intend to model the subject's uncertainty regarding the state of the system using a coherent conditional probability, we initially restrict our attention to those events and gambles for which a betting interpretation makes sense. In Section 2.2.1$_{16}$ we called such an event or gamble determinable, in the sense that 'a bet based upon it can be decided without question'. In the present setting, it is clear that a gamble or event is definitely *not* determinable if its value cannot be determined after any (finite) period of time. For an example of such an event, we turn to our running example.

**Joseph's Example 3.6.** In our formalism, we idealise the event 'Joseph's machine displays heads all the time' as

$$\{\omega \in \Omega \colon (\forall t \in \mathbb{R}_{\geq 0})\ \omega(t) = \mathtt{H}\}.$$

Clearly, this event depends on the state of the system at all time points – that is, on $\mathbb{R}_{\geq 0}$. Thus, there is no time point $t$ in $\mathbb{R}_{\geq 0}$ such that we can determine if the event occurs – or, in other words, decide a bet based upon it – in case we know the state of the system at all times before $t$. Differently put, it can be that we have to wait 'infinitely long' before we can determine whether the event occurs or not.

Generally speaking, it is (also) practically impossible to observe the state of the system at infinitely many time points. For this reason, in the setting of jump processes we take the position that an event or a gamble can only be determinable if it depends on the state of the system at a finite number of time points. In Chapter 5215 further on, we will deal with variables that depend on the state of the system at more than a finite number of time points by means of – arguably more natural – limit arguments instead of through coherence and a direct betting interpretation. Similar to what we saw earlier for the discrete-time case in Bruno's Example 2.6251, it is not necessarily the case that every event or gamble that depends on the state of the system at a finite number of time points is determinable. For example, as in the discrete-time case, there might be a physical horizon after which we cannot realistically observe the system. Additionally, specific to the continuous-time case is that we cannot measure time up to arbitrary precision. In order not to get caught in these subtleties, we choose to adhere to an 'idealised' version of determinability: henceforth, we consider variables and gambles determinable if and only if they determine on the state of the system at a finite number of time points.

**Joseph's Example 3.7.** Consider two time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s < r$. In our formalism, the event 'Joseph's machine displays heads during the time interval $[s, r]$' is

$$H_{[s,r]} \coloneqq \{\omega \in \Omega \colon (\forall t \in [s, r])\ \omega(t) = \mathtt{H}\}.$$

Clearly, this event depends on the state of the system at all time points in the interval $[s, r]$, which is *not* a finite number of time points. Thus, the event $H_{[s,r]}$ is not determinable.

It is customary to let $X_t$ denote the state of the system at time $t$, and we follow in this tradition here. For any time point $t$ in $\mathbb{R}_{\geq 0}$, the projector variable $X_t$ is the $\mathscr{X}$-valued variable that projects the path $\omega$ on its state at time $t$, defined by

$$X_t(\omega) \coloneqq \omega(t) \quad \text{for all } \omega \in \Omega. \tag{3.3}$$

### Events that depend on a single time point

Let us start with the most basic case of events that only depend on the state at a single time point. For any time point $t$ in $\mathbb{R}_{\geq 0}$ and any state $x$ in $\mathscr{X}$, we denote the event 'the state at time $t$ is $x$' by

$$\{X_t = x\} := \{\omega \in \Omega \colon \omega(t) = x\}. \tag{3.4}$$

More generally, for any subset $B$ of $\mathscr{X}$, we define the event 'the state at time $t$ belongs to $B$' as

$$\{X_t \in B\} := \{\omega \in \Omega \colon \omega(t) \in B\}.$$

Note that in both cases, the projector variable $X_t$ allows us to easily denote such events. Furthermore, it is easy to see that

$$\{X_t \in B\} = \bigcup_{x \in B} (X_t = x),$$

where by the definition of the union of a family of sets, the empty union is equal to the empty set.

**Joseph's Example 3.8.** In our formalism, the event 'Joseph's machine displays heads at time $t$' is $\{X_t = \mathtt{H}\}$. Observe that this is similar to how we denoted the event 'the $n$-th toss of Bruno's machine is heads' by $(X_n = \mathtt{H})$. For any two time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s < r$, we also have that

$$H_{[s,r]} = \bigcap_{t \in [s,r]} \{X_t = \mathtt{H}\}.$$

Again, this is similar to what we did in Bruno's Example 2.3332, where we saw that the event $H_n$, 'the first $n$ tosses of Bruno's machine are heads', is equal to $\bigcap_{k=1}^{n} \{X_k = \mathtt{H}\}$. ♪

### Sequences of time points and corresponding states

To simplify the notation regarding events that depend on more than a single time point, we will avail ourselves of the notational conventions used by Krak et al. (2017). A *sequence of time points* is a finite sequence of increasing non-negative real numbers, that is, a sequence $(t_1, \dots, t_n)$ in $\mathbb{R}_{\geq 0}$ such that $t_1 < \cdots < t_n$. We could denote a generic sequence of time points $(t_1, \dots, t_n)$ by $t_{1:n}$, but we choose to denote a generic sequence by $u$. In order not to deal with edge cases in the statements of definitions and results, we also consider the empty sequence $()$ as a sequence of time points and call all other sequences of time points non-empty. Thus, the set of sequences of time points is

$$\mathscr{U} := \{(t_1, \dots, t_n) \colon n \in \mathbb{Z}_{\geq 0}; t_1, \dots, t_n \in \mathbb{R}_{\geq 0}; t_1 < \cdots < t_n\} \tag{3.5}$$

and the set of all non-empty sequences of time points is

$$\mathscr{U}_{\neq ()} := \{(t_1, \dots, t_n) \colon n \in \mathbb{N}; t_1, \dots, t_n \in \mathbb{R}_{\geq 0}; t_1 < \cdots < t_n\}. \tag{3.6}$$

We denote the first and last time points of a non-empty sequence $u = (t_1, \dots, t_n)$ by $\min u \coloneqq t_1$ and $\max u \coloneqq t_n$, respectively; in order to conveniently deal with the edge case that $u$ is the empty sequence of time points $()$, we let $\min() \coloneqq 0 =: \max()$. For any two non-empty sequences of time points $u$ and $v$ in $\mathscr{U}_{\neq()}$, we write $u \preccurlyeq v$ whenever $v$ only contains time points that either belong to $u$ or that succeed the last time point of $u$; formally,

$$u \preccurlyeq v \Leftrightarrow (\forall t \in v)\ t \in u \text{ or } t \in [\max u, +\infty[. \tag{3.7}$$

Note that $\preccurlyeq$ is a partial order (see Schechter, 1997, Section 3.8) over $\mathscr{U}_{\neq()}$. Additionally, for all $u$ and $v$ in $\mathscr{U}_{\neq()}$, we write

$$u \prec v \Leftrightarrow \max u < \min v. \tag{3.8}$$

Note that $\prec$ is a strict (or irreflexive) partial order over $\mathscr{U}_{\neq()}$; in contrast to what our notation suggests, $\prec$ is *not* the strict partial order corresponding to $\preccurlyeq$. Out of convenience, we also impose that $() \preccurlyeq v$ and $() \prec v$ for all $v$ in $\mathscr{U}_{\neq()}$. This way, for any – possibly empty – sequence of time points $u$ in $\mathscr{U}$, we can define the sets $\mathscr{U}_{\succcurlyeq u} \coloneqq \{v \in \mathscr{U} \colon v \succcurlyeq u\}$ and $\mathscr{U}_{\succ u} \coloneqq \{v \in \mathscr{U}_{\neq()} \colon v \succ u\}$. Additionally, if $t$ is a time point in $\mathbb{R}_{\geq 0}$ and $u$ a sequence of time points, then we write $t \succ u$ whenever $(t) \succ u$.

Because a sequence of time points is an ordered set, we may use set-theoretic operations on sequences of time points, in the understanding that the result of these operation are sequences of time points – that is, ordered sets of real numbers – as well. For example, if $u$ and $v$ are sequences of time points, then $u \cup v$ denotes the sequence of time points that is made up of the time points in $u$ and $v$; similarly, $u \setminus v$ is the sequences that is made up of the time points in $u$ that do not belong to $v$.

Next, we consider tuples of states indexed by time sequences. For a non-empty sequence of time points $u = (t_1, \dots, t_n)$, we consider $n$-tuples of states, that is, elements $(x_{t_1}, \dots, x_{t_n})$ of $\mathscr{X}^n$. In order not to constantly refer to the length $|u|$ of a generic non-empty sequence of time points $u$, we write $\mathscr{X}_u$ instead of $\mathscr{X}^{|u|}$. Similarly, we denote an element of $\mathscr{X}_u$ – that is, a generic $|u|$-tuple of states – by $x_u$ instead of $(x_{t_1}, \dots, x_{t_n})$ or $(x_t)_{t \in u}$. For any $t$ in $u = (t_1, \dots, t_n)$, we then let $x_t$ be the corresponding component of $x_u = (x_{t_1}, \dots, x_{t_n})$; similarly, for any subsequence $v$ of $u$, we let $x_v \coloneqq (x_t)_{t \in v}$ be the corresponding subtuple of $x_u$. For ease of notation, we extend these conventions to the empty sequence of time points as well; if $u$ is the empty sequence of time points $()$, then $x_u = x_{()}$ is the empty tuple, which we will denote by $\diamond$, and $\mathscr{X}_u = \mathscr{X}_{()}$ is the singleton $\{\diamond\}$ containing this empty tuple.

Fix any sequence of time points $u$ in $\mathscr{U}$. For any path $\omega$ in $\tilde{\Omega}$, we denote the restriction of $\omega$ to $u$ by

$$\omega|_u \coloneqq \big(\omega(t)\big)_{t \in u} = \big(\omega(t_1), \dots, \omega(t_n)\big), \tag{3.9}$$

where we enumerate the time points in $u$ as $(t_1, \dots, t_n)$. Note that $\omega|_u$ is the empty tuple $\diamond$ whenever $u$ is the empty time sequence $()$. Additionally, we let $X_u := (X_t)_{t \in u} = (X_{t_1}, \dots, X_{t_n})$ be the $\mathscr{X}_u$-valued variable defined by

$$X_u(\omega) := \omega|_u \quad \text{for all } \omega \in \Omega. \tag{3.10}$$

**Events that depend on a finite number of time points**

With this notation, we can now formally define the event 'the states at the times in $u$ are equal to $x_u$' as

$$\{X_u = x_u\} := \{\omega \in \Omega \colon \omega|_u = x_u\}. \tag{3.11}$$

Similarly, for any subset $B$ of $\mathscr{X}_u$, we set

$$\{X_u \in B\} := \{\omega \in \Omega \colon \omega|_u \in B\}. \tag{3.12}$$

Note that

$$\{X_u \in B\} = \bigcup_{x_u \in B} \{X_u = x_u\} = \bigcup_{x_u \in B} \bigcap_{t \in u} \{X_t = x_t\}, \tag{3.13}$$

where by the definition of the intersection of a family of sets, the empty intersection is the entire set. Observe that in case $u$ is the empty sequence of time points $()$, $\{X_{()} = x_{()}\} = \Omega = \{X_{()} \in \mathscr{X}_{()}\}$.

Most authors call an event of the form $\{X_u \in B\}$ a *cylinder event* (see, for example, Billingsley, 1995, Section 36), but we will often also refer to it as a *finitary event* because it depends on the state of the system at a finite number of time points. For the same reason, we conclude that an event $\{X_u = x_u\}$ is determinable; conversely, it is easy to verify that any determinable event must have a (non-unique) representation of the form $\{X_u \in B\}$. In other words, every event in the set

$$\mathscr{F} := \big\{\{X_u \in B\} \colon u \in \mathscr{U}, B \subseteq \mathscr{X}_u\big\}^{3}$$

of finitary events is determinable, and vice versa.

**Conditioning events and the corresponding fields of events**

Let us fix the conditioning events. One option would be to use all the events of the form $\{X_u \in B\}$, but we follow (Krak et al., 2017) and only condition on the simpler events of the form $\{X_u = x_u\}$. Thus, we end up with the set

$$\mathscr{H} := \big\{\{X_u = x_u\} \colon u \in \mathscr{U}, x_u \in \mathscr{X}_u\big\} \tag{3.14}$$

of conditioning events. We can only use the events in $\mathscr{H}$ as conditioning events in case they are all non-empty, and the following result establishes that this is not an issue.

[3] Note the subtle difference between $\mathscr{F}$ and $\mathcal{F}$, the 'calligraphic capital F' of the Stix font family and Times font families, respectively.

**Corollary 3.9.** *Every conditioning event $\{X_u = x_u\}$ in $\mathscr{H}$ is non-empty.*

*Proof.* The statement is trivial in case $u$ is the empty sequence of time points $()$. When $u$ is not the empty time sequence, the statement follows immediately from Lemma 3.558. □

Fix any event $\{X_u = x_u\}$ in $\mathscr{H}$, any time point $t$ in $\mathbb{R}_{\geq 0}$ such that $t \succ u$ and any state $x$ in $\mathscr{X}$. Because

$$\{X_u = x_u\} \cap \{X_t = x\} = \{X_v = x_v\},$$

with $v := u \cup (t)$ and $x_t := x$, the event

$$\{X_u = x_u, X_t = x\} := \{X_u = x_u\} \cap \{X_t = x\} \tag{3.15}$$

again belongs to the set $\mathscr{H}$. We will often encounter events of this form. In these cases, it usually pays to think of $t$ as the current time point and $x$ as the current state of the system, and $\{X_u = x_u\}$ as an incomplete specification of the system's history – that is, the values of $\omega(s)$ for all time points $s < t$.

Next, we decide on the events that we should consider given a state history $\{X_u = x_u\}$ in $\mathscr{H}$. In case $\{X_u = x_u\} = \{X_{()} = x_{()}\} = \Omega$, we consider the set $\mathscr{F}$ of all determinable (or finitary) events. In case $\{X_u = x_u\} \neq \Omega$, or equivalently, $u \neq ()$, things are a bit more involved. In this work, we choose to interpret the conditioning event $\{X_u = x_u\}$ as 'we have observed the state of the system at the time points in $u$, and only at those time points'. Under this interpretation, it is no longer possible to observe the state of the system at previous time points – that is, the time points in $[0, \max u] \setminus u$. This way, given the state history $\{X_u = x_u\}$, we can only decide upon bets that depend on the state of the system at time points that either belong to $u$ or that succeed the time points in $u$. More concretely, we should only consider those events in $\mathcal{F}$ that have a representation of the form $\{X_v \in B\}$ with $v \succcurlyeq u$ and $B \subseteq \mathscr{X}_v$. For any sequence of time points $u$ in $\mathscr{U}$, we collect these events in

$$\mathscr{F}_u := \big\{\{X_v \in B\} \colon v \in \mathscr{U}_{\succcurlyeq u}, B \subseteq \mathscr{X}_v\big\}. \tag{3.16}$$

Note that $\mathscr{F}_{()} = \mathscr{F}$. Quite conveniently, the set $\mathscr{F}_u$ turns out to constitute a field of events.

**Lemma 3.10.** *For any sequence of time points $u$ in $\mathscr{U}$, the set $\mathscr{F}_u$ is a field of events.*

*Proof.* We need to verify that $\mathscr{F}_u$ satisfies (F1)32–(F3)32, the three requirements of Definition 2.3232.

First, we observe that (F1)32 holds, as the sure event $\Omega$ belongs to $\mathscr{F}_u$. We let $t = \max u$, so $t \succcurlyeq u$ by construction and

$$\mathscr{F}_u \ni \{X_t \in \mathscr{X}\} = \{X_{()} \in \mathscr{X}_{()}\} = \Omega.$$

Similarly, $\varnothing$ belongs to $\mathscr{F}_u$ because $\{X_t \in \varnothing\} = \{X_{()} \in \varnothing\} = \varnothing$. Furthermore, we see that (F2)$_{32}$ is trivially satisfied for $A = \Omega$ and $A = \varnothing$, and so is (F3)$_{32}$ if one of the two events is either the sure or the impossible event.

Next, we fix any $A = \{X_{v_1} \in B_1\}$ and $B = \{X_{v_2} \in B_2\}$ in $\mathscr{F}_u$ such that $\varnothing \neq A \neq \Omega$ and $\varnothing \neq B \neq \Omega$, and verify that (F2)$_{32}$ and (F3)$_{32}$ hold. To that end, we recall that $\{X_{()} \in \mathscr{X}_{()}\} = \Omega$ and $\{X_{()} \in \varnothing\} = \varnothing$, so the conditions on $A$ and $B$ imply that $v_1 \neq ()$ and $v_2 \neq ()$.

In order to verify (F2)$_{32}$, we observe that

$$\begin{aligned} A^{\mathrm{c}} &= \Omega \setminus \{\omega \in \Omega \colon \omega|_{v_1} \in B_1\} = \{\omega \in \Omega \colon \omega|_{v_1} \notin B_1\} \\ &= \{\omega \in \Omega \colon \omega|_{v_1} \in B_1^{\mathrm{c}}\} = \{X_{v_1} \in B_1^{\mathrm{c}}\}, \end{aligned}$$

with $B_1^{\mathrm{c}} = \mathscr{X}_{v_1} \setminus B_1$. As $B_1^{\mathrm{c}} \subseteq \mathscr{X}_{v_1}$, it follows from Equation (3.16) that $A^{\mathrm{c}} = \{X_{v_1} \in B_1^{\mathrm{c}}\}$ belongs to $\mathscr{F}_u$, as required.

To verify (F3)$_{32}$, we observe that

$$A \cup B = \{\omega \in \Omega \colon \omega|_{v_1} \in B_1\} \cup \{\omega \in \Omega \colon \omega|_{v_2} \in B_2\}.$$

Let $v_3$ be the ordered union of $v_1$ and $v_2$. Note that $v_1 \subseteq v_3$ and $v_2 \subseteq v_3$, and that $v_3 \succcurlyeq u$ because both $v_1 \succcurlyeq u$ and $v_2 \succcurlyeq u$. Now let

$$B_3 \coloneqq \{x_{v_3} \in \mathscr{X}_{v_3} \colon x_{v_1} \in B_1\} \cup \{x_{v_3} \in \mathscr{X}_{v_3} \colon x_{v_2} \in B_2\}.$$

It then follows that

$$A \cup B = \{\omega \in \Omega \colon \omega|_{v_1} \in B_1\} \cup \{\omega \in \Omega \colon \omega|_{v_2} \in B_2\} = \{\omega \in \Omega \colon \omega|_{v_3} \in B_3\} = \{X_{v_3} \in B_3\}.$$

Because $v_3 \succcurlyeq u$ and $B_3 \subseteq \mathscr{X}_{v_3}$, we may conclude from this equality that $A \cup B$ belongs to $\mathscr{F}_u$, as required. □

A convenient property of these finitary events is that if we have any number of them, they all have a representation using the same finite sequence of time points.

**Lemma 3.11.** *Consider a sequence of time points $u$ in $\mathscr{U}$. Then for all natural numbers $n$ and all finitary events $A_1, \dots, A_n$ in $\mathscr{F}_u$, there are a sequence of time points $v$ in $\mathscr{U}_{\succ u}$ and subsets $B_1, \dots, B_n$ of $\mathscr{X}_{u \cup v}$ such that*

$$A_k = \{X_{u \cup v} \in B_k\} \quad \textit{for all } k \in \{1, \dots, n\}.$$

*Proof.* By construction of $\mathscr{F}_u$, for every $k$ in $\{1, \dots, n\}$ there is a sequence of time points $v_k$ in $\mathscr{U}$ such that $v_k \succcurlyeq u$ and a subset $B'_k$ of $\mathscr{X}_{v_k}$ such that $A_k = \{X_{v_k} \in B'_k\}$. Let $v$ be the ordered sequence of time points that consists of those time points in $v_1$, …, $v_n$ that do not belong to $u$ as well; if $v$ turns out to be non-empty, then we choose an arbitrary time point $t$ in $\mathbb{R}_{\geq 0}$ such that $t > u$ and set $v \coloneqq (t)$. Thus, by construction, $v$ is non-empty and $v > u$. For any $k$ in $\{1, \dots, n\}$, we also let

$$B_k \coloneqq \{x_{u \cup v} \in \mathscr{X}_{u \cup v} \colon x_{v_k} \in B'_n\}.$$

The statement now follows if we see that, for all $k$ in $\{1, \dots, n\}$,

$$A_k = \{X_{v_k} \in B'_k\} = \left\{\omega \in \Omega \colon \omega|_{v_k} \in B'_k\right\} = \left\{\omega \in \Omega \colon \omega|_{u \cup v} \in B_k\right\} = \{X_{u \cup v} \in B_k\}. \quad \square$$

### 3.1.3 Jump processes as coherent conditional probabilities

Before we can finally state the formal definition of a jump process, we need to introduce a last bit of notation. For any state space $\mathscr{X}$, we define the domain

$$\mathscr{D}_{\mathscr{X}} := \big\{(A \,|\, X_u = x_u) \colon u \in \mathscr{U}, x_u \in \mathscr{X}_u, A \in \mathscr{F}_u\big\}. \tag{3.17}$$

Again, whenever the state space $\mathscr{X}$ is clear from the context, we write $\mathscr{D}$ instead of $\mathscr{D}_{\mathscr{X}}$. In Eq. (3.17), we have chosen to write $(A \,|\, X_u = x_u)$ instead of the more correct $\big(A, \{X_u = x_u\}\big)$. We do this not only because this makes the notation less obfuscated, but also because this allows us to unambiguously use notation like $(A \,|\, X_u = x_u, X_t = x)$, where we use the notational convention introduced in Eq. $(3.15)_{63}$. In the remainder, we will silently extend this convenient notation to general intersections of representations of finitary events. For example, if $u$ is a sequence of time points in $\mathscr{U}$ and $\{X_{v_1} \in B_1\}, \dots, \{X_{v_n} \in B_n\}$ are events in $\mathscr{F}_u$, then it will be convenient to write

$$\{X_{v_1} \in B_1, \dots, X_{v_n} \in B_n\} := \{X_{v_1} \in B_1\} \cap \dots \cap \{X_{v_n} \in B_n\}.$$

Because $\mathscr{F}_u$ is a field due to Lemma $3.10_{63}$, this event again belongs to $\mathscr{F}_u$.

Recall from Corollary $3.9_{63}$ that every state history $\{X_u = x_u\}$ in $\mathscr{H}$ is non-empty. Therefore, $\mathscr{D}_{\mathscr{X}}$ is a subset of $\mathcal{P}(\Omega_{\mathscr{X}}) \times \mathcal{P}(\Omega_{\mathscr{X}})_{\supset\emptyset}$, so $\mathscr{D}_{\mathscr{X}}$ can be the domain of a coherent conditional probability. This brings us to the following definition, taken from (Krak et al., 2017, Definition 4.3).

**Definition 3.12.** A *jump process $P$* with state space $\mathscr{X}$ is a coherent conditional probability on $\mathscr{D}_{\mathscr{X}}$. We denote the collection of all jump processes with state space $\mathscr{X}$ by $\mathbb{P}_{\mathscr{X}}$, or simply $\mathbb{P}$ whenever the state space is clear from the context.

Because a jump process is a coherent conditional probability, we can use all the related machinery that we introduced in Section $2.4.1_{40}$. Specifically interesting is that we can define a coherent conditional probability on a *smaller* domain $\mathscr{D}'$, and then use Theorem $2.54_{45}$ to extend it to a coherent conditional probability on $\mathscr{D}$, that is, a jump process. Let us illustrate this with our running example.

**Joseph's Example 3.13.** Cecilia, a colleague of Bruno and Joseph's, is breaking Joseph's heart. She has managed to shake his confidence by letting him know that she is convinced that the machine will always display heads – or, equivalently, she does not believe that the machine will ever display tails. We now construct a coherent conditional probability that captures these beliefs of hers.

Let $\omega_{\mathtt{H}}$ be the càdlàg path that is always $\mathtt{H}$, so $\omega_{\mathtt{H}}(t) := \mathtt{H}$ for all $t$ in $\mathbb{R}_{\geq 0}$. Consider the real-valued map $P'_{\mathtt{H}}$ on

$$\mathscr{D}_{\mathtt{H}} := \big\{(A \,|\, X_u = x_u) \in \mathscr{D} \colon \omega_{\mathtt{H}} \in \{X_u = x_u\}\big\}$$

that is defined for all $(A \mid X_u = x_u)$ in $\mathscr{D}_{\mathtt{H}}$ by

$$P'_{\mathtt{H}}(A \mid X_u = x_u) := \begin{cases} 1 & \text{if } \omega_{\mathtt{H}} \in A, \\ 0 & \text{otherwise.} \end{cases}$$

To check that $P'_{\mathtt{H}}$ is a coherent conditional probability on $\mathscr{D}_{\mathtt{H}}$, we fix a natural number $n$, some real numbers $\mu_1, \dots, \mu_n$ and some $(A_1 \mid C_1), \dots, (A_n \mid C_n)$ in $\mathscr{D}_{\mathtt{H}}$, and verify that

$$\max\left\{ \sum_{k=1}^{n} \mu_k \mathbb{I}_{C_k}(\omega)\big(\mathbb{I}_{A_k}(\omega) - P'_{\mathtt{H}}(A_k \mid C_k)\big) \colon \omega \in \bigcup_{k=1}^{n} C_k \right\} \geq 0. \tag{3.18}$$

To this end, we observe that, by construction of $P'_{\mathtt{H}}$, for any $k$ in $\{1, \dots, n\}$,

$$\mu_k \mathbb{I}_{C_k}(\omega_{\mathtt{H}})\big(\mathbb{I}_{A_k}(\omega_{\mathtt{H}}) - P'_{\mathtt{H}}(A_k \mid C_k)\big) = \mu_k(\mathbb{I}_{A_k}(\omega_{\mathtt{H}}) - P'_{\mathtt{H}}(A_k \mid C_k)) = 0.$$

Because $\omega_{\mathtt{H}}$ is an element of $C_k$ for all $k$ in $\{1, \dots, n\}$, this verifies Eq. (3.18).

Because $P'_{\mathtt{H}}$ is a coherent conditional probability on $\mathscr{D}_{\mathtt{H}}$, it follows from Theorem 2.54$_{45}$ that there is at least one coherent conditional probability $P_{\mathtt{H}}$ on $\mathscr{D}$ that extends $P'_{\mathtt{H}}$. By construction, any such extension $P_{\mathtt{H}}$ is a jump process with state space $\mathscr{X} = \{\mathtt{H}, \mathtt{T}\}$. ♭

A similar constructive method also comes in handy in the following example, which will be essential in Joseph's Example 3.43$_{92}$ further on.

**Joseph's Example 3.14.** Drawing inspiration from Krak et al. (2017, Example 4.3), we make the following claim: for any $[0,1]$-valued function $\phi$ on $\mathbb{R}_{>0}$, there is a jump process $P_\phi$ such that

$$P_\phi(X_t = x \mid X_0 = x) = \phi(t) \quad \text{for all } x \in \mathscr{X}, t \in \mathbb{R}_{>0}. \tag{3.19}$$

To verify this claim, we consider the real-valued map $P'_\phi$ on

$$\mathscr{D}_\phi := \big\{(X_t = x \mid X_0 = x) \colon x \in \mathscr{X}, t \in \mathbb{R}_{>0}\big\} \subseteq \mathscr{D},$$

defined for all $(X_t = x \mid X_0 = x)$ in $\mathscr{D}_\phi$ by

$$P'_\phi(X_t = x \mid X_0 = x) := \phi(t).$$

As in Joseph's Example 3.13$_\curvearrowleft$, we set out to check that $P'_\phi$ is a coherent conditional probability on $\mathscr{D}_\phi$. To verify that $P'_\phi$ satisfies the condition in Definition 2.51$_{44}$, we fix a natural number $n$, real numbers $\mu_1, \dots, \mu_n$ and some $(X_{t_1} = x_1 \mid X_0 = x_1), \dots, (X_{t_n} = x_n \mid X_0 = x_n)$ in $\mathscr{D}_\phi$. We have to show that

$$\max\left\{ \sum_{k=1}^{n} \mu_k \mathbb{I}_{x_k}\big(\omega(0)\big)\Big(\mathbb{I}_{x_k}\big(\omega(t_k)\big) - P'_\phi(X_{t_k} = x_k \mid X_0 = x_k)\Big) \colon \omega \in C \right\} \geq 0, \tag{3.20}$$

with $C \coloneqq \bigcup_{k=1}^{n} \{X_0 = x_k\}$. Define the index sets

$$\mathcal{K}_{\mathtt{H}} \coloneqq \{k \in \{1,\dots,n\} \colon x_k = \mathtt{H}\} \quad \text{and} \quad \mathcal{K}_{\mathtt{T}} \coloneqq \{k \in \{1,\dots,n\} \colon x_k = \mathtt{T}\},$$

and observe that, for all $\omega$ in $C$,

$$\begin{aligned}
\sum_{k=1}^{n} & \mu_k \mathbb{I}_{x_k}(\omega(0)) \Big( \mathbb{I}_{x_k}(\omega(t_k)) - P'_\phi(X_{t_k} = x_k \mid X_0 = x_k) \Big) \\
&= \sum_{k=1}^{n} \mu_k \mathbb{I}_{x_k}(\omega(0)) \Big( \mathbb{I}_{x_k}(\omega(t_k)) - \phi(t_k) \Big) \\
&= \sum_{x \in \mathscr{X}} \sum_{k \in \mathcal{K}_x} \mu_k \mathbb{I}_{x}(\omega(0)) \Big( \mathbb{I}_{x}(\omega(t_k)) - \phi(t_k) \Big). \qquad (3.21)
\end{aligned}$$

Clearly, we may assume without loss of generality that for $x$ in $\mathscr{X} = \{\mathtt{H}, \mathtt{T}\}$ and $k$, $\ell$ in $\mathcal{K}_x$ such that $k \neq \ell$, $t_k \neq t_\ell$; otherwise, we simply add together the corresponding terms in the second summation.

Fix any state $x$ in $\mathscr{X}$ such that the corresponding index set $\mathcal{K}_x$ is non-empty, and let $y$ be the other state in the binary state space $\mathscr{X}$. We let $u \coloneqq (0) \cup (t_k)_{k \in \mathcal{K}_x}$ be the sequence of time points that starts with 0 and continues with the time points in $\{t_k \colon k \in \mathcal{K}_x\}$. Furthermore, we divide $\mathcal{K}_x$ into $\mathcal{K}_x^{\geq 0} \coloneqq \{k \in \mathcal{K}_x \colon \mu_k \geq 0\}$ and $\mathcal{K}_x^{<0} \coloneqq \mathcal{K}_x \setminus \mathcal{K}_x^{\geq 0}$, and let $x_u$ be the unique tuple of states in $\mathscr{X}_u$ such that $x_0 \coloneqq x$, $x_{t_k} \coloneqq x$ for all $k$ in $\mathcal{K}_x^{\geq 0}$ and $x_{t_k} \coloneqq y$ for all $k$ in $\mathcal{K}_x^{<0}$. By Lemma 3.5$_{58}$, there is a path $\omega^\star$ in $\Omega$ such that $\omega^\star|_u = x_u$. Because $x = \omega^\star(0) \neq y$ by construction, the path $\omega^\star$ belongs to $C$, and it is clear that

$$\sum_{k \in \mathcal{K}_y} \mu_k \mathbb{I}_{y}(\omega^\star(0)) \Big( \mathbb{I}_{y}(\omega^\star(t_k)) - \phi(t_k) \Big) = 0 \qquad (3.22)$$

and

$$\begin{aligned}
\sum_{k \in \mathcal{K}_x} & \mu_k \mathbb{I}_{x}(\omega^\star(0)) \Big( \mathbb{I}_{x}(\omega^\star(t_k)) - \phi(t_k) \Big) \\
&= \sum_{k \in \mathcal{K}_x^{\geq 0}} \mu_k \mathbb{I}_{x}(\omega^\star(0)) \Big( \mathbb{I}_{x}(\omega^\star(t_k)) - \phi(t_k) \Big) \\
&\qquad + \sum_{k \in \mathcal{K}_x^{<0}} \mu_k \mathbb{I}_{x}(\omega^\star(0)) \Big( \mathbb{I}_{x}(\omega^\star(t_k)) - \phi(t_k) \Big) \\
&= \sum_{k \in \mathcal{K}_x^{\geq 0}} \mu_k \big(1 - \phi(t_k)\big) + \sum_{k \in \mathcal{K}_x^{<0}} \mu_k \big(-\phi(t_k)\big) \geq 0, \qquad (3.23)
\end{aligned}$$

where for the inequality we used that $\phi$ is $[0,1]$-valued. Finally, it follows from Eqs. (3.21) to (3.23) that

$$\max\left\{ \sum_{k=1}^{n} \mu_k \mathbb{I}_{x_k}(\omega(0)) \Big( \mathbb{I}_{x_k}(\omega(t_k)) - P'_\phi(X_{t_k} = x_k \mid X_0 = x_k) \Big) \colon \omega \in C \right\} \geq 0,$$

as required.

Because $P'_\phi$ is a coherent conditional probability on $\mathscr{D}_\phi$, it follows from Theorem 2.54₄₅ that there is at least one coherent conditional probability $P_\phi$ on $\mathscr{D}$ that extends $P'_\phi$. By construction, any such extension $P_\phi$ is a jump process with state space $\mathscr{X} = \{\mathtt{H}, \mathtt{T}\}$ that satisfies Eq. (3.19)₆₆, which is what we needed to verify. ♬

**The conditional expectation of determinable variables**

Note that the domain $\mathscr{D}$ of a jump process $P$ is a structure of fields. As we have seen in Section 2.4.2₄₆, this implies that the jump process $P$ – that is, the coherent conditional probability $P$ on $\mathscr{D}$ – induces a conditional expectation operator $E_P$ on

$$\mathbb{JS} := \mathbb{CS}(\mathscr{D}) = \big\{(f \,|\, X_u = x_u) \colon u \in \mathscr{U}, x_u \in \mathscr{X}_u, f \in \mathbb{S}(\mathscr{F}_u)\big\}, \tag{3.24}$$

where we choose to write $(f \,|\, X_u = x_u)$ instead of $\big(f, \{X_u = x_u\}\big)$.

The domain $\mathbb{JS}$ contains precisely those combinations of (determinable) gambles and conditioning events for which a betting interpretation makes sense. To see why this is the case, we recall from Section 3.1.2₅₈ that we consider a gamble to be determinable if and only if it depends on the state of the system at a finite number of time points. For any sequence of time points $v$ in $\mathscr{U}$ and any gamble $f$ on $\mathscr{X}_v$, we write the functional composition $f \circ X_v$[4] as $f(X_v)$; in other words, $f(X_v)$ is the gamble on $\Omega$ that maps any càdlàg path $\omega$ in $\Omega$ to $f(X_v(\omega)) = f(\omega|_v)$. Thus, a gamble $g$ on $\Omega$ is determinable if and only if there is a sequence of time points $v$ in $\mathscr{U}$ and a gamble $f$ on $\mathscr{X}_v$ such that $g = f(X_v)$.

Fix some state history $\{X_u = x_u\}$ in $\mathscr{H}$. Due to our betting interpretation and our interpretation of $\{X_u = x_u\}$, we should only consider a determinable gamble $g$ if it has a representation of the form $f(X_v)$ with $v \succcurlyeq u$. Observe that

$$f(X_v) = \sum_{y_v \in \mathscr{X}_v} f(y_v) \mathbb{I}_{\{X_v = y_v\}},$$

and that, for all $y_v$ in $\mathscr{X}$, $\{X_v = y_v\}$ belongs to $\mathscr{F}_u$ because $v \succcurlyeq u$. Thus, a determinable gamble $g = f(X_v)$ is clearly an $\mathscr{F}_u$-simple variable. The following result establishes the converse, in the sense that every $\mathscr{F}_u$-simple variable has a representation of the form $f(X_{u \cup v})$ with $u \cup v \succcurlyeq u$, and hence is determinable.

**Lemma 3.15.** *Let $u$ be a sequence of time points in $\mathscr{U}$. For any $\mathscr{F}_u$-simple variable $g$ in $\mathbb{S}(\mathscr{F}_u)$, there is a sequence $v$ in $\mathscr{U}_{>u}$ and a gamble $f$ on $\mathscr{X}_{u \cup v}$ such that $g = f(X_{u \cup v}) = f(X_u, X_v)$.*

[4] Throughout this dissertation, we use $\circ$ to denote functional composition: for any two functions $f \colon \operatorname{domain} f \to \operatorname{codomain} f$ and $g \colon \operatorname{domain} g \to \operatorname{codomain} g$ such that $\operatorname{codomain} g \subseteq \operatorname{domain} f$, $f \circ g := f\big(g(\bullet)\big)$ is the map from domain $g$ to codomain $f$ that maps any $x$ in domain $g$ to $f\big(g(x)\big)$.

*Proof.* Because $g$ is $\mathscr{F}_u$-simple, it has a representation of the form

$$g = \sum_{k=1}^{n} a_k \mathbb{I}_{A_k},$$

where $A_k$ is an event in $\mathscr{F}_u$ for all $k$ in $\{1,\dots,n\}$. Recall from Lemma 3.11$_{64}$ that there is a sequence of time points $v$ that succeeds $u$ and subsets $B_1,\dots,B_n$ of $\mathscr{X}_{u\cup v}$ such that $A_k = \{X_{u\cup v} \in B_k\}$ for all $k$ in $\{1,\dots,n\}$. Consequently,

$$g = \sum_{k=1}^{n} a_k \mathbb{I}_{\{X_{u\cup v}\in B_k\}} = \sum_{k=1}^{n} a_k \sum_{y_{u\cup v}\in B_k} \mathbb{I}_{\{X_{u\cup v}=y_{u\cup v}\}}.$$

From this, we infer that the statement holds for the gamble $f$ on $\mathscr{X}_{u\cup v}$ defined by

$$f(y_{u\cup v}) := \sum_{\substack{k\in\{1,\dots,n\}\\ y_{u\cup v}\in B_k}} a_k \quad \text{for all } y_{u\cup v} \in \mathscr{X}_{u\cup v}. \qquad \square$$

Due to the preceding, we can focus on conditional expectations of the form $E_P(f(X_u,X_v) \,|\, X_u = x_u)$ with $v \succ u$ or $E_P(f(X_v) \,|\, X_u = x_u)$ with $v \succcurlyeq u$. Because, by definition,

$$E_P(f(X_v) \,|\, X_u = x_u) = \sum_{y_v\in\mathscr{X}_v} f(y_v) P(X_v = y_v \,|\, X_u = x_u),$$

we take a closer look at probabilities of the form $P(X_v = y_v \,|\, X_u = x_u)$.

### 3.1.4 Properties of jump processes

Seeing that jump processes are coherent conditional probabilities, it follows from Corollary 2.55$_{45}$ that they satisfy (CP1)$_{41}$–(CP9)$_{42}$, the standard laws and properties of (conditional) probability. Some special cases of these properties will be of use to us later on, which is why we list them in the following result.

**Proposition 3.16.** *Consider a jump process $P$ and a state history $\{X_u = x_u\}$ in $\mathscr{H}$.*[5] *Then*

JP1. *for any $v$ in $\mathscr{U}_{\succcurlyeq u}$ and any subset $B$ of $\mathscr{X}_v$,*

$$P(X_v \in B \,|\, X_u = x_u) = P(X_w \in C \,|\, X_u = x_u),$$

*with $w := v \setminus u$ and $C := \big\{z_w \in \mathscr{X}_w : (\exists y_v \in B)\ y_{u\cap v} = x_{u\cap v}, y_w = z_w\big\}$;*

JP2. *for any $v$ in $\mathscr{U}_{\succcurlyeq u}$ and any subset $B$ of $\mathscr{X}_v$,*

$$P(X_v \in B \,|\, X_u = x_u) = \sum_{y_v\in B} P(X_v = y_v \,|\, X_u = x_u);$$

[5] Here and in the remainder, we use 'state history $\{X_u = x_u\}$ in $\mathscr{H}$' as a shorthand for '$u$ in $\mathscr{U}$ and $x_u$ in $\mathscr{X}_u$'.

JP3. *for any* $v = (t_1, \ldots, t_n)$ *in* $\mathscr{U}_{\succ u}$ *and any* $y_v$ *in* $\mathscr{X}_v$,

$$P(X_v = y_v \,|\, X_u = x_u) = \prod_{k=1}^{n} P(X_{t_k} = y_{t_k} \,|\, X_u = x_u, X_{t_{1:k-1}} = y_{t_{1:k-1}}),$$

*where we let* $t_{1:k-1} := (t_1, \ldots, t_{k-1})$ *for all* $k$ *in* $\{1, \ldots, n\}$;

JP4. *for any time points* $s, t$ *in* $\mathbb{R}_{\geq 0}$ *such that* $u \prec s < t$ *and any state* $x$ *in* $\mathscr{X}$,

$$P(X_t = x \,|\, X_u = x_u) = \sum_{y \in \mathscr{X}} P(X_t = x \,|\, X_u = x_u, X_s = y) P(X_s = y \,|\, X_u = x_u).$$

*Proof.* We prove the four listed properties one by one.

JP1. Observe that $\{X_u = x_u\} \subseteq (X_{u \cap v} = x_{u \cap v})$. Consequently, it follows from (CP1)$_{41}$ that $P(X_{u \cap v} = x_{u \cap v} \,|\, X_u = x_u) = 1$. Due to (CP9)$_{42}$, this implies that

$$P(X_v \in B \,|\, X_u = x_u) = P(\{X_v \in B\} \cap \{X_{u \cap v} = x_{u \cap v}\} \,|\, X_u = x_u).$$

The statement now follows because

$$\{X_v \in B\} \cap \{X_{u \cap v} = x_{u \cap v}\} = \{X_w \in C\},$$

with $w$ and $C$ as defined in the statement.

JP2. Note that for any $y_v$ and $z_v$ in $B$ such that $y_v \neq z_v$, $\{X_v = y_v\} \cap \{X_v = z_v\} = \emptyset$. The statement follows from this and repeated application of (CP3)$_{41}$.

JP3. Observe that $\{X_v = y_v\} = \{X_{t_{1:n-1}} = y_{t_{1:n-1}}, X_{t_n} = y_{t_n}\}$, and that

$$\{X_u = x_u, X_{t_{1:n-1}} = x_{t_{1:n-1}}\} = \{X_u = x_u\} \cap \{X_{t_{1:n-1}} = x_{t_{1:n-1}}\}$$

is a conditioning event in $\mathscr{H}$. Consequently, it follows from (CP4)$_{41}$ that

$$P(X_v = y_v \,|\, X_u = x_u) = P(X_{t_n} = y_{t_n} \,|\, X_u = x_u, X_{t_{1:n-1}} = x_{t_{1:n-1}}) P(X_{t_{1:n-1}} = x_{t_{1:n-1}} \,|\, X_u = x_u).$$

If we execute the same trick $n-1$ times, it is clear that we eventually obtain the statement.

JP4. Observe that $\{X_t = x\} = \bigcup_{y \in \mathscr{X}} \{X_s = y\} \cap \{X_t = x\}$, so repeated application of (CP3)$_{41}$ yields

$$P(X_t = x \,|\, X_u = x_u) = \sum_{y \in \mathscr{X}} P(\{X_s = y\} \cap \{X_t = x\} \,|\, X_u = x_u).$$

The property follows immediately from this and (CP4)$_{41}$. □

An immediate consequence of the properties (JP1)–(JP3) is that the conditional probability of any finitary event can be written in terms of a specific type of conditional probabilities.

**Lemma 3.17.** *Consider a conditioning event* $\{X_u = x_u\}$ *in* $\mathscr{H}$ *and an event* $A$ *in* $\mathscr{F}_u$. *Then there is a non-empty sequence* $v = (t_1, \dots, t_n)$ *that succeeds* $u$ *and a subset* $B$ *of* $\mathscr{X}_v$ *such that for any jump process* $P$,

$$P(A \,|\, X_u = x_u) = \sum_{y_v \in B} \prod_{k=1}^{n} P(X_{t_k} = y_{t_k} \,|\, X_u = x_u, X_{t_{1:k-1}} = y_{t_{1:k-1}}),$$

*where for any* $k$ *in* $\{1, \dots, n\}$, *we let* $t_{1:k-1} := (t_1 \dots, t_{k-1})$.

*Proof.* Recall from Lemma 3.11$_{64}$ that there is a non-empty sequence of time points $v$ with $v > u$ and a subset $B'$ of $\mathscr{X}_{u \cup v}$ such that $A = \{X_{u \cup v} \in B'\}$. Observe now that, due to (JP1)$_{69}$,

$$P(A \,|\, X_u = x_u) = P(X_{u \cup v} \in B' \,|\, X_u = x_u) = P(X_v \in B \,|\, X_u = x_u),$$

with $B := \{y_v \colon y_{u \cup v} \in B', y_u = x_u\} \subseteq \mathscr{X}_v$. To obtain the expression of the statement, we simply apply (JP2)$_{69}$ and (JP3)$_{\curvearrowleft}$. □

The properties (JP1)$_{69}$–(JP4)$_{\curvearrowleft}$ listed in Proposition 3.16$_{69}$ also have consequences for conditional expectations. Let us start with a more or less immediate consequence of (JP1)$_{69}$.

**Corollary 3.18.** *Let* $P$ *be a jump process. For any* $\{X_u = x_u\}$ *in* $\mathscr{H}$, *any* $v$ *in* $\mathscr{U}_{\succcurlyeq u}$ *and any gamble* $f$ *on* $\mathscr{X}_v$,

$$E_P(f(X_v) \,|\, X_u = x_u) = E_P(f(x_{u \cap v}, X_{v \setminus u}) \,|\, X_u = x_u).$$

*Proof.* Let $u' := u \cap v$ and $w := v \setminus u$. It follows from (JP1)$_{69}$ that, for any $y_v$ in $\mathscr{X}_v$,

$$P(X_v = y_v \,|\, X_u = x_u) = \begin{cases} P(X_w = y_w \,|\, X_u = x_u) & \text{if } y_{u'} = x_{u'}, \\ P(\emptyset \,|\, X_u = x_u) & \text{otherwise.} \end{cases}$$

Recall from (CP5)$_{42}$ that $P(\emptyset \,|\, X_u = x_u) = 0$. Using this, we see that

$$\begin{aligned} E_P(f(X_v) \,|\, X_u = x_u) &= \sum_{y_v \in \mathscr{X}_v} f(y_v) P(X_v = y_v \,|\, X_u = x_u) \\ &= \sum_{y_w \in \mathscr{X}_w} f(x_{u'}, y_w) P(X_w = y_w \,|\, X_u = x_u) \\ &= E_P(f(x_{u'}, X_w) \,|\, X_u = x_u), \end{aligned}$$

as stated. □

### The law of iterated expectations

A more important consequence of Proposition 3.16$_{69}$ is that we can establish a law of iterated expectations similar to that of Eq. (2.23)$_{39}$. In order to facilitate an elegant statement of this law, we define the gamble

$$E_P(f(X_v) \,|\, X_u) := \sum_{x_u \in \mathscr{X}_u} E_P(f(X_v) \,|\, X_u = x_u) \mathbb{I}_{(X_u = x_u)}$$

for all $u$ in $\mathscr{U}$, all $v$ in $\mathscr{U}_{\neq()}$ such that $v \succcurlyeq u$ and all $f$ in $\mathbb{G}(\mathscr{X}_v)$. Observe that $E_P(f(X_v)\,|\,X_u)$ only depends on the state of the system at the time points in $u$, so it is an $\mathscr{F}_u$-simple variable.

**Theorem 3.19.** *Let $P$ be a jump process, and fix some $u$ in $\mathscr{U}$. Then for all $v$ and $w$ in $\mathscr{U}_{\neq()}$ such that $u \prec v$ and $u \cup v \preccurlyeq w$, $f$ in $\mathbb{G}(\mathscr{X}_w)$ and $x_u$ in $\mathscr{X}_u$,*

$$E_P(f(X_w)\,|\,X_u = x_u) = E_P\big(E_P(f(X_w)\,|\,X_{u\cup v})\,\big|\,X_u = x_u\big).$$

*Proof.* First, we use (JP1)$_{69}$ and (CP5)$_{42}$, to yield

$$\begin{aligned}
&E_P\big(E_P(f(X_w)\,|\,X_{u\cup v})\,\big|\,X_u = x_u\big)\\
&\quad= \sum_{y_{u\cup v}\in\mathscr{X}_{u\cup v}} P(X_{u\cup v} = y_{u\cup v}\,|\,X_u = x_u)E_P(f(X_w)\,|\,X_{u\cup v} = y_{u\cup v})\\
&\quad= \sum_{y_v\in\mathscr{X}_v} P(X_v = y_v\,|\,X_u = x_u)E_P(f(X_w)\,|\,X_u = x_u, X_v = y_v).
\end{aligned}$$

From this and Corollary 3.18$_{\frown}$, we infer that

$$\begin{aligned}
&E_P\big(E_P(f(X_w)\,|\,X_{u\cup v})\,\big|\,X_u = x_u\big)\\
&\quad= \sum_{y_v\in\mathscr{X}_v} P(X_v = y_v\,|\,X_u = x_u)E_P(f(x_{u'}, y_{v'}, X_{w'})\,|\,X_u = x_u, X_v = y_v),
\end{aligned}$$

where we let $u' := u \cap w$, $v' := v \cap w$ and $w' := w \setminus (u \cup v)$. Next, we use (CP4)$_{41}$, to yield

$$\begin{aligned}
&E_P\big(E_P(f(X_w)\,|\,X_{u\cup v})\,\big|\,X_u = x_u\big)\\
&\quad= \sum_{y_v\in\mathscr{X}_v} P(X_v = y_v\,|\,X_u = x_u)\\
&\qquad\qquad \sum_{z_{w'}\in\mathscr{X}_{w'}} f(x_{u'}, y_{v'}, z_{w'})P(X_{w'} = z_{w'}\,|\,X_u = x_u, X_v = y_v)\\
&\quad= \sum_{y_v\in\mathscr{X}_v}\sum_{z_{w'}\in\mathscr{X}_{w'}} f(x_{u'}, y_{v'}, z_{w'})P(X_v = y_v, X_{w'} = z_{w'}\,|\,X_u = x_u).
\end{aligned}$$

Because $v$ can be split into $v'$ and $(v \setminus v')$, it follows from this that

$$\begin{aligned}
&E_P\big(E_P(f(X_w)\,|\,X_{u\cup v})\,\big|\,X_u = x_u\big)\\
&\quad= \sum_{z_{w'}\in\mathscr{X}_{w'}}\sum_{y_{v'}\in\mathscr{X}_{v'}} f(x_{u'}, y_{v'}, z_{w'})\\
&\qquad\qquad \sum_{y_{v\setminus v'}\in\mathscr{X}_{v\setminus v'}} P(X_{v'} = y_{v'}, X_{v\setminus v'} = y_{v\setminus v'}, X_{w'} = z_{w'}\,|\,X_u = x_u)\\
&\quad= \sum_{z_{w'}\in\mathscr{X}_{w'}}\sum_{y_{v'}\in\mathscr{X}_{v'}} f(x_{u'}, y_{v'}, z_{w'})P(X_{v'} = y_{v'}, X_{v\setminus v'} \in \mathscr{X}_{v\setminus v'}, X_{w'} = z_{w'}\,|\,X_u = x_u)\\
&\quad= \sum_{z_{w'}\in\mathscr{X}_{w'}}\sum_{y_{v'}\in\mathscr{X}_{v'}} f(x_{u'}, y_{v'}, z_{w'})P(X_{v'} = y_{v'}, X_{w'} = z_{w'}\,|\,X_u = x_u),
\end{aligned}$$

where the second equality holds due to (JP2)69. Because $v' \cup w' = w \setminus u$, it follows immediately that

$$E_P\big(E_P(f(X_w)\,|\,X_{u\cup v})\,\big|\,X_u = x_u\big) = \sum_{y_{w\setminus u}\in\mathscr{X}_{w\setminus u}} f(x_{u'}, y_{w\setminus u})P(X_{w\setminus u} = y_{w\setminus u}\,|\,X_u = x_u)$$
$$= E_P(f(x_{u'}, X_{w\setminus u})\,|\,X_u = x_u) = E_P(f(X_w)\,|\,X_u = x_u),$$

where the final equality holds due to Corollary 3.1871. □

### 3.1.5 The initial and transition probabilities

From Lemma 3.1771, we learn that the conditional probabilities of some specific events seem to play an important role. This includes probabilities of the form

$$P(X_r = y\,|\,X_u = x_u, X_t = x),$$

with $u \prec t < r$. These probabilities are customarily called the *transition probabilities*. Additionally, the *initial probabilities* of the form

$$P(X_0 = x) = P(X_0 = x\,|\,\Omega)$$

also play an essential role. The following result establishes that a jump process is uniquely determined by its initial and transition probabilities.

**Proposition 3.20.** *Two jump processes $P_1$ and $P_2$ with the same state space $\mathscr{X}$ are equal if and only if*

$$P_1(X_0 = x) = P_2(X_0 = x) \quad \textit{for all } x \in \mathscr{X}$$

*and, for all $u$ in $\mathscr{U}$, $t$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $u \prec t < r$ and $x_u$ in $\mathscr{X}_u$,*

$$P_1(X_r = y\,|\,X_u = x_u, X_t = x) = P_2(X_r = y\,|\,X_u = x_u, X_t = x) \quad \textit{for all } x, y \in \mathscr{X}.$$

*Proof.* That the two conditions of the statement are necessary is obvious. To verify that these conditions are also sufficient, we observe that the two jump processes $P_1$ and $P_2$ are equal if and only if

$$P_1(A\,|\,X_u = x_u) = P_2(A\,|\,X_u = x_u) \quad \text{for all } (A\,|\,X_u = x_u) \in \mathscr{D}.$$

Hence, we fix any $\{X_u = x_u\}$ in $\mathscr{H}$ and $A$ in $\mathscr{F}_u$, and show that the equality above holds. By Lemma 3.1771, there is a non-empty sequence $v = (t_1, \ldots, t_n)$ of time points that succeeds $u$ and a subset $B$ of $\mathscr{X}_v$ such that, with $\ell$ in $\{1,2\}$,

$$P_\ell(A\,|\,X_u = x_u) = \sum_{y_v\in B}\prod_{k=1}^{n} P_\ell(X_{t_k} = y_{t_k}\,|\,X_u = x_u, X_{t_{1:k-1}} = y_{t_{1:k-1}})$$

where, for any $k$ in $\{1,\ldots,n\}$, we let $t_{1:k-1} := (t_1, \ldots, t_{k-1})$. Consequently, it suffices to show that, for any $y_v$ in $B$ and $k$ in $\{1,\ldots,n\}$,

$$P_1(X_{t_k} = y_{t_k}\,|\,X_u = x_u, X_{t_{1:k-1}} = y_{t_{1:k-1}})$$
$$= P_2(X_{t_k} = y_{t_k}\,|\,X_u = x_u, X_{t_{1:k-1}} = y_{t_{1:k-1}}). \quad (3.25)$$

This equality follows immediately from the conditions of the statement, except for the edge case that $k = 1$, $u$ is the empty time sequence and $t_1 > 0$. Note that in this edge case, $(X_u = x_u, X_{v_1} = y_{v_1}) = \Omega$. Therefore, it then follows from (JP4)$_{70}$ that, with $\ell$ in $\{1, 2\}$,

$$P_\ell(X_{t_1} = y_{t_1} \mid X_u = x_u, X_{v_1} = y_{v_1}) = P_\ell(X_{t_1} = y_{t_1}) \\ = \sum_{y \in \mathscr{X}} P_\ell(X_{t_1} = y_{t_1} \mid X_0 = y) P_\ell(X_0 = y).$$

Since all these terms are equal for $P_1$ and $P_2$ due to the conditions of the statement, this verifies the equality in Eq. (3.25)$_\curvearrowleft$ for the edge case as well. □

While the initial and transition probabilities uniquely determine a jump process, it is important to stress that we cannot choose them arbitrarily. More precisely, it follows from Definition 2.51$_{44}$, Theorem 2.54$_{45}$ and Proposition 3.20$_\curvearrowleft$ that it is necessary and sufficient for the initial and transition probabilities to form a coherent conditional probability on (a subset of)

$$\big\{(X_0 = x \mid \Omega) \colon x \in \mathscr{X}\big\} \cup \big\{(X_r = y \mid X_u = x_u, X_t = x) \colon \\ u \in \mathscr{U}; x_u \in \mathscr{X}_u; t, r \in \mathbb{R}_{\geq 0}, u \prec t < r; x, y \in \mathscr{X}\big\}.$$

## 3.2 Markovian jump processes

Specifying *all* of the initial and transition probabilities in such a way that they form a coherent conditional probability is certainly non-trivial. One option to deal with this is to first only specify the probabilities of events in some subset $\mathscr{D}'$ of $\mathscr{D}$ for which checking coherence becomes easy, and then use Theorem 2.54$_{45}$ to extend this coherent conditional probability on $\mathscr{D}'$ to a – not necessarily unique – jump process. A second, more conventional option is to impose the following two simplifying assumptions on the transition probabilities.

First and foremost is the Markov property, which holds if the probability of being in a future state only depends on the current state, and not on the state history. For the formal classical formulation, we refer to (Chung, 1960, Section II.4) or (Doob, 1953, Section I.6); here we adhere to the formulation of Krak et al. (2017, Definition 5.1), which is tailored to our use of coherent conditional probabilities.

**Definition 3.21.** A jump process $P$ is *Markovian* – or, alternatively, has the *Markov property* – if for all time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and all state histories $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t < r$,

$$P(X_r = y \mid X_u = x_u, X_t = x) = P(X_r = y \mid X_t = x) \quad \text{for all } x, y \in \mathscr{X}.$$

We collect all Markovian jump processes with state space $\mathscr{X}$ in $\mathbb{P}^{\mathrm{M}}_{\mathscr{X}}$, and write $\mathbb{P}^{\mathrm{M}}$ whenever the state space is clear from the context.

The second simplifying assumption is that the transition probabilities are (time-)homogeneous, which is sometimes also referred to as stationary. Under this assumption, the transition probabilities only depend on the length of the time period between the current and the future time point, and not on the current and future time points themselves. Our formal definition is taken from (Krak et al., 2017, Definition 5.2).

**Definition 3.22.** A Markovian jump process $P$ is *homogeneous* if for all time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $t < r$,

$$P(X_r = y \mid X_t = x) = P(X_{r-t} = y \mid X_0 = x) \quad \text{for all } x, y \in \mathscr{X}.$$

We denote the set of all homogeneous Markovian jump processes with state space $\mathscr{X}$ by $\mathbb{P}^{\mathrm{HM}}_{\mathscr{X}}$, or simply $\mathbb{P}^{\mathrm{HM}}$ in case the state space is clear from the context.

### 3.2.1 Operators and semi-groups

To appreciate just how much these two properties simplify the transition probabilities, we need to introduce some additional mathematical machinery: operators.

#### The real vector space of gambles on $\mathscr{X}$

Recall from Section 2.1.313 that the set $\mathbb{G}(\mathscr{X})$ of all gambles on the state space $\mathscr{X}$ is a real vector space. Furthermore, we also recall that, for any subset $A$ of $\mathscr{X}$, its indicator $\mathbb{I}_A \colon \mathscr{X} \to \{0,1\}$ assumes 1 on $A$ and 0 elsewhere. In order to unburden our notation, we write $\mathbb{I}_x$ instead of $\mathbb{I}_{\{x\}}$, where $x$ is a state in the state space $\mathscr{X}$.

We equip the real vector space $\mathbb{G}(\mathscr{X})$ with the inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{G}(\mathscr{X})$, given by $\langle f, g \rangle \coloneqq \sum_{x \in \mathscr{X}} f(x) g(x)$ for all $f$ and $g$ in $\mathbb{G}(\mathscr{X})$. The inner product $\langle \cdot, \cdot \rangle$ induces the *Euclidean norm* $\|\bullet\|_2$, given by

$$\|f\|_2 \coloneqq \langle f, f \rangle^{1/2} = \left( \sum_{x \in \mathscr{X}} f(x)^2 \right)^{1/2} \quad \text{for all } f \in \mathbb{G}(\mathscr{X}),$$

but – as we will see later on – it is usually more fitting and convenient to use the *supremum norm* $\|\bullet\|_\infty$ (see Schechter, 1997, Section 23.3), defined as

$$\|f\|_\infty \coloneqq \max\{|f(x)| \colon x \in \mathscr{X}\} = \max|f| \quad \text{for all } f \in \mathbb{G}(\mathscr{X}). \tag{3.26}$$

Note that

$$\|f\|_\infty \leq \|f\|_2 \leq \sqrt{|\mathscr{X}|}\,\|f\|_\infty \quad \text{for all } f \in \mathbb{G}(\mathscr{X}),$$

so these two norms induce the same topology: convergence with respect to the supremum norm implies convergence with respect to the Euclidian norm, and vice versa.

It is well-known that $\mathbb{G}(\mathscr{X})$ is complete with respect to $\|\bullet\|_\infty$ – and also with respect to $\|\bullet\|_2$, for that matter – thus $\mathbb{G}(\mathscr{X})$ is a Banach space (see Schechter, 1997, Sections 22.8 and 22.11). Because we will always use the supremum norm $\|\bullet\|_\infty$ on $\mathbb{G}(\mathscr{X})$, we will henceforth simply write $\|\bullet\|$. We bestow the real vector space $\mathbb{G}(\mathscr{X})$ of gambles on $\mathscr{X}$ with the topology that is induced by (the metric induced by) the supremum norm $\|\bullet\|$, and every limit statement we make regarding gambles on $\mathscr{X}$ is with respect to the supremum norm.

Recall (see Schechter, 1997, Section 22.2) that a norm $\|\bullet\|_{\mathcal{V}} \colon \mathcal{V} \to \mathbb{R}_{\geq 0}$ on a real vector space $\mathcal{V}$ is a non-negative real-valued function that has the following three properties:

N1. $\|\mu f\|_{\mathcal{V}} = |\mu| \|f\|_{\mathcal{V}}$ for all $\mu$ in $\mathbb{R}$ and $f$ in $\mathcal{V}$,

N2. $\|f+g\|_{\mathcal{V}} \leq \|f\|_{\mathcal{V}} + \|g\|_{\mathcal{V}}$ for all $f$ and $g$ in $\mathcal{V}$,

N3. $\|f\|_{\mathcal{V}} = 0$ if and only if $f = 0$.

**Operators on** $\mathbb{G}(\mathscr{X})$

We will repeatedly use maps from $\mathbb{G}(\mathscr{X})$ to $\mathbb{G}(\mathscr{X})$, which we will call *operators (on* $\mathbb{G}(\mathscr{X})$*)*. One example of an operator is the identity operator $I$, defined by $If := f$ for all $f$ in $\mathbb{G}(\mathscr{X})$. An operator $M$ is *non-negatively homogeneous* if $M(\lambda f) = \lambda M f$ for all $f$ in $\mathbb{G}(\mathscr{X})$ and $\lambda$ in $\mathbb{R}_{\geq 0}$. It is easy to see that for any real number $\mu$ and any two non-negatively homogeneous operators $M$ and $N$ on $\mathbb{G}(\mathscr{X})$, $\mu M$, $M+N$ and $MN$ – which is to be interpreted as applying $M$ after $N$ – are non-negatively homogeneous operators as well. Consequently, the non-negatively homogeneous operators on $\mathbb{G}(\mathscr{X})$ constitute a linear subspace of the set of operators on $\mathbb{G}(\mathscr{X})$.

An operator $M$ is *super-additive* if $M(f+g) \geq Mf + Mg$ for all $f$ and $g$ in $\mathbb{G}(\mathscr{X})$, and *additive* if this relation holds with equality instead of inequality. A *linear operator* $M$ is an operator that is non-negatively homogeneous and additive; clearly, the identity operator $I$ is linear. Note that if $M$ is linear, then $M(\lambda f) = \lambda M f$ for negative real numbers $\lambda$ as well.

Let $n$ be the cardinality of $\mathscr{X}$. Given some basis $(g_1, \dots, g_n)$ for the $n$-dimensional real vector space $\mathbb{G}(\mathscr{X})$, any linear operator $M$ is represented by a unique square matrix (and vice versa). If we fix an ordering $x_1, \dots, x_n$ on the state space $\mathscr{X}$, an obvious basis for $\mathbb{G}(\mathscr{X})$ is $(\mathbb{I}_{x_1}, \dots, \mathbb{I}_{x_n})$, and the $(k,\ell)$-component of the square matrix that represents $M$ is $[M\mathbb{I}_{x_\ell}](x_k)$. We do not frequently need this matrix representation, but we do take inspiration from it and define the 'component' $M(x,y) := [M\mathbb{I}_y](x)$ for all $x$ an $y$ in $\mathscr{X}$. Note that a linear operator $M$ is completely defined by its 'components', because for all $f$ in $\mathbb{G}(\mathscr{X})$ and $x$ in $\mathscr{X}$,

$$[Mf](x) = \left[M\left(\sum_{y\in\mathscr{X}} f(y)\mathbb{I}_y\right)\right](x) = \sum_{y\in\mathscr{X}} f(y)[M\mathbb{I}_y](x) = \sum_{y\in\mathscr{X}} f(y)M(x,y),$$

where the second equality follows from the linearity of $M$.

The supremum norm $\|\bullet\|$ on $\mathbb{G}(\mathscr{X})$ induces an operator norm $\|\bullet\|_{\mathrm{op}}$ on the real vector space of non-negatively homogeneous operators,[6] defined for any non-negatively homogeneous operator $M$ on $\mathbb{G}(\mathscr{X})$ by

$$\|M\|_{\mathrm{op}} := \sup\big\{\|Mf\| : f \in \mathbb{G}(\mathscr{X}), \|f\| = 1\big\}. \tag{3.27}$$

Note that this definition is a straightforward generalisation of the definition of the operator norm for *linear* operators (see Schechter, 1997, Section 23.1) to *general* operators. It is well-known (see Schechter, 1997, Section 23.3) that for any linear operator $M$ on $\mathbb{G}(\mathscr{X})$,

$$\|M\|_{\mathrm{op}} = \max\left\{ \sum_{y\in\mathscr{X}} \big|M(x,y)\big| : x \in \mathscr{X} \right\}. \tag{3.28}$$

It follows almost immediately from Eq. (3.27) (see De Bock, 2017b, Appendix) that for all non-negatively homogeneous operators $M$ and $N$ on $\mathbb{G}(\mathscr{X})$,

N4. $\|Mf\| \leq \|M\|_{\mathrm{op}}\|f\|$ for all $f$ in $\mathbb{G}(\mathscr{X})$, and

N5. $\|MN\|_{\mathrm{op}} \leq \|M\|_{\mathrm{op}}\|N\|_{\mathrm{op}}$.

We bestow the real vector space of non-negatively homogeneous operators with the topology that is induced by (the metric induced by) the operator norm $\|\bullet\|_{\mathrm{op}}$, and every limit statement we make regarding non-negatively homogeneous operators is with respect to the operator norm. For instance, we write $\lim_{n\to+\infty} M_n = M$ if and only if $\lim_{n\to+\infty}\|M_n - M\|_{\mathrm{op}} = 0$. Note that in this case, due to (N4), $\lim_{n\to+\infty} M_n f = Mf$ for any gamble $f$ on $\mathscr{X}$, in the sense that $\lim_{n\to+\infty}\|M_n f - Mf\| = 0$.

### Semi-groups of operators

From its very beginning – see, for example, (Kolmogorov, 1931; Elfving, 1937; Doeblin, 1938; Fréchet, 1938) – the study of jump processes has been entwined with what we now call semi-groups. A (one-parameter) *semi-group* $(M_t)_{t\in\mathbb{R}_{\geq 0}}$ is a family of operators indexed by $\mathbb{R}_{\geq 0}$ such that

SG1. $M_0 = I$;

[6] We can extend this to the real vector space of all operators on $\mathbb{G}(\mathscr{X})$. In this more general case, for any operator $M$ on $\mathbb{G}(\mathscr{X})$ we define

$$\|M\|_{\mathrm{op}} := \sup\left\{ \frac{\|Mf\|}{\|f\|} : f \in \mathbb{G}(\mathscr{X}), f \neq 0 \right\},$$

which is equal to Eq. (3.27) whenever $M$ is non-negatively homogeneous. If we restrict ourselves to (the real vector space of) bounded operators – that is, those operators $M$ with $\|M\|_{\mathrm{op}} < +\infty$ – then $\|\bullet\|_{\mathrm{op}}$ is a semi-norm (see Schechter, 1997, Section 22.2) because it satisfies (N1)↶ and (N2)↶ but not necessarily (N3)↶. If we furthermore restrict ourselves to (the real vector space of) bounded operators with $M0 = 0$, then $\|\bullet\|_{\mathrm{op}}$ also satisfies (N3)↶, and thus is a norm.

SG2. $M_{t+r} = M_t M_r$ for all $t$ and $r$ in $\mathbb{R}_{\geq 0}$.

Property (SG2) is the crucial *semi-group property*. A semi-group $(M_t)_{t \in \mathbb{R}_{\geq 0}}$ is called *point-wise continuous* – sometimes also called *strongly continuous*, see (Dunford et al., 1958, Chapter VIII, Section 1) or (Engel et al., 2000, Chapter 1, Section 5) – if for all $t$ in $\mathbb{R}_{\geq 0}$ and $f$ in $\mathbb{G}(\mathscr{X})$,

SG3. $\lim_{r \searrow t} M_r f = M_t f$ and, if $t > 0$, $\lim_{s \nearrow t} M_s f = M_t f$.

A semi-group $(M_t)_{t \in \mathbb{R}_{\geq 0}}$ of non-negatively homogeneous operators is called *continuous* if for all $t$ in $\mathbb{R}_{\geq 0}$,

SG4. $\lim_{r \searrow t} M_r = M_t$ and, if $t > 0$, $\lim_{s \nearrow t} M_s = M_t$.

In this definition, we restrict us to semi-groups of non-negatively homogeneous operators because this ensures that the limit statement – which implicitly uses the operator norm – is well-defined. Note that continuity clearly implies point-wise continuity, due to (N4)↶. Even more, one can show that for a semi-group of linear operators on $\mathbb{G}(\mathscr{X})$, point-wise continuity is not only necessary but also sufficient for continuity because $\mathscr{X}$ is finite.

There is a vast body of literature that deals with semi-groups of operators on a general Banach (or Hilbert) space. Following Pazy (1983), we divide this work into two main topics: linear semi-groups and contraction semi-groups.

**Linear semi-groups and the operator exponential** The basic case of a semi-group $(M_t)_{t \in \mathbb{R}_{\geq 0}}$ of linear operators has been studied extensively. Most of the important results are collected in the seminal work of Hille et al. (1957); a more recent reference is the work of Engel et al. (2000), who not only give a clear overview but also provide interesting historical background. In the interest of conciseness, however, we limit ourselves to what we will need in the remainder. The linearity assumption is key to the following essential and well-known result; it was first established by Nathan (1935), Nagumo (1936), and Yosida (1936), but can also be found in (Hille et al., 1957, Theorem 9.6.1) or (Engel et al., 2000, Theorem 3.7).

**Proposition 3.23.** *Let $(M_t)_{t \in \mathbb{R}_{\geq 0}}$ be a continuous semi-group of linear operators on $\mathbb{G}(\mathscr{X})$. Then the generator*

$$G := \lim_{t \searrow 0} \frac{M_t - I}{t}$$

*is a linear operator on $\mathbb{G}(\mathscr{X})$, and it is the unique linear operator such that*

$$M_t = e^{tG} = \lim_{n \to +\infty} \left( I + \frac{t}{n} G \right)^n \quad \text{for all } t \in \mathbb{R}_{\geq 0}.$$

*Conversely, if $G$ is a linear operator on $\mathbb{G}(\mathscr{X})$, then for all $t$ in $\mathbb{R}_{\geq 0}$,*

$$e^{tG} = \lim_{n \to +\infty} \left( I + \frac{t}{n} G \right)^n$$

*is a linear operator on* $\mathbb{G}(\mathscr{X})$*, and* $(e^{tG})_{t\in\mathbb{R}_{\geq 0}}$ *is a continuous semi-group.*

Crucial to Proposition 3.23 is the (linear) operator exponential, which has some additional interesting properties. For a detailed discussion of these properties, we refer to (Van Loan, 1975) or (Norris, 1997, Sections 2.1 and 2.10). Here, we only repeat that for any linear operator $G$ on $\mathbb{G}(\mathscr{X})$ and any non-negative real number $t$,

$$e^{tG} := \lim_{n\to+\infty}\left(I+\frac{t}{n}G\right)^n = \sum_{n=0}^{+\infty}\frac{t^nG^n}{n!} \tag{3.29}$$

is a well-defined linear operator. Furthermore, the linear operator exponential is the unique solution to the following initial value problem (see Van Loan, 1975, Section 4).

**Proposition 3.24.** *Consider a linear operator* $G$*. Then* $(M_t)_{t\in\mathbb{R}_{\geq 0}} = (e^{tG})_{t\in\mathbb{R}_{\geq 0}}$ *is the unique solution to the (linear operator) initial value problem*

$$\lim_{r\to t}\frac{M_r - M_t}{r-t} = GM_t \quad \textit{with } M_0 = I, \tag{3.30}$$

*where, of course, we only take the limit from the right for* $t=0$*.*

**Contraction semi-groups** If we relax the linearity condition, things become much more difficult. Almost all of the available literature studies semi-groups of general – not necessarily non-negatively homogeneous – operators that are contractive. An operator $M\colon \mathscr{B}\to\mathscr{B}$ on a Banach space $\mathscr{B}$ with norm $\|\bullet\|_{\mathscr{B}}$ is a *contraction* – sometimes also called a *non-expansive operator* – if

$$\|Mf - Mg\|_{\mathscr{B}} \leq \|f-g\|_{\mathscr{B}} \quad \text{for all } f,g\in\mathscr{B}.$$

Because the operators in the semi-group need not be non-negatively homogeneous, we cannot use the operator norm $\|\bullet\|_{\text{op}}$ and therefore cannot impose continuity as in (SG4).[7] For this reason, one usually resorts to point-wise continuity in the sense of (SG3). To the best of our knowledge, Proposition 3.23 does not generalise from continuous semi-groups of linear operators to point-wise continuous semi-groups of contractions on a general (finite-dimensional) Banach space. Due to the weaker notion of continuity, the limit in the definition of the 'generator' need not exist on the entire Banach space $\mathscr{B}$. Furthermore, it is customary to construct a semi-group from a 'generator' through its 'resolvent' instead of through an exponential-like expression. We refer to (Crandall et al., 1969; Miyadera, 1992) for more details. In Section 3.3.3107 further on, we will encounter some contraction

[7] As we have previously noted near Eq. (3.27)77, instead of the set of non-negatively homogeneous operators we could consider the set of all bounded operators that map 0 to 0. To the best of our knowledge, this is not done in the literature.

semi-groups of non-linear operators on $\mathbb{G}(\mathscr{X})$ with some powerful additional properties. It is due to these properties that we nevertheless *can* generalise Proposition $3.23_{78}$ for this specific type of contraction semi-groups in Proposition $3.74_{114}$ and Theorem $3.75_{114}$ further on.

### 3.2.2 Transition and rate operators

When studying the transition probabilities of a single jump process, we always encounter two specific types of linear operators: transition operators and rate operators. For the reader who is already familiar with stochastic processes, we should remark that – given a suitable basis – transition and rate operators are represented by what is known as transition (or stochastic) matrices and rate (or intensity) matrices, respectively. In this dissertation, we prefer to use linear operators, which essentially goes back to (Whittle, 2000, Chapter 9, Section 2).

#### Transition operators

A transition operator is a linear operator on $\mathbb{G}(\mathscr{X})$ that is bounded below by the minimum.

**Definition 3.25.** A *transition operator* $T$ is an operator $T\colon \mathbb{G}(\mathscr{X}) \to \mathbb{G}(\mathscr{X})$ such that

T1. $Tf \geq \min f$ for all $f$ in $\mathbb{G}(\mathscr{X})$;

T2. $T(\mu f) = \mu T f$ for all $\mu$ in $\mathbb{R}$ and $f$ in $\mathbb{G}(\mathscr{X})$;

T3. $T(f+g) = Tf + Tg$ for all $f$ and $g$ in $\mathbb{G}(\mathscr{X})$.

We denote the set of all transition operators by $\mathfrak{T}_{\mathscr{X}}$, or simply by $\mathfrak{T}$ whenever the state space $\mathscr{X}$ is clear from the context.

Let $T$ be a transition operator. Note that (T1)–(T3) are 'gamble-valued versions' of $(E1)_{22}$–$(E3)_{22}$. For this reason, it follows from Proposition $2.15_{22}$ that for any $x$ in $\mathscr{X}$,

$$[T\bullet](x)\colon \mathbb{G}(\mathscr{X}) \to \mathbb{R}\colon f \mapsto [Tf](x)$$

is a coherent expectation on $\mathbb{G}(\mathscr{X})$. Consequently, it follows from Proposition $2.18_{23}$ that there is a (unique) probability mass function $p_x$ on $\mathscr{X}$ such that $[T\bullet](x) = E_{p_x}$. With this in mind, we observe that

$$T(x,y) = [T\mathbb{I}_y](x) = E_{p_x}(\mathbb{I}_y) = p_x(y) \quad \text{for all } x, y \in \mathscr{X}. \tag{3.31}$$

From these equalities, we can conclude that the matrix representation of a transition operator with respect to the natural basis $\{\mathbb{I}_x\colon x \in \mathscr{X}\}$ is a transition (or stochastic) matrix, that is, a matrix with non-negative components and rows that sum to 1. From this, Eq. $(3.28)_{77}$ and the properties of probability mass functions, we conclude that for any transition operator $T$, T4. $\|T\|_{\mathrm{op}} = 1$.

**Joseph's Example 3.26.** A straightforward example of a transition operator is the identity operator $I$. With respect to the basis $(\mathbb{I}_{\mathtt{H}}, \mathbb{I}_{\mathtt{T}})$, any linear operator $M$ on $\mathbb{G}(\{\mathtt{H},\mathtt{T}\})$ is represented by the matrix

$$\begin{pmatrix} M(\mathtt{H},\mathtt{H}) & M(\mathtt{H},\mathtt{T}) \\ M(\mathtt{T},\mathtt{H}) & M(\mathtt{T},\mathtt{T}) \end{pmatrix}.$$

It is clear that for a transition operator $T$, this matrix is of the form

$$\begin{pmatrix} p_{\mathtt{H}}(\mathtt{H}) & p_{\mathtt{H}}(\mathtt{T}) \\ p_{\mathtt{T}}(\mathtt{H}) & p_{\mathtt{T}}(\mathtt{T}) \end{pmatrix},$$

where $p_{\mathtt{H}}$ and $p_{\mathtt{T}}$ are probability mass functions on $\{\mathtt{H},\mathtt{T}\}$. Conversely, any matrix of this form defines a transition operator. ♫

### Rate operators

We want to take a closer look at semi-groups of transition operators, and rate operators are indispensable for this.

**Definition 3.27.** A *rate operator* $Q$ is an operator $Q\colon \mathbb{G}(\mathscr{X}) \to \mathbb{G}(\mathscr{X})$ such that

R1. $Q\mathbb{I}_{\mathscr{X}} = 0$;

R2. $[Q\mathbb{I}_y](x) \geq 0$ for all $x$ and $y$ in $\mathscr{X}$ such that $x \neq y$;

R3. $Q(\mu f) = \mu Q f$ for all $\mu$ in $\mathbb{R}$ and $f$ in $\mathbb{G}(\mathscr{X})$;

R4. $Q(f+g) = Qf + Qg$ for all $f$ and $g$ in $\mathbb{G}(\mathscr{X})$.

We let $\mathfrak{Q}_{\mathscr{X}}$ denote the set of all rate operators, and write $\mathfrak{Q}$ whenever the state space $\mathscr{X}$ is clear from the context.

Let $Q$ be a rate operator. It follows immediately from (R1) that, for any $x$ in $\mathscr{X}$,

$$0 = [Q\mathbb{I}_{\mathscr{X}}](x) = \sum_{y\in\mathscr{X}} Q(x,y)\mathbb{I}_{\mathscr{X}}(y) = \sum_{y\in\mathscr{X}} Q(x,y).$$

From these equalities, it follows that – with respect to the natural basis $(\mathbb{I}_x)_{x\in\mathscr{X}}$ – a rate operator is represented by a rate (or intensity) matrix, that is, a matrix with non-negative off-diagonal components and rows that sum to 0. Furthermore, we infer from these equalities and (R2) that, for any $x$ in $\mathscr{X}$,

$$0 \leq \sum_{y\in\mathscr{X}\setminus\{x\}} Q(x,y) = -Q(x,x).$$

Due to Eq. $(3.28)_{77}$, we conclude from the preceding equality that

R5. $\|Q\|_{\mathrm{op}} = 2\max\{-[Q\mathbb{I}_x](x)\colon x\in\mathscr{X}\} = 2\max\{-Q(x,x)\colon x\in\mathscr{X}\}$.

**Joseph's Example 3.28.** It is easy to verify that with respect to the basis $(\mathbb{I}_{\mathtt{H}}, \mathbb{I}_{\mathtt{T}})$, a rate operator $Q$ on $\mathbb{G}(\{\mathtt{H},\mathtt{T}\})$ is represented by a matrix of the form

$$\begin{pmatrix} Q(\mathtt{H},\mathtt{H}) & Q(\mathtt{H},\mathtt{T}) \\ Q(\mathtt{T},\mathtt{H}) & Q(\mathtt{T},\mathtt{T}) \end{pmatrix} = \begin{pmatrix} -\lambda_{\mathtt{H}} & \lambda_{\mathtt{H}} \\ \lambda_{\mathtt{T}} & -\lambda_{\mathtt{T}} \end{pmatrix},$$

with $\lambda_{\mathtt{H}}$ and $\lambda_{\mathtt{T}}$ two (arbitrary) non-negative real numbers. By (R5)↶,

$$\|Q\|_{\mathrm{op}} = 2\max\{\lambda_{\mathtt{H}}, \lambda_{\mathtt{T}}\}.$$

### Semi-groups of transition operators

Rate operators and transition operators are tightly linked. For starters, a rate operator can be used to construct a transition operator and vice versa.

**Lemma 3.29.** *Let $Q$ be a rate operator. For all $\Delta$ in $\mathbb{R}_{\geq 0}$ such that $\Delta\|Q\|_{\mathrm{op}} \leq 2$, $(I+\Delta Q)$ is a transition operator.*

*Proof.* This is essentially well-known, but also a special case of Lemma 3.72<sub></sub>112 further on. □

**Lemma 3.30.** *Let $T$ be a transition operator. For all $\Delta$ in $\mathbb{R}_{>0}$, $(T-I)/\Delta$ is a rate operator.*

*Proof.* This is essentially well-known, but also a special case of Lemma 3.73 (p. 113) further on. □

More importantly, it essentially follows from Proposition 3.23 (p. 78) and Lemmas 3.29 and 3.30 that semi-groups of transition operators are generated by rate operators through the operator exponential.

**Corollary 3.31.** *If $(T_t)_{t\in\mathbb{R}_{\geq 0}}$ is a continuous semi-group of transition operators, then its generator*

$$Q := \lim_{t\searrow 0} \frac{T_t - I}{t}$$

*is a rate operator, and it is the unique linear operator such that $T_t = e^{tQ}$ for all $t$ in $\mathbb{R}_{\geq 0}$. Conversely, for any rate operator $Q$, $(e^{tQ})_{t\in\mathbb{R}_{\geq 0}}$ is a continuous semi-group of transition operators.*

*Proof.* To verify the first part of the statement, we recall from Proposition 3.23 (p. 78) that $Q$ is the unique linear operator such that $T_t = e^{tQ}$ for all $t$ in $\mathbb{R}_{\geq 0}$. That $Q$ is a rate operator follows from its definition because by Lemma 3.30, $(T_t-I)/t$ is a rate operator for all $t$ in $\mathbb{R}_{>0}$ and the properties (R1)↶–(R4)↶ of rate operators are preserved when taking limits.

To verify the second part of the statement, we recall from Proposition 3.23 (p. 78) that $(e^{tQ})_{t\in\mathbb{R}_{\geq 0}}$ is a continuous semi-group, and that for all $t$ in $\mathbb{R}_{\geq 0}$,

$$e^{tQ} = \lim_{n\to+\infty}\left(I + \frac{t}{n}Q\right)^n.$$

Recall from Lemma 3.29 that whenever $t\|Q\|_{\mathrm{op}} \leq 2n$, $(I + \frac{t}{n}Q)$ is a transition operator. Furthermore, it is easy to see that the composition of two or more transition operators is again a transition operator. Therefore, $(I + \frac{t}{n}Q)^n$ is a transition operator whenever $t\|Q\|_{\mathrm{op}} \leq 2n$. Because the properties (T1)$_{80}$–(T3)$_{80}$ of transition operators are preserved when taking limits, this proves that $e^{tQ}$ is a transition operator. □

**Joseph's Example 3.32.** Recall from Joseph's Example 3.28 that, for $\mathscr{X} = \{\mathtt{H}, \mathtt{T}\}$, any rate operator $Q$ is uniquely characterised by the two parameters $\lambda_{\mathtt{H}}$ and $\lambda_{\mathtt{T}}$. Using Eq. (3.29)$_{79}$ and after a bit of work,[8] one can obtain the following analytical expression for the operator exponential of $Q$:

$$e^{tQ} = I + \frac{1 - e^{-t(\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}})}}{\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}}} Q \quad \text{for all } t \in \mathbb{R}_{\geq 0}, \tag{3.32}$$

where the second term is only added in case $\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}} > 0$. Alternatively, we can check if this expression solves the initial value problem of Proposition 3.24$_{79}$. This is trivial in case $\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}} = 0$, so from here on we assume that $\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}} > 0$. Fix some $t$ in $\mathbb{R}_{\geq 0}$. On the one hand,

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{tQ} = e^{-t(\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}})} Q.$$

On the other hand,

$$\begin{aligned} Qe^{tQ} &= Q + \frac{1 - e^{-t(\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}})}}{\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}}} Q^2 = Q - \left(1 - e^{-t(\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}})}\right) Q \\ &= e^{-t(\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}})} Q, \end{aligned}$$

where for the second equality we used that $Q^2 = -(\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}})Q$. In conclusion, $e^{tQ}$ satisfies the initial value problem of Proposition 3.24$_{79}$.

To verify that $e^{tQ}$ is a transition operator, we observe that

$$\begin{aligned} \left[e^{tQ}\right](\mathtt{H}, \mathtt{H}) &= 1 - \frac{\lambda_{\mathtt{H}}}{\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}}} \left(1 - e^{-t(\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}})}\right), \\ \left[e^{tQ}\right](\mathtt{H}, \mathtt{T}) &= \frac{\lambda_{\mathtt{H}}}{\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}}} \left(1 - e^{-t(\lambda_{\mathtt{H}} + \lambda_{\mathtt{T}})}\right), \end{aligned}$$

and similarly for the other two components. Clearly, the components $[e^{tQ}](\mathtt{H}, \mathtt{H})$ and $[e^{tQ}](\mathtt{H}, \mathtt{T})$ are non-negative and sum to 1. In other words, they correspond to a probability mass function, as required by Eq. (3.31)$_{80}$. ♁

[8] This expression is well-known; to the best of our knowledge, Fréchet (1938, Chapter II, Section II) was (one of) the first to derive it.

### 3.2.3 From transition probabilities to the transition operator

Finally, it is time to return to the main subject of this chapter: jump processes. Let us find out how transition operators help when studying (the transition probabilities of) jump processes.

To that end, we let $P$ be a jump process. First, we observe that it follows immediately from $(CP1)_{41}$–$(CP3)_{41}$ that

$$P(X_0 = \bullet) \colon \mathscr{X} \to [0,1] \colon x \mapsto P(X_0 = x)$$

is a probability mass function. For this reason, we call $P(X_0 = \bullet)$ the *initial mass function*[9] of $P$.

Next, for any two time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $t \leq r$, we define the operator $T_{t,r} \colon \mathbb{G}(\mathscr{X}) \to \mathbb{G}(\mathscr{X})$ by

$$\big[T_{t,r} f\big](x) := E_P(f(X_r) \,|\, X_t = x) \quad \text{for all } f \in \mathbb{G}(\mathscr{X}) \text{ and } x \in \mathscr{X}. \tag{3.33}$$

It follows immediately from $(ES1)_{37}$–$(ES3)_{37}$ that $T_{t,r}$ is a transition operator. Note that, for any $x$ and $y$ in $\mathscr{X}$,

$$T_{t,r}(x,y) = \big[T_{t,r}\mathbb{I}_y\big](x) = E_P(\mathbb{I}_y(X_r) \,|\, X_t = x) = P(X_r = y \,|\, X_t = x). \tag{3.34}$$

Because the components of $T_{t,r}$ are the transition probabilities, Krak et al. (2017, Definition 4.5) call $T_{t,r}$ the *transition operator from $t$ to $r$*.

The transition operator $T_{t,r}$ thus essentially provides us with a concise way to denote the transition probabilities of the form $P(X_r = y \,|\, X_t = x)$. However, more intricate transition probabilities are of importance as well, which is why Krak et al. (2017, Definition 4.6) define *history-dependent transition operators*. For any two time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and any state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t \leq r$, we define the operator $T_{t,r}^{\{X_u = x_u\}} \colon \mathbb{G}(\mathscr{X}) \to \mathbb{G}(\mathscr{X})$ by

$$\big[T_{t,r}^{\{X_u = x_u\}} f\big](x) := E_P(f(X_r) \,|\, X_u = x_u, X_t = x) \text{ for all } f \in \mathbb{G}(\mathscr{X}), x \in \mathscr{X}. \tag{3.35}$$

Again, $T_{t,r}^{\{X_u = x_u\}}$ is a transition operator, and

$$T_{t,r}^{\{X_u = x_u\}}(x,y) = P(X_r = y \,|\, X_u = x_u, X_t = x) \quad \text{for all } x, y \in \mathscr{X}. \tag{3.36}$$

As a special case, we observe that $T_{t,r}^{\Omega} = T_{t,r}^{(X_{()} = x_{()})} = T_{t,r}$. Furthermore, we note that it follows almost immediately from Eq. (3.36), $(CP1)_{41}$, $(CP9)_{42}$ and $(CP5)_{42}$ that

$$T_{t,t}^{\{X_u = x_u\}} = I \quad \text{for all } t \in \mathbb{R}_{\geq 0} \text{ and } \{X_u = x_u\} \in \mathscr{H} \text{ such that } u < t. \tag{3.37}$$

[9] The initial mass function is colloquially referred to as the 'initial distribution'.

### The transition operators of homogeneous Markovian jump processes

Due to Eq. (3.36)↶, a jump process $P$ is Markovian if and only if for all time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and any conditioning event $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t < r$,

$$T_{t,r}^{\{X_u = x_u\}} = T_{t,r}. \tag{3.38}$$

Similarly, it follows from Eq. (3.34)↶ that a Markovian jump process $P$ is homogeneous if and only if for all time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $t < r$,

$$T_{t,r} = T_{0,(r-t)}. \tag{3.39}$$

For this reason, for a homogeneous Markovian jump processes $P$, it suffices to consider the sequence $(T_t)_{t\in\mathbb{R}_{\geq 0}}$ of transition operators defined by

$$T_t := T_{0,t} \quad \text{for all } t \in \mathbb{R}_{\geq 0}.$$

It is relatively easy to establish that the transition operators of a homogeneous Markovian jump process form a semi-group; for a proof, see the first part of the proof of (Krak et al., 2017, Proposition 5.1).

**Lemma 3.33.** *For any homogeneous Markovian jump process $P$, the corresponding sequence $(T_t)_{t\in\mathbb{R}_{\geq 0}}$ of transition operators is a semi-group.*

With Corollary 3.31$_{82}$ in mind, it is clearly useful to establish a necessary and sufficient condition for this semi-group $(T_t)_{t\in\mathbb{R}_{\geq 0}}$ to be continuous.

**Lemma 3.34.** *Let $P$ be a homogeneous Markovian jump process. The corresponding semi-group $(T_t)_{t\in\mathbb{R}_{\geq 0}}$ is continuous if and only if*

$$\lim_{t\searrow 0} P(X_t = x \,|\, X_0 = x) = 1 \quad \textit{for all } x \in \mathscr{X}.$$

*Proof.* We do not prove this result directly, seeing that it is a special case of more general results that we prove further on. We may use these results because every transition operator is a lower transition operator – see Definition 3.60$_{107}$. Recall from Lemma 3.33 that $(T_t)_{t\in\mathbb{R}_{\geq 0}}$ is a semi-group of transition operators. For this reason, it follows from Lemma 3.76$_{114}$ further on that $(T_t)_{t\in\mathbb{R}_{\geq 0}}$ is continuous if and only if $\lim_{t\searrow 0}[T_t \mathbb{I}_x](x) = 1$ for all $x$ in $\mathscr{X}$. With the help of Eq. (3.34)↶, we immediately obtain the condition of the statement. □

We can use Corollary 3.31$_{82}$ and Lemmas 3.33 and 3.34 to establish that the transition probabilities of a homogeneous Markovian jump process are generated by a unique rate operator. This result is essentially well-known; in this form, it is similar to (Krak et al., 2017, Theorem 5.4).[10]

[10] The difference between Theorem 3.35↷ and (Krak et al., 2017, Theorem 5.4) is that we do not limit ourselves to 'well-behaved' processes a priori, but only require a (weaker) continuity condition.

**Theorem 3.35.** *Let $P$ be a homogeneous Markovian jump process. If*

$$\lim_{t \searrow 0} P(X_t = x \mid X_0 = x) = 1 \quad \textit{for all } x \in \mathscr{X}, \tag{3.40}$$

*then there is a unique probability mass function $p_0$ on $\mathscr{X}$ such that*

$$P(X_0 = x) = p_0(x) \quad \textit{for all } x \in \mathscr{X}$$

*and a unique rate operator $Q$ such that for all time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and all $\{X_u = x_u\}$ in $\mathscr{H}$ with $u \prec t \leq r$,*

$$T_{t,r}^{\{X_u = x_u\}} = e^{(r-t)Q}. \tag{3.41}$$

*Proof.* For the first part of the statement, we let $p_0$ be the real-valued function on $\mathscr{X}$ defined by $P(X_0 = x)$ for all $x$ in $\mathscr{X}$. It follows from (CP2)$_{41}$ that $p_0$ is non-negative, and – because the events $(\{X_0 = x\})_{x \in \mathscr{X}}$ form a (disjoint) partition of $\Omega$ – from (CP1)$_{41}$ and (CP3)$_{41}$ that $p_0$ is normed; thus, $p_0$ is a probability mass function. That $p_0$ is unique is obvious.

For the second part of the statement, we observe that due to Corollary 3.31$_{82}$, Lemmas 3.33$_{\curvearrowleft}$ and 3.34$_{\curvearrowleft}$ and Eqs. (3.38)$_{\curvearrowleft}$ and (3.39)$_{\curvearrowleft}$,

$$Q := \lim_{t \searrow 0} \frac{T_t - I}{t}$$

is the unique linear operator that satisfies Eq. (3.41). □

The continuity condition of Eq. (3.40) in Theorem 3.35 is a rather natural one, but it is not trivially satisfied. Let us investigate an example of a jump process that does not satisfy this continuity condition.

**Example 3.36.** Consider a state space $\mathscr{X}$ with at least 2 states. Let $T_0 := I$ and, for all $t$ in $\mathbb{R}_{>0}$, let $T_t$ be the transition operator defined by

$$[T_t f](x) := \frac{1}{|\mathscr{X}|} \sum_{y \in \mathscr{X}} f(y) \quad \text{for all } f \in \mathbb{G}(\mathscr{X}), x \in \mathscr{X}.$$

By Theorem 5.2 in (Krak et al., 2017), there is a homogeneous Markovian jump process $P$ such that for all $x, y$ in $\mathscr{X}$ and $t, \Delta$ in $\mathbb{R}_{\geq 0}$,

$$P(X_t = y \mid X_0 = x) = [T_t \mathbb{I}_y](x) = \begin{cases} 1 & \text{if } t = 0 \text{ and } x = y, \\ 0 & \text{if } t = 0 \text{ and } x \neq y, \\ \frac{1}{|\mathscr{X}|} & \text{if } t > 0. \end{cases}$$

Then because $|\mathscr{X}| \geq 2$, it is obvious that

$$\lim_{t \searrow 0} P(X_t = x \mid X_0 = x) = \frac{1}{|\mathscr{X}|} \neq 1 \quad \text{for all } x \in \mathscr{X},$$

so the continuity condition of Eq. (3.40) is not satisfied. Furthermore, it is clear that for all $x, y$ in $\mathscr{X}$, the 'transition rate from $x$ to $y$' is unbounded:

$$\lim_{t \searrow 0} \left[ \frac{T_{0,t} - I}{t} \right](x, y) = \begin{cases} -\infty & \text{if } x = y, \\ +\infty & \text{if } x \neq y. \end{cases}$$ ◇

### 3.2.4 Constructing Markovian jump processes

Theorem 3.35↶ establishes that under a very mild continuity assumption, the transition probabilities of a homogeneous Markovian jump process are uniquely characterised by a single parameter: the rate operator $Q$. Based on Corollary 3.31[82], it seems only natural to expect the converse to hold as well, in the sense that every rate operator characterises a (unique) Markov chain. Krak et al. (2017, Corollary 5.3) prove that this is the case, and we repeat their result here.

**Theorem 3.37.** *Fix a probability mass function $p_0$ on $\mathscr{X}$ and a rate operator $Q$. Then there is a unique jump process $P$ such that*

$$P(X_0 = x) = p_0(x) \quad \textit{for all } x \in \mathscr{X}$$

*and, for all time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and any conditioning event $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t \leq r$,*

$$T_{t,r}^{\{X_u = x_u\}} = e^{(r-t)Q}.$$

*Clearly, this jump process $P$ is Markovian and homogeneneous and satisfies the continuity condition from Eq. (3.40)↶ in Theorem 3.35↶.*

The importance of Theorem 3.37 can hardly be overstated. Most importantly, together with Proposition 3.20[73] and Theorem 3.35↶, it establishes thatany homogeneous Markovian jump process that satisfies the continuity condition from Eq. (3.40)↶ in Theorem 3.35↶ is uniquely characterised by two parameters: its initial mass function $p_0$ and its rate operator $Q$. For this reason, for all initial mass functions $p_0$ and rate operators $Q$, we let $P_{p_0,Q}$ denote the unique jump process that is characterised by $p_0$ and $Q$ in the sense of Theorem 3.37. In order not to clutter our notation, we will denote the conditional expectation operator $E_{P_{p_0,Q}}$ corresponding to $P_{p_0,Q}$ by $E_{p_0,Q}$.

**Joseph's Example 3.38.** Bruno and Joseph have several other colleagues besides Alice and Cecilia, and one of these is Deborah. Deborah looks a bit like a zebra and has a knack for conjuring tricks, but also knows a thing or two about nuclear decay. More precisely, she knows that if the mass in Joseph's machine only contains a single radioactive isotope, then the decay events can be described by a so-called Poisson process (see, for example, Norris, 1997, Section 2.4). Note that $X_t$ differs from $X_r$ if and only if there have been an odd number of decays – that is, Poisson events – in the interval $]t, r]$. Hence, Deborah knows that the jump process $P$ that models (her beliefs about) Joseph's machine should satisfy

$$P(X_r = \mathtt{T} \mid X_u = x_u, X_t = \mathtt{H}) = P(X_r = \mathtt{H} \mid X_u = x_u, X_t = \mathtt{T}) = \frac{1}{2}\Big(1 - e^{-2(r-t)\lambda}\Big),$$

where the positive real number $\lambda$ is related to the half-life of the radioactive isotope. Comparing this to the expression for $e^{(r-t)Q}$ in Joseph's Example 3.32₈₃ and using the laws of probability, she comes to the conclusion that

$$T_{t,r}^{\{X_u=x_u\}} = e^{(t-r)Q_\lambda} \quad \text{with} \quad Q_\lambda := \begin{pmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{pmatrix}.$$

Theorem 3.37↶ now guarantees that there is such a jump process. ♧

Up until now, we have focussed exclusively on homogeneous Markovian jump processes. If we drop the homogeneity assumption, then the transition probabilities of the form $P(X_r = y \,|\, X_t = x)$ *do* depend on the current time point $t$, and we need to consider a 'two-parameter semi-group'

$$T_{t,r} \quad \text{with } t, r \in \mathbb{R}_{\geq 0} \text{ such that } t \leq r.$$

It is common practice to look at the (left and right) derivative of $T_{t,r}$ for both $t$ and $r$ for a given Markovian jump process, or to construct such a two-parameter semi-group – and the corresponding Markovian jump process – using a time-dependent rate operator $Q_t$. We will not pursue this any further here, but refer the interested reader to (Krak et al., 2017, Section 5.2) and references therein.

## 3.3 Imprecise jump processes

Now that we have (precise) jump processes covered, it is finally time to move on to imprecise jump processes. As we have previously mentioned in Section 2.4.3₄₉, we will do so from a sensitivity analysis point of view. More specifically, we take the position that there is a 'true' jump process that accurately models the system, but that the subject's beliefs only allow a partial specification in the form of a *set* of jump processes.

**Definition 3.39.** An *imprecise jump process* $\mathscr{P}$ with state space $\mathscr{X}$ is a non-empty subset of $\mathbb{P}_{\mathscr{X}}$.

Because every jump process $P$ in an imprecise jump process $\mathscr{P}$ has the same domain $\mathscr{D}$ that is a structure of fields, imprecise jump processes fall squarely in the scope of Section 2.4.3₄₉. Recall that there, we defined the conditional lower and upper probabilities $\underline{P}_{\mathscr{P}}$ and $\overline{P}_{\mathscr{P}}$ on $\mathscr{D}$ as the lower and upper envelope of $\mathscr{P}$ given by

$$\underline{P}_{\mathscr{P}}(A \,|\, C) := \inf\{P(A \,|\, C) \colon P \in \mathscr{P}\} \qquad \text{for all } (A \,|\, C) \in \mathscr{D} \tag{3.42}$$

and

$$\overline{P}_{\mathscr{P}}(A \,|\, C) := \sup\{P(A \,|\, C) \colon P \in \mathscr{P}\} \qquad \text{for all } (A \,|\, C) \in \mathscr{D}. \tag{3.43}$$

Furthermore, we also defined there the conditional lower and upper expectation $\underline{E}_{\mathscr{P}}$ and $\overline{E}_{\mathscr{P}}$ on $\mathbb{JS}$ as the lower and upper envelopes of the set $\{E_P : P \in \mathscr{P}\}$ of conditional expectations corresponding to the coherent conditional probabilities in $\mathscr{P}$; in short, $\underline{E}_{\mathscr{P}}$ and $\overline{E}_{\mathscr{P}}$ are defined by

$$\underline{E}_{\mathscr{P}}(f \mid C) := \inf\{E_P(f \mid C) : P \in \mathscr{P}\} \qquad \text{for all } (f \mid C) \in \mathbb{JS} \tag{3.44}$$

and

$$\overline{E}_{\mathscr{P}}(f \mid C) := \sup\{E_P(f \mid C) : P \in \mathscr{P}\} \qquad \text{for all } (f \mid C) \in \mathbb{JS}. \tag{3.45}$$

Finally, we recall from Section 2.4.3[49] that it suffices to study conditional lower expectations, on account of the fact that conditional lower probabilities are a special case and that upper envelopes – of expectations as well as probabilities – can be obtained through conjugacy.

### 3.3.1 Consistent jump processes

Imprecise jump processes appear naturally in situations where we are unable to precisely specify the defining parameters – that is, the initial mass function and the rate operator – of a homogeneous Markovian jump process. Let us illustrate this with our running example.

**Joseph's Example 3.40.** Recall from Joseph's Example 3.38[87] that Deborah knows that when Joseph's machine contains a single isotope, the machine can be modelled by a homogeneous Markovian jump process with rate operator

$$Q_\lambda = \begin{pmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{pmatrix},$$

where the parameter $\lambda$ is related to the half-life of the radioactive isotope.

Deborah is in charge of the stock of radioactive isotopes in her workplace, and in this capacity she has noticed that recently, the stock of $n$ specific isotopes has lowered without anyone actually using it for their research. For this reason, she is absolutely certain that Joseph has used at least one of these $n$ isotopes. Let us denote the corresponding parameters by $\lambda_1, \dots, \lambda_n$. If Deborah were to believe that Joseph has put a single radioactive isotope in his machine, then her beliefs would be modelled by the imprecise jump process $\mathscr{P}$ that consists of all homogeneous Markovian jump process (with state space $\{\mathtt{H}, \mathtt{T}\}$) whose rate operator $Q$ belongs to $\{Q_{\lambda_1}, \dots, Q_{\lambda_n}\}$. However, Deborah does not want to exclude the possibility that Joseph uses two radioactive isotopes: one when his machine displays heads, and the other when his machine displays tails. In this case, her beliefs are modelled by the imprecise jump process $\mathscr{P}$ that consist of all homogeneous Markovian jump process (with state space $\{\mathtt{H}, \mathtt{T}\}$) whose rate operator $Q$ belongs to

$$\mathscr{Q} := \left\{ \begin{pmatrix} -\lambda_{\mathtt{H}} & \lambda_{\mathtt{H}} \\ \lambda_{\mathtt{T}} & -\lambda_{\mathtt{T}} \end{pmatrix} : \lambda_{\mathtt{H}}, \lambda_{\mathtt{T}} \in \{\lambda_1, \dots, \lambda_n\} \right\}.$$ ♄

Imprecise jump processes also arise naturally in less artificial situations. For example, in Queuing Network Example $7.36_{368}$ in Chapter $7_{337}$ further on, we will use them to model a closed queueing network where the defining parameters – that is, the mean service times of the servers – are only known to belong to intervals; more specifically, we resort there to imprecise jump process to obtain guaranteed lower and upper bounds on some performance measures. Similarly, in Chapter $8_{403}$ further on, we will investigate spectrum allocation policies for optical links, and there we will use an imprecise jump process to obtain policy-independent bounds on the key performance indicators. Joseph's Example 3.40↶ and the two preceding examples all have something in common: the imprecise jump processes are naturally characterised by sets of initial mass functions and sets of rate operators. Sets of rate operators also arise naturally when learning the rate operator of a homogeneous Markovian jump process in an imprecise framework. In that case, the estimator for the rate operator is a set of rate operators, and this set of rate operators characterises a set of (homogeneous and Markovian) jump processes (Krak et al., 2018).

**Consistent homogeneous Markovian jump processes**

The most basic way to construct an imprecise jump process is using the procedure of Joseph's Example 3.40↶. Let us generalise this procedure for any non-empty set $\mathscr{M}$[11] of probability mass functions on $\mathscr{X}$ and any non-empty set $\mathscr{Q}$ of rate operators. As in Joseph's Example 3.40↶, we construct an imprecise jump process using $\mathscr{M}$ and $\mathscr{Q}$ by considering all homogeneous Markovian jump processes that are characterised by an initial mass function $p_0$ in $\mathscr{M}$ and a rate operator $Q$ in $\mathscr{Q}$:

$$\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}} = \left\{P_{p_0,Q} \colon p_0 \in \mathscr{M}, Q \in \mathscr{Q}\right\}. \tag{3.46}$$

By definition, $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is fully characterised by the two parameters $\mathscr{M}$ and $\mathscr{Q}$, and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}} = \{P_{p_0,Q}\}$ whenever $\mathscr{M} = \{p_0\}$ and $\mathscr{Q} = \{Q\}$. Thus, the imprecise jump process $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is a proper generalisation of the homogeneous Markovian jump process $P_{p_0,Q}$ to imprecise jump processes However, $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is *not* the only imprecise jump process that generalises $P_{p_0,Q}$ in this way. In fact, Krak et al. (2017) propose two additional such generalisations: the first is the set of all – not necessarily homogeneous – Markovian jump processes that are characterised by $\mathscr{M}$ and $\mathscr{Q}$, and the second is the set of all – not necessarily

[11] In Section $2.2.4_{27}$, we have used $\mathcal{M}$ to denote a set of coherent expectations on $\mathbb{G}(\mathscr{X})$, which motivates our use of $\mathscr{M}$ to denote a set of probability mass functions. Recall from Proposition $2.18_{23}$ that because $\mathscr{X}$ is finite, every coherent expectation $E$ on $\mathbb{G}(\mathscr{X})$ is in one-to-one correspondence with a probability mass function $p$ through the relation $p(x) = E(\mathbb{I}_x)$ for all $x$ in $\mathscr{X}$. Thus, because $\mathscr{X}$ is finite, sets of coherent expectations on $\mathbb{G}(\mathscr{X})$ and sets of probability mass functions on $\mathscr{X}$ are essentially equivalent.

Markovian – jump process that are characterised by $\mathscr{M}$ and $\mathscr{Q}$. To formalise these two definitions, we need to introduce the notion of *consistency*.

### Consistency with a set of initial mass functions

First, we consider the notion of consistency with a set $\mathscr{M}$ of initial mass functions (see Krak et al., 2017, Definition 6.2).

**Definition 3.41.** Consider a set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. A jump process $P$ is *consistent with the set $\mathscr{M}$ of initial mass functions* if $P(X_0 = \bullet)$, the initial mass function of $P$, belongs to $\mathscr{M}$. We let $\mathbb{P}_{\mathscr{M}}$ denote the set of all jump processes that are consistent with the set $\mathscr{M}$ of initial mass functions.

Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty set $\mathscr{Q}$ of rate operators. By Eq. (3.46)$_{\curvearrowleft}$ and Theorem 3.37$_{87}$, every $P$ in $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is a homogeneous Markovian jump process that is consistent with the set $\mathscr{M}$ of initial mass functions. Consequently,

$$\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathbb{P}^{\mathrm{HM}} \cap \mathbb{P}_{\mathscr{M}}. \tag{3.47}$$

### Dynamics of the history-dependent transition operators

Consistency with a set $\mathscr{Q}$ of rate operators is a bit more involved. Our definition is inspired by the following important result regarding homogeneous Markovian jump processes.

**Proposition 3.42.** *Consider a rate operator $Q$ and a jump process $P$. Then $P$ is a homogeneous Markovian jump process with rate operator $Q$ – in the sense that $P = P_{p_0,Q}$ with $p_0 := P(X_0 = \bullet)$ – if and only if for all $t$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t$,*

$$\lim_{r \searrow t} \frac{T^{\{X_u=x_u\}}_{t,r} - I}{r-t} = Q \quad \textit{and, if } t > 0, \quad \lim_{s \nearrow t} \frac{T^{\{X_u=x_u\}}_{s,t} - I}{t-s} = Q. \tag{3.48}$$

*Proof.* The direct implication follows immediately from Theorem 3.37$_{87}$, Proposition 3.24$_{79}$ and Eq. (3.37)$_{84}$.

To prove the converse implication, we assume that $P$ is a jump process that satisfies Eq. (3.48). It then follows from (Krak et al., 2017, Definition 6.1) that $P$ is consistent with $\{Q\}$ in the sense of Krak et al. (2017, Definition 6.1) – or, by Proposition 3.57$_{104}$ further on, in the sense of Definition 3.50$_{99}$. Hence, it follows from (Krak et al., 2017, Proposition 8.1) – essentially Proposition 3.80$_{117}$ further on – that for all $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ with $u \prec t \leq r$

$$\big(\forall f \in \mathbb{G}(\mathscr{X})\big)\ T^{\{X_u=x_u\}}_{t,r} f \geq e^{(r-t)Q} f.$$

Recall that $T^{\{X_u=x_u\}}_{t,r}$ and $e^{(r-t)Q}$ are transition operators, so $T^{\{X_u=x_u\}}_{t,r}(-f) = -T^{\{X_u=x_u\}}_{t,r} f$ and $e^{(r-t)Q}(-f) = -e^{(r-t)Q} f$ due to (T2)$_{80}$. Hence,

$$\big(\forall f \in \mathbb{G}(\mathscr{X})\big)\ T^{\{X_u=x_u\}}_{t,r} f = e^{(r-t)Q} f.$$

From this and Theorem 3.37$_{87}$, we conclude that $P = P_{p_0,Q}$, as required. □

Crucially, we learn from Proposition 3.42 that for a homogeneous Markovian jump process, the rate operator fully characterizes the 'dynamics' of its (history-dependent) transition operators. That is, the rate operator $Q$ determines the (history-dependent) transition operators – or equivalently, the transition probabilities – for the time points $s$ and $r$ with $s < t < r$ that are sufficiently close to the current time point $t$, in the sense that

$$T_{t,r}^{\{X_u = x_u\}} \approx I + (r-t)Q \quad \text{and} \quad T_{s,t}^{\{X_u = x_u\}} \approx I + (t-s)Q. \tag{3.49}$$

It is this characterisation of the dynamics that we want to extend to sets of rate operators through consistency. Thus, we say that a jump process $P$ is 'consistent with the set $\mathcal{Q}$ of rate operators' if this set $\mathcal{Q}$ characterizes its dynamics – or, more precisely, the dynamics of its history-dependent transition operators. There are a couple of ways to formalise this statement. One of the more obvious ones is to relax Eq. (3.48), in the sense that these limits should belong to the set $\mathcal{Q}$. However, this way we implicitly assume that these limits exist, and this need not be the case. Krak et al. (2017, Example 4.5) give one example of a – not all too crazy – Markovian jump process for which these limits do not always exist, but let us look at an example in the setting of our running example.

**Joseph's Example 3.43.** Recall from Joseph's Example 3.14$_{66}$ that for any $[0,1]$-valued function $\phi$ on $\mathbb{R}_{>0}$, there is a jump process $P_\phi$ such that

$$P_\phi(X_t = x \mid X_0 = x) = \phi(t) \quad \text{for all } x \in \mathscr{X}, t \in \mathbb{R}_{>0}.$$

In this example, we consider a jump process $P_\phi$ corresponding to the function

$$\phi\colon \mathbb{R}_{>0} \to [0,1]\colon t \mapsto \phi(t) := \begin{cases} 1-t & \text{if } \frac{1}{2n} \leq t < \frac{1}{2n-1} \text{ for some } n \in \mathbb{N}, \\ 1-\frac{1}{2}t & \text{if } \frac{1}{2n+1} \leq t < \frac{1}{2n} \text{ for some } n \in \mathbb{N}, \\ 0 & \text{if } t \geq 1. \end{cases}$$

Then for any $t$ in $\mathbb{R}_{>0}$, the matrix representation of the corresponding transition operator $T_{0,t}$ is

$$\begin{pmatrix} \phi(t) & 1-\phi(t) \\ 1-\phi(t) & \phi(t) \end{pmatrix}. \tag{3.50}$$

Let $Q$ be any linear operator with matrix representation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Fix any $t$ in $\mathbb{R}_{>0}$. Then by Eqs. (3.28)$_{77}$ and (3.50)$_{\curvearrowleft}$,

$$\left\|Q - \frac{T_{0,t} - I}{t}\right\|_{\mathrm{op}} = \max\left\{\left|a - \frac{\phi(t)-1}{t}\right| + \left|b - \frac{1-\phi(t)}{t}\right|, \left|c - \frac{1-\phi(t)}{t}\right| + \left|d - \frac{\phi(t)-1}{t}\right|\right\}. \quad (3.51)$$

Let us assume that $t < 1$, so we can more easily substitute the definition of $\phi$ in this expression. If there is some $n$ in $\mathbb{N}$ such that $\frac{1}{2n} \leq t < \frac{1}{2n-1}$, then

$$\left\|Q - \frac{T_{0,t} - I}{t}\right\|_{\mathrm{op}} = \max\{|a+1| + |b-1|, |c-1| + |d+1|\}.$$

On the other hand, if there is some $n$ in $\mathbb{N}$ such that $\frac{1}{2n+1} \leq t < \frac{1}{2n}$, then

$$\left\|Q - \frac{T_{0,t} - I}{t}\right\|_{\mathrm{op}} = \max\left\{\left|a + \frac{1}{2}\right| + \left|b - \frac{1}{2}\right|, \left|c - \frac{1}{2}\right| + \left|d + \frac{1}{2}\right|\right\}.$$

Clearly, there is no choice of the parameters $a$, $b$, $c$, $d$ such that

$$\lim_{t \searrow 0} \left\|Q - \frac{T_{0,t} - I}{t}\right\|_{\mathrm{op}} = 0,$$

so this is an example where the limit in Eq. (3.48)$_{91}$ does not exist. ♪

Alternatively, we could relax Eq. (3.49)$_{\curvearrowleft}$ by requiring that the approximations hold for some rate operators $Q_r$ and $Q_s$ in $\mathcal{Q}$. However, this approach would only ensure that the limits in Eq. (3.48)$_{91}$ are in the (topological) closure of $\mathcal{Q}$,[12] so it would then be a bit of a stretch to say that $\mathcal{Q}$ characterizes the dynamics of $T_{t,t}^{\{X_u=x_u\}}$. To mitigate these issues, we will use a combination of both approaches, essentially following the alternative path proposed by Krak et al. (2017).

Our approach differs from that of Krak et al. (2017) on two fronts. First and foremost, we define our set-valued derivative using – a notion of – the distance between a rate operator and a set of rate operators instead of through the limit points of convergent sequences of rate operators corresponding to sequences of time points. Second, we use a notion of (set-valued) differentiability to ensure that the set-valued derivative exists, instead of a priori restricting us to the 'well-behaved' jump processes that Krak et al. (2017) consider. This being said, we will establish in Proposition 3.57$_{104}$ further on that our notion of the set-valued derivative coincides with that of Krak et al. (2017) whenever we deal with jump processes that have bounded rate; consequently, the same holds for consistency with a bounded set of rate operators.

[12]More precisely, this condition only ensures that for every non-increasing sequence $(r_n)_{n\in\mathbb{N}}$ in $]t, +\infty[$ with $\lim_{n\to+\infty} r_n = t$, the limit (or accumulation) points of the sequence $((T_{t,r_n}^{\{X_u=x_u\}} - I)/(r_n - t))_{n\in\mathbb{N}}$ belong to the closure of $\mathcal{Q}$, and similarly for the limit from the left.

**Set-valued directional derivatives**

First, let us introduce some additional notation. For any jump process $P$, any time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and any state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t < r$, we let

$$Q_{t,r}^{\{X_u=x_u\}} := \frac{T_{t,r}^{\{X_u=x_u\}} - I}{r-t}. \tag{3.52}$$

Because $T_{t,r}^{\{X_u=x_u\}}$ is a transition operator, it follows from Lemma 3.30$_{82}$ that $Q_{t,r}^{\{X_u=x_u\}}$ is a rate operator. In order to study the dynamics of a jump process $P$, we look at the 'limit' of $Q_{t,r}^{\{X_u=x_u\}}$ as $r$ decreases to $t$ and the 'limit' of $Q_{s,t}^{\{X_u=x_u\}}$ as $s$ increases to $t$. As we have mentioned before, these limits do not necessarily exist; more precisely, $Q_{t,r}^{\{X_u=x_u\}}$ can have multiple accumulation points as $r$ decreases to $t$, and similarly for $Q_{s,t}^{\{X_u=x_u\}}$ as $s$ increases to $t$.

The intuitive idea behind our set-valued directional derivatives is the following. We generalise Eq. (3.48)$_{91}$ in the sense that we consider the accumulation points of $Q_{t,r}^{\{X_u=x_u\}}$ as $r$ decreases to $t$ and those of $Q_{s,t}^{\{X_u=x_u\}}$ as $s$ increases to $r$. However, in the spirit of Eq. (3.49)$_{92}$, we only call these sets of accumulation points 'set-valued directional derivatives' whenever they allow us to approximate the history dependent transition operators they were derived from, in the sense that for time points $r$ and $s$ sufficiently close to $t$, there are accumulation points $Q_r$ and $Q_s$ such that

$$T_{t,r}^{\{X_u=x_u\}} \approx I + (r-t)Q_r \text{ and } T_{s,t}^{\{X_u=x_u\}} \approx I + (t-s)Q_s. \tag{3.53}$$

To formalise this intuitive idea, we will use a notion of the 'distance' between a single rate operator and a set of rate operators.[13] For any rate operator $Q$ in $\mathfrak{Q}$ and any subset $\mathcal{Q}$ of $\mathfrak{Q}$, we define the *distance from $Q$ to $\mathcal{Q}$* as

$$d_{\mathfrak{Q}}(Q, \mathcal{Q}) := \inf\{\|Q - Q'\|_{\text{op}} : Q' \in \mathcal{Q}\}; \tag{3.54}$$

note that this distance is a real number whenever $\mathcal{Q}$ is non-empty and equal to $+\infty$ whenever $\mathcal{Q}$ is empty.

There is more than one way to explain how the distance function $d_{\mathfrak{Q}}$ leads to a notion of set-valued directional derivatives. Here, we will give a definition that is inspired by the notion of Painlevé-Kuratowski convergence for sequences of sets (see Rockafellar et al., 1998, Chapter 4). An alternative and perhaps more direct way to arrive at the same notion of set-valued directional derivatives is to use the Hausdorff semi-metric that corresponds to the distance function $d_{\mathfrak{Q}}$. The link between the latter definition of set-valued directional derivatives and that of – standard – directional derivatives is arguably more immediate but also more technical; for this reason, we have relegated it to Appendix 3.A.2$_{128}$.

[13] This is a special case of the general notion of the distance between an element of a metric space and a subset of the same metric space, see (Schechter, 1997, Section 4.40), (Burago et al., 2011, Exercise 1.4.2) or (Conci et al., 2017, Section 2).

Consider a jump process $P$, and fix a current time point $t$ in $\mathbb{R}_{\geq 0}$ and state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t$. The set of accumulation points of $Q_{t,r}^{\{X_u=x_u\}}$ as $r$ decreases to $t$ consists of those rate operators $Q$ in $\mathfrak{Q}$ that come arbitrarily close to $Q_{t,r}^{\{X_u=x_u\}}$ time and time again as $r$ approaches $t$. More formally, we draw inspiration from the notion of Painlevé-Kuratowski convergence for sequences of sets,[14] and collect the right-sided accumulation points of $Q_{t,\bullet}^{\{X_u=x_u\}}$ in

$$\partial_+ T_{t,t}^{\{X_u=x_u\}} := \left\{Q \in \mathfrak{Q} \colon \lim_{r \searrow t} d_{\mathfrak{Q}}\big(Q, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\big) = 0\right\}. \tag{3.55}$$

Similarly, if $t > 0$, we collect the left-sided accumulation points of $Q_{\bullet,t}^{\{X_u=x_u\}}$ in

$$\partial_- T_{t,t}^{\{X_u=x_u\}} := \left\{Q \in \mathfrak{Q} \colon \lim_{s \nearrow t} d_{\mathfrak{Q}}\big(Q, \{Q_{s',t}^{\{X_u=x_u\}}\}_{s' \in ]s,t[}\big) = 0\right\}. \tag{3.56}$$

**Joseph's Example 3.44.** We consider again the jump process $P_\phi$ as defined in Joseph's Example 3.43₉₂. Let $Q$ be any rate operator with matrix representation

$$\begin{pmatrix} -\lambda_{\mathrm{H}} & \lambda_{\mathrm{H}} \\ \lambda_{\mathrm{T}} & -\lambda_{\mathrm{T}} \end{pmatrix}.$$

Fix any $t$ in $]0,1[$. Then by Eq. (3.51)₉₃,

$$\|Q - Q_{0,t}\|_{\mathrm{op}} = \max\left\{2\left|\lambda_{\mathrm{H}} + \frac{\phi(t)-1}{t}\right|, 2\left|\lambda_{\mathrm{T}} + \frac{\phi(t)-1}{t}\right|\right\}.$$

As in Joseph's Example 3.43₉₂, we substitute the definition of $\phi$ in this expression. If there is some $n$ in $\mathbb{N}$ such that $\frac{1}{2n} \leq t < \frac{1}{2n-1}$, then

$$\|Q - Q_{0,t}\|_{\mathrm{op}} = \max\{2|\lambda_{\mathrm{H}} - 1|, 2|\lambda_{\mathrm{T}} - 1|\}. \tag{3.57}$$

On the other hand, if there is some $n$ in $\mathbb{N}$ such that $\frac{1}{2n+1} \leq t < \frac{1}{2n}$, then

$$\|Q - Q_{0,t}\|_{\mathrm{op}} = \max\left\{2\left|\lambda_{\mathrm{H}} - \frac{1}{2}\right|, 2\left|\lambda_{\mathrm{T}} - \frac{1}{2}\right|\right\}. \tag{3.58}$$

From this and Eq. (3.54)↶, it follows that

$$\begin{aligned} &d_{\mathfrak{Q}}\big(Q, \{Q_{0,s}\}_{s \in ]0,t[}\big) \\ &\qquad = \min\left\{\max\{2|\lambda_{\mathrm{H}} - 1|, 2|\lambda_{\mathrm{T}} - 1|\}, \max\left\{2\left|\lambda_{\mathrm{H}} - \frac{1}{2}\right|, 2\left|\lambda_{\mathrm{T}} - \frac{1}{2}\right|\right\}\right\}. \end{aligned}$$

[14] The notion of Painlevé-Kuratowski convergence is as follows (see Rockafellar et al., 1998, Definition 4.1 and Exercise 4.2): the limit of a sequence $(\mathcal{Q}_n)_{n \in \mathbb{N}}$ of subsets of $\mathfrak{Q}$ exists (in the Painlevé-Kuratowski sense) if the sets

$$\left\{Q \in \mathfrak{Q} \colon \liminf_{n \to +\infty} d_{\mathfrak{Q}}(Q, \mathcal{Q}_n) = 0\right\} \quad \text{and} \quad \left\{Q \in \mathfrak{Q} \colon \limsup_{n \to +\infty} d_{\mathfrak{Q}}(Q, \mathcal{Q}_n) = 0\right\},$$

coincide, where the first set is the 'limit superior' and the second set the 'limit inferior'.

Because this expression holds for any $t$ in $]0,1[$, it follows from Eq. (3.55) that $\partial_+ T_{0,0}$ contains precisely two rate operators: one with $\lambda_{\mathrm{H}} = 1 = \lambda_{\mathrm{T}}$ and one with $\lambda_{\mathrm{H}} = \frac{1}{2} = \lambda_{\mathrm{T}}$.

As explained right before Eq. (3.53)$_{94}$, we should only call the set $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ of accumulation points the 'set-valued right-sided derivative' if these accumulation points approximate the history dependent transition operators for sufficiently close (future) time points. Formally, we should require that

$$\big(\forall \epsilon \in \mathbb{R}_{>0}\big)\big(\exists \delta \in \mathbb{R}_{>0}\big)\big(\forall r \in \,]t, t+\delta[\big)\big(\exists Q \in \partial_+ T_{t,t}^{\{X_u=x_u\}}\big)\ \big\|Q_{t,r}^{\{X_u=x_u\}} - Q\big\|_{\mathrm{op}} < \epsilon.$$

It is not difficult to see that this condition is equivalent to

$$\big(\forall \epsilon \in \mathbb{R}_{>0}\big)\big(\exists \delta \in \mathbb{R}_{>0}\big)\big(\forall r \in \,]t, t+\delta[\big)\ d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_u=x_u\}}, \partial_+ T_{t,t}^{\{X_u=x_u\}}\big) < \epsilon;$$

that is, we should require that

$$\lim_{r \searrow t} d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_u=x_u\}}, \partial_+ T_{t,t}^{\{X_u=x_u\}}\big) = 0. \tag{3.59}$$

Important to realise is that this condition is not necessarily satisfied.

**Joseph's Example 3.45.** For this example, we consider the function

$$\phi\colon \mathbb{R}_{>0} \to [0,1]\colon t \mapsto \phi(t) := \begin{cases} 0 & \text{if } \frac{1}{t} \in \mathbb{N}, \\ 1 & \text{otherwise.} \end{cases}$$

As we have shown in Joseph's Example 3.14$_{66}$, there is a jump process $P_\phi$ such that

$$P_\phi(X_t = x \,|\, X_0 = x) = \phi(t) \quad \text{for all } x \in \mathscr{X}, t \in \mathbb{R}_{>0}.$$

Let $Q$ be any rate operator characterised by $\lambda_{\mathrm{H}}$ and $\lambda_{\mathrm{T}}$. Then following a similar argument as in Joseph's Example 3.44, we find that, for all $t$ in $\mathbb{R}_{>0}$,

$$\|Q - Q_{0,t}\|_{\mathrm{op}} = \begin{cases} \max\{2|\lambda_{\mathrm{H}} - \frac{1}{t}|, 2|\lambda_{\mathrm{T}} - \frac{1}{t}|\} & \text{if } \frac{1}{t} \in \mathbb{N}, \\ \max\{2\lambda_{\mathrm{H}}, 2\lambda_{\mathrm{T}}\} & \text{otherwise.} \end{cases} \tag{3.60}$$

From this and Eq. (3.54)$_{94}$, it follows – after some easy manipulations – that

$$d_{\mathfrak{Q}}\big(Q, \{Q_{0,s}\}_{s \in ]0,t[}\big) = \inf\Big(\{\max\{2\lambda_{\mathrm{H}}, 2\lambda_{\mathrm{T}}\}\} \cup \Big\{\max\{2|\lambda_{\mathrm{H}} - n|, 2|\lambda_{\mathrm{T}} - n|\}\colon n \in \mathbb{N}, \frac{1}{n} < t\Big\}\Big).$$

From this, we infer that

$$\lim_{t \searrow 0} d_{\mathfrak{Q}}\big(Q, \{Q_{0,s}\}_{s \in ]0,t[}\big) = 0 \Leftrightarrow \lambda_{\mathrm{H}} = 0 = \lambda_{\mathrm{T}}.$$

Therefore, it follows from Eq. (3.55)95 that $\partial_+ T_{0,0}$ is the singleton containing the rate operator $Q^\star$ with $\lambda_{\mathrm{H}} = 0 = \lambda_{\mathrm{T}}$.

Because $\partial_+ T_{0,0} = \{Q^\star\}$, it follows from Eqs. (3.54)94 and (3.60)↶ that, for all $t$ in $\mathbb{R}_{>0}$,

$$d_{\mathfrak{Q}}\big(Q_{0,t}, \partial_+ T_{0,0}\big) = \|Q^\star - Q_{0,t}\|_{\mathrm{op}} = \begin{cases} \frac{2}{t} & \text{if } \frac{1}{t} \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, this implies that

$$\liminf_{t\searrow 0} d_{\mathfrak{Q}}\big(Q_{0,t}, \partial_+ T_{0,0}\big) = 0 \quad \text{and} \quad \limsup_{t\searrow 0} d_{\mathfrak{Q}}\big(Q_{0,t}, \partial_+ T_{0,0}\big) = +\infty,$$

so this is an example where Eq. (3.59)↶ does not hold. ♫

Our intuitive definition of the set-valued directional derivatives leads to the following formal one.[15]

**Definition 3.46.** Consider a jump process $P$, a time point $t$ in $\mathbb{R}_{\geq 0}$ and a state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t$. We say that $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-*differentiable from the right* if

$$\partial_+ T_{t,t}^{\{X_u=x_u\}} \neq \emptyset \quad \text{and} \quad \lim_{r\searrow t} d_{\mathfrak{Q}}\Big(Q_{t,r}^{\{X_u=x_u\}}, \partial_+ T_{t,t}^{\{X_u=x_u\}}\Big) = 0,$$

and we then call $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ the *right-sided* $d_{\mathfrak{Q}}$*-derivative of* $T_{t,t}^{\{X_u=x_u\}}$. Similarly, if $t > 0$, we say that $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$*-differentiable from the left* if

$$\partial_- T_{t,t}^{\{X_u=x_u\}} \neq \emptyset \quad \text{and} \quad \lim_{s\nearrow t} d_{\mathfrak{Q}}\Big(Q_{s,t}^{\{X_u=x_u\}}, \partial_- T_{t,t}^{\{X_u=x_u\}}\Big) = 0,$$

and we then call $\partial_- T_{t,t}^{\{X_u=x_u\}}$ the *left-sided* $d_{\mathfrak{Q}}$*-derivative of* $T_{t,t}^{\{X_u=x_u\}}$. Finally, $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$*-differentiable* if it is $d_{\mathfrak{Q}}$-differentiable from the right and, if applicable, also from the left.

**Joseph's Example 3.47.** Let us return to the jump process $P_\phi$ as defined in Joseph's Example 3.43₉₂. Recall from Joseph's Example 3.44₉₅ that $\partial_+ T_{0,0}$ contains precisely two rate operators: $Q_1$ with $\lambda_{\mathrm{H}} = 1 = \lambda_{\mathrm{T}}$ and $Q_{1/2}$ with $\lambda_{\mathrm{H}} = 1/2 = \lambda_{\mathrm{T}}$. Consequently, it follows from Eq. (3.54)94 that, for all $t$ in $]0,1[$,

$$d_{\mathfrak{Q}}\big(Q_{0,t}, \partial_+ T_{0,0}\big) = \min\Big\{\big\|Q_1 - Q_{0,t}\big\|_{\mathrm{op}}, \big\|Q_{1/2} - Q_{0,t}\big\|_{\mathrm{op}}\Big\}.$$

From this and Eqs. (3.57)95 and (3.58)95, it follows immediately that $d_{\mathfrak{Q}}(Q_{0,t}, \partial_+ T_{0,0}) = 0$ for all $t$ in $]0,1[$, and therefore

$$\lim_{t\searrow 0} d_{\mathfrak{Q}}\big(Q_{0,t}, \partial_+ T_{0,0}\big) = 0.$$

[15] In the interest of clarity and conciseness, we have deliberatly chosen to cut some – formal – corners with this notation and terminology. Norte that we explicitly demand that the sets of accumulation points are non-empty, but we could also leave this implicit because $d_{\mathfrak{Q}}(Q, \emptyset) = +\infty$ for all $Q$ in $\mathfrak{Q}$.

Hence, $T_{0,0}$ is $d_{\mathfrak{Q}}$-differentiable from the right, with right-sided $d_{\mathfrak{Q}}$-derivative $\partial_+ T_{0,0} = \{Q_1, Q_{1/2}\}$. ♃

Of course, this definition of the directional $d_{\mathfrak{Q}}$-derivatives is only sensible if they coincide with the 'normal' directional derivatives whenever the latter exist. The following result establishes that this is indeed the case; it is similar to a result of Krak et al. (2017, Corollary 4.8) for their 'outer directional derivatives'. For the sake of conciseness, we have relegated the proof to Appendix 3.A.1₁₂₅.

**Lemma 3.48.** *Consider a jump process $P$, a current time point $t$ in $\mathbb{R}_{\geq 0}$, a state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t$ and a rate operator $Q$ in $\mathfrak{Q}$. Then*

$$\lim_{r \searrow t} \frac{T_{t,r}^{\{X_u=x_u\}} - I}{r - t} = Q$$

*if and only if $\partial_+ T_{t,t}^{\{X_u=x_u\}} = \{Q\}$ and $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable from the right. The same holds for the left-sided ($d_{\mathfrak{Q}}$-)derivative whenever $t > 0$.*

### Consistency with a set of rate operators

Before we formally establish the notion of consistency with a set of rate operators, it will be instructive to put the newly-introduced directional $d_{\mathfrak{Q}}$-derivatives to work in the setting of homogeneous Markov chains (see also Krak et al., 2017, Proposition 5.5).

**Corollary 3.49.** *Consider a probability mass function $p_0$ on $\mathscr{X}$ and a rate operator $Q$. Then $P_{p_0,Q}$ is the unique jump process $P$ such that (i) the initial mass function $P(X_0 = \bullet)$ is $p_0$; and (ii) for all $t$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t$, $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable with*

$$\partial_+ T_{t,t}^{\{X_u=x_u\}} = \{Q\} \quad \textit{and, if } t > 0, \quad \partial_- T_{t,t}^{\{X_u=x_u\}} = \{Q\}.$$

*Proof.* Follows from Theorem 3.37₈₇, Proposition 3.42₉₁ and Lemma 3.48. □

Due to the foregoing result, it is arguably sensible to say that a jump process is consistent with the set $\mathcal{Q}$ of rate operators – or, alternatively, that the dynamics of $P$ are characterized by $\mathcal{Q}$ – if the set-valued directional derivatives of its history-dependent transition operators are contained in $\mathcal{Q}$. Thus, we obtain the following definition, essentially taken from (Krak et al., 2017, Definition 6.1).[16]

[16] The difference between our definition and that of Krak et al. (2017, Definition 6.1) is that due to our more general definition of the 'set-valued directional derivative', we do not have to a priori limit ourselves to jump processes that have bounded rate, in the sense of Definition 3.53₁₀₁ further on.

**Definition 3.50.** Consider a non-empty set $\mathcal{Q}$ of rate operators. The jump process $P$ is *consistent with the set* $\mathcal{Q}$ if for all $t$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t$, $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable with

$$\partial_+ T_{t,t}^{\{X_u=x_u\}} \subseteq \mathcal{Q} \quad \text{and, if } t > 0, \quad \partial_- T_{t,t}^{\{X_u=x_u\}} \subseteq \mathcal{Q}.$$

We let $\mathbb{P}_{\mathcal{Q}}$ denote the set of all jump processes that are consistent with the set $\mathcal{Q}$ of rate operators.

Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty set $\mathcal{Q}$ of rate operators. Due to Theorem 3.37<sub></sub>$_{87}$ and Corollary 3.49, every jump process $P$ in $\mathbb{P}_{\mathscr{M},\mathcal{Q}}^{\mathrm{HM}}$ is a homogeneous and Markovian jump process that is consistent with $\mathcal{Q}$. Consequently,

$$\mathbb{P}_{\mathscr{M},\mathcal{Q}}^{\mathrm{HM}} \subseteq \mathbb{P}^{\mathrm{HM}} \cap \mathbb{P}_{\mathcal{Q}}. \tag{3.61}$$

Taking into account Eq. (3.47)$_{91}$, it follows that

$$\mathbb{P}_{\mathscr{M},\mathcal{Q}}^{\mathrm{HM}} \subseteq \mathbb{P}^{\mathrm{HM}} \cap \mathbb{P}_{\mathscr{M}} \cap \mathbb{P}_{\mathcal{Q}}. \tag{3.62}$$

To turn this inclusion into an equality, we need an additional assumption.

One possible such assumption is to restrict the jump processes on the right-hand side of Eq. (3.62) to those that satisfy the continuity condition of Theorem 3.35$_{86}$. It is not all too difficult to verify that this continuity condition is equivalent to requiring that the history-dependent transition operators are 'continuous'; that is, the jump processes on the right-hand side of Eq. (3.62) should additionally belong to

$$\mathbb{P}^{\mathrm{C}} := \Big\{ P \in \mathbb{P} \colon (\forall t \in \mathbb{R}_{\geq 0})(\forall \{X_u = x_u\} \in \mathscr{H}, u \prec t) \lim_{r \searrow t} T_{t,r}^{\{X_u=x_u\}} = I,$$
$$(\forall t \in \mathbb{R}_{>0})(\forall \{X_u = x_u\} \in \mathscr{H}, u \prec t) \lim_{s \nearrow t} T_{s,t}^{\{X_u=x_u\}} = I \Big\}.$$

The following result establishes this formally.

**Lemma 3.51.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty set $\mathcal{Q}$ of rate operators. Then*

$$\mathbb{P}_{\mathscr{M},\mathcal{Q}}^{\mathrm{HM}} = \mathbb{P}^{\mathrm{HM}} \cap \mathbb{P}_{\mathscr{M}} \cap \mathbb{P}_{\mathcal{Q}} \cap \mathbb{P}^{\mathrm{C}}.$$

*Proof.* Observe that for all $P$ in $\mathbb{P}^{\mathrm{C}}$, $x$ in $\mathscr{X}$ and $r$ in $\mathbb{R}_{>0}$,

$$|P(X_r = x \,|\, X_0 = x) - 1| = |T_{0,r}(x,x) - I(x,x)| \leq \|T_{0,r} - I\|_{\mathrm{op}},$$

where the first equality follows from Eq. (3.34)$_{84}$. From this and the definition of $\mathbb{P}^{\mathrm{C}}$, it follows that

$$\big(\forall P \in \mathbb{P}^{\mathrm{C}}\big)\big(\forall x \in \mathscr{X}\big) \lim_{r \searrow 0} P(X_r = x \,|\, X_0 = x) = 1.$$

Fix any jump process $P$ in $\mathbb{P}^{\mathrm{HM}} \cap \mathbb{P}^{\mathrm{C}}$. Due to the preceding, it follows from Theorem 3.35₈₆ and Theorem 3.37₈₇ that there is a unique probability mass function $p_0$ on $\mathscr{X}$ and a unique rate operator $Q$ such that $P = P_{p_0,Q}$. It follows from this and Corollary 3.49₉₈ that

$$\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}} \supseteq \mathbb{P}^{\mathrm{HM}} \cap \mathbb{P}_{\mathscr{M}} \cap \mathbb{P}_{\mathscr{Q}} \cap \mathbb{P}^{\mathrm{C}}.$$

Next, we fix any $P$ in $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$. Then by construction, $P = P_{p_0,Q}$ for some $p_0$ in $\mathscr{M}$ and $Q$ in $\mathscr{Q}$. Consequently, it follows from Theorem 3.37₈₇ and Proposition 3.24₇₉ that for all $t$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t$,

$$\lim_{r \searrow t} \frac{T^{\{X_u=x_u\}}_{t,r} - I}{r-t} = Q \quad \text{and, if } t > 0, \quad \lim_{s \nearrow t} \frac{T^{\{X_u=x_u\}}_{s,t} - I}{t-s} = Q;$$

due to a standard argument from analysis, this implies that

$$\lim_{r \searrow t} T^{\{X_u=x_u\}}_{t,r} = I \quad \text{and, if } t > 0, \quad \lim_{s \nearrow t} T^{\{X_u=x_u\}}_{s,t} = I.$$

Hence, $P$ belongs to $\mathbb{P}^{\mathrm{C}}$. In other words, $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathbb{P}^{\mathrm{C}}$. From this and Eq. (3.62)↶, we infer that

$$\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathbb{P}^{\mathrm{HM}} \cap \mathbb{P}_{\mathscr{M}} \cap \mathbb{P}_{\mathscr{Q}} \cap \mathbb{P}^{\mathrm{C}}.$$

We have shown that $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is included in and includes $\mathbb{P}^{\mathrm{HM}} \cap \mathbb{P}_{\mathscr{M}} \cap \mathbb{P}_{\mathscr{Q}} \cap \mathbb{P}^{\mathrm{C}}$, which implies the equality of the statement. ☐

Alternatively, and arguably more sensibly, we can require that $\mathscr{Q}$ is bounded, in the sense that

$$\|\mathscr{Q}\|_{\mathrm{op}} := \sup\{\|Q\|_{\mathrm{op}} : Q \in \mathscr{Q}\} < +\infty, \tag{3.63}$$

with some slight abuse of notation. Whenever $\mathscr{Q}$ is bounded, it turns out that

$$\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}} = \mathbb{P}^{\mathrm{HM}} \cap \mathbb{P}_{\mathscr{M}} \cap \mathbb{P}_{\mathscr{Q}}; \tag{3.64}$$

for a proof, see Proposition 3.56₁₀₃ further on.

### Bounded sets of rate operators

So why is it sensible to assume that $\mathscr{Q}$ is bounded? Intuitively, this 'bounds' the dynamics of the history-dependent transition operators of every jump process $P$ that is consistent with $\mathscr{Q}$. If we do not assume that $\mathscr{Q}$ is bounded, then the transition probabilities can change 'arbitrarily fast'. We can use our running example to illustrate this.

**Joseph's Example 3.52.** Fix an arbitrary initial mass function $p_0$ on $\mathscr{X} = \{\mathtt{H}, \mathtt{T}\}$. For any non-negative real number $\lambda$, we let $P_\lambda$ be the homogeneous Markovian jump process with initial mass function $p_0$ and rate operator $Q_\lambda$, with $Q_\lambda$ as defined in Joseph's Example 3.38₈₇:

$$Q_\lambda := \begin{pmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{pmatrix}.$$

Consider the set

$$\mathcal{Q} := \{Q_\lambda \colon \lambda \in [1, +\infty[\}$$

of rate operators $Q_\lambda$ with parameter $\lambda$ greater than or equal to 1. Observe that

$$\|\mathcal{Q}\|_{\mathrm{op}} = \sup\{\|Q_\lambda\|_{\mathrm{op}} \colon \lambda \in [1, +\infty[\} = \sup\{2\lambda \colon \lambda \in [1, +\infty[\} = +\infty$$

due to (R5)$_{81}$, so $\mathcal{Q}$ is unbounded.

Observe that by Eq. (3.46)$_{90}$ and the definition of $\mathcal{Q}$,

$$\mathbb{P}^{\mathrm{HM}}_{\{p_0\},\mathcal{Q}} = \{P_{p_0,Q} \colon Q \in \mathcal{Q}\} = \{P_{p_0,Q_\lambda} \colon \lambda \in [1, +\infty[\} = \{P_\lambda \colon \lambda \in [1, +\infty[\}.$$

Thus, for every $\lambda$ in $[1, +\infty[$, there is a jump process $P_\lambda$ in $\mathbb{P}^{\mathrm{HM}}_{\{p_0\},\mathcal{Q}}$.

Fix some $\lambda$ in $[1, +\infty[$. For all $t$ in $\mathbb{R}_{\geq 0}$, we let $T_{0,t}$ denote the transition operator from 0 to $t$ corresponding to $P_\lambda$ as defined by Eq. (3.33)$_{84}$. Then $T_{0,t} = e^{tQ_\lambda}$ for all $t$ in $\mathbb{R}_{\geq 0}$ by definition of $P_\lambda$ – see also Theorem 3.37$_{87}$ – and it therefore follows from Corollary 3.31$_{82}$ that

$$\lim_{t \searrow 0} \frac{T_{0,t} - I}{t} = Q_\lambda.$$

Because $I = T_{0,0}$ due to Eq. (3.37)$_{84}$, it follows that for all $x$ and $y$ in $\mathcal{X} = \{\mathtt{H}, \mathtt{T}\}$,

$$\lim_{t \searrow 0} \frac{P_\lambda(X_t = y \mid X_0 = x) - P(X_0 = y \mid X_0 = x)}{t} = \begin{cases} \lambda & \text{if } x \neq y \\ -\lambda & \text{if } x = y. \end{cases}$$

Because there is such a jump process $P_\lambda$ in $\mathbb{P}^{\mathrm{HM}}_{\{p_0\},\mathcal{Q}}$ for every $\lambda$ in $[1, +\infty[$, we infer from this equality that the rate of change of every transition probability can be made arbitrarily large (in absolute value). ♪

The following notion, taken from (Krak et al., 2017, Definition 4.4 and Proposition 4.5), is tightly related to consistency with a bounded set of rate operators; we state our definition using the 'history-dependent' rate operators, so in the spirit of (Krak et al., 2017, Proposition 4.5), but also provide a characterisation in terms of probabilities similar to (Krak et al., 2017, Definition 4.4).

**Definition 3.53.** A jump process $P$ has *bounded rate*[17] if for all $t$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t$,

$$\limsup_{r \searrow t} \left\| Q^{\{X_u = x_u\}}_{t,r} \right\|_{\mathrm{op}} < +\infty \quad \text{and, if } t > 0, \quad \limsup_{s \nearrow t} \left\| Q^{\{X_u = x_u\}}_{s,t} \right\|_{\mathrm{op}} < +\infty.$$

[17] As mentioned previously, Krak et al. (2017, Definition 4.4) use the term 'well-behaved' instead of the term bounded rate.

**Lemma 3.54.** *Consider a jump process $P$. Then $P$ has bounded rate if and only if for all $t$ in $\mathbb{R}_{\geq 0}$, $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t$ and $x$ in $\mathscr{X}$,*

$$\limsup_{r \searrow t} \frac{1}{r-t}\big(1 - P(X_r = x \,|\, X_u = x_u, X_t = x)\big) < +\infty$$

*and, if $t > 0$,*

$$\limsup_{s \nearrow t} \frac{1}{t-s}\big(1 - P(X_t = x \,|\, X_u = x_u, X_s = x)\big) < +\infty.$$

*Proof.* Fix any $s, r$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec s < r$. Recall from our discussion right after Eq. (3.52)$_{94}$ that $Q_{s,r}^{\{X_u=x_u\}}$ is a rate operator. Hence, it follows from (R5)$_{81}$ and Eqs. (3.52)$_{94}$ and (3.36)$_{84}$ that

$$\begin{aligned}
\big\|Q_{s,r}^{\{X_u=x_u\}}\big\|_{\mathrm{op}} &= 2\max\big\{-[Q_{s,r}^{\{X_u=x_u\}}\mathbb{I}_x](x) : x \in \mathscr{X}\big\} \\
&= 2\max\left\{-\frac{\big[T_{s,r}^{\{X_u=x_u\}}\mathbb{I}_x\big](x) - \mathbb{I}_x(x)}{r-s} : x \in \mathscr{X}\right\} \\
&= 2\max\left\{\frac{1}{r-s}\big(1 - P(X_r = x \,|\, X_u = x_u, X_s = x)\big) : x \in \mathscr{X}\right\}.
\end{aligned}$$

Because the state space $\mathscr{X}$ is finite, the statement follows almost immediately from this equality. □

Crucially, any jump process that is consistent with a bounded set of rate operators turns out to have bounded rate.

**Lemma 3.55.** *Consider a non-empty and bounded set $\mathscr{Q}$ of rate operators and a jump process $P$ that is consistent with $\mathscr{Q}$. Then for all $t$ in $\mathbb{R}_{\geq 0}$ and all $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t$,*

$$\limsup_{r \searrow t} \big\|Q_{t,r}^{\{X_u=x_u\}}\big\|_{\mathrm{op}} \leq \|\mathscr{Q}\|_{\mathrm{op}} \text{ and, if } t > 0, \limsup_{s \nearrow t} \big\|Q_{t,r}^{\{X_u=x_u\}}\big\|_{\mathrm{op}} \leq \|\mathscr{Q}\|_{\mathrm{op}}.$$

*Consequently, $P$ has bounded rate.*

*Proof.* That $P$ has bounded rate follows immediately from the two inequalities of the statement. Here we will only prove the first inequality of the statement, the proof for the second one is entirely analoguous.

First, we recall from the definition of consistency that $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$ differentiable from the right, and that $\partial_+ T_{t,t}^{\{X_u=x_u\}} \subseteq \mathscr{Q}$. Furthermore, we observe that for any $r$ in $\mathbb{R}_{\geq 0}$ with $r > t$ and any $Q$ in $\partial_+ T_{t,t}^{\{X_u=x_u\}}$,

$$\big\|Q_{t,r}^{\{X_u=x_u\}}\big\|_{\mathrm{op}} \leq \big\|Q_{t,r}^{\{X_u=x_u\}} - Q\big\|_{\mathrm{op}} + \|Q\|_{\mathrm{op}} \leq \big\|Q_{t,r}^{\{X_u=x_u\}} - Q\big\|_{\mathrm{op}} + \|\mathscr{Q}\|_{\mathrm{op}}, \tag{3.65}$$

where the final inequality holds because $Q \in \partial_+ T_{t,t}^{\{X_u=x_u\}} \subseteq \mathscr{Q}$.

Next, we fix any positive real number $\epsilon$. Because $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable from the right, there is a positive real number $\delta$ such that

$$\big(\forall r\in\,]t,t+\delta[\big)\ d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_u=x_u\}},\partial_+T_{t,t}^{\{X_u=x_u\}}\big)<\frac{\epsilon}{2}.$$

Furthermore, it follows from this and the definition of $d_{\mathfrak{Q}}$ that for any $r$ in $]t,t+\delta[$, there is a $Q_r$ in $\partial_+T_{t,t}^{\{X_u=x_u\}}$ such that

$$d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_u=x_u\}},\partial_+T_{t,t}^{\{X_u=x_u\}}\big)\le\big\|Q_{t,r}^{\{X_u=x_u\}}-Q_r\big\|_{\mathrm{op}}<d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_u=x_u\}},\partial_+T_{t,t}^{\{X_u=x_u\}}\big)+\frac{\epsilon}{2}.$$

Combining these two inequalities, we see that

$$\big(\forall r\in\,]t,t+\delta[\big)\ \big\|Q_{t,r}^{\{X_u=x_u\}}-Q_r\big\|_{\mathrm{op}}<\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.$$

It follows immediately from the preceding inequality and Eq. (3.65)↶ that

$$\big(\forall r\in\,]t,t+\delta[\big)\ \big\|Q_{t,r}^{\{X_u=x_u\}}\big\|_{\mathrm{op}}<\epsilon+\|\mathscr{Q}\|_{\mathrm{op}}.$$

Because $\epsilon$ was an arbitrary positive real number, we infer from this inequality that

$$\limsup_{r\searrow t}\big\|Q_{t,r}^{\{X_u=x_u\}}\big\|_{\mathrm{op}}\le\|\mathscr{Q}\|_{\mathrm{op}}<+\infty,$$

which verifies the first inequality of the statement. □

With the help of the preceding result and Lemma 3.51<sub>99</sub>, we can easily establish that Eq. (3.62)$_{99}$ holds with equality whenever $\mathscr{Q}$ is bounded.

**Proposition 3.56.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ or rate operators. Then*

$$\mathbb{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}=\mathbb{P}^{\mathrm{HM}}\cap\mathbb{P}_{\mathscr{M}}\cap\mathbb{P}_{\mathscr{Q}}.$$

*Proof.* Essential to this result is that any $P$ in $\mathbb{P}_{\mathscr{Q}}$ also belongs to $\mathbb{P}^{\mathrm{C}}$. To verify this, we fix any $P$ in $\mathbb{P}_{\mathscr{Q}}$, $t$ in $\mathbb{R}_{\ge0}$ and $\{X_u=x_u\}$ in $\mathscr{H}$ such that $u\prec t$. Then by Lemma 3.55↶,

$$\big(\forall\epsilon\in\mathbb{R}_{>0}\big)\big(\exists\delta\in\mathbb{R}_{>0}\big)\big(\forall r\in\,]t,t+\delta[\big)\ \big\|Q_{t,r}^{\{X_u=x_u\}}\big\|_{\mathrm{op}}<\|\mathscr{Q}\|_{\mathrm{op}}+\epsilon.$$

Observe that due to (N1)$_{76}$ and the definition of $Q_{t,r}^{\{X_u=x_u\}}$,

$$\big\|T_{t,r}^{\{X_u=x_u\}}-I\big\|_{\mathrm{op}}=(r-t)\|Q_{t,r}^{\{X_u=x_u\}}\|_{\mathrm{op}}.$$

From the preceding two observations, it follows almost immediately that

$$\big(\forall\epsilon\in\mathbb{R}_{>0}\big)\big(\exists\delta\in\mathbb{R}_{>0}\big)\big(\forall r\in\,]t,t+\delta[\big)\ \big\|T_{t,r}^{\{X_u=x_u\}}-I\big\|_{\mathrm{op}}<\epsilon,$$

so $\lim_{r\searrow t}T_{t,r}^{\{X_u=x_u\}}=I$. If applicable, a similar argument shows the same for the limit from the left. Because this holds any $P$ in $\mathbb{P}_{\mathscr{Q}}$, $t$ in $\mathbb{R}_{\ge0}$ and $\{X_u=x_u\}$ in $\mathscr{H}$ such that $u\prec t$, it follows immediately that $\mathbb{P}_{\mathscr{Q}}\subseteq\mathbb{P}^{\mathrm{C}}$. Therefore,

$$\mathbb{P}^{\mathrm{HM}}\cap\mathbb{P}_{\mathscr{M}}\cap\mathbb{P}_{\mathscr{Q}}=\mathbb{P}^{\mathrm{HM}}\cap\mathbb{P}_{\mathscr{M}}\cap\mathbb{P}_{\mathscr{Q}}\cap\mathbb{P}^{\mathrm{C}}.$$

The equality of the statement follows immediately from this and Lemma 3.51$_{99}$. □

As we have mentioned before, the notion of consistency with a set of rate operators favoured by Krak et al. (2017, Definition 6.1) is slightly different from ours. More precisely, they only define their 'outer directional derivatives' for jump processes that have bounded rate and only define consistency for jump process that have bounded rate, whereas we rely on the notion of $d_{\mathfrak{Q}}$-differentiability to ensure that the directional $d_{\mathfrak{Q}}$-derivatives exist. This being said, it follows from Lemma 3.55$_{102}$ and the next result that for bounded sets of rate operators, our notion of consistency coincides with theirs. The following result is based on (Krak et al., 2017, Propositions 4.6 and 4.7); its proof is rather lengthy, which is why we have relegated it to Appendix 3.A.3$_{131}$.

**Proposition 3.57.** *Consider a jump process $P$ that has bounded rate. Fix a time point $t$ in $\mathbb{R}_{\geq 0}$ and state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t$. Then $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable, $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ is bounded and*

$$\partial_+ T_{t,t}^{\{X_u=x_u\}} = \left\{Q \in \mathfrak{Q} \colon \left(\exists (r_n)_{n\in\mathbb{N}} \searrow t\right) \lim_{n\to+\infty} Q_{t,r_n}^{\{X_u=x_u\}} = Q\right\};^{18}$$

*whenever $t > 0$, $\partial_- T_{t,t}^{\{X_u=x_u\}}$ is bounded, and*

$$\partial_- T_{t,t}^{\{X_u=x_u\}} = \left\{Q \in \mathfrak{Q} \colon \left(\exists (s_n)_{n\in\mathbb{N}} \nearrow t\right) \lim_{n\to+\infty} Q_{s_n,t}^{\{X_u=x_u\}} = Q\right\}.^{19}$$

**Imprecise jump processes through consistency**

Proposition 3.56$_{\curvearrowleft}$ suggests two additional imprecise jump processes that are characterised by $\mathscr{M}$ and $\mathscr{Q}$: we can simply replace the set $\mathbb{P}^{\mathrm{HM}}$ of homogeneous Markovian jump process with the set $\mathbb{P}^{\mathrm{M}}$ of all (not necessarily homogeneous) Markovian jump processes or with the set $\mathbb{P}$ of all (not necessarily Markovian) jump processes. Thus, for any non-empty set $\mathscr{M}$ of initial distributions and any non-empty and *bounded* set $\mathscr{Q}$ of rate operators, we follow Krak et al. (2017, Definition 6.4) in defining three imprecise jump processes through consistency with $\mathscr{M}$ and $\mathscr{Q}$:

(i) $\mathbb{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}} = \mathbb{P}^{\mathrm{HM}} \cap \mathbb{P}_{\mathscr{M}} \cap \mathbb{P}_{\mathscr{Q}}$ consists of all consistent jump processes that are *Markovian and homogeneous*;

(ii) $\mathbb{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{M}} := \mathbb{P}^{\mathrm{M}} \cap \mathbb{P}_{\mathscr{M}} \cap \mathbb{P}_{\mathscr{Q}}$ consists of all consistent jump processes that are *Markovian*; and

(iii) $\mathbb{P}_{\mathscr{M},\mathscr{Q}} := \mathbb{P}_{\mathscr{M}} \cap \mathbb{P}_{\mathscr{Q}}$ consists of *all* consistent jump processes.

Whenever $\mathscr{M} = \{p_0\}$ and $\mathscr{Q} = \{Q\}$, it follows from Corollary 3.49$_{98}$ that

$$\mathbb{P}_{\mathscr{M},\mathscr{Q}} = \mathbb{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{M}} = \mathbb{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}} = \left\{P_{p_0,Q}\right\}. \tag{3.66}$$

[18] In this expression, we let $(r_n)_{n\in\mathbb{N}} \searrow t$ denote a sequence in $\mathbb{R}_{\geq 0}$ such that $t < r_{n+1} < r_n$ for all $n$ in $\mathbb{N}$ and $\lim_{n\to+\infty} r_n = t$.

[19] In this expression, we let $(s_n)_{n\in\mathbb{N}} \nearrow t$ denote a sequence in $\mathbb{R}_{\geq 0}$ such that $u \prec s_n < s_{n+1} < t$ for all $n$ in $\mathbb{N}$ and $\lim_{n\to+\infty} s_n = t$

Thus, each of the three imprecise jump processes $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ is a proper generalisation of the homogeneous Markovian jump process $P_{p_0,Q}$. In fact, we will see in Section 3.3.5$_{118}$ further on that these three imprecise jump processes are all Markovian and homogeneous, so we can rightfully call each of them a homogeneous Markovian imprecise jump processes. This being said, it is easy to see that in general,

$$\mathbb{P}_{\mathscr{M},\mathscr{Q}} \supseteq \mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \supseteq \mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}. \tag{3.67}$$

In order not to clutter our notation unnecessarily, we denote the conditional lower expectations corresponding to $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ by $\underline{E}_{\mathscr{M},\mathscr{Q}}$, $\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, respectively; we also use similar notation for the conjugate conditional upper expectation and the corresponding conditional lower and upper probabilities. It follows immediately from Eq. (3.67) that, for all $\{X_u = x_u\}$ in $\mathscr{H}$ and $f$ in $\mathbb{S}(\mathscr{F}_u)$,

$$\underline{E}_{\mathscr{M},\mathscr{Q}}(f \,|\, X_u = x_u) \leq \underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f \,|\, X_u = x_u) \leq \underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}(f \,|\, X_u = x_u). \tag{3.68}$$

These inequalities can be strict, and they often turn out to be so. We refer to Joseph's Example 4.13$_{171}$ further on or (Krak et al., 2017, Examples 6.2, 9.1 and 9.2) for examples where this is indeed the case. Whenever this is the case, the inclusions in Eq. (3.67) are also strict, meaning that the imprecise jump processes $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ are distinct!

### 3.3.2 The law of iterated lower expectations

Recall from Theorem 3.19$_{72}$ that a jump processes satisfies a law of iterated expectations. For imprecise jump processes, a similar concept exists, but it is a bit more involved.

To see why, we let $\mathscr{P}$ be any imprecise jump process. For all $u$ in $\mathscr{U}$, $v$ in $\mathscr{U}_{\neq()}$ such that $v \succcurlyeq u$ and $f$ in $\mathscr{X}_v$, we define the gamble

$$\underline{E}_{\mathscr{P}}(f(X_v) \,|\, X_u) := \sum_{x_u \in \mathscr{X}_u} \underline{E}_{\mathscr{P}}(f(X_v) \,|\, X_u = x_u) \mathbb{I}_{\{X_u = x_u\}}.$$

Note that this gamble is similar to the gamble $E_P(f(X_v) \,|\, X_u)$, and that, for all $P$ in $\mathscr{P}$ and $x_u$ in $\mathscr{X}_u$,

$$\underline{E}_{\mathscr{P}}(f(X_v) \,|\, X_u = x_u) \leq E_P(f(X_v) \,|\, X_u = x_u).$$

Fix some $u$ in $\mathscr{U}$, some $v$ and $w$ in $\mathscr{U}_{\neq()}$ such that $u \prec v$ and $u \cup v \preccurlyeq w$ and a gamble $f$ in $\mathbb{G}(\mathscr{X}_w)$. From the preceding inequality, Theorem 3.19$_{72}$ and (ES4)$_{37}$, we infer that for any jump process $P$ in the imprecise jump process $\mathscr{P}$ and any $x_u$ in $\mathscr{X}_u$,

$$\begin{aligned} E_P(f(X_w) \,|\, X_u = x_u) &= E_P(E_P(f(X_w) \,|\, X_{u \cup v}) \,|\, X_u = x_u) \\ &\geq E_P(\underline{E}_{\mathscr{P}}(f(X_w) \,|\, X_{u \cup v}) \,|\, X_u = x_u). \end{aligned}$$

Because inequalities are preserved under taking the infimum, we conclude that in general,

$$\underline{E}_{\mathscr{P}}(f(X_w) \mid X_u = x_u) \geq \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f(X_w) \mid X_{u\cup v}) \,\big|\, X_u = x_u\big).$$

In the special case that this inequality always holds with equality, we say that $\mathscr{P}$ satisfies the law of iterated lower expectations.

**Definition 3.58.** An imprecise jump process $\mathscr{P}$ satisfies the *law of iterated lower expectations* if for all $u$ in $\mathscr{U}$, all $v$ and $w$ in $\mathscr{U}_{\neq()}$ such that $u < v$ and $u \cup v \preccurlyeq w$, all $f$ in $\mathbb{G}(\mathscr{X}_w)$ and all $x_u$ in $\mathscr{X}_u$,

$$\underline{E}_{\mathscr{P}}(f(X_w) \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f(X_w) \mid X_{u\cup v}) \,\big|\, X_u = x_u\big). \tag{3.69}$$

The law of iterated lower expectations is rather strong, in the sense that many imprecise jump processes do not satisfy it. Fortunately, the following weaker form of this law already goes a long way.

**Definition 3.59.** An imprecise jump process $\mathscr{P}$ satisfies the *sum-product law of iterated lower expectations* if for all $u$ in $\mathscr{U}$, $v$ in $\mathscr{U}_{\neq()}$ and $t$ in $\mathbb{R}_{\geq 0}$ such that $u < v < t$, all $f$ in $\mathbb{G}(\mathscr{X})$, all $g, h$ in $\mathbb{G}(\mathscr{X}_v)$ such that $h \geq 0$ and and all $x_u$ in $\mathscr{X}_u$,

$$\begin{aligned}\underline{E}_{\mathscr{P}}&(f(X_t)h(X_v) + g(X_v) \mid X_u = x_u)\\ &= \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v) \,\big|\, X_u = x_u\big).\end{aligned} \tag{3.70}$$

In Theorem 3.88~120~ further on, we show that $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ satisfies the law of iterated lower expectations whenever $\mathscr{Q}$ is convex and has separately specified rows – in the sense of Definition 3.67~110~ further on. Similarly, Theorem 3.89~120~ further on establishes that both $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ satisfy the sum-product law of iterated lower expectations whenever $\mathscr{Q}$ has separately specified rows – so it need not be convex.

Krak et al. (2017, Section 9) essentially argue that whenever $\mathscr{P}$ satisfies the law of iterated lower expectations, we can compute conditional lower expectations of the general form $\underline{E}_{\mathscr{P}}(f(X_u, X_v) \mid X_u = x_u)$ through backwards recursion. This recursive method goes back to imprecise Markov chains – see (De Cooman et al., 2008, Section 4.5) and (De Cooman et al., 2009, Section 3) – and we will discuss it in depth in Section 4.1.4~170~ further on – and in particular in Proposition 4.11~170~. The gist of this method is that we can compute a lower expectation of the form $\underline{E}_{\mathscr{P}}(f(X_u, X_v) \mid X_u = x_u)$ by repeatedly computing (conditional) lower expectations of the following two basic types: 'transition lower expectations' of the form

$$\underline{E}_{\mathscr{P}}(g(X_r) \mid X_u = x_u, X_t = x) \quad \text{with } u < t < r,$$

and, if applicable, one 'initial lower expectation' of the form $\underline{E}_{\mathscr{P}}(g(X_0))$. Furthermore, for many variables, as we will see in Section 4.1.3~162~ further on – and in particular in Proposition 4.8₁₆₆ – we can use a similar method for Markovian imprecise jump processes that only satisfy the weaker sum-product law of iterated lower expectations. For this reason, these basic types of (conditional) lower expectations are of paramount importance. In the remainder of this section, we will take a closer look at the conditional ones.

### 3.3.3 Lower transition operators and lower rate operators

Whereas we used linear operators to study the conditional expectation – or, alternatively, the transition probabilities – for a single jump process, we need more general operators to study the conditional lower expectation for imprecise jump processes. More precisely, we need lower transition operators and lower rate operators, the non-negatively homogeneous and super-additive generalisations of transition operators and rate operators, respectively.

#### Lower transition operators

A lower transition operator is a generalisation of a transition operator that is only required to be non-negatively homogeneous and super-additive, instead of homogeneous and additive – see (De Cooman et al., 2008, Section 8) or (De Cooman et al., 2009, Section 3).

**Definition 3.60.** A *lower transition operator* $\underline{T}$ is an operator $\underline{T}\colon \mathbb{G}(\mathscr{X}) \to \mathbb{G}(\mathscr{X})$ such that

LT1. $\underline{T}f \geq \min f$ for all $f$ in $\mathbb{G}(\mathscr{X})$;

LT2. $\underline{T}(\mu f) = \mu \underline{T}f$ for all $\mu$ in $\mathbb{R}_{\geq 0}$ and $f$ in $\mathbb{G}(\mathscr{X})$;

LT3. $\underline{T}(f+g) \geq \underline{T}f + \underline{T}g$ for all $f$ and $g$ in $\mathbb{G}(\mathscr{X})$.

The operator $\overline{T}$ defined by $\overline{T}f := -\overline{T}(-f)$ for all $f$ in $\mathbb{G}(\mathscr{X})$ is called the *(conjugate) upper transition operator*. We collect all lower transition operators on $\mathscr{X}$ in $\underline{\mathfrak{T}}(\mathscr{X})$, and shorten this to $\underline{\mathfrak{T}}$ whenever the state space is clear from the context.

To see why this is a useful generalisation, we recall that for any transition operator $T$, $[T\bullet](x)$ is a coherent expectation due to the similarity of $(T1)_{80}$–$(T3)_{80}$ to $(E1)_{22}$–$(E3)_{22}$. This time around, (LT1)–(LT3) are similar to $(LE1)_{30}$–$(LE3)_{30}$. For this reason, and because $\mathbb{G}(\mathscr{X})$ is a linear space, we obtain the following corollary of Proposition $2.30_{31}$.

**Corollary 3.61.** *For any lower transition operator* $\underline{T}$ *and state* $x$ *in* $\mathscr{X}$,

$$[\underline{T}\bullet](x)\colon \mathbb{G}(\mathscr{X}) \to \mathbb{R}\colon f \mapsto [\underline{T}f](x)$$

*is a coherent lower expectation on* $\mathbb{G}(\mathscr{X})$.

Due to this corollary, it follows immediately from Proposition 2.29[30] that, for any lower transition operator $\underline{T}$,

LT4. $\min f \leq \underline{T} f \leq \overline{T} f \leq \max f$ for all $f$ in $\mathbb{G}(\mathscr{X})$;

LT5. $\underline{T}(f+\mu) = \underline{T} f + \mu$ for all $f$ in $\mathbb{G}(\mathscr{X})$ and $\mu$ in $\mathbb{R}$;

LT6. $\underline{T} f \leq \underline{T} g$ for all $f, g$ in $\mathbb{G}(\mathscr{X})$ with $f \leq g$;

LT7. $|\underline{T} f - \underline{T} g| \leq \overline{T}(|f-g|)$ for all $f, g$ in $\mathbb{G}(\mathscr{X})$.

Furthermore, it is well-known – see for example (Hermans et al., 2012, T8) or De Bock (2017b, L10 and L11) – that (LT7) and (LT4) more or less immediately imply that

LT8. $\|\underline{T} f - \underline{T} g\| \leq \|f-g\|$ for all $f, g$ in $\mathbb{G}(\mathscr{X})$;

LT9. $\|\underline{T} M - \underline{T} N\|_{\mathrm{op}} \leq \|M-N\|_{\mathrm{op}}$ for any two non-negatively homogeneous operators $M$ and $N$ on $\mathbb{G}(\mathscr{X})$.

Because $-\|f\| \leq \min f \leq \max f \leq \|f\|$, it follows from (LT4) and Eq. (3.27)[77] that $\|\underline{T}\|_{\mathrm{op}} \leq 1$. On the other hand, another immediate consequence of (LT4) is that $\underline{T}\mathbb{1}_{\mathscr{X}} = \mathbb{1}_{\mathscr{X}}$. Because $\|\mathbb{1}_{\mathscr{X}}\| = 1$, we conclude that

LT10. $\|\underline{T}\|_{\mathrm{op}} = 1$,

which generalises (T4)[81]. Finally, it is well-known – see, for example, (De Cooman et al., 2009, Appendix) or (Krak et al., 2017, Proposition 7.1) – that

LT11. for any two lower transition operators $\underline{T}$ and $\underline{S}$, their composition $\underline{T}\,\underline{S}$ is a lower transition operator as well.

Throughout this dissertation, we will often use this last property without explicitly mentioning it.

**Vacuous Example 3.62.** Let $\mathscr{X}$ be any non-empty and finite set. It is easy to verify that the operator $\underline{T}$ on $\mathbb{G}(\mathscr{X})$ defined by

$$\underline{T} f := \min f \quad \text{for all } f \in \mathbb{G}(\mathscr{X})$$

is a lower transition operator. ¤

With Section 2.2.4[27] in mind, we see that lower transition operators are not only non-linear generalisations of transition operators, but can also be interpreted as lower envelopes of sets of transition operators. More precisely, for any non-empty subset $\mathscr{T}$ of $\mathfrak{T}$, the corresponding lower envelope $\underline{T}_{\mathscr{T}} \colon \mathbb{G}(\mathscr{X}) \to \mathbb{G}(\mathscr{X})$ defined for all $f$ in $\mathbb{G}(\mathscr{X})$ and $x$ in $\mathscr{X}$ by

$$\left[\underline{T}_{\mathscr{T}} f\right](x) := \inf\left\{[Tf](x) \colon T \in \mathscr{T}\right\}$$

is a lower transition operator by Definition 2.25[28]. Conversely, for any lower transition operator $\underline{T}$, Theorem 2.28[30] implies that the set

$$\mathscr{T}_{\underline{T}} := \left\{T \in \mathfrak{T} \colon \left(\forall f \in \mathbb{G}(\mathscr{X})\right) \underline{T} f \leq T f\right\}$$

of dominating transition operators is non-empty, and that $\underline{T}$ is the lower envelope of $\mathscr{T}_{\underline{T}}$.

### Lower rate operators

Since we take lower envelopes of transition operators, it makes sense to also take lower envelopes of rate operators. Concretely, a non-empty and bounded[20] set $\mathcal{Q}$ of rate operators induces the lower envelope $\underline{Q}_{\mathcal{Q}} \colon \mathbb{G}(\mathcal{X}) \to \mathbb{G}(\mathcal{X})$, defined for all $f$ in $\mathbb{G}(\mathcal{X})$ and $x$ in $\mathcal{X}$ by

$$\big[\underline{Q}_{\mathcal{Q}} f\big](x) := \inf\big\{[Qf](x) \colon Q \in \mathcal{Q}\big\}. \tag{3.71}$$

It is easy to see – but will be formally established in Proposition 3.65 – that this lower envelope is what is known as a lower rate operator, see (Škulj, 2015, Section 2.5) or (De Bock, 2017b, Definition 5).

**Definition 3.63.** A *lower rate operator* $\underline{Q}$ is an operator $\underline{Q} \colon \mathbb{G}(\mathcal{X}) \to \mathbb{G}(\mathcal{X})$ such that

LR1. $\underline{Q}\mathbb{I}_{\mathcal{X}} = 0$;

LR2. $[\underline{Q}\mathbb{I}_y](x) \geq 0$ for all $x$ and $y$ in $\mathcal{X}$ such that $x \neq y$;

LR3. $\underline{Q}(\mu f) = \mu \underline{Q} f$ for all $\mu$ in $\mathbb{R}_{\geq 0}$ and $f$ in $\mathbb{G}(\mathcal{X})$;

LR4. $\underline{Q}(f+g) \geq \underline{Q}f + \underline{Q}g$ for all $f$ and $g$ in $\mathbb{G}(\mathcal{X})$.

The corresponding *(conjugate) upper rate operator* $\overline{Q}$ is defined by $\overline{Q}f := -\underline{Q}(-f)$ for all $f$ in $\mathbb{G}(\mathcal{X})$. We collect all lower rate operators in $\underline{\mathfrak{Q}}(\mathcal{X})$, and shorten this to $\underline{\mathfrak{Q}}$ whenever the state space is clear from the context.

In other words, a lower rate operator is the non-linear – that is, non-negatively homogeneous and super-additive – generalisation of a rate operator, in the same way that a lower transition operator is the non-linear generalisation of a transition operator.

**Joseph's Example 3.64.** Taking into account Joseph's Example 3.2882, it is not difficult to see that for $\mathcal{X} = \{\mathtt{H}, \mathtt{T}\}$, any lower rate operator $\underline{Q}$ is of the form

$$\begin{pmatrix} f(\mathtt{H}) \\ f(\mathtt{T}) \end{pmatrix} \mapsto \underline{Q}f = \begin{pmatrix} \min\big\{\lambda_{\mathtt{H}}(f(\mathtt{T}) - f(\mathtt{H})) \colon \lambda_{\mathtt{H}} \in \{\underline{\lambda}_{\mathtt{H}}, \overline{\lambda}_{\mathtt{H}}\}\big\} \\ \min\big\{\lambda_{\mathtt{T}}(f(\mathtt{H}) - f(\mathtt{T})) \colon \lambda_{\mathtt{T}} \in \{\underline{\lambda}_{\mathtt{T}}, \overline{\lambda}_{\mathtt{T}}\}\big\} \end{pmatrix},$$

where $\underline{\lambda}_{\mathtt{H}}$, $\overline{\lambda}_{\mathtt{H}}$ ,$\underline{\lambda}_{\mathtt{T}}$ and $\overline{\lambda}_{\mathtt{T}}$ are non-negative real numbers such that $\underline{\lambda}_{\mathtt{H}} \leq \overline{\lambda}_{\mathtt{H}}$ and $\underline{\lambda}_{\mathtt{T}} \leq \overline{\lambda}_{\mathtt{T}}$.

[20] This requirement ensures that $\underline{Q}_{\mathcal{Q}}$ is an operator – that is, that $\underline{Q}_{\mathcal{Q}} f$ takes values in $\mathbb{G}(\mathcal{X})$ – because

$$\begin{aligned} \|\underline{Q}_{\mathcal{Q}} f\| &= \max\big\{|[\underline{Q}_{\mathcal{Q}} f](x)| \colon x \in \mathcal{X}\big\} \leq \max\Big\{\sup\big\{|[Qf](x)| \colon Q \in \mathcal{Q}\big\} \colon x \in \mathcal{X}\Big\} \\ &\leq \sup\big\{\max\big\{|[Qf](x)| \colon x \in \mathcal{X}\big\} \colon Q \in \mathcal{Q}\big\} = \sup\big\{\|Qf\| \colon Q \in \mathcal{Q}\big\} \\ &\leq \sup\big\{\|Q\|_{\mathrm{op}} \colon Q \in \mathcal{Q}\big\}\|f\| < +\infty, \end{aligned}$$

where the penultimate inequality follows from (N4)77.

Krak et al. (2017, Proposition 7.5) formally establish that the lower envelope $\underline{Q}_{\mathcal{Q}}$ of a set $\mathcal{Q}$ of rate operators is a lower rate operator.

**Proposition 3.65.** *For any non-empty bounded set $\mathcal{Q}$ of rate operators, the corresponding lower envelope $\underline{Q}_{\mathcal{Q}}$ is a lower rate operator.*

Just like for lower transition operators, we are also interested in the set

$$\mathcal{Q}_{\underline{Q}} := \left\{Q \in \mathfrak{Q} \colon (\forall f \in \mathbb{G}(\mathscr{X}))\ \underline{Q}f \leq Qf\right\}. \tag{3.72}$$

of rate operators that dominate a lower rate operator $\underline{Q}$. Krak et al. (2017, Proposition 7.6) establish that this set $\mathcal{Q}_{\underline{Q}}$ of dominating rate operators has the following interesting properties.

**Lemma 3.66.** *Let $\underline{Q}$ be a lower rate operator. Then the induced set $\mathcal{Q}_{\underline{Q}}$ of dominating rate operators is non-empty and bounded. Furthermore, for every $f$ in $\mathbb{G}(\mathscr{X})$, there is a rate operator $Q$ in $\mathcal{Q}_{\underline{Q}}$ such that $\underline{Q}f = Qf$. Consequently, $\underline{Q}$ is the lower envelope of $\mathcal{Q}_{\underline{Q}}$.*

The preceding result establishes that the set $\mathcal{Q}_{\underline{Q}}$ of dominating rate operators is the largest set of rate operators that has $\underline{Q}$ as a lower envelope. This set has a number of other interesting properties. Before we state these, we need to introduce the concept of separately specified rows (Krak et al., 2017, Definition 7.3), which goes back to (Hartfiel, 1998), see also (Škulj, 2009, Section 2) and Hermans et al., 2014, Definition 11.6.

**Definition 3.67.** A non-empty set of rate operators $\mathcal{Q} \subseteq \mathfrak{Q}$ has *separately specified rows* if for every selection $(Q_x)_{x\in\mathscr{X}}$ in $\mathcal{Q}$, there is a rate operator $Q$ in $\mathcal{Q}$ such that

$$Q(x,y) = Q_x(x,y) \quad \text{for all } x,y \in \mathscr{X}.$$

**Joseph's Example 3.68.** Let us use the matrix representation of rate operators to clarify what it means for a set of rate operators to have separately specified rows. Fix non-negative real numbers $\lambda_1$ and $\lambda_2$ such that $\lambda_1 \neq \lambda_2$. Then

$$\left\{\begin{pmatrix}-\lambda_1 & \lambda_1\\ \lambda_2 & -\lambda_2\end{pmatrix}, \begin{pmatrix}-\lambda_2 & \lambda_2\\ \lambda_1 & -\lambda_1\end{pmatrix}\right\}$$

does *not* have separately specified rows, whereas

$$\left\{\begin{pmatrix}-\lambda_1 & \lambda_1\\ \lambda_2 & -\lambda_2\end{pmatrix}, \begin{pmatrix}-\lambda_2 & \lambda_2\\ \lambda_1 & -\lambda_1\end{pmatrix}, \begin{pmatrix}-\lambda_1 & \lambda_1\\ \lambda_1 & -\lambda_1\end{pmatrix}, \begin{pmatrix}-\lambda_2 & \lambda_2\\ \lambda_2 & -\lambda_2\end{pmatrix}, \right\}$$

does. ♪

It is not all too difficult to prove that the set of dominating rate operators is closed and convex and has separately specified rows (see Krak et al., 2017, Proposition 7.7); the same essentially holds for the set $\mathscr{T}_{\underline{T}}$ of dominating transition operators, see (Hermans et al., 2014, Proposition 11.4)

**Lemma 3.69.** *For any lower rate operator $\underline{Q}$, the corresponding set $\mathcal{Q}_{\underline{Q}}$ of dominating rate operators is closed and convex and has separately specified rows.*

Finally, Krak et al. (2017, Proposition 7.8) prove that $\mathcal{Q}_{\underline{Q}}$ is the only bounded, closed and convex set of rate operators with separately specified rows that has $\underline{Q}$ as lower envelope.

**Lemma 3.70.** *Consider a non-empty, bounded, closed and convex set $\mathcal{Q}$ of rate operators with separately specified rows. Then $\mathcal{Q}$ is equal to the set $\mathcal{Q}_{\underline{Q}}$ of rate operators that dominate the lower envelope $\underline{Q} := \underline{Q}_{\mathcal{Q}}$.*

**Joseph's Example 3.71.** Let $\underline{Q}$ be a lower rate operator that is characterised by the parameters $\underline{\lambda}_{\mathrm{H}}$, $\overline{\lambda}_{\mathrm{H}}$ ,$\underline{\lambda}_{\mathrm{T}}$ and $\overline{\lambda}_{\mathrm{T}}$, as in Joseph's Example 3.64$_{109}$. Note that

$$\mathcal{Q}_{\underline{Q}} = \left\{ \begin{pmatrix} -\lambda_{\mathrm{H}} & \lambda_{\mathrm{H}} \\ \lambda_{\mathrm{T}} & -\lambda_{\mathrm{T}} \end{pmatrix} : \lambda_{\mathrm{H}} \in [\underline{\lambda}_{\mathrm{H}}, \overline{\lambda}_{\mathrm{H}}], \lambda_{\mathrm{T}} \in [\underline{\lambda}_{\mathrm{T}}, \overline{\lambda}_{\mathrm{T}}] \right\}.$$

It is easy to see that the set $\mathcal{Q}_{\underline{Q}}$ of dominating rate operators has most of the established properties: it is non-empty, bounded and convex, and it has separately specified rows. To verify that $\mathcal{Q}_{\underline{Q}}$ is furthermore closed, one could use that a sequence of linear operators $(Q_n)_{n\in\mathbb{N}}$ converges if and only each of the component sequences $(Q_n(x,y))_{n\in\mathbb{N}}$ converges as well.

Alternatively, we can consider the set of rate operators

$$\mathcal{Q} := \left\{ \begin{pmatrix} -\lambda_{\mathrm{H}} & \lambda_{\mathrm{H}} \\ \lambda_{\mathrm{T}} & -\lambda_{\mathrm{T}} \end{pmatrix} : \lambda_{\mathrm{H}} \in \{\underline{\lambda}_{\mathrm{H}}, \overline{\lambda}_{\mathrm{H}}\}, \lambda_{\mathrm{T}} \in \{\underline{\lambda}_{\mathrm{T}}, \overline{\lambda}_{\mathrm{T}}\} \right\},$$

which is non-empty, bounded, (trivially) closed and has separately specified rows, but is *not* convex. Nonetheless, its lower envelope $\underline{Q}_{\mathcal{Q}}$ is also equal to $\underline{Q}$.

De Bock (2017b, (R5)–(R12)) lists a lot of interesting properties of lower rate operators. For now, we only repeat the following two properties:

LR5. $\underline{Q}f \leq \overline{Q}f$ for all $f$ in $\mathbb{G}(\mathscr{X})$;

LR6. $\underline{Q}(f+\mu) = \underline{Q}f$ for all $f$ in $\mathbb{G}(\mathscr{X})$ and $\mu$ in $\mathbb{R}$.

Lemmas 3.66 and 3.69 allow us to furthermore establish that for any lower rate operator $\underline{Q}$,

LR7. $\|\underline{Q}\|_{\mathrm{op}} = 2\max\{-[\underline{Q}\mathbb{I}_x](x) : x \in \mathscr{X}\} = \max\{\|Q\|_{\mathrm{op}} : Q \in \mathcal{Q}_{\underline{Q}}\} = \|\mathcal{Q}_{\underline{Q}}\|_{\mathrm{op}}$.

*Proof of* (LR7). First and foremost, we will need that

$$\|Q\|_{\mathrm{op}} \leq \|\underline{Q}\|_{\mathrm{op}} \quad \text{for all } Q \in \mathcal{Q}_{\underline{Q}}, \tag{3.73}$$

as is established by Krak et al. (2017, Lemma F.3).

Next, we fix any arbitrary positive real number $\epsilon$. By definition of the operator norm, there is a gamble $f$ on $\mathscr{X}$ with $\|f\| = 1$ such that

$$\|\underline{Q}\|_{\mathrm{op}} < \|\underline{Q}f\| + \epsilon.$$

By Lemma 3.66₁₁₀, there is a rate operator $Q$ in $\mathcal{Q}_{\underline{Q}}$ such that $\underline{Q}f = Qf$. Therefore,

$$\|\underline{Q}\|_{\mathrm{op}} < \|\underline{Q}f\| + \epsilon = \|Qf\| + \epsilon \leq \|Q\|_{\mathrm{op}}\|f\| + \epsilon = \|Q\|_{\mathrm{op}} + \epsilon,$$

where for the second inequality we have used (N4)$_{77}$. Because $Q$ is a rate operator in $\mathcal{Q}_{\underline{Q}}$ and $\epsilon$ an arbitrary positive real number, we infer from this that $\|\underline{Q}\|_{\mathrm{op}} \leq \sup\{\|Q\|_{\mathrm{op}} \colon Q \in \mathcal{Q}_{\underline{Q}}\}$. We combine this with Eq. (3.73)$_{\curvearrowleft}$, to conclude that

$$\|\underline{Q}\|_{\mathrm{op}} = \sup\{\|Q\|_{\mathrm{op}} \colon Q \in \mathcal{Q}_{\underline{Q}}\} = \|\mathcal{Q}_{\underline{Q}}\|_{\mathrm{op}}. \tag{3.74}$$

The first equality of the statement now follows after some straightforward manipulations:

$$\begin{aligned}\|\underline{Q}\|_{\mathrm{op}} &= \sup\{\|Q\|_{\mathrm{op}} \colon Q \in \mathcal{Q}_{\underline{Q}}\} = \sup\{2\max\{-[Q\mathbb{I}_x](x) \colon x \in \mathscr{X}\} \colon Q \in \mathcal{Q}_{\underline{Q}}\} \\ &= 2\max\{\sup\{-[Q\mathbb{I}_x](x) \colon Q \in \mathcal{Q}_{\underline{Q}}\} \colon x \in \mathscr{X}\} = 2\max\{-[\underline{Q}\mathbb{I}_x](x) \colon x \in \mathscr{X}\},\end{aligned}$$

where we have used (R5)$_{81}$ for the second equality and Lemma 3.69$_{\curvearrowleft}$ for the final equality.

Finally, we prove the second equality of the statement, that is, that the supremum is a maximum. To this end, we recall from Lemma 3.66₁₁₀ that for any $x$ in $\mathscr{X}$, there is a rate operator $Q_x$ in $\mathcal{Q}_{\underline{Q}}$ such that $Q_x\mathbb{I}_x = \underline{Q}\mathbb{I}_x$. Observe now that

$$\begin{aligned}\|\underline{Q}\|_{\mathrm{op}} &= 2\max\{-[\underline{Q}\mathbb{I}_x](x) \colon x \in \mathscr{X}\} = 2\max\{-[Q_x\mathbb{I}_x](x) \colon x \in \mathscr{X}\} \\ &\leq \max\{\|Q_x\|_{\mathrm{op}} \colon x \in \mathscr{X}\},\end{aligned}$$

where for the inequality we have used (R5)$_{81}$. On the other hand, we observe that due to Eq. (3.74),

$$\|\underline{Q}\|_{\mathrm{op}} = \sup\{\|Q\|_{\mathrm{op}} \colon Q \in \mathcal{Q}_{\underline{Q}}\} \geq \max\{\|Q_x\|_{\mathrm{op}} \colon x \in \mathscr{X}\},$$

where the inequality holds because $\{Q_x \colon x \in \mathscr{X}\} \subseteq \mathcal{Q}_{\underline{Q}}$. From the two preceding inequalities, we can conclude that the supremum in Eq. (3.74) is reached, and therefore is a maximum. This proves the second equality of the statement. □

Note that (LR7)$_{\curvearrowleft}$ generalises (R5)$_{81}$. Furthermore, it improves on a result by De Bock (2017b, R9), and this allows us to strengthen some of his results. For example, we can fully generalise Lemma 3.29$_{82}$.

**Lemma 3.72.** *Let $\underline{Q}$ be a lower rate operator. For all $\Delta$ in $\mathbb{R}_{\geq 0}$ such that $\Delta\|\underline{Q}\|_{\mathrm{op}} \leq 2$, $(I + \Delta\underline{Q})$ is a lower transition operator.*

*Proof.* That $\underline{T} := (I + \Delta\underline{Q})$ satisfies (LT2)$_{107}$ and (LT3)$_{107}$ follows immediately from (LR3)$_{109}$ and (LR4)$_{109}$, so we only need to check if $\underline{T}$ satisfies (LT1)$_{107}$. To that end, we fix an arbitrary $f$ in $\mathbb{G}(\mathscr{X})$. Observe that, for any $x$ in $\mathscr{X}$,

$$\begin{aligned}
[\underline{T}f](x) &= f(x) + \Delta[\underline{Q}f](x) = f(x) + \Delta\big[\underline{Q}(f - \min f)\big](x) \\
&= f(x) + \Delta\left[\underline{Q}\left(\sum_{y\in\mathscr{X}} (f(y) - \min f)\mathbb{I}_y\right)\right](x) \\
&\geq f(x) + \Delta\sum_{y\in\mathscr{X}} (f(y) - \min f)\big[\underline{Q}\mathbb{I}_y\big](x) \geq f(x) + \Delta(f(x) - \min f)\big[\underline{Q}\mathbb{I}_x\big](x) \\
&\geq f(x) - \Delta\frac{\|\underline{Q}\|_{\mathrm{op}}}{2}(f(x) - \min f),
\end{aligned}$$

where we have used (LR6)$_{111}$ for the second equality, (LR3)$_{109}$ and (LR4)$_{109}$ for the first inequality, (LR2)$_{109}$ for the second inequality and (LR7)$_{111}$ for the final inequality. Seeing that $\Delta\|\underline{Q}\|_{\mathrm{op}} \leq 2$ by the assumption of the statement, we infer from the preceding inequality that for all $x$ in $\mathscr{X}$,

$$[\underline{T}f](x) \geq f(x) - \Delta\frac{\|\underline{Q}\|_{\mathrm{op}}}{2}(f(x) - \min f) \geq f(x) - (f(x) - \min f) = \min f;$$

hence, $\underline{T}f \geq \min f$. Because $f$ was an arbitrary element of $\mathbb{G}(\mathscr{X})$, we have shown that $\underline{T}$ satisfies (LT1)$_{107}$. □

The preceding result allows us to establish the following two convenient properties, which strengthen (De Bock, 2017b, R11 and R12). For any lower rate operator $\underline{Q}$,

LR8. $\|\underline{Q}f - \underline{Q}g\| \leq \|\underline{Q}\|_{\mathrm{op}}\|f - g\|$ for all $f$ and $g$ in $\mathbb{G}(\mathscr{X})$;

LR9. $\|\underline{Q}M - \underline{Q}N\|_{\mathrm{op}} \leq \|\underline{Q}\|_{\mathrm{op}}\|M - N\|_{\mathrm{op}}$ for any two non-negatively homogeneous operators $M$ and $N$.

*Proof.* Our proof is entirely the same as that of (De Bock, 2017b, R11 and R12). The two properties clearly hold in case $\|\underline{Q}\|_{\mathrm{op}} = 0$, so we may assume $\|\underline{Q}\|_{\mathrm{op}} > 0$ without loss of generality. If we let $\Delta := 2/\|\underline{Q}\|_{\mathrm{op}}$ and $\underline{T} := (I + \Delta\underline{Q})$, then $\underline{T}$ is a lower transition operator by Lemma 3.72↶. The two properties now follow almost immediately from (LT8)$_{108}$ and (LT9)$_{108}$, if we observe that $\underline{Q} = (\underline{T}-I)/\Delta = \|\underline{Q}\|_{\mathrm{op}}(\underline{T}-I)/2$. □

Finally, and quite expectedly, Lemma 3.30$_{82}$ also generalises to the non-linear setting (see De Bock, 2017b, Proposition 6).

**Lemma 3.73.** *Let $\underline{T}$ be a lower transition operator. For all $\Delta$ in $\mathbb{R}_{>0}$, $(\underline{T}-I)/\Delta$ is a lower rate operator.*

### Generated semi-groups of lower transition operators

Lemmas 3.72↶ and 3.73 are not the only two connections between lower transition operators and lower rate operators. De Bock (2017b, Propositions 8 to 10) proves that, analogously to the linear case, lower rate operators generate a continuous semi-group of lower transition operators through the (non-linear) operator exponential.

**Proposition 3.74.** *Let $\underline{Q}$ be a lower rate operator. For any non-negative real number $t$ in $\mathbb{R}_{\geq 0}$,*

$$e^{t\underline{Q}} := \lim_{n\to+\infty}\left(I + \frac{t}{n}\underline{Q}\right)^n$$

*is a lower transition operator. Furthermore, $(e^{t\underline{Q}})_{t\in\mathbb{R}_{\geq 0}}$ is a continuous semi-group.*

Note that the preceding result generalises the second part of Corollary 3.3182 from rate operators and transition operators to lower rate operators and lower transition operators. This leads one to believe – or at least hope – that the first part generalises as well, and we establish that this is indeed the case. Because the proof of this result is rather long, we have chosen to relegate it to Appendix 3.B133.

**Theorem 3.75.** *Let $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ be a continuous semi-group of lower transition operators. Then*

$$\underline{Q} := \lim_{t\searrow 0}\frac{\underline{T}_t - I}{t}$$

*is a lower rate operator, and $\underline{Q}$ is the unique lower rate operator such that $\underline{T}_t = e^{t\underline{Q}}$ for all $t$ in $\mathbb{R}_{\geq 0}$.*

The condition that the semi-group is continuous might seem rather strong. The following result establishes that for a semi-group of lower transition operators, continuity is equivalent to continuity from the right at $t = 0$, and actually even to the continuity from the right at $t = 0$ of $[\underline{T}_t \mathbb{I}_x](x)$ for all $x$ in $\mathscr{X}$.

**Lemma 3.76.** *Consider a semi-group $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ of lower transition operators. Then the following three statements are equivalent.*

(i) *The semi-group $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ is continuous.*

(ii) *The semi-group $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ is continuous from the right at $t = 0$, in the sense that $\lim_{t\searrow 0}\underline{T}_t = I$.*

(iii) *For every state $x$ in $\mathscr{X}$, $\lim_{t\searrow 0}[\underline{T}_t\mathbb{I}_x](x) = 1$.*

*Proof.* Clearly, (i) implies (ii). Thus, to prove that (i) is equivalent to (ii), it suffices to prove that the latter implies the former. To see that this is indeed the case, we observe that for all $t$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $t < r$,

$$\|\underline{T}_r - \underline{T}_t\|_{\mathrm{op}} = \|\underline{T}_t\underline{T}_{r-t} - \underline{T}_t I\|_{\mathrm{op}} \leq \|\underline{T}_{r-t} - I\|_{\mathrm{op}},$$

where the inequality holds due to (LT9)$_{108}$. From this inequality and (ii)$_\curvearrowleft$, we infer that, for all $t$ in $\mathbb{R}_{\geq 0}$,

$$\lim_{r\searrow t} \|\underline{T}_r - \underline{T}_t\|_{\mathrm{op}} = 0 \quad \text{and, if } t > 0, \quad \lim_{s\nearrow t} \|\underline{T}_t - \underline{T}_s\|_{\mathrm{op}} = 0.$$

Therefore, (i)$_\curvearrowleft$ implies (ii)$_\curvearrowleft$, as required.

Next, we prove that (ii)$_\curvearrowleft$ and (iii)$_\curvearrowleft$ are equivalent. Recall from Lemma 3.73$_{113}$ with $\Delta = 1$ that for any $t$ in $\mathbb{R}_{>0}$, $(\underline{T}_t - I)$ is a lower rate operator. Thus, it follows from (LR7)$_{111}$ that for all $t$ in $\mathbb{R}_{>0}$,

$$\|\underline{T}_t - I\|_{\mathrm{op}} = 2\max\left\{-\left[(\underline{T}_t - I)\mathbb{I}_x\right](x) \colon x \in \mathscr{X}\right\} = 2\max\left\{-[\underline{T}_t\mathbb{I}_x](x) + 1 \colon x \in \mathscr{X}\right\}.$$

Because the state space $\mathscr{X}$ is finite, and because $[\underline{T}_t\mathbb{I}_x](x) \leq 1$ due to (LT4)$_{108}$, we infer from this equality that (ii)$_\curvearrowleft$ and (iii)$_\curvearrowleft$ are equivalent. □

Finally, De Bock (2017b, Proposition 9) proves that $(e^{t\underline{Q}})_{t\in\mathbb{R}_{\geq 0}}$ is the unique solution to a version of the operator initial value problem of Proposition 3.24$_{79}$ for lower rate operators.

**Proposition 3.77.** *Let $\underline{Q}$ be a lower rate operator. Then $(M_t)_{t\in\mathbb{R}_{\geq 0}} = (e^{t\underline{Q}})_{t\in\mathbb{R}_{\geq 0}}$ is the unique solution to the (non-negatively homogeneous operator) initial value problem*

$$\lim_{r\to t} \frac{M_r - M_t}{r - t} = \underline{Q}M_t \quad \textit{with } M_0 = I,$$

*where we only take the limit from the right for $t = 0$.*

To be more precise, De Bock (2017b, Proposition 9) only states that $(e^{t\underline{Q}})_{t\in\mathbb{R}_{\geq 0}}$ is a feasible solution for this initial value problem; that $(e^{t\underline{Q}})_{t\in\mathbb{R}_{\geq 0}}$ is the unique solution then follows from the following result by Škulj (2015, Corollary 2) (see also De Bock, 2017b, Section 6).

**Proposition 3.78.** *Let $\underline{Q}$ be a lower rate operator and $f$ a gamble on $\mathscr{X}$. Then the initial value problem*

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t = \underline{Q} f_t \quad \textit{with } f_0 = f$$

*has a unique solution that is given by $f_t = e^{t\underline{Q}} f$ for all $t$ in $\mathbb{R}_{\geq 0}$.*

**Joseph's Example 3.79.** Recall that in Joseph's Example 3.64$_{109}$, we defined the lower rate operator $\underline{Q}$ as

$$\begin{pmatrix} f(\mathrm{H}) \\ f(\mathrm{T}) \end{pmatrix} \mapsto \underline{Q}f = \begin{pmatrix} \min\left\{\lambda_{\mathrm{H}}(f(\mathrm{T}) - f(\mathrm{H})) \colon \lambda_{\mathrm{H}} \in \{\underline{\lambda}_{\mathrm{H}}, \overline{\lambda}_{\mathrm{H}}\}\right\} \\ \min\left\{\lambda_{\mathrm{T}}(f(\mathrm{H}) - f(\mathrm{T})) \colon \lambda_{\mathrm{T}} \in \{\underline{\lambda}_{\mathrm{T}}, \overline{\lambda}_{\mathrm{T}}\}\right\} \end{pmatrix}.$$

One can use Proposition 3.74$_\curvearrowleft$ (see Erreygers et al., 2017b, Example 2) to obtain that, for any $f$ in $\mathbb{G}(\mathscr{X})$,

$$e^{t\underline{Q}} f = f + \frac{1 - e^{-t\lambda_f}}{\lambda_f} \underline{Q} f \quad \text{for all } t \in \mathbb{R}_{\geq 0}, \tag{3.75}$$

where $\lambda_f \coloneqq \overline{\lambda}_{\mathrm{H}} + \underline{\lambda}_{\mathrm{T}}$ if $f(\mathrm{H}) \geq f(\mathrm{T})$ and $\lambda_f \coloneqq \underline{\lambda}_{\mathrm{H}} + \overline{\lambda}_{\mathrm{T}}$ if $f(\mathrm{H}) < f(\mathrm{T})$, and the second term is only added if $\lambda_f > 0$. Note the striking similarity between this expression – for which we are indebted to Troffaes et al. (2019, Eq. (16)) – and that of the operator exponential $e^{tQ}$ of the rate operator $Q$ as seen in Joseph's Example 3.32$_{83}$.

Similar to what we did there, we will verify that the expression for $e^{t\underline{Q}}f$ satisfies the initial value problem of Proposition 3.78. To that end, we fix some $f$ in $\mathbb{G}(\mathscr{X})$. Setting $t=0$ in Eq. (3.75), we obtain that $e^{0\underline{Q}}f = f$, as required. To verify that the differential equation holds, we fix some $t$ in $\mathbb{R}_{\geq 0}$. From here on, we assume that $\lambda_f > 0$; the alternate and straightforward case that $\lambda_f = 0$ is left as an exercise to the reader.

On the one hand, straightforward derivation of Eq. (3.75) yields

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{t\underline{Q}} f = e^{-t\lambda_f} \underline{Q} f.$$

We now set out to verify if this is equal to $\underline{Q} e^{t\underline{Q}} f$. In case $f(\mathrm{H}) \geq f(\mathrm{T})$, we see that

$$\begin{aligned}
\left[e^{t\underline{Q}}f\right](\mathrm{H}) - \left[e^{t\underline{Q}}f\right](\mathrm{T}) &= f(\mathrm{H}) + \left(1 - e^{t\lambda_f}\right)\frac{\overline{\lambda}_{\mathrm{H}}}{\overline{\lambda}_{\mathrm{H}} + \underline{\lambda}_{\mathrm{T}}}\left(f(\mathrm{T}) - f(\mathrm{H})\right) \\
&\qquad - f(\mathrm{T}) - \left(1 - e^{t\lambda_f}\right)\frac{\underline{\lambda}_{\mathrm{T}}}{\overline{\lambda}_{\mathrm{H}} + \underline{\lambda}_{\mathrm{T}}}\left(f(\mathrm{H}) - f(\mathrm{T})\right) \\
&= \left(f(\mathrm{H}) - f(\mathrm{T})\right)e^{-t\lambda_f},
\end{aligned}$$

and similar equalities hold in case $f(\mathrm{H}) < f(\mathrm{T})$. Therefore, we infer that $e^{t\underline{Q}}(\mathrm{H}) \geq e^{t\underline{Q}}(\mathrm{T})$ if $f(\mathrm{H}) \geq f(\mathrm{T})$ and $e^{t\underline{Q}}(\mathrm{H}) < e^{t\underline{Q}}(\mathrm{T})$ otherwise. In case $f(\mathrm{H}) \geq f(\mathrm{T})$, this implies that

$$\begin{aligned}
\left[\underline{Q} e^{t\underline{Q}} f\right](\mathrm{H}) &= \overline{\lambda}_{\mathrm{H}}\Big(\left[e^{t\underline{Q}}f\right](\mathrm{T}) - \left[e^{t\underline{Q}}f\right](\mathrm{H})\Big) = e^{-t\lambda_f}\overline{\lambda}_{\mathrm{H}}\left(f(\mathrm{T}) - f(\mathrm{H})\right) \\
&= e^{-t\lambda_f}\left[\underline{Q}f\right](\mathrm{H})
\end{aligned}$$

and analoguously,

$$\left[\underline{Q} e^{t\underline{Q}} f\right](\mathrm{T}) = e^{-t\lambda_f}\left[\underline{Q}f\right](\mathrm{T}).$$

Similar expressions hold in case $f(\mathrm{H}) < f(\mathrm{T})$, and we may therefore conclude that

$$\underline{Q} e^{t\underline{Q}} f = e^{-t\lambda_f}\underline{Q} f = \frac{\mathrm{d}}{\mathrm{d}t} e^{t\underline{Q}} f,$$

as it should be because of Proposition 3.78.

### 3.3.4 Back to imprecise jump processes

Interestingly, and quite crucially, the semi-group $(e^{t\underline{Q}})_{t\in\mathbb{R}_{\geq 0}}$ generated by the lower envelope $\underline{Q} \coloneqq \underline{Q}_{\mathscr{Q}}$ of a non-empty set of rate operators $\mathscr{Q}$ is related to the conditional lower expectation with respect to (subsets of) $\mathbb{P}_{\mathscr{Q}}$. For starters, Krak et al. (2017, Proposition 8.1) establish that it gives a lower bound – and, through conjugacy, also an upper bound – for the 'transition expectations' of a single consistent jump process.

**Proposition 3.80.** *Let $\mathscr{Q}$ be a non-empty bounded set of rate operators with lower envelope $\underline{Q} := \underline{Q}_{\mathscr{Q}}$, and $P$ a jump process that is consistent with $\mathscr{Q}$. Then for all $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t \leq r$, all $f$ in $\mathbb{G}(\mathscr{X})$ and all $x$ in $\mathscr{X}$,*

$$E_P(f(X_r) \mid X_u = x_u, X_t = x) \geq \left[e^{(r-t)\underline{Q}} f\right](x).$$

Recall that in Section 3.3.1$_{89}$, we defined the three imprecise jump processes $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ through consistency with $\mathscr{Q}$. Due to the preceding result, the semi-group generated by the lower envelope $\underline{Q}$ therefore also gives a lower bound for the 'transition lower expectations' with respect to $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Crucially, Krak et al. (2017, Corollary 8.3) establish that for the last two of these imprecise jump processes, this lower bound is reached whenever $\mathscr{Q}$ has separately specified rows.

**Proposition 3.81.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows, with lower envelope $\underline{Q} := \underline{Q}_{\mathscr{Q}}$. Then for all $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t \leq r$, all $f$ in $\mathbb{G}(\mathscr{X})$ and all $x$ in $\mathscr{X}$,*

$$\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f(X_r) \mid X_u = x_u, X_t = x) = \left[e^{(r-t)\underline{Q}} f\right](x)$$

*and*

$$\underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_r) \mid X_u = x_u, X_t = x) = \left[e^{(r-t)\underline{Q}} f\right](x),$$

*and, more generally, for any imprecise jump process $\mathscr{P}$ with $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$,*

$$\underline{E}_{\mathscr{P}}(f(X_r) \mid X_u = x_u, X_t = x) = \left[e^{(r-t)\underline{Q}} f\right](x).$$

Recall from Section 3.3.2$_{105}$ that we are also interested in the 'initial lower expectations' of the form $\underline{E}_{\mathscr{P}}(f(X_0))$. Here, we are especially interested in these lower expectations for imprecise jump processes that are consistent with a non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. Crucial is the lower expectation $\underline{E}_{\mathscr{M}}$[21] on $\mathbb{G}(\mathscr{X})$, defined by

$$\underline{E}_{\mathscr{M}}(f) := \inf\{E_p(f) : p \in \mathscr{M}\} \quad \text{for all } f \in \mathbb{G}(\mathscr{X}). \tag{3.76}$$

[21] Recall that in Section 2.2.4$_{27}$, we used $\underline{E}_{\mathcal{M}}$ to denote the lower envelope of a set $\mathcal{M}$ of coherent expectations. However, as we have mentioned before in Footnote 11$_{90}$, due to Proposition 2.18$_{23}$, every coherent expectation on $\mathbb{G}(\mathscr{X})$ is in one-to-one correspondence with a probability mass function because $\mathscr{X}$ is finite.

For every probability mass function $p$ on $\mathscr{X}$, $E_p$ is a coherent expectation on $\mathbb{G}(\mathscr{X})$ due to Corollary 2.17$_{23}$. Hence, $\underline{E}_{\mathscr{M}}$ is a coherent lower expectation by Definition 2.25$_{28}$. Krak et al. (2017, Proposition 9.3) establish that $\underline{E}_{\mathscr{M}}(f)$ is equal to the lower expectation of $f(X_0)$ for the three imprecise jump processes $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$.

**Proposition 3.82.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate operators. Then for all $f$ in $\mathbb{G}(\mathscr{X})$,*

$$\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}(f(X_0)) = \underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f(X_0)) = \underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_0)) = \underline{E}_{\mathscr{M}}(f),$$

*and, more generally, for any imprecise jump process $\mathscr{P}$ with $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M}}$,*

$$\underline{E}_{\mathscr{P}}(f(X_0)) = \underline{E}_{\mathscr{M}}(f).$$

Furthermore, (Krak et al., 2017, Proposition 9.4) also establish that for any future time point $t$, the (unconditional) lower expectation of $f(X_t)$ for the two imprecise jump process $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ is equal to $\underline{E}_{\mathscr{M}}(e^{t\underline{Q}}f)$ whenever $\mathscr{Q}$ has separately specified rows and lower envelope $\underline{Q} := \underline{Q}_{\mathscr{Q}}$.

**Proposition 3.83.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows, with lower envelope $\underline{Q} := \underline{Q}_{\mathscr{Q}}$. Then for any $t$ in $\mathbb{R}_{\geq 0}$ and $f$ in $\mathbb{G}(\mathscr{X})$,*

$$\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f(X_t)) = \underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_t)) = \underline{E}_{\mathscr{M}}(e^{t\underline{Q}}f),$$

*and, more generally, for any imprecise jump process $\mathscr{P}$ with $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$,*

$$\underline{E}_{\mathscr{P}}(f(X_t)) = \underline{E}_{\mathscr{M}}(e^{t\underline{Q}}f).$$

### 3.3.5 Homogeneous Markovian imprecise jump processes

The equality in Proposition 3.81$_{\curvearrowleft}$ is reminiscent of the (precise) Markov and homogeneity properties. That is to say, an imprecise jump process $\mathscr{P}$ that satisfies the conditions of Proposition 3.81$_{\curvearrowleft}$ satisfies imprecise versions of the Markov and homogeneity properties. In order to formalise this statement, we need to properly establish imprecise generalisations of these two properties.

**Definition 3.84.** An imprecise jump process $\mathscr{P}$ is *Markovian* – or, alternatively, has the *(imprecise) Markov property* – if for all time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and state histories $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t < r$, all $f$ in $\mathbb{G}(\mathscr{X})$ and all $x$ in $\mathscr{X}$,

$$\underline{E}_{\mathscr{P}}(f(X_r) \mid X_u = x_u, X_t = x) = \underline{E}_{\mathscr{P}}(f(X_r) \mid X_t = x).$$

Recall from Definition 3.21$_{74}$ that in the precise case, we impose the Markov property on the transition probabilities. However, because the Markovian character of the transition probabilities carries over to the conditional expectations through linearity, we might as well have imposed it on the conditional expectations. Thus, the preceding definition is a proper generalisation of the (precise) Markov property, in the sense that it reduces to Definition 3.21$_{74}$ in case $\mathscr{P} = \{P\}$. For a similar reason, the following properly generalises Definition 3.22$_{75}$.

**Definition 3.85.** A Markovian imprecise jump process $\mathscr{P}$ is *homogeneous* if for all time points $t$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $t < r$, all $f$ in $\mathbb{G}(\mathscr{X})$ and all $x$ in $\mathscr{X}$,

$$\underline{E}_{\mathscr{P}}(f(X_r) \mid X_t = x) = \underline{E}_{\mathscr{P}}(f(X_{r-t}) \mid X_0 = x).$$

Whereas in the precise case every – sufficiently continuous – homogeneous Markovian jump process is of the same 'type', in the sense that they are characterised by a single initial mass function and a single rate operator, this is not the case for homogeneous Markovian imprecise jump processes. Let us start with what is arguably the most straightforward example of a homogeneous Markovian imprecise jump process.

**Proposition 3.86.** *Consider a non-empty set $\mathscr{M}$ of initial probability mass functions and a non-empty set $\mathscr{Q}$ of rate operators. Then $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is Markovian and homogeneous. More generally, any imprecise jump process $\mathscr{P}$ that is a subset of $\mathbb{P}^{\mathrm{HM}}$ is Markovian and homogeneous.*

*Proof.* Recall from Eq. (3.47)$_{91}$ that $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is a subset of $\mathbb{P}^{\mathrm{HM}}$. For this reason, it suffices to prove that any subset $\mathscr{P}$ of $\mathbb{P}^{\mathrm{HM}}$ is Markovian and homogeneous. Because the corresponding lower envelope $\underline{E}_{\mathscr{P}}$ is taken over processes that are all Markovian and homogeneous, it is clear that the imprecise jump process $\mathscr{P}$ is Markovian and homogeneous as well. □

However, due to Proposition 3.81$_{117}$, we know that whenever $\mathscr{Q}$ is bounded and has separately specified rows, we do not need to impose homogeneity nor Markovianity on the processes in an imprecise jump process in order for the latter to be Markovian and homogeneous itself.

**Corollary 3.87.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows. Then $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ are Markovian and homogeneous. More generally, any imprecise jump process $\mathscr{P}$ with $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$ is Markovian and homogeneous.*

*Proof.* Follows immediately from Proposition 3.81$_{117}$. □

The preceding two results demonstrate that homogeneous Markovian imprecise jump process come in all shapes and sizes. For example, for any combination of a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows, it follows from Proposition 3.86 and Corollary 3.87 that $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ are all Markovian and homogeneous. This clearly illustrates that $\mathscr{M}$ and $\mathscr{Q}$ do not characterise a *unique* homogeneous Markovian imprecise jump process.

This being said, we can discriminate between various (homogeneous Markovian) imprecise jump processes based on whether or not they satisfy the law of iterated lower expectations or the sum-product law of iterated lower expectations. Let us start with the former. Krak et al. (2017, Theorem 6.5) show that $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ satisfies the law of iterated lower expectations whenever $\mathscr{Q}$ is bounded and convex, but their proof for this result contains an error. As we explain in Appendix 3.C$_{138}$, one way to fix their proof is to additionally require that $\mathscr{Q}$ has separately specified rows.

**Theorem 3.88.** *Let $\mathscr{M}$ be an arbitrary non-empty set of initial mass functions, and $\mathscr{Q}$ a non-empty, bounded and convex set of rate operators that has separately specified rows. Then $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ satisfies the law of iterated lower expectations.*

Unfortunately, the imprecise jump process $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ does not satisfy the law of iterated lower expectations, at least not in general; for a counterexample, see (Krak et al., 2017, Example 9.2).[22] Be that as it may, whenever $\mathscr{Q}$ is bounded and has separately specified rows, both $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ satisfy the weaker sum-product law of iterated lower expectations; because our proof is rather lengthy, we have relegated it to Appendix 3.D$_{151}$.

**Theorem 3.89.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows. Then $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ satisfy the sum-product law of iterated lower expectations.*

### Back to lower transition operators

Because transition operators were convenient tools when studying homogeneous Markovian jump processes, we now turn to lower transition operators to study homogeneous Markovian *imprecise* jump processes. To that end, we let $\mathscr{P}$ be a homogeneous Markovian imprecise jump process. For any time

[22] Krak et al. (2017) only make explicit that this example 'illustrates – by means of a counterexample – that (Krak et al., 2017, Algorithm 3) is not in general applicable' to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$. However, due to (Krak et al., 2017, Corollary 8.3), it also illustrates that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ does not satisfy the law of iterated lower expectations.

point $t$ in $\mathbb{R}_{\geq 0}$, we define the corresponding operator $\underline{T}_t\colon \mathbb{G}(\mathscr{X}) \to \mathbb{G}(\mathscr{X})$ by

$$[\underline{T}_t f](x) \coloneqq \underline{E}_{\mathscr{P}}(f(X_t) \,|\, X_0 = x) \quad \text{for all } f \in \mathbb{G}(\mathscr{X}), x \in \mathscr{X}. \tag{3.77}$$

Observe that, by construction, $\underline{T}_t$ is a lower transition operator.

Recall from Lemma 3.33$_{85}$ that for a homogeneous Markovian jump process, $(T_t)_{t\in\mathbb{R}_{\geq 0}}$ is a semi-group of transition operators. This generalises to the family $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ corresponding to a homogeneous Markovian imprecise jump process $\mathscr{P}$ whenever the latter satisfies the sum-product law of iterated lower expectations.

**Lemma 3.90.** *Let $\mathscr{P}$ be a homogeneous Markovian imprecise jump process. If $\mathscr{P}$ satisfies the sum-product law of iterated lower expectations, then the corresponding sequence $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ of lower transition operators is a semi-group.*

*Proof.* That $\underline{T}_t$ is a lower transition operator follows immediately from its definition and that of $\underline{E}_{\mathscr{P}}$ as a lower envelope. It remains for us to show that $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ is a semi-group. Let us start with verifying (SG1)$_{77}$. To that end, we fix any $f$ in $\mathbb{G}(\mathscr{X})$ and $x$ in $\mathscr{X}$. Observe that, for any $P$ in $\mathscr{P}$,

$$E_P(f(X_0) \,|\, X_0 = x) = \sum_{y\in\mathscr{X}} f(y) P(X_0 = y \,|\, X_0 = x) = f(x),$$

where we have used (CP1)$_{41}$ and (CP5)$_{42}$ for the second equality. Clearly, this implies that

$$\big[\underline{T}_0 f\big](x) = \underline{E}_{\mathscr{P}}(f(X_0) \,|\, X_0 = x) = \inf\big\{E_P(f(X_0) \,|\, X_0 = x)\colon P \in \mathscr{P}\big\} = f(x).$$

Because this equality holds for every gamble $f$ on $\mathscr{X}$ and any state $x$, we may conclude that $\underline{T}_0 = I$, which verifies (SG1)$_{77}$.

Next, we verify (SG2)$_{78}$. To that end, we fix any time points $s$ and $t$ in $\mathbb{R}_{\geq 0}$. It $s = 0$ or $t = 0$, it follows immediately from the previous that $\underline{T}_{s+t} = \underline{T}_s \underline{T}_t$. Hence, from now on we assume that $s > 0$ and $t > 0$. Fix any $f$ in $\mathbb{G}(\mathscr{X})$ and $x$ in $\mathscr{X}$. Because $\mathscr{P}$ satisfies the sum-product law of iterated lower expectations, it follows from Eq. (3.70)$_{106}$ in Definition 3.59$_{106}$ with $v = (s)$, $g = 0$ and $h = 1$ that

$$\big[\underline{T}_{s+t} f\big](x) = \underline{E}_{\mathscr{P}}(f(X_{s+t}) \,|\, X_0 = x) = \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(f(X_{s+t}) \,|\, X_0, X_s) \,|\, X_0 = x).$$

Note that because $\mathscr{P}$ is Markovian and homogeneous,

$$\begin{aligned}
\underline{E}_{\mathscr{P}}(f(X_{s+t}) \,|\, X_0, X_s) &= \sum_{y\in\mathscr{X}} \sum_{z\in\mathscr{X}} \underline{E}_{\mathscr{P}}(f(X_{s+t}) \,|\, X_0 = y, X_s = z) \mathbb{I}_{(X_0 = y, X_s = z)} \\
&= \sum_{y\in\mathscr{X}} \sum_{z\in\mathscr{X}} \underline{E}_{\mathscr{P}}(f(X_{s+t}) \,|\, X_s = z) \mathbb{I}_{(X_0 = y, X_s = z)} \\
&= \sum_{y\in\mathscr{X}} \sum_{z\in\mathscr{X}} \underline{E}_{\mathscr{P}}(f(X_t) \,|\, X_0 = z) \mathbb{I}_{(X_0 = y, X_s = z)} \\
&= \sum_{y\in\mathscr{X}} \sum_{z\in\mathscr{X}} [\underline{T}_t f](z) \mathbb{I}_{(X_0 = y, X_s = z)} \\
&= \sum_{z\in\mathscr{X}} [\underline{T}_t f](z) \mathbb{I}_{(X_s = z)} = \big[\underline{T}_t f\big](X_s).
\end{aligned}$$

We combine the two preceding equalities, to yield

$$[\underline{T}_{s+t}f](x) = \underline{E}_{\mathscr{P}}\big([\underline{T}_t f](X_r) \,\big|\, X_0 = x\big) = [\underline{T}_s\underline{T}_t f](x).$$

Seeing that this equality holds for every gamble $f$ on $\mathscr{X}$ and every state $x$ in $\mathscr{X}$, we conclude that $\underline{T}_{s+t} = \underline{T}_s\underline{T}_t$, which verifies (SG2)$_{78}$. □

### Unique characterisation

We end this chapter with some observations regarding the characterisation of homogeneous Markovian jump processes that satisfy the (sum-product) law of iterated lower expectations. We start with a generalisation of Theorem 3.35$_{86}$.

**Corollary 3.91.** *Let $\mathscr{P}$ be a homogeneous Markovian imprecise jump process that satisfies the sum-product law of iterated lower expectations. If*

$$\lim_{t\searrow 0}\underline{P}_{\mathscr{P}}(X_t = x \,|\, X_0 = x) = 1 \quad \textit{for all } x \in \mathscr{X},$$

*then there is a unique coherent lower expectation $\underline{E}_0$ on $\mathbb{G}(\mathscr{X})$ such that for all $f$ in $\mathbb{G}(\mathscr{X})$,*

$$\underline{E}_{\mathscr{P}}(f(X_0)) = \underline{E}_0(f) \tag{3.78}$$

*and there is a unique lower rate operator $\underline{Q}$ such that for all $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and all $\{X_u = x_u\}$ in $\mathscr{H}$ with $u < t \leq r$, all $f$ in $\mathbb{G}(\mathscr{X})$ and all $x$ in $\mathscr{X}$,*

$$\underline{E}_{\mathscr{P}}(f(X_r) \,|\, X_u = x_u, X_t = x) = \big[e^{(r-t)\underline{Q}}f\big](x). \tag{3.79}$$

*Proof.* It follows almost immediately from Proposition 2.42$_{37}$ and Definition 2.25$_{28}$ that the real-valued map $\underline{E}_0$ on $\mathbb{G}(\mathscr{X})$ defined by

$$\underline{E}_0(f) \coloneqq \underline{E}_{\mathscr{P}}(f(X_0)) \quad \text{for all } f \in \mathbb{G}(\mathscr{X})$$

is a coherent lower expectation, and it is easy to see that it is the unique coherent lower expectation that satisfies the condition of the statement.

To verify the second part of the statement, we recall from Lemma 3.90$_{\curvearrowleft}$ that the family $(\underline{T}_\tau)_{\tau\in\mathbb{R}_{\geq 0}}$ of lower transition operators as defined in Eq. (3.77)$_{\curvearrowleft}$ is a semi-group of lower transition operators. This semi-group is continuous due to Lemma 3.76$_{114}$; thus, it follows immediately from Theorem 3.75$_{114}$ that there is a unique lower rate operator $\underline{Q}$ such that $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}} = (e^{t\underline{Q}})_{t\in\mathbb{R}_{\geq 0}}$. Consequently, for any $t, r$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ with $u < t \leq r$, $f$ in $\mathbb{G}(\mathscr{X})$ and $x$ in $\mathscr{X}$,

$$\underline{E}_{\mathscr{P}}(f(X_r) \,|\, X_u = x_u, X_t = x) = \underline{E}_{\mathscr{P}}(f(X_{r-t}) \,|\, X_0 = x) = \big[\underline{T}_{t-r}f\big](x), = \big[e^{(t-t)\underline{Q}}f\big](x),$$

where the first equality holds because $\mathscr{P}$ is homogeneous and Markovian. This verifies the second part of the statement, and finalises our proof. □

Not only do we require a weak form of continuity, but also that the imprecise jump process $\mathscr{P}$ should satisfy the sum-product law of iterated lower expectations. If $\mathscr{P}$ furthermore satisfies the law of iterated lower expectations, the corresponding conditional lower expectation $\underline{E}_{\mathscr{P}}$ is actually completely determined by the unique coherent lower expectation $\underline{E}_0$ and the lower rate operator $\underline{Q}$; indeed, this follows from Corollary 3.91 and the following result.

**Corollary 3.92.** *Consider two imprecise jump processes $\mathscr{P}_1$ and $\mathscr{P}_2$ with corresponding conditional lower expectations $\underline{E}_1 := \underline{E}_{\mathscr{P}_1}$ and $\underline{E}_2 := \underline{E}_{\mathscr{P}_2}$. If $\mathscr{P}_1$ and $\mathscr{P}_2$ both satisfy the law of iterated lower expectations, then their corresponding conditional lower expectations $\underline{E}_1$ and $\underline{E}_2$ are equal if and only if for all $f$ in $\mathbb{G}(\mathscr{X})$,*

$$\underline{E}_1(f(X_0)) = \underline{E}_2(f(X_0))$$

*and, for all $t$ and $r$ in $\mathbb{R}_{\geq 0}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t < r$, all $f$ in $\mathbb{G}(\mathscr{X})$ and all $x$ in $\mathscr{X}$,*

$$\underline{E}_1(f(X_r) \,|\, X_u = x_u, X_t = x) = \underline{E}_2(f(X_r) \,|\, X_u = x_u, X_t = x).$$

*Proof.* The direct implication is trivial, and the converse implication follows immediately from Proposition 4.11$_{170}$ further on, which we may use without any issues because it only needs the law of iterated lower expectations. □

Corollary 3.92 is similar to Proposition 3.20$_{73}$. The difference is that in the imprecise case, it is not necessarily the two imprecise jump processes that are equal but only their corresponding conditional lower expectations. That this is possible should not come as a surprise, because we have already seen in Bruno's Example 2.27$_{29}$ that different sets of expectations can have the same lower envelope.

Next, because Corollary 3.91 can be seen as a generalisation of Theorem 3.35$_{86}$, we wonder whether Theorem 3.37$_{87}$ also generalises. To that end, we consider any non-empty set $\mathscr{M}$ of distributions and any non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows. We already know that $\mathscr{M}$ and $\mathscr{Q}$ do not characterise a unique homogeneous and Markovian imprecise jump processes whenever $\mathscr{Q}$ is not a singleton. More precisely, in Corollaries 3.86$_{119}$ and 3.87$_{119}$ we have established that the imprecise jump processes $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ – which are distinct in general, as we have argued right after Eq. (3.68)$_{105}$ – are all Markovian and homogeneous. Even more, it follows from Theorem 3.88$_{120}$, Propositions 3.81$_{117}$ and 3.82$_{118}$ and Corollary 3.92 that $\underline{E}_{\mathscr{M},\mathscr{Q}} = \underline{E}_{\mathscr{M}',\mathscr{Q}'}$ for any non-empty set $\mathscr{M}'$ of initial mass functions such that $\underline{E}_{\mathscr{M}} = \underline{E}_{\mathscr{M}'}$ and any non-empty and bounded set $\mathscr{Q}'$ of rate operators that has separately specified rows such that $\underline{Q}_{\mathscr{Q}} = \underline{Q}_{\mathscr{Q}'}$. However, the following result establishes that for a given initial coherent lower expectation $\underline{E}_0$ and a given lower rate operator $\underline{Q}$, there always is a largest imprecise jump process that satisfies Eqs. (3.78) and (3.79); in this sense, it is the 'converse' of Corollary 3.91$_{122}$ in a similar manner as Theorem 3.37$_{87}$ is the converse of Theorem 3.35$_{86}$.

**Proposition 3.93.** *Consider an initial coherent lower expectation $\underline{E}_0$ on $\mathbb{G}(\mathscr{X})$ and a lower rate operator $\underline{Q}$. Then there is at least one imprecise jump process $\mathscr{P}$ such that*

$$\big(\forall f \in \mathbb{G}(\mathscr{X})\big)\ \underline{E}_{\mathscr{P}}(f(X_0)) = \underline{E}_0(f)$$

*and*

$$\big(\forall t, r \in \mathbb{R}_{\geq 0}, t \leq r\big)\big(\forall u \in \mathscr{U}, u \prec t\big)\big(\forall x_u \in \mathscr{X}_u\big)\big(\forall f \in \mathbb{G}(\mathscr{X})\big)\big(\forall x \in \mathscr{X}\big)$$
$$\underline{E}_{\mathscr{P}}(f(X_r) \mid X_u = x_u, X_t = x) = \big[e^{(r-t)\underline{Q}} f\big](x),$$

*and the largest such $\mathscr{P}$ is $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ with $\mathscr{Q} := \mathscr{Q}_{\underline{Q}}$ and*

$$\mathscr{M} := \Big\{p \in \Sigma_{\mathscr{X}} \colon \big(\forall f \in \mathbb{G}(\mathscr{X})\big)\ E_p(f) \geq \underline{E}_0(f)\Big\}.$$

*Clearly, any such imprecise jump process is Markovian and homogeneous.*

*Proof.* Recall from Theorem 2.28$_{30}$ that because $\underline{E}_0$ is a coherent lower expectation on $\mathbb{G}(\mathscr{X})$, the set $\mathcal{M}_{\underline{E}_0}$ of coherent expectations on $\mathbb{G}(\mathscr{X})$ that dominate $\underline{E}_0$ is non-empty and that

$$\big(\forall f \in \mathbb{G}(\mathscr{X})\big)\ \underline{E}_0(f) = \min\{E(f) \colon E \in \mathcal{M}_{\underline{E}_0}\}.$$

By Proposition 2.18$_{23}$, every coherent expectation $E$ on $\mathbb{G}(\mathscr{X})$ corresponds to a unique probability mass function $p$ on $\mathscr{X}$ – in the sense that $E = E_p$. Because furthermore $E_p$ is a coherent expectation on $\mathbb{G}(\mathscr{X})$ for all $p$ in $\Sigma_{\mathscr{X}}$ due to Corollary 2.17$_{23}$, it is clear that $\mathscr{M}$ is non-empty and that

$$\big(\forall f \in \mathbb{G}(\mathscr{X})\big)\ \underline{E}_0(f) = \min\{E_p(f) \colon p \in \mathscr{M}\}. \tag{3.80}$$

Furthermore, we recall from Lemma 3.66$_{110}$ that $\mathscr{Q}$ is non-empty and bounded and that its lower envelope is $\underline{Q}_{\mathscr{Q}} = \underline{Q}$, and from Lemma 3.69$_{111}$ that $\mathscr{Q}$ has separately specified rows. Therefore, it follows immediately from Propositions 3.81$_{117}$ and 3.82$_{118}$ that $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ satisfies the two conditions of the statement. Thus, it remains for us to prove that $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ is the largest imprecise jump process that satisfies the two conditions of the statement. To this end, we let $\mathscr{P}$ be any imprecise jump process that satisfies the two conditions, and prove that $\mathscr{P}$ is contained in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$.

First, observe that for all $P$ in $\mathscr{P}$,

$$\big(\forall f \in \mathbb{G}(\mathscr{X})\big)\ E_{P(X_0=\bullet)}(f) = E_P(f(X_0)) \geq \underline{E}_{\mathscr{P}}(f(X_0)) = \underline{E}_0(f).$$

Due to Eq. (3.80), this implies that $P(X_0 = \bullet)$ belongs to $\mathscr{M}$ for all $P$ in $\mathscr{P}$. In other words, $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{M}}$.

Next, we recall that Krak et al. (2017, Theorem 8.4) have proven that $\mathbb{P}_{\mathscr{Q}}$ is the largest imprecise jump process that satisfies the second condition of the statement. Thus, $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$. Because we have previously shown that $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{M}}$, we may conclude that $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{M}} \cap \mathbb{P}_{\mathscr{Q}} = \mathbb{P}_{\mathscr{M},\mathscr{Q}}$, which is what we needed to prove. □

## 3.A Set-valued directional derivatives and the Hausdorff function

The first part of the appendix for this chapter consists of three parts. In Appendix 3.A.1, we establish some properties of the directional $d_{\mathfrak{Q}}$-derivatives. Next, we provide an alternative definition of $d_{\mathfrak{Q}}$-differentiability from the point of view of the Hausdorff function in Appendix 3.A.2$_{128}$. We subsequently use this alternative definition to prove Proposition 3.57$_{104}$ in Appendix 3.A.3$_{131}$.

### 3.A.1 Properties of the set-valued directional derivatives

We establish two properties of the sets $\partial_+ T_{t,r}^{\{X_u=x_u\}}$ and $\partial_- T_{t,r}^{\{X_u=x_u\}}$ of accumulation points in general; in particular, these hold for the directional $d_{\mathfrak{Q}}$-derivatives. First, we establish an alternative expression for the sets of accumulation points.

**Lemma 3.94.** *Consider a jump process $P$, a current time point $t$ in $\mathbb{R}_{\geq 0}$ and a state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t$. Then for any rate operator $Q$ in $\mathfrak{Q}$, the following three statements are equivalent.*

(i) *$Q$ belongs to $\partial_+ T_{t,t}^{\{X_u=x_u\}}$;*

(ii) $\big(\forall r \in \,]t,+\infty[\big)\ d_{\mathfrak{Q}}\big(Q, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in\,]t,r[}\big) = 0$;

(iii) $\big(\forall r \in \,]t,+\infty[\big)\big(\forall \epsilon \in \mathbb{R}_{>0}\big)\big(\exists r' \in \,]t,r[\big)\ \big\|Q - Q_{t,r'}^{\{X_u=x_u\}}\big\|_{\mathrm{op}} < \epsilon$.

*A similar statement holds for $\partial_- T_{t,t}^{\{X_u=x_u\}}$ whenever $t > 0$.*

*Proof.* Note that for any rate operator $Q$ in $\mathfrak{Q}$ and any time point $r$ in $]t,+\infty[$,

$$d_{\mathfrak{Q}}\big(Q, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in\,]t,r[}\big)$$

is non-decreasing as $r$ decreases to $t$. From this and Eq. (3.55)$_{95}$, it follows immediately that (i) is equivalent to (ii). That (ii) is equivalent to (iii) follows immediately from Eq. (3.54)$_{94}$. □

Second, we establish that the sets $\partial_+ T_{t,r}^{\{X_u=x_u\}}$ and $\partial_- T_{t,r}^{\{X_u=x_u\}}$ are (topologically) closed; the following result is similar to (Krak et al., 2017, Proposition 4.6), and our proof follows theirs quite closely.

**Lemma 3.95.** *Consider a jump process $P$, a current time point $t$ in $\mathbb{R}_{\geq 0}$ and a state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t$. Then $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ is closed, and the same holds for $\partial_- T_{t,t}^{\{X_u=x_u\}}$ whenever $t > 0$.*

*Proof.* We only prove the statement for the set of accumulation points from the right; the proof for the set of accumulation points from the left is analoguous. Because $(\mathfrak{Q}, \|\bullet\|_{\mathrm{op}})$ is a normed space, the subset $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ of $\mathfrak{Q}$ is closed if and only if for any sequence $(Q_n)_{n\in\mathbb{N}}$ in $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ that converges with respect to $\|\bullet\|_{\mathrm{op}}$, the limit $\lim_{n\to+\infty} Q_n$ belongs to $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ (see Schechter, 1997, Section 15.34).

Thus, we fix any such convergent sequence $(Q_n)_{n\in\mathbb{N}}$ in $\partial_+ T_{t,t}^{\{X_u=x_u\}}$. To prove that the limit $Q := \lim_{n\to+\infty} Q_n$ belongs to $\partial_+ T_{t,t}^{\{X_u=x_u\}}$, we fix some time point $r$ in $]t,+\infty[$ and some positive real number $\epsilon$ in $\mathbb{R}_{>0}$. Because $(Q_n)_{n\in\mathbb{N}}$ converges to $Q$, there is a natural number $n$ such that

$$\|Q_n - Q\|_{\mathrm{op}} < \frac{\epsilon}{2}. \tag{3.81}$$

Furthermore, because $Q_n$ belongs to $\partial_+ T_{t,t}^{\{X_u=x_u\}}$, it follows from Lemma 3.94 that

$$\left(\exists r' \in\, ]t,r[\right) \left\|Q_n - Q_{t,r'}^{\{X_u=x_u\}}\right\|_{\mathrm{op}} < \frac{\epsilon}{2}. \tag{3.82}$$

It now follows from Eqs. (3.81) and (3.82) that there is some $r'$ in $]t,r[$ such that

$$\left\|Q - Q_{t,r'}^{\{X_u=x_u\}}\right\|_{\mathrm{op}} \leq \|Q_n - Q\|_{\mathrm{op}} + \left\|Q_n - Q_{t,r'}^{\{X_u=x_u\}}\right\|_{\mathrm{op}} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Since $r$ in $]t,+\infty[$ and $\epsilon$ in $\mathbb{R}_{>0}$ were arbitrary, it follows from Lemma 3.94 that $Q = \lim_{n\to+\infty} Q_n$ belongs to $\partial_+ T_{t,t}^{\{X_u=x_u\}}$, as required. □

With these two properties of the sets of accumulation points, we can start investigating the properties of the directional $d_{\mathfrak{Q}}$-derivatives. First, we establish that these are non-empty and closed.

**Corollary 3.96.** *Consider a jump process $P$, a current time point $t$ in $\mathbb{R}_{\geq 0}$ and a state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t$. If $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable from the right, then $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ is non-empty and closed. Similarly, if $t > 0$ and $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable from the left, then $\partial_- T_{t,t}^{\{X_u=x_u\}}$ is non-empty and closed.*

*Proof.* That $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ is non-empty follows immediately from Definition 3.46[97]. That $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ is closed follows immediately from Lemma 3.95. As before, the proof for the left-sided $d_{\mathfrak{Q}}$-derivative is entirely analoguous. □

Next, we establish that the directional $d_{\mathfrak{Q}}$-derivatives of $T_{t,r}^{\{X_u=x_u\}}$ coincide with the 'standard' directional derivatives whenever the latter exist.

**Lemma 3.48.** *Consider a jump process $P$, a current time point $t$ in $\mathbb{R}_{\geq 0}$, a state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t$ and a rate operator $Q$ in $\mathfrak{Q}$. Then*

$$\lim_{r\searrow t} \frac{T_{t,r}^{\{X_u=x_u\}} - I}{r-t} = Q$$

*if and only if $\partial_+ T_{t,t}^{\{X_u=x_u\}} = \{Q\}$ and $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable from the right. The same holds for the left-sided ($d_{\mathfrak{Q}}$-)derivative whenever $t > 0$.*

*Proof.* We only prove the statement for the right-sided derivative, the proof for the left-sided derivative is analogous. Crucial to our proof will be the observation that for any $r$ in $]t,+\infty[$,

$$\left\|Q_{t,r}^{\{X_u=x_u\}} - Q\right\|_{\mathrm{op}} = d_{\mathfrak{Q}}\left(Q_{t,r}^{\{X_u=x_u\}}, \{Q\}\right). \tag{3.83}$$

First, we prove the converse implication. To this end, we assume that $\partial_+ T_{t,t}^{\{X_u=x_u\}} = \{Q\}$ and that $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable from the right. Then

$$\lim_{r\searrow t} d_{\mathfrak{Q}}\left(Q_{t,r}^{\{X_u=x_u\}}, \{Q\}\right) = 0.$$

It follows immediately from this and Eq. (3.83) that $\lim_{r\searrow t} Q_{t,r}^{\{X_u=x_u\}} = Q$, which is what we needed to prove.

The proof of the direct implication is a bit more involved. Assume that the limit $\lim_{r\searrow t} Q_{t,r}^{\{X_u=x_u\}}$ exists and is equal to $Q$. Due to Eq. (3.83), this implies that

$$\lim_{r\searrow t} d_{\mathfrak{Q}}\left(Q_{t,r}^{\{X_u=x_u\}}, \{Q\}\right) = 0. \tag{3.84}$$

Thus, it remains for us to show that $\partial_+ T_{t,t}^{\{X_u=x_u\}} = \{Q\}$. To this end, we observe that Eq. (3.84) implies that

$$(\forall \epsilon \in \mathbb{R}_{>0})(\exists \delta \in \mathbb{R}_{>0})(\forall r \in ]t,t+\delta[)\ \left\|Q_{t,r}^{\{X_u=x_u\}} - Q\right\|_{\mathrm{op}} < \frac{\epsilon}{2}. \tag{3.85}$$

Due to Lemma $3.94_{125}$ (iii), we infer from this that $Q$ belongs to $\partial_+ T_{t,t}^{\{X_u=x_u\}}$.

To prove that $\partial_+ T_{t,t}^{\{X_u=x_u\}} = \{Q\}$, we assume *ex absurdo* that there is a $Q'$ in $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ such that $Q' \neq Q$. Then by Lemma $3.94_{125}$ (iii),

$$(\forall r \in ]t,+\infty[)(\forall \epsilon \in \mathbb{R}_{>0})(\exists r' \in ]t,r[)\ \left\|Q' - Q_{t,r'}^{\{X_u=x_u\}}\right\|_{\mathrm{op}} < \frac{\epsilon}{2}. \tag{3.86}$$

Observe that for any $r$ in $]t,+\infty[$,

$$\begin{aligned}\|Q' - Q\|_{\mathrm{op}} &= \left\|Q' - Q_{t,r}^{\{X_u=x_u\}} + Q_{t,r}^{\{X_u=x_u\}} - Q\right\|_{\mathrm{op}} \\ &\leq \left\|Q' - Q_{t,r}^{\{X_u=x_u\}}\right\|_{\mathrm{op}} + \left\|Q - Q_{t,r}^{\{X_u=x_u\}}\right\|_{\mathrm{op}}.\end{aligned}$$

Fix any $\epsilon$ in $\mathbb{R}_{>0}$. From the previous inequality and Eqs. (3.85) and (3.86), it follows that there are $\delta$ in $\mathbb{R}_{>0}$ and $r'$ in $]t,t+\delta[$ such that

$$\|Q' - Q\|_{\mathrm{op}} \leq \left\|Q' - Q_{t,r'}^{\{X_u=x_u\}}\right\|_{\mathrm{op}} + \left\|Q - Q_{t,r'}^{\{X_u=x_u\}}\right\|_{\mathrm{op}} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Because this inequality holds for any positive real number $\epsilon$, we infer that $\|Q' - Q\|_{\mathrm{op}} = 0$, but this is a contradiction because $Q' \neq Q$ and therefore $\|Q' - Q\|_{\mathrm{op}} > 0$ due to $(\mathrm{N3})_{76}$. For this reason, $\partial_+ T_{t,t}^{\{X_u=x_u\}} = \{Q\}$, as required. □

### 3.A.2 An alternative motivation for $d_{\mathfrak{Q}}$-differentiability

When is it justified to call the non-empty set $\mathcal{Q}$ of rate operators the (or a) set-valued directional derivative of $T_{t,t,}^{\{X_u=x_u\}}$? Let us focus on the right-sided derivative. Intuitively, we want that the distance between $Q_{t,r}^{\{X_u=x_u\}}$ and $\mathcal{Q}$ vanishes as $r$ approaches $t$:

$$\lim_{r\searrow t} d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_u=x_u\}}, \mathcal{Q}\big) = 0. \tag{3.87}$$

However, this requirement does not characterise the set-valued derivative uniquely. For example, $\mathfrak{Q}$ always satisfies this requirement because $d_{\mathfrak{Q}}(Q_{t,r}^{\{X_u=x_u\}}, \mathfrak{Q}) = 0$ for all $r$ in $]t,+\infty[$.

To solve this problem, it makes sense to require that $\mathcal{Q}$ should contain no 'unnecessary information'; that is, as $r$ approaches $t$, $Q_{t,r}^{\{X_u=x_u\}}$ should come arbitrarily close to every $Q$ in $\mathcal{Q}$ over and over again. More precisely put, the set $\mathcal{Q}$ contains no unnecessary information if for any $Q$ in $\mathcal{Q}$, the distance between $Q$ and $\{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in]t,r[}$ vanishes as $r$ approaches $t$; that is, we require that

$$(\forall Q \in \mathcal{Q})\ \lim_{r\searrow t} d_{\mathfrak{Q}}\big(Q, \big\{Q_{t,r'}^{\{X_u=x_u\}}\big\}_{r'\in]t,r[}\big) = 0. \tag{3.88}$$

#### The Hausdorff function

To formalise this intuitive definition, we use the Hausdorff function, sometimes also called the Pompeiu-Hausdorff function, a well-known semi-metric on the powerset of any metric space – see (Burago et al., 2011, Definition 1.1.4 and Section 7.3.1), (Conci et al., 2017, Section 3) or Proposition 3.97 further on. More precisely, we need the Hausdorff function on the normed space $(\mathfrak{Q}, \|\bullet\|_{\mathrm{op}})$, and this is the non-negative extended real-valued function $h_{\mathfrak{Q}}$ on $\mathcal{P}(\mathfrak{Q})_{\supset\emptyset} \times \mathcal{P}(\mathfrak{Q})_{\supset\emptyset}$ that is defined for any couple $(\mathcal{Q}_1, \mathcal{Q}_2)$ of non-empty subsets of $\mathfrak{Q}$ by

$$h_{\mathfrak{Q}}(\mathcal{Q}_1, \mathcal{Q}_2) := \max\left\{ \sup_{Q_1\in\mathcal{Q}_1} d_{\mathfrak{Q}}(Q_1, \mathcal{Q}_2), \sup_{Q_2\in\mathcal{Q}_2} d_{\mathfrak{Q}}(Q_2, \mathcal{Q}_1) \right\}.$$

Burago et al. (2011, Proposition 7.3.3) establish some important properties for the generic Hausdorff function, which we repeat only for $h_{\mathfrak{Q}}$ here.

**Proposition 3.97.** *The extended real valued function $h_{\mathfrak{Q}}$ is a semi-metric on $\mathcal{P}(\mathfrak{Q})_{\supset\emptyset}$, in the sense that*

HF1. $h_{\mathfrak{Q}}(\mathcal{Q}_1, \mathcal{Q}_2) = h_{\mathfrak{Q}}(\mathcal{Q}_2, \mathcal{Q}_1)$ *for all* $\mathcal{Q}_1, \mathcal{Q}_2 \in \mathcal{P}(\mathfrak{Q})_{\supset\emptyset}$;

HF2. $h_{\mathfrak{Q}}(\mathcal{Q}_1, \mathcal{Q}_2) \leq h_{\mathfrak{Q}}(\mathcal{Q}_1, \mathcal{Q}_3) + h_{\mathfrak{Q}}(\mathcal{Q}_3, \mathcal{Q}_2)$ *for all* $\mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3 \in \mathcal{P}(\mathfrak{Q})_{\supset\emptyset}$.

*Furthermore, for all $\mathcal{Q}_1$ and $\mathcal{Q}_2$ in $\mathcal{P}(\mathfrak{Q})_{\supset\emptyset}$,*

HF3. $h_{\mathfrak{Q}}(\mathcal{Q}_1, \mathrm{cl}(\mathcal{Q}_1)) = 0$, *where* $\mathrm{cl}(\mathcal{Q}_1)$ *denotes the topological closure of* $\mathcal{Q}_1$;

HF4. $\mathcal{Q}_1 = \mathcal{Q}_2$ *whenever $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are closed and* $h_{\mathfrak{Q}}(\mathcal{Q}_1, \mathcal{Q}_2) = 0$.

**Formalising our intuitive definition**

Let us return to our intuitive definition of the set-valued right-sided derivative. It is relatively straightforward to verify that Eq. (3.87)↶ holds if and only if

$$\lim_{r \searrow t} \sup\Big\{d_{\mathfrak{Q}}(Q, \mathcal{Q}) \colon Q \in \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\Big\} = 0.$$

We now recognize this supremum as one of the two suprema in the definition of $h_{\mathfrak{Q}}(\mathcal{Q}, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[})$. Furthermore, after some tedious but straightforward manipulations, we see that Eq. (3.88)↶ holds if and only if

$$\lim_{r \searrow t} \sup\Big\{d_{\mathfrak{Q}}(Q, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}) \colon Q \in \mathcal{Q}\Big\} = 0.$$

Here, we recognize the other of the two suprema in the definition of $h_{\mathfrak{Q}}(\mathcal{Q}, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[})$. In conclusion, our intuitive definition comes down to requiring that $\mathcal{Q}$ should satisfy

$$\lim_{r \searrow t} h_{\mathfrak{Q}}(\mathcal{Q}, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}) = 0. \tag{3.89}$$

It follows immediately from (HF1)↶–(HF3)↶ that if a non-empty subset $\mathcal{Q}$ of $\mathfrak{Q}$ satisfies the preceding limit statement, then so does its topological closure $\operatorname{cl}(\mathcal{Q})$. Thus, it makes sense to only consider closed subsets $\mathcal{Q}$ of $\mathfrak{Q}$ as candidates for the set-valued directional derivatives. The following result establishes that in doing so, we actually get uniqueness.

**Lemma 3.98.** *Consider a jump process $P$, a time point $t$ in $\mathbb{R}_{\geq 0}$ and a state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t$. If there is a non-empty and closed subset $\mathcal{Q}_+$ of $\mathfrak{Q}$ such that*

$$\lim_{r \searrow t} h_{\mathfrak{Q}}\Big(\mathcal{Q}_+, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\Big) = 0,$$

*then this set is unique. A similar statement holds for the limit from the left whenever $t > 0$.*

*Proof.* We only prove the statement for the limit from the right; the proof of the statement for the limit from the left is entirely analoguous. Assume *ex absurdo* that there are two non-empty and closed subsets $\mathcal{Q}_1$ and $\mathcal{Q}_2$ of $\mathfrak{Q}$ with $\mathcal{Q}_1 \neq \mathcal{Q}_2$ such that

$$\lim_{r \searrow t} h_{\mathfrak{Q}}\Big(\mathcal{Q}_2, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\Big) = 0.$$

Note that for any $r$ in $]t, +\infty[$,

$$h_{\mathfrak{Q}}(\mathcal{Q}_1, \mathcal{Q}_2) \leq h_{\mathfrak{Q}}\Big(\mathcal{Q}_1, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\Big) + h_{\mathfrak{Q}}\Big(\mathcal{Q}_2, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\Big)$$

due to (HF1)↶ and (HF2)↶. Taking the limit for $r$ going to $t$ on both sides of this inequality, we find that $h_{\mathfrak{Q}}(\mathcal{Q}_1, \mathcal{Q}_2) = 0$. Because $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are both closed, it follows from this and (HF4)↶ that $\mathcal{Q}_1 = \mathcal{Q}_2$. This clearly contradicts our assumption that $\mathcal{Q}_1 \neq \mathcal{Q}_2$, so the set $\mathcal{Q}_+$, if it exists, is unique. □

### Back to $d_{\mathfrak{Q}}$-differentiability

The intuitive argument that motivates Eq. (3.89)↶ is very similar to our intuitive motivation for the notion of $d_{\mathfrak{Q}}$-differentiability. We have already mentioned in Section 3.3.1₈₉ that these two formalisations are equivalent, and this statement is justified by the following result.

**Proposition 3.99.** *Consider a jump process $P$, a time point $t$ in $\mathbb{R}_{\geq 0}$ and a state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u < t$. Then $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable from the right if and only if there is a non-empty and closed subset $\mathcal{Q}_+$ of $\mathfrak{Q}$ such that*

$$\lim_{r \searrow t} h_{\mathfrak{Q}}\left(\mathcal{Q}_+, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\right) = 0; \tag{3.90}$$

*whenever this is the case, $\mathcal{Q}_+ = \partial_+ T_{t,t}^{\{X_u=x_u\}}$. Furthermore, if $t > 0$ then $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable from the left if and only if there is a non-empty and closed subset $\mathcal{Q}_-$ of $\mathfrak{Q}$ such that*

$$\lim_{s \nearrow t} h_{\mathfrak{Q}}\left(\mathcal{Q}_-, \{Q_{s',t}^{\{X_u=x_u\}}\}_{s' \in ]s,t[}\right) = 0;$$

*whenever this is the case, $\mathcal{Q}_- = \partial_- T_{t,t}^{\{X_u=x_u\}}$.*

*Proof.* We only prove the statement for the limit from the right; the proof for the limit from the left is analoguous.

First, we prove the direct implication; that is, we assume that $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable, and prove that $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ is the unique non-empty and closed subset of $\mathfrak{Q}$ such that

$$\lim_{r \searrow t} h_{\mathfrak{Q}}\left(\partial_+ T_{t,t}^{\{X_u=x_u\}}, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\right) = 0. \tag{3.91}$$

Recall from Corollary 3.96₁₂₆ that $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ is a non-empty and closed subset of $\mathfrak{Q}$. Thus, it remains for us to verify Eq. (3.91) and establish uniqueness. To this end, we observe that by definition of the Hausdorff function $h_{\mathfrak{Q}}$, Eq. (3.91) holds if and only if

$$\lim_{r \searrow t} \sup\left\{d_{\mathfrak{Q}}\left(Q, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\right) : Q \in \partial_+ T_{t,t}^{\{X_u=x_u\}}\right\} = 0 \tag{3.92}$$

and

$$\lim_{r \searrow t} \sup\left\{d_{\mathfrak{Q}}\left(Q_{t,r'}^{\{X_u=x_u\}}, \partial_+ T_{t,t}^{\{X_u=x_u\}}\right) : r' \in ]t,r[\right\} = 0. \tag{3.93}$$

Observe that Eq. (3.92) follows almost immediately from Lemma 3.94₁₂₅ (ii). To verify Eq. (3.93), we recall from Definition 3.46₉₇ that because $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable from the right,

$$\lim_{r \searrow t} d_{\mathfrak{Q}}\left(Q_{t,r}^{\{X_u=x_y\}}, \partial_+ T_{t,t}^{\{X_u=x_u\}}\right) = 0.$$

By definition, this means that

$$\left(\forall \epsilon \in \mathbb{R}_{>0}\right)\left(\exists \delta \in \mathbb{R}_{>0}\right)\left(\forall r \in ]t, t+\delta[\right) \; d_{\mathfrak{Q}}\left(Q_{t,r}^{\{X_u=x_y\}}, \partial_+ T_{t,t}^{\{X_u=x_u\}}\right) < \epsilon.$$

Clearly, this implies that

$$\big(\forall \epsilon \in \mathbb{R}_{>0}\big)\big(\exists \delta \in \mathbb{R}_{>0}\big)\big(\forall r \in \,]t, t+\delta[\big) \sup\{d_{\mathfrak{Q}}\big(Q_{t,r'}^{\{X_u=x_y\}}, \partial_+ T_{t,t}^{\{X_u=x_u\}}\big) \colon r' \in \,]t, r[\} < \epsilon,$$

which in turn implies Eq. (3.93)↶, as required.

To summarise, we have shown that $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ is a non-empty and closed subset of $\mathfrak{Q}$ that satisfies Eq. (3.91)↶, or equivalently, that $\mathcal{Q}_+ = \partial_+ T_{t,t}^{\{X_u=x_u\}}$ is a non-empty and closed subset of $\mathfrak{Q}$ that satisfies Eq. (3.90)↶. Lemma 3.98$_{129}$ furthermore tells us that this set is unique, which finalises our proof of the direct implication.

Second, we prove the converse implication; that is, we assume that there is a non-empty and closed subset $\mathcal{Q}_+$ of $\mathfrak{Q}$ that satisfies Eq. (3.90)↶, and set out to prove that $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable from the right. To this end, we observe that due to Eq. (3.90)↶ and the definition of $h_{\mathfrak{Q}}$,

$$\lim_{r \searrow t} \sup\Big\{d_{\mathfrak{Q}}\big(Q, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\big) \colon Q \in \mathcal{Q}_+\Big\} = 0.$$

From this, it follows immediately that

$$(\forall Q \in \mathcal{Q}_+) \lim_{r \searrow t} d_{\mathfrak{Q}}\big(Q, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r' \in ]t,r[}\big) = 0,$$

and therefore $\mathcal{Q}_+ \subseteq \partial_+ T_{t,t}^{\{X_u=x_u\}}$. Hence, for all $r$ in $]t, +\infty[$,

$$\begin{aligned} d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_u=x_u\}}, \partial_+ T_{t,t}^{\{X_u=x_u\}}\big) &= \inf\Big\{\big\|Q_{t,r}^{\{X_u=x_u\}} - Q\big\|_{\mathrm{op}} \colon Q \in \partial_+ T_{t,t}^{\{X_u=x_u\}}\Big\} \\ &\leq \inf\Big\{\big\|Q_{t,r}^{\{X_u=x_u\}} - Q\big\|_{\mathrm{op}} \colon Q \in \mathcal{Q}_+\Big\} \\ &= d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_u=x_u\}}, \mathcal{Q}_+\big). \end{aligned} \tag{3.94}$$

Observe now that due to Eq. (3.90)↶ and the definition of $h_{\mathfrak{Q}}$,

$$\lim_{r \searrow t} \sup\Big\{d_{\mathfrak{Q}}\big(Q_{t,r'}^{\{X_u=x_u\}}, \mathcal{Q}_+\big) \colon r' \in \,]t, r[\Big\} = 0$$

and therefore

$$\lim_{r \searrow t} d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_u=x_u\}}, \mathcal{Q}_+\big) = 0.$$

Because $d_{\mathfrak{Q}}$ is non-negative, we infer from this and Eq. (3.94) that

$$\lim_{r \searrow t} d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_u=x_u\}}, \partial_+ T_{t,t}^{\{X_u=x_u\}}\big) = 0.$$

Thus, $T_{t,r}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable. □

### 3.A.3 Proof of Proposition 3.57

One reason why we have gone through the trouble of establishing Proposition 3.99↶, is that it comes in handy in the proof of Proposition 3.57$_{104}$.

**Proposition 3.57.** *Consider a jump process $P$ that has bounded rate. Fix a time point $t$ in $\mathbb{R}_{\geq 0}$ and state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec t$. Then $T_{t,t}^{\{X_u=x_u\}}$ is $d_{\mathfrak{Q}}$-differentiable, $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ is bounded and*

$$\partial_+ T_{t,t}^{\{X_u=x_u\}} = \Big\{Q \in \mathfrak{Q} \colon \big(\exists (r_n)_{n\in\mathbb{N}} \searrow t\big) \lim_{n\to+\infty} Q_{t,r_n}^{\{X_u=x_u\}} = Q\Big\};$$[23]

*whenever $t > 0$, $\partial_- T_{t,t}^{\{X_u=x_u\}}$ is bounded, and*

$$\partial_- T_{t,t}^{\{X_u=x_u\}} = \Big\{Q \in \mathfrak{Q} \colon \big(\exists (s_n)_{n\in\mathbb{N}} \nearrow t\big) \lim_{n\to+\infty} Q_{s_n,t}^{\{X_u=x_u\}} = Q\Big\}.$$[24]

*Proof.* We will only prove the statement regarding the right-sided $d_{\mathfrak{Q}}$-derivative, the proof for the statement pertaining to the left-sided $d_{\mathfrak{Q}}$-derivative is analoguous. To this end, we let

$$\mathcal{Q} := \Big\{Q \in \mathfrak{Q} \colon \big(\exists (r_n)_{n\in\mathbb{N}} \searrow t\big) \lim_{n\to+\infty} Q_{t,r_n}^{\{X_u=x_u\}} = Q\Big\}.$$

Krak et al. (2017, Proposition 4.6) prove that because $P$ has bounded rate, $\mathcal{Q}$ is non-empty, bounded and closed. From this and Proposition 3.99$_{130}$, it follows that to prove the statement, it suffices to verify that

$$\lim_{r\searrow t} h_{\mathfrak{Q}}\big(\mathcal{Q}, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in]t,r[}\big) = 0,$$

which holds if and only if

$$\lim_{r\searrow t} \sup\Big\{d_{\mathfrak{Q}}\big(Q_{t,r'}^{\{X_u=x_u\}}, \mathcal{Q}\big) \colon r' \in ]t,r[\Big\} = 0 \tag{3.95}$$

and

$$\lim_{r\searrow t} \sup\Big\{d_{\mathfrak{Q}}\big(Q, \{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in]t,r[}\big) \colon Q \in \mathcal{Q}\Big\} = 0. \tag{3.96}$$

First, we verify that Eq. (3.95) holds. To this end, we fix any positive real number $\epsilon$. Krak et al. (2017, Proposition 4.7) prove that there is a real number $\delta$ in $\mathbb{R}_{>0}$ such that

$$\big(\forall r' \in ]t, t+\delta[\big)\big(\exists Q \in \mathcal{Q}\big) \left\|Q_{t,r'}^{\{X_u=x_u\}} - Q\right\|_{\mathrm{op}} < \epsilon.$$

It follows from this and the definition of $d_{\mathfrak{Q}}$ that

$$\big(\forall r' \in ]t, t+\delta[\big)\; d_{\mathfrak{Q}}\big(Q_{t,r'}^{\{X_u=x_u\}}, \mathcal{Q}\big) < \epsilon,$$

and therefore

$$\big(\forall r \in ]t, t+\delta[\big)\; \sup\Big\{d_{\mathfrak{Q}}\big(Q_{t,r'}^{\{X_u=x_u\}}, \mathcal{Q}\big) \colon r' \in ]t,r[\Big\} < \epsilon.$$

Because $\epsilon$ in $\mathbb{R}_{>0}$ was arbitrary, we infer from this that Eq. (3.95) holds.

[23] In this expression, we let $(r_n)_{n\in\mathbb{N}} \searrow t$ denote a sequence in $\mathbb{R}_{\geq 0}$ such that $t < r_{n+1} < r_n$ for all $n$ in $\mathbb{N}$ and $\lim_{n\to+\infty} r_n = t$.

[24] In this expression, we let $(s_n)_{n\in\mathbb{N}} \nearrow t$ denote a sequence in $\mathbb{R}_{\geq 0}$ such that $u \prec s_n < s_{n+1} < t$ for all $n$ in $\mathbb{N}$ and $\lim_{n\to+\infty} s_n = t$

Next, we verify that Eq. (3.96)$_{\curvearrowleft}$ holds. Observe that the supremum in this limit statement is non-decreasing as $r$ approaches $t$, or equivalently, that it is non-increasing for increasing $r$; thus,

$$\lim_{r\searrow t}\sup\left\{d_{\mathfrak{Q}}\left(Q,\{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in\,]t,r[}\right)\colon Q\in\mathcal{Q}\right\} = \sup\left\{\sup\left\{d_{\mathfrak{Q}}\left(Q,\{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in\,]t,r[}\right)\colon Q\in\mathcal{Q}\right\}\colon r\in\,]t,+\infty[\right\}.$$

We may always change the order of suprema, so

$$\lim_{r\searrow t}\sup\left\{d_{\mathfrak{Q}}\left(Q,\{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in\,]t,r[}\right)\colon Q\in\mathcal{Q}\right\} = \sup\left\{\sup\left\{d_{\mathfrak{Q}}\left(Q,\{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in\,]t,r[}\right)\colon r\in\,]t,+\infty[\right\}\colon Q\in\mathcal{Q}\right\}. \tag{3.97}$$

Fix any positive real number $\epsilon$ in $\mathbb{R}_{>0}$ and any $Q$ in $\mathcal{Q}$. Note that, by construction of $\mathcal{Q}$, there is a decreasing sequence $(r_n)_{n\in\mathbb{N}}$ in $]t,+\infty[$ that converges to $t$ such that $\lim_{n\to+\infty}Q_{t,r_n}^{\{X_u=x_u\}}=Q$. Thus,

$$(\exists N\in\mathbb{N})(\forall n\in\mathbb{N},n\geq N)\;\left\|Q_{t,r_n}^{\{X_u=x_u\}}-Q\right\|_{\mathrm{op}}<\epsilon.$$

Because $(r_n)_{n\in\mathbb{N}}$ is decreasing and converges to $t$, we infer from this that

$$\left(\forall r\in\,]t,+\infty[\right)\;d_{\mathfrak{Q}}\left(Q,\{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in\,]t,r[}\right)<\epsilon,$$

and therefore

$$\sup\left\{d_{\mathfrak{Q}}\left(Q,\{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in\,]t,r[}\right)\colon r\in\,]t,+\infty[\right\}<\epsilon.$$

Note that this holds for all $Q$ in $\mathcal{Q}$, so

$$\sup\left\{\sup\left\{d_{\mathfrak{Q}}\left(Q,\{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in\,]t,r[}\right)\colon r\in\,]t,+\infty[\right\}\colon Q\in\mathcal{Q}\right\}<\epsilon.$$

It follows immediately from this and Eq. (3.97) that

$$\lim_{r\searrow t}\sup\left\{d_{\mathfrak{Q}}\left(Q,\{Q_{t,r'}^{\{X_u=x_u\}}\}_{r'\in\,]t,r[}\right)\colon Q\in\mathcal{Q}\right\}<\epsilon.$$

Because this inequality holds for any positive real number $\epsilon$, we can conclude that Eq. (3.96)$_{\curvearrowleft}$ holds. □

## 3.B Proof of Theorem 3.75

This appendix is devoted to the proof of Theorem 3.75$_{114}$, which we repeat for convenience.

**Theorem 3.75.** *Let $(\underline{T}_t)_{t\in\mathbb{R}_{\geq0}}$ be a continuous semi-group of lower transition operators. Then*

$$\underline{Q}\coloneqq\lim_{t\searrow0}\frac{\underline{T}_t-I}{t}$$

*is a lower rate operator, and $\underline{Q}$ is the unique lower rate operator such that $\underline{T}_t=e^{t\underline{Q}}$ for all $t$ in $\mathbb{R}_{\geq0}$.*

Our proof is inspired by that of (Krak et al., 2017, Theorem 5.4), which establishes a similar result for semi-groups of transition operators. In our proof, we will need that the 'derivative' of a continuous semi-group of transformations is always bounded, as is established by the following intermediary result.

**Lemma 3.100.** *If $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ is a continuous semi-group of lower transition operators, then*

$$\limsup_{t\searrow 0}\left\|\frac{\underline{T}_t - I}{t}\right\|_{\mathrm{op}} < +\infty.$$

In our proof of this lemma, we will need the following observation.

**Lemma 3.101.** *For any real number $a$,*

$$e^a = \lim_{n\to+\infty}\left(1+\frac{a}{n+1}\right)^n.$$

*Proof.* Recall that, by definition of the exponential function,

$$e^a = \lim_{n\to+\infty}\left(1+\frac{a}{n}\right)^n.$$

To prove the equality in the statement, we observe that for any natural number $n$ such that $n+1 \neq -a$,

$$\left(1+\frac{a}{n+1}\right)^n = \frac{\left(1+\frac{a}{n+1}\right)^{n+1}}{\left(1+\frac{a}{n+1}\right)}.$$

Note that in the right-hand side, the numerator converges to $e^a$ in the limit for $n$ going to $+\infty$ and the denominator converges to 1. Therefore, taking the limit for $n$ going to $+\infty$ on both sides of the equality proves the statement. □

*Proof of Lemma 3.100.* Our proof is one by contradiction, so we assume *ex absurdo* that

$$\limsup_{t\searrow 0}\left\|\frac{\underline{T}_t - I}{t}\right\|_{\mathrm{op}} = +\infty. \tag{3.98}$$

Because $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ is a continuous semi-group,

$$\big(\forall \epsilon\in\mathbb{R}_{>0}\big)\big(\exists\delta\in\mathbb{R}_{>0}\big)\big(\forall\Delta\in\,]0,\delta[\big)\ \|\underline{T}_\Delta - I\|_{\mathrm{op}} < \epsilon. \tag{3.99}$$

Observe that for all $\Delta$ in $\mathbb{R}_{>0}$ and $x$ in $\mathscr{X}$,

$$\big|[\underline{T}_\Delta \mathbb{I}_x](x) - 1\big| \leq \|\underline{T}_\Delta\mathbb{I}_x - I\mathbb{I}_x\| \leq \|\underline{T}_\Delta - I\|_{\mathrm{op}},$$

where for the second equality we have used that $\|\mathbb{I}_x\| = 1$. From the preceding inequalities and Eq. (3.99), we infer that

$$\big(\forall \epsilon\in\mathbb{R}_{>0}\big)\big(\exists\delta\in\mathbb{R}_{>0}\big)\big(\forall\Delta\in\,]0,\delta[\big)\big(\forall x\in\mathscr{X}\big)\ \big|[\underline{T}_\Delta\mathbb{I}_x](x) - 1\big| < \epsilon. \tag{3.100}$$

In particular, if we fix an arbitrary positive real number $\epsilon$ such that $\epsilon < 1/2$, then due to Eq. (3.100), there is some $\Delta$ in $\mathbb{R}_{>0}$ such that

$$\big|[\underline{T}_\Delta\mathbb{I}_x](x) - 1\big| < \epsilon \quad \text{for all } x\in\mathscr{X},$$

and therefore

$$1 - \epsilon < [\underline{T}_\Delta \mathbb{I}_x](x) \quad \text{for all } x \in \mathscr{X}. \tag{3.101}$$

In the remainder of this proof, we use the assumption in Eq. (3.98) to contradict this inequality.

To that end, we fix two positive real numbers $\epsilon_1$ and $\epsilon_2$ such that $\epsilon_1 + 2\epsilon_2 < 1 - 2\epsilon$ (this is always possible because we have required that $\epsilon < 1/2$). Additionally, we also fix a positive real number $\lambda$ such that $e^{-\Delta\lambda} < 1 - 2\epsilon - \epsilon_1 - 2\epsilon_2$ (this is possible because we have chosen $\epsilon_1$ and $\epsilon_1$ such that $\epsilon_1 + 2\epsilon_2 < 1 - 2\epsilon$).

By Lemma 3.101, there is a natural number $N_{\epsilon_1}$ such that

$$\left(1 - \frac{\Delta\lambda}{n+1}\right)^n < e^{-\Delta\lambda} + \epsilon_1 \quad \text{for all } n \geq N_{\epsilon_1}. \tag{3.102}$$

Additionally, by Eq. (3.99), there is a positive real number $\delta_{\epsilon_2}$ such that

$$\|\underline{T}_\delta - I\|_{\text{op}} < \epsilon_2 \quad \text{for all } \delta \in \left]0, \delta_{\epsilon_2}\right[. \tag{3.103}$$

Similarly, due to Eq. (3.100) there is a positive real number $\overline{\delta}$ – where, without loss of generality, we assume that $\overline{\delta} \leq \Delta$ – such that

$$\left|[\underline{T}_\delta \mathbb{I}_x](x) - 1\right| < \frac{1}{2} \quad \text{for all } \delta \in \left]0, \overline{\delta}\right[. \tag{3.104}$$

The time has come to use our ex-absurdo assumption; it follows from Eq. (3.98) that there is a decreasing sequence $(\delta_k)_{k\in\mathbb{N}}$ in $]0, \overline{\delta}[$ that converges to 0 such that

$$\lambda \leq \left\| \frac{\underline{T}_{\delta_k} - I}{\delta_k} \right\| = 2\max\left\{ \frac{1 - [\underline{T}_{\delta_k} \mathbb{I}_x](x)}{\delta_k} : x \in \mathscr{X} \right\} \quad \text{for all } k \in \mathbb{N}, \tag{3.105}$$

where the equality follows from Lemma 3.73$_{113}$ and (LR7)$_{111}$. For any $k$ in $\mathbb{N}$, we let $n_k := \lfloor \Delta/\delta_k \rfloor$. Because $(\delta_k)_{k\in\mathbb{N}}$ is sequence of positive real numbers that is lower than $\Delta$ (since $\overline{\delta} \leq \Delta$) and converges to 0, $(n_k)_{k\in\mathbb{N}}$ is a non-decreasing sequence of natural numbers that diverges to $+\infty$. Furthermore, $n_k\delta_k \leq \Delta < (n_k+1)\delta_k$ by construction, whence

$$0 \leq \Delta - n_k\delta_k < \delta_k \quad \text{for all } k \in \mathbb{N}.$$

Because $(\delta_k)_{k\in\mathbb{N}}$ converges to 0, we infer from the preceding inequalities that $(\Delta - n_k\delta_k)_{k\in\mathbb{N}}$ is a sequence of non-negative real numbers that converges to 0 as well. Due to the aforementioned properties of $(n_k)_{k\in\mathbb{N}}$ and $(\Delta - n_k\delta_k)_{k\in\mathbb{N}}$, we can pick some natural number $k^\star$ such that $n_{k^\star} \geq N_{\epsilon_1}$ and $\Delta - n_{k^\star}\delta_{k^\star} < \delta_{\epsilon_2}$; to ease our notation, we set $\delta^\star := \delta_{k^\star}$ and $n^\star := n_{k^\star}$. Thus, we have ensured that

$$(1 - \delta^\star\lambda)^{n^\star} < \left(1 - \frac{\Delta\lambda}{n^\star + 1}\right)^{n^\star} < e^{-\Delta\lambda} + \epsilon_1, \tag{3.106}$$

where the first inequality holds because $\delta^\star\lambda < 1$ (which follows from Eqs. (3.104) and (3.105)) and $\Delta < (n^\star+1)\delta^\star$, and the second inequality follows from Eq. (3.102). We have also chosen $n^\star$ and $\delta^\star$ such that

$$\|\underline{T}_\Delta - \underline{T}_{n^\star\delta^\star}\|_{\text{op}} = \|\underline{T}_{n^\star\delta^\star}\underline{T}_{\Delta - n^\star\delta^\star} - \underline{T}_{n^\star\delta^\star} I\|_{\text{op}} \leq \|\underline{T}_{\Delta - n^\star\delta^\star} - I\|_{\text{op}} < \epsilon_2, \tag{3.107}$$

where we have used (LT9)$_{108}$ for the first inequality and the second inequality follows from (SG1)$_{77}$ if $\Delta - n^\star\delta^\star = 0$ and from Eq. (3.103) otherwise.

We are finally ready to put together all the pieces of the puzzle. Let $x^\star$ be one of the states in $\mathscr{X}$ that reaches the maximum in Eq. (3.105) for $k = k^\star$. Because $\delta^\star < \overline{\delta}$ by construction, it follows from Eqs. (3.104) and (3.105) that

$$0 \leq \frac{1}{2}\lambda\delta^\star \leq M := 1 - [\underline{T}_{\delta^\star}\mathbb{I}_{x^\star}](x^\star) < \frac{1}{2}. \tag{3.108}$$

Observe that for every other state $y$ in $\mathscr{X} \setminus \{x^\star\}$,

$$[\underline{T}_{\delta^\star}\mathbb{I}_{x^\star}](y) \leq [\overline{T}_{\delta^\star}\mathbb{I}_{x^\star}](y) \leq [\overline{T}_{\delta^\star}(1 - \mathbb{I}_y)](y) = 1 - [\underline{T}_{\delta^\star}\mathbb{I}_y](y) \leq M,$$

where the inequalities follow from (LT4)$_{108}$ and (LT6)$_{108}$, and the equality follows from (LT5)$_{108}$. From the two preceding inequalities, we infer that

$$\underline{T}_{\delta^\star}\mathbb{I}_{x^\star} \leq M + (1-2M)\mathbb{I}_{x^\star}.$$

It follows from this, (LT6)$_{108}$, (LT5)$_{108}$ and (LT2)$_{107}$ – which we may use because $M < 1/2$ by construction – that

$$\begin{aligned}\underline{T}_{2\delta^\star}\mathbb{I}_{x^\star} &= \underline{T}_{\delta^\star}(\underline{T}_{\delta^\star}\mathbb{I}_{x^\star}) \leq \underline{T}_{\delta^\star}\big(M + (1-2M)\mathbb{I}_{x^\star}\big) = M + (1-2M)\underline{T}_{\delta^\star}\mathbb{I}_{x^\star} \\ &\leq M(1 + (1-2M)) + (1-2M)^2\mathbb{I}_{x^\star}.\end{aligned}$$

We apply this same trick $n^\star - 1$ times more, to yield

$$\begin{aligned}\underline{T}_{n^\star\delta^\star}\mathbb{I}_{x^\star} &\leq M\left(\sum_{\ell=0}^{n^\star-1}(1-2M)^\ell\right) + (1-2M)^{n^\star}\mathbb{I}_{x^\star} \\ &= M\frac{1-(1-2M)^{n^\star}}{1-(1-2M)} + (1-2M)^{n^\star}\mathbb{I}_{x^\star} \\ &= \frac{1}{2}\Big(1-(1-2M)^{n^\star}\Big) + (1-2M)^{n^\star}\mathbb{I}_{x^\star}.\end{aligned}$$

More specifically, we see that

$$[\underline{T}_{n^\star\delta^\star}\mathbb{I}_{x^\star}](x^\star) \leq \frac{1}{2}\Big(1 + (1-2M)^{n^\star}\Big) \leq \frac{1}{2}\Big(1 + \big(1-\delta^\star\lambda\big)^{n^\star}\Big) < \frac{1}{2}\Big(1 + e^{-\Delta\lambda} + \epsilon_1\Big),$$

where the second inequality holds because $0 \leq \delta^\star\lambda \leq 2M < 1$ due to Eq. (3.108) and the final inequality follows from Eq. (3.106). Next, we observe that

$$\Big|[\underline{T}_\Delta\mathbb{I}_{x^\star}](x^\star) - [\underline{T}_{n^\star\delta^\star}\mathbb{I}_{x^\star}](x^\star)\Big| \leq \|\underline{T}_\Delta\mathbb{I}_{x^\star} - \underline{T}_{n^\star\delta^\star}\mathbb{I}_{x^\star}\| \leq \|\underline{T}_\Delta - \underline{T}_{n^\star\delta^\star}\| < \epsilon_2,$$

where the penultimate inequality holds because $\|\mathbb{I}_x\| = 1$ and the final inequality holds due to Eq. (3.107). From the previous two inequalities, we now infer that

$$[\underline{T}_\Delta\mathbb{I}_{x^\star}](x^\star) < [\underline{T}_{n^\star\delta^\star}\mathbb{I}_{x^\star}](x^\star) + \epsilon_2 < \frac{1}{2}\Big(1 + e^{-\Delta\lambda} + \epsilon_1\Big) + \epsilon_2 < 1 - \epsilon,$$

where the final inequality holds because we have chosen $\lambda$ such that $e^{-\Delta\lambda} < 1 - 2\epsilon - \epsilon_1 - 2\epsilon_2$. This clearly contradicts Eq. (3.101), which proves the statement. □

To prove Theorem $3.75_{114}$, we now combine the preceding result with Proposition $A.6_{448}$ in Appendix $A.3_{448}$.

*Proof of Theorem $3.75_{114}$.* Fix any decreasing sequence $(\Delta_n)_{n\in\mathbb{N}}$ of positive real numbers that converges to 0. For any natural number $n$, we set $\underline{Q}_n := (\underline{T}_{\Delta_n} - I)/\Delta_n$. Recall from Lemma $3.73_{113}$ that for any natural number $n$, $\underline{Q}_n$ is a lower rate operator. Additionally, because $(\Delta_n)_{n\in\mathbb{N}}$ decreases to 0, it follows from Lemma $3.100_{134}$ that

$$B := \sup\{\|\underline{Q}_n\|_{\text{op}} : n \in \mathbb{N}\} = \sup\left\{\left\|\frac{1}{\Delta_n}\big(\underline{T}_{\Delta_n - I}\big)\right\|_{\text{op}} : n \in \mathbb{N}\right\} < +\infty.$$

In other words, $(\underline{Q}_n)_{n\in\mathbb{N}}$ is a sequence in $\underline{\mathfrak{Q}}_B := \{\underline{Q} \in \underline{\mathfrak{Q}} : \|\underline{Q}\|_{\text{op}} \leq B\}$.

In Proposition $A.6_{448}$, we establish that $\underline{\mathfrak{Q}}_B$ is sequentially compact. Therefore, there is an increasing sequence $(n_k)_{k\in\mathbb{N}}$ of natural numbers and a lower rate operator $\underline{Q}$ in $\underline{\mathfrak{Q}}_B$ such that the subsequence $(\underline{Q}_{n_k})_{k\in\mathbb{N}}$ converges to $\underline{Q}$, in the sense that

$$\lim_{k\to+\infty}\left\|\frac{1}{\Delta_{n_k}}\big(\underline{T}_{\Delta_{n_k}} - I\big) - \underline{Q}\right\|_{\text{op}} = \lim_{k\to+\infty}\|\underline{Q}_{n_k} - \underline{Q}\|_{\text{op}} = 0. \tag{3.109}$$

We now show that, for any $t$ in $\mathbb{R}_{\geq 0}$, $\underline{T}_t = e^{t\underline{Q}}$. To that end, we fix a time point $t$ in $\mathbb{R}_{\geq 0}$ and any positive real number $\epsilon$, and any four positive real numbers $\epsilon_1, \ldots, \epsilon_4$ such that $t(\epsilon_1 + \epsilon_2) + \epsilon_3 + \epsilon_4 \leq \epsilon$. By Eq. (3.109), there is a natural number $K$ such that for any natural number $k \geq K$,

$$\|\underline{T}_{\Delta_{n_k}} - (I + \Delta_{n_k}\underline{Q})\|_{\text{op}} = \Delta_{n_k}\left\|\frac{1}{\Delta_{n_k}}\big(\underline{T}_{\Delta_{n_k}} - I\big) - \underline{Q}\right\|_{\text{op}} < \Delta_{n_k}\epsilon_1. \tag{3.110}$$

Additionally, it follows from the operator differential equation in Proposition $3.77_{115}$[25] with $t = 0$ that there is a positive real number $\delta_2$ such that

$$\big(\forall \Delta \in ]0, \delta_2[\big)\ \|e^{\Delta\underline{Q}} - (I + \Delta\underline{Q})\|_{\text{op}} = \Delta\left\|\frac{1}{\Delta}\big(e^{\Delta\underline{Q}} - I\big) - \underline{Q}\right\|_{\text{op}} < \Delta\epsilon_2. \tag{3.111}$$

Because $(\underline{T}_t)_{t\in\mathbb{R}_{\geq 0}}$ is a continuous semi-group by the condition of the statement, from $(\text{SG4})_{78}$ with $t = 0$ it follows that there is a positive real number $\delta_3$ such that

$$\big(\forall \Delta \in ]0, \delta_3[\big)\ \|\underline{T}_\Delta - I\|_{\text{op}} < \epsilon_3. \tag{3.112}$$

Similarly, because $(e^{t\underline{Q}})_{t\in\mathbb{R}_{\geq 0}}$ is a continuous semi-group due to Proposition $3.74_{114}$, there is a positive real number $\delta_4$ such that

$$\big(\forall \Delta \in ]0, \delta_4[\big)\ \|e^{\Delta\underline{Q}} - I\|_{\text{op}} < \epsilon_4. \tag{3.113}$$

For any natural number $k$, we let $\ell_k := \lfloor t/\Delta_{n_k}\rfloor$. Observe that because $(\Delta_{n_k})_{k\in\mathbb{N}}$ is a decreasing sequence of positive real numbers that converges to 0, $(\ell_k)_{k\in\mathbb{N}}$ is a non-decreasing sequence of non-negative integers that diverges to $+\infty$. Even more, because $\ell_k\Delta_{n_k} \leq t < (\ell_k + 1)\Delta_{n_k}$ by construction, $0 \leq t - \ell_k\Delta_{n_k} < \Delta_{n_k}$ and therefore

[25] Note that the result that we are currently proving actually precedes Proposition $3.77_{115}$ in the main text, but this is not an issue because Proposition $3.74_{114}$ simply restates (De Bock, 2017b, Proposition 9).

$\lim_{k\to+\infty}(t - \ell_k \Delta_{n_k}) = 0$ because $(\Delta_{n_k})_{k\in\mathbb{N}}$ converges to 0. Thus, there is a smallest natural number $k^\star$ such that

$$k^\star \geq K, \quad \Delta_{n_{k^\star}} < \delta_2 \quad \text{and} \quad (t - \ell_{k^\star}\Delta_{n_{k^\star}}) < \min\{\delta_3, \delta_4\}.$$

To ease our notation, we set $\Delta \coloneqq \Delta_{n_{k^\star}}$, $\ell \coloneqq \ell_{k^\star}$ and $\delta \coloneqq t - \ell_{k^\star}\Delta_{n_{k^\star}}$. Note that $t = \ell\Delta + \delta$, so $\underline{T}_t = \left(\underline{T}_\Delta\right)^\ell \underline{T}_\delta$ and $e^{t\underline{Q}} = \left(e^{\Delta\underline{Q}}\right)^\ell e^{\delta\underline{Q}}$ by repeated application of the semi-group property. Consequently,

$$\|\underline{T}_t - e^{t\underline{Q}}\|_{\text{op}} = \|\left(\underline{T}_\Delta\right)^\ell \underline{T}_\delta - \left(e^{\Delta\underline{Q}}\right)^\ell e^{\delta\underline{Q}}\|_{\text{op}} \leq \ell\|\underline{T}_\Delta - e^{\Delta\underline{Q}}\|_{\text{op}} + \|\underline{T}_\delta - e^{\delta\underline{Q}}\|_{\text{op}},$$

where we have used Proposition 3.74 114 and Lemma E.4 in (Krak et al., 2017) – see also Lemma 4.14 176 further on – for the inequality. We execute some straightforward manipulations, to yield

$$\|\underline{T}_t - e^{t\underline{Q}}\|_{\text{op}} \leq \ell\|\underline{T}_\Delta - (I + \Delta\underline{Q})\|_{\text{op}} + \ell\|e^{\Delta\underline{Q}} - (I + \Delta\underline{Q})\|_{\text{op}} + \|\underline{T}_\delta - I\|_{\text{op}} + \|e^{\delta\underline{Q}} - I\|_{\text{op}}.$$

Because, by construction, $\Delta = \Delta_{n_{k^\star}}$, $k^\star \geq K$, $\Delta < \delta_2$ and $\delta < \min\{\delta_3, \delta_4\}$, it follows from Eqs. (3.110)↶ to (3.113)↶ (or from Eqs. (3.110)↶ and (3.111)↶, Proposition 3.74 114 and (SG1) 77 if $\delta = 0$) that

$$\|\underline{T}_t - e^{t\underline{Q}}\|_{\text{op}} < \ell\Delta\epsilon_1 + \ell\Delta\epsilon_2 + \epsilon_3 + \epsilon_4 \leq t(\epsilon_1 + \epsilon_2) + \epsilon_3 + \epsilon_4 \leq \epsilon,$$

where the second inequality holds because $\ell\Delta \leq t$ by construction and the final inequality is precisely our condition on $\epsilon_1, \ldots, \epsilon_4$. Because $\epsilon$ was an arbitrary positive real number, we conclude that $\underline{T}_t = e^{t\underline{Q}}$ for all $t$ in $\mathbb{R}_{\geq 0}$.

The first part now follows from the previous. More precisely, because $\underline{T}_t = e^{t\underline{Q}}$ for all $t$ in $\mathbb{R}_{\geq 0}$, it follows from the operator differential equation in Proposition 3.77 115[26] with $t = 0$ that

$$\lim_{t\searrow 0} \frac{\underline{T}_t - I}{\Delta} = \lim_{t\searrow 0} \frac{e^{t\underline{Q}} - I}{\Delta} = \underline{Q}.$$

Finally, the preceding equality also shows that the lower rate operator $\underline{Q}$ is unique. More precisely, if $\underline{R}$ is another lower rate operator such that $\underline{T}_t = e^{t\underline{R}}$ for all $t$ in $\mathbb{R}_{\geq 0}$, then it follows from the previous equality that $\underline{Q} = \underline{R}$. □

## 3.C Theorems 6.3 and 6.5 in (Krak et al., 2017)

The proof that Krak et al. (2017) give for their Theorem 6.5 relies on their Theorem 6.3, and it is the latter result that is incorrect. The reason why Theorem 6.3 in (Krak et al., 2017) is wrong is a bit subtle, but it essentially comes down to a missing condition in the (original) statement; as we will now see, this error can be solved by additionally assuming that $\mathcal{Q}$ has separately specified rows.

[26] Again, it is not a problem that in the main text, Proposition 3.77 115 is established after the result that we are proving.

**Theorem 3.102.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$ and a non-empty, bounded and convex subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$ that has separately specified rows. Fix a jump process $P_0$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, a sequence of time points $u$ in $\mathscr{U}$, and, for all $x_u$ in $\mathscr{X}_u$, a jump process $P_{x_u}$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Then there is a jump process $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ such that for all $x_u$ in $\mathscr{X}_u$ and $u_1, u_2$ in $\mathscr{U}$ with $u_2 \neq ()$, $u_1 \cup u_2 \subseteq u$ and $u_1 < u_2$,*

$$P(X_{u_2} = x_{u_2} \,|\, X_{u_1} = x_{u_1}) = P_0(X_{u_2} = x_{u_2} \,|\, X_{u_1} = x_{u_1})$$

*and for all $A$ in $\mathscr{F}_u$,*

$$P(A \,|\, X_u = x_u) = P_{x_u}(A \,|\, X_u = x_u).$$

Besides invoking Theorem 3.102 instead of Theorem 6.3 in (Krak et al., 2017), our proof for Theorem $3.88_{120}$ is exactly the same as the one that Krak et al. (2017) give for their Theorem 6.5; because this is a trivial change, we will not repeat this argument here. We will, however, prove Theorem 3.102. In fact, we actually prove an even more general result first – we need this result in our proof for Proposition $7.69_{395}$ further on – and then use this result to prove Theorem 3.102. Needless to say, this more general result is inspired by Theorem 6.3 in (Krak et al., 2017), and our proof follows theirs quite closely.

**Theorem 3.103.** *Consider a non-empty set $\mathscr{M}$ of initial probability mass functions on $\mathscr{X}$ and a non-empty, bounded, convex subset $\mathscr{Q}$ of rate operators that has separately specified rows. Fix a jump process $P_0$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, a sequence of time points $u$ in $\mathscr{U}$, and, for all $x_u$ in $\mathscr{X}_u$, a jump process $P_{x_u}$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ and some $y_u^{x_u}$ in $\mathscr{X}_u$ with $y_{\max u}^{x_u} = x_{\max u}$. Then there is a jump process $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ such that for all $x_u$ in $\mathscr{X}_u$ and $u_1, u_2$ in $\mathscr{U}$ with $u_2 \neq ()$, $u_1 \cup u_2 \subseteq u$ and $u_1 < u_2$,*

$$P(X_{u_2} = x_{u_2} \,|\, X_{u_1} = x_{u_1}) = P_0(X_{u_2} = x_{u_2} \,|\, X_{u_1} = x_{u_1})$$

*and for all $v$ in $\mathscr{U}_{>u}$ and $B \subseteq \mathscr{X}_v$,*

$$P(X_v \in B \,|\, X_u = x_u) = P_{x_u}(X_v \in B \,|\, X_u = y_u^{x_u}).$$

*Proof.* The statement is trivial whenever $u = ()$: the first part of the statement does not come into play because there are no $u_1, u_2$ in $\mathscr{U}$ that satisfy the conditions, and the second part holds trivially because $\mathscr{X}_u = \mathscr{X}_{()} = \{x_{()}\}$ – let $P := P_{x_{()}}$. Henceforth, we assume without loss of generality that $u \neq ()$. Our proof consists of the following three parts: (i) we construct a coherent conditional probability $P$ on $\mathscr{D}$ that satisfies the equalities in the statement; (ii) we show that $P$ is consistent with $\mathscr{M}$; and (iii) we show that $P$ is consistent with $\mathscr{Q}$.

To construct the coherent conditional probability $P$ on $\mathscr{D}$, we will construct a coherent conditional probability on a subset of $\mathscr{D}$ and then extend it to $\mathscr{D}$ by virtue of Theorem $2.54_{45}$. To this end, we let

$$\mathscr{D}_0 := \{(X_w \in B \,|\, X_v = z_v) \in \mathscr{D} \colon v < \max u, \max w \leq \max u\} \tag{3.114}$$

and

$$\mathscr{D}_1 := \{(X_w \in B \,|\, X_v = z_v) \in \mathscr{D} \colon u \subseteq v < w\}. \tag{3.115}$$

Furthermore, for any sequence of time points $v = (t_1, \dots, t_n)$ in $\mathscr{U}$ with $u \subseteq v$, we let $v \ominus u$ denote the sequence of time points that consists of those time points $t$ in $v$ such that $t > \max u$ – that is, $v \ominus u := ()$ if $\max v = \max u$, and $v \ominus u := (t_k, \dots, t_n)$ otherwise, with $k$ the smallest index in $\{1, \dots, n\}$ such that $t_k > \max u$. With this notation, we let $\check{P}$ be the real-valued function on $\check{\mathscr{D}} := \mathscr{D}_0 \cup \mathscr{D}_1$ that is defined for all $(A \,|\, X_v = z_v)$ in $\check{\mathscr{D}}$ by

$$\check{P}(A \,|\, X_v = z_v) := \begin{cases} P_0(A \,|\, X_v = z_v) & \text{if } v \prec \max u, \\ P_{z_u}(A \,|\, X_u = y_u^{z_u}, X_{v \ominus u} = z_{v \ominus u}) & \text{otherwise.} \end{cases} \tag{3.116}$$

Next, we show that $\check{P}$ is a coherent conditional probability. By Definition 2.51$_{44}$, we need to show that for all $n$ in $\mathbb{N}$, $\mu_1, \dots, \mu_n$ in $\mathbb{R}$ and $(A_1 \,|\, C_1), \dots, (A_n \,|\, C_n)$ in $\check{\mathscr{D}}$,

$$\max\left\{ \sum_{k=1}^{n} \mu_k \mathbb{I}_{C_k}(\omega)\big(\mathbb{I}_{A_k}(\omega) - \check{P}(A_k \,|\, C_k)\big) \colon \omega \in \bigcup_{k=1}^{n} C_k \right\} \geq 0. \tag{3.117}$$

Thus, we fix arbitrary $n$ in $\mathbb{N}$, $\mu_1, \dots, \mu_n$ in $\mathbb{R}$ and $(A_1 \,|\, C_1), \dots, (A_n \,|\, C_n)$ in $\check{\mathscr{D}}$. By definition of $\check{\mathscr{D}}$, for all $k$ in $\{1, \dots, n\}$, there are $v_k, w_k$ in $\mathscr{U}$, $z_{v_k}^k$ in $\mathscr{X}_{v_k}$ and $B_k \subseteq \mathscr{X}_{w_k}$ such that

$$A_k = \{X_{w_k} \in B_k\} \quad \text{and} \quad C_k = \{X_{v_k} = z_{v_k}^k\}.$$

We collect those indices $k$ for which $(A_k \,|\, C_k)$ belongs to $\mathscr{D}_0$ in the index set

$$\mathscr{K}_0 := \{k \in \{1, \dots, n\} \colon (A_k \,|\, C_k) \in \mathscr{D}_0\}.$$

First, we consider the case that $\mathscr{K}_0$ is non-empty. Then because $P_0$ is a stochastic process by assumption, it is a coherent conditional probability on $\mathscr{D} \supseteq \mathscr{D}_0$. Hence, by Definition 2.51$_{44}$,

$$\max\left\{ \sum_{k \in \mathscr{K}_0} \mu_k \mathbb{I}_{C_k}(\omega)\big(\mathbb{I}_{A_k}(\omega) - P_0(A_k \,|\, C_k)\big) \colon \omega \in \bigcup_{k \in \mathscr{K}_0} C_k \right\} \geq 0.$$

Because $\check{P}$ is equal to $P_0$ on $\mathscr{D}_0$ by Eq. (3.116), it follows from the preceding inequality that there is a path $\omega_0$ in $\bigcup_{k \in \mathscr{K}_0} C_k \subseteq \bigcup_{k=1}^{n} C_k$ such that

$$\sum_{k \in \mathscr{K}_0} \mu_k \mathbb{I}_{C_k}(\omega_0)\big(\mathbb{I}_{A_k}(\omega_0) - \check{P}(A_k \,|\, C_k)\big) \geq 0. \tag{3.118}$$

If $\mathscr{K}_0$ is empty, then we let $\omega_0$ be an arbitrary path in $\bigcup_{k=1}^{n} C_k$; obviously, this path $\omega_0$ satisfies Eq. (3.118) because the sum over an empty index set is equal to zero by convention.

Next, we let $z_u^0 := \omega_0|_u$. Note that for all $k$ in $\{1, \dots, n\} \setminus \mathscr{K}_0$, $(A_k \,|\, C_k)$ belongs to $\mathscr{D}_1$, and therefore $u \subseteq v_k$. Hence, for all $k$ in $\{1, \dots, n\} \setminus \mathscr{K}_0$, $v_k \oslash u := v_k \setminus (v_k \ominus u)$ contains those time points $t$ in $v_k$ such that $t \leq \max u$, and this includes all the time points in $u$. Consider the index set

$$\mathscr{K}_1 := \{k \in \{1, \dots, n\} \setminus \mathscr{K}_0 \colon z_{v_k \oslash u}^k = \omega_0|_{v_k \oslash u}\}.$$

Note that, by construction, $z_u^k = z_u^0$ for all $k$ in $\mathscr{K}_1$. Again, we start with the case that this index set $\mathscr{K}_1$ is non-empty. Then we let

$$y_u^0 := y_u^{z_u^0} \quad \text{and} \quad C_k^1 := \{X_u = y_u^0, X_{v_k \ominus u} = z_{v_k \ominus u}^k\} \quad \text{for all } k \in \mathscr{K}_1.$$

The jump process $P_{z_u^0}$ is a coherent conditional probability on $\mathscr{D}$ by Definition 3.12$_{65}$, so it follows from Definition 2.51$_{44}$ that

$$\max\left\{\sum_{k\in\mathscr{K}_1}\mu_k\mathbb{I}_{C_k^1}(\omega)\big(\mathbb{I}_{A_k}(\omega)-P_{z_u^0}\big(A_k\,\big|\,C_k^1\big)\big)\colon \omega\in\bigcup_{k\in\mathscr{K}_1}C_k^1\right\}\geq 0.$$

Take any path $\omega_1^\star$ in $\bigcup_{k\in\mathscr{K}_1}C_k^1$ for which the sum in the preceding expression is non-negative, and observe that $\omega_1^\star|_u=y_u^0$. We replace $\omega_1^\star$ on $[0,\max u]$ by $\omega_0$:

$$\omega_1\colon\mathbb{R}_{\geq 0}\to\mathscr{X}\colon t\mapsto\omega_1(t):=\begin{cases}\omega_0(t) & \text{if } t\leq\max u,\\ \omega_1^\star(t) & \text{otherwise.}\end{cases}$$

Recall that $z^0_{\max u}=y^0_{\max u}$ by the assumptions of the statement, and that, by construction, $\omega_1^\star(\max u)=y^0_{\max u}$ and $\omega_0(\max u)=z^0_{\max u}$; because furthermore $\omega_0$ and $\omega_1^\star$ are càdlàg, it is clear that the path $\omega_1$ is càdlàg too. Furthermore, for all $k$ in $\mathscr{K}_1$,

$$\omega_1^\star\in C_k^1=\left\{X_u=y_u^0,X_{v_k\ominus u}=z^k_{v_k\ominus u}\right\}\Leftrightarrow\omega_1\in C_k=\left\{X_{v_k}=z^k_{v_k}\right\},$$

and, because $u\subseteq v_k\prec w_k$ and hence $u\prec w_k$,

$$\omega_1^\star\in\left\{X_{w_k}\in B_k\right\}=A_k\Leftrightarrow\omega_1\in\left\{X_{w_k}\in B_k\right\}=A_k.$$

Thus, we have constructed a path $\omega_1$ in $\bigcup_{k\in\mathscr{K}_1}C_k\subseteq\bigcup_{k=1}^n C_k$ such that

$$\begin{aligned}\sum_{k\in\mathscr{K}_1}\mu_k\mathbb{I}_{C_k}(\omega_1)\big(\mathbb{I}_{A_k}(\omega_1)-\check{P}(A_k\,|\,C_k)\big)&=\sum_{k\in\mathscr{K}_1}\mu_k\mathbb{I}_{C_k}(\omega_1)\big(\mathbb{I}_{A_k}(\omega_1)-P_{z_u^k}\big(A_k\,\big|\,C_k^1\big)\big)\\&=\sum_{k\in\mathscr{K}_1}\mu_k\mathbb{I}_{C_k^1}(\omega_1^\star)\big(\mathbb{I}_{A_k}(\omega_1^\star)-P_{z_u^0}\big(A_k\,\big|\,C_k^1\big)\big)\\&\geq 0,\end{aligned}\tag{3.119}$$

where, for all $k$ in $\mathscr{K}_1$, for the first equality we used that $\check{P}(A_k\,|\,C_k)=P_{z_u^k}(A_k\,|\,C_k^1)$ by Eq. (3.116)$_\curvearrowleft$ and for the second equality we used that $z_u^k=z_u^0$. If $\mathscr{K}_1$ is empty, then we simply let $\omega_1:=\omega_0$; here too, $\omega_1$ trivially satisfies Eq. (3.119) and belongs to $\bigcup_{k=1}^n C_k$ because $\omega_0$ does so by assumption.

Observe that, for all $k$ in $\mathscr{K}_0$, $v_k\prec\max u$ and $\max w_k\leq\max u$ by definition of $\mathscr{D}_0$, and therefore

$$\omega_0\in\left\{X_{v_k}=z^k_{v_k}\right\}=C_k\Leftrightarrow\omega_1\in C_k\quad\text{and}\quad\omega_0\in\{X_{w_k}\in B_k\}=A_k\Leftrightarrow\omega_1\in A_k;$$

hence, it follows from Eq. (3.118)$_\curvearrowleft$ that

$$\sum_{k\in\mathscr{K}_0}\mu_k\mathbb{I}_{C_k}(\omega_1)\big(\mathbb{I}_{A_k}(\omega_1)-\check{P}(A_k\,|\,C_k)\big)\geq 0.\tag{3.120}$$

Furthermore, for all $k$ in $\{1,\dots,n\}\setminus(\mathscr{K}_0\cup\mathscr{K}_1)$, $z^k_{v_k\oslash u}\neq\omega_0|_{v_k\oslash u}=\omega_1|_{v_k\oslash u}$, so $\omega_1$ does not belong to $C_k=\{X_{v_k}=z^k_{v_k}\}$, or equivalently,

$$\mathbb{I}_{C_k}(\omega_1)=0.\tag{3.121}$$

Recall from before that the path $\omega_1$ belongs to $\bigcup_{k=1}^{n} C_k$. Therefore, it follows from Eqs. (3.119)↶ to (3.121)↶ that

$$\sum_{k=1}^{n} \mu_k \mathbb{I}_{C_k}(\omega_1)\big(\mathbb{I}_{A_k}(\omega_1) - \check{P}(A_k \,|\, C_k)\big) \geq 0;$$

this inequality implies Eq. $(3.117)_{140}$, as required.

Because $\check{P}$ is a coherent conditional probability on $\check{\mathscr{D}} \subseteq \mathscr{D}$, there is a coherent conditional probability $P$ on $\mathscr{D}$ that extends $\check{P}$ by Theorem $2.54_{45}$. Then $P$ coincides with $\check{P}$ on $\check{\mathscr{D}}$, so it follows immediately from Eq. $(3.116)_{140}$ that $P$ satisfies the two equalities in the statement. This settles the first part of the proof.

In the second part of this proof, we need to show that $P$ is consistent with $\mathscr{M}$. To this end, we observe that, because $u \neq ()$ by assumption, $(X_0 = x \,|\, \Omega)$ belongs to $\mathscr{D}_0$ for all $x$ in $\mathscr{X}$. Because $P$ coincides with $\check{P}$ on $\check{\mathscr{D}} \supseteq \mathscr{D}_0$ by construction and $\check{P}$ coincides with $P_0$ on $\mathscr{D}_0$ by Eq. $(3.116)_{140}$, we find that

$$P(X_0 = x) = \check{P}(X_0 = x) = P_0(X_0 = x) \quad \text{for all } x \in \mathscr{X},$$

so the initial probability mass function $P(X_0 = \bullet)$ of $P$ is equal to that of $P_0$. The jump process $P_0$ is consistent with $\mathscr{M}$ by assumption, so this implies that $P$ is consistent with $\mathscr{M}$, as required.

In the third and last part of this proof, we show that $P$ is consistent with $\mathscr{Q}$ – be warned, this will take a lot more work than proving consistency with $\mathscr{M}$. For all $\{X_v = z_v\}$ in $\mathscr{H}$ and $t, r$ in $\mathbb{R}_{\geq 0}$ such that $u < t \leq r$, we denote the history-dependent transition operator corresponding to $P$ by $T_{t,r}^{\{X_v = z_v\}}$, the one corresponding to $P_0$ by $T_{0,t,r}^{\{X_v = z_v\}}$ and, for all $y_u$ in $\mathscr{X}_u$, the one corresponding to $P_{y_u}$ by $T_{y_u,t,r}^{\{X_v = z_v\}}$. As we will see, these history-dependent transition operators are interconnected.

Fix some $\{X_v = z_v\}$ in $\mathscr{H}$ and $t, r$ in $\mathbb{R}_{\geq 0}$ such that $v < t < r$. First, we assume that $t < \max u$ and $r \leq \max u$. Then for all $z_t, y_r$ in $\mathscr{X}$, $(X_r = y_r \,|\, X_v = z_v, X_t = z_t)$ belongs to $\mathscr{D}_0$, so by Eq. $(3.116)_{140}$,

$$P(X_r = y_r \,|\, X_v = z_v, X_t = z_t) = P_0(X_r = y_r \,|\, X_v = z_v, X_t = z_t). \tag{3.122}$$

By Eq. $(3.36)_{84}$, this implies that

$$T_{t,r}^{\{X_v = z_v\}} = T_{0,t,r}^{\{X_v = z_v\}}. \tag{3.123}$$

Next, we assume that $t \geq \max u$; in this case, things are a bit more involved. Let $w \coloneqq u \setminus (v \cup (t))$, and let $P^\star$ be any coherent extension of $P$ to $\mathcal{P}(\Omega) \times \mathcal{P}(\Omega)_{\supset\emptyset}$. Then for all $z_t, y_r$ in $\mathscr{X}$,

$$\begin{aligned} P(X_r = y_r \,|\, X_v = z_v, X_t = z_t) &= P^\star(X_r = y_r \,|\, X_v = z_v, X_t = z_t) \\ &= P^\star\Big(\{X_r = y_r\} \cap \{X_w \in \mathscr{X}_w\} \,\Big|\, X_v = z_v, X_t = z_t\Big) \\ &= \sum_{z_w \in \mathscr{X}_w} P^\star(X_r = y_r, X_w = z_w \,|\, X_v = z_v, X_t = z_t), \end{aligned} \tag{3.124}$$

where we used $(\text{CP3})_{41}$ for the last equality. Let us investigate the terms in this sum separately. By definition, $w$ only contains those time points in $u$ that do not belong to $v$ and that precede $t$. Because furthermore $u \subseteq w \cup v \cup (t)$, this implies that for all $z_w$ in $\mathscr{X}_w$,

$$(X_r = y_r \,|\, X_w = z_w, X_v = z_v, X_t = z_t) \in \mathscr{D}_1 \subseteq \mathscr{D},$$

and therefore

$$\begin{aligned}
&P^\star(X_r = y_r, X_w = z_w \,|\, X_v = z_v, X_t = z_t) \\
&\quad = P^\star(X_r = y_r \,|\, X_w = z_w, X_v = z_v, X_t = z_t) P^\star(X_w = z_w \,|\, X_v = z_v, X_t = z_t), \\
&\quad = P(X_r = y_r \,|\, X_w = z_w, X_v = z_v, X_t = z_t) P^\star(X_w = z_w \,|\, X_v = z_v, X_t = z_t) \\
&\quad = \check{P}(X_r = y_r \,|\, X_w = z_w, X_v = z_v, X_t = z_t) P^\star(X_w = z_w \,|\, X_v = z_v, X_t = z_t),
\end{aligned} \tag{3.125}$$

where for the first equality we used (CP4)$_{41}$, for the second equality we used that $P^\star$ extends $P$ and for the third equality we used that $P$ extends $\check{P}$. Because $u \subseteq w \cup v \cup (t)$, we can rewrite Eq. (3.125) even more by substituting Eq. (3.116)$_{140}$, but this step warrants some extra care; we distinguish two subcases: $t > \max u$ and $t = \max u$. In any case, we let $u' := u \setminus (t)$. First, we consider the case that $t > \max u$. Then $u = u'$ and $(w \cup v \cup (t)) \ominus u = (v \ominus u) \cup (t)$, so it follows from Eq. (3.125) that

$$\begin{aligned}
&\check{P}(X_r = y_r \,|\, X_w = z_w, X_v = z_v, X_t = z_t) \\
&\quad = P_{z_u}(X_r = y_r \,|\, X_u = y_u^{z_u}, X_{v\ominus u} = z_{v\ominus u}, X_t = z_t) \\
&\quad = P_{z_u}\big(X_r = y_r \,\big|\, X_{u'} = y_{u'}^{z_u}, X_{v\ominus u} = z_{v\ominus u}, X_t = z_t\big),
\end{aligned} \tag{3.126}$$

where for the second equality we used that $u = u'$. Second, we consider the case that $t = \max u$. Then $(w \cup v \cup (t)) \ominus u = ()$, so it follows from Eq. (3.125) that

$$\check{P}(X_r = y_r \,|\, X_w = z_w, X_v = z_v, X_t = z_t) = P_{z_u}(X_r = y_r \,|\, X_u = y_u^{z_u});$$

the subscript $t$ for $z_t$ is essential here, because otherwise $z_u$ is not properly defined. Note that $u = u' \cup (t)$ and that $y_t^{z_u} = y_{\max u}^{z_u} = z_{\max u} = z_t$ by the assumptions in the statement, so

$$\begin{aligned}
&\check{P}(X_r = y_r \,|\, X_w = z_w, X_v = z_v, X_t = z_t) \\
&\quad = P_{z_u}(X_r = y_r \,|\, X_u = y_{u'}^{z_u}, X_t = z_t) \\
&\quad = P_{z_u}\big(X_r = y_r \,\big|\, X_{u'} = y_{u'}^{z_u}, X_{v\ominus u} = z_{v\ominus u}, X_t = z_t\big),
\end{aligned} \tag{3.127}$$

where for the last equality we also used that $v \ominus u = ()$. We substitute Eqs. (3.126) and (3.127) in Eq. (3.125), to yield

$$\begin{aligned}
&P^\star(X_r = y_r, X_w = z_w \,|\, X_v = z_v, X_t = z_t) \\
&\quad = P_{z_u}(X_r = y_r \,|\, X_{u'} = y_{u'}^{z_u}, X_{v\ominus u} = z_{v\ominus u}, X_t = z_t) P^\star(X_w = z_w \,|\, X_v = z_v, X_t = z_t).
\end{aligned}$$

Finally, we substitute the preceding equality into Eq. (3.124)$_{\curvearrowleft}$, to yield

$$P(X_r = y_r \,|\, X_v = z_v, X_t = z_t) = \sum_{z_w \in \mathscr{X}_w} P_{z_u}(X_r = y_r \,|\, C(z_w), X_t = z_t) p_t^{z_t}(z_w), \tag{3.128}$$

where we let

$$p_t^{z_t} \colon \mathscr{X}_w \to \mathbb{R} \colon z_w \mapsto p_t^{z_t}(z_w) := P^\star(X_w = z_w \,|\, X_v = z_v, X_t = z_t)$$

and, for all $z_w$ in $\mathscr{X}_w$, we let

$$C(z_w) := \left\{X_{u'} = y_{u'}^{z_u}, X_{v \ominus u} = z_{v \ominus u}\right\}$$

Note that, by the laws of (conditional) probability, $\sum_{z_w \in \mathscr{X}_w} p_t^{z_t}(z_w) = 1$ and $p_t^{z_t}(z_w) \geq 0$ for all $z_w$ in $\mathscr{X}_w$, so the sum on the right-hand side is a convex combination. Due to Eq. (3.36)$_{84}$, it follows immediately from Eq. (3.128)$_{\curvearrowleft}$ that for all $f$ in $\mathbb{G}(\mathscr{X})$ and $z_t$ in $\mathscr{X}$,

$$\left[T_{t,r}^{\{X_v = z_v\}} f\right](z_t) = \sum_{z_w \in \mathscr{X}_w} p_t^{z_t}(z_w)\left[T_{z_u,t,r}^{C(z_w)} f\right](z_t); \tag{3.129}$$

as mentioned before, the subscript $t$ for $z_t$ is essential in case $t = \max u$, because otherwise $z_u$ is not well-defined.[27]

A convenient intermediary result is to use Eqs. (3.122)$_{142}$ and (3.128)$_{\curvearrowleft}$ to prove that $P$ has bounded rate. By Lemma 3.54$_{102}$, we need to show that for all $t$ in $\mathbb{R}_{\geq 0}$, $\{X_v = z_v\}$ in $\mathscr{H}$ such that $u < t$ and $x$ in $\mathscr{X}$,

$$\limsup_{r \searrow t} \frac{1}{r-t}\big(1 - P(X_r = x \,|\, X_v = z_v, X_t = x)\big) < +\infty, \tag{3.130}$$

and, if $t > 0$,

$$\limsup_{s \nearrow t} \frac{1}{t-s}\big(1 - P(X_t = x \,|\, X_v = z_v, X_s = x)\big) < +\infty. \tag{3.131}$$

Thus, we fix any such $t$, $\{X_v = z_v\}$ and $x$. We start by proving Eq. (3.130). First, we assume that $t < \max u$; then it follows immediately from Eq. (3.122)$_{142}$ with $z_t = x = y_r$ that

$$\limsup_{r \searrow t} \frac{1}{r-t}\big(1 - P(X_r = x \,|\, X_v = z_v, X_t = x)\big) = \limsup_{r \searrow t} \frac{1}{r-t}\big(1 - P_0(X_r = x \,|\, X_v = z_v, X_t = x)\big),$$

Since $P_0$ is consistent with the bounded set $\mathscr{Q}$ of rate operators by assumption, $P_0$ has bounded rate by Lemma 3.55$_{102}$. Thus, it follows from Lemma 3.54$_{102}$ and the preceding equality that Eq. (3.130) holds in this case. Next, we consider the case that $t \geq \max u$. Fix some $r$ in $\mathbb{R}_{\geq 0}$ such that $r > t$, and let $z_t := x$. Then by Eq. (3.128)$_{\curvearrowleft}$ with $y_r = x$,

$$\begin{aligned}
\limsup_{r \searrow t} &\frac{1}{r-t}\big(1 - P(X_r = x \,|\, X_v = z_v, X_t = x)\big) \\
&= \limsup_{r \searrow t} \frac{1}{r-t}\left(1 - \sum_{z_w \in \mathscr{X}_w} p_t^x(z_w) P_{z_u}(X_r = x \,|\, C(z_w), X_t = x)\right) \\
&\leq \sum_{z_w \in \mathscr{X}_w} p_t^x(z_w) \limsup_{r \searrow t} \frac{1}{r-t}\big(1 - P_{z_u}(X_r = x \,|\, C(z_w), X_t = x)\big).
\end{aligned}$$

[27] It is at this point that we start to deviate from the proof of Krak et al. (2017): compare Eq. (3.129) to their Eqs. (D.18) and (D.19).

Recall that every jump process $P_{z_u}$ on the right-hand side of this inequality is consistent with the bounded set $\mathcal{Q}$ of rate operators by assumption, so this jump process has bounded rate by Lemma 3.55$_{102}$. Thus, it follows from Lemma 3.54$_{102}$ and the preceding inequality that Eq. (3.130)$_{\curvearrowleft}$ holds in this case. The proof for Eq. (3.131)$_{\curvearrowleft}$ is analoguous to the proof for Eq. (3.130)$_{\curvearrowleft}$, the only difference being that we need to distinguish the cases $t \leq \max u$ and $t > \max u$.

Because the jump process $P$ has bounded rate, we know from Proposition 3.57$_{104}$ that, for all $t$ in $\mathbb{R}_{\geq 0}$ and $\{X_v = z_v\}$ in $\mathscr{H}$ such that $v < t$, $T_{t,t}^{\{X_v=z_v\}}$ is $d_{\mathfrak{Q}}$-differentiable. To prove that $P$ is consistent with $\mathcal{Q}$, we still need to show that $\partial T_{t,t}^{\{X_v=z_v\}}$ belongs to $\mathcal{Q}$. To do so, we fix an arbitrary rate operator $Q$ in $\partial T_{t,t}^{\{X_v=z_v\}}$, and we set out to show that $Q$ belongs to $\mathcal{Q}$. First, we assume that $Q$ belongs to $\partial_- T_{t,t}^{\{X_v=z_v\}}$. Then by Proposition 3.57$_{104}$, there is a sequence $(s_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{\geq 0}$ such that $\max v < s_n < s_{n+1} < t$ for all $n$ in $\mathbb{N}$, $\lim_{n\to+\infty} s_n = t$ and

$$\lim_{n\to+\infty} \frac{T_{s_n,t}^{\{X_v=z_v\}} - I}{t - s_n} = Q. \tag{3.132}$$

Let us start with the case that $t \leq \max u$. Then for all $n$ in $\mathbb{N}$, it follows from Eq. (3.123)$_{142}$ that

$$\frac{T_{s_n,t}^{\{X_v=z_v\}} - I}{t - s_n} = \frac{T_{0,s_n,t}^{\{X_v=z_v\}} - I}{t - s_n},$$

and therefore

$$\lim_{n\to+\infty} \frac{T_{0,s_n,t}^{\{X_v=z_v\}} - I}{t - s_n} = Q.$$

Because $P_0$ is consistent with the bounded set $\mathcal{Q}$ by assumption, it follows from Propositions 3.55$_{102}$ and 3.57$_{104}$ that $Q$ belongs to $\partial_- T_{0,t,t}^{\{X_v=z_v\}} \subseteq \mathcal{Q}$, as required.

Next, we deal with the case that $t > \max u$; without loss of generality, we may assume that $s_1 > \max u$. Let $w := u \setminus v$, and observe that for all $z_w$ in $\mathscr{X}_w$, $z_u$ is well-defined because $u \subseteq w \cup v$. Thus, it follows from Eq. (3.129)$_{\curvearrowleft}$ that, for all $z$ in $\mathscr{X}$, $f$ in $\mathbb{G}(\mathscr{X})$ and $n$ in $\mathbb{N}$,

$$\frac{\left[T_{s_n,t}^{\{X_v=z_v\}} f\right](z) - f(z)}{t - s_n} = \sum_{z_w\in\mathscr{X}_w} p_{s_n}^{z}(z_w) \frac{\left[T_{z_u,s_n,t}^{C(z_w)} f\right](z) - f(z)}{t - s_n}; \tag{3.133}$$

note that we actually invoke Eq. (3.129)$_{\curvearrowleft}$ with $z_{s_n} = z$, but we may drop the 'index' $s_n$ of $z_{s_n}$ because $z_u$ is always well-defined as the time point $s_n$ does not occur in $u$. Fix some $z$ in $\mathscr{X}$. Then the corresponding sequence $(p_{s_n}^{z})_{n\in\mathbb{N}}$ is a sequence of probability mass functions on $\mathscr{X}_w$. The set $\mathscr{X}_w$ of probability mass functions on $\mathscr{X}_w$ is well-known to be sequentially compact – it is in essence a bounded subset of $\mathbb{R}^{|\mathscr{X}_w|}$ – so there is a probability mass function $p_\star^z$ on $\mathscr{X}_w$ and a subsequence $(s_{1,n})_{n\in\mathbb{N}}$ of $(s_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} p_{s_{1,n}}^{z} = p_\star^z$ (where the convergence is with respect to the supremum norm on $\mathscr{X}_w$). Next, we fix some $z_w$ in $\mathscr{X}_w$. Because $P_{z_u}$ has bounded rate due to Lemma 3.55$_{102}$, the corresponding sequence

$$\left(\frac{T_{z_u,s_{1,n},t}^{C(z_w)} - I}{t - s_{1,n}}\right)_{n\in\mathbb{N}}$$

is bounded, so it has a convergent subsequence due to Corollary A.7$_{449}$. Thus, there is a rate operator $Q^{\star}_{z,z_w}$ and a subsequence $(s_{2,n})_{n\in\mathbb{N}}$ of $(s_{1,n})_{n\in\mathbb{N}}$ such that

$$\lim_{n\to+\infty}\frac{T^{C(z_w)}_{z_u,s_{2,n},t}-I}{t-s_{2,n}}=Q^{\star}_{z,z_w};$$

because $(s_{2,n})_{n\in\mathbb{N}}\nearrow t$, it follows from Proposition 3.57$_{104}$ that this limit $Q^{\star}_{z,z_w}$ belongs to $\partial_- T^{C(z_w)}_{z_w,t,t}\subseteq\mathcal{Q}$. We repeat essentially the same procedure for all other $z_w$ in $\mathscr{X}_w$, albeit that first we construct a subsequence $(s_{3,n})_{n\in\mathbb{N}}$ of $(s_{2,n})_{n\in\mathbb{N}}$, then a subsequence $(s_{4,n})_{n\in\mathbb{N}}$ of $(s_{3,n})_{n\in\mathbb{N}}$, and so on. This way, we end up with a subsequence $(s_{K,n})_{n\in\mathbb{N}}$ of $(s_n)_{n\in\mathbb{N}}$, with $K:=|\mathscr{X}_w|+1$, such that $\lim_{n\to+\infty}p^z_{s_{K,n}}=p^z_{\star}$ and, for all $z_w$ in $\mathscr{X}_w$, with a rate operator $Q^{\star}_{z,z_w}$ in $\mathcal{Q}$ such that

$$\lim_{n\to+\infty}\frac{T^{C(z_w)}_{z_u,s_{K,n},t}-I}{t-s_{K,n}}=Q^{\star}_{z,z_w}.$$

Let $Q^{\star}_z:=\sum_{z_w\in\mathscr{X}_w}p^z_{\star}(z_w)Q^{\star}_{z,z_w}$. Clearly, $Q^{\star}_z$ is a rate operator in $\mathcal{Q}$ because it is defined as a convex combination of rate operators in $\mathcal{Q}$ and $\mathcal{Q}$ is convex; furthermore, it is not difficult to verify that

$$\lim_{n\to+\infty}\sum_{z_w\in\mathscr{X}_w}p^z_{s_{K,n}}(z_w)\frac{T^{C(z_w)}_{z_u,s_{K,n},t}-I}{t-s_{K,n}}=Q^{\star}_z.$$

Because $(s_{K,n})_{n\in\mathbb{N}}$ is a subsequence of $(s_n)_{n\in\mathbb{N}}$, it follows from the preceding equality and Eqs. (3.132)$_\curvearrowleft$ and (3.133)$_\curvearrowleft$ that, for all $f$ in $\mathbb{G}(\mathscr{X})$,

$$[Qf](z)=\lim_{n\to+\infty}\sum_{z_w\in\mathscr{X}_w}p^z_{s_{K,n}}(z_w)\frac{\left[T^{C(z_w)}_{z_u,s_{K,n},t}f\right](z)-f(z)}{t-s_{K,n}}=\left[Q^{\star}_z f\right](z).$$

We repeat this procedure for all other $z$ in $\mathscr{X}$; thus, for all $z$ in $\mathscr{X}$, we find some $Q^{\star}_z$ in $\mathcal{Q}$ such that

$$\big(\forall f\in\mathbb{G}(\mathscr{X})\big)\ [Qf](z)=\left[Q^{\star}_z f\right](z). \tag{3.134}$$

Let $Q^{\star}\colon\mathbb{G}(\mathscr{X})\to\mathbb{G}(\mathscr{X})$ be the operator defined for all $f$ in $\mathbb{G}(\mathscr{X})$ and $z$ in $\mathscr{X}$ by $[Q^{\star}f](z):=[Q^{\star}_z f](z)$. Because $(Q^{\star}_z)_{z\in\mathscr{X}}$ is a selection in $\mathcal{Q}$ and $\mathcal{Q}$ has separately specified rows, $Q^{\star}$ belongs to $\mathcal{Q}$. Observe that, by definition of $Q^{\star}$ and Eq. (3.134),

$$\begin{aligned}\left\|Q-Q^{\star}\right\|_{\mathrm{op}}&=\sup\{\left\|Qf-Q^{\star}f\right\|\colon f\in\mathbb{G}(\mathscr{X}),\|f\|=1\}\\&=\sup\Big\{\max\{\left|[Qf](z)-[Q^{\star}f](z)\right|\colon z\in\mathscr{X}\}\colon f\in\mathbb{G}(\mathscr{X}),\|f\|=1\Big\}\\&=\sup\Big\{\max\{\left|[Qf](z)-[Q^{\star}_z f](z)\right|\colon z\in\mathscr{X}\}\colon f\in\mathbb{G}(\mathscr{X}),\|f\|=1\Big\}\\&=0.\end{aligned}$$

Thus, we have shown that $Q=Q^{\star}$; because $Q^{\star}$ belongs to $\mathcal{Q}$ by construction, this proves that $Q$ belongs to $\mathcal{Q}$ whenever $Q$ belongs to $\partial_- T^{\{X_v=z_v\}}_{t,t}$.

Finally, we deal with the case that $Q$ belongs to $\partial_+ T^{\{X_v=z_v\}}_{t,t}$; the argument is similar to the one as before, but arguably not quite similar enough to justify omission.

By Proposition 3.57$_{104}$, there is a sequence $(r_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_{\geq 0}$ such that $t < r_{n+1} < r_n$ for all $n$ in $\mathbb{N}$, $\lim_{n\to+\infty} r_n = t$ and

$$\lim_{n\to+\infty} \frac{T_{t,r_n}^{\{X_v=z_v\}} - I}{r_n - t} = Q. \tag{3.135}$$

Let us start with the case that $t < \max u$; without loss of generality, we may assume that $r_1 < \max u$. Then for all $n$ in $\mathbb{N}$, it follows from Eq. (3.123)$_{142}$ that

$$\frac{T_{t,r_n}^{\{X_v=z_v\}} - I}{r_n - t} = \frac{T_{0,t,r_n}^{\{X_v=z_v\}} - I}{r_n - t},$$

and therefore

$$\lim_{n\to+\infty} \frac{T_{0,t,r_n}^{\{X_v=z_v\}} - I}{r_n - t} = Q.$$

Because $P_0$ is consistent with the bounded set $\mathcal{Q}$ by assumption, it follows from Propositions 3.55$_{102}$ and 3.57$_{104}$ that $Q$ belongs to $\partial_+ T_{0,t,t}^{\{X_v=z_v\}} \subseteq \mathcal{Q}$, as required.

Next, we deal with the case that $t \geq \max u$. Let $w := u \setminus (v \cup t)$ and $u' := u \setminus (t)$. Then it follows from Eq. (3.129)$_{144}$ that, for all $z_t$ in $\mathcal{X}$, $f$ in $\mathbb{G}(\mathcal{X})$ and $n$ in $\mathbb{N}$,

$$\frac{\left[T_{t,r_n}^{\{X_v=z_v\}} f\right](z_t) - f(z_t)}{r_n - t} = \sum_{z_w\in\mathcal{X}_w} p_t^{z_t}(z_w) \frac{\left[T_{z_u,t,r_n}^{C(z_w)} f\right](z_t) - f(z_t)}{r_n - t}. \tag{3.136}$$

Fix some $z_t$ in $\mathcal{X}$, and then some $z_w$ in $\mathcal{X}_w$. Because $P_{z_u}$ has bounded rate due to Lemma 3.55$_{102}$, the corresponding sequence

$$\left(\frac{T_{z_u,t,r_n}^{C(z_w)} - I}{r_n - t}\right)_{n\in\mathbb{N}}$$

is bounded, so it has a convergent subsequence due to Corollary A.7$_{449}$. Thus, there is a rate operator $Q^\star_{z_t,z_w}$ and a subsequence $(r_{1,n})_{n\in\mathbb{N}}$ of $(r_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to+\infty} \frac{T_{z_u,t,r_{1,n}}^{C(z_w)} - I}{r_{1,n} - t} = Q^\star_{z_t,z_w};$$

because $(r_{2,n})_{n\in\mathbb{N}} \searrow t$, it follows from Proposition 3.57$_{104}$ that this limit $Q^\star_{z_t,z_w}$ belongs to $\partial_+ T_{z_w,t,t}^{C(z_w)} \subseteq \mathcal{Q}$. We repeat the same procedure for all other $z_w$ in $\mathcal{X}_w$, albeit that first we construct a subsequence $(r_{2,n})_{n\in\mathbb{N}}$ of $(r_{1,n})_{n\in\mathbb{N}}$, then a subsequence $(r_{3,n})_{n\in\mathbb{N}}$ of $(r_{2,n})_{n\in\mathbb{N}}$, and so on. This way, we end up with a subsequence $(r_{K,n})_{n\in\mathbb{N}}$ of $(r_n)_{n\in\mathbb{N}}$, with $K := |\mathcal{X}_w|$, and, for all $z_w$ in $\mathcal{X}_w$, with a rate operator $Q^\star_{z_t,z_w}$ in $\mathcal{Q}$ such that

$$\lim_{n\to+\infty} \frac{T_{z_u,r_{K,n},t}^{C(z_w)} - I}{t - r_{K,n}} = Q^\star_{z_t,z_w}.$$

Let $Q^\star_{z_t} := \sum_{z_w\in\mathcal{X}_w} p_t^{z_t}(z_w) Q^\star_{z_w}$. Observe that $Q^\star_{z_t}$ is a rate operator in $\mathcal{Q}$ because it is defined as a convex combination of rate operators in $\mathcal{Q}$ and $\mathcal{Q}$ is convex, and that

$$\lim_{n\to+\infty} \sum_{z_w\in\mathcal{X}_w} p_t^{z_t}(z_w) \frac{T_{z_u,t,r_{K,n}}^{C(z_w)} - I}{r_{K,n} - t} = Q^\star_{z_t}.$$

Because $(r_{K,n})_{n\in\mathbb{N}}$ is a subsequence of $(r_n)_{n\in\mathbb{N}}$, it follows from the preceding equality and Eqs. (3.135) and (3.136) that, for all $f$ in $\mathbb{G}(\mathscr{X})$,

$$[Qf](z_t) = \lim_{n\to+\infty} \sum_{z_w\in\mathscr{X}_w} p_t^{z_t}(z_w) \frac{\big[T_{z_u,t,r_{K,n}}^{C(z_w)} f\big](z_t) - f(z_t)}{r_{K,n}-t} = \big[Q^\star_{z_t} f\big](z_t).$$

We repeat this procedure for all other $z_t$ in $\mathscr{X}$; thus, for all $z_t$ in $\mathscr{X}$, we find some $Q^\star_{z_t}$ in $\mathscr{Q}$ such that

$$\big(\forall f\in\mathbb{G}(\mathscr{X})\big)\ [Qf](z_t) = \big[Q^\star_{z_t} f\big](z_t). \tag{3.137}$$

Let $Q^\star\colon \mathbb{G}(\mathscr{X})\to\mathbb{G}(\mathscr{X})$ be the operator defined for all $f$ in $\mathbb{G}(\mathscr{X})$ and $z_t$ in $\mathscr{X}$ by $[Q^\star f](z_t) := [Q^\star_{z_t} f](z_t)$. Because $(Q^\star_{z_t})_{z_t\in\mathscr{X}}$ is a selection in $\mathscr{Q}$ and $\mathscr{Q}$ has separately specified rows, $Q^\star$ belongs to $\mathscr{Q}$. As before, it follows from the definition of $Q^\star$ and Eq. (3.137) that $\|Q-Q^\star\|_{\mathrm{op}}=0$, so $Q=Q^\star$. Because $Q^\star$ belongs to $\mathscr{Q}$ by construction, this proves that $Q$ belongs to $\mathscr{Q}$ whenever $Q$ belongs to $\partial_+ T_{t,t}^{\{X_v=z_v\}}$, as required. □

*Proof of Theorem 3.102*$_{139}$. Let $P$ be the process as given by Theorem 3.103$_{139}$, with $y_u^{x_u} := x_u$. The first equality in the statement follows immediately from the first equality in Theorem 3.103$_{139}$. To verify the second equality in the statement, we fix some $x_u$ in $\mathscr{X}$ and $A$ in $\mathscr{F}_u$. Then by Lemma 3.11$_{64}$, there is a sequence $v$ in $\mathscr{U}_{>u}$ and a subset $B'$ of $\mathscr{X}_{u\cup v}$ such that $A=(X_{u\cup v}\in B')$. Let

$$B := \big\{z_v\in\mathscr{X}_v\colon (\exists y_{u\cup v}\in B')\ y_u = x_u, y_v = z_v\big\}.$$

Then by (JP1)$_{69}$,

$$P(A\,|\,X_u=x_u) = P(X_{u\cup v}\in B'\,|\,X_u=x_u) = P(X_v\in B\,|\,X_u=x_u),$$

and similarly

$$P_{x_u}(A\,|\,X_u=x_u) = P_{x_u}(X_{u\cup v}\in B'\,|\,X_u=x_u) = P_{x_u}(X_v\in B\,|\,X_u=x_u).$$

Because $u\prec v$, these equalities and the second equality in Theorem 3.103$_{139}$ imply the second equality in the statement, as required. □

In our proof for Theorem 3.89$_{120}$ in Appendix 3.D$_{151}$ further on, we need the following result that is similar to – yet not quite a special case of – Theorem 3.103$_{139}$.

**Lemma 3.104.** *Consider a non-empty set $\mathscr{M}$ of initial probability mass functions on $\mathscr{X}$ and a non-empty and bounded subset $\mathscr{Q}$ of rate operators. Fix a jump process $P_1$ in $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and a sequence of time points $u$ in $\mathscr{U}$. Then for all $P_0$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, there is a jump process $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ such that for all $x_u$ in $\mathscr{X}_u$ and $u_1, u_2$ in $\mathscr{U}$ with $u_2\neq()$, $u_1\cup u_2\subseteq u$ and $u_1\prec u_2$,*

$$P(X_{u_2}=x_{u_2}\,|\,X_{u_1}=x_{u_1}) = P_0(X_{u_2}=x_{u_2}\,|\,X_{u_1}=x_{u_1})$$

*and for all $v$ in $\mathscr{U}_{>u}$ and $B\subseteq\mathscr{X}_v$,*

$$P(X_v\in B\,|\,X_u=x_u) = P_1(X_v\in B\,|\,X_u=x_u).$$

*If additionally $P_0$ belongs to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$, then so does $P$.*

*Proof.* Our proof is similar to that of Theorem 3.103$_{139}$, so it consists of the following three parts: (i) we construct a coherent conditional probability $P$ on $\mathscr{D}$ that satisfies the equalities in the statement; (ii) we show that $P$ is consistent with $\mathscr{M}$; (iii) we show that $P$ is consistent with $\mathscr{Q}$; and (iv) we show that $P$ is Markovian if $P_0$ is Markovian. Here too, the statemant is trivial whenever $u = ()$, so we may assume without loss of generality that $u \neq ()$.

For all $x_u$ in $\mathscr{X}_u$, we let $y_u^{x_u} := x_u$ and $P_{x_u} := P_1$. The first part of this proof is then entirely similar to the first part of our proof for Theorem 3.103$_{139}$.

In the second part of this proof, we need to show that $P$ is consistent with $\mathscr{M}$. Here too, the argument is exactly the same as our argument in our proof for Theorem 3.103$_{139}$, so we will not repeat this argument here.

In the third part of this proof, we show that $P$ is consistent with $\mathscr{Q}$. This part of the proof is not the same as the third part of our proof for Theorem 3.103$_{139}$, but it will be a lot easier. For all $\{X_v = z_v\}$ in $\mathscr{H}$ and $t, r$ in $\mathbb{R}_{\geq 0}$ such that $u < t < r$, we denote the history-dependent transition operator corresponding to $P$ by $T_{t,r}^{\{X_v=z_v\}}$, the one corresponding to $P_0$ by $T_{0,t,r}^{\{X_v=z_v\}}$ and the one corresponding to $P_1$ by $T_{1,t,r}^{\{X_v=z_v\}}$. Furthermore, we let

$$Q_{t,r}^{\{X_v=z_v\}} := \frac{T_{t,r}^{\{X_v=z_v\}} - I}{r-t}, \quad Q_{0,t,r}^{\{X_v=z_v\}} := \frac{T_{0,t,r}^{\{X_v=z_v\}} - I}{r-t} \text{ and } Q_{1,t,r}^{\{X_v=z_v\}} := \frac{T_{1,t,r}^{\{X_v=z_v\}} - I}{r-t}.$$

If $t < \max u$ and $r \leq \max u$, then it follows from Eq. (3.123)$_{142}$ that

$$T_{t,r}^{\{X_v=z_v\}} = T_{0,t,r}^{\{X_v=z_v\}}. \tag{3.138}$$

If on the other hand $t \geq \max u$, then it follows from Eq. (3.129)$_{144}$ that for all $z_t$ in $\mathscr{X}$ and $f$ in $\mathbb{G}(\mathscr{X})$,

$$\left[T_{t,r}^{\{X_v=z_v\}} f\right](z_t) = \sum_{z_w \in \mathscr{X}_w} p_t^{z_t}(z_w)\left[T_{z_u,t,r}^{C(z_w)} f\right](z_t) = \sum_{z_w \in \mathscr{X}_w} p_t^{z_t}(z_w)\left[T_{1,t,r}^{C(z_w)} f\right](z_t),$$

where we let $w := u \setminus (v \cup (t))$ and where $p_t^{z_t}$ is a probability mass function on $\mathscr{X}_w$. Note that $P_1$ is Markovian by assumption, so it follows from the preceding equality and Eq. (3.38)$_{85}$ that for all $z_t$ in $\mathscr{X}$ and $f$ in $\mathbb{G}(\mathscr{X})$

$$\begin{aligned}\left[T_{t,r}^{\{X_v=z_v\}} f\right](z_t) &= \sum_{z_w \in \mathscr{X}_w} p_t^{z_t}(z_w)\left[T_{1,t,r} f\right](z_t) = \left[T_{1,t,r} f\right](z_t) \sum_{z_w \in \mathscr{X}_w} p_t^{z_t}(z_w) \\ &= \left[T_{1,t,r} f\right](z_t),\end{aligned}$$

where for the third equality we used (MF2)$_{23}$. Because this equality holds for all $z_t$ in $\mathscr{X}$ and $f$ in $\mathbb{G}(\mathscr{X})$, we conclude that

$$T_{t,r}^{\{X_v=z_v\}} = T_{1,t,r}. \tag{3.139}$$

Because $P_0$ and $P_1$ are consistent with $\mathscr{Q}$ by assumption, it follows almost immediately from Eqs. (3.138) and (3.139) that $P$ is consistent with $\mathscr{Q}$. Recall from Definition 3.50$_{99}$ that we need to establish that for all $\{X_v = z_v\}$ in $\mathscr{H}$ and $t$ in $\mathbb{R}_{\geq 0}$ such that $v < t$, $T_{t,t}^{\{X_v=z_v\}}$ is $d_{\mathfrak{Q}}$-differentiable with

$$\partial_+ T_{t,t}^{\{X_v=z_v\}} \subseteq \mathscr{Q} \quad \text{and, if } t > 0, \quad \partial_- T_{t,t}^{\{X_v=z_v\}} \subseteq \mathscr{Q}.$$

Thus, we fix any $\{X_v = z_v\}$ in $\mathscr{H}$ and $t$ in $\mathbb{R}_{\geq 0}$ such that $v < t$. If $t < \max u$, then it follows immediately from Eq. (3.138)↶ that

$$\big(\forall r \in \,]t, \max u]\big)\; Q_{t,r}^{\{X_v=z_v\}} = Q_{0,t,r}^{\{X_v=z_v\}}.$$

Hence, if $t < \max u$, then

$$\emptyset \neq \partial_+ T_{t,t}^{\{X_v=z_v\}} = \partial_+ T_{0,t,t}^{\{X_v=z_v\}} \subseteq \mathcal{Q}$$

and

$$\lim_{r\searrow t} d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_v=z_v\}}, \partial_+ T_{t,t}^{\{X_v=z_v\}}\big) = \lim_{r\searrow t} d_{\mathfrak{Q}}\big(Q_{0,t,r}^{\{X_v=z_v\}}, \partial_+ T_{0,t,t}^{\{X_v=z_v\}}\big) = 0,$$

where we used that $P_0$ is consistent with $\mathcal{Q}$ by assumption – see Definitions 3.50₉₉ and 3.46₉₇. If on the other hand $t \geq \max u$, then it follows immediately from Eq. (3.139)↶ that

$$\big(\forall r \in \,]t, +\infty[\big)\; Q_{t,r}^{\{X_v=z_v\}} = Q_{1,t,r}.$$

Hence, if $t \geq \max u$, then

$$\emptyset \neq \partial_+ T_{t,t}^{\{X_v=z_v\}} = \partial_+ T_{1,t,t} \subseteq \mathcal{Q}$$

and

$$\lim_{r\searrow t} d_{\mathfrak{Q}}\big(Q_{t,r}^{\{X_v=z_v\}}, \partial_+ T_{t,t}^{\{X_v=z_v\}}\big) = \lim_{r\searrow t} d_{\mathfrak{Q}}\big(Q_{1,t,r}, \partial_+ T_{1,t,t}\big) = 0,$$

where this time we used that $P_1$ is consistent with $\mathcal{Q}$ by assumption. This shows that $T_{t,t}^{\{X_v=z_v\}}$ is $d_{\mathfrak{Q}}$-differentiable from the right with $\partial_+ T_{t,t}^{\{X_v=z_v\}} \subseteq \mathcal{Q}$. If $t > 0$, then a similar argument – but distinguishing the cases $t \leq \max u$ and $t > \max u$ – shows that $T_{t,t}^{\{X_v=z_v\}}$ is $d_{\mathfrak{Q}}$-differentiable from the left with $\partial_- T_{t,t}^{\{X_v=z_v\}} \subseteq \mathcal{Q}$.

In the fourth and final part of this proof, we establish that if $P_0$ is Markovian, then so is $P$. By Definition 3.21₇₄, we need to show that for all $\{X_v = z_v\}$ in $\mathscr{H}$, $t, r$ in $\mathbb{R}_{\geq 0}$ such that $v < t < r$ and $x, y$ in $\mathscr{X}$

$$P(X_r = y \,|\, X_v = z_v, X_t = x) = P(X_r = y \,|\, X_t = x).$$

Hence, we fix any $\{X_v = z_v\}$ in $\mathscr{H}$, $t, r$ in $\mathbb{R}_{\geq 0}$ such that $v < t < r$ and $x, y$ in $\mathscr{X}$. If $r \leq \max u$, then it follows immediately from Eqs. (3.36)₈₄ and (3.138)↶ (the former twice) that

$$\begin{aligned} P(X_r = y \,|\, X_v = z_v, X_t = x) = T_{t,r}^{\{X_v=z_v\}}(x, y) &= T_{0,t,r}^{\{X_v=z_v\}}(x, y) \\ &= P_0(X_r = y \,|\, X_v = z_v, X_t = x) \\ &= P_0(X_r = y \,|\, X_t = x) \\ &= P(X_r = y \,|\, X_t = x), \end{aligned}$$

where for the fourth equality we used that $P_0$ is Markovian. Similarly, if $t \geq \max u$, then it follows from Eqs. (3.36)₈₄ and (3.139)↶ (the former twice) that,

$$\begin{aligned} P(X_r = y \,|\, X_v = z_v, X_t = x) = T_{t,r}^{\{X_v=z_v\}}(x, y) = T_{1,t,r}(x, y) &= P_1(X_r = y \,|\, X_t = x) \\ &= P(X_r = y \,|\, X_t = x). \end{aligned}$$

Finally, we deal with the remaining case that $t < \max u < r$. Let $s := \max u$. Then by (JP4)$_{70}$,

$$P(X_r = y \,|\, X_v = z_v, X_t = x) = \sum_{z \in \mathscr{X}} P(X_r = y \,|\, X_v = z_v, X_t = x, X_s = z) P(X_s = z \,|\, X_v = z_v, X_t = x).$$

Because $t < s \leq \max u$ and $\max u \leq s < r$, it follows from this and Eqs. (3.36)$_{84}$, (3.138)$_{149}$ and (3.139)$_{149}$ that

$$\begin{aligned} P(X_r = y \,|\, X_v = z_v, X_t = x) &= \sum_{z \in \mathscr{X}} T_{s,r}^{\{X_v = z_v, X_t = x\}}(z, y) T_{t,s}^{\{X_v = z_v\}}(x, z) \\ &= \sum_{z \in \mathscr{X}} T_{1,s,r}(z, y) T_{0,t,s}^{\{X_v = z_v\}}(x, z) \\ &= \sum_{z \in \mathscr{X}} P_1(X_r = y \,|\, X_s = z) P_0(X_s = z \,|\, X_v = z_v, X_t = x) \\ &= \sum_{z \in \mathscr{X}} P_1(X_r = y \,|\, X_s = z) P_0(X_s = z \,|\, X_t = x), \end{aligned}$$

wher for the final equality we used that $P_0$ is Markovian. Essentially the same argument shows that

$$P(X_r = y \,|\, X_t = x) = \sum_{z \in \mathscr{X}} P_1(X_r = y \,|\, X_s = z) P_0(X_s = z \,|\, X_t = x),$$

and therefore

$$P(X_r = y \,|\, X_v = z_v, X_t = x) = P(X_r = y \,|\, X_t = x),$$

as required. □

## 3.D Proof of Proposition 3.89

In the fourth and final part of this appendix, we prove Theorem 3.89$_{120}$, which we repeat for convenience.

**Theorem 3.89.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows. Then $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ satisfy the sum-product law of iterated lower expectations.*

In our proof for Theorem 3.89$_{120}$, we rely on the following intermediary results. The first intermediary result that we need is that any imprecise jump process $\mathscr{P}$ satisfies Eq. (3.70)$_{106}$ in Definition 3.59$_{106}$ but with inequality instead of equality.

**Lemma 3.105.** *Consider an imprecise jump process $\mathscr{P}$. Then for all $\{X_u = x_u\}$ in $\mathscr{H}$, all $v$ in $\mathscr{U}_{\neq ()}$ and $t$ in $\mathbb{R}_{\geq 0}$ such that $u \prec v \prec t$, all $f$ in $\mathbb{G}(\mathscr{X})$ and all $g, h$*

*in* $\mathbb{G}(\mathscr{X}_v)$ *such that* $h \geq 0$,

$$\underline{E}_{\mathscr{P}}(f(X_t)h(X_v) + g(X_v) \mid X_u = x_u) \geq \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v) \,\big|\, X_u = x_u\big).$$

In our proof for Lemma 3.105$_{\curvearrowleft}$, and also in our proof for Theorem 3.89$_{120}$ further on, we need the following intermediary result.

**Lemma 3.106.** *Consider a jump process* $P$. *Then for all* $\{X_u = x_u\}$ *in* $\mathscr{H}$, *all* $v$ *in* $\mathscr{U}_{\neq()}$ *and* $t$ *in* $\mathbb{R}_{\geq 0}$ *such that* $u < v < t$, *all* $f$ *in* $\mathbb{G}(\mathscr{X})$ *and all* $g, h$ *in* $\mathbb{G}(\mathscr{X}_v)$ *such that* $h \geq 0$,

$$E_P(f(X_t)h(X_v) + g(X_v) \mid X_u = x_u) = E_P\big(E_P(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v) \,\big|\, X_u = x_u\big).$$

*Proof.* It follows immediately from Theorem 3.19$_{72}$ with $w = v \cup (t)$ that

$$E_P(f(X_t)h(X_v) + g(X_v) \mid X_u = x_u) = E_P\big(E_P(f(X_t)h(X_v) + g(X_v) \mid X_{u\cup v}) \,\big|\, X_u = x_u\big). \quad (3.140)$$

Observe that for any $y_{u\cup v}$ in $\mathscr{X}_{u\cup v}$,

$$\begin{aligned} E_P(f(X_t)h(X_v) + g(X_v) \mid X_{u\cup v} = y_{u\cup v}) &= E_P(f(X_t)h(y_v) + g(y_v) \mid X_{u\cup v} = y_{u\cup v}) \\ &= E_P(f(X_t) \mid X_{u\cup v} = y_{u\cup v})h(y_v) + g(y_v), \end{aligned}$$

where for the first equality we used Corollary 3.18$_{71}$ and for the second equality we used (ES1)$_{37}$–(ES3)$_{37}$. Consequently,

$$E_P(f(X_t)h(X_v) + g(X_v) \mid X_{u\cup v}) = E_P(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v).$$

We substitute this equality into Eq. (3.140), to yield

$$E_P(f(X_t)h(X_v) + g(X_v) \mid X_u = x_u) = E_P\big(E_P(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v) \,\big|\, X_u = x_u\big),$$

as required. ☐

*Proof.* Fix any $P$ in $\mathscr{P}$. Then by Lemma 3.106,

$$E_P(f(X_t)h(X_v) + g(X_v) \mid X_u = x_u) = E_P\big(E_P(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v) \,\big|\, X_u = x_u\big).$$

Note that $E_P(f(X_t) \mid X_{u\cup v}) \geq \underline{E}_{\mathscr{P}}(f(X_t) \mid X_{u\cup v})$; because furthermore $h \geq 0$ by assumption, this implies that

$$E_P(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v) \geq \underline{E}_{\mathscr{P}}(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v).$$

It follows from the previous equality and (ES4)$_{37}$ that

$$E_P(f(X_t)h(X_v) + g(X_v) \mid X_u = x_u) \geq E_P\big(\underline{E}_{\mathscr{P}}(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v) \,\big|\, X_u = x_u\big).$$

Because this inequality holds for all $P$ in $\mathscr{P}$ and inequalities are preserved when taking the infimum, this proves the inequality of the statement. ☐

The second intermediary result that we need is taken from (Krak et al., 2017, Proposition 8.2).

**Lemma 3.107.** *Consider a non-empty set $\mathscr{M}$ of initial probability mass functions and a bounded set $\mathscr{Q}$ of rate operators that has separately specified rows, with corresponding lower rate operator $\underline{Q} := \underline{Q}_{\mathscr{Q}}$. Fix some time points $s$ and $t$ in $\mathbb{R}_{\geq 0}$ such that $s < t$ and some $f$ in $\mathbb{G}(\mathscr{X})$. Then for all $\epsilon$ in $\mathbb{R}_{>0}$, there is a Markovian jump process $P$ in $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ such that*

$$\left|E_P(f(X_t) \mid X_s = x) - \left[e^{(t-s)\underline{Q}} f\right](x)\right| < \epsilon \quad \textit{for all } x \in \mathscr{X}.$$

Finally, we combine the preceding intermediary results to cook up our proof of Theorem $3.89_{120}$.

*Proof of Theorem $3.89_{120}$.* In order not to needlessly repeat ourselves, we let $\mathscr{P}$ be equal to $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ or $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$. By Definition $3.59_{106}$, we need to show that for all $u$ in $\mathscr{U}$, $v$ in $\mathscr{U}_{\neq ()}$ and $t$ in $\mathbb{R}_{\geq 0}$ such that $u \prec v \prec t$, all $f$ in $\mathbb{G}(\mathscr{X})$, all $g, h$ in $\mathbb{G}(\mathscr{X}_v)$ such that $h \geq 0$ and and all $x_u$ in $\mathscr{X}_u$,

$$\begin{aligned}\underline{E}_{\mathscr{P}}(f(X_t)h(X_v) + g(X_v) \mid X_u = x_u) \\ = \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v) \,\big|\, X_u = x_u\big).\end{aligned} \tag{3.141}$$

Hence, we fix any such $u$, $v$, $t$, $f$, $g$, $h$ and $x_u$.

Recall from Lemma $3.105_{151}$ that

$$\begin{aligned}\underline{E}_{\mathscr{P}}(f(X_t)h(X_v) + g(X_v) \mid X_u = x_u) \\ \geq \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f(X_t) \mid X_{u\cup v})h(X_v) + g(X_v) \,\big|\, X_u = x_u\big).\end{aligned} \tag{3.142}$$

To prove Eq. (3.141), we need to prove that the converse inequality holds as well.

To this end, we fix any positive real number $\epsilon$. Furthermore, we let

$$B := \max\{|h(y_v)| : y_v \in \mathscr{X}_v\},$$

and we fix arbitrary positive real numbers $\epsilon_0$ and $\epsilon_1$ such that $\epsilon_0 + B\epsilon_1 \leq \epsilon$.

Let $s := \max v$, and recall from Proposition $3.81_{117}$ that

$$\underline{E}_{\mathscr{P}}(f(X_t) \mid X_{u\cup v}) = \left[e^{(t-s)\underline{Q}} f\right](X_s). \tag{3.143}$$

To simplify our notation, we let

$$f' : \mathscr{X}_{u\cup v} \to \mathbb{R} : y_{u\cup v} \mapsto \left[e^{(t-s)\underline{Q}} f\right](y_s)h(y_v) + g(y_v).$$

By definition of $\underline{E}_{\mathscr{P}}$, there is a jump process $P_0$ in $\mathscr{P}$ such that

$$\left|\underline{E}_{\mathscr{P}}(f'(X_{u\cup v}) \mid X_u = x_u) - E_{P_0}(f'(X_{u\cup v}) \mid X_u = x_u)\right| < \epsilon_0, \tag{3.144}$$

where we have used that $\underline{E}_{\mathscr{P}}(f'(X_{u\cup v}) \mid X_u = x_u)$ is real because, by $(\mathrm{ES1})_{37}$, it is bounded below by $\min f'$ and bounded above by $\max f'$.

Next, we recall from Lemma 3.107 that there is a Markovian jump process $P_1$ in $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P}$ such that

$$\left|E_{P_1}(f(X_t)\,|\,X_s) - \left[e^{(t-s)\underline{Q}}f\right](X_s)\right| < \epsilon_1. \tag{3.145}$$

We now let

$$f'' \colon \mathscr{X}_{u\cup v} \to \mathbb{R} \colon y_{u\cup v} \mapsto E_{P_1}(f(X_t)\,|\,X_{u\cup v} = y_{u\cup v})h(y_v) + g(y_v).$$

Then clearly

$$\begin{aligned}
\left|f'(X_{u\cup v}) - f''(X_{u\cup v})\right| &= \left|\left[e^{(t-s)\underline{Q}}f\right](X_s)h(X_v) - E_{P_1}(f(X_t)\,|\,X_{u\cup v})h(X_v)\right| \\
&= \left|h(X_v)\left(\left[e^{(t-s)\underline{Q}}f\right](X_s) - E_{P_1}(f(X_t)\,|\,X_{u\cup v})\right)\right| \\
&= |h(X_v)|\left|\left[e^{(t-s)\underline{Q}}f\right](X_s) - E_{P_1}(f(X_t)\,|\,X_{u\cup v})\right| \\
&\leq B\left|\left[e^{(t-s)\underline{Q}}f\right](X_s) - E_{P_1}(f(X_t)\,|\,X_{u\cup v})\right|,
\end{aligned}$$

where the inequality holds due to the definition of $B$. Note that, because $P_1$ is Markovian, for all $y_{u\cup v}$ in $\mathscr{X}_v$,

$$\begin{aligned}
E_{P_1}(f(X_t)\,|\,X_{u\cup v} = y_{u\cup v}) &= \sum_{z\in\mathscr{X}} f(z)P_1(X_t = z\,|\,X_{u\cup v} = y_{u\cup v}) \\
&= \sum_{z\in\mathscr{X}} f(z)P_1(X_t = z\,|\,X_s = y_s) \\
&= E_{P_1}(f(X_t)\,|\,X_s = y_s).
\end{aligned}$$

This implies that

$$E_{P_1}(f(X_t)\,|\,X_{u\cup v}) = E_{P_1}(f(X_t)\,|\,X_s),$$

so it follows from the preceding inequality and Eq. (3.145) that

$$\left|f'(X_{u\cup v}) - f''(X_{u\cup v})\right| \leq B\epsilon_1. \tag{3.146}$$

From Eqs. (3.144) and (3.146), we infer that

$$\begin{aligned}
\underline{E}_{\mathscr{P}}(f'(X_{u\cup v})\,|\,X_u = x_u) &> E_{P_0}(f'(X_{u\cup v})\,|\,X_u = x_u) - \epsilon_0 \\
&\geq E_{P_0}(f''(X_{u\cup v}) - B\epsilon_1\,|\,X_u = x_u) - \epsilon_0 \\
&= E_{P_0}(f''(X_{u\cup v})\,|\,X_u = x_u) - \epsilon_0 - B\epsilon_1,
\end{aligned}$$

where we have also used (ES4)$_{37}$ for the second inequality and (ES3)$_{37}$ and (ES1)$_{37}$ for the equality. Because $\epsilon_0 + B\epsilon_1 \leq \epsilon$ by assumption, we conclude that

$$\underline{E}_{\mathscr{P}}(f'(X_{u\cup v})\,|\,X_u = x_u) > E_{P_0}(f''(X_{u\cup v})\,|\,X_u = x_u) - \epsilon. \tag{3.147}$$

By Lemma 3.104$_{148}$ (with $u\cup v$ here in the role of $u$ there, $u$ here in the role of $u_1$ there and $v$ here in the role of $u_2$ there), there is a jump process $P$ in $\mathscr{P}$ such that

$$(\forall y_v \in \mathscr{X}_v)\ P(X_v = y_v\,|\,X_u = x_u) = P_0(X_v = y_v\,|\,X_u = x_u), \tag{3.148}$$

and for all $y_{u\cup v}$ in $\mathscr{X}_{u\cup v}$,

$$(\forall z \in \mathscr{X})\ P(X_t = z\,|\,X_{u\cup v} = y_{u\cup v}) = P_1(X_t = z\,|\,X_{u\cup v} = y_{u\cup v}). \tag{3.149}$$

Clearly, Eq. (3.149)↶ implies that for all $y_{u\cup v} \in \mathscr{X}_{u\cup v}$,

$$\begin{aligned} E_P(f(X_t)\,|\,X_{u\cup v} = y_{u\cup v}) &= \sum_{z\in\mathscr{X}} f(z)P(X_t = z\,|\,X_{u\cup v} = y_{u\cup v}) \\ &= \sum_{z\in\mathscr{X}} f(z)P_1(X_t = z\,|\,X_{u\cup v} = y_{u\cup v}) \\ &= E_{P_1}(f(X_t)\,|\,X_{u\cup v} = y_{u\cup v}). \end{aligned}$$

Consequently,

$$E_P(f(X_t)\,|\,X_{u\cup v}) = E_{P_1}(f(X_t)\,|\,X_{u\cup v}).$$

Similarly, it follows from Corollary $3.18_{71}$ and Eq. (3.148)↶ that

$$\begin{aligned} E_P(f''(X_{u\cup v})\,|\,X_u = x_u) = E_P(f''(x_u, X_v)\,|\,X_u = x_u) &= E_{P_0}(f''(x_u, X_v)\,|\,X_u = x_u) \\ &= E_{P_0}(f''(X_{u\cup v})\,|\,X_u = x_u). \end{aligned}$$

Thus, we find that

$$\begin{aligned} E_{P_0}(f''(X_{u\cup v})\,|\,X_u = x_u) &= E_P(f''(X_{u\cup v})\,|\,X_u = x_u) \\ &= E_P(E_{P_1}(f(X_t)\,|\,X_{u\cup v})h(X_v) + g(X_v)\,|\,X_u = x_u) \\ &= E_P(E_P(f(X_t)\,|\,X_{u\cup v})h(X_v) + g(X_v)\,|\,X_u = x_u) \\ &= E_P(f(X_t)h(X_v) + g(X_v)\,|\,X_u = x_u), \end{aligned} \tag{3.150}$$

where for the final equality we used Lemma $3.106_{152}$.

At last, we are ready to combine our intermediary findings. It follows from Eq. (3.147)↶ and Eq. (3.150) that

$$\begin{aligned} \underline{E}_{\mathscr{P}}(f'(X_{u\cup v})\,|\,X_u = x_u) &> E_{P_0}(f''(X_{u\cup v})\,|\,X_u = x_u) - \epsilon \\ &= E_P(f(X_t)h(X_v) + g(X_v)\,|\,X_u = x_u) - \epsilon \\ &\geq \underline{E}_{\mathscr{P}}(f(X_t)h(X_v) + g(X_v)\,|\,X_u = x_u) - \epsilon, \end{aligned}$$

where the last inequality holds because $P$ belongs to $\mathscr{P}$. We rewrite the left-hand side of this inequality with the help of Eq. $(3.143)_{153}$, to yield

$$\begin{aligned} &\underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f(X_t)\,|\,X_{u\cup v})h(X_v) + g(X_v)\,\big|\,X_u = x_u\big) \\ &\qquad > \underline{E}_{\mathscr{P}}(f(X_t)h(X_v) + g(X_v)\,|\,X_u = x_u) - \epsilon. \end{aligned}$$

Because $\epsilon$ is an arbitrary positive real number, we infer from this inequality that

$$\begin{aligned} &\underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f(X_t)\,|\,X_{u\cup v})h(X_v) + g(X_v)\,\big|\,X_u = x_u\big) \\ &\qquad \geq \underline{E}_{\mathscr{P}}(f(X_t)h(X_v) + g(X_v)\,|\,X_u = x_u). \end{aligned}$$

Finally, Eq. $(3.141)_{153}$ follows immediately from this inequality and Eq. $(3.142)_{153}$. □

# 4 Computing lower expectations of simple variables

In the previous chapter, we introduced three homogeneous and imprecise Markovian jump processes through consistency with a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate operators: $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$. In this chapter, we examine if and how we can calculate the corresponding conditional lower expectations $\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\underline{E}_{\mathscr{M},\mathscr{Q}}$. More precisely, in Section 4.1 we will discover that the various laws of iterated (lower) expectations play a crucial role, as well as the semi-group $(e^{t\underline{Q}})_{t\in\mathbb{R}_{\geq 0}}$ of lower transition operators generated by the lower envelope $\underline{Q} := \underline{Q}_{\mathscr{Q}}$ of $\mathscr{Q}$. For this reason, in Section 4.2$_{173}$ we turn our attention to numerical methods that approximate $e^{t\underline{Q}}f$ – or, to be more exact, solve the initial value problem of Proposition 3.78$_{115}$. In particular, we propose two methods to choose the step size for the well-known Euler method in such a way that the approximation error is guaranteed to be lower than some desired maximal error. These two methods are related to the notion of ergodicity, and we will dig into this relation in Section 4.3$_{188}$.

Section 4.1 is loosely based on (Krak et al., 2017, Section 9), for the most part. One important exception is Section 4.1.3$_{162}$, which contains novel results that sprout from the sum-product law of iterated lower expectations. In Sections 4.2$_{173}$ and 4.3$_{188}$, we more or less follow (Erreygers et al., 2017a,b).

## 4.1 Why we need a law of iterated (lower) expectations

Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate operators. For a given state history $\{X_u = x_u\}$ in $\mathscr{H}$ and an $\mathscr{F}_u$-simple variable $g$, we want to determine the (conditional) lower expectations

$$\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}(g\,|\,X_u = x_u),\quad \underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(g\,|\,X_u = x_u)\quad\text{and}\quad \underline{E}_{\mathscr{M},\mathscr{Q}}(g\,|\,X_u = x_u).$$

Recall from Lemma 3.15$_{68}$ that the $\mathscr{F}_u$-simple variable $g$ has a representation of the form $f(X_u, X_v)$, with $v$ a sequence of time points in $\mathscr{U}_{>u}$ and $f$ a gamble on $\mathscr{X}_{u\cup v}$. We can consider an even less general case due to the following immediate consequence of Corollary 3.18$_{71}$.

**Corollary 4.1.** *Consider an imprecise jump process $\mathscr{P}$. For any $\{X_u = x_u\}$ in $\mathscr{H}$, any $v$ in $\mathscr{U}_{\succcurlyeq u}$ and any gamble $f$ on $\mathscr{X}_v$,*

$$\underline{E}_{\mathscr{P}}(f(X_v) \mid X_u = x_u) = \underline{E}_{\mathscr{P}}(f(x_{u\cap v}, X_{v\setminus u}) \mid X_u = x_u).$$

It therefore follows immediately that

$$\underline{E}_{\mathscr{P}}(g \mid X_u = x_u) = \underline{E}_{\mathscr{P}}(f(X_u, X_v) \mid X_u = x_u) = \underline{E}_{\mathscr{P}}(f(x_u, X_v) \mid X_u = x_u),$$

where $\mathscr{P}$ is equal to $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ or $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. For this reason, we can henceforth focus on determining

$$\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}(f(X_v) \mid X_u = x_u), \underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f(X_v) \mid X_u = x_u) \text{ and } \underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_v) \mid X_u = x_u)$$

for all $v$ in $\mathscr{U}_{\neq()}$ such that $v \succ u$ and all $f$ in $\mathbb{G}(\mathscr{X}_v)$. Note that in case $v = ()$, determining the lower expectations of $f(X_v)$ is trivial because then $f(X_v)$ is a constant.

### 4.1.1 A single initial mass function and rate operator

In order not to dive into the deep end of the pool immediately, we start off gently with the simple case of a single initial mass function $p_0$ and rate operator $Q$. Recall from Eq. (3.66)$_{104}$ that in this case, we might as well study the unique (homogeneous and Markovian) jump process $P_{p_0,Q}$ that is characterised by $p_0$ and $Q$ in the sense of Theorem 3.37$_{87}$; as explained in Section 3.2.4$_{87}$, we let $E_{p_0,Q} = E_{P_{p_0,Q}}$ denote the corresponding conditional expectation.

One of the reasons that homogeneous and Markovian jump processes are so popular, is that we can compute $E_{p_0,Q}(f(X_v) \mid X_u = x_u)$ through backwards recursion due to Theorem 3.19$_{72}$, the law of iterated expectations. More precisely, one can use Theorems 3.19$_{72}$ and 3.37$_{87}$ to obtain the recursive method of Algorithm 4.1$_{\curvearrowright}$; alternatively, Algorithm 4.1$_{\curvearrowright}$ is an obvious specialisation of Algorithm 4.3$_{171}$ further on. Crucially, in this recursive method we only need the operator exponential $(e^{tQ})_{t\in\mathbb{R}_{\geq 0}}$ of $Q$ and the coherent expectation $E_{p_0}$ that corresponds to the initial mass function $p_0$. Let us illustrate this with our running example, drawing inspiration from (Krak et al., 2017, Example 9.1).

**Joseph's Example 4.2.** Recall from Joseph's Example 3.28$_{82}$ that for the binary state space $\mathscr{X} = \{\mathtt{H}, \mathtt{T}\}$, any rate operator $Q$ has a matrix representation of the form

$$\begin{pmatrix} Q(\mathtt{H},\mathtt{H}) & Q(\mathtt{H},\mathtt{T}) \\ Q(\mathtt{T},\mathtt{H}) & Q(\mathtt{T},\mathtt{T}) \end{pmatrix} = \begin{pmatrix} -\lambda_{\mathtt{H}} & \lambda_{\mathtt{H}} \\ \lambda_{\mathtt{T}} & -\lambda_{\mathtt{T}} \end{pmatrix},$$

**Algorithm 4.1:** Iteratively computing $E_{p_0,Q}(f(X_v)\,|\,X_u = x_u)$

**Input:** An initial mass function $p_0$, a rate operator $Q$, a state history $\{X_u = x_u\}$ in $\mathscr{H}$, a sequence of time points $v = (t_1, \ldots, t_n)$ in $\mathscr{U}_{>u}$ and a gamble $f$ on $\mathscr{X}_v$.

**Output:** $E_{p_0,Q}(f(X_v)\,|\,X_u = x_u)$

$f_n := f$
**for** $k \in \{n-1, \ldots, 1\}$ **do**
  $t_{1:k} := (t_1, \ldots, t_k)$
  $\Delta := t_{k+1} - t_k$
  $f_k \colon \mathscr{X}_{t_{1:k}} \to \mathbb{R}$
  **for** $y_{t_{1:k}} \in \mathscr{X}_{t_{1:k}}$ **do**
    $f'_{k+1} := f_{k+1}(y_{t_{1:k}}, \bullet) \colon \mathscr{X} \to \mathbb{R} \colon z \mapsto f_{k+1}(y_{t_{1:k}}, z)$
    $f_k(y_{t_{1:k}}) := \left[e^{\Delta Q} f'_{k+1}\right](y_{t_k})$
**if** $u \neq ()$ **then** ▷ The $n$-th step requires some special care
  $\Delta := t_1 - \max u$
  **return** $\left[e^{\Delta Q} f_1\right](x_{\max u})$
**else if** $t_1 = 0$ **then**
  **return** $E_{p_0}(f_1)$
**else**
  $f_0 \colon \mathscr{X} \to \mathbb{R}$
  **for** $x \in \mathscr{X}$ **do**
    $f_0(x) := \left[e^{t_1 Q} f_1\right](x)$
  **return** $E_{p_0}(f_0)$

where $\lambda_{\mathrm{H}}$ and $\lambda_{\mathrm{T}}$ are two non-negative real numbers. In Joseph's Example 3.32$_{83}$, we furthermore obtained an analytical expression for the operator exponential $(e^{tQ})_{t \in \mathbb{R}_{\geq 0}}$ generated by such a rate operator $Q$.

Consider any initial mass function $p_0$, and a rate operator $Q$ with $\lambda_{\mathrm{H}}$ and $\lambda_{\mathrm{T}}$ both positive. We want to determine the probability that the state at time $t_1$ is equal to the state at time $t_2$, with $t_1$ and $t_2$ in $\mathbb{R}_{\geq 0}$ such that $0 < t_1 < t_2$. More formally, we want to determine

$$P_{p_0,Q}(X_{t_1} = X_{t_2}) = E_{p_0,Q}(f(X_{t_1}, X_{t_2})) = E_{p_0,Q}(f(X_{t_1}, X_{t_2})\,|\,X_{()} = x_{()}),$$

with $f$ the function on $\mathscr{X}^2 = \mathscr{X}_{(t_1,t_2)}$ such that, for any two states $x$ and $y$ in $\mathscr{X}$, $f(x,y) = 1$ if $x = y$ and 0 otherwise.

Let us determine this probability with Algorithm 4.1; note that $u = ()$ and $v = (t_1, t_2)$. First, we need to determine the components of the gamble $g_1$; that is, for any $x$ in $\mathscr{X}$, we need to determine

$$f_1(x) = \left[e^{\Delta_1 Q} f(x, \bullet)\right](x),$$

with $\Delta_1 := (t_2 - t_1)$. By Eq. (3.32)83,

$$f_1(\mathrm{H}) = f(\mathrm{H},\mathrm{H}) + \frac{1 - e^{-\Delta_1(\lambda_\mathrm{H}+\lambda_\mathrm{T})}}{\lambda_\mathrm{H}+\lambda_\mathrm{T}}\lambda_\mathrm{H}\big(f(\mathrm{H},\mathrm{T}) - f(\mathrm{H},\mathrm{H})\big)$$
$$= 1 - \big(1 - e^{-\Delta_1(\lambda_\mathrm{H}+\lambda_\mathrm{T})}\big)\frac{\lambda_\mathrm{H}}{\lambda_\mathrm{H}+\lambda_\mathrm{T}},$$

and by symmetry,

$$f_1(\mathrm{T}) = 1 - \big(1 - e^{-\Delta_1(\lambda_\mathrm{H}+\lambda_\mathrm{T})}\big)\frac{\lambda_\mathrm{T}}{\lambda_\mathrm{H}+\lambda_\mathrm{T}}.$$

Because $\{X_u = x_u\} = \Omega$ and hence $u = ()$, we also need to determine the components of the gamble $f_0$ on $\mathscr{X}$, given for all $x$ in $\mathscr{X}$ by

$$f_0(x) = \big[e^{t_1 Q} f_1\big](x).$$

Again, it follows from Eq. (3.32)83 and some algebra that

$$f_0(\mathrm{H}) = 1 - \big(1 - e^{-\Delta_1(\lambda_\mathrm{H}+\lambda_\mathrm{T})}\big)\frac{\lambda_\mathrm{H}}{\lambda_\mathrm{H}+\lambda_\mathrm{T}}\left(1 + \big(1 - e^{-t_1(\lambda_\mathrm{H}+\lambda_\mathrm{T})}\big)\frac{\lambda_\mathrm{T}-\lambda_\mathrm{H}}{\lambda_\mathrm{H}+\lambda_\mathrm{T}}\right)$$

and

$$f_0(\mathrm{T}) = 1 - \big(1 - e^{-\Delta_1(\lambda_\mathrm{H}+\lambda_\mathrm{T})}\big)\frac{\lambda_\mathrm{T}}{\lambda_\mathrm{H}+\lambda_\mathrm{T}}\left(1 + \big(1 - e^{-t_1(\lambda_\mathrm{H}+\lambda_\mathrm{T})}\big)\frac{\lambda_\mathrm{H}-\lambda_\mathrm{T}}{\lambda_\mathrm{H}+\lambda_\mathrm{T}}\right).$$

Finally, we can determine the desired probability, because

$$P_{p_0,Q}(X_{t_1} = X_{t_2}) = E_{p_0}(f_0) = p_0(\mathrm{H})f_0(\mathrm{H}) + p_0(\mathrm{T})f_0(\mathrm{T}).$$ ♗

For state spaces that consist of three or more states, it is usually infeasible if not impossible to obtain an analytical expression for the operator exponential $e^{tQ}$ of the rate operator. Fortunately, one can resort to one of the many numerical methods that numerically approximate the operator exponential; for a broad overview of such numerical methods, we refer to (Moler et al., 2003). Alternatively, we can also use the numerical methods that we will discuss in Section 4.2173 further on.

### 4.1.2 Finite sets of initial mass functions and rate operators

Instead of a single initial mass function $p_0$ and rate operator $Q$, we now consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate operators. To wet our feet, we will initially assume that $\mathscr{Q}$ is a finite set; in order not to deal with an obvious degenerate case, we assume that $\mathscr{Q}$ is not a singleton. Out of convenience, we furthermore assume that $\mathscr{M}$ is a finite set as well; this is not essential, but it lets us focus on what really matters.

Computing lower expectations with respect to $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is not all too difficult. To unearth why, we recall from Eq. (3.46)$_{90}$ that

$$\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}} = \{P_{p_0,Q} \colon p_0 \in \mathscr{M}, Q \in \mathscr{Q}\}.$$

Consequently, for all state histories $\{X_u = x_u\}$ in $\mathscr{H}$, all sequences of time points $v$ in $\mathscr{U}_{>u}$ and all gambles $f$ on $\mathscr{X}_v$,

$$\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}(f(X_v) \mid X_u = x_u) = \inf\{E_{p_0,Q}(f(X_v) \mid X_u = x_u) \colon p_0 \in \mathscr{M}, Q \in \mathscr{Q}\}.$$

Because we have assumed that $\mathscr{M}$ and $\mathscr{Q}$ are both finite, we can compute $\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}(f(X_v) \mid X_u = x_u)$ by exhaustive search. That is, we compute $E_{p_0,Q}(f(X_v) \mid X_u = x_u)$ for every $(p_0, Q)$ in $\mathscr{M} \times \mathscr{Q}$, and subsequently pick the smallest of these precise expectations. Obviously, this method is only computationally tractable for small enough sets $\mathscr{M}$ and $\mathscr{Q}$.

**Joseph's Example 4.3.** Recall from Joseph's Example 3.40$_{89}$ that Deborah's beliefs are modelled by the imprecise jump process $\mathbb{P}_{\mathscr{M},\mathscr{Q}_1}$, with

$$\mathscr{Q}_1 := \left\{ \begin{pmatrix} -\lambda_{\mathrm{H}} & \lambda_{\mathrm{H}} \\ \lambda_{\mathrm{T}} & -\lambda_{\mathrm{T}} \end{pmatrix} \colon \lambda_{\mathrm{H}}, \lambda_{\mathrm{T}} \in \{\lambda_1, \ldots, \lambda_n\} \right\}$$

and $\mathscr{M}$ the set of all initial mass functions. Here, we will assume that $\mathscr{M} := \{p_0\}$ with $p_0 := \mathbb{I}_{\mathrm{H}}$ for the sake of simplicity.

It will be instructive to look at a numerical example, so we should fix some parameter values. Say there are $n := 4$ radioactive isotopes, with parameters $\lambda_1 := 1$, $\lambda_2 := 3/2$, $\lambda_3 = 4/3$ and $\lambda_4 = 5/4$. The parameter $\lambda_k$ of the $k$-th radioactive isotope can be interpreted in two ways: (i) on average, there are $\lambda_k$ decays per time unit; and (ii) the average time between two radioactive decays is $1/\lambda_k$ time units.

Deborah is interested in a lower bound on the probability that Joseph's machine displays the same outcome after one time unit and after two time units. In our formal framework, she is interested in

$$\begin{aligned}\underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_1}(X_1 = X_2) &= \inf\{P_{p_0,Q}(X_1 = X_2) \colon Q \in \mathscr{Q}_1\} \\ &= \inf\{E_{p_0,Q}(f(X_1, X_2)) \colon Q \in \mathscr{Q}_1\} = \underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_1}(f(X_1, X_2)),\end{aligned}$$

with $f$ the gamble on $\mathscr{X}_{(t_1,t_2)} = \mathscr{X}^2$ as defined in Joseph's Example 4.2$_{158}$. To compute this lower probability by means of exhaustive search, we need to compute the probability $P_{p_0,Q}(X_1 = X_2)$ for every rate operator $Q$ in $\mathscr{Q}_1$ that is defined by a couple of parameters $(\lambda_{\mathrm{H}}, \lambda_{\mathrm{T}})$ in $\{\lambda_1, \ldots, \lambda_4\}^2$. Using the procedure of Joseph's Example 4.2$_{158}$, we obtain the values – up to four significant digits – listed below.

| | $\lambda_{\mathrm{T}} = \lambda_1$ | $\lambda_{\mathrm{T}} = \lambda_2$ | $\lambda_{\mathrm{T}} = \lambda_3$ | $\lambda_{\mathrm{T}} = \lambda_4$ |
|---|---|---|---|---|
| $\lambda_{\mathrm{H}} = \lambda_1$ | 0.5677 | 0.5654 | 0.5631 | 0.5629 |
| $\lambda_{\mathrm{H}} = \lambda_2$ | 0.5504 | 0.5249 | 0.5293 | 0.5329 |
| $\lambda_{\mathrm{H}} = \lambda_3$ | 0.5506 | 0.5326 | 0.5347 | 0.5371 |
| $\lambda_{\mathrm{H}} = \lambda_4$ | 0.5524 | 0.5383 | 0.5393 | 0.5410 |

Thus, by taking the minimum of these values, we see that

$$\underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_1}(X_1 = X_2) \approx 0.5249.$$

Can we also compute the (conditional) lower expectations with respect to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ by means of the same procedure? In general, the answer to this question seems to be negative, as there is no straightforward way to construct all consistent Markovian jump processes, let alone all consistent non-Markovian jump processes. This is not as big as a problem as one might think though, because in many – practically relevant – cases, the set $\mathscr{Q}$ of rate operators has separately specified rows. Whenever this is the case, as we are about to explain, we can invoke the results in Chapter $3_{53}$ to compute the lower expectation of many – and in some cases even all – variables in the domain of $\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\underline{E}_{\mathscr{M},\mathscr{Q}}$. In Section 4.1.3, we will investigate how separately specified rows allow us to compute the (conditional) lower expectation of variables that have a 'sum-product representation'. In Section $4.1.4_{170}$, we will furthermore assume that $\mathscr{Q}$ is convex, which allows us to compute all (conditional) lower expectations corresponding to $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$.

### 4.1.3 A set $\mathscr{Q}$ of rate operators that has separately specified rows

Consider a set $\mathscr{Q}$ of rate operators that has separately specified rows. Note that this set can be finite, but it does not have to be. Furthermore, we also let go of the requirement that $\mathscr{M}$ is a finite set. Our intention is to identify variables for which we can determine the lower expectation.

First, we recall from Proposition $3.82_{118}$ that, for any $f$ in $\mathbb{G}(\mathscr{X})$,

$$\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}(f(X_0)) = \underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f(X_0)) = \underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_0)) = \underline{E}_{\mathscr{M}}(f)$$

Of course, this equation is only useful in case we can actually compute this infimum. Fortunately, this typically is not an issue, because more often than not, $\mathscr{M}$ is specified by means of linear inequality constraints; whenever this is the case, we can use any linear programming method to (numerically) solve the optimisation problem above.

With regards to conditional lower expectations, we recall from Proposition $3.81_{117}$ that because the set $\mathscr{Q}$ has separately specified rows and lower envelope $\underline{Q} \coloneqq \underline{Q}_{\mathscr{Q}}$,

$$\begin{aligned}\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f(X_r) \mid X_u = x_u, X_t = x) &= \underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_r) \mid X_u = x_u, X_t = x) \\ &= \left[e^{(r-t)\underline{Q}} f\right](x)\end{aligned}$$

for any current time point $t$ and state $x$, any future time point $r$ and gamble $f$ on $\mathscr{X}$ and any state history $\{X_u = x_u\}$. Furthermore, we can also determine the (marginal) lower expectation of any variable that depends on the state of the system at a single time point: by Proposition $3.83_{118}$,

$$\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f(X_t)) = \underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_t)) = \underline{E}_{\mathscr{M}}(e^{t\underline{Q}} f)$$

for any time point $t$ in $\mathbb{R}_{\geq 0}$ and any gamble $f$ on $\mathscr{X}$. For now, let us assume that we can actually compute $\left[e^{(r-t)\underline{Q}} f\right](x)$; this is certainly so in the case of a binary state space, as we have seen in Joseph's Example $3.79_{115}$. We will return to this assumption in Section $4.2_{173}$ further on.

But what about the lower expectation corresponding to $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$? Well, even in the case of basic (non-trivial) conditional lower expectations, computing this lower expectation becomes intractable whenever $\mathscr{Q}$ is *not* finite. This is because by Eq. $(3.46)_{90}$ and Theorem $3.37_{87}$,

$$\begin{aligned}&\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}(f(X_r) \mid X_u = x_u, X_t = x) \\ &\quad = \inf\left\{E_{p_0,Q}(f(X_r) \mid X_u = x_u, X_t = x) \colon p_0 \in \mathscr{M}, Q \in \mathscr{Q}\right\} \\ &\quad = \inf\left\{\left[e^{(r-t)Q} f\right](x) \colon Q \in \mathscr{Q}\right\}.\end{aligned}$$

Thus, to compute even the simplest conditional lower expectation, one needs to solve a constrained non-linear optimisation problem,[1] and more general conditional lower expectations are even worse. Krak et al. (2017, Section 6.3) argue that one way to solve this problem, is to discretise the sets $\mathscr{M}$ and $\mathscr{Q}$ and then carry out an exhaustive search over the discretised parameter space. However, this approach is only feasible for small state spaces because its computational complexity explodes as the size of the state space $\mathscr{X}$ increases.

**Joseph's Example 4.4.** In Joseph's Example $4.3_{161}$, we considered the set $\mathscr{Q}_1$ as defined in Joseph's Example $3.40_{89}$. An obvious more general set of rate operators is

$$\mathscr{Q}_2 \coloneqq \left\{\begin{pmatrix} -\lambda_{\mathtt{H}} & \lambda_{\mathtt{H}} \\ \lambda_{\mathtt{T}} & -\lambda_{\mathtt{T}} \end{pmatrix} \colon \lambda_{\mathtt{H}}, \lambda_{\mathtt{T}} \in [\underline{\lambda}, \overline{\lambda}]\right\},$$

with $\underline{\lambda} \coloneqq \min\{\lambda_1, \dots, \lambda_n\}$ and $\overline{\lambda} \coloneqq \max\{\lambda_1, \dots, \lambda_n\}$. Note that $\mathscr{Q}_2$ is the convex hull of $\mathscr{Q}_1$, and that $\mathscr{Q}_2$ has separately specified rows.

[1] In case the set $\mathscr{Q}$ is convex, the minimum need not be reached by one of the extreme points of $\mathscr{Q}$ (Krak et al., 2017, Example 6.2).

To illustrate the 'discretise and search' approach, we determine $\underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_2}(X_1 = X_2)$ for the same numerical values as in Joseph's Example 4.3161. We discretise the parameter space with a 'rectangular grid', that is, we let $\lambda_{\mathrm{H}}$ and $\lambda_{\mathrm{T}}$ assume one of the 1 000 equidistant points in the interval $[\underline{\lambda}, \overline{\lambda}] = [1, 3/2]$, including the start and endpoint. Next, we compute $P_{p_0,Q}(X_1 = X_2)$ for every rate operator $Q$ corresponding to the values for $\lambda_{\mathrm{H}}$ and $\lambda_{\mathrm{T}}$; this way, we obtain a table with 1 000 rows and 1 000 columns. The minimum of these values is

$$\underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_2}(X_1 = X_2) \approx 0.5249.$$

Note that, up to four significant digits, $\underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_2}(X_1 = X_2)$ and $\underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_1}(X_1 = X_2)$ are equal. In our discretisation, the minimal value is the one for the homogeneous and Markovian jump process $P_{p_0,Q}$ with rate operator $Q$ characterised by $\lambda_H = \overline{\lambda} = \lambda_2$ and $\lambda_{\mathrm{T}} = \overline{\lambda} = \lambda_2$, as in Joseph's Example 4.3161. At first sight, one might think that this is to be expected. In particular, for any $0 \leq t_1 < t_2$, one might think that to minimise the probability of $\{X_{t_1} = X_{t_2}\}$, one should choose the largest rates, as this increases the probability of having a jump after $t_1$ but before $t_2$. However, this large rate also increases the probability of jumping back to $X_{t_1}$ before $t_2$, so this argument works both ways. In general, the lower probability of the event $\{X_{t_1} = X_{t_2}\}$ is not the same for $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_1}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_2}$; for example, for $t_1 = 1$ and $t_2 = 4$, we find that, up to four significant digits,

$$\underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_1}(X_1 = X_4) \approx 0.4995 \quad \text{and} \quad \underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_2}(X_1 = X_4) \approx 0.4989.$$ ♎︎

Because the 'discretise and exhaustive search' method to compute $\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is infeasible for large state spaces, we henceforth focus on computing (conditional) lower expectations with respect to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. For these models, we have only given expressions for the (conditional) lower expectation of variables that depend on the state of the system at a single point in time. Fortunately, we can also compute the (conditional) lower expectation for more general variables. An important class of such variables are those that have a sum-product representation over a finite number of time points.

**Definition 4.5.** Consider a sequence $v = (t_1, \ldots, t_n)$ in $\mathscr{U}_{\neq()}$. A real variable $f$ in $\mathbb{V}(\Omega)$ has a *sum-product representation over* $v$ if there are gambles $g_1, \ldots, g_n$ in $\mathbb{G}(\mathscr{X})$ and non-negative gambles $h_1, \ldots, h_{n-1}$ in $\mathbb{G}(\mathscr{X})$ such that

$$\begin{aligned} f &= \sum_{k=1}^{n} g_k(X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell}) \\ &= g_1(X_{t_1}) + h_1(X_{t_1}) g_2(X_{t_2}) + \cdots + h_1(X_{t_1}) \cdots h_{n-1}(X_{t_{n-1}}) g_n(X_{t_n}). \end{aligned}$$

**Lemma 4.6.** *Consider a sequence $v = (t_1, \ldots, t_n)$ in $\mathscr{U}_{\neq()}$ and a real variable $f$ in $\mathbb{V}(\Omega)$ with a sum-product representation over $v$. Then for any sequence $u$ in $\mathscr{U}$ such that $u \preccurlyeq v$, $f$ is $\mathscr{F}_u$-simple.*

*Proof.* Because $f$ has a sum-product representation over $v$, there are gambles $g_1, \dots, g_n$ in $\mathbb{G}(\mathscr{X})$ and non-negative gambles $h_1, \dots, h_{n-1}$ in $\mathbb{G}(\mathscr{X})$ such that

$$f = \sum_{k=1}^{n} g_k(X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell}).$$

For any $k$ in $\{1, \dots, n\}$, we let $t_{1:k} := (t_1, \dots, t_k)$. Then clearly

$$f = \sum_{k=1}^{n} \sum_{x_{t_{1:k}} \in \mathscr{X}_{t_{1:k}}} \left( g_k(x_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(x_{t_\ell}) \right) \mathbb{I}_{\{X_{t_{1:k}} = x_{t_{1:k}}\}}.$$

Because $v \succcurlyeq u$ by assumption, $t_{1:k} \succcurlyeq u$ for all $k$ in $\{1, \dots, n\}$; hence, by definition of $\mathscr{F}_u$, every event of the form $\{X_{t_{1:k}} = x_{t_{1:k}}\}$ in the right-hand side of the equality above belongs to $\mathscr{F}_u$. Consequently, $f$ is $\mathscr{F}_u$-simple due to the equality above. □

**Lemma 4.7.** *Consider a sequence $v = (t_1, \dots, t_n)$ in $\mathscr{U}_{\neq ()}$ and a real variable $f$ in $\mathbb{V}(\Omega)$ with a sum-product representation*

$$f = \sum_{k=1}^{n} g_k(X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell})$$

*over $v$. Then*

$$-f = \sum_{k=1}^{n} [-g_k](X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell}),$$

*so $-f$ has a sum-product representation over $v$ too.*

*Proof.* It is obvious that

$$-f = -\sum_{k=1}^{n} g_k(X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell}) = \sum_{k=1}^{n} [-g_k](X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell}).$$

For all $k$ in $\{1, \dots, n\}$, $g_k$ belongs to $\mathbb{G}(\mathscr{X})$ by assumption, so $-g_k$ belongs to $\mathbb{G}(\mathscr{X})$ because $\mathbb{G}(\mathscr{X})$ is a real vector space. This proves that $-f$ has a sum-product representation over $v$, as required. □

This last result is relevant because, due to conjugacy, it allows us to focus exclusively on determining lower expectations of variables with a sum product representation. Indeed, to determine the upper expectation

$$\overline{E}_{\mathscr{P}}(f \mid X_u = x_u) = -\underline{E}_{\mathscr{P}}(-f \mid X_u = x_u)$$

of a variable $f$ with sum-product representation over $v$, it suffices to determine the lower expectation $\underline{E}_{\mathscr{P}}(-f \mid X_u = x_u)$ of $-f$, which, conveniently, also has a sum-product representation over $v$.

As suggested by our terminology, there is a link between variables with a sum-product representation and the sum-product law of iterated lower expectations – see Definition 3.59$_{106}$. The following result establishes this link: given a Markovian imprecise jump process $\mathscr{P}$ that satisfies the sum-product law of iterated lower expectations, we can use a backward recursive scheme to compute $\underline{E}_{\mathscr{P}}(f(X_v) \mid X_u = x_u)$ for any variable $f(X_v)$ with sum-product representation over $v \succ u$. T'Joens et al. (2019, Theorem 2) prove a related result in the setting of imprecise Markov chains (see also De Bock et al., 2021, Theorem 1), and their recursive algorithm is inspired by earlier work by De Cooman et al. (2010) and De Bock (2015, Chapter 7) on inference in credal networks under epistemic irrelevance (De Bock, 2017a). In order not to burden the main text, we have relegated our proof for this result to Appendix 4.A$_{198}$.

**Proposition 4.8.** *Consider a Markovian imprecise jump process $\mathscr{P}$ that satisfies the sum-product law of iterated lower expectations. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, a sequence of time-points $v = (t_1, \dots, t_n)$ in $\mathscr{U}_{\succ u}$, and an $\mathscr{F}_u$-simple variable $f$ with sum-product representation*

$$f = \sum_{k=1}^{n} g_k\big(X_{t_k}\big) \prod_{\ell=1}^{k-1} h_\ell\big(X_{t_\ell}\big)$$

*over $v$. Then*

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}(f_1(X_{t_1}) \mid X_u = x_u),$$

*where $f_1 \colon \mathscr{X} \to \mathbb{R}$ is recursively defined by the initial condition $f_n := g_n$ and, for all $k$ in $\{1, \dots, n-1\}$, by the recursive relation*

$$f_k \colon \mathscr{X} \to \mathbb{R} \colon x \mapsto \underline{E}_{\mathscr{P}}(f_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = x) h_k(x) + g_k(x).$$

It might not be immediately obvious, but Proposition 4.8 is extremely relevant from a practical point of view whenever $\mathscr{Q}$ has separately specified rows. In this case, both $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ are Markovian by Corollary 3.87$_{119}$, and they satisfy the sum-product law of iterated lower expectations by Theorem 3.89$_{120}$, so we may use Proposition 4.8. In combination with Propositions 3.81$_{117}$ to 3.83$_{118}$, we find a backwards recursive method to compute the (conditional) lower expectation with respect to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ of any real variable $f$ with sum-product representation. Quite remarkably, we can use the same method to compute the lower expectation of $f$ with respect to any imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$; the following result validates this claim, and also slightly relaxes the requirement that $v \succ u$. Our proof is rather long, so we have relegated it to Appendix 4.A$_{198}$,

**Theorem 4.9.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, and a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows. Fix an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$, a sequence of time-points $v = (t_1, \dots, t_n)$ in $\mathscr{U}_{\neq ()}$ and a real variable $f$ in $\mathbb{V}(\Omega)$ that has a sum-product representation*

$$f(X_v) = \sum_{k=1}^{n} g_k(X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell})$$

*over* $v$. *Let* $f_1\colon \mathscr{X} \to \mathbb{R}$ *be recursively defined by the initial condition* $f_n := g_n$ *and, for all* $k$ *in* $\{1, \dots, n-1\}$, *by the recursive relation*

$$f_k\colon \mathscr{X} \to \mathbb{R}\colon x \mapsto \big[e^{(t_{k+1}-t_k)\underline{Q}_{\mathscr{Q}}} f_{k+1}\big](x) h_k(x) + g_k(x).$$

*Then for all* $x$ *in* $\mathscr{X}$,

$$\underline{E}_{\mathscr{P}}(f \mid X_{t_1} = x) = f_1(x),$$

*and for all* $\{X_u = x_u\}$ *in* $\mathscr{H}$ *such that* $s := \max u \le t_1$,

$$\begin{aligned}\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) &= \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f \mid X_{t_1}) \,\big|\, X_u = x_u\big)\\ &= \begin{cases} f_1(x_s) & \text{if } u \neq () \text{ and } s = t_1\\ \big[e^{(t_1-s)\underline{Q}_{\mathscr{Q}}} f_1\big](x_s) & \text{if } u \neq () \text{ and } s < t_1,\\ \underline{E}_{\mathscr{M}}(f_1) & \text{if } u = () \text{ and } t_1 = 0,\\ \underline{E}_{\mathscr{M}}\big(e^{t_1\underline{Q}_{\mathscr{Q}}} f_1\big) & \text{if } u = () \text{ and } t_1 > 0,\end{cases}\end{aligned}$$

*and therefore*

$$\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{M},\mathscr{Q}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}(f \mid X_u = x_u).$$

Theorem 4.9 translates to the backwards recursive method in Algorithm 4.2. Let us use our running example to illustrate this method.

**Joseph's Example 4.10.** Let us revisit the situation in Joseph's Example 4.3161. Because $\mathscr{Q}_1$ clearly has separately specified rows, we can use Algorithm 4.2 to compute the lower expectation with respect to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}_1}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}_1}$ of variables that have a sum-product representation. Unfortunately, we cannot use Algorithm 4.2 to compute the lower probabilities $\underline{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}_1}(X_1 = X_2)$ and $\underline{P}_{\mathscr{M},\mathscr{Q}_1}(X_1 = X_2)$, because the $\mathscr{F}$-simple variable $\mathbb{I}_{\{X_1=X_2\}}$ does not have a sum-product representation (over $v = (1,2)$). A related $\mathscr{F}$-simple variable that does have a sum-product representation (over $v = (1,2)$) is

$$f(X_1, X_2) := \mathbb{I}_{\mathtt{H}}(X_1)\mathbb{I}_{\mathtt{H}}(X_2) = g_1(X_1) + h_1(X_1) g_2(X_2),$$

with $g_1 := 0$, $g_2 := \mathbb{I}_{\mathtt{H}}$ and $h_1 := \mathbb{I}_{\mathtt{H}}$. Thus, we can use Algorithm 4.2 to compute the lower probability

$$\underline{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}_1}(X_1 = \mathtt{H}, X_2 = \mathtt{H}) = \underline{P}_{\mathscr{M},\mathscr{Q}_1}(X_1 = \mathtt{H}, X_2 = \mathtt{H}) = \underline{E}_{\mathscr{M},\mathscr{Q}_1}(f(X_1, X_2)).$$

Here, we will do so for the numerical values in Joseph's Example 4.3161.

First, we need to determine the components of the gamble $f_1$; that is, for any $x$ in $\mathscr{X}$, we need to determine

$$f_1(x) = \big[e^{\underline{Q}} g_2(x)\big](x) h_1(x) + g_1(x).$$

**Algorithm 4.2:** Iteratively computing $\underline{E}_{\mathscr{P}}(f \mid X_u = x_u)$

**Input:** A non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows, an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$, a state history $\{X_u = x_u\}$ in $\mathscr{H}$, a sequence of time points $v = (t_1, \dots, t_n)$ in $\mathscr{U}_{\neq ()}$ such that $\max u \leq t_1$ and a variable

$$f = \sum_{k=1}^{n} g_k(X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell})$$

that has a sum-product representation over $v$.

**Output:** $\underline{E}_{\mathscr{P}}(f \mid X_u = x_u)$

$f_n \coloneqq g_n$
**for** $k \in \{n-1, \dots, 1\}$ **do** ▷ Determine $f_1$ as defined in Theorem 4.9$_{166}$
  $\Delta \coloneqq t_{k+1} - t_k$
  $f_k \colon \mathscr{X} \to \mathbb{R}$
  **for** $x \in \mathscr{X}$ **do**
    $f_k(x) \coloneqq \big[e^{\Delta \underline{Q}_{\mathscr{Q}}} f_{k+1}\big](x) h_k(x) + g_k(x)$
**if** $u \neq ()$ **and** $\max u = t_1$ **then**
  **return** $f_1(x_{\max u})$
**else if** $u \neq ()$ **and** $\max u < t_1$ **then**
  $\Delta \coloneqq t_1 - \max u$
  **return** $\big[e^{\Delta \underline{Q}_{\mathscr{Q}}} f_1\big](x_{\max u})$
**else if** $t_1 = 0$ **then**
  **return** $\underline{E}_{\mathscr{M}}(f_1)$
**else**
  $f_0 \colon \mathscr{X} \to \mathbb{R}$
  **for** $x \in \mathscr{X}$ **do**
    $f_0(x) \coloneqq \big[e^{t_1 \underline{Q}_{\mathscr{Q}}} f_1\big](x)$
  **return** $\underline{E}_{\mathscr{M}}(f_0)$

Observe that $g_2(\mathtt{H}) - g_2(\mathtt{T}) = 1$, so by Eq. (3.75)$_{115}$,

$$\begin{aligned} f_1(\mathtt{H}) &= \Big(g_2(\mathtt{H}) + \frac{1 - e^{-(\overline{\lambda} + \underline{\lambda})}}{\overline{\lambda} + \underline{\lambda}} \overline{\lambda}\big(g_2(\mathtt{T}) - g_2(\mathtt{H})\big)\Big) h_1(\mathtt{H}) + g_1(\mathtt{H}) \\ &= 1 - \big(1 - e^{-(\overline{\lambda} + \underline{\lambda})}\big) \frac{\overline{\lambda}}{\overline{\lambda} + \underline{\lambda}}, \end{aligned}$$

and

$$f_1(\mathtt{T}) = \Big(g_2(\mathtt{T}) + \frac{1-e^{-(\overline{\lambda}+\underline{\lambda})}}{\overline{\lambda}+\underline{\lambda}}\underline{\lambda}\big(g_2(\mathtt{H}) - g_2(\mathtt{T})\big)\Big)h_1(\mathtt{T}) + g_1(\mathtt{T})$$
$$= 0.$$

Because $\{X_u = x_u\} = \Omega$ and hence $u = ()$, we also need to determine the components of the gamble $f_0$ on $\mathscr{X}$, given for all $x$ in $\mathscr{X}$ by

$$f_0(x) = \big[e^{\underline{Q}} f_1\big](x).$$

Because $f_1(\mathtt{H}) - f_1(\mathtt{T}) \geq 0$, it follows from Eq. $(3.75)_{115}$ that

$$f_0(\mathtt{H}) = \left(1 - \big(1 - e^{-(\overline{\lambda}+\underline{\lambda})}\big)\frac{\overline{\lambda}}{\overline{\lambda}+\underline{\lambda}}\right)^2$$

and

$$f_0(\mathtt{T}) = \big(1 - e^{-(\overline{\lambda}+\underline{\lambda})}\big)\frac{\underline{\lambda}}{\overline{\lambda}+\underline{\lambda}}\left(1 - \big(1 - e^{-(\overline{\lambda}+\underline{\lambda})}\big)\frac{\overline{\lambda}}{\overline{\lambda}+\underline{\lambda}}\right).$$

Therefore,

$$\underline{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}_1}(X_1 = \mathtt{H}, X_2 = \mathtt{H}) = \underline{P}_{\mathscr{M},\mathscr{Q}_1}(X_1 = \mathtt{H}, X_2 = \mathtt{H})$$
$$= E_{p_0}(f_0) = f_0(\mathtt{H}) = \left(1 - \big(1 - e^{-(\overline{\lambda}+\underline{\lambda})}\big)\frac{\overline{\lambda}}{\overline{\lambda}+\underline{\lambda}}\right)^2$$
$$\approx 0.2018.$$

Let $Q$ be the rate operator in $\mathscr{Q}_1$ with $Q(\mathtt{H},\mathtt{T}) = \overline{\lambda}$ and $Q(\mathtt{T},\mathtt{H}) = \underline{\lambda}$. Then using the procedure in Joseph's Example $4.2_{158}$, we find that

$$P_{p_0,Q}(X_1 = \mathtt{H}, X_2 = \mathtt{H}) = \underline{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}_1}(X_1 = \mathtt{H}, X_2 = \mathtt{H}) = \underline{P}_{\mathscr{M},\mathscr{Q}_1}(X_1 = \mathtt{H}, X_2 = \mathtt{H}).$$

Therefore, $\underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_1}(X_1 = \mathtt{H}, X_2 = \mathtt{H}) \leq \underline{P}_{\mathscr{M},\mathscr{Q}_1}(X_1 = \mathtt{H}, X_2 = \mathtt{H})$. From this and Eq. $(3.68)_{105}$, we infer that

$$\underline{E}_{\mathscr{M},\mathscr{Q}_1}(f(X_1,X_2)) = \underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}_1}(f(X_1,X_2)) = \underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_1}(f(X_1,X_2)).$$

Quite remarkably, this is a non-trivial case in which the inequalities in Eq. $(3.68)_{105}$ are actually equalities!

In Joseph's Example $4.4_{163}$, we defined a set $\mathscr{Q}_2$ of rate operators that includes $\mathscr{Q}_1$. Note that this set $\mathscr{Q}_2$ has separately specified rows and that $\mathscr{Q}_1$ and $\mathscr{Q}_2$ have the same lower envelope $\underline{Q}$. Thus, it follows immediately from Algorithm $4.2_{\frown}$ that the lower probability of $\{X_1 = \mathtt{H}, X_2 = \mathtt{H}\}$ for $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}_2}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}_2}$ is equal to that for $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}_1}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}_1}$! ♨

### 4.1.4 A set $\mathscr{Q}$ of rate operators that is convex and has separately specified rows

If the simple variable $f(X_v)$ does not have a sum-product representation, then determining the lower expectation $\underline{E}_{\mathscr{P}}(f(X_v)\,|\,X_u = x_u)$ by means of backwards recursion is possible whenever $\mathscr{P}$ satisfies the law of iterated lower expectations. Krak et al. (2017, Section 9) essentially already make this point, and their argument goes back to similar results for imprecise Markov chains (De Cooman et al., 2008, 2009); the proof of this result has been relegated to Appendix $4.B_{204}$.

**Proposition 4.11.** *Consider an imprecise jump process $\mathscr{P}$ that satisfies the law of iterated lower expectations. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, a sequence of time points $v = (t_1, \ldots, t_n)$ in $\mathscr{U}_{>u}$ and a gamble $f$ on $\mathscr{X}_v$. Then*

$$\underline{E}_{\mathscr{P}}(f(X_v)\,|\,X_u = x_u) = \underline{E}_{\mathscr{P}}(f_1(X_{t_1})\,|\,X_u = x_u),$$

*where $f_1 \colon \mathscr{X} \to \mathbb{R}$ is defined recursively: the initial condition is $f_n := f$, and for all $k$ in $\{1, \ldots, n-1\}$, we let $t_{1:k} := (t_1, \ldots, t_k)$ and let $f_k$ be the gamble on $\mathscr{X}_{t_{1:k}}$ defined by*

$$f_k(y_{t_{1:k}}) := \underline{E}_{\mathscr{P}}(f_{k+1}(y_{t_{1:k}}, X_{t_{k+1}})\,|\,X_u = x_u, X_{t_{1:k}} = y_{t_{1:k}}) \quad \textit{for all } y_{t_{1:k}} \in \mathscr{X}_{t_{1:k}}.$$

Recall from Theorem $3.88_{120}$ that $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ satisfies the law of iterated lower expectations whenever $\mathscr{Q}$ is bounded and convex and has separately specified rows – but $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ or $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ may not. Thus, under these conditions, we may use Proposition 4.11 to iteratively determine $\underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_v)\,|\,X_u = x_u)$. Under the same conditions, we can also rely on Propositions $3.81_{117}$ to $3.83_{118}$ to reduce every step in the backward recursion to determining $e^{\Delta \underline{Q}_{\mathscr{Q}}} g$ or $\underline{E}_{\mathscr{M}}(g)$, and we then obtain the following backwards iterative method. For the proof of this result, see Appendix $4.B_{204}$.

**Theorem 4.12.** *Consider a non-empty set $\mathscr{M}$ of initial probability mass functions, and a non-empty, bounded and convex set $\mathscr{Q}$ of rate operators that has separately specified rows. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, a sequence of time points $v = (t_1, \ldots, t_n)$ in $\mathscr{U}_{>u}$ and a gamble $f$ on $\mathscr{X}_v$. Then*

$$\underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_v)\,|\,X_u = x_u) = \begin{cases} \left[e^{(t_1 - \max u)\underline{Q}_{\mathscr{Q}}} f_1\right](x_{\max u}) & \textit{if } u \neq (), \\ \underline{E}_{\mathscr{M}}(f_1) & \textit{if } u = () \textit{ and } t_1 = 0, \\ \underline{E}_{\mathscr{M}}\left(e^{(t_1 - \max u)\underline{Q}_{\mathscr{Q}}} f_1\right) & \textit{if } u = () \textit{ and } t_1 > 0, \end{cases}$$

*where $f_1 \colon \mathscr{X} \to \mathbb{R}$ is defined recursively: the initial condition is $f_n := f$, and for all $k$ in $\{1, \ldots, n-1\}$, we let $t_{1:k} := (t_1, \ldots, t_k)$ and let $f_k$ be the gamble on $\mathscr{X}_{t_{1:k}}$ defined by*

$$f_k(y_{t_{1:k}}) := \left[e^{(t_{k+1} - t_k)\underline{Q}_{\mathscr{Q}}} f_{k+1}(y_{t_{1:k}}, \bullet)\right](y_{t_k}) \quad \textit{for all } y_{t_{1:k}} \in \mathscr{X}_{t_{1:k}}.$$

Thus, whenever $\mathscr{Q}$ is bounded and convex and has separately specified rows, we can determine $\underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_v)\,|\,X_u = x_u)$ by means of the backwards recursive procedure of Algorithm 4.3. Note that Algorithm 4.1 [159] is the pre-

---

**Algorithm 4.3:** Iteratively computing $\underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_v)\,|\,X_u = x_u)$

---

**Input:** A non-empty set $\mathscr{M}$ of initial mass functions, a non-empty, bounded and convex set $\mathscr{Q}$ of rate operators that has separately specified rows, a state history $\{X_u = x_u\}$ in $\mathscr{H}$, a sequence of time points $v = (t_1, \dots, t_n)$ in $\mathscr{U}_{>u}$ and a gamble $f$ on $\mathscr{X}_v$.

**Output:** $\underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_v)\,|\,X_u = x_u)$

$f_n := f$
**for** $k \in \{n-1, \dots, 1\}$ **do** ▷ Determine $f_1$ as defined in Theorem 4.12 ↶
  $t_{1:k} := (t_1, \dots, t_k)$
  $\Delta := t_{k+1} - t_k$
  $f_k \colon \mathscr{X}_{t_{1:k}} \to \mathbb{R}$
  **for** $y_{t_{1:k}} \in \mathscr{X}_{t_{1:k}}$ **do**
    $f'_{k+1} := f_{k+1}(y_{t_{1:k}}, \bullet) \colon \mathscr{X} \to \mathbb{R} \colon z \mapsto f_{k+1}(y_{t_{1:k}}, z)$
    $f_k(y_{t_{1:k}}) := \big[e^{\Delta \underline{Q}_{\mathscr{Q}}} f'_{k+1}\big](y_{t_k})$
**if** $u \neq ()$ **then**
  $\Delta := t_1 - \max u$
  **return** $\big[e^{\Delta \underline{Q}_{\mathscr{Q}}} f_1\big](x_{\max u})$
**else if** $t_1 = 0$ **then**
  **return** $\underline{E}_{\mathscr{M}}(f_1)$
**else**
  $f_0 \colon \mathscr{X} \to \mathbb{R}$
  **for** $x \in \mathscr{X}$ **do**
    $f_0(x) := \big[e^{t_1 \underline{Q}_{\mathscr{Q}}} f_1\big](x)$
  **return** $\underline{E}_{\mathscr{M}}(f_0)$

---

cise specialisation of Algorithm 4.3, and that the similarities between Algorithm 4.2 [168] and Algorithm 4.3 follow from the similarities between Theorem 4.9 [166] and Theorem 4.12 ↶. As before, we will use our running example to make clear how Algorithm 4.3 works in practice.

**Joseph's Example 4.13.** Alice, Bruno, Cecilia, Deborah and Joseph are part of the in-crowd at their workplace. They form a tight-knit group, and frequently organise end-of-the-week celebrations on Friday afternoon. Amongst themselves, they refer to the co-workers that do not participate in these celebrations as 'the lonely people'. One of these lonely people is Eleanor, who stopped attending following one particularly joyful edition where, after telling him that she had once picked up the rice in a church where a wedding had been, she was picked on by Bruno. Despite not attending these Friday get-togethers any more, Eleanor still gets on quite well with Deborah. Seeing that Eleanor's contact with most other co-workers is essentially restricted to awkward conversations in the vicinity of the coffee machine, Deborah sees it as her duty to enlighten Eleanor with all the latest tittle-tattle. On one of these gossip-sharing occasions, Deborah tells Eleanor that Joseph has used one or more of the $n$ radioactive isotopes for his machine.

Eleanor believes that Joseph might have done something more intricate than using a single radioactive isotope, possibly using a mixture of isotopes or switching between isotopes. Thus, we take it that her beliefs are accurately modelled by $\underline{E}_{\mathscr{M},\mathscr{Q}_2}$, with $\mathscr{M} = \{\mathbb{I}_{\mathtt{H}}\}$ as in Joseph's Example 4.3$_{161}$ and $\mathscr{Q}_2$ as defined in Joseph's Example 4.4$_{163}$. Note that by construction, $\mathscr{Q}_2$ is convex and has separately specified rows. Like Deborah, Eleanor is interested in a (tight) lower bound on the probability that Josephs machine displays the same outcome one minute and two minutes after he switches it on. To illustrate the workings of Algorithm 4.3$_{\curvearrowleft}$, we compute this lower probability $\underline{P}_{\mathscr{M},\mathscr{Q}_2}(X_1 = X_2)$ for the numerical values of Joseph's Example 4.3$_{161}$.

Basically, we repeat the steps of Joseph's Example 4.2$_{158}$ but now using the lower rate operator $\underline{Q} := \underline{Q}_{\mathscr{Q}_2}$ instead of a rate operator $Q$. First, we need to determine the components of the gamble $g_1$; that is, for any $x$ in $\mathscr{X}$, we need to determine

$$f_1(x) = \left[e^{\underline{Q}} f(x, \bullet)\right](x).$$

Observe that $f(\mathtt{H},\mathtt{T}) - f(\mathtt{H},\mathtt{H}) = -1$ and $f(\mathtt{H},\mathtt{T}) - f(\mathtt{T},\mathtt{T}) = -1$, so by Eq. (3.75)$_{115}$,

$$\begin{aligned} f_1(\mathtt{H}) &= f(\mathtt{H},\mathtt{H}) + \frac{1 - e^{-(\overline{\lambda}+\underline{\lambda})}}{\overline{\lambda} + \underline{\lambda}} \overline{\lambda}\big(f(\mathtt{H},\mathtt{T}) - f(\mathtt{H},\mathtt{H})\big) \\ &= 1 - \left(1 - e^{-(\overline{\lambda}+\underline{\lambda})}\right) \frac{\overline{\lambda}}{\overline{\lambda} + \underline{\lambda}}, \end{aligned}$$

and similarly,

$$f_1(\mathtt{T}) = 1 - \left(1 - e^{-(\underline{\lambda}+\overline{\lambda})}\right) \frac{\overline{\lambda}}{\underline{\lambda} + \overline{\lambda}}.$$

Because $\{X_u = x_u\} = \Omega$ and hence $u = ()$, we also need to determine the components of the gamble $f_0$ on $\mathscr{X}$, given for all $x$ in $\mathscr{X}$ by

$$f_0(x) = \left[e^{\underline{Q}} f_1\right](x).$$

Because $f_1(\mathtt{H}) = f_1(\mathtt{T})$, it follows from Eq. (3.75)$_{115}$ that

$$f_0(\mathtt{H}) = f_0(\mathtt{T}) = f_1(\mathtt{H}) = 1 - \left(1 - e^{-(\overline{\lambda}+\underline{\lambda})}\right) \frac{\overline{\lambda}}{\overline{\lambda} + \underline{\lambda}}.$$

Therefore,

$$\underline{P}_{\mathscr{M},\mathscr{Q}_2}(X_1 = X_2) = E_{p_0}(f_0) = f_0(\mathtt{H}) = 1 - \left(1 - e^{-(\overline{\lambda}+\underline{\lambda})}\right) \frac{\overline{\lambda}}{\overline{\lambda} + \underline{\lambda}} \approx 0.4493.$$

Recall from Joseph's Example 4.4₁₆₃ that $\underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_2}(X_1 = X_2) \approx 0.5249$, so

$$\begin{aligned}\underline{P}_{\mathscr{M},\mathscr{Q}_2}(X_1 = X_2) &= \underline{E}_{\mathscr{M},\mathscr{Q}_2}(f(X_1, X_2)) \\ &< \underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_2}(f(X_1, X_2)) = \underline{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}_2}(X_1 = X_2).\end{aligned}$$

Thus, in this case one of the inequalities in Eq. (3.68)₁₀₅ is strict.

### 4.1.5 A set $\mathscr{Q}$ of rate operators that does not have separately specified rows or is not convex

In practice, it may happen that $\mathscr{Q}$ is not convex or does not have separately specified rows, or we might want to determine $\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ instead of $\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ or $\underline{E}_{\mathscr{M},\mathscr{Q}}$. This is not a problem though, because even if $\mathscr{Q}$ does not have separately specified rows (or is not convex), we can always compute a conservative bound on the (conditional) lower expectation with respect to $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. If $\mathscr{Q}$ is bounded, then it follows from Proposition 3.65₁₁₀ that $\underline{Q} := \underline{Q}_{\mathscr{Q}}$ is a lower rate operator; consequently, it follows from Lemma 3.69₁₁₁ that the set $\mathscr{Q}' := \mathscr{Q}_{\underline{Q}}$ of rate operators that dominate $\underline{Q}$ has separately specified rows and is convex. Hence, we can use Algorithms 4.2₁₆₈ and 4.3₁₇₁ to compute the lower expectation with respect to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}'}$ and/or $\mathbb{P}_{\mathscr{M},\mathscr{Q}'}$. Because $\mathscr{Q}'$ contains $\mathscr{Q}$, the corresponding lower expectation $\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}'}$ is more conservative than – or alternatively, provides a lower bound for – the lower expectations $\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ and $\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$; similarly, $\underline{E}_{\mathscr{M},\mathscr{Q}'}$ is more conservative than $\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\underline{E}_{\mathscr{M},\mathscr{Q}}$. The sets $\mathscr{Q}$ and $\mathscr{Q}'$ furthermore have the same lower envelope due to Lemma 3.66₁₁₀, so it is as if you simply apply the algorithms for $\underline{Q} := \underline{Q}_{\mathscr{Q}}$, even if $\mathscr{Q}$ does not satisfy the required properties.

What *is* essential though, is whether we can compute $[e^{\Delta \underline{Q}} f](x)$. The reader will probably agree that it is extremely difficult – if not impossible – to obtain an analytical expression for $[e^{\Delta \underline{Q}} f](x)$ in case the state space $\mathscr{X}$ consists of more than two states. Hence, it is vital that we have efficient numerical methods for accurately *computing* $[e^{\Delta \underline{Q}} f](x)$.

## 4.2 Computing the exponential of a lower rate operator

Consider a lower rate operator $\underline{Q}$, and let $f$ be a gamble on $\mathscr{X}$. Recall from Proposition 3.78₁₁₅ that

$$\mathbb{R}_{\geq 0} \to \mathbb{G}(\mathscr{X}) \colon t \mapsto e^{t\underline{Q}} f$$

is the unique solution of the initial value problem

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t = \underline{Q} f_t \quad \text{with } f_0 = f, \tag{4.1}$$

where for $t = 0$ we only impose the derivative from the right. Thus, one way to compute $\left[e^{t\underline{Q}} f\right](x)$ is to numerically integrate this non-linear ordinary differential equation.

Numerical methods to approximate the solution of ordinary differential equations are available in abundance. It is certainly outside the scope of this dissertation to give an overview of all methods; for a thorough yet introductory-level overview, we refer to (Hairer et al., 2008; Iserles, 2009). To understand the remainder, it suffices to understand that every numerical integration method boils down to the iterative procedure of Algorithm 4.4.

**Algorithm 4.4:** Iteratively approximate $e^{t\underline{Q}} f$

**Input:** A lower rate operator $\underline{Q}$, a time point $t$ in $\mathbb{R}_{\geq 0}$ and a gamble $f$ on $\mathscr{X}$.

**Output:** An approximation $\tilde{f}_n$ of $e^{t\underline{Q}} f$

```
f̃_0 := f; t_0 := 0; k := 0
while t_k < t do
    Pick a time step Δ_k in ]0, t − t_k].
    Compute f̃_{k+1} using Q, f̃_k and Δ_k.
    t_{k+1} := t_k + Δ_k; k := k + 1
n := k
return f̃_n
```

A numerical integration method consists of two parts: a mechanism to choose the next time step $\Delta_k$, and a mechanism to compute the approximation $\tilde{f}_{k+1}$ of $e^{t_{k+1}\underline{Q}} f$ using the previous approximation $\tilde{f}_k$, the time step $\Delta_k$ and the lower rate operator $\underline{Q}$. For most methods, there is a conservative (theoretical) upper bound on the error $\|e^{t\underline{Q}} f - \tilde{f}_n\|$ that is made by the approximation. However, in the words of Iserles (2009, p. 7), 'in an overwhelming majority of practical cases [this upper bound] is too large by many orders of magnitude', meaning that it is usually much too conservative; for this reason, he says that such an upper bound 'must not be used in practical estimations of numerical error!' In fact, most numerical integration methods include a mechanism to numerically estimate the error. Even more, this estimate of the error is used to determine the step size in such a way that in the end, the (estimated) error is lower than some desired maximal error $\epsilon$ – see (Hairer et al., 2008, Section II.3) or (Iserles, 2009, Chapter 6) for the general principle.

Going against this tradition, both Škulj (2015, Section 4) and Krak et al. (2017, Algorithm 1) propose approximation methods that use a (theoretical) upper bound on the error $\|e^{t\underline{Q}} f - \tilde{f}_n\|$ to determine the step size. Škulj (2015, Section 4) proposes three methods: one with a fixed step size, a second with an adaptive step size and a third that is a combination of the previous two.

He proposes to compute the approximation $\tilde{f}_{k+1}$ as

$$\tilde{f}_{k+1} = e^{\Delta_k Q_k} \tilde{f}_k \approx e^{\Delta_k \underline{Q}} e^{t_k \underline{Q}} f = e^{(t_k + \Delta_k)\underline{Q}} f = e^{t_{k+1}\underline{Q}} f,$$

where $Q_k$ is a transition rate operator in $\mathcal{Q}_{\underline{Q}}$ such that $\underline{Q}\tilde{f}_k = Q_k \tilde{f}_k$. One drawback of this method is that it needs the operator exponential $e^{\Delta_k Q}$, which – in general – has to be numerically approximated as well. Škulj (2015) mentions that his methods turn out to be quite computationally heavy, even if the uniform and adaptive methods are combined.

Krak et al. (2017, Algorithm 1) propose an alternative method that uses $n$ steps with a fixed step size $\Delta := t/n$. They put forward a method to choose the number of steps $n$ such that the error $\|e^{t\underline{Q}} f - \tilde{f}_n\|$ is guaranteed to be lower than the desired maximal error $\epsilon$, and they compute the approximation as

$$\tilde{f}_{k+1} = (I + \Delta \underline{Q})\tilde{f}_k = \tilde{f}_k + \Delta \underline{Q} \tilde{f}_k. \tag{4.2}$$

This, of course, is *Euler's method*, which is perhaps the most basic and well-known numerical integration method there is – see (Hairer et al., 2008, Section I.7) or (Iserles, 2009, Section 1.2). Crucially, Krak et al. (2017) guarantee that the approximation error $\|e^{t\underline{Q}} f - \tilde{f}_n\|$ is lower than or equal to $\epsilon$, and argue that their method is computationally more efficient than Škulj's (2015) for two reasons. The first one is that when comparing both methods with a fixed step size, theirs requires fewer iterations – or, equivalently, can take larger steps. Second, in every step of the iteration, they only need to compute $(I + \Delta_k \underline{Q})\tilde{f}_k$, whereas Škulj's (2015) method requires one to first determine $Q_k$ and subsequently compute $e^{\Delta_k Q_k} \tilde{f}_k$, which has a much larger computational footprint.

Most numerical integration methods require repeated evaluation of $\underline{Q} f$, so it is absolutely critical that we can actually do this. As we have seen in Section 3.3.3$_{109}$, a lower rate operator is typically defined as the lower envelope of a non-empty and bounded set $\mathcal{Q}$ of rate operators, preferably one that is convex and has separately specified rows. Usually, such a set $\mathcal{Q}$ is defined through linear inequalities; examples can be found in Power Network Example 6.45$_{313}$ in Chapter 6$_{273}$ or in Section 8.3.3$_{428}$ in Chapter 8$_{403}$ further on. Evaluating its lower envelope $\underline{Q}_{\mathcal{Q}}$ is then not an issue, because this amounts to solving a simple constrained linear optimisation problem, for example with the help of linear programming methods. From here on, we will therefore assume that evaluating the lower rate operator $\underline{Q}_{\mathcal{Q}}$ is tractable – that is, that the optimisation problem in Eq. (3.71)$_{109}$ can be solved efficiently.

In the remainder of this section, we improve on the approximation method suggested by Krak et al. (2017, Algorithm 1) in two ways. First, we obtain tighter bounds for the approximation error in Section 4.2.1$_{\frown}$. Second, we use these tighter bounds to choose a fixed step size for Euler's method in Section 4.2.2$_{181}$, and propose a mechanism to increase the step size on the fly in Section 4.2.3$_{184}$.

### 4.2.1 Bounding the approximation error

Consider a lower rate operator $\underline{Q}$, a time point $t$ in $\mathbb{R}_{\geq 0}$ and a gamble $f$ on $\mathscr{X}$. Suppose that we approximate $e^{t\underline{Q}}f$ by means of Euler's method, taking $n$ steps of size $\Delta_1, \dots, \Delta_n$ in $\mathbb{R}_{>0}$ such that $t = \sum_{k=1}^{n} \Delta_n$. This means that we iteratively compute

$$\tilde{f}_1 = (I + \Delta_1 \underline{Q}) f, \quad \tilde{f}_2 = (I + \Delta_2 \underline{Q}) \tilde{f}_1, \quad \dots, \quad \tilde{f}_n = (I + \Delta_n \underline{Q}) \tilde{f}_{n-1}.$$

In order to be able to choose these step sizes in an informed manner later on, we determine a theoretical upper bound on the approximation error $\|e^{t\underline{Q}}f - \tilde{f}_n\|$.

#### Approximating the semi-group

It will be instructive to first look at approximations of the semi-group itself. For any sequence $\underline{T}_1, \dots, \underline{T}_n$ of non-negatively homogeneous operators, we define the non-negatively homogeneous operator

$$\prod_{k=1}^{n} \underline{T}_k \coloneqq \underline{T}_n \cdots \underline{T}_1,$$

where we choose to interpret $\prod_{k=1}^{n} \underline{T}_k$ as first applying $\underline{T}_1$, then $\underline{T}_2$, and so on; this interpretation is a bit unconventional and might feel counter-intuitive, but it will tremendously simplify our notation in the remainder of this section.

Instead of looking at the approximation error

$$\left\| e^{t\underline{Q}} f - \tilde{f}_n \right\| = \left\| e^{t\underline{Q}} f - \prod_{k=1}^{n} (I + \Delta_k \underline{Q}) f \right\|,$$

for some given $f$, we will look at

$$\left\| e^{t\underline{Q}} - \prod_{k=1}^{n} (I + \Delta_k \underline{Q}) \right\|_{\mathrm{op}} = \left\| \prod_{k=1}^{n} e^{\Delta_k \underline{Q}} - \prod_{k=1}^{n} (I + \Delta_k \underline{Q}) \right\|_{\mathrm{op}}, \tag{4.3}$$

where the equality follows from the semi-group property (SG2)$_{78}$. Essential to this will be the following result, taken from (Krak et al., 2017, Lemma E.4).

**Lemma 4.14.** *Let $\underline{T}_1, \dots, \underline{T}_n$ and $\underline{S}_1, \dots, \underline{S}_n$ be two finite sequences of lower transition operators. Then*

$$\left\| \prod_{k=1}^{n} \underline{T}_k - \prod_{k=1}^{n} \underline{S}_k \right\|_{\mathrm{op}} \leq \sum_{k=1}^{n} \| \underline{T}_k - \underline{S}_k \|_{\mathrm{op}}.$$

We use this result to prove a stronger version of Lemma E.5 in (Krak et al., 2017).

**Lemma 4.15.** *Consider a lower rate operator $\underline{Q}$, a natural number $n$ and for every $k$ in $\{1,\dots,n\}$, a non-negative real number $\Delta_k$ such that $\Delta_k\|\underline{Q}\|_{\mathrm{op}} \leq 2$. If we let $\Delta := \sum_{k=1}^{n} \Delta_k$, then*

$$\left\| (I + \Delta\underline{Q}) - \prod_{k=1}^{n} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}} \leq \frac{1}{2}\Delta^2 \|\underline{Q}\|_{\mathrm{op}}^2.$$

Our proof is essentially the same as that of Lemma E.5 in Krak et al. (2017), but we repeat it here for the sake of completeness.

*Proof.* The statement is trivially true in case $n = 1$, so we may assume without loss of generality that $n > 1$. Observe that

$$\begin{aligned}
&\left\| (I + \Delta\underline{Q}) - \prod_{k=1}^{n} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}} \\
&\quad = \left\| \left( I + \Big(\sum_{k=1}^{n-1} \Delta_k\Big)\underline{Q} \right) + \Delta_n\underline{Q} - \prod_{k=1}^{n-1} (I + \Delta_k\underline{Q}) - \Delta_n\underline{Q} \prod_{k=1}^{n-1} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}} \\
&\quad \leq \left\| \left( I + \Big(\sum_{k=1}^{n-1} \Delta_k\Big)\underline{Q} \right) - \prod_{k=1}^{n-1} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}} + \left\| \Delta_n\underline{Q} - \Delta_n\underline{Q} \prod_{k=1}^{n-1} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}}, \qquad (4.4)
\end{aligned}$$

where the inequality follows from (N2)$_{76}$. We take a closer look at the second term. It follows from (N1)$_{76}$ and (LR9)$_{113}$ that

$$\left\| \Delta_n\underline{Q} - \Delta_n\underline{Q} \prod_{k=1}^{n-1} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}} \leq \Delta_n \|\underline{Q}\|_{\mathrm{op}} \left\| I - \prod_{k=1}^{n-1} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}},$$

Observe that $I = I^{n-1}$ is a lower transition operator, as well as $(I + \Delta_k\underline{Q})$ for all $k$ in $\{1,\dots,n-1\}$ due to Lemma 3.72$_{112}$. Therefore, it follows from Lemma 4.14$_{\curvearrowleft}$ that

$$\left\| \Delta_n\underline{Q} - \Delta_n\underline{Q} \prod_{k=1}^{n-1} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}} \leq \Delta_n \|\underline{Q}\|_{\mathrm{op}} \sum_{k=1}^{n-1} \left\| I - (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}} = \Delta_n \|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=1}^{n-1} \Delta_k,$$

where for the final equality we have used (N1)$_{76}$. From the previous inequality and Eq. (4.4), we infer that

$$\begin{aligned}
&\left\| (I + \Delta\underline{Q}) - \prod_{k=1}^{n} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}} \\
&\qquad \leq \left\| \left( I + \Big(\sum_{k=1}^{n-1} \Delta_k\Big)\underline{Q} \right) - \prod_{k=1}^{n-1} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}} + \Delta_n \|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=1}^{n-1} \Delta_k.
\end{aligned}$$

If we apply the same trick $n-1$ more times – or, alternatively, use mathematical induction – we obtain that

$$\begin{aligned}
\left\| (I + \Delta\underline{Q}) - \prod_{k=1}^{n} (I + \Delta_k\underline{Q}) \right\|_{\mathrm{op}} &\leq \|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=2}^{n} \Delta_k \sum_{\ell=1}^{k-1} \Delta_\ell \\
&\leq \frac{1}{2} \left( \sum_{k=1}^{n} \Delta_k \right)^2 \|\underline{Q}\|_{\mathrm{op}}^2 = \frac{1}{2}\Delta^2 \|\underline{Q}\|_{\mathrm{op}}^2,
\end{aligned}$$

as required. □

Next, we use the preceding result to prove a stronger version of Lemma E.9 in (Krak et al., 2017).

**Lemma 4.16.** *Consider a lower rate operator $\underline{Q}$. Then for any $\Delta$ in $\mathbb{R}_{\geq 0}$,*

$$\|e^{\Delta\underline{Q}} - (I + \Delta\underline{Q})\|_{\mathrm{op}} \leq \frac{1}{2}\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2.$$

Again, our proof is essentially that of Lemma E.9 in (Krak et al., 2017), but we repeat it for the sake of completeness.

*Proof.* Fix any positive real number $\epsilon$. Due to Proposition $3.74_{114}$, there is a natural number $n$ such that $\Delta\|\underline{Q}\|_{\mathrm{op}} \leq 2n$ and

$$\left\|e^{\Delta\underline{Q}} - \left(I + \delta\underline{Q}\right)^n\right\|_{\mathrm{op}} \leq \epsilon,$$

with $\delta := \Delta/n$. It follows from this inequality and Lemma 4.15 that

$$\begin{aligned}\left\|e^{\Delta\underline{Q}} - (I + \Delta\underline{Q})\right\|_{\mathrm{op}} &= \left\|e^{\Delta\underline{Q}} - (I + \delta\underline{Q})^n + (I + \delta\underline{Q})^n - (I + \Delta\underline{Q})\right\|_{\mathrm{op}} \\ &\leq \left\|e^{\Delta\underline{Q}} - (I + \delta\underline{Q})^n\right\|_{\mathrm{op}} + \left\|(I + \delta\underline{Q})^n - (I + \Delta\underline{Q})\right\|_{\mathrm{op}} \\ &\leq \epsilon + \|(I + \delta\underline{Q})^n - (I + \Delta\underline{Q})\|_{\mathrm{op}} \leq \epsilon + \frac{1}{2}\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2.\end{aligned}$$

Because this inequality holds for any arbitrary positive real number $\epsilon$, we have proven the statement. □

It is Lemma 4.16 that we will need in the remainder, but it would be strange not to mention that it also implies the following result.

**Corollary 4.17.** *Consider a lower rate operator $\underline{Q}$ and a time point $t$ in $\mathbb{R}_{\geq 0}$. Then for any sequence $\Delta_1, \ldots, \Delta_n$ of non-negative real numbers such that $\sum_{k=1}^n \Delta_k = t$ and $\Delta_k\|\underline{Q}\|_{\mathrm{op}} \leq 2$ for all $k$ in $\{1, \ldots, n\}$,*

$$\left\|e^{t\underline{Q}} - \prod_{k=1}^n (I + \Delta_k\underline{Q})\right\|_{\mathrm{op}} \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=1}^n \Delta_k^2.$$

*Proof.* Follows almost immediately from Eq. $(4.3)_{176}$, Lemma $3.72_{112}$, Lemma $4.14_{176}$ and Lemma 4.16 □

### Two convenient semi-norms

Remember that our goal is to obtain a theoretical bound on the approximation error $\|e^{t\underline{Q}}f - \prod_{k=1}^{n}(I+\Delta_k\underline{Q})f\|$. To this end, we will use semi-norms, that is, non-negative real-valued functions on $\mathbb{G}(\mathscr{X})$ that satisfy (N1)$_{76}$ and (N2)$_{76}$ but not necessarily (N3)$_{76}$ (see Schechter, 1997, Section 22.2). Concretely, we need the *variation semi-norm* $\|\bullet\|_{\mathrm{v}}$ and the *centred semi-norm* $\|\bullet\|_{\mathrm{c}}$, defined for all $f$ in $\mathbb{G}(\mathscr{X})$ by

$$\|f\|_{\mathrm{v}} := \|f - \min f\| = \max f - \min f \tag{4.5}$$

and

$$\|f\|_{\mathrm{c}} := \left\| f - \frac{\max f + \min f}{2} \right\| = \frac{1}{2}(\max f - \min f) = \frac{1}{2}\|f\|_{\mathrm{v}}. \tag{4.6}$$

Verifying that $\|\bullet\|_{\mathrm{v}}$ and $\|\bullet\|_{\mathrm{c}}$ are semi-norms but not norms is straightforward, and it follows almost immediately from the definition above that

N6. $\|f+\mu\|_{\mathrm{v}} = \|f\|_{\mathrm{v}}$ and $\|f+\mu\|_{\mathrm{c}} = \|f\|_{\mathrm{c}}$ for all $f$ in $\mathbb{G}(\mathscr{X})$ and $\mu$ in $\mathbb{R}$.

More importantly, these semi-norms are relevant due to the following three properties.

**Lemma 4.18.** *For any lower transition operators $\underline{T}$, $\underline{T}_1$, $\underline{T}_2$, $\underline{S}_1$, $\underline{S}_2$ and any gamble $f$ on $\mathscr{X}$,*

LT12. $\|\underline{T}f\|_{\mathrm{v}} \leq \|f\|_{\mathrm{v}}$;

LT13. $\|\underline{T}f\|_{\mathrm{c}} \leq \|f\|_{\mathrm{c}}$;

LT14. $\|\underline{T}f - \underline{S}f\| \leq \|\underline{T} - \underline{S}\|_{\mathrm{op}}\|f\|_{\mathrm{c}}$;

LT15. $\|\underline{T}_1\underline{T}_2 f - \underline{S}_1\underline{S}_2 f\| \leq \|\underline{T}_2 f - \underline{S}_2 f\| + \|\underline{T}_1 - \underline{S}_1\|_{\mathrm{op}}\|\underline{S}_2 f\|_{\mathrm{c}}$.

*Proof.* (LT12) follows immediately from (LT4)$_{108}$, because

$$\|\underline{T}f\|_{\mathrm{v}} = \max \underline{T}f - \min \underline{T}f \leq \max f - \min f = \|f\|_{\mathrm{v}}.$$

The same argument proves (LT13), but this also follows immediately from (LT12) and Eq. (4.6).

Next, we prove (LT14). To that end, we let $\mu_f := (\max f + \min f)/2$ and observe that

$$\begin{aligned}\|\underline{T}f - \underline{S}f\| &= \|\underline{T}f - \mu_f - \underline{S}f + \mu_f\| = \|\underline{T}(f-\mu_f) - \underline{S}(f-\mu_f)\| \\ &= \|(\underline{T}-\underline{S})(f-\mu_f)\| \leq \|\underline{T}-\underline{S}\|_{\mathrm{op}}\|f-\mu_f\| = \|\underline{T}-\underline{S}\|_{\mathrm{op}}\|f\|_{\mathrm{c}},\end{aligned}$$

where the second equality follows from (LT5)$_{108}$, the inequality follows from (N4)$_{77}$ and the last equality follows from Eq. (4.6).

To prove (LT15), it suffices to observe that

$$\begin{aligned}\|\underline{T}_1\underline{T}_2 f - \underline{S}_1\underline{S}_2 f\| &= \|\underline{T}_1\underline{T}_2 f - \underline{T}_1\underline{S}_2 f + \underline{T}_1\underline{S}_2 f - \underline{S}_1\underline{S}_2 f\| \\ &\leq \|\underline{T}_1\underline{T}_2 f - \underline{T}_1\underline{S}_2 f\| + \|\underline{T}_1\underline{S}_2 f - \underline{S}_1\underline{S}_2 f\| \\ &\leq \|\underline{T}_2 f - \underline{S}_2 f\| + \|\underline{T}_1\underline{S}_2 f - \underline{S}_1\underline{S}_2 f\| \\ &\leq \|\underline{T}_2 f - \underline{S}_2 f\| + \|\underline{T}_1 - \underline{S}_1\|_{\mathrm{op}}\|\underline{S}_2 f\|_{\mathrm{c}},\end{aligned}$$

where the first inequality follows from (N2)$_{76}$, the second inequality follows from (LT8)$_{108}$ and the third inequality follows from (LT14). □

We use the centred semi-norm $\|\bullet\|_{c}$ in the following result, which establishes the theoretical bound on the approximation error that we were after.

**Lemma 4.19.** *Consider a lower rate operator $\underline{Q}$, a non-negative real number $t$ and a gamble $f$ on $\mathscr{X}$. Let $\Delta_0, \dots, \Delta_{n-1}$ be a sequence of non-negative real numbers such that $\sum_{k=0}^{n-1}\Delta_k = t$ and, for all $k$ in $\{0,\dots,n-1\}$, $\Delta_k\|\underline{Q}\|_{\mathrm{op}} \leq 2$. Then*

$$\left\| e^{t\underline{Q}}f - \prod_{k=0}^{n-1}(I+\Delta_k\underline{Q})f \right\| \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=0}^{n-1}\Delta_k^2\|\tilde{f}_k\|_{c},$$

*with $\tilde{f}_0 := f$ and, for every $k$ in $\{0,\dots,n-1\}$, $\tilde{f}_{k+1} := (I+\Delta_k\underline{Q})\tilde{f}_k$.*

*Proof.* Our proof is one by mathematical induction. Thus, first we verify that the statement is true in case $n=1$. To that end, we observe that $t=\Delta_0$ and

$$\left\| e^{t\underline{Q}}f - (I+\Delta_0\underline{Q})f \right\| \leq \left\| e^{\Delta_0\underline{Q}} - (I+\Delta_0\underline{Q}) \right\|_{\mathrm{op}} \|f\|_{c} \leq \frac{1}{2}\Delta_0^2\|\underline{Q}\|_{\mathrm{op}}^2\|f\|_{c},$$

where the first inequality follows from (LT14) and the second from Lemma 4.16$_{178}$. It is now a matter of straightforward verification that this agrees with the inequality of the statement.

For the inductive step, we assume that the statement holds for some natural number $n$, and verify that the statement then also holds for $n+1$. Let $t_n := \sum_{k=0}^{n-1}\Delta_k$, and observe that

$$\begin{aligned}
&\left\| e^{t\underline{Q}}f - \prod_{k=0}^{n}(I+\Delta_k\underline{Q})f \right\| \\
&\quad= \left\| e^{\Delta_n\underline{Q}}e^{t_n\underline{Q}}f - (I+\Delta_n\underline{Q})\prod_{k=0}^{n-1}(I+\Delta_k\underline{Q})f \right\| \\
&\quad\leq \left\| e^{\Delta_n\underline{Q}}e^{t_n\underline{Q}}f - e^{\Delta_n\underline{Q}}\prod_{k=0}^{n-1}(I+\Delta_k\underline{Q})f \right\| + \left\| e^{\Delta_n\underline{Q}}\tilde{f}_n - (I+\Delta_n\underline{Q})\tilde{f}_n \right\| \\
&\quad\leq \left\| e^{t_n\underline{Q}}f - \prod_{k=0}^{n-1}(I+\Delta_k\underline{Q})f \right\| + \left\| e^{\Delta_n\underline{Q}}\tilde{f}_n - (I+\Delta_n\underline{Q})\tilde{f}_n \right\| \\
&\quad\leq \left\| e^{t_n\underline{Q}}f - \prod_{k=0}^{n-1}(I+\Delta_k\underline{Q})f \right\| + \left\| e^{\Delta_n\underline{Q}} - (I+\Delta_n\underline{Q}) \right\|_{\mathrm{op}}\|\tilde{f}_n\|_{c},
\end{aligned}$$

where for the first equality we have used Proposition 3.74$_{114}$ and (SG2)$_{78}$, the first inequality holds due to (N2)$_{76}$ and because $\tilde{f}_n = \prod_{k=0}^{n-1}(I+\Delta_k\underline{Q})f$, the second inequality follows from (LT8)$_{108}$ and the final inequality follows from (LT14). To verify the induction step, we simply use the induction hypothesis for the first term and

Lemma 4.16[178] for the second term:

$$\left\| e^{t\underline{Q}} f - \prod_{k=0}^{n} (I + \Delta_k \underline{Q}) f \right\| \leq \frac{1}{2} \|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=0}^{n-1} \Delta_k^2 \|\tilde{f}_k\|_{\mathrm{c}} + \left\| e^{\Delta_n \underline{Q}} - (I + \Delta_n \underline{Q}) \right\|_{\mathrm{op}} \|\tilde{f}_n\|_{\mathrm{c}}$$
$$\leq \frac{1}{2} \|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=0}^{n-1} \Delta_k^2 \|\tilde{f}_k\|_{\mathrm{c}} + \frac{1}{2} \Delta_n^2 \|\underline{Q}\|_{\mathrm{op}}^2 \|\tilde{f}_n\|_{\mathrm{c}}$$
$$= \frac{1}{2} \|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=0}^{n} \Delta_k^2 \|\tilde{f}_k\|_{\mathrm{c}}.$$

This is the inequality of the statement, so we have verified the induction step. □

### 4.2.2 The Euler method with a fixed step size

We can use Lemma 4.19[↶] to establish approximation methods that ensure that the approximation error is lower than the desired maximal error $\epsilon$ as follows. The most straightforward method is to use a fixed step size $\Delta_k = \Delta$ for every iteration. The following result establishes how to choose this fixed step size $\Delta$; it is a strengthened version of (Erreygers et al., 2017b, Lemma 23).

**Proposition 4.20.** *Consider a lower rate operator $\underline{Q}$, a non-negative real number $t$, a gamble $f$ on $\mathscr{X}$ and a desired maximal error $\epsilon$ in $\mathbb{R}_{>0}$. Fix some natural number $n$ such that*

$$n \geq \max\left\{ \frac{1}{2} t \|\underline{Q}\|_{\mathrm{op}}, \frac{1}{2\epsilon} t^2 \|\underline{Q}\|_{\mathrm{op}}^2 \|f\|_{\mathrm{c}} \right\},$$

*and let $\Delta := t/n$, $\tilde{f}_0 := f$ and, for any $k$ in $\{0, \dots, n-1\}$, $\tilde{f}_{k+1} := \tilde{f}_k + \Delta \underline{Q} \tilde{f}_k$. Then*

$$\left\| e^{t\underline{Q}} f - \tilde{f}_n \right\| = \left\| e^{t\underline{Q}} f - \left(I + \Delta \underline{Q}\right)^n f \right\| \leq \frac{1}{2} \Delta^2 \|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=0}^{n-1} \|\tilde{f}_k\|_{\mathrm{c}} \leq \epsilon.$$

*Proof.* Due to Lemma 3.72[112], the condition on $n$ guarantees that $(I + \Delta \underline{Q})^k$ is a lower transition operator. Thus, $\|\tilde{f}_k\|_{\mathrm{c}} \leq \|f\|_{\mathrm{c}}$ by repeated application of (LT13)[179]. From this and Lemma 4.19[↶], we infer that

$$\left\| e^{t\underline{Q}} f - \left(I + \Delta \underline{Q}\right)^n f \right\| \leq \frac{1}{2} \Delta^2 \|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=0}^{n-1} \|\tilde{f}_k\|_{\mathrm{c}} \leq \frac{1}{2} \frac{t^2}{n} \|\underline{Q}\|_{\mathrm{op}}^2 \|f\|_{\mathrm{c}} \leq \epsilon,$$

where the final inequality holds due to the condition on $n$. □

**Joseph's Example 4.21.** We again consider the lower rate operator $\underline{Q}$ that is defined as the lower envelope of $\mathscr{Q}_2$ in Joseph's Example 4.13[171]. We also use the numerical values of Joseph's Example 4.3[161], so for any $f$ in $\mathbb{G}(\mathscr{X})$,

$$\begin{pmatrix} f(\mathrm{H}) \\ f(\mathrm{T}) \end{pmatrix} \mapsto \underline{Q} f = \begin{pmatrix} \min\left\{ \lambda_{\mathrm{H}}(f(\mathrm{T}) - f(\mathrm{H})) \colon \lambda_{\mathrm{H}} \in \{\underline{\lambda}, \overline{\lambda}\} \right\} \\ \min\left\{ \lambda_{\mathrm{T}}(f(\mathrm{H}) - f(\mathrm{T})) \colon \lambda_{\mathrm{T}} \in \{\underline{\lambda}, \overline{\lambda}\} \right\} \end{pmatrix}, \tag{4.7}$$

with $\underline{\lambda} := 1$ and $\overline{\lambda} := 3/2$.

Using (LR7)$_{111}$, we find that $\|\underline{Q}\|_{\mathrm{op}} = 2\max\{\overline{\lambda}, \overline{\lambda}\} = 2\overline{\lambda} = 3$. Suppose we want to approximate $e^{t\underline{Q}}f$ with $t = 1$ and $f = \mathbb{I}_{\mathrm{H}}$ using an approximation of the form $(I + \frac{t}{n}\underline{Q})^n f$ up to a desired maximal error $\epsilon = 1 \cdot 10^{-3}$. By Proposition 4.20, we would need at least

$$\max\left\{\frac{1}{2}t\|\underline{Q}\|_{\mathrm{op}}, \frac{1}{2\epsilon}t^2\|\underline{Q}\|_{\mathrm{op}}^2\|f\|_{\mathrm{c}}\right\} = \max\left\{\frac{3}{4}, \frac{9}{4\cdot 10^{-3}}\right\} = 2\,250$$

iterations $n$.

Proposition 4.20 extends a similar result of Krak et al. (2017, Proposition 8.5) on two fronts. First, the lower bound

$$\max\left\{t\|\underline{Q}\|_{\mathrm{op}}, \frac{1}{\epsilon}t^2\|\underline{Q}\|_{\mathrm{op}}^2\|f\|_{\mathrm{c}}\right\}$$

of Krak et al. (2017, Proposition 8.5) is twice as large as ours – excluding the edge cases that $t = 0$ or $\|\underline{Q}\|_{\mathrm{op}} = 0$. Second, Proposition 4.20 establishes that we can compute an upper bound

$$\epsilon_{\mathrm{otf}} := \frac{1}{2}\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2\sum_{k=0}^{n-1}\|\tilde{f}_k\|_{\mathrm{c}}$$

on the approximation error 'on the fly' – meaning during the iterations – that is at least as good as and possibly better than the desired maximal error $\epsilon$. To prevent any unnecessary iterations, we identify three obvious cases where $e^{t\underline{Q}}f$ is equal to $f$.

**Corollary 4.22.** *Consider a lower rate operator $\underline{Q}$, a non-negative real number $t$ and a gamble $f$ on $\mathscr{X}$. Then $e^{t\underline{Q}}f = f$ whenever $\|\underline{Q}\|_{\mathrm{op}} = 0$ or $t = 0$ or $\|f\|_{\mathrm{c}} = 0$.*

*Proof.* By (N3)$_{76}$, $\|\underline{Q}\|_{\mathrm{op}} = 0$ if and only $\underline{Q} = 0$. Whenever this is the case, it follows from Proposition 3.78$_{115}$ that $e^{t\underline{Q}}f = f$. In case $t = 0$, it follows from Proposition 3.74$_{114}$ and (SG1)$_{77}$ that $e^{t\underline{Q}}f = If = f$. Finally, we observe that $\|f\|_{\mathrm{c}} = 0$ if and only if $\min f = \max f = f$. Whenever this is the case, it follows from Proposition 3.74$_{114}$ and (LT4)$_{108}$ that $\min f \leq e^{t\underline{Q}}f \leq \max f$; consequently, $e^{t\underline{Q}}f = f$. □

Using Proposition 4.20 and Corollary 4.22, we obtain the approximation method of Algorithm 4.5. The tighter error bound $\epsilon_{\mathrm{otf}}$ is computed on the fly in line 7, but this step is optional. The main reason for not executing this step is that, as we will see in the following example, it adds non-negligible computational overhead.

**Joseph's Example 4.23.** Following Joseph's Example 4.21, we use Algorithm 4.5 to approximate $e^{t\underline{Q}}f$ – with the lower rate operator $\underline{Q}$ as defined in Eq. (4.7) – with $t = 1$ and $f = \mathbb{I}_{\mathrm{H}}$ up to the desired maximal error $\epsilon = 1\cdot 10^{-3}$.

**Algorithm 4.5:** Euler's method with a fixed step size, guaranteed error bound and on-the-fly error bound

**Input:** A lower rate operator $\underline{Q}$, a time point $t$ in $\mathbb{R}_{\geq 0}$, a gamble $f$ on $\mathscr{X}$ and a desired maximal error $\epsilon$ in $\mathbb{R}_{>0}$.

**Output:** An approximation $\tilde{f}_n$ of $e^{t\underline{Q}}f$ with $\|e^{t\underline{Q}}f - \tilde{f}_n\| \leq \epsilon_{\mathrm{otf}} \leq \epsilon$.

1 $n := \left\lceil \max\{\frac{1}{2}t\|\underline{Q}\|_{\mathrm{op}}, \frac{1}{2\epsilon}t^2\|\underline{Q}\|_{\mathrm{op}}^2\|f\|_{\mathrm{c}}\}\right\rceil$
2 $\tilde{f}_0 := f$; $\epsilon_{\mathrm{otf}} := 0$
3 **if** $n > 0$ **and** $\|f\|_{\mathrm{c}} > 0$ **then** ▷ Note that $n = 0$ whenever $\|\underline{Q}\|_{\mathrm{op}} = 0$.
4     $\Delta := t/n$
5     **for** $k \in \{0, \dots, n-1\}$ **do**
6         $\tilde{f}_{k+1} := \tilde{f}_k + \Delta\underline{Q}\tilde{f}_k$
7         $\epsilon_{\mathrm{otf}} := \epsilon_{\mathrm{otf}} + \frac{1}{2}\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2\|\tilde{f}_k\|_{\mathrm{c}}$
8 **return** $\tilde{f}_n, \epsilon_{\mathrm{otf}}$

Executing Algorithm 4.5 results in $n = 2\,250$ direct Euler steps with step size $\Delta = 1/2\,250$, and we eventually obtain the approximation

$$e^{t\underline{Q}}f = e^{\underline{Q}}\mathbb{I}_{\mathtt{H}} \approx \tilde{f}_{2250} = \begin{pmatrix}0.4492\\0.3672\end{pmatrix} \quad \text{with} \quad \epsilon_{\mathrm{otf}} \approx 0.3672 \cdot 10^{-3}.$$

Note that our on-the-fly error bound $\epsilon_{\mathrm{otf}}$ is well below the desired maximal error $\epsilon = 1 \cdot 10^{-3}$. Recall from Joseph's Example $3.79_{115}$ that we have an analytical expression for $e^{\underline{Q}}\mathbb{I}_{\mathtt{H}}$; because $\mathbb{I}_{\mathtt{H}}(\mathtt{H}) > \mathbb{I}_{\mathtt{H}}(\mathtt{T})$,

$$e^{\underline{Q}}\mathbb{I}_{\mathtt{H}} = \begin{pmatrix}1\\0\end{pmatrix} + \frac{1 - e^{(\overline{\lambda}+\underline{\lambda})}}{\overline{\lambda}+\underline{\lambda}}\begin{pmatrix}-\overline{\lambda}\\\underline{\lambda}\end{pmatrix}.$$

Thus, in this case we can determine the actual error $\epsilon_{\mathrm{act}} := \|e^{\underline{Q}}\mathbb{I}_{\mathtt{H}} - \tilde{f}_{2250}\|$, which turns out to be $6.841 \cdot 10^{-5}$, up to four significant digits. Note that the (actual) numerical error $\epsilon_{\mathrm{act}}$ is significantly smaller than the on-the-fly error bound $\epsilon_{\mathrm{otf}}$, and more than an order of magnitude smaller than the desired maximal error $\epsilon$, as is to be expected from the discussion at the beginning of Section $4.2_{173}$. In fact, we find that using Euler's method with $n = 150$ steps of size $\Delta = 1/150$ already results in a numerical error that is approximately equal to the desired maximal error $\epsilon = 1 \cdot 10^{-3}$.

Computing the on-the-fly error bound $\epsilon_{\mathrm{otf}}$ comes at a cost, though. If we execute Joseph's Example $4.21_{181}$ without the step in line 7, our (naive and unoptimized) Python implementation takes about 29.73 ms to run (on average over 100 runs). However, in case we do compute the on-the-fly error, our Python implementation takes about 35.86 ms to run (on average over 100 runs), so about 20 % longer; see also Table $4.1_{186}$. ♘

### 4.2.3 The Euler method with an adaptive step size

Due to Joseph's Example 4.23[182], we are inched to believe that there is significant margin for improvement on Euler's method. To that end, we take a step back and suppose that we approximate $e^{t\underline{Q}}f$ using Euler's method by taking $n$ steps with – possibly different – step sizes $\Delta_0, \ldots, \Delta_{n-1}$ such that $t = \sum_{k=0}^{n-1} \Delta_k$ and $\Delta_k \|\underline{Q}\|_{\mathrm{op}} \leq 2$ for all $k$ in $\{0, \ldots, n-1\}$. That is, we iteratively compute

$$\tilde{f}_1 = (I + \Delta_0 \underline{Q}) f, \quad \tilde{f}_2 = (I + \Delta_1 \underline{Q}) \tilde{f}_1, \quad \ldots, \quad \tilde{f}_n = (I + \Delta_{n-1} \underline{Q}) \tilde{f}_{n-1}.$$

By Lemma 4.19[180], the approximation error then has a guaranteed upper bound:

$$\|e^{t\underline{Q}}f - \tilde{f}_n\| \leq \frac{1}{2} \|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=0}^{n-1} \Delta_k^2 \|\tilde{f}_k\|_{\mathrm{c}}.$$

This upper bound is a sum of the local errors of the iterations, where the local error of the $k$-th iteration corresponds to the term

$$\frac{1}{2} \|\underline{Q}\|_{\mathrm{op}}^2 \Delta_k^2 \|\tilde{f}_k\|_{\mathrm{c}}.$$

If we keep track of this local error for every iteration, we can use this information when choosing our step size. One of the more straightforward ways to do this, is to choose the step size $\Delta_k$ such that the local error is proportional to it. More formally, we choose $\Delta_k$ such that $\Delta_k \|\underline{Q}\|_{\mathrm{op}} \leq 2$ and

$$\frac{1}{2} \Delta_k^2 \|\underline{Q}\|_{\mathrm{op}}^2 \|\tilde{f}_k\|_{\mathrm{c}} \leq \epsilon \frac{\Delta_k}{t} \Leftrightarrow \Delta_k \leq \frac{2\epsilon}{t \|\underline{Q}\|_{\mathrm{op}}^2 \|\tilde{f}_k\|_{\mathrm{c}}}.$$

This way, the step size $\Delta_k$ of the $k$-th iteration is at least as large as the step size $\Delta_{k-1}$ of the previous iteration, because by (LT13)[179],

$$\begin{aligned} \|\tilde{f}_k\|_{\mathrm{c}} = \|(I + \Delta_{k-1}\underline{Q})\tilde{f}_{k-1}\|_{\mathrm{c}} \leq \|\tilde{f}_{k-1}\|_{\mathrm{c}} \leq \cdots \leq \|\tilde{f}_1\|_{\mathrm{c}} = \|(I + \Delta_0 \underline{Q})\tilde{f}_0\|_{\mathrm{c}} \\ \leq \|\tilde{f}_0\|_{\mathrm{c}} = \|f\|_{\mathrm{c}}. \end{aligned}$$

Recall from Eq. (4.6)[179] that $\|\tilde{f}_k\|_{\mathrm{c}} = (\max \tilde{f}_k - \min \tilde{f}_k)^{1/2}$, so this means that we need to determine $\max \tilde{f}_k$ and $\min \tilde{f}_k$ every iteration. Because this incurs some computational overhead, we might choose to only re-evaluate the step size every $m$ iterations instead of every iteration. Taking care to ensure that our step sizes sum up to $t$, we obtain the method of Algorithm 4.6[↷].

Note that we keep track of a tighter on-the-fly error bound $\epsilon_{\mathrm{otf}}$ in lines 10 and 15. In contrast to Algorithm 4.5[↶], this adds negligible computational overhead because we already need to compute $\|\tilde{f}_k\|_{\mathrm{c}}$ every $m$ iterations to re-evaluate the step size. This is really the sole reason to keep track of this error, because if this method works as intended and $m$ is not all too large, then we expect that $\epsilon_{\mathrm{otf}}$ is almost equal to $\epsilon$.

**Algorithm 4.6:** Euler's method with a variable step size, guaranteed error bound and on-the-fly error bound

**Input:** A lower rate operator $\underline{Q}$, a time point $t$ in $\mathbb{R}_{\geq 0}$, a gamble $f$ on $\mathscr{X}$, a maximal error $\epsilon$ in $\mathbb{R}_{>0}$ and a natural number $m$.

**Output:** An approximation $\tilde{f}_n$ of $e^{t\underline{Q}}f$ with $\|e^{t\underline{Q}}f - \tilde{f}_n\| \leq \epsilon_{\mathrm{otf}} \leq \epsilon$.

$t_{\mathrm{rem}} \coloneqq t$; $\tilde{f}_0 \coloneqq f$; $\epsilon_{\mathrm{otf}} \coloneqq 0$; $\ell \coloneqq m$; $k \coloneqq 0$; $\Delta_{\mathrm{fin}} \coloneqq 0$
**if** $\|\underline{Q}\|_{\mathrm{op}} > 0$ **then**
    **while** $t_{\mathrm{rem}} > 0$ **and** $\|\tilde{f}_k\|_{\mathrm{c}} > 0$ **do**
        $\Delta \coloneqq \min\{t_{\mathrm{rem}}, 2/\|\underline{Q}\|_{\mathrm{op}}, (2\epsilon)/(t\|\underline{Q}\|^2_{\mathrm{op}}\|\tilde{f}_k\|_{\mathrm{c}})\}$
        **if** $\ell\Delta \geq t_{\mathrm{rem}}$ **then** ▷ Check the termination condition
            $\ell \coloneqq \lfloor t_{\mathrm{rem}}/\Delta \rfloor$
            $\Delta_{\mathrm{fin}} \coloneqq t_{\mathrm{rem}} - \ell\Delta$
            $t_{\mathrm{rem}} \coloneqq 0$
        **else** $t_{\mathrm{rem}} \coloneqq t_{\mathrm{rem}} - \ell\Delta$
        $\epsilon_{\mathrm{otf}} \coloneqq \epsilon_{\mathrm{otf}} + \frac{1}{2}\ell\Delta^2\|\underline{Q}\|^2_{\mathrm{op}}\|\tilde{f}_k\|_{\mathrm{c}}$
        **for** $i \in \{0, \ldots, \ell - 1\}$ **do**
            $\tilde{f}_{k+1} \coloneqq \tilde{f}_k + \Delta\underline{Q}\tilde{f}_k$
            $k \coloneqq k + 1$
    **if** $\Delta_{\mathrm{fin}} > 0$ **then** ▷ The final step
        $\epsilon_{\mathrm{otf}} \coloneqq \epsilon_{\mathrm{otf}} + \frac{1}{2}\Delta^2_{\mathrm{fin}}\|\underline{Q}\|^2_{\mathrm{op}}\|\tilde{f}_{k-\ell}\|_{\mathrm{c}}$
        $\tilde{f}_{k+1} \coloneqq \tilde{f}_k + \Delta_{\mathrm{fin}}\underline{Q}\tilde{f}_k$
        $k \coloneqq k + 1$
$n \coloneqq k$
**return** $\tilde{f}_n, \epsilon_{\mathrm{otf}}$

**Joseph's Example 4.24.** Like in Joseph's Example 4.23$_{182}$, we use Algorithm 4.6 to obtain an approximation for $e^{\underline{Q}}\mathbb{I}_{\mathrm{H}}$. We do this for $m = 1$ and $m = 5$, and report our findings in Table 4.1↷. In this table, we see that with the adaptive step size we only need about 1/3 of the iterations that we need with the fixed step size, and that the on-the-fly error bound $\epsilon_{\mathrm{otf}}$ is approximately equal to the desired maximal error $\epsilon$; in other words, the algorithm works as intended.

However, fewer iterations do not necessarily imply a shorter runtime. Qualitatively, we can conclude the following from Table 4.1↷. First, keeping track of the on-the-fly error bound $\epsilon_{\mathrm{otf}}$ increases the duration, as expected. Second, Algorithm 4.6 is faster than Algorithm 4.5$_{183}$, at least if we choose $m$ large enough. And third, both methods yield an actual error that is significantly lower than the desired maximal error. ೨

The following result establishes what we have empirically observed in Joseph's Example 4.24: that Algorithm 4.6 works as intended, and in particular that the total number of iterations $n$ is at most equal to the number of

**Table 4.1** Comparison of the approximations in Joseph's Example 4.23~182~ and Joseph's Example 4.24.

$n$ denotes the number of iterations, $\epsilon_{\text{act}}$ and $\epsilon_{\text{otf}}$ denote the actual and on-the-fly error, respectively, and $D_{\text{rel}}$ is the relative runtime (averaged over 100 runs).

| Algorithm | $n$ | $\epsilon_{\text{act}}/\epsilon$ | $\epsilon_{\text{otf}}/\epsilon$ | $D_{\text{rel}}$ |
|---|---|---|---|---|
| 4.5~183~ | 2250 | 0.06841 | 0.3672 | 1.235 |
| 4.5~183~ (no $\epsilon_{\text{otf}}$) | 2250 | 0.06841 | / | 1 |
| 4.5~183~ ($n = 150$) | 150 | 1.027 | 5.518 | 0.0456 |
| 4.6 ($m = 1$) | 827 | 0.3052 | 0.9987 | 1.008 |
| 4.6 ($m = 5$) | 832 | 0.6183 | 0.9987 | 0.5080 |

iterations that is required by Algorithm 4.5~183~.

**Proposition 4.25.** *Consider a lower rate operator $\underline{Q}$, a non-negative real number $t$ and a gamble $f$ on $\mathscr{X}$. We fix some desired maximal error $\epsilon$ in $\mathbb{R}_{>0}$ and a parameter $m$ in $\mathbb{N}$, and use Algorithm 4.6 to determine $\tilde{f}_n$ and $\epsilon_{\text{otf}}$. Then*

$$\|e^{t\underline{Q}}f - \tilde{f}_n\| \leq \epsilon_{\text{otf}} \leq \epsilon$$

*and*

$$n \leq \left\lceil \max\left\{\frac{1}{2}t\|\underline{Q}\|_{\text{op}}, \frac{1}{2\epsilon}t^2\|\underline{Q}\|_{\text{op}}^2\|f\|_{\text{c}}\right\}\right\rceil.$$

*Proof.* In case $t = 0$, $\|\underline{Q}\|_{\text{op}} = 0$ or $\|f\|_{\text{c}} = 0$, the statement follows immediately from Corollary 4.22~182~. Therefore, we may assume without loss of generality that $t > 0$, $\|\underline{Q}\|_{\text{op}} > 0$ and $\|f\|_{\text{c}} > 0$; consequently, $n \geq 1$. To facilitate our proof, for every $k$ in $\{0, \ldots, n-1\}$, we denote the value of $\Delta$ that is used in the computation of $\tilde{f}_{k+1}$ on line 12 – or, in case $k = n-1$, possibly $\Delta_{\text{fin}}$ on line 16 – by $\Delta_k$.

Note that the condition of the while loop and the selection of the step sizes ensures that $\sum_{k=0}^{n-1}\Delta_k \leq t$. Set $t' := \sum_{k=0}^{n-1}\Delta_k$, and let $\delta := t - t'$. Note that $\delta \geq 0$, and that $\delta > 0$ can only occur if $\|\tilde{f}_n\|_{\text{c}} = 0$ – meaning that $\tilde{f}_n$ is a constant function. In this case,

$$\|e^{t\underline{Q}}f - \tilde{f}_n\| = \|e^{\delta\underline{Q}}e^{t'\underline{Q}}f - \tilde{f}_n\| = \|e^{\delta\underline{Q}}e^{t'\underline{Q}}f - e^{\delta\underline{Q}}\tilde{f}_n\| \leq \|e^{t'\underline{Q}}f - \tilde{f}_n\|, \tag{4.8}$$

where we have used Proposition 3.74~114~ and (SG2)~78~ for the first equality, (LT4)~108~ for the second equality and (LT8)~108~ for the inequality.

We denote the number of times that we run through the while loop by $w+1$, with $w$ being a non-negative integer. Due to line 12, for all $j$ in $\{0, \ldots, w-1\}$, the approximations of $\tilde{f}_{jm+1}, \ldots, \tilde{f}_{(j+1)m}$ all use the same step size $\delta_j := \Delta_{jm} = \Delta_{jm+1} = \cdots = \Delta_{(j+1)m-1}$. Additionally, it follows from line 5 and following that in the final run of the while loop, we determine the $n - k^\star$ approximations $\tilde{f}_{k^\star+1}, \ldots, \tilde{f}_n$, with

$k^\star := mw$. Note that by choice of $\Delta$ and $\Delta_{\text{fin}}$ in the last while loop, $0 < n - k^\star \leq m$. It follows from these observations and lines 10 and 15 that

$$\begin{aligned}\epsilon_{\text{otf}} &= \frac{1}{2} m \|\underline{Q}\|_{\text{op}}^2 \sum_{j=0}^{w-1} \delta_j^2 \|\tilde{f}_{jm}\|_{\text{c}} + \frac{1}{2} \|\underline{Q}\|_{\text{op}}^2 \|\tilde{f}_{k^\star}\|_{\text{c}} \sum_{k=k^\star}^{n-1} \Delta_k^2 \\ &= \frac{1}{2} \|\underline{Q}\|_{\text{op}}^2 \sum_{j=0}^{w-1} \|\tilde{f}_{jm}\|_{\text{c}} \sum_{i=0}^{m-1} \Delta_{jm+i}^2 + \frac{1}{2} \|\underline{Q}\|_{\text{op}}^2 \|\tilde{f}_{k^\star}\|_{\text{c}} \sum_{k=k^\star}^{n-1} \Delta_k^2. \end{aligned} \tag{4.9}$$

On lines 4 and 7, we see that for every $k$ in $\{0, \ldots, n-1\}$, the time step $\Delta_k$ is chosen in such a way that $\Delta_k \|\underline{Q}\|_{\text{op}} \leq 2$. Consequently, it follows from Lemma $3.72_{112}$ and repeated application of $\text{(LT13)}_{179}$ that

$$\|\tilde{f}_n\|_{\text{c}} = \|(I + \Delta_{n-1}\underline{Q})\tilde{f}_{n-1}\|_{\text{c}} \leq \|\tilde{f}_{n-1}\|_{\text{c}} \leq \cdots \leq \|\tilde{f}_0\|_{\text{c}} = \|f\|_{\text{c}}. \tag{4.10}$$

It follows immediately from these inequalities and Eq. (4.9) that

$$\epsilon_{\text{otf}} \geq \frac{1}{2} \|\underline{Q}\|_{\text{op}}^2 \sum_{k=0}^{n-1} \Delta_k^2 \|\tilde{f}_k\|_{\text{c}} \geq \|e^{t'\underline{Q}} f - \tilde{f}_n\|, \tag{4.11}$$

where the last inequality holds due to Lemma $4.19_{180}$. To establish the first part of the statement, we furthermore recall from line 4 that for every $j$ in $\{0, \ldots, w-1\}$ and $i$ in $\{0, \ldots, m-1\}$,

$$\Delta_{jm+i} \leq \frac{2\epsilon}{t \|\underline{Q}\|_{\text{op}}^2 \|\tilde{f}_{jm}\|_{\text{c}}};$$

and similarly, from lines 4 and 7 that for every $k$ in $\{k^\star, \ldots, n-1\}$,

$$\Delta_k \leq \frac{2\epsilon}{t \|\underline{Q}\|_{\text{op}}^2 \|\tilde{f}_{k^\star}\|_{\text{c}}}.$$

From the two preceding inequalities and Eq. (4.9), we infer that

$$\epsilon_{\text{otf}} \leq \sum_{j=0}^{w-1} \sum_{i=0}^{m-1} \Delta_{jm+i} \frac{\epsilon}{t} + \sum_{k=k^\star}^{n-1} \Delta_k \frac{\epsilon}{t} = \frac{\epsilon}{t} \sum_{k=0}^{n-1} \Delta_k = \epsilon \frac{t'}{t} \leq \epsilon. \tag{4.12}$$

Finally, it follows from Eq. (4.8)↶ – which is only needed in case $t' < t$ – and Eqs. (4.11) and (4.12) that

$$\|e^{t\underline{Q}} f - \tilde{f}_n\| \leq \|e^{t'\underline{Q}} f - \tilde{f}_n\| \leq \epsilon_{\text{otf}} \leq \epsilon,$$

which proves the first part of the statement.

To prove the second part of the statement, we let

$$\Delta^\star := \min\left\{\frac{2}{\|\underline{Q}\|_{\text{op}}}, \frac{2\epsilon}{t \|\underline{Q}\|_{\text{op}}^2 \|f\|_{\text{c}}}\right\}.$$

and observe that

$$n^\star := \left\lceil \max\left\{\frac{1}{2} t \|\underline{Q}\|_{\text{op}}, \frac{1}{2\epsilon} t^2 \|\underline{Q}\|_{\text{op}}^2 \|f\|_{\text{c}}\right\} \right\rceil = \left\lceil \frac{t}{\Delta^\star} \right\rceil.$$

Consequently, $n^\star \Delta^\star \geq t$.

In case $w \geq 1$, it follows from Eq. (4.10)↶ that

$$\frac{2\epsilon}{t\|\underline{Q}\|_{\mathrm{op}}^2\|\tilde{f}_{(w-1)m}\|_{\mathrm{c}}} \geq \frac{2\epsilon}{t\|\underline{Q}\|_{\mathrm{op}}^2\|\tilde{f}_{(w-2)m}\|_{\mathrm{c}}} \geq \cdots \geq \frac{2\epsilon}{t\|\underline{Q}\|_{\mathrm{op}}^2\|\tilde{f}_0\|_{\mathrm{c}}} = \frac{2\epsilon}{t\|\underline{Q}\|_{\mathrm{op}}^2\|f\|_{\mathrm{c}}}.$$

Seeing that in these cases we are not yet in the final run of the while loop, the upper bound $t_{\mathrm{rem}}$ for $\Delta$ in line 4 is never reached; therefore, the preceding inequalities imply that

$$\Delta_{k^\star-1} = \Delta_{mw-1} \geq \Delta_{mw-2} \geq \cdots \geq \Delta_{(w-1)m+i} \geq \cdots \geq \Delta_0 \geq \Delta^\star. \tag{4.13}$$

The final execution of the while loop requires some extra care. Let $t_{\mathrm{rem}} := t - \sum_{k=0}^{k^\star-1}\Delta_k$ and

$$\Delta := \min\left\{t_{\mathrm{rem}}, \frac{2}{\|\underline{Q}\|_{\mathrm{op}}}, \frac{2\epsilon}{t\|\underline{Q}\|_{\mathrm{op}}^2\|\tilde{f}_{k^\star}\|_{\mathrm{c}}}\right\}.$$

First, let us consider the case $\Delta < t_{\mathrm{rem}}$, Then because

$$\frac{2\epsilon}{t\|\underline{Q}\|_{\mathrm{op}}^2\|\tilde{f}_{k^\star}\|_{\mathrm{c}}} \geq \frac{2\epsilon}{t\|\underline{Q}\|_{\mathrm{op}}^2\|f\|_{\mathrm{c}}}$$

due to Eq. (4.10)↶, we have that

$$\Delta = \Delta_k \geq \Delta^\star \quad \text{for all } k \in \{k^\star, \ldots, n-2\}, \quad \text{and} \quad 0 < \Delta_{n-1} \leq \Delta.$$

For this reason,

$$\sum_{k=k^\star}^{n-1}\Delta_k > \sum_{k=k^\star}^{n-2}\Delta_k \geq (n-1-k^\star)\Delta^\star. \tag{4.14}$$

In the alternative case that $\Delta = t_{\mathrm{rem}} > 0$, it follows from Algorithm 4.6$_{185}$ that $n-1 = k^\star$, and we therefore obtain the same strict inequality:

$$\sum_{k=k^\star}^{n-1}\Delta_k = \Delta_{n-1} = \Delta = t_{\mathrm{rem}} > 0 = (n-1-k^\star)\Delta^\star. \tag{4.15}$$

In any case, we finally observe that

$$\begin{aligned} n^\star\Delta^\star \geq t \geq t' = \sum_{k=0}^{n-1}\Delta_k = \sum_{k=0}^{k^\star-1}\Delta_k + \sum_{k=k^\star}^{n-1}\Delta_k \geq k^\star\Delta^\star + \sum_{k=k^\star}^{n-1}\Delta_k \\ > k^\star\Delta^\star + (n-1-k^\star)\Delta^\star = (n-1)\Delta^\star, \end{aligned}$$

where we have used Eq. (4.13) for the second inequality and Eqs. (4.14) and (4.15) for the third inequality. Dividing both sides of this inequality by $\Delta^\star$, we find that $n^\star + 1 > n$; because $n^\star$ and $n$ are both natural numbers, this is equivalent to $n^\star \geq n$, which is precisely the second part of the statement. □

## 4.3 Ergodicity

Consider a lower rate operator $\underline{Q}$, and suppose we are approximating $e^{t\underline{Q}}f$ using Algorithm 4.6$_{185}$. Let us assume that we have done $k$ iterations, and have fixed the step size

$$\Delta := \min\left\{t_{\mathrm{rem}}, \frac{2}{\|\underline{Q}\|_{\mathrm{op}}}, \frac{2\epsilon}{t\|\underline{Q}\|_{\mathrm{op}}^2\|\tilde{f}_k\|_{\mathrm{c}}}\right\}$$

for the next $m$ iterations, with $\|\tilde{f}_k\|_c > 0$. In case $t_{\text{rem}}$ is sufficiently large, re-evaluating the step size after these $m$ iterations only yields a larger step size in case

$$\|\tilde{f}_{k+m}\|_c = \left\|(I + \Delta \underline{Q})^m \tilde{f}_k\right\|_c < \|\tilde{f}_k\|_c.$$

This inequality is always satisfied whenever

$$\big(\forall \Delta \in \mathbb{R}_{>0}, \Delta\|\underline{Q}\|_{\text{op}} \leq 2\big)\big(\forall g \in \mathbb{G}(\mathscr{X}), \|g\|_c > 0\big)\ \|(I + \Delta \underline{Q})^m g\|_c < \|g\|_c. \quad (4.16)$$

In fact, since the preceding inequality is invariant under translation or positive scaling of $g$ – use (N1)$_{76}$, (LT2)$_{107}$ and (LT5)$_{108}$ – it suffices if

$$\big(\forall \Delta \in \mathbb{R}_{>0}, \Delta\|\underline{Q}\|_{\text{op}} \leq 2\big)\big(\forall g \in \mathbb{G}(\mathscr{X}), 0 \leq g \leq 1\big)\ \|(I + \Delta \underline{Q})^m g\|_v < 1. \quad (4.17)$$

Readers that are familiar with (the ergodicity of) lower transition operators – see (Hermans et al., 2012) or (Škulj et al., 2013) – will probably recognise this condition, as it essentially states that the coefficient of ergodicity of the lower transition operator $(I + \Delta \overline{Q})^m$ is strictly smaller than 1. It is this link between our approximation methods and the coefficient of ergodicity – and ergodicity in general – that we set out to investigate in this section.

For this reason, we discuss the coefficient of ergodicity for lower transition operators, as well as the associated notion of ergodicity, in Section 4.3.1. In Section 4.3.2$_{191}$, we look at the related notion of ergodicity for lower rate operators, and furthermore examine the crucial connection between these two notions. Finally, we return to our methods to numerically approximate $e^{t\underline{Q}}f$ in Section 4.3.3$_{194}$, where we use ergodicity to obtain an alternative theoretical error bound.

### 4.3.1 Ergodicity of lower transition operators

Hermans et al. (2012, Definition 2) put in place the essential notion of ergodicity of lower transition operators as follows.

**Definition 4.26.** A lower transition operator $\underline{T}$ is *ergodic* if for all $f$ in $\mathbb{G}(\mathscr{X})$, $\lim_{n\to\infty} \underline{T}^n f$ exists and is a constant function.

The term ergodicity is well-known in the study of transition operators (or matrices), but there is no universally-adopted meaning of the term. To take away any possible confusion, we mention that for a transition operator $T$, our notion of ergodicity is equal to Tornambè's (1995, Definition 4.7) – and with Iosifescu's (1980, Section 2.6.2) notion of indecomposability.

The condition of Definition 4.26 is in a form that makes it hard to check, at least in general. Fortunately, Hermans et al. (2012, Proposition 3) establish the following necessary and sufficient condition.

**Proposition 4.27.** *The lower transition operator $\underline{T}$ is ergodic if and only if it is* regularly absorbing*, meaning that it is*

(i) top class regular, *in the sense that*

$$\mathscr{X}_{\underline{T}} := \left\{x \in \mathscr{X} : (\exists n \in \mathbb{N})(\forall y \in \mathscr{X})\ \left[\overline{T}^n \mathbb{I}_x\right](y) > 0\right\} \neq \emptyset; \tag{4.18}$$

(ii) top class absorbing, *in the sense that*

$$\left(\forall y \in \mathscr{X} \setminus \mathscr{X}_{\underline{T}}\right)\left(\exists n \in \mathbb{N}\right)\ \left[\underline{T}^n \mathbb{I}_{\mathscr{X}_{\underline{T}}}\right](y) > 0.$$

If $\underline{T}$ is ergodic, then $\mathscr{X}_{\underline{T}}$ as defined by Eq. (4.18) is called the *top class*; this terminology stems from the fact that $\mathscr{X}_{\underline{T}}$ is then the unique maximal communication class induced by the accessibility relation corresponding to $\underline{T}$ – see (Kemeny et al., 1960, Section 2.4) for transition operators and (De Cooman et al., 2009, Section 4.1) for lower transition operators.

At first sight, one might think that it takes a bit of work to check the two conditions in Proposition 4.27, but Hermans et al. (2012, Section 5) establish a convenient procedure that simplifies this process considerably. Since we do not really need this method here, we will not go into detail. That being said, we do use some of their intermediary results in the proof of Theorem $4.36_{194}$ further on, and it is for this reason that we repeat some of their intermediary results in Appendix $4.C.1_{206}$.

### The coefficient of ergodicity

Because a function $f$ on $\mathscr{X}$ is constant if and only if $\|f\|_{\mathrm{v}} = 0 = \|f\|_{\mathrm{c}}$, it is clear that if the lower transition operator $\underline{T}$ is ergodic, then

$$\lim_{n\to+\infty} \|\underline{T}^n f\|_{\mathrm{v}} = 0 \quad \text{for all } f \in \mathbb{G}(\mathscr{X}).$$

With the help of $(\mathrm{LT4})_{108}$, one can verify that this condition is not only necessary but also sufficient for ergodicity. For this reason, we take a closer look at $\|\underline{T}^n f\|_{\mathrm{v}}$ in general and $\|\underline{T} f\|_{\mathrm{v}}$ in particular.

For any lower transition operator $\underline{T}$, Škulj et al. (2013, Definition 1) define its *coefficient of ergodicity*

$$\rho(\underline{T}) := \max\left\{\|\underline{T} f\|_{\mathrm{v}} : f \in \mathbb{G}(\mathscr{X}), 0 \leq f \leq 1\right\}. \tag{4.19}$$

This coefficient has some interesting properties. For any two lower transition operators $\underline{T}$ and $\underline{S}$,

EC1. $0 \leq \rho(\underline{T}) \leq 1$;

EC2. $\rho(\underline{T}\,\underline{S}) \leq \rho(\underline{T})\rho(\underline{S})$;

EC3. $\|\underline{T} f\|_{\mathrm{v}} \leq \rho(\underline{T})\|f\|_{\mathrm{v}}$ for all $f$ in $\mathbb{G}(\mathscr{X})$.

*Proof.* The first property, (EC1), follows immediately from $(\mathrm{LT4})_{108}$. The second property, (EC2), follows almost immediately from (EC3) and $(\mathrm{LT4})_{108}$ (see also Škulj et al., 2013, Corollary 15). Thus, what remains for us is to prove (EC3).

First, we consider the case that $\|f\|_{\mathrm{v}} = 0$. Then $\|\underline{T}f\|_{\mathrm{v}} = 0$ by (LT4)$_{108}$, such that (EC3)$_{\curvearrowleft}$ holds. Second, we consider the case that $\|f\|_{\mathrm{v}} > 0$. Note that $0 \leq (f - \min f)^{1/\|f\|_{\mathrm{v}}} \leq 1$. Combining this with – in that order – (N6)$_{179}$, (LT5)$_{108}$, (LT2)$_{107}$, (N1)$_{76}$ and Eq. (4.19)$_{\curvearrowleft}$, we find that

$$\|\underline{T}f\|_{\mathrm{v}} = \|\underline{T}f - \min f\|_{\mathrm{v}} = \|\underline{T}(f - \min f)\|_{\mathrm{v}} = \left\| \|f\|_{\mathrm{v}} \underline{T}\left(\frac{f - \min f}{\|f\|_{\mathrm{v}}}\right) \right\|_{\mathrm{v}}$$
$$= \left\| \underline{T}\left(\frac{f - \min f}{\|f\|_{\mathrm{v}}}\right) \right\|_{\mathrm{v}} \|f\|_{\mathrm{v}} \leq \rho(\underline{T}) \|f\|_{\mathrm{v}},$$

as required □

Hermans et al. (2012, Proposition 7) establish a second necessary and sufficient condition for ergodicity, this time using the coefficient of ergodicity; a similar result appears in (Škulj et al., 2013, Theorem 21).

**Theorem 4.28.** *A lower transition operator $\underline{T}$ is ergodic if and only if there is some natural number $n$ such that $\rho(\underline{T}^n) < 1$.*

Let us return to Eq. (4.17)$_{189}$, which was our reason for looking into ergodicity. With our newly-introduced notation, it follows from (EC3)$_{\curvearrowleft}$ that Eq. (4.17)$_{189}$ holds whenever

$$\left(\forall \Delta \in \mathbb{R}_{>0}, \Delta\|\underline{Q}\|_{\mathrm{op}} \leq 2\right)\ \rho\left((I + \Delta\underline{Q})^m\right) < 1.$$

Due to Theorem 4.28, this condition implies that $(I + \Delta\underline{Q})$ is ergodic for all step sizes $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\|\underline{Q}\|_{\mathrm{op}} \leq 2$.

### 4.3.2 Ergodicity of lower rate operators

As will become apparent, whether or not $(I + \Delta\underline{Q})$ is ergodic is tightly connected with the behaviour of $e^{t\underline{Q}}f$ for large $t$. De Bock (2017b) was the first to study this limit behaviour, and we recall some of his findings here. For example, he shows that $e^{t\underline{Q}}$ converges to a lower transition operator as $t$ recedes to $+\infty$ (see De Bock, 2017b, Theorem 12). From this and (N4)$_{77}$, we infer that the limit $\lim_{t\to+\infty} e^{t\underline{Q}}f$ exists for all $f$ in $\mathbb{G}(\mathscr{X})$. Of course, this in turn implies that the limit $\lim_{t\to+\infty}[e^{t\underline{Q}}f](x)$ exists for every state $x$ in $\mathscr{X}$, but take note that these limit values might differ for different $x$. De Bock (2017b, Definition 2) calls the lower rate operator $\underline{Q}$ ergodic whenever the limits are equal for all states, or equivalently, whenever $e^{t\underline{Q}}f$ converges to a constant function.

**Definition 4.29.** The lower rate operator $\underline{Q}$ is *ergodic* if for all $f$ in $\mathbb{G}(\mathscr{X})$, $\lim_{t\to\infty} e^{t\underline{Q}}f$ is a constant function.

Note that Definition 4.29 is similar to Definition 4.26$_{189}$. Here as well, we should remark that different authors use the term ergodicity to refer to various concepts for rate operators. For a rate operator $Q$, the definition above is equal to Tornambè's 1995, Definition 4.17.

**Joseph's Example 4.30.** For the sake of generality, we again consider a general lower rate operator $\underline{Q}$ on the state space $\mathcal{X} = \{\mathtt{H}, \mathtt{T}\}$ which, as we have seen in Joseph's Example 3.64<sub>109</sub>, is of the form

$$\begin{pmatrix} f(\mathtt{H}) \\ f(\mathtt{T}) \end{pmatrix} \mapsto \underline{Q}f = \begin{pmatrix} \min\{\lambda_{\mathtt{H}}(f(\mathtt{T}) - f(\mathtt{H})) \colon \lambda_{\mathtt{H}} \in \{\underline{\lambda}_{\mathtt{H}}, \overline{\lambda}_{\mathtt{H}}\}\} \\ \min\{\lambda_{\mathtt{T}}(f(\mathtt{H}) - f(\mathtt{T})) \colon \lambda_{\mathtt{T}} \in \{\underline{\lambda}_{\mathtt{T}}, \overline{\lambda}_{\mathtt{T}}\}\} \end{pmatrix},$$

where $\underline{\lambda}_{\mathtt{H}}$, $\overline{\lambda}_{\mathtt{H}}$ ,$\underline{\lambda}_{\mathtt{T}}$ and $\overline{\lambda}_{\mathtt{T}}$ are non-negative real numbers such that $\underline{\lambda}_{\mathtt{H}} \leq \overline{\lambda}_{\mathtt{H}}$ and $\underline{\lambda}_{\mathtt{T}} \leq \overline{\lambda}_{\mathtt{T}}$. Furthermore, we recall from Eq. (3.75)<sub>115</sub> in Joseph's Example 3.79<sub>115</sub> that

$$e^{t\underline{Q}} f = f + \frac{1 - e^{-t\lambda_f}}{\lambda_f} \underline{Q} f \quad \text{for all } t \in \mathbb{R}_{\geq 0} \text{ and } f \in \mathbb{G}(\mathcal{X}),$$

where $\lambda_f \coloneqq \overline{\lambda}_{\mathtt{H}} + \underline{\lambda}_{\mathtt{T}}$ if $f(\mathtt{H}) \geq f(\mathtt{T})$ and $\lambda_f \coloneqq \underline{\lambda}_{\mathtt{H}} + \overline{\lambda}_{\mathtt{T}}$ if $f(\mathtt{H}) < f(\mathtt{T})$, and the second term is only added if $\lambda_f > 0$. With this expression, we can check if $\underline{Q}$ satisfies the condition of Definition 4.29.

Let us start with $f = \mathbb{I}_{\mathtt{H}}$. If $\lambda_{\mathbb{I}_{\mathtt{H}}} = \overline{\lambda}_{\mathtt{H}} + \underline{\lambda}_{\mathtt{T}} = 0$, $\lim_{t \to +\infty} e^{t\underline{Q}} \mathbb{I}_{\mathtt{H}} = \mathbb{I}_{\mathtt{H}}$. Because $\mathbb{I}_{\mathtt{H}}$ is not a constant function, we may conclude from this that $\underline{Q}$ is *not* ergodic if $\overline{\lambda}_{\mathtt{H}} + \underline{\lambda}_{\mathtt{T}} = 0$. Using a similar argument with $\mathbb{I}_{\mathtt{T}}$, one can verify that whenever $\underline{\lambda}_{\mathtt{H}} + \overline{\lambda}_{\mathtt{T}} = 0$, $\underline{Q}$ is not ergodic either.

Thus, we now assume that $\overline{\lambda}_{\mathtt{H}} + \underline{\lambda}_{\mathtt{T}} > 0$ and $\underline{\lambda}_{\mathtt{H}} + \overline{\lambda}_{\mathtt{T}} > 0$. Observe now that for any $f$ in $\mathbb{G}(\mathcal{X})$ such that $f(\mathtt{H}) \geq f(\mathtt{T})$,

$$\lim_{t \to +\infty} e^{t\underline{Q}} f = f + \frac{1}{\overline{\lambda}_{\mathtt{H}} + \underline{\lambda}_{\mathtt{T}}} \begin{pmatrix} \overline{\lambda}_{\mathtt{H}}\big(f(\mathtt{T}) - f(\mathtt{H})\big) \\ \underline{\lambda}_{\mathtt{T}}\big(f(\mathtt{H}) - f(\mathtt{T})\big) \end{pmatrix} = \frac{1}{\overline{\lambda}_{\mathtt{H}} + \underline{\lambda}_{\mathtt{T}}} \big(\underline{\lambda}_{\mathtt{T}} f(\mathtt{H}) + \overline{\lambda}_{\mathtt{H}} f(\mathtt{T})\big),$$

where for the first equality we have used that $\lim_{t \to +\infty} 1 - e^{-t(\overline{\lambda}_{\mathtt{H}} + \underline{\lambda}_{\mathtt{T}})} = 1$. Similarly, for any $f$ in $\mathbb{G}(\mathcal{X})$ such that $f(\mathtt{H}) < f(\mathtt{T})$,

$$\lim_{t \to +\infty} e^{t\underline{Q}} f = \frac{1}{\underline{\lambda}_{\mathtt{H}} + \overline{\lambda}_{\mathtt{T}}} \big(\underline{\lambda}_{\mathtt{H}} f(\mathtt{T}) + \overline{\lambda}_{\mathtt{T}} f(\mathtt{H})\big).$$

The two preceding equalities show that $e^{t\underline{Q}} f$ converges to a constant function for all $f$ in $\mathbb{G}(\mathcal{X})$. In conclusion, we have shown that $\underline{Q}$ is ergodic if and only if $\overline{\lambda}_{\mathtt{H}} + \underline{\lambda}_{\mathtt{T}} > 0$ and $\underline{\lambda}_{\mathtt{H}} + \overline{\lambda}_{\mathtt{T}} > 0$.

The preceding example illustrates that the condition in Definition 4.29 is in a form that is not the most convenient to check. Fortunately, De Bock (2017b) establishes a necessary and sufficient condition that is a bit more easy to grasp. The notions of upper and lower reachability allow for an elegant statement of this condition (see Krak et al., 2017, Definitions 7 and 8).

**Definition 4.31.** Consider a lower rate operator $\underline{Q}$. For any two states $x$ and $y$ in $\mathcal{X}$, we say that $x$ is *upper reachable* from $y$, and denote this by $y \hookrightarrow x$, if there is some sequence $(x_0, \ldots, x_n)$ in $\mathcal{X}$ with $n$ in $\mathbb{Z}_{\geq 0}$, $x_0 = y$ and $x_n = x$ such that $[\overline{Q}\mathbb{I}_{x_k}](x_{k-1}) > 0$ for all $k \in \{1, \ldots, n\}$.

Note that any state $x$ is always upper reachable from itself. Lower reachability is a bit more involved.

**Definition 4.32.** Consider a lower rate operator $\underline{Q}$. For any state $x$ in $\mathscr{X}$ and any non-empty subset $A$ of $\mathscr{X}$, we say that $A$ is *lower reachable* from $x$, and denote this by $x \hookrightarrow A$, if $x$ belongs to $B_n$, where $(B_k)_{k\in\mathbb{Z}_{\geq 0}}$ is the sequence that is defined by the initial condition $B_0 := A$ and by the recursive relation

$$B_{k+1} := B_k \cup \left\{y \in \mathscr{X} \setminus B_k : [\underline{Q}\mathbb{I}_{B_k}](y) > 0\right\} \quad \text{for all } k \in \mathbb{Z}_{\geq 0},$$

and $n \leq |\mathscr{X} \setminus A|$ is the first index $k$ for which $B_k = B_{k+1}$.

The notions of upper and lower reachability might seem a bit daunting at first, but checking them is actually quite simple in practice (see De Bock, 2017b, Algorithms 1 and 2). More importantly, and as announced, the following result makes clear that they are linked to ergodicity in a fundamental way (see De Bock, 2017b, Theorem 19).

**Proposition 4.33.** *A lower rate operator $\underline{Q}$ is ergodic if and only if*

$$\mathscr{X}_{\underline{Q}} := \left\{x \in \mathscr{X} : (\forall y \in \mathscr{X})\ y \hookrightarrow x\right\} \neq \emptyset; \tag{4.20}$$

*and*

$$(\forall y \in \mathscr{X} \setminus \mathscr{X}_{\underline{Q}})\ y \hookrightarrow \mathscr{X}_{\underline{Q}}.$$

Whenever a lower rate operator $\underline{Q}$ is ergodic, we call the set $\mathscr{X}_{\underline{Q}}$ as defined by Eq. (4.20) the *top class*; the reason for this is that in this case, $\mathscr{X}_{\underline{Q}}$ is the maximal class for the upper reachability relation $\bullet \hookrightarrow \bullet$ (see De Bock, 2017b, p. 174).

Proposition 4.33 is reminiscent of Proposition 4.27$_{189}$, and they turn out to be even more similar than one would expect at first sight. To make this similarity more obvious, we repeat (De Bock, 2017b, Propositions 17 and 18).

**Lemma 4.34.** *Consider a lower rate operator $\underline{Q}$. Then for any $t$ in $\mathbb{R}_{>0}$, all states $x$ and $y$ in $\mathscr{X}$ and any non-empty subset $A$ of $\mathscr{X}$,*

$$-\left[e^{t\underline{Q}}(-\mathbb{I}_x)\right](y) =: \left[e^{t\overline{Q}}\mathbb{I}_x\right](y) > 0 \Leftrightarrow y \hookrightarrow x$$

*and*

$$\left[e^{t\underline{Q}}\mathbb{I}_A\right](y) > 0 \Leftrightarrow y \hookrightarrow A.$$

**Joseph's Example 4.35.** Consider again the general lower rate operator $\underline{Q}$ as used in Joseph's Example 4.30$_{\curvearrowleft}$ and introduced in Joseph's Example 3.64$_{109}$. In Joseph's Example 4.30$_{\curvearrowleft}$, it cost us quite a bit of work to determine whether $\underline{Q}$ was ergodic or not. Here, we check if $\underline{Q}$ is ergodic using the simpler conditions of Proposition 4.33.

Because any state is upper reachable from itself, $\mathtt{H} \hookrightarrow \mathtt{H}$ and $\mathtt{T} \hookrightarrow \mathtt{T}$. Consequently, $\mathtt{H}$ belongs to $\mathscr{X}_{\underline{Q}}$ if and only if $\mathtt{T} \hookrightarrow \mathtt{H}$, and it is not difficult too see that due to Definition 4.31<sub>192</sub>, this can only be the case if $[\overline{Q}\mathbb{I}_{\mathtt{H}}](\mathtt{T}) = \overline{\lambda}_{\mathtt{T}} > 0$. Similarly,

$$\mathtt{T} \in \mathscr{X}_{\underline{Q}} \Leftrightarrow \mathtt{H} \hookrightarrow \mathtt{T} \Leftrightarrow [\overline{Q}\mathbb{I}_{\mathtt{T}}](\mathtt{H}) = \overline{\lambda}_{\mathtt{H}} > 0.$$

Hence, $\mathscr{X}_{\underline{Q}}$ is non-empty if and only if $\overline{\lambda}_{\mathtt{H}} > 0$ *or* $\overline{\lambda}_{\mathtt{T}} > 0$.

In case both $\overline{\lambda}_{\mathtt{H}} > 0$ *and* $\overline{\lambda}_{\mathtt{T}} > 0$, the second condition of Proposition 4.33 is trivially satisfied, and $\underline{Q}$ is ergodic. What remains is the case that one of the two is zero, say $\overline{\lambda}_{\mathtt{H}}$. In this case, $\mathscr{X}_{\underline{Q}} = \{\mathtt{H}\}$, and the second condition is satisfied if and only if $\mathtt{T} \rightarrow \{\mathtt{H}\}$. By Definition 4.32, this can only be the case if $[\underline{Q}\mathbb{I}_{\mathtt{H}}](\mathtt{T}) = \underline{\lambda}_{\mathtt{T}} > 0$. By symmetry, in case $\overline{\lambda}_{\mathtt{T}} = 0$ and $\overline{\lambda}_{\mathtt{H}} > 0$, the second condition of Proposition 4.33 is satisfied if and only if

$$\mathtt{H} \rightarrow \{\mathtt{T}\} \Leftrightarrow [\underline{Q}\mathbb{I}_{\mathtt{T}}](\mathtt{H}) = \underline{\lambda}_{\mathtt{H}} > 0.$$

In summary, we have verified that $\underline{Q}$ is ergodic if and only if either (i) $\overline{\lambda}_{\mathtt{H}} > 0$ and $\overline{\lambda}_{\mathtt{T}} > 0$, (ii) $\overline{\lambda}_{\mathtt{H}} = 0$ and $\underline{\lambda}_{\mathtt{T}} > 0$, or (iii) $\overline{\lambda}_{\mathtt{T}} = 0$ and $\underline{\lambda}_{\mathtt{H}} > 0$. Thus, the lower rate operator $\underline{Q}$ is ergodic if and only if $\overline{\lambda}_{\mathtt{H}} + \underline{\lambda}_{\mathtt{T}} > 0$ and $\underline{\lambda}_{\mathtt{H}} + \overline{\lambda}_{\mathtt{T}} > 0$, which – of course – agrees with what we found in Joseph's Example 4.30<sub>192</sub>.

It essentially follows from Definitions 4.26<sub>189</sub> and 4.29<sub>191</sub> (see De Bock, 2017b, Proposition 13) that the lower rate operator $\underline{Q}$ is ergodic if and only if for all $t$ in $\mathbb{R}_{>0}$, the generated lower transition operator $e^{t\underline{Q}}$ is ergodic as well. It turns out that whenever this is the case, the approximation $(I + \Delta\underline{Q})$ is ergodic as well for any step size $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\|\underline{Q}\|_{\mathrm{op}} < 2$.

**Theorem 4.36.** *A lower rate operator $\underline{Q}$ is ergodic if and only if there is some natural number $n < |\mathscr{X}|$ such that for some (and then all) natural number(s) $k \geq n$ and some (and then all) step size(s) $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\|\underline{Q}\|_{\mathrm{op}} < 2$:*

$$\rho\big((I + \Delta\underline{Q})^k\big) < 1.$$

Because our proof of this result is somewhat lengthy,[2] we have relegated it to Appendix 4.D<sub>210</sub>. Suffice to say that in our proof, we rely heavily on Propositions 4.27<sub>189</sub> and 4.33 and Theorem 4.28<sub>191</sub>.

### 4.3.3 Back to numerical integration

Theorem 4.36 guarantees that for an ergodic lower rate operator, Eq. (4.17)<sub>189</sub> is satisfied for sufficiently large $m$. In particular, if the lower rate operator $\underline{Q}$ is ergodic, then there is some natural number $n < |\mathscr{X}|$ such that $\rho((I+\Delta\underline{Q})^m) < 1$

[2]This might be a bit of an understatement, as it is almost 3 pages long, excluding the necessary technical lemmas. Be that as it may, the proof is actually a simplified version of the proof of a more general result in (Erreygers et al., 2017a,b, Theorem 8).

for all $m \geq n$ and all $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\|\underline{Q}\|_{\mathrm{op}} < 2$. Consequently, if we choose $m \geq |\mathscr{X}| - 1$ then re-evaluating the step size $\Delta$ in Algorithm 4.6$_{185}$ will – except maybe for the last re-evaluation – result in a new step size that is strictly greater than the previous one. Therefore, we conclude that if the lower rate operator $\underline{Q}$ is ergodic, then using Algorithm 4.6$_{185}$ is certainly justified; Algorithm 4.6$_{185}$ will need fewer iterations – that is, evaluations of $\underline{Q}$ – than Algorithm 4.5$_{183}$, provided $m$ is sufficiently large.

Another nice consequence of the ergodicity of a lower rate operator $\underline{Q}$ is that we can prove an alternative a priori guaranteed upper bound for the error of approximations with a fixed step size.

**Proposition 4.37.** *Consider a lower rate operator $\underline{Q}$. Fix some gamble $f$ on $\mathscr{X}$, two natural numbers $m$ and $n$, and a step size $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\|\underline{Q}\|_{\mathrm{op}} < 2$. For any positive real number $\beta$ such that $\rho((I+\Delta\underline{Q})^m) \leq \beta < 1$,*

$$\left\|e^{t\underline{Q}}f - (I+\Delta\underline{Q})^n f\right\| \leq \epsilon_{\mathrm{e}} := \frac{1}{2} m\Delta^2 \|\underline{Q}\|_{\mathrm{op}}^2 \|f\|_{\mathrm{c}} \frac{1-\beta^k}{1-\beta} \leq \epsilon'_{\mathrm{e}} := \frac{m\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2\|f\|_{\mathrm{c}}}{2(1-\beta)},$$

*where we let $t := n\Delta$ and $k := \lceil n/m \rceil$.*

*Proof.* Recall from Lemma 4.19$_{180}$ that

$$\left\|e^{t\underline{Q}}f - (I+\Delta\underline{Q})^n f\right\| \leq \frac{1}{2}\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2 \sum_{\ell=0}^{n-1} \|(I+\Delta\underline{Q})^\ell f\|_{\mathrm{c}}$$

Note that by definition, $km \geq n$. We add some non-negative terms to the right-hand side of the previous inequality, to yield

$$\begin{aligned}\left\|e^{t\underline{Q}}f - (I+\Delta\underline{Q})^n f\right\| &\leq \frac{1}{2}\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2 \sum_{\ell=0}^{km-1} \|(I+\Delta\underline{Q})^\ell f\|_{\mathrm{c}} \\ &= \frac{1}{2}\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2 \sum_{\ell=0}^{k-1}\sum_{i=0}^{m} \|(I+\Delta\underline{Q})^i (I+\Delta\underline{Q})^{m\ell} f\|_{\mathrm{c}}\end{aligned}$$

Because $(I+\Delta\underline{Q})^i$ is a lower transition operator due to Lemma 3.72$_{112}$ and (LT11)$_{108}$, it follows from (LT13)$_{179}$ that

$$\left\|e^{t\underline{Q}}f - (I+\Delta\underline{Q})^n f\right\| \leq \frac{1}{2}m\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2 \sum_{\ell=0}^{k-1} \|(I+\Delta\underline{Q})^{m\ell} f\|_{\mathrm{c}}.$$

Furthermore, it follows from (EC3)$_{190}$ and Eq. (4.6)$_{179}$ that

$$\left\|e^{t\underline{Q}}f - (I+\Delta\underline{Q})^n f\right\| \leq \frac{1}{2}m\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2 \|f\|_{\mathrm{c}} \sum_{\ell=0}^{k-1} \rho\big((I+\Delta\underline{Q})^{m\ell}\big),$$

and from (EC2)$_{190}$ that

$$\begin{aligned}\left\|e^{t\underline{Q}}f - (I+\Delta\underline{Q})^n f\right\| &\leq \frac{1}{2}m\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2 \|f\|_{\mathrm{c}} \sum_{\ell=0}^{k-1} \rho\big((I+\Delta\underline{Q})^{m}\big)^\ell \\ &\leq \frac{1}{2}m\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2 \|f\|_{\mathrm{c}} \sum_{\ell=0}^{k-1} \beta^\ell.\end{aligned}$$

The statement follows immediately from the preceding inequality because, by assumption, $0 < \beta < 1$ □

Interestingly enough, the upper bound $\epsilon'_{\mathrm{e}}$ is not dependent on $t$ (or $n$) at all! This is a significant improvement on the upper bound of Proposition 4.20<sub>181</sub>, as that upper bound is proportional to $t^2$. However, it is clear from the proof of Proposition 4.37 that the a priori error bounds $\epsilon_{\mathrm{e}}$ and $\epsilon'_{\mathrm{e}}$ are always greater than the on-the-fly error bound $\epsilon_{\mathrm{otf}}$.

Recall from Theorem 4.36[194] that there always is a natural number $m < |\mathscr{X}|$ such that $\rho((I + \Delta\underline{Q})^m) < 1$ for all $\Delta$ in $\mathbb{R}_{>0}$ with $\Delta\|\underline{Q}\|_{\mathrm{op}} < 2$. Thus, given such an $m$, we could theoretically improve Algorithm 4.5[183], in the sense that we might be able to get away with fewer iterations, or equivalently, with a larger step size $\Delta$. To see how this works in a bit more detail, we recall that on line 1 of Algorithm 4.5[183], we choose $n$ such that

$$\epsilon_{\mathrm{act}} := \|e^{t\underline{Q}}f - (I + \Delta\underline{Q})^n f\| \leq \epsilon_n := \frac{1}{2}n\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2\|f\|_{\mathrm{c}} \leq \epsilon.$$

However, courtesy of Proposition 4.37, we also know that

$$\epsilon_{\mathrm{act}} \leq \epsilon_{\mathrm{e}} = \frac{1}{2}m\Delta^2\|\underline{Q}\|_{\mathrm{op}}^2\|f\|_{\mathrm{c}}\frac{1-\beta^k}{1-\beta}.$$

Note that this latter is a tighter a priori upper bound on the error if and only if

$$m\frac{1-\beta^k}{1-\beta} < n.$$

Whenever this is the case, we could go looking for the smallest natural number $n$ in $\mathbb{N}$ that yields

$$\frac{1}{2}m\Delta_n^2\|\underline{Q}\|_{\mathrm{op}}^2\|f\|_c(1-\beta_n^{k_n}) \leq (1-\beta_n)\epsilon,$$

where $k_n = \lceil n/m \rceil$ and $\Delta_n = t/n$ depend on $n$, and $\beta_n < 1$ is an upper bound on $\rho((I + \Delta_n\underline{Q})^m)$. This method would yield a smaller $n$, but the time we gain by having to execute fewer iterations does not necessarily compensate the time lost by looking for this smaller $n$.

### Approximating the coefficient of ergodicity

Were we to actually implement this improvement, we would need to be able to determine (an upper bound on) the coefficient of ergodicity $\rho((I + \Delta\underline{Q})^m)$. In general, this is certainly non-trivial, and it may not even be possible. For this reason, we will have to make do with the following upper and lower bounds for the coefficient of ergodicity that can always be computed.

**Proposition 4.38.** *Let $\underline{T}$ be a lower transition operator. Then*

$$\rho(\underline{T}) \leq \max\big\{\max\{[\overline{T}\mathbb{I}_A](x) - [\underline{T}\mathbb{I}_A](y) \colon x, y \in \mathscr{X}\} \colon \emptyset \neq A \subset \mathscr{X}\big\},$$

*and*

$$\rho(\underline{T}) \geq \max\big\{\max\{[\underline{T}\mathbb{I}_A](x) - [\underline{T}\mathbb{I}_A](y) \colon x, y \in \mathscr{X}\} \colon \emptyset \neq A \subset \mathscr{X}\big\}.$$

*Proof.* Fix some lower transition operator $\underline{T}$. The lower bound on $\rho(\underline{T})$ follows from the fact that for any $\emptyset \neq A \subset \mathscr{X}$, $0 \leq \mathbb{I}_A \leq 1$. To obtain the upper bound, we observe that

$$\begin{aligned}\rho(\underline{T}) &= \max\{\|\underline{T}f\|_{\mathrm{v}} \colon f \in \mathbb{G}(\mathscr{X}), 0 \leq f \leq 1\}\\ &= \max\Big\{\max\{\big|[\underline{T}f](x) - [\underline{T}f](y)\big| \colon x, y \in \mathscr{X}\} \colon f \in \mathbb{G}(\mathscr{X}), 0 \leq f \leq 1\Big\}\\ &= \max\Big\{\max\{\big|[\underline{T}f](x) - [\underline{T}f](y)\big| \colon f \in \mathbb{G}(\mathscr{X}), 0 \leq f \leq 1\} \colon x, y \in \mathscr{X}\Big\}.\end{aligned}$$

Recall from Corollary 3.61$_{107}$ that for any $x$ in $\mathscr{X}$, $[\underline{T}\bullet](x)$ is a coherent lower expectation on $\mathbb{G}(\mathscr{X})$. Thus,

$$\rho(\underline{T}) = \max\Big\{d\big([\underline{T}\bullet](x), [\underline{T}\bullet](y)\big) \colon x, y \in \mathscr{X}\Big\},$$

where we use the metric $d$ as defined in Eq. (A.1)$_{443}$. From this and Lemma A.1$_{443}$, it follows that

$$\begin{aligned}\rho(\underline{T}) &\leq \max\Big\{\max\{[\overline{T}\mathbb{I}_A](x) - [\underline{T}\mathbb{I}_A](y) \colon \emptyset \neq A \subset \mathscr{X}\} \colon x, y \in \mathscr{X}\Big\}\\ &= \max\Big\{\max\{[\overline{T}\mathbb{I}_A](x) - [\underline{T}\mathbb{I}_A](y) \colon x, y \in \mathscr{X}\} \colon \emptyset \neq A \subset \mathscr{X}\Big\},\end{aligned}$$

where for the last equality we simply change the order of two maxima. This verifies the upper bound of the statement, as required. □

The upper bound in Proposition 4.38 corresponds to what Škulj et al. (2013, Section 5.1) call the 'uniform coefficient of ergodicity'. Observe that for a transition operator $T$, the upper and lower bounds in Proposition 4.38 are equal. In this case, it furthermore follows from (Škulj et al., 2013, Proposition 1) that

$$\rho(T) = \max\left\{\frac{1}{2}\sum_{z \in \mathscr{X}} |T(x,z) - T(y,z)| \colon x, y \in \mathscr{X}\right\},$$

which is (one of) the coefficient(s) of ergodicity for transition operators – see (Seneta, 1979), (Diener et al., 1995, Section 1.11), (Anderson, 1991, Section 6.1) or (Škulj et al., 2013, Section 5). Besides (linear) transition operators, there are other lower transition operators for which the lower bound of Proposition 4.38 is equal to the coefficient of ergodicity. Škulj et al. (2013, Proposition 22 and Corollary 23) show that this is also the case for lower transition operators defined using Choquet integrals with respect to 2-monotone lower probabilities.

The upper bound in Proposition 4.38 is particularly useful in combination with Proposition $4.37_{195}$, at least in case it is strictly smaller than 1. In (Erreygers et al., 2017a, Proposition 11), we mistakenly claimed that this is always the case for lower transition operators of the form $(I+\Delta\underline{Q})^m$, with $\underline{Q}$ an ergodic lower rate operator and $m$ sufficiently large. We can use our running example to illustrate that this may not be the case though.

**Joseph's Example 4.39.** As in Joseph's Example $4.30_{192}$ and Joseph's Example $4.35_{193}$, we consider a general lower rate operator $\underline{Q}$, which by Joseph's Example $3.64_{109}$ is uniquely characterised by the four parameters $\underline{\lambda}_{\mathtt{H}}$, $\overline{\lambda}_{\mathtt{H}}$, $\underline{\lambda}_{\mathtt{T}}$ and $\overline{\lambda}_{\mathtt{T}}$. In this counterexample, we take $\underline{\lambda}_{\mathtt{H}}=0=\underline{\lambda}_{\mathtt{T}}$, $\overline{\lambda}_{\mathtt{H}}>0$ and $\overline{\lambda}_{\mathtt{T}}>0$.

Because $\overline{\lambda}_{\mathtt{H}}+\underline{\lambda}_{\mathtt{T}}>0$ and $\underline{\lambda}_{\mathtt{H}}+\overline{\lambda}_{\mathtt{T}}>0$, $\underline{Q}$ is ergodic as per the discussion in Joseph's Example $4.30_{192}$. Furthermore, it follows immediately from $(LR7)_{111}$ that $\|\underline{Q}\|_{\mathrm{op}}=2\max\{\overline{\lambda}_{\mathtt{H}},\overline{\lambda}_{\mathtt{T}}\}>0$. Fix any natural number $m$ and a step size $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\overline{\lambda}_{\mathtt{H}}<1$ and $\Delta\overline{\lambda}_{\mathtt{T}}<1$. Then $(I+\Delta\underline{Q})$ is a lower transition operator due to Lemma $3.72_{112}$. Let $\underline{T}\coloneqq(I+\Delta\underline{Q})^{\overline{m}}$, and observe that by construction, $\rho(\underline{T})<1$ due to Theorem $4.36_{194}$.

Let us determine the upper bound on $\rho(\underline{T})$ given in Proposition 4.38. First, we observe that the only two non-empty strict subsets of $\mathscr{X}$ are $\{\mathtt{H}\}$ and $\{\mathtt{T}\}$. Second, after some straightforward calculations, we obtain that

$$\begin{aligned}\big[(I+\Delta\overline{Q})\mathbb{I}_{\mathtt{H}}\big](\mathtt{H})&=1-\Delta\underline{\lambda}_{\mathtt{H}}=1, & 0<\big[(I+\Delta\overline{Q})\mathbb{I}_{\mathtt{H}}\big](\mathtt{T})&=\Delta\overline{\lambda}_{\mathtt{T}}<1,\\ 0<\big[(I+\Delta\overline{Q})\mathbb{I}_{\mathtt{T}}\big](\mathtt{H})&=\Delta\overline{\lambda}_{\mathtt{H}}<1, & \big[(I+\Delta\overline{Q})\mathbb{I}_{\mathtt{T}}\big](\mathtt{T})&=1-\Delta\underline{\lambda}_{\mathtt{T}}=1,\end{aligned}$$

and

$$\begin{aligned}1>\big[(I+\Delta\underline{Q})\mathbb{I}_{\mathtt{H}}\big](\mathtt{H})&=1-\Delta\overline{\lambda}_{\mathtt{H}}>0, & \big[(I+\Delta\underline{Q})\mathbb{I}_{\mathtt{H}}\big](\mathtt{T})&=\Delta\underline{\lambda}_{\mathtt{T}}=0,\\ \big[(I+\Delta\underline{Q})\mathbb{I}_{\mathtt{T}}\big](\mathtt{H})&=\Delta\underline{\lambda}_{\mathtt{H}}=0, & 1>\big[(I+\Delta\underline{Q})\mathbb{I}_{\mathtt{T}}\big](\mathtt{T})&=1-\Delta\overline{\lambda}_{\mathtt{T}}>0.\end{aligned}$$

From this, we infer that $\mathbb{I}_{\mathtt{H}}\leq(I+\Delta\overline{Q})\mathbb{I}_{\mathtt{H}}\leq1$ and $0\leq(I+\Delta\underline{Q})\mathbb{I}_{\mathtt{H}}\leq\mathbb{I}_{\mathtt{H}}$. Because $(I+\Delta\underline{Q})$ is a lower transition operator, it follows from these two inequalities and repeated application of $(LT6)_{108}$ and $(LT4)_{108}$ that

$$\big[(I+\Delta\overline{Q})^m\mathbb{I}_{\mathtt{H}}\big](\mathtt{H})=\big[\overline{T}\mathbb{I}_{\mathtt{H}}\big](\mathtt{H})=1\quad\text{and}\quad\big[(I+\Delta\underline{Q})^m\mathbb{I}_{\mathtt{H}}\big](\mathtt{T})=\big[\underline{T}\mathbb{I}_{\mathtt{H}}\big](\mathtt{T})=0.$$

Thus, we see that

$$\max\Big\{\max\{[\overline{T}\mathbb{I}_A](x)-[\underline{T}\mathbb{I}_A](y)\colon x,y\in\mathscr{X}\}\colon\emptyset\neq A\subset\mathscr{X}\Big\}\\ \geq\big[\overline{T}\mathbb{I}_{\mathtt{H}}\big](\mathtt{H})-\big[\underline{T}\mathbb{I}_{\mathtt{H}}\big](\mathtt{T})=1.$$

## 4.A Proofs of Proposition 4.8 and Theorem 4.9

In this appendix, we prove the two results in Section $4.1.3_{162}$. First, we prove Proposition $4.8_{166}$.

**Proposition 4.8.** *Consider a Markovian imprecise jump process $\mathscr{P}$ that satisfies the sum-product law of iterated lower expectations. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, a sequence of time-points $v = (t_1, \dots, t_n)$ in $\mathscr{U}_{>u}$, and an $\mathscr{F}_u$-simple variable $f$ with sum-product representation*

$$f = \sum_{k=1}^{n} g_k(X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell})$$

*over $v$. Then*

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}(f_1(X_{t_1}) \mid X_u = x_u),$$

*where $f_1 \colon \mathscr{X} \to \mathbb{R}$ is recursively defined by the initial condition $f_n := g_n$ and, for all $k$ in $\{1, \dots, n-1\}$, by the recursive relation*

$$f_k \colon \mathscr{X} \to \mathbb{R} \colon x \mapsto \underline{E}_{\mathscr{P}}(f_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = x) h_k(x) + g_k(x).$$

*Proof.* First, we verify that for all $k$ in $\{1, \dots, n-1\}$, $f_k$ as defined in the statement is indeed a gamble on $\mathscr{X}$, or in other words, that $f_k$ is bounded. To this end, we fix some $k$ in $\{1, \dots, n-1\}$ and assume that $f_{k+1}$ is a gamble on $\mathscr{X}$ – for $k = n-1$, this is true because $f_n = g_n$ by the initial condition in the statement. For all $x$ in $\mathscr{X}$ and $P$ in $\mathscr{P}$, it follows from (ES1)$_{37}$ that

$$-\max|f_{k+1}| \leq \min f_{k+1} \leq E_P(f_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = x) \leq \max f_{k+1} \leq \max|f_{k+1}|.$$

Because inequalities are preserved when taking infima, it follows that, for all $x$ in $\mathscr{X}$,

$$|\underline{E}_{\mathscr{P}}(f_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = x)| \leq \max|f_{k+1}|,$$

and therefore

$$|f_k| \leq (\max|f_{k+1}|)(\max h_k) + \max|g_k|,$$

where we also used that $h_k \geq 0$. Because $f_{k+1}$, $g_k$ and $h_k$ are bounded, it follows from this inequality that $f_k$ is bounded as well, as required.

Next, we verify the equality in the statement. For any $k$ in $\{1, \dots, n-1\}$, we let $t_{1:k} := (t_1, \dots, t_k)$ and let $\tilde{g}_k$ and $\tilde{h}_k$ be the gambles on $\mathscr{X}_{t_{1:k}}$ that are defined for all $y_{t_{1:k}}$ in $\mathscr{X}_{t_{1:k}}$ by

$$\tilde{g}_k(y_{t_{1:k}}) := \sum_{i=1}^{k} g_i(y_{t_i}) \prod_{\ell=1}^{i-1} h_\ell(y_{t_\ell}) \quad \text{and} \quad \tilde{h}_k(y_{t_{1:k}}) := \prod_{\ell=1}^{k} h_\ell(y_{t_\ell}).$$

Note that $\tilde{h}_k \geq 0$ because $h_\ell \geq 0$ for all $\ell$ in $\{1, \dots, k\}$ by assumption. In order to elegantly deal with an edge case, we let $t_{1:0} := ()$ and let $\tilde{g}_0 := 0$ and $\tilde{h}_0 := 1$ be constant gambles in $\mathbb{G}(\mathscr{X}_{t_{1:0}})$.

We set out to verify that for all $k$ in $\{0, \dots, n-1\}$,

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\big(f_{k+1}(X_{t_{k+1}}) \tilde{h}_k(X_{t_{1:k}}) + \tilde{g}_k(X_{t_{1:k}}) \,\big|\, X_u = x_u\big); \tag{4.21}$$

note that for $k = 0$, this is the equality in the statement because $\tilde{g}_0 = 0$ and $\tilde{h}_0 = 1$ by construction. Our proof will be one by induction. For the base case that $k = n-1$, we observe that by construction,

$$f = f_n(X_{t_n}) \tilde{h}_{n-1}(X_{t_{1:n-1}}) + \tilde{g}_{n-1}(X_{t_{1:n-1}}) = f_{k+1}(X_{t_{k+1}}) \tilde{h}_k(X_{t_{1:k}}) + \tilde{g}_k(X_{t_{1:k}});$$

clearly, this verifies Eq. (4.21)$_{\curvearrowleft}$ in case $k = n-1$. For the inductive step, we fix some $\ell$ in $\{0, \dots, n-2\}$, and assume that Eq. (4.21)$_{\curvearrowleft}$ holds for $k = \ell + 1$, so

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\big(f_{\ell+2}(X_{t_{\ell+2}})\tilde{h}_{\ell+1}(X_{t_{1:\ell+1}}) + \tilde{g}_{\ell+1}(X_{t_{1:\ell+1}}) \,\big|\, X_u = x_u\big).$$

Because $\tilde{h}_{\ell+1} \geq 0$ and because $\mathscr{P}$ satisfies the sum-product law of iterated lower expectations by assumption, it follows from this and Eq. (3.70)$_{106}$ in Definition 3.59$_{106}$ – with $t = t_{\ell+2}$ and $v = t_{1:\ell+1}$ – that

$$\begin{aligned}\underline{E}_{\mathscr{P}}&(f \mid X_u = x_u)\\ &= \underline{E}_{\mathscr{P}}\big(f'_{\ell+1}(X_u, X_{t_{1:\ell+1}})\tilde{h}_{\ell+1}(X_{t_{1:\ell+1}}) + \tilde{g}_{\ell+1}(X_{t_{1:\ell+1}}) \,\big|\, X_u = x_u\big),\end{aligned} \tag{4.22}$$

where we let

$$f'_{\ell+1} \colon \mathscr{X}_{u \cup t_{1:\ell+1}} \to \mathbb{R} \colon y_{u \cup t_{1:\ell+1}} \mapsto \underline{E}_{\mathscr{P}}\big(f_{\ell+2}(X_{t_{\ell+2}}) \,\big|\, X_u = y_u, X_{t_{1:\ell+1}} = y_{t_{1:\ell+1}}\big).$$

The imprecise jump process $\mathscr{P}$ is Markovian by assumption, so by Definition 3.84$_{118}$,

$$f'_{\ell+1}(X_u, X_{t_{1:\ell+1}}) = \underline{E}_{\mathscr{P}}\big(f_{\ell+2}(X_{t_{\ell+2}}) \,\big|\, X_u, X_{t_{1:\ell+1}}\big) = \underline{E}_{\mathscr{P}}\big(f_{\ell+2}(X_{t_{\ell+2}}) \,\big|\, X_{t_{\ell+1}}\big).$$

Furthermore, we observe that, by construction, $\tilde{h}_{\ell+1}(X_{t_{1:\ell+1}}) = h_{\ell+1}(X_{t_{\ell+1}})\tilde{h}_{\ell}(X_{t_{1:\ell}})$ and $\tilde{g}_{\ell+1}(X_{t_{1:\ell+1}}) = g_{\ell+1}(X_{t_{\ell+1}})\tilde{h}_{\ell}(X_{t_{1:\ell}}) + \tilde{g}_{\ell}(X_{t_{1:\ell}})$. Consequently,

$$\begin{aligned}f'_{\ell+1}&(X_u, X_{t_{1:\ell+1}})\tilde{h}_{\ell+1}(X_{t_{1:\ell+1}}) + \tilde{g}_{\ell+1}(X_{t_{1:\ell+1}})\\ &= \underline{E}_{\mathscr{P}}\big(f_{\ell+2}(X_{t_{\ell+2}}) \,\big|\, X_{t_{\ell+1}}\big)\tilde{h}_{\ell+1}(X_{t_{1:\ell+1}}) + \tilde{g}_{\ell+1}(X_{t_{1:\ell+1}})\\ &= \underline{E}_{\mathscr{P}}\big(f_{\ell+2}(X_{t_{\ell+2}}) \,\big|\, X_{t_{\ell+1}}\big)h_{\ell+1}(X_{t_{\ell+1}})\tilde{h}_{\ell}(X_{t_{1:\ell}})\\ &\qquad + g_{\ell+1}(X_{t_{\ell+1}})\tilde{h}_{\ell}(X_{t_{1:\ell}}) + \tilde{g}_{\ell}(X_{t_{1:\ell}})\\ &= \big(\underline{E}_{\mathscr{P}}\big(f_{\ell+2}(X_{t_{\ell+2}}) \,\big|\, X_{t_{\ell+1}}\big)h_{\ell+1}(X_{t_{\ell+1}}) + g_{\ell+1}(X_{t_{\ell+1}})\big)\tilde{h}_{\ell}(X_{t_{1:\ell}}) + \tilde{g}_{\ell}(X_{t_{1:\ell}})\\ &= f_{\ell+1}(X_{t_{\ell+1}})\tilde{h}_{\ell}(X_{t_{1:\ell}}) + \tilde{g}_{\ell}(X_{t_{1:\ell}}).\end{aligned}$$

We substitute this equality in Eq. (4.22), to yield

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\big(f_{\ell+1}(X_{t_{\ell+1}})\tilde{h}_{\ell}(X_{t_{1:\ell}}) + \tilde{g}_{\ell}(X_{t_{1:\ell}}) \,\big|\, X_u = x_u\big),$$

which is Eq. (4.21)$_{\curvearrowleft}$ for $k = \ell$, as required. □

Second, we use Proposition 4.8$_{166}$ to prove the following intermediary result, which we will need in our proof for Theorem 4.9$_{166}$ further on.

**Lemma 4.40.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Fix a sequence of time-points $v = (t_1, \dots, t_n)$ in $\mathscr{U}_{\neq()}$ and a real variable $f$ in $\mathbb{V}(\Omega)$ that has a sum-product representation*

$$f(X_v) = \sum_{k=1}^{n} g_k(X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell})$$

*over $v$. Let $f_1\colon \mathscr{X} \to \mathbb{R}$ be recursively defined by the initial condition $f_n := g_n$ and, for all $k$ in $\{1,\dots,n-1\}$, by the recursive relation*

$$f_k\colon \mathscr{X} \to \mathbb{R}\colon x \mapsto \left[e^{(t_{k+1}-t_k)\underline{Q}_{\mathscr{Q}}} f_{k+1}\right](x)\, h_k(x) + g_k(x).$$

*Then for all $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec v$, and with $s := \max u$,*

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = \begin{cases} \left[e^{(t_1-s)\underline{Q}_{\mathscr{Q}}} f_1\right](x_s) & \text{if } u \neq (),\\ \underline{E}_{\mathscr{M}}(f_1) & \text{if } u = () \text{ and } t_1 = 0,\\ \underline{E}_{\mathscr{M}}\left(e^{t_1\underline{Q}_{\mathscr{Q}}} f_1\right) & \text{if } u = () \text{ and } t_1 > 0,\end{cases} \tag{4.23}$$

*and therefore*

$$\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{M},\mathscr{Q}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}(f \mid X_u = x_u).$$

*Proof.* First, we let $\mathscr{P}$ be equal to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ or $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Because $\mathscr{Q}$ has separately specified rows by assumption, $\mathscr{P}$ is Markovian by Corollary 3.87<sub></sub>119 and satisfies the sum-product law of iterated lower expectations by Theorem 3.89 [120]. Hence, if follows from Proposition 4.8 [166] and Proposition 3.81 [117] that

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}(f_1(X_{t_1}) \mid X_u = x_u).$$

By Propositions 3.81 [117] to 3.83 [118], this equality implies Eq. (4.23), as required.

Next, we prove the second part of the statement, so we let $\mathscr{P}$ be any imprecise jump process such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Then clearly

$$\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f \mid X_u = x_u) \leq \underline{E}_{\mathscr{P}}(f \mid X_u = x_u) \leq \underline{E}_{\mathscr{M},\mathscr{Q}}(f \mid X_u = x_u).$$

From the first part of our proof, we know that Eq. (4.23) holds for $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Consequently, the preceding inequalities imply that Eq. (4.23) holds for $\mathscr{P}$ as well, and this also proves the other equalities in the statement. □

In our proof for Theorem 4.9 [166], we also need the following intermediary result, which essentially follows from Lemma 4.40 [↶].

**Lemma 4.41.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Fix some grid $v = (t_1,\dots,t_n)$ over $[s,r]$ and a real variable $f$ with sum-product representation*

$$f = \sum_{k=1}^{n} g_k(X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell})$$

*over $v$. Then for all $x$ in $\mathscr{X}$ and $\{X_u = x_u\}$ in $\mathscr{H}$ such that $u \prec v$,*

$$\underline{E}_{\mathscr{P}}\left(f \,\middle|\, X_u = x_u, X_{t_1} = x\right) = f_1(x),$$

*where $f_1\colon \mathscr{X} \to \mathbb{R}$ is recursively defined by the initial condition $f_n := g_n$ and, for all $k$ in $\{1,\dots,n-1\}$, by the recursive relation*

$$f_k\colon \mathscr{X} \to \mathbb{R}\colon x \mapsto \left[e^{(t_{k+1}-t_k)\underline{Q}_{\mathscr{Q}}} f_{k+1}\right](x)\, h_k(x) + g_k(x).$$

*Proof.* Let us first consider the edge case $n = 1$. Then $f = g_1(X_{t_1}) = f_1(X_{t_1})$, so it follows immediately from Corollary 4.1158 that

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u, X_{t_1} = x) = \underline{E}_{\mathscr{P}}(f_1(X_{t_1}) \mid X_u = x_u, X_{t_1} = x) = \underline{E}_{\mathscr{P}}(f_1(x) \mid X_u = x_u, X_{t_1} = x).$$

For any jump process $P$ in $\mathscr{P}$, it follows from (ES1)37 that

$$E_P(f_1(x) \mid X_u = x_u, X_{t_1} = x) = f_1(x),$$

and therefore

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u, X_{t_1} = x) = f_1(x),$$

as required.

Next, we consider the more involved case that $n \geq 2$. Let $t_{2:n} := (t_2, \ldots, t_n)$, and let $f_{2:n}$ be the gamble on $\mathscr{X}_{t_{2:n}}$ defined by

$$f_{2:n}(x_{t_{2:n}}) := \sum_{k=2}^{n} g_k(x_{t_k}) \prod_{\ell=2}^{k-1} h_\ell(x_{t_\ell}) \quad \text{for all } x_{t_{2:n}} \in \mathscr{X}_{t_{2:n}}.$$

Note that $f = g_1(X_{t_1}) + h_1(X_{t_1}) f_{2:n}(X_{t_{2:n}})$, and therefore

$$\begin{aligned}\underline{E}_{\mathscr{P}}(f \mid X_u = x_u, X_{t_1} = x) &= \underline{E}_{\mathscr{P}}\big(g_1(X_{t_1}) + h_1(X_{t_1}) f_{2:n}(X_{t_{2:n}}) \,\big|\, X_u = x_u, X_{t_1} = x\big) \\ &= \underline{E}_{\mathscr{P}}\big(g_1(x) + h_1(x) f_{2:n}(X_{t_{2:n}}) \,\big|\, X_u = x_u, X_{t_1} = x\big),\end{aligned}$$

where the second equality follows from Corollary 4.1158. For any jump process $P$ in $\mathscr{P}$, it follows from (ES3)37, (ES1)37 and (ES2)37 that

$$\begin{aligned}&E_P\big(g_1(x) + h_1(x) f_{2:n}(X_{t_{2:n}}) \,\big|\, X_u = x_u, X_{t_1} = x\big) \\ &\qquad = g_1(x) + h_1(x) E_P\big(f_{2:n}(X_{t_{2:n}}) \,\big|\, X_u = x_u, X_{t_1} = x\big).\end{aligned}$$

Because $h_1(x) \geq 0$, this implies that

$$\begin{aligned}&\underline{E}_{\mathscr{P}}(f(X_v) \mid X_u = x_u, X_{t_1} = x) \\ &\qquad = g_1(x) + h_1(x) \underline{E}_{\mathscr{P}}\big(f_{2:n}(X_{t_{2:n}}) \,\big|\, X_u = x_u, X_{t_1} = x\big). \end{aligned} \tag{4.24}$$

Now recall from Lemma 4.40200 – with $u \cup (t_1)$ here in the role of $u$ there and with $t_{2:n}$ here in the role of $v$ there – that

$$\underline{E}_{\mathscr{P}}\big(f_{2:n}(X_{t_{2:n}}) \,\big|\, X_u = x_u, X_{t_1} = x\big) = \big[e^{(t_2 - t_1)\underline{Q}} f_2\big](x).$$

We substitute this equality into Eq. (4.24), to yield

$$\underline{E}_{\mathscr{P}}(f(X_v) \mid X_u = x_u, X_{t_1} = x) = g_1(x) + h_1(x) \big[e^{(t_2 - t_1)\underline{Q}} f_2\big](x) = f_1(x),$$

as required. □

Finally, we use Lemmas 4.40200 and 4.41↶ in our – pretty convoluted – proof for Theorem 4.9166.

**Theorem 4.9.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, and a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows. Fix an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$, a sequence of time-points $v = (t_1, \dots, t_n)$ in $\mathscr{U}_{\neq ()}$ and a real variable $f$ in $\mathbb{V}(\Omega)$ that has a sum-product representation*

$$f(X_v) = \sum_{k=1}^{n} g_k(X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_\ell})$$

*over $v$. Let $f_1 \colon \mathscr{X} \to \mathbb{R}$ be recursively defined by the initial condition $f_n := g_n$ and, for all $k$ in $\{1, \dots, n-1\}$, by the recursive relation*

$$f_k \colon \mathscr{X} \to \mathbb{R} \colon x \mapsto \left[e^{(t_{k+1}-t_k)\underline{Q}_{\mathscr{Q}}} f_{k+1}\right](x) h_k(x) + g_k(x).$$

*Then for all $x$ in $\mathscr{X}$,*

$$\underline{E}_{\mathscr{P}}(f \mid X_{t_1} = x) = f_1(x),$$

*and for all $\{X_u = x_u\}$ in $\mathscr{H}$ such that $s := \max u \leq t_1$,*

$$\begin{aligned}\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) &= \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f \mid X_{t_1}) \,\big|\, X_u = x_u\big) \\ &= \begin{cases} f_1(x_s) & \text{if } u \neq () \text{ and } s = t_1 \\ \left[e^{(t_1 - s)\underline{Q}_{\mathscr{Q}}} f_1\right](x_s) & \text{if } u \neq () \text{ and } s < t_1, \\ \underline{E}_{\mathscr{M}}(f_1) & \text{if } u = () \text{ and } t_1 = 0, \\ \underline{E}_{\mathscr{M}}\big(e^{t_1 \underline{Q}_{\mathscr{Q}}} f_1\big) & \text{if } u = () \text{ and } t_1 > 0, \end{cases}\end{aligned}$$

*and therefore*

$$\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{M},\mathscr{Q}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}(f \mid X_u = x_u).$$

*Proof.* It follows immediately from Lemma 4.41₂₀₁ that for all $x$ in $\mathscr{X}$,

$$\underline{E}_{\mathscr{P}}(f \mid X_{t_1} = x) = f_1(x). \tag{4.25}$$

Next, we fix any state history $\{X_u = x_u\}$ in $\mathscr{H}$ such that $s := \max u \leq t_1$. First, let us deal with the case that $u \neq ()$ and $s = t_1$. Then it follows immediately from Lemma $4.41_{201}$ that

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = f_1(x_s).$$

One the other hand, for all $P$ in $\mathscr{P}$, it follows from Lemma $2.39_{36}$, $(\mathrm{ES1})_{37}$, Corollary $3.18_{71}$ and Eq. (4.25) that

$$f_1(x_s) = E_P(f_1(x_s) \mid X_u = x_u) = E_P(f_1(X_1) \mid X_u = x_u) = E_P\big(\underline{E}_{\mathscr{P}}(f \mid X_s) \,\big|\, X_u = x_u\big).$$

It now follows immediately from the preceding two equalities that

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f \mid X_s) \,\big|\, X_u = x_u\big) = f_1(x_s),$$

as required.

The remaining case that either $u \neq ()$ and $s < t_1$ or $u = ()$ is much more straightforward. In this case, it follows immediately from Lemma 4.40<sub>200</sub> that

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = \begin{cases} [e^{(t_1 - s)\underline{Q}_{\mathscr{Q}}} f_1](x_s) & \text{if } u \neq (), \\ \underline{E}_{\mathscr{M}}(f_1) & \text{if } u = () \text{ and } t_1 = 0, \\ \underline{E}_{\mathscr{M}}(e^{t_1 \underline{Q}_{\mathscr{Q}}} f_1) & \text{if } u = () \text{ and } t_1 > 0. \end{cases}$$

Because $f_1(X_{t_1}) = \underline{E}_{\mathscr{P}}(f \mid X_{t_1})$ due to Eq. (4.25), it follows from the preceding equality and Propositions $3.81_{117}$ to $3.83_{118}$ that

$$\underline{E}_{\mathscr{P}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f \mid X_{t_1}) \,\big|\, X_u = x_u\big),$$

as required. □

## 4.B Proofs of Proposition 4.11 and Theorem 4.12

This appendix contains proofs for the two results in Section $4.1.4_{170}$. First, we prove Proposition $4.11_{170}$.

**Proposition 4.11.** *Consider an imprecise jump process $\mathscr{P}$ that satisfies the law of iterated lower expectations. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, a sequence of time points $v = (t_1, \dots, t_n)$ in $\mathscr{U}_{>u}$ and a gamble $f$ on $\mathscr{X}_v$. Then*

$$\underline{E}_{\mathscr{P}}(f(X_v) \mid X_u = x_u) = \underline{E}_{\mathscr{P}}(f_1(X_{t_1}) \mid X_u = x_u),$$

*where $f_1 \colon \mathscr{X} \to \mathbb{R}$ is defined recursively: the initial condition is $f_n := f$, and for all $k$ in $\{1, \dots, n-1\}$, we let $t_{1:k} := (t_1, \dots, t_k)$ and let $f_k$ be the gamble on $\mathscr{X}_{t_{1:k}}$ defined by*

$$f_k(y_{t_{1:k}}) := \underline{E}_{\mathscr{P}}(f_{k+1}(y_{t_{1:k}}, X_{t_{k+1}}) \mid X_u = x_u, X_{t_{1:k}} = y_{t_{1:k}}) \quad \textit{for all } y_{t_{1:k}} \in \mathscr{X}_{t_{1:k}}.$$

*Proof.* Observe that in case $|v| = n = 1$, the statement follows immediately from Corollary $4.1_{158}$. Hence, from here on we assume that $n \geq 2$.

Fix some $k$ in $\{1, \dots, n-1\}$. It follows from the law of iterated lower expectations that

$$\underline{E}_{\mathscr{P}}(f(X_v) \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f(X_v) \mid X_u, X_{t_{1:k}}) \,\big|\, X_u = x_u\big).$$

Using Corollary $4.1_{158}$, we infer from this that

$$\underline{E}_{\mathscr{P}}(f(X_v) \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\big(\tilde{f}_k(X_{t_{1:k}}) \,\big|\, X_u = x_u\big),$$

where $\tilde{f}_k$ is the gamble on $\mathscr{X}_{t_{1:k}}$ defined by

$$\tilde{f}_k(y_{t_{1:k}}) := \underline{E}_{\mathscr{P}}(f(X_v) \mid X_u = x_u, X_{t_{1:k}} = y_{t_{1:k}}) \quad \text{for all } y_{t_{1:k}} \in \mathscr{X}_{t_{1:k}}.$$

To prove the statement, it clearly suffices to show that $\tilde{f}_k = f_k$ for all $k$ in $\{1, \dots, n-1\}$, with $f_k$ as defined in the statement. To this end, we observe that it follows from Corollary $4.1_{158}$ that for all $y_{t_{1:n-1}}$ in $\mathscr{X}_{t_{1:n-1}}$,

$$\begin{aligned} \tilde{f}_{n-1}(y_{t_{1:n-1}}) &= \underline{E}_{\mathscr{P}}(f(x_u, y_{t_{1:n-1}}, X_{t_n}) \mid X_u = x_u, X_{t_{1:n-1}} = y_{t_{1:n-1}}) \\ &= \underline{E}_{\mathscr{P}}(f_n(y_{t_{1:n-1}}, X_{t_n}) \mid X_u = x_u, X_{t_{1:n-1}} = y_{t_{1:n-1}}) = f_{n-1}(y_{t_{1:n-1}}). \end{aligned}$$

Thus, $\tilde{f}_{n-1} = f_{n-1}$. Second, we fix some $k$ in $\{1,\ldots,n-2\}$ and $y_{t_{1:k}}$ in $\mathscr{X}_{t_{1:k}}$. Then it follows from the law of iterated lower expectations and Corollary 4.1<sub>158</sub> that

$$\begin{aligned}\tilde{f}_k(y_{t_{1:k}}) &= \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(f(X_v)\,|\,X_u, X_{t_{1:k}}, X_{t_{k+1}})\,\big|\,X_u = x_u, X_{t_{1:k}} = y_{t_{1:k}}\big)\\ &= \underline{E}_{\mathscr{P}}\big(\tilde{f}_{k+1}(y_{t_{1:k}}, X_{t_{k+1}})\,\big|\,X_u = x_u, X_{t_{1:k}} = y_{t_{1:k}}\big).\end{aligned}$$

Because this equality holds for all $k$ in $\{1,\ldots,n-2\}$ and $y_{t_{1:k}}$ in $\mathscr{X}_{t_{1:k}}$, and because we have previously show that $\tilde{f}_{n-1} = f_{n-1}$, we conclude that $\tilde{f}_k = f_k$ for all $k$ in $\{1,\ldots,n-1\}$, as required. □

Second, we use Proposition 4.11<sub>170</sub> to prove Theorem 4.12<sub>170</sub>.

**Theorem 4.12.** *Consider a non-empty set $\mathscr{M}$ of initial probability mass functions, and a non-empty, bounded and convex set $\mathscr{Q}$ of rate operators that has separately specified rows. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, a sequence of time points $v = (t_1,\ldots,t_n)$ in $\mathscr{U}_{>u}$ and a gamble $f$ on $\mathscr{X}_v$. Then*

$$\underline{E}_{\mathscr{M},\mathscr{Q}}(f(X_v)\,|\,X_u = x_u) = \begin{cases} \big[e^{(t_1-\max u)\underline{Q}_{\mathscr{Q}}} f_1\big](x_{\max u}) & \text{if } u \neq (),\\ \underline{E}_{\mathscr{M}}(f_1) & \text{if } u = () \text{ and } t_1 = 0,\\ \underline{E}_{\mathscr{M}}\big(e^{(t_1-\max u)\underline{Q}_{\mathscr{Q}}} f_1\big) & \text{if } u = () \text{ and } t_1 > 0,\end{cases}$$

*where $f_1\colon \mathscr{X} \to \mathbb{R}$ is defined recursively: the initial condition is $f_n := f$, and for all $k$ in $\{1,\ldots,n-1\}$, we let $t_{1:k} := (t_1,\ldots,t_k)$ and let $f_k$ be the gamble on $\mathscr{X}_{t_{1:k}}$ defined by*

$$f_k(y_{t_{1:k}}) := \big[e^{(t_{k+1}-t_k)\underline{Q}_{\mathscr{Q}}} f_{k+1}(y_{t_{1:k}}, \bullet)\big](y_{t_k}) \quad \text{for all } y_{t_{1:k}} \in \mathscr{X}_{t_{1:k}}.$$

*Proof.* Because $\mathscr{Q}$ is bounded and convex and has separately specified rows by assumption, Theorem 3.88<sub>120</sub> guarantees that $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ satisfies the law of iterated lower expectations. Hence, the statement follows (almost) immediately from Proposition 4.11<sub>170</sub> and Propositions 3.81<sub>117</sub> to 3.83<sub>118</sub> – see also the the proof of Lemma 4.40<sub>200</sub>. □

## 4.C Some additional results regarding ergodicity

The main goal of this appendix is to prove Theorem 4.36<sub>194</sub>, and we will do so in Appendix 4.D<sub>210</sub>. First, however, we take a closer look at ergodic lower transition operators in Appendix 4.C.1↷ and upper reachability with respect to lower rate operator in Appendix 4.C.2<sub>209</sub>.

Throughout this appendix, we will repeatedly make use of the following almost trivial observation.

**Lemma 4.42.** *Consider a lower transition operator $\underline{T}$. Then for any subset $A$ of $\mathscr{X}$,*

$$\underline{T}\mathbb{I}_A = 1 - \overline{T}\mathbb{I}_{A^c} \quad \text{and} \quad \overline{T}\mathbb{I}_A = 1 - \underline{T}\mathbb{I}_{A^c}.$$

*Proof.* Observe that the second equality follows from the first. The first equality holds because due to (LT5)<sub>108</sub>,

$$\underline{T}\mathbb{I}_A = -\overline{T}(-\mathbb{I}_A) = 1 - \overline{T}(1-\mathbb{I}_A) = 1 - \overline{T}\mathbb{I}_{A^c}.$$ □

### 4.C.1 A qualitative study of ergodicity

We take a closer look at ergodic lower transition operators from the qualitative point of view of Hermans et al. (2012). Recall from Proposition 4.27$_{189}$ that a lower transition operator $\underline{T}$ is ergodic if and only if it is regularly absorbing, meaning that it is top class regular and top class absorbing.

The observant reader might have noticed that our definitions of top class regularity and top class absorption in Proposition 4.27$_{189}$ differ slightly from those of Hermans et al. (2012, Proposition 3), but they are actually entirely equivalent. For top class regularity, we demand that there is some $n$ in $\mathbb{N}$ such that $\overline{T}^n\mathbb{I}_x \gg 0$. By (LT6)$_{108}$, it then holds that $\overline{T}^k\mathbb{I}_x \gg 0$ for any $k \geq n$, which is what Hermans et al. (2012, (TCR)) demand. For top class absorption, Hermans et al. (2012, (TCA)) demand that

$$\big(\forall y \in \mathscr{X} \setminus \mathscr{X}_{\underline{T}}\big)\big(\exists n \in \mathbb{N}\big)\ [\overline{T}^n\mathbb{I}_A](y) < 1,$$

where $A = \mathscr{X} \setminus \mathscr{X}_{\underline{T}}$. Recall from Lemma 4.42$_{\curvearrowleft}$ that $[\overline{T}^n\mathbb{I}_A](y) = 1 - [\underline{T}^n\mathbb{I}_{\mathscr{X}_{\underline{T}}}](y)$, so their demand is equivalent to ours.

The following two technical lemmas are related to ergodicity, and will come in handy in the proof of Theorem 4.36$_{194}$; the first is taken from (Hermans et al., 2012, Proposition 4).

**Lemma 4.43.** *Consider a lower transition operator $\underline{T}$, a natural number $n$ and two states $x$ and $y$ in $\mathscr{X}$. Then $[\overline{T}^n\mathbb{I}_x](y) > 0$ if and only if there is a sequence $(x_0,\dots,x_n)$ in $\mathscr{X}$ with $x_0 = y$ and $x_n = x$ such that $[\overline{T}\mathbb{I}_{x_k}](x_{k-1}) > 0$ for all $k$ in $\{1,\dots,n\}$.*

**Lemma 4.44.** *If the lower transition operator $\underline{T}$ is top class regular, then for any state $x$ in the top class $\mathscr{X}_{\underline{T}}$, any state $y$ in $\mathscr{X}_{\underline{T}}^{\mathrm{c}} = \mathscr{X} \setminus \mathscr{X}_{\underline{T}}$ and any natural number $k$ in $\mathbb{N}$,*

$$\big[\overline{T}^k\mathbb{I}_y\big](x) = 0 \qquad \text{and} \qquad \big[\underline{T}^k\mathbb{I}_{\mathscr{X}_{\underline{T}}}\big](x) = 1.$$

*Proof.* First we prove the first equality. Our proof will be one by contradiction, so we assume *ex absurdo* that there is some $k$ in $\mathbb{N}$ such that $c_x := [\overline{T}^k\mathbb{I}_y](x) \neq 0$. Note that $\overline{T}^k\mathbb{I}_y \geq 0$ due to (LT4)$_{108}$, so $c_x > 0$. Because $x$ is in the top class $\mathscr{X}_{\underline{T}}$, there is a natural number $n_x$ such that

$$\big[\overline{T}^{n_x}\mathbb{I}_x\big](z) > 0 \quad \text{for all } z \in \mathscr{X}. \tag{4.26}$$

Observe that $\overline{T}^k\mathbb{I}_y \geq c_x\mathbb{I}_x$. From this, (LT6)$_{108}$ and (LT2)$_{107}$, we infer that for all $z$ in $\mathscr{X}$,

$$\big[\overline{T}^{k+n_x}\mathbb{I}_y\big](z) = \big[\overline{T}^{n_x}\overline{T}^k\mathbb{I}_y\big](z) \geq \big[\overline{T}^{n_x}(c_x\mathbb{I}_x)\big](z) = c_x\big[\overline{T}^{n_x}\mathbb{I}_x\big](z) > 0,$$

where for the last inequality we have used that $c_x > 0$ and Eq. (4.26). However, because this inequality holds for all $z$ in $\mathscr{X}$, we infer that $y$ belongs to the top class $\mathscr{X}_{\underline{T}}$, which contradicts the initial condition on $y$.

Next, we prove the second statement. Recall from Lemma 4.42$_{205}$ that

$$\underline{T}^k \mathbb{I}_{\mathscr{X}_{\underline{T}}} = 1 - \overline{T}^k \mathbb{I}_{\mathscr{X}_{\underline{T}}^{\mathrm{c}}}.$$

From the conjugacy of $\underline{T}^k$ and $\overline{T}^k$ and (LT3)$_{107}$, it furthermore follows that

$$\overline{T}^k \mathbb{I}_{\mathscr{X}_{\underline{T}}^{\mathrm{c}}} = \overline{T}^k \left( \sum_{z \in \mathscr{X}_{\underline{T}}^{\mathrm{c}}} \mathbb{I}_z \right) \leq \sum_{z \in \mathscr{X}_{\underline{T}}^{\mathrm{c}}} \overline{T}^k \mathbb{I}_z.$$

From the – already proven – first equality of the statement, we know that $\sum_{z \in \mathscr{X}_{\underline{T}}^{\mathrm{c}}} [\overline{T}^k \mathbb{I}_z](x) = 0$, whence

$$\left[\underline{T}^k \mathbb{I}_{\mathscr{X}_{\underline{T}}}\right](x) = 1 - \left[\overline{T}^k \mathbb{I}_{\mathscr{X}_{\underline{T}}^{\mathrm{c}}}\right](x) \geq 1 - \sum_{z \in \mathscr{X}_{\underline{T}}^{\mathrm{c}}} \left[\overline{T}^k \mathbb{I}_z\right](x) = 1.$$

Note that by (LT4)$_{108}$, $[\underline{T}^k \mathbb{I}_{\mathscr{X}_{\underline{T}}}](x) \leq 1$. By combining the two preceding inequalities, we find that the the second equality of the statement holds: $[\underline{T}^k \mathbb{I}_{\mathscr{X}_{\underline{T}}}](x) = 1$. □

The condition in Proposition 4.27$_{189}$ for top class absorption is in a form that is not easily verified. Luckily for us, Hermans et al. (2012, Proposition 6) give an equivalent condition that is more straightforward; the following lemma establishes this condition in a form that is tailored to how we will use it.

**Lemma 4.45.** *Let $\underline{T}$ be a lower transition operator that is top class regular. Then $\underline{T}$ is top class absorbing if and only if $B_n = \mathscr{X}$, where $(B_k)_{k \in \mathbb{Z}_{\geq 0}}$ is the sequence defined by the initial condition $B_0 := \mathscr{X}_{\underline{T}}$ and, for all $k$ in $\mathbb{Z}_{\geq 0}$, by the recursive relation*

$$B_{k+1} := B_k \cup \left\{ x \in \mathscr{X} \setminus B_k \colon \left[\underline{T} \mathbb{I}_{B_k}\right](x) > 0 \right\} = \left\{ x \in \mathscr{X} \colon \left[\underline{T} \mathbb{I}_{B_k}\right](x) > 0 \right\},$$

*and where $n \leq |\mathscr{X} \setminus \mathscr{X}_{\underline{T}}|$ is the first index such that $B_n = B_{n+1}$.*

*Proof.* Let $\underline{T}$ be a top class regular lower transition operator with top class $\mathscr{X}_{\underline{T}}$. By (Hermans et al., 2012, Proposition 6), $\underline{T}$ is top class absorbing if and only if $A_n = \emptyset$, where $A_n$ is the set determined by the initial condition $A_0 := \mathscr{X} \setminus \mathscr{X}_{\underline{T}}$ and, for all $k$ in $\mathbb{Z}_{\geq 0}$, by the recursive relation

$$A_{k+1} := \left\{ x \in A_k \colon \left[\overline{T} \mathbb{I}_{A_k}\right](x) = 1 \right\},$$

and where $n \leq |\mathscr{X} \setminus \mathscr{X}_{\underline{T}}|$ is the first index for which $A_n = A_{n+1}$. Observe that for all $k$ in $\mathbb{Z}_{\geq 0}$, $A_{k+1} \subseteq A_k$ by construction.

We now let $B'_k := \mathscr{X} \setminus A_k$ for all $k$ in $\mathbb{Z}_{\geq 0}$. Note that $B'_k = B'_{k+1}$ if and only if $A_k = A_{k+1}$. Consequently, $\underline{T}$ is top class absorbing if and only if $B'_n = \mathscr{X}$, where $n$ is the smallest index for which $B'_n = B'_{n+1}$. Thus, the statement follows if we can shown that $B'_k = B_k$ for all $k$ in $\mathbb{Z}_{\geq 0}$.

Note that both $B'_0$ and $B_0$ are equal to $\mathscr{X}_{\underline{T}}$, so $B'_k = B_k$ for $k = 0$. Thus, it remains for us to show that $B'_{k+1} = B_{k+1}$ for all $k$ in $\mathbb{Z}_{\geq 0}$. Fix any such $k$, and observe that $0 \leq \underline{T}\mathbb{I}_{A_k} \leq 1$ due to (LT4)$_{108}$; therefore,

$$\begin{aligned} B'_{k+1} = \mathscr{X} \setminus A_{k+1} &= \big(\mathscr{X} \setminus A_k\big) \cup \Big\{x \in A_k \colon \big[\overline{T}\mathbb{I}_{A_k}\big](x) < 1\Big\} \\ &= B'_k \cup \Big\{x \in A_k \colon \big[\overline{T}\mathbb{I}_{A_k}\big](x) < 1\Big\}. \end{aligned}$$

From this and Lemma 4.42$_{205}$, it now follows that

$$B'_{k+1} = B'_k \cup \Big\{x \in A_k \colon \big[\underline{T}\mathbb{I}_{\mathscr{X}\setminus A_k}\big](x) > 0\Big\} = B'_k \cup \Big\{x \in \mathscr{X} \setminus B'_k \colon \big[\underline{T}\mathbb{I}_{B'_k}\big](x) > 0\Big\}.$$

This verifies that $B'_{k+1} = B_{k+1}$, as required.

Finally, we verify the second equality of the statement. For starters, we observe that $B_k \subseteq B_{k+1}$ for all $k$ in $\mathbb{Z}_{\geq 0}$, so $\underline{T}\mathbb{I}_{B_k} \leq \underline{T}\mathbb{I}_{B_{k+1}}$ due to (LT6)$_{108}$. Furthermore, it follows from Lemma 4.44$_{206}$ and (LT4)$_{108}$ that

$$\mathbb{I}_{B_0} = \mathbb{I}_{\mathscr{X}_{\underline{T}}} \leq \underline{T}\mathbb{I}_{\mathscr{X}_{\underline{T}}} = \underline{T}\mathbb{I}_{B_0} \quad \text{for all } x \in \mathscr{X}_{\underline{T}} = B_0.$$

Due to these two observations, we conclude that $[\underline{T}\mathbb{I}_{B_k}](x) > 0$ for all $x$ in $B_k$ and $k$ in $\mathbb{Z}_{\geq 0}$. Clearly, this verifies the second equality of the statement. □

We conclude this section on ergodic lower transition operators with a technical lemma that is related to Lemma 4.45.

**Lemma 4.46.** *Consider a lower transition operator $\underline{T}$, a natural number $n$ and a subset $A$ of $\mathscr{X}$. Then*

$$c_1 \cdots c_n \mathbb{I}_{A_n} \leq \underline{T}^k \mathbb{I}_A \leq \mathbb{I}_{A_n},$$

*where $A_n \subseteq \mathscr{X}$ is derived from the initial condition $A_0 := A$ and the recursive relation*

$$A_k := \big\{x \in \mathscr{X} \colon \big[\underline{T}\mathbb{I}_{A_{k-1}}\big](x) > 0\big\} \quad \textit{for all } k \in \{1, \dots, n\},$$

*and the non-negative real numbers $c_1, \dots, c_n$ are defined as*

$$c_k := \min\Big\{\big[\underline{T}\mathbb{I}_{A_{k-1}}\big](x) \colon x \in A_k\Big\} \quad \textit{for all } k \in \{1, \dots, n\},$$

*with the convention that the minimum of an empty set is zero. Furthermore, $A_n \neq \varnothing$ if and only if $c_k > 0$ for all $k$ in $\{1, \dots, n\}$.*

*Proof.* Fix any $k$ in $\{1, \dots, n\}$. On the one hand, it follows from (LT4)$_{108}$ that $\underline{T}\mathbb{I}_{A_{k-1}} \leq \mathbb{I}_{A_k}$. On the other hand, $\underline{T}\mathbb{I}_{A_{k-1}} \geq c_k \mathbb{I}_{A_k}$, with $0 \leq c_k \leq 1$ due to (LT4)$_{108}$. In summary, $c_k \mathbb{I}_{A_k} \leq \underline{T}\mathbb{I}_{A_{k-1}} \leq \mathbb{I}_{A_k}$. We now repeatedly use these inequalities in combination with (LT2)$_{107}$ and (LT6)$_{108}$, to yield

$$\underline{T}^n \mathbb{I}_A = \underline{T}^{n-1}(\underline{T}\mathbb{I}_A) \geq c_1 \underline{T}^{n-1}\mathbb{I}_{A_1} \geq \cdots \geq c_1 \cdots c_n \mathbb{I}_{A_n}$$

and

$$\underline{T}^n \mathbb{I}_A = \underline{T}^{n-1}(\underline{T}\mathbb{I}_A) \leq \underline{T}^{n-1}\mathbb{I}_{A_1} \leq \cdots \leq \mathbb{I}_{A_n}.$$

This verifies the first part of the statement.

To prove the second part of the statement, we observe that $c_k = 0$ if and only if $A_k = \emptyset$. Thus, it is clear that $A_n \neq \emptyset$ if $c_k > 0$ for all $k$ in $\{1,\dots,n\}$. To verify that the converse implication, we assume that $c_k = 0$ for some $k$ in $\{1,\dots,n\}$. Because then $A_k = \emptyset$, it follows from (LT4)$_{108}$ that $\underline{T}\mathbb{I}_{A_k} = 0$, whence $A_{k+1} = \emptyset$. The same argument proves that $A_\ell = \emptyset$ for all $\ell$ in $\{k+1,\dots,n\}$ as well, and in particular also for $\ell = n$. $\square$

### 4.C.2 Upper reachability

We now turn to ergodic lower rate operators, and more precisely to upper reachability. Let us start with a result that follows more or less immediately from Definition 4.31$_{192}$.

**Lemma 4.47.** *Let $\underline{Q}$ be a lower rate operator, and fix any two distinct states $x$ and $y$ in $\mathscr{X}$ such that $x \neq y$. Then $y \overline{\hookrightarrow} x$ if and only if there is a sequence $(x_0,\dots,x_n)$ in $\mathscr{X}$ with $n$ in $\mathbb{N}$, $x_0 = y$ and $x_n = x$ in which every state occurs at most once and such that $[\overline{Q}\mathbb{I}_{x_k}](x_{k-1}) > 0$ for all $k$ in $\{1,\dots,n\}$. Consequently, $n < |\mathscr{X}|$.*

*Proof.* The direct implication follows almost immediately from Definition 4.31$_{192}$. To prove it, we suppose that $y \overline{\hookrightarrow} x$, and observe that by Definition 4.31$_{192}$, there is some sequence $(x_0,\dots,x_n)$ in $\mathscr{X}$ with $n$ in $\mathbb{Z}_{\geq 0}$, $x_0 = y$ and $x_n = x$ such that for all $k$ in $\{1,\dots,n\}$, $[\overline{Q}\mathbb{I}_{x_k}](x_{k-1}) > 0$. Assume that there is a state $z$ in $\mathscr{X}$ that occurs more than once in this sequence. Then we can simply delete every element of the sequence from right after the the first occurrence of $z$ up to and including the last occurrence of $z$, and still have a valid sequence. If we continue this way, then we end up with a sequence in which every state occurs at most once. As every state occurs at most once, the length $n+1$ of the sequence is lower than or equal to $|\mathscr{X}|$. Consequently, $n < |\mathscr{X}|$. Furthermore, since the deletions we executed do not alter the first and the last element of the sequence, and since $x \neq y$, we have that $n \neq 0$ – that is, $n$ is a natural number.

The converse implication holds because the requirements of Definition 4.31$_{192}$ are trivially satisfied. $\square$

In our proof of Theorem 4.36$_{194}$, we will use the previous lemma in the following slightly different form.

**Lemma 4.48.** *Let $\underline{Q}$ be a lower rate operator, and fix any two states $x$ and $y$ in $\mathscr{X}$ such that $y \overline{\hookrightarrow} x$. Then there is a non-negative integer $n < |\mathscr{X}|$ such that for all natural numbers $k \geq n$ and all $\Delta$ in $\mathbb{R}_{>0}$ with $\Delta\|\underline{Q}\| < 2$, there is a sequence $(x_0,\dots,x_k)$ in $\mathscr{X}$ with $x_0 = y$ and $x_k = x$ such that*

$$\left[(I+\Delta\overline{Q})\mathbb{I}_{x_\ell}\right](x_{\ell-1}) > 0 \quad \textit{for all } \ell \in \{1,\dots,k\}.$$

*Proof.* We first consider the special case that $x = y$. For all $\Delta \in \mathbb{R}_{>0}$ such that $\Delta\|\underline{Q}\| < 2$,

$$\left[(I+\Delta\overline{Q})\mathbb{I}_x\right](x) = \mathbb{I}_x(x) + \Delta\left[\overline{Q}\mathbb{I}_x\right](x) = 1 + \Delta\left[\overline{Q}\mathbb{I}_x\right](x) > 0, \tag{4.27}$$

where the inequality follows from (LR5)$_{111}$, (LR7)$_{111}$ and the requirement that $\Delta\|\underline{Q}\| < 2$. Consequently, the sequence $(x_0,\ldots,x_k)$ with $x_\ell = x$ for all $k$ in $\{1,\ldots,k\}$ is an example of a sequence that satisfies the condition of the statement.

Next, we consider the case that $y \neq x$. From Lemma 4.47, we know that there is a sequence $S_y := (x_0,\ldots,x_n)$ in $\mathscr{X}$ with $n$ in $\mathbb{N}$, $x_0 = y$ and $x_n = x$ in which every state occurs at most once – so $n < |\mathscr{X}|$ – and such that $[\overline{Q}\mathbb{I}_{x_\ell}](x_{\ell-1}) > 0$ for all $\ell$ in $\{1,\ldots,n\}$. Fix an arbitrary natural number $k \geq n$ and a step size $\Delta$ in $\mathbb{R}_{>0}$ with $\Delta\|\underline{Q}\| < 2$. Note that for all $\ell$ in $\{1,\ldots,n\}$,

$$0 < \Delta[\overline{Q}\mathbb{I}_{x_\ell}](x_{\ell-1}) = \mathbb{I}_{x_\ell}(x_{\ell-1}) + \Delta[\overline{Q}\mathbb{I}_{x_\ell}](x_{\ell-1}) = [(I+\Delta\overline{Q})\mathbb{I}_{x_\ell}](x_{\ell-1}),$$

where the inequality holds because $0 < \Delta$ and $[\overline{Q}\mathbb{I}_{x_\ell}](x_{\ell-1}) > 0$, and the first equality holds because $x_\ell \neq x_{\ell-1}$. Also, from Eq. (4.27)↶ we know that $[(I+\Delta\overline{Q})\mathbb{I}_x](x) > 0$. Hence, appending the sequence $S_y$ with $(k-n)$ times $x$ yields a sequence $(x_0,\ldots,x_k)$ in $\mathscr{X}$ with $x_0 = y$ and $x_k = x$ such that $[(I+\Delta\overline{Q})\mathbb{I}_{x_\ell}](x_{\ell-1}) > 0$ for all $\ell$ in $\{1,\ldots,k\}$, as required. □

## 4.D Proof of Theorem 4.36

With the help of the technical lemmas of the two preceding sections, we can prove Theorem 4.36$_{194}$, which we repeat here for good measure.

**Theorem 4.36.** *A lower rate operator $\underline{Q}$ is ergodic if and only if there is some natural number $n < |\mathscr{X}|$ such that for some (and then all) natural number(s) $k \geq n$ and some (and then all) step size(s) $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\|\underline{Q}\|_{\mathrm{op}} < 2$:*

$$\rho\big((I+\Delta\underline{Q})^k\big) < 1.$$

*Proof.* First, we prove the direct implication. To this end, we assume that $\underline{Q}$ is ergodic, let $n := |\mathscr{X}| - 1$ and fix some natural number $k \geq n$ and a step size $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\|\underline{Q}\| < 2$. Note that by construction, $\underline{T} := (I+\Delta\underline{Q})$ is a lower transition operator due to Lemma 3.72$_{112}$.

For the direct implication, we need to prove that $\rho(\underline{T}^k) < 1$. We will provide a proof by contradiction, so we assume *ex absurdo* that $\rho(\underline{T}^k) = 1$. Due to Eq. (4.19)$_{190}$ and (LT12)$_{179}$, there is some $f$ in $\mathbb{G}(\mathscr{X})$ with $\min f = 0$ and $\max f = 1$ such that $\|\underline{T}^k f\|_{\mathrm{v}} = 1$. It furthermore follows from (LT4)$_{108}$ that there are states $y_0$ and $y_1$ in $\mathscr{X}$ such that $[\underline{T}^k f](y_0) = 0$ and $[\underline{T}^k f](y_1) = 1$.

We define the – obviously non-empty – set

$$\{f = 0\} := \big\{x \in \mathscr{X} \colon f(x) = 0\big\} \ni y_0,$$

and distinguish two cases: either $\mathscr{X}_{\underline{Q}} \cap \{f = 0\} \neq \emptyset$ or $\mathscr{X}_{\underline{Q}} \cap \{f = 0\} = \emptyset$.

First, we consider the case $\mathscr{X}_{\underline{Q}} \cap \{f = 0\} \neq \emptyset$, and fix any arbitrary element $x_0$ in this intersection. Note that, by construction, $\mathbb{I}_{x_0} \leq 1 - f$. Using the conjugacy of $\underline{T}^k$ and $\overline{T}^k$ and (LT6)$_{108}$, we find that

$$\overline{T}^k\mathbb{I}_{x_0} \leq \overline{T}^k(1-f) = 1 + \overline{T}^k(-f) = 1 - \underline{T}^k f,$$

where for the first equality we have used (LT5)$_{108}$. From the previous inequality and (LT4)$_{108}$, it follows that

$$0 \leq [\overline{T}^k \mathbb{I}_{x_0}](y_1) \leq 1 - [\underline{T}^k f](y_1) = 0,$$

and hence $[\overline{T}^k \mathbb{I}_{x_0}](y_1) = 0$. Because $x_0$ is a state in the top class $\mathscr{X}_{\underline{Q}}$, it follows from Proposition 4.33$_{193}$ that $y_1 \hookrightarrow x_0$. Seeing that $k \geq n = |\mathscr{X}| - 1$, it therefore follows from Lemma 4.48$_{209}$ that there is a sequence $(z_0, \ldots, z_k)$ in $\mathscr{X}$ with $z_0 = y_1$ and $z_k = x_0$ such that $[\overline{T} \mathbb{I}_{z_\ell}](z_{\ell-1}) > 0$ for all $\ell$ in $\{1, \ldots, k\}$. Because of Lemma 4.43$_{206}$, this implies that $[\overline{T}^k \mathbb{I}_{x_0}](y_1) > 0$, contradicting our earlier finding that $[\overline{T}^k \mathbb{I}_{x_0}](y_1) = 0$. Thus, we have established that $\rho(\underline{T}^k) < 1$ in case $\mathscr{X}_{\underline{Q}} \cap \{f = 0\} \neq \varnothing$.

Next, we consider the case $\mathscr{X}_{\underline{Q}} \cap \{f = 0\} = \varnothing$. We let $c := \min\{f(x) \colon x \in \mathscr{X}_{\underline{Q}}\} > 0$ – note that due to Proposition 4.33$_{193}$, $\mathscr{X}_{\underline{Q}} \neq \varnothing$ because $\underline{Q}$ is ergodic – and observe that $c\mathbb{I}_{\mathscr{X}_{\underline{Q}}} \leq f$. Thus, using (LT2)$_{107}$ and (LT6)$_{108}$, we find that $c\underline{T}^k \mathbb{I}_{\mathscr{X}_{\underline{Q}}} \leq \underline{T}^k f$. From Lemma 4.46$_{208}$, we furthermore know that

$$c_1 \cdots c_k \mathbb{I}_{A_k} \leq \underline{T}^k \mathbb{I}_{\mathscr{X}_{\underline{Q}}},$$

where $A_0 := \mathscr{X}_{\underline{Q}}$ and, for all $\ell$ in $\{1, \ldots, k\}$,

$$A_\ell := \Big\{x \in \mathscr{X} \colon [\underline{T} \mathbb{I}_{A_{\ell-1}}](x) > 0\Big\} \qquad \text{and} \qquad c_\ell := \min\Big\{[\underline{T} \mathbb{I}_{A_{\ell-1}}](x) \colon x \in A_\ell\Big\} \geq 0.$$

Combining the two obtained inequalities yields

$$c c_1 \cdots c_k \mathbb{I}_{A_k}(y_0) \leq c[\underline{T}^k \mathbb{I}_{\mathscr{X}_{\underline{Q}}}](y_0) \leq [\underline{T}^k f](y_0) = 0.$$

Because by the second part of Lemma 4.46$_{208}$ either $A_k = \varnothing$ or $c_\ell > 0$ for all $\ell$ in $\{1, \ldots, k\}$, we infer from this inequality that $y_0 \notin A_k$.

We now prove that $A_k = \mathscr{X}$, which contradicts the previous because $y_0 \in \mathscr{X} = A_k$. To this end, observe that for all $\ell$ in $\{0, \ldots, k-1\}$,

$$A_{\ell+1} = \big\{x \in A_\ell \colon [(I + \Delta \underline{Q})\mathbb{I}_{A_\ell}](x) > 0\big\} \cup \big\{x \in \mathscr{X} \setminus A_\ell \colon [(I + \Delta \underline{Q})\mathbb{I}_{A_\ell}](x) > 0\big\}.$$

Note that for all $x$ in $A_\ell$, $\mathbb{I}_{A_\ell} \geq \mathbb{I}_x$. Thus, it follows from (LT6)$_{108}$ that $(I + \Delta \underline{Q})\mathbb{I}_{A_\ell} \geq (I + \Delta \underline{Q})\mathbb{I}_x$, and more particularly that

$$\big[(I + \Delta \underline{Q})\mathbb{I}_{A_\ell}\big](x) \geq 1 + \Delta\big[\underline{Q}\mathbb{I}_x\big](x) > 0 \quad \text{for all } x \in A_\ell,$$

where the inequality follows from (LR7)$_{111}$ because $\Delta \|\underline{Q}\| < 2$. Additionally, we observe that for all $x$ in $\mathscr{X} \setminus A_\ell$,

$$\big[(I + \Delta \underline{Q})\mathbb{I}_{A_\ell}\big](x) = \Delta\big[\underline{Q}\mathbb{I}_{A_\ell}\big](x) > 0 \Leftrightarrow \big[\underline{Q}\mathbb{I}_{A_\ell}\big](x) > 0,$$

where we have used that $\Delta$ is a positive real number. Combining these two observations, we see that for all $\ell$ in $\{0, \ldots, k-1\}$,

$$A_{\ell+1} = A_\ell \cup \Big\{x \in \mathscr{X} \setminus A_\ell \colon \big[\underline{Q}\mathbb{I}_{A_\ell}\big](x) > 0\Big\}.$$

From this recursive relation, it is obvious that $A_0, \ldots, A_k$ is equal to the first $(k+1)$ terms of the sequence $\{B_\ell\}_{\ell \in \mathbb{Z}_{\geq 0}}$ that is defined in Definition 4.32$_{193}$ for $B_0 = \mathscr{X}_{\underline{Q}}$.

Because $\underline{Q}$ is ergodic and $k \geq |\mathscr{X}|-1 \geq |\mathscr{X}\setminus\mathscr{X}_{\underline{Q}}|$, it follows from Definition 4.32$_{193}$ and Proposition 4.33$_{193}$ that $A_k = B_k = \mathscr{X}$. In particular, $y_0$ is an element of $A_k$, and this contradicts what we have previously found. Thus, $\rho(\underline{T}^k)<1$ in case $\mathscr{X}_{\underline{Q}}\cap\{f=0\}=\emptyset$ as well, as required.

Next, we prove the converse implication. More precisely, we assume there is a natural number $k$ and a step size $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\|\underline{Q}\|<2$ and $\rho(\underline{T}^k)<1$, where $\underline{T} := (I+\Delta\underline{Q})$ is a lower transition operator. It now follows from Theorem 4.28$_{191}$ that $\underline{T}$ is ergodic, and then from Proposition 4.27$_{189}$ that $\underline{T}$ is regularly absorbing, meaning that

(i) $\mathscr{X}_{\underline{T}} := \big\{x\in\mathscr{X} \colon (\exists m\in\mathbb{N})(\forall y\in\mathscr{X})\ \big[\overline{T}^m\mathbb{I}_x\big](y)>0\big\}\neq\emptyset$ and

(ii) $(\forall y\in\mathscr{X}\setminus\mathscr{X}_{\underline{T}})(\exists m\in\mathbb{N})\ \big[\underline{T}^m\mathbb{I}_{\mathscr{X}_{\underline{T}}}\big](y)>0$.

We now use this and Proposition 4.33$_{193}$ to show that $\underline{Q}$ is ergodic. First, we show that $\mathscr{X}_{\underline{Q}}=\mathscr{X}_{\underline{T}}$, and we will do this by showing that $\mathscr{X}_{\underline{T}}\subseteq\mathscr{X}_{\underline{Q}}$ and $\mathscr{X}_{\underline{Q}}\subseteq\mathscr{X}_{\underline{T}}$.

Fix any $x$ in $\mathscr{X}_{\underline{T}}$ and $y$ in $\mathscr{X}$. By (i), there is a natural number $m$ such that $[\overline{T}^m\mathbb{I}_x](y)>0$. Due to Lemma 4.43$_{206}$, there is a sequence $(x_0,\dots,x_m)$ in $\mathscr{X}$ with $x_0=y$ and $x_m=x$ such that $[\overline{T}\mathbb{I}_{x_\ell}](x_{\ell-1})>0$ for all $\ell$ in $\{1,\dots,m\}$. Without loss of generality, we may assume that $x_\ell\neq x_{\ell-1}$ for all $\ell$ in $\{1,\dots,m\}$. If this is not the case, then we simply shorten the sequence by replacing every instance of consecutive equal consecutive entries with a single entry; note that if all states in the sequence $(x_0,\dots,x_m)$ are equal to $x_0$, then we end up with the monuple $(x_0)$ with $m=0$. Then for all $\ell$ in $\{1,\dots,m\}$,

$$0<\big[\overline{T}\mathbb{I}_{x_\ell}\big](x_{\ell-1})=\big[(I+\Delta\overline{Q})\mathbb{I}_{x_\ell}\big](x_{\ell-1})=\mathbb{I}_{x_\ell}(x_{\ell-1})+\Delta\big[\overline{Q}\mathbb{I}_{x_\ell}\big](x_{\ell-1})=\Delta\big[\overline{Q}\mathbb{I}_{x_\ell}\big](x_{\ell-1}).$$

Because $\Delta>0$, we infer from this strict inequality that $(x_0,\dots,x_m)$ is a sequence in $\mathscr{X}$ with $x_0=y$ and $x_m=x$ such that $[\overline{Q}\mathbb{I}_{x_\ell}](x_{\ell-1})>0$ for all $\ell$ in $\{1,\dots,m\}$. By Definition 4.31$_{192}$, this means that $y\hookrightarrow x$; because $y$ is an arbitrary state in $\mathscr{X}$, we have shown that $x$ is in the top class $\mathscr{X}_{\underline{Q}}$. Even more, because $x$ was an arbitrary element in $\mathscr{X}_{\underline{T}}$, this shows that $\mathscr{X}_{\underline{T}}\subseteq\mathscr{X}_{\underline{Q}}$.

To verify that $\mathscr{X}_{\underline{Q}}\subseteq\mathscr{X}_{\underline{T}}$ as well, we fix an arbitrary state $x$ in $\mathscr{X}_{\underline{Q}}$ and let $m := |\mathscr{X}|-1$. Note that $\mathscr{X}_{\underline{Q}}=\mathscr{X}_{\underline{T}}=\mathscr{X}$ whenever $|\mathscr{X}|=1$, so we may assume without loss of generality that $m=|\mathscr{X}|-1\geq 1$. It follows from the definition of $\mathscr{X}_{\underline{Q}}$ and Lemma 4.48$_{209}$ that for any state $y$ in $\mathscr{X}$, there is a sequence $(x_0,\dots,x_m)$ in $\mathscr{X}$ with $x_0=y$ and $x_m=x$ such that

$$0<\big[(I+\Delta\overline{Q})\mathbb{I}_{x_\ell}\big](x_{\ell-1})=\big[\overline{T}\mathbb{I}_{x_\ell}\big](x_{\ell-1})\quad\text{for all } \ell\in\{1,\dots,m\}.$$

Therefore, it follows from Lemma 4.43$_{206}$ that $[\overline{T}^m\mathbb{I}_x](y)>0$ for all $y$ in $\mathscr{X}$; by (i), this implies that $x$ belongs to the top class $\mathscr{X}_{\underline{T}}$. Since $x$ was an arbitrary state in $\mathscr{X}_{\underline{Q}}$, this verifies that $\mathscr{X}_{\underline{Q}}\subseteq\mathscr{X}_{\underline{T}}$.

To summarise, we have shown that $\mathscr{X}_{\underline{Q}}=\mathscr{X}_{\underline{T}}$. For this reason, it follows from (i) that

$$\mathscr{X}_{\underline{Q}}=\mathscr{X}_{\underline{T}}\neq\emptyset,$$

which settles the first condition of Proposition 4.33$_{193}$.

To prove that $\underline{Q}$ also satisfies the second condition of Proposition 4.33$_{193}$, we fix any $y$ in $\mathscr{X}\setminus\mathscr{X}_{\underline{Q}}=\mathscr{X}\setminus\mathscr{X}_{\underline{T}}$, and prove that $y\rightharpoondown\mathscr{X}_{\underline{Q}}$. Since $\underline{T}$ is ergodic, we know from

Proposition 4.27<sub>189</sub> that it is top class regular and top class absorbing. It therefore follows from Lemma 4.45<sub>207</sub> that $B_{\ell^\star} = \mathscr{X}$, where $(B_\ell)_{\ell\in\mathbb{Z}_{\geq 0}}$ is the sequence that is derived from the initial condition $B_0 \coloneqq \mathscr{X}_{\underline{T}}$ and the recursive relation

$$B_{\ell+1} \coloneqq B_\ell \cup \big\{x \in \mathscr{X} \setminus B_\ell \colon \big[\underline{T}\mathbb{I}_{B_\ell}\big](x) > 0\big\} \quad \text{for all } \ell \in \mathbb{Z}_{\geq 0},$$

and where $\ell^\star$ is the first index such that $B_{\ell^\star} = B_{\ell^\star+1}$. Observe that $B_0 = \mathscr{X}_{\underline{Q}}$ and that for all $\ell$ in $\mathbb{Z}_{\geq 0}$,

$$B_{\ell+1} = B_\ell \cup \big\{x \in \mathscr{X} \setminus B_\ell \colon \big[(I + \Delta\underline{Q})\mathbb{I}_{B_\ell}\big](x) > 0\big\} = B_\ell \cup \big\{x \in \mathscr{X} \setminus B_\ell \colon \big[\underline{Q}\mathbb{I}_{B_\ell}\big](x) > 0\big\},$$

where the second equality holds because $\mathbb{I}_{B_{\ell-1}}(x) = 0$ and $\Delta > 0$. It now follows from this and Definition 4.32<sub>193</sub> that $y \hookrightarrow \mathscr{X}_{\underline{Q}}$.

Seeing that we have shown that $\mathscr{X}_{\underline{Q}} \neq \emptyset$ and $y \hookrightarrow \mathscr{X}_{\underline{Q}}$ for all $y$ in $\mathscr{X} \setminus \mathscr{X}_{\underline{Q}}$, it now follows from Proposition 4.27<sub>189</sub> that $\underline{Q}$ is ergodic, which is what we needed to prove. □

# 5 Extension to idealised inferences

The inferences that we can make using an imprecise jump process are rather limited: by construction, we only deal with finitary events, that is, those events that depend on the state of the system at a finite number of time points. Therefore, we have only defined the lower and upper expectation of finitary variables – which are all gambles – and, as a particular case, the lower and upper probability of finitary events. In many applications, however, finitary variables do not suffice. More often than not, one is interested in inferences – or, more precisely, in events and variables – that depend on the state of the system at the time points in some closed time interval $[s, r] \subset \mathbb{R}_{\geq 0}$, or even on the state of the system at *all* time points in $\mathbb{R}_{\geq 0}$. As always, we can make this more concrete with our running example.

**Joseph's Example 5.1.** Recall from Joseph's Example 3.13[65] that Cecilia is convinced that Joseph's machine will always display heads. For this reason, she should assign probability one to the event

$$\bigcap_{t \in \mathbb{R}_{\geq 0}} \{X_t = \mathtt{H}\} = \{\omega_{\mathtt{H}}\},$$

where $\omega_{\mathtt{H}}$ is the path that is always $\mathtt{H}$. By the laws of probability, she should then assign probability zero to the complement

$$\bigcup_{t \in \mathbb{R}_{\geq 0}} \{X_t = \mathtt{T}\} = \Omega \setminus \{\omega_{\mathtt{H}}\}.$$

The probability of the second event is an example of what is commonly known as a *hitting probability*. For this reason, we call the event that a subset $A$ of $\mathscr{X}$ is ever 'hit' – that is, that some situation $\omega(t)$ along the path $\omega$ belongs to $A$ at some time point $t$ – a hitting event.

Often, we are also interested in the time it takes to hit the subset $A$ of $\mathscr{X}$. This is captured by the *hitting time* $\tau_A$: the variable that is equal to the earliest time point such that the state of the system at that time point belongs to $A$. In the present setting, the hitting time of $\{\mathtt{H}\}$ is

$$\tau^{\mathtt{H}} : \Omega \to \overline{\mathbb{R}} : \omega \mapsto \inf\{t \in \mathbb{R}_{\geq 0} : \omega(t) = \mathtt{H}\}.$$

Note that $\tau^{\mathtt{H}}$ depends on the value of the path $\omega$ at all time points in $\mathbb{R}_{\geq 0}$, so it is *not* finitary. Even worse, so to speak, $\tau^{\mathtt{H}}$ is not a gamble but an extended real variable: $\tau^{\mathtt{H}}(\omega) = +\infty$ if $\omega$ never hits H because the infimum of the empty set in $\mathbb{R}$ (or $\mathbb{R}_{\geq 0}$) is $+\infty$.

In the setting of jump processes, we will only deal with *idealised* events and variables, meaning that they are 'point-wise limits' of finitary events and variables, respectively – that is, the 'point-wise limit' of events in $\mathscr{F}_u$ and variables in $\mathbb{S}(\mathscr{F}_u)$, respectively. Because they are defined as limits of finitary variables, idealised variables may be unbounded; in fact, as the preceding example illustrates, it makes sense to allow them to be extended real-valued. Because these idealised variables are defined through a limit, it is reasonable to determine their expectation through limit arguments as well, and it is precisely this we will do in this chapter.

We will not immediately set about to extending the domain of the lower and upper expectations corresponding to an imprecise jump process. Instead, we start this chapter in Section 5.1 with some general theory on how to extend the domain of the expectation that corresponds to a probability charge. In Section 5.2228, we use this theory to extend the domain of the conditional expectation corresponding to a (countably additive) jump process. Finally, in Section 5.3238 we extend the domain of the conditional lower and upper expectations corresponding to an imprecise jump process.

## 5.1 Extension through limit arguments

Suppose we have a probability charge $P$ on some field $\mathcal{F}$ of events over a possibility space $\mathcal{X}$. Recall that in Section 2.3.335, we have defined the corresponding expectation $E_P$ on the linear space $\mathbb{S}(\mathcal{F})$ of $\mathcal{F}$-simple variables – which, by Definition 2.3836, is a subspace of the set of all gambles – through the Dunford integral. In this section, we seek to extend the domain of the expectation $E_P$ to more general variables, that is, to bounded real variables that are not $\mathcal{F}$-simple, to unbounded real variables and even to extended real variables.

We go about this as follows. First, we explain in Section 5.1.1↷ why the natural extension is ill-suited for this purpose. Next, in Sections 5.1.2219 and 5.1.3224 we explain how Daniell (1918) uses limit arguments to extend $E_P$. Finally, in Section 5.1.4227 we briefly investigate the relationship of this extension with coherence.

For the duration of this section, we return to the discrete-time setting in our running example because it is conceptually easier.

**Bruno's Example 5.2.** Recall from Bruno's Example 2.3332 that the field

$$\mathcal{F} = \left\{\{X_{1:n} \in A\} \colon n \in \mathbb{N}, A \subseteq \{\mathtt{H}, \mathtt{T}\}^n\right\}$$

consists of all events that depend on the outcome of a finite – but unbounded – number of coin flips. Furthermore, we recall from Eq. (2.17)35 in Bruno's Example 2.3735 that for a given probability mass function $q$ on the state space $\{\mathtt{H},\mathtt{T}\}$, the real-valued map $P$ on $\mathcal{F}$ defined by

$$P(X_{1:n} \in A) = \sum_{y_{1:n} \in A} \prod_{k=1}^{n} q(y_k) \quad \text{for all } \{X_{1:n} \in A\} \in \mathcal{F} \tag{5.1}$$

is a probability charge.

### 5.1.1 Why not the natural extension?

From Proposition 2.4338, we know that $E_P$ is a coherent expectation on $\mathbb{S}(\mathcal{F})$. Thus, due to Proposition 2.2025, we can always extend the coherent expectation $E_P$ to a coherent expectation on $\mathbb{G}(\mathcal{X})$, and we have seen in Proposition 2.2226 that the natural extension $\underline{\mathcal{E}}_P := \underline{\mathcal{E}}_{E_P}$ provides tight lower and upper bounds on these coherent extensions. Moreover, because $\mathbb{S}(\mathcal{F})$ is a real vector space that includes all constant gambles – see Lemma 2.3936 – it follows from Proposition 2.2427 that

$$\underline{\mathcal{E}}_P(g) = \sup\{E_P(h) \colon h \in \mathbb{S}(\mathcal{F}), h \le g\} \quad \text{for all } g \in \mathbb{G}(\mathcal{X}) \tag{5.2}$$

and, by conjugacy, that

$$\overline{\mathcal{E}}_P(g) = \inf\{E_P(h) \colon h \in \mathbb{S}(\mathcal{F}), h \ge g\} \quad \text{for all } g \in \mathbb{G}(\mathcal{X}). \tag{5.3}$$

Note that for any $\mathcal{F}$-simple variable $g$, $\underline{\mathcal{E}}_P(g) = \overline{\mathcal{E}}_P(g) = E_P(g)$. Whenever the lower and upper expectation of an arbitrary – so not necessarily $\mathcal{F}$-simple – gamble $g$ coincide, we call that gamble $g$ *C-integrable*, where the 'C' refers to coherence. In this case, we denote the common value of $\underline{\mathcal{E}}_P(g)$ and $\overline{\mathcal{E}}_P(g)$ by $E_P^{\mathrm{C}}(g)$, and simply call it the expectation of $g$.[1] This way, we can extend the domain of $E_P$ from the set of $\mathcal{F}$-simple variables $\mathbb{S}(\mathcal{F})$ to the set of *C-integrable gambles*

$$\mathbb{D}_P^{\mathrm{C}} := \{g \in \mathbb{G}(\mathcal{X}) \colon \underline{\mathcal{E}}_P(g) = \overline{\mathcal{E}}_P(g)\}.$$

In other words, $E_P^{\mathrm{C}}$ is the restriction of $\underline{\mathcal{E}}_P$ to $\mathbb{D}_P^{\mathrm{C}}$, those gambles for which the natural extension $\underline{\mathcal{E}}_P$ is self-conjugate. By (Troffaes et al., 2014, Proposition 8.2), $E_P^{\mathrm{C}}$ is a coherent expectation on $\mathbb{D}_P^{\mathrm{C}}$.

Unfortunately, a lot of practically relevant gambles are *not* C-integrable. This is illustrated by the following example, for which we are indebted to Troffaes (2013).

[1] Troffaes et al. (2014, Definition 8.1) use different terminology: they use the term *$E_P$-integrable* instead of C-integrable, and call $E_P^{\mathrm{C}}(g)$ the $E_P$-integral of $g$.

**Bruno's Example 5.3.** Cecilia is not only sceptical about Joseph's machine, but also about Bruno's coin flipping machine. More precisely, Cecilia is convinced that Bruno's machine will always flip heads – or, equivalently, she has very strong doubts that a flip of the machine will ever come out tails. In our formalism, the event that Bruno's machine never flips tails corresponds to the event

$$H_{\lim} \coloneqq \{(x_n)_{n\in\mathbb{N}} \in \mathcal{X} \colon (\forall n \in \mathbb{N})\ x_n = \mathtt{H}\} = \{(x_n)_{n\in\mathbb{N}} \in \mathcal{X} \colon (\exists n \in \mathbb{N})\ x_n = \mathtt{T}\}^{\mathrm{c}}$$
$$= \bigcap_{n\in\mathbb{N}} \{X_n = \mathtt{H}\} = \bigcap_{n\in\mathbb{N}} \bigcap_{k=1}^{n} \{X_k = \mathtt{H}\} = \bigcap_{n\in\mathbb{N}} H_n,$$

where $\{X_n = \mathtt{H}\}$ and $H_n$ are as defined in Bruno's Example $2.33_{32}$. From the last equality, we infer that $H_{\lim}$ is a countable intersection of events in the field $\mathcal{F}$. However, because $H_{\lim}$ depends on the result of *all* coin flips and not just on a finite number of them, $H_{\lim}$ itself does not belong to the field $\mathcal{F}$.

We now set out to determine whether or not $H_{\lim}$ is a C-integrable event, meaning that $\mathbb{I}_{H_{\lim}}$ is a C-integrable gamble. Observe that $\emptyset \subseteq H_{\lim} \subseteq H_n$ for any natural number $n$, so $0 = \mathbb{I}_\emptyset \leq \mathbb{I}_{H_{\lim}} \leq \mathbb{I}_{H_n}$. One can furthermore check that if $h$ is an $\mathcal{F}$-simple variable such that $h \leq \mathbb{I}_{H_{\lim}}$, then also $h \leq \mathbb{I}_\emptyset = 0$; similarly, if $h$ is an $\mathcal{F}$-simple variable such that $h \geq \mathbb{I}_{H_{\lim}}$, then there is some natural number $n$ such that $\mathbb{I}_{H_n} \leq h$. Thus, it follows from this, $(\text{LE6})_{30}$ and Eqs. $(5.2)_{\curvearrowleft}$ and $(5.3)_{\curvearrowleft}$ that

$$\underline{\mathcal{E}}_P(\mathbb{I}_{H_{\lim}}) = P(\emptyset) = 0 \quad \text{and} \quad \overline{\mathcal{E}}_P(\mathbb{I}_{H_{\lim}}) = \inf\{P(H_n) \colon n \in \mathbb{N}\} = \inf\{q^n \colon n \in \mathbb{N}\}.$$

In case $0 \leq q < 1$, we infer from this that $\mathbb{I}_{H_{\lim}}$ is C-integrable and $E_P^{\mathrm{C}}(\mathbb{I}_{H_{\lim}}) = 0$. If on the other hand $q = 1$, then $\mathbb{I}_{H_{\lim}}$ is *not* C-integrable because $\underline{\mathcal{E}}_P(\mathbb{I}_{H_{\lim}}) = 0$ and $\overline{\mathcal{E}}_P(\mathbb{I}_{H_{\lim}}) = 1$. So even though in the latter case $P(H_n) = E_P(\mathbb{I}_{H_n})$ – the probability that the first $n$ flips of the machine are all heads – is equal to one for every natural number $n$, we cannot conclude on the basis of coherence alone that the probability of $H_{\lim}$ is one as well. In fact, this probability can be can be any real number $r$ in $[0,1]$, because it follows from Proposition $2.22_{26}$ that for every $r$ in $[0,1] = [\underline{\mathcal{E}}_P(\mathbb{I}_{H_{\lim}}), \overline{\mathcal{E}}_P(\mathbb{I}_{H_{\lim}})]$, there is a coherent extension $E^\star$ of $E_P$ from $\mathbb{S}(\mathcal{F})$ to $\mathbb{G}(\mathcal{X})$ such that $E^\star(\mathbb{I}_{H_{\lim}}) = r$. ∲

In this example, the event $H_{\lim}$ is an 'idealisation' because it concerns an *infinite* number of coin flips. Note that it is the 'limit' $\bigcap_{n\in\mathbb{N}} H_n$ of the non-increasing sequence of events $(H_n)_{n\in\mathbb{N}}$, where $H_n$ only depends on $n$ consecutive coin flips. In spite of this, this limit behaviour of the events is *not* carried over to their probabilities whenever $q = 1$. In short, this example reveals that coherence might not be the right tool when dealing with idealisations, especially so if we want that 'monotone limits' of events are carried over to monotone limits of probabilities – and, more generally, monotone limits of variables carry over to expectations.

A second argument against extending using the coherence framework in general and the natural extension in particular, is that this framework is motivated using a gambling interpretation, and it does not always make sense to use such an interpretation for every gamble in $\mathbb{G}(\mathcal{X})$. More precisely, we have previously mentioned in Section $2.2.1_{16}$ that this gambling interpretation only makes sense for determinable gambles. The event $H_{\text{lim}}$ in Bruno's Example 5.3 is clearly not determinable, because we can never be sure that the next toss will not be tails instead of heads.

A third argument for not using the coherence framework, is that it is restricted to gambles. This makes sense because the interpretation of coherence does not seem compatible with unbounded real variables – let alone extended real variables – because no rational subject should be disposed to agree to a transaction that has no lower bound on the utility that it can cost her. Nonetheless, the theory of coherence has been extended to unbounded real variables – but, to the best of our knowledge, *not* to extended real variables. Crisma et al. (1997) have extended the notion of coherence for expectations to real variables, but their work is subsumed by that of Troffaes et al. (2014, Chapter 13), who treat coherence for (conditional) lower expectations on real variables.

Troffaes et al. (2014, Section 13.11) also argue that in this more general setting, the natural extension might be too conservative to be of any practical use. More specifically, they conclude that 'perhaps, starting from the well-understood bounded case, and taking limits, might lead to more practical answers.' They subsequently put their money where their mouth is, and thoroughly explain how this works: in (Troffaes et al., 2014, Chapter 15), they extend a coherent lower expectation $\underline{E}$ on the set $\mathbb{G}(\mathcal{X})$ of all gambles to the set of 'previsible' real variables. However, their results are not (immediately) applicable to our setting, because we seek to extend the coherent expectation $E_P$ from the set $\mathbb{S}(\mathcal{F})$ of $\mathcal{F}$-*simple* variables to (a subset of) the *extended* real variables. We could follow the approach of Troffaes et al. (2014, Chapter 15) starting from the natural extension $\underline{\mathcal{E}}_P$ of $E_P$ to $\mathbb{G}(\mathcal{X})$, but we choose not to because (i) we already know from Bruno's Example 5.3 that the natural extension $\underline{\mathcal{E}}_P$ lacks continuity properties, and (ii) this does not permit us to deal with extended real variables. This being said, we will relate our method of extending to the coherence framework in Section $5.1.4_{227}$ further on, and we will also reflect on this again in Section $5.3_{238}$.

### 5.1.2 Daniell extension through monotone convergence

There are *a lot* of ways to extend the expectation $E_P$ that corresponds to the probability charge $P$. For example, Bhaskara Rao et al. (1983, Definition 4.4.11) extend $E_P$ to the linear space of the 'Dunford integrable' real-valued variables by means of limit arguments (see also Troffaes et al., 2014, Definition 8.29). However, this extension is ill-suited for our purposes because it is restricted to real variables.

A well-known alternative that can deal with extended real variables is the standard way to go in measure-theoretic probability theory: extend the charge to a probability measure using Carathéodory's Theorem, and subsequently use the Lebesgue integral with respect to this measure to extend the domain of $E_P$. This extension essentially uses limit arguments, as is detailed in Appendix $\text{C}_{461}$ (see also D. Williams, 1991; Billingsley, 1995; Shiryaev, 2016). However, in this dissertation we follow Daniell's (1918) slightly different approach. Apart from my being slightly contrarian, my main reason for doing so is that the limit arguments are more obvious. Nonetheless, in Theorem $\text{C.19}_{471}$ of Appendix $\text{C.3}_{470}$, we establish that, for all intents and purposes, these two approaches are equivalent.

Daniell (1918) extends the domain of the expectation $E_P$ that corresponds to a probability charge $P$ on the basis of limit arguments. His work is actually more general, because he extends a general 'elementary integral operator' over some abstract 'vector lattice of functions'. Royden (1968, Chapter 13) and Taylor (1985, Chapter 6) also stick to this general setting in their more recent treatments of Daniell's approach. Our exposition is to a large extent in line with Taylor's (1985), but there are two notable differences. The first key difference is that we do not consider a general 'elementary integral', but instead take the special case of an expectation with respect to a (countably additive) probability charge as starting point. The second difference is that our extension is extended-real-valued, whereas Daniell (1918), Royden (1968), and Taylor (1985) restrict the domain in order to end up with a real-valued extension.

As a consequence of these two differences, we cannot simply borrow all results from the aforementioned references, but almost all of the results in this section are straightforward adaptations of well-known results. For this reason, we have chosen to nonetheless refer to Taylor's (1985) results throughout the main text; a justification for why we can use these results can be found in Appendix $\text{B.1}_{451}$. In order to not unnecessarily burden the main text with technicalities, we have also relegated to that appendix the proofs of most of the results in the remainder of this section.

Taking these caveats into account, we construct the Daniell extension as follows. Our starting point is a *countably additive* probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Next, a limit argument extends the domain of $E_P$ to extended real variables that are the limits of monotone sequences of $\mathcal{F}$-simple variables. Finally, in Section $5.1.3_{224}$ we use this first extension to extend the domain $E_P$ even further, this time by approximating any extended real variable with limits of monotone sequences of extended real variables that are themselves the limit of a monotone sequence of $\mathcal{F}$-simple variables.

**Monotone sequences of $\mathcal{F}$-simple variables**

Essential to Daniell's (1918) extension are sequences of (extended) real variables that converge in some sense. In this context, the basic notion of convergence is that of point-wise convergence. A sequence $(f_n)_{n\in\mathbb{N}}$ of extended real variables *converges point-wise* if for all $x$ in $\mathcal{X}$, the limit $\lim_{n\to+\infty} f_n(x)$ exists. Whenever this is the case, we say that $(f_n)_{n\in\mathbb{N}}$ converges point-wise to

$$\operatorname*{p\text{-}w\,lim}_{n\to+\infty} f_n\colon \mathcal{X}\to\overline{\mathbb{R}}\colon x\mapsto \lim_{n\to+\infty} f_n(x). \tag{5.4}$$

The second notion of convergence deals with *monotone sequences* of extended real variables, that is, sequences that are non-decreasing, non-increasing or both. Formally, we call a sequence $(f_n)_{n\in\mathbb{N}}$ of extended real variables *non-decreasing* if $f_n\le f_{n+1}$ for all $n$ in $\mathbb{N}$. For such a non-decreasing sequence, it is clear that the limit $\lim_{n\to+\infty} f_n(x)$ exists for all $x$ in $\mathcal{X}$, but it can be equal to $+\infty$ for some $x$ in $\mathcal{X}$; hence, $(f_n)_{n\in\mathbb{N}}$ converges point-wise. Therefore, we say that the non-decreasing sequence $(f_n)_{n\in\mathbb{N}}$ *converges monotonically* to $f:=\operatorname{p\text{-}w\,lim}_{n\to+\infty} f_n$ and denote this by $(f_n)_{n\in\mathbb{N}}\nearrow f$. Conversely, we call a sequence $(f_n)_{n\in\mathbb{N}}$ of extended real variables *non-increasing* if $(-f_n)_{n\in\mathbb{N}}$ is non-decreasing; in this case we also say that $(f_n)_{n\in\mathbb{N}}$ *converges monotonically* to the extended real variable $f:=\operatorname{p\text{-}w\,lim}_{n\to+\infty}$, but denote this by $(f_n)_{n\in\mathbb{N}}\searrow f$.

Consider any $\mathcal{F}$-simple variable $f$, and suppose $(f_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence of $\mathcal{F}$-simple variables that converges monotonically to $f$. Observe that for all $n$ in $\mathbb{N}$, $f_n\le f_{n+1}\le f$, so $E_P(f_n)\le E_P(f_{n+1})\le E_P(f)$ due to (ES4)$_{37}$. It follows from this and (ES1)$_{37}$ that $(E_P(f_n))_{n\in\mathbb{N}}$ is a non-decreasing sequence of real numbers that is bounded above by $E_P(f)\le\max f<+\infty$, so this sequence converges to a real number and

$$\lim_{n\to+\infty} E_P(f_n)\le E_P(f).$$

Whenever this inequality holds with equality for all $f$ in $\mathbb{S}(\mathcal{F})$ and $(f_n)_{n\in\mathbb{N}}$ in $\mathbb{S}(\mathcal{F})$ with $(f_n)\nearrow f$, we call the probability charge $P$ countably additive. It might be a bit counter-intuitive to refer to this property of monotone limits using the term 'countable additivity', but this terminology is well-established. The more conventional necessary and sufficient conditions of Lemma C.3$_{463}$ in Appendix C$_{461}$ explain the origins of this terminology.

**Definition 5.4.** Consider a probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then the following three conditions are equivalent. Whenever $P$ satisfies one (and hence all) of them, we call $P$ *countably additive*.

(i) For any $\mathcal{F}$-simple variable $f$ and any sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables such that $(f_n)_{n\in\mathbb{N}}\nearrow f$, $E_P(f)=\lim_{n\to+\infty} E_P(f_n)$.

(ii) For any $\mathcal{F}$-simple variable $f$ and any sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables such that$(f_n)_{n\in\mathbb{N}} \searrow f$, $E_P(f) = \lim_{n\to+\infty} E_P(f_n)$.

(iii) For any sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables such that $(f_n)_{n\in\mathbb{N}} \searrow 0$, $\lim_{n\to+\infty} E_P(f_n) = 0$.

**Bruno's Example 5.5.** Billingsley (1995, Theorem 2.3) shows that any probability charge on the field $\mathcal{F}$ of events that depend on the outcome of a finite number of coin flips – as defined in Bruno's Example 2.33[32] – is countably additive. In particular, this holds for our probability charge $P$ on $\mathcal{F}$ as defined in Eq. (2.17)[35]. ∲

### $\mathcal{F}$-over and $\mathcal{F}$-under variables

The first step towards Daniell's (1918) extension is to extend the domain of $E_P$ to extended real variables that are the limits of monotone sequences of $\mathcal{F}$-simple variables. If $(f_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence of $\mathcal{F}$-simple variables, then we call the extended real variable $f := \text{p-w}\lim_{n\to+\infty} f_n$ an *$\mathcal{F}$-over variable*. Similarly, an *$\mathcal{F}$-under variable* is an extended real variable $f$ such that there is at least one non-increasing sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables that converges point-wise to $f$. Note that an $\mathcal{F}$-over variable $f$ can attain $+\infty$ but not $-\infty$ because $f$ is bounded below; conversely, an $\mathcal{F}$-under variable can attain $-\infty$ but not $+\infty$ because it is bounded above. Hence, an extended real variable $f$ that is both an $\mathcal{F}$-over and $\mathcal{F}$-under variable is bounded below and above, so simply a gamble.

We denote the set of all $\mathcal{F}$-over and $\mathcal{F}$-under variables by $\overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})$ and $\overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$, respectively. Observe that

$$\overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F}) = -\overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F}) = \left\{-f : f \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})\right\} \tag{5.5}$$

because $(f_n)_{n\in\mathbb{N}} \nearrow f$ implies that $(-f_n)_{n\in\mathbb{N}} \searrow -f$ and vice versa. Furthermore, because any $\mathcal{F}$-simple variable $f$ is the limit of the constant (and hence non-decreasing and non-increasing) sequence $(f)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables, we conclude that $\mathbb{S}(\mathcal{F})$ is included in $\overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})$ and $\overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$ – so $\mathbb{S}(\mathcal{F}) \subseteq \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F}) \cap \overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$. For ease of notation, we denote the set of all $\mathcal{F}$-over and all $\mathcal{F}$-under variables by $\overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathcal{F}) := \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F}) \cup \overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$.

Take any $\mathcal{F}$-over or $\mathcal{F}$-under variable $f$. By definition, it is the limit of some sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables that is non-decreasing or non-increasing. In either case, it follows from (ES4)[37] that $(E_P(f_n))_{n\in\mathbb{N}}$ is a non-decreasing or non-increasing sequence of real numbers, so the limit $\lim_{n\to+\infty} E_P(f_n)$ exists. Note that this limit need not be real-valued, but can be equal to $-\infty$ if $(f_n)_{n\in\mathbb{N}}$ is non-increasing and $+\infty$ if $(f_n)_{n\in\mathbb{N}}$ is non-decreasing. Crucial to Daniell's extension is that if $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ are two monotone sequences of $\mathcal{F}$-simple variables that converge to the same variable $f$, then the limits of their respective expectations should be equal. The following result – essentially due to Taylor (1985, Section 6-2) – establishes that this holds if $P$ is countably additive.

**Lemma 5.6.** *Consider a countably additive probability charge $P$ on a field of events $\mathscr{F}$ over some possibility space $\mathscr{X}$, and some $f$ in $\overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathscr{F})$. If $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ are monotone sequences of $\mathscr{F}$-simple variables that both converge point-wise to $f$, then*

$$\lim_{n\to+\infty} E_P(f_n) = \lim_{n\to+\infty} E_P(g_n).$$

Because the limit $\lim_{n\to+\infty} E_P(f_n)$ is the same for any monotone sequence $(f_n)_{n\in\mathbb{N}}$ that converges point-wise to the same limit variable $f$, it seems sensible to accept this limit value as the expectation of $f$. In this way, we obtain the extension $E_P^{\mathrm{mc}}$ of $E_P$ to $\overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathscr{F})$ defined by

$$E_P^{\mathrm{mc}}(f) := \lim_{n\to+\infty} E_P(f_n) \quad \text{for all } f \in \overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathscr{F}), \tag{5.6}$$

where $(f_n)_{n\in\mathbb{N}}$ is any monotone sequence of $\mathscr{F}$-simple variables that converges point-wise to $f$. Observe that $E_P^{\mathrm{mc}}$ coincides with $E_P$ on $\mathbb{S}(\mathscr{F})$ because – as we have seen before – any $\mathscr{F}$-simple variable $f$ is the monotone limit of the constant sequence $(f)_{n\in\mathbb{N}}$.

Let us confirm that this extension does not suffer from the same issue as the coherence approach in Bruno's Example $5.3_{218}$.

**Bruno's Example 5.7.** Recall from Bruno's Example $5.3_{218}$ that

$$H_{\mathrm{lim}} = \bigcap_{n\in\mathbb{N}} H_n.$$

We know that $H_n$ belongs to $\mathscr{F}$ from Bruno's Example $2.33_{32}$, and it is clear that $H_n \supseteq H_{n+1}$ for all $n$ in $\mathbb{N}$. For this reason, $(\mathbb{I}_{H_n})_{n\in\mathbb{N}}$ is a non-increasing sequence of $\mathscr{F}$-simple variables. It is easy to verify that $(\mathbb{I}_{H_n})_{n\in\mathbb{N}}$ converges point-wise to $\mathbb{I}_{H_{\mathrm{lim}}}$, which makes $\mathbb{I}_{H_{\mathrm{lim}}}$ an $\mathscr{F}$-under variable. Therefore,

$$E_P^{\mathrm{mc}}(\mathbb{I}_{H_{\mathrm{lim}}}) = \lim_{n\to+\infty} E_P(\mathbb{I}_{H_n}) = \lim_{n\to+\infty} P(H_n) = \lim_{n\to+\infty} q^n = \begin{cases} 0 & \text{if } q<1 \\ 1 & \text{if } q=1. \end{cases}$$

Thus, $\mathbb{I}_{H_{\mathrm{lim}}}$ belongs to the domain $\overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathscr{F})$ of the extension $E_P^{\mathrm{mc}}$, regardless of the value of $p(\mathtt{H}) = q$. ∲

It is important to realise that $E_P^{\mathrm{mc}}$ need not be real-valued, as is illustrated by the next example.

**Bruno's Example 5.8.** To illustrate that $E_P^{\mathrm{mc}}$ need not be real-valued, we consider the hitting time of $\mathtt{T}$:

$$\tau^{\mathtt{T}} : \mathscr{X} \to \overline{\mathbb{R}} : \phi \mapsto \tau^{\mathtt{T}}(\phi) := \inf\{n \in \mathbb{N} : \phi_n = \mathtt{T}\}.$$

Observe that $\tau^{\mathtt{T}}$ is always real-valued, except when *every* coin flip in the outcome is H. In the latter case, the set in the definition above is empty, so the variable $\tau^{\mathtt{T}}$ assumes the value $+\infty$. We now set out to establish that $\tau^{\mathtt{T}}$ is an $\mathcal{F}$-over variable.

To this end, we consider for any natural number $n$ the real variable $\tau^{\mathtt{T}}_{\wedge n} := \tau^{\mathtt{T}} \wedge n$. Observe that, by construction, $\tau^{\mathtt{T}}_{\wedge n}$ only depends on the first $n$ flips of the machine. It is essentially for this reason that $\tau^{\mathtt{T}}_{\wedge n}$ is an $\mathcal{F}$-simple variable. To establish this formally, we observe that

$$\tau^{\mathtt{T}}_{\wedge n} = \sum_{k=1}^{n} k\mathbb{I}_{\{X_{1:k}=y^k_{1:k}\}} + n\mathbb{I}_{\{X_{1:n}=(\mathtt{H},\dots,\mathtt{H})\}},$$

where for any natural number $k$, $y^k_{1:k}$ is the $k$-tuple in $\{\mathtt{H},\mathtt{T}\}^k$ such that $y^k_k = \mathtt{T}$ and $y^k_\ell = \mathtt{H}$ for all $\ell < k$. From this and Eqs. (2.16)<sub></sub>35 and (2.19)36, we infer that

$$E_P\big(\tau^{\mathtt{T}}_{\wedge n}\big) = \sum_{k=1}^{n} kq^{k-1}(1-q) + nq^n = \begin{cases} n & \text{if } q = 1,\\ \frac{1-q^n}{1-q} & \text{if } 0 \leq q < 1.\end{cases} \tag{5.7}$$

We know that $(\tau^{\mathtt{T}}_{\wedge n})_{n\in\mathbb{N}}$ is a sequence of $\mathcal{F}$-simple variables. It is furthermore easy to see that this sequence is non-decreasing and that it converges point-wise to $\tau^{\mathtt{T}}$; in other words, $\tau^{\mathtt{T}}$ is an $\mathcal{F}$-over variable. Consequently, it follows from Eqs. (5.6)↶ and (5.7) that

$$E^{\mathrm{mc}}_P\big(\tau^{\mathtt{T}}\big) = \lim_{n\to+\infty} E_P\big(\tau^{\mathtt{T}}_{\wedge n}\big) = \begin{cases} +\infty & \text{if } q = 1,\\ \frac{1}{1-q} & \text{if } 0 \leq q < 1.\end{cases}$$ ∲

### 5.1.3 Daniell extension through inner and outer approximations

In many cases, including that of jump processes, $\mathcal{F}$-over and $\mathcal{F}$-under variables do not make up all idealised variables that one might be interested in. For this reason, we take the second and final step in Daniell's extension, which consists in approximating general extended real variables from above and below. This step is similar to what we did in Proposition 2.24₂₇, where we obtained the natural extension $\underline{\mathcal{E}}$ (and its conjugate $\overline{\mathcal{E}}$) of a coherent lower expectation $\underline{E}$ through inner and outer approximation. The difference is that instead of approximating a gamble $g$ from above and below with gambles $h$ in the domain $\mathcal{G}$ of $\underline{E}$, we now approach the extended real variable $f$ from above with $\mathcal{F}$-over variables and from below with $\mathcal{F}$-under variables. Formally, we define the *inner Daniell extension* $E^{\mathrm{i}}_P \colon \overline{\mathbb{V}}(\mathcal{X}) \to \overline{\mathbb{R}}$ by

$$E^{\mathrm{i}}_P(f) := \sup\big\{E^{\mathrm{mc}}_P(h) \colon h \in \overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F}), h \leq f\big\} \qquad \text{for all } f \in \overline{\mathbb{V}}(\mathcal{X}) \tag{5.8}$$

and the *outer Daniell extension* $E^{\mathrm{o}}_P \colon \overline{\mathbb{V}}(\mathcal{X}) \to \overline{\mathbb{R}}$

$$E^{\mathrm{o}}_P(f) := \inf\big\{E^{\mathrm{mc}}_P(h) \colon h \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F}), h \geq f\big\} \qquad \text{for all } f \in \overline{\mathbb{V}}(\mathcal{X}). \tag{5.9}$$

Because Eqs. (5.8)$_\curvearrowleft$ and (5.9)$_\curvearrowleft$ are so similar to the two expressions for the natural extension in Proposition 2.24$_{27}$, it should not come as a surprise that the inner and outer Daniell extensions satisfy similar properties as coherent lower and upper expectations, respectively. We refer to (Taylor, 1985, Section 6-3) or Lemma B.6$_{455}$ in Appendix B.1$_{451}$ for a list of properties. In order to understand the remainder, it suffices to know that

$$E_P^{\mathrm{i}}(f) = E_P^{\mathrm{o}}(f) = E_P^{\mathrm{mc}}(f) \quad \text{for all } f \in \overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F}) \tag{5.10}$$

and, because $E_P$ and $E_P^{\mathrm{mc}}$ coincide for $\mathcal{F}$-simple variables, therefore also

$$E_P^{\mathrm{i}}(f) = E_P^{\mathrm{o}}(f) = E_P^{\mathrm{mc}}(f) = E_P(f) \quad \text{for all } f \in \mathbb{S}(\mathcal{F}). \tag{5.11}$$

The inner and outer Daniell extension typically coincide on more variables than just the $\mathcal{F}$-over and $\mathcal{F}$-under ones. Whenever the inner and outer Daniell extensions $E_P^{\mathrm{i}}(f)$ and $E_P^{\mathrm{o}}(f)$ of an extended real variable $f$ coincide, we call $f$ *D-integrable*. In that case, we denote the common value of $E_P^{\mathrm{i}}(f)$ and $E_P^{\mathrm{o}}(f)$ by $E_P^{\mathrm{D}}(f)$ and call this the *Daniell expectation* – sometimes also Daniell integral – of $f$. In this way, $E_P^{\mathrm{D}}$ is an extended real-valued functional with domain

$$\mathbb{D}_P^{\mathrm{D}} := \left\{ f \in \overline{\mathbb{V}}(\mathcal{X}) \colon E_P^{\mathrm{i}}(f) = E_P^{\mathrm{o}}(f) \right\}. \tag{5.12}$$

Note that, due to Eqs. (5.10) and (5.11),

$$\mathbb{S}(\mathcal{F}) \subseteq \overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F}) \subseteq \mathbb{D}_P^{\mathrm{D}}. \tag{5.13}$$

Our definition of the domain $\mathbb{D}_P^{\mathrm{D}}$ of the Daniell extension $E_P^{\mathrm{D}}$ is more general than Taylor's (1985, Section 6–3), because he ensures that $E_P^{\mathrm{D}}$ is real-valued by restricting its domain to

$$\tilde{\mathbb{D}}_P^{\mathrm{D}} := \left\{ f \in \overline{\mathbb{V}}(\mathcal{X}) \colon -\infty < E_P^{\mathrm{i}}(f) = E_P^{\mathrm{o}}(f) < +\infty \right\}.$$

Taylor (1985, Theorem 6-3 II) lists several properties of $E_P^{\mathrm{D}}$ on $\tilde{\mathbb{D}}_P^{\mathrm{D}}$. For the sake of conciseness, we only generalise some of these properties to $\mathbb{D}_P^{\mathrm{D}}$.

**Theorem 5.9.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then*

DE1. $\mathbb{S}(\mathcal{F}) \subseteq \mathbb{D}_P^{\mathrm{D}}$ *and* $E_P^{\mathrm{D}}(f) = E_P(f)$ *for all $f$ in* $\mathbb{S}(\mathcal{F})$*;*

DE2. $\mathbb{S}(\mathcal{F}) \subseteq \overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F}) \subseteq \mathbb{D}_P^{\mathrm{D}}$ *and* $E_P^{\mathrm{D}}(f) = E_P^{\mathrm{mc}}(f)$ *for all $f$ in* $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F})$.

*Furthermore, for all D-integrable extended real variables $f$ and $g$ in $\mathbb{D}_P^{\mathrm{D}}$ and all real numbers $\mu$ in $\mathbb{R}$,*

DE3. $\inf f \leq E_P^{\mathrm{D}}(f) \leq \sup f$*;*

DE4. $\mu f$ *is D-integrable and* $E_P^{\mathrm{D}}(\mu f) = \mu E_P^{\mathrm{D}}(f)$*;*

DE5. $f + g$ *is D-integrable and* $E_P^{\mathrm{D}}(f+g) = E_P^{\mathrm{D}}(f) + E_P^{\mathrm{D}}(g)$ *whenever $f + g$ and $E_P^{\mathrm{D}}(f) + E_P^{\mathrm{D}}(g)$ are well-defined;*

DE6. $E_P^{\mathrm{D}}(f) \le E_P^{\mathrm{D}}(g)$ *whenever* $f \le g$.

*Finally, for all D-integrable real variables* $f$ *and* $g$ *in* $\tilde{\mathbb{D}}_P^{\mathrm{D}}$,

DE7. $f \vee g$ *and* $f \wedge g$ *also belong to* $\tilde{\mathbb{D}}_P^{\mathrm{D}}$.

Observe that the properties (DE3)–(DE6) of the Daniell expectation $E_P^{\mathrm{D}}$ are generalisations of the properties $(\mathrm{ES1})_{37}$–$(\mathrm{ES4})_{37}$ of the Dunford expectation $E_P$.

**Limit theorems**

Even more important, however, are the following two quintessential limit properties. The first is known as the *Monotone Convergence Theorem*; we extend Taylor's (1985, Theorem 6-3 III) statement from $\tilde{\mathbb{D}}_P^{\mathrm{D}}$ to $\mathbb{D}_P^{\mathrm{D}}$ here.

**Theorem 5.10.** *Consider a countably additive probability charge* $P$ *on a field of events* $\mathcal{F}$ *over some possibility space* $\mathcal{X}$. *Let* $(f_n)_{n\in\mathbb{N}}$ *be a non-decreasing sequence of D-integrable variables with* $E_P^{\mathrm{D}}(f_1) > -\infty$. *Then the point-wise limit of* $(f_n)_{n\in\mathbb{N}}$ *is D-integrable, and*

$$E_P^{\mathrm{D}}\left(\operatorname*{p\text{-}w\lim}_{n\to+\infty} f_n\right) = \lim_{n\to+\infty} E_P^{\mathrm{D}}(f_n).$$

*The same holds in case* $(f_n)_{n\in\mathbb{N}}$ *is non-increasing and* $E_P^{\mathrm{D}}(f_1) < +\infty$.

In the second convergence theorem, we substitute the requirement that the sequence $(f_n)_{n\in\mathbb{N}}$ of extended real variables should be bounded for the requirement that it should be monotone. This second convergence result is known as *Lebesgue's Dominated Convergence Theorem* (see Taylor, 1985, Theorem 6-3 IV).

**Theorem 5.11.** *Consider a countably additive probability charge* $P$ *on a field of events* $\mathcal{F}$ *over some possibility space* $\mathcal{X}$. *Let* $(f_n)_{n\in\mathbb{N}}$ *be a sequence of D-integrable variables that converges point-wise. If there is a D-integrable variable* $g$ *with* $E_P^{\mathrm{D}}(g) < +\infty$ *such that* $|f_n| \le g$ *for all* $n$ *in* $\mathbb{N}$, *then the point-wise limit of* $(f_n)_{n\in\mathbb{N}}$ *is D-integrable, and*

$$E_P^{\mathrm{D}}\left(\operatorname*{p\text{-}w\lim}_{n\to+\infty} f_n\right) = \lim_{n\to+\infty} E_P^{\mathrm{D}}(f_n).$$

Recall from (DE2) that the set $\mathbb{D}_P^{\mathrm{D}}$ of D-integrable variables includes all $\mathcal{F}$-simple, $\mathcal{F}$-over and $\mathcal{F}$-under variables. In other words, these variables are D-integrable for *every* countably additive probability $P$ on $\mathcal{F}$; quite remarkably, we can identify a lot more variables for which this holds as well. To show this, we recall from the beginning of Section $5.1.2_{219}$ that, by Theorem $\mathrm{C.19}_{471}$ in Appendix $\mathrm{C.3}_{470}$, the Daniell extension $E_P^{\mathrm{D}}$ essentially coincides with the usual extension of $E_P$ in measure-theoretic probability theory. More precisely put, Theorem C.19<sub>471</sub> establishes that the domain of the measure-theoretical extension of $E_P$ is contained in $\mathbb{D}_P^{\mathrm{D}}$, and that this extension coincides with $E_P^{\mathrm{D}}$ on that domain. For this reason, we can use standard results from measure theory to identify a larger class of variables that are D-integrable for any countably additive probability charge $P$ on $\mathcal{F}$.

As we did for the $\mathcal{F}$-over and $\mathcal{F}$-under variables, we consider sequences of $\mathcal{F}$-simple variables that converge point-wise. Here, we do not require that these sequences should be monotone; instead, all that we need is that these sequences are uniformly bounded below or above. A sequence $(f_n)_{n\in\mathbb{N}}$ of extended real variables is called *uniformly bounded below* if there is some real number $\beta$ such that $f_n \geq \beta$ for all $n$ in $\mathbb{N}$. Furthermore, $(f_n)_{n\in\mathbb{N}}$ is said to be *uniformly bounded above* if $(-f_n)_{n\in\mathbb{N}}$ is uniformly bounded below, and simply called *uniformly bounded* if it is uniformly bounded above and below. We collect all extended real variables that are the point-wise limit of a sequence of $\mathcal{F}$-simple variables that is either bounded above or bounded below in

$$\overline{\mathbb{V}}_{\mathrm{lim}}(\mathcal{F}) \coloneqq \overline{\mathbb{V}}_{\mathrm{b}}(\mathcal{F}) \cup \overline{\mathbb{V}}^{\mathrm{b}}(\mathcal{F}), \tag{5.14}$$

where we let $\overline{\mathbb{V}}_{\mathrm{b}}(\mathcal{F})$ denote the set of all extended real variables $f$ in $\overline{\mathbb{V}}$ for which there is a sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variable that is uniformly bounded below and that converges point-wise to $f$, and similarly, we let $\overline{\mathbb{V}}^{\mathrm{b}}(\mathcal{F})$ denote the set of all extended real variables $f$ in $\overline{\mathbb{V}}$ for which there is a sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables that is uniformly bounded above and that converges point-wise to $f$. Note that every $f$ in $\overline{\mathbb{V}}_{\mathrm{b}}(\mathcal{F})$ is bounded below – in the sense that $\inf f > -\infty$ – and every $f$ in $\overline{\mathbb{V}}^{\mathrm{b}}(\mathcal{F})$ is bounded above – in the sense that $\sup f < +\infty$. Furthermore, it is clear that $\overline{\mathbb{V}}_{\mathrm{b}}(\mathcal{F}) = -\overline{\mathbb{V}}^{\mathrm{b}}(\mathcal{F})$, so $\overline{\mathbb{V}}_{\mathrm{lim}}(\mathcal{F})$ is negation invariant.

The following result establishes that every variable in $\overline{\mathbb{V}}_{\mathrm{lim}}(\mathcal{F})$ is always D-integrable. Our proof relies on Theorem C.19<sub>471</sub>, and on the well-known fact that the point-wise limit of a sequence of measurable extended real variables is a measurable extended real variable – see Lemma C.13<sub>467</sub> in Appendix C.2.1<sub>467</sub>. For this reason, we have relegated it to Appendix C.3<sub>470</sub>. Note that in Corollary C.25<sub>474</sub> in Appendix C.3<sub>470</sub>, we also show that $\mathbb{D}_P^{\mathrm{D}}$ contains every $\sigma(\mathcal{F})$-measurable variable – see Appendices C.1.1<sub>464</sub> and C.2.1<sub>467</sub> – that is either bounded below or bounded above.

**Theorem 5.12.** *Consider a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then for any countably additive probability charge $P$ on $\mathcal{F}$,*

$$\mathbb{S}(\mathcal{F}) \subseteq \overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F}) \subseteq \overline{\mathbb{V}}_{\mathrm{lim}}(\mathcal{F}) \subseteq \mathbb{D}_P^{\mathrm{D}}.$$

### 5.1.4 Daniell extension from the point of view of coherence

We have accomplished what we set out to do: we have defined an extension of $E_P$ to point-wise limits of sequences of $\mathcal{F}$-simple variables, and this extension carries over *monotone* limit behaviour of events and variables to their probabilities and expectations in a consistent manner. To tie things up, we take a look at the Daniell extension $E_P^{\mathrm{D}}$ from the point of view of coherence.

In Section 5.1.1$_{217}$, we defined the class $\mathbb{D}_P^{\mathrm{C}}$ of C-integrable gambles as those gambles $g$ in $\mathbb{G}(\mathcal{X})$ whose lower and upper natural extension $\underline{\mathcal{E}}_P(g)$ and $\overline{\mathcal{E}}_P(g)$ coincide. The following result establishes that C-integrable gambles are always D-integrable; we have relegated our proof to Appendix B.2$_{458}$.

**Proposition 5.13.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then for any gamble $g$ in $\mathbb{G}(\mathcal{X})$, $\underline{\mathcal{E}}_P(g) \leq E_P^{\mathrm{i}}(g)$ and $\overline{\mathcal{E}}_P(g) \geq E_P^{\mathrm{o}}(g)$. Consequently, $\mathbb{D}_P^{\mathrm{C}}$ is included in $\mathbb{D}_P^{\mathrm{D}}$, and*

$$E_P^{\mathrm{C}}(g) = E_P^{\mathrm{D}}(g) \quad \text{for all } g \in \mathbb{D}_P^{\mathrm{C}}.$$

By definition, the domain $\mathbb{D}_P^{\mathrm{D}}$ of the Daniell expectation $E_P^{\mathrm{D}}$ contains those extended real variables $f$ whose inner and outer Daniell expectations $E_P^{\mathrm{i}}(f)$ and $E_P^{\mathrm{o}}(f)$ coincide; this is similar to the definition of $E_P^{\mathrm{C}}$, although the lower and upper natural extensions $\underline{\mathcal{E}}_P$ and $\overline{\mathcal{E}}_P$ are replaced by the inner and outer Daniell extensions $E_P^{\mathrm{i}}$ and $E_P^{\mathrm{o}}$. As $E_P^{\mathrm{C}}$ is a coherent expectation, this leads us wondering whether the Daniell expectation $E_P^{\mathrm{D}}$, when restricted to gambles, is coherent as well. That this is indeed the case follows almost immediately from Proposition 2.15$_{22}$ and Theorem 5.9$_{225}$; a formal proof for this result can be found in Appendix B.2$_{458}$.

**Proposition 5.14.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then the restriction of the Daniell extension $E_P^{\mathrm{D}}$ to $\mathbb{G}(\mathcal{X}) \cap \mathbb{D}_P^{\mathrm{D}}$ is a coherent expectation.*

## 5.2 Extending the domain of a jump process

It is about time we put Daniell's extension method to work in the setting of jump processes. In Section 5.2.1, we show that for any jump process $P$ that has (uniformly) bounded rate and for any state history $\{X_u = x_u\}$ in $\mathscr{H}$, the probability charge $P(\bullet \,|\, X_u = x_u)$ on $\mathscr{F}_u$ is countably additive; hence, we can then use Daniell's method to extend the domain of the conditional expectation $E_P$ corresponding to this jump process $P$. Subsequently, in Section 5.2.2$_{231}$ we consider the number of jumps as a first example of a variable in this extended domain; we show that this real variable is the pointwise limit of a non-decreasing sequence of simple variables, and also provide an upper bound on its expectation.

### 5.2.1 Countably additive jump processes

Recall from Section 3.1.3$_{65}$ that the domain $\mathscr{D}$ of a jump process $P$ is a structure of fields. Due to Corollary 2.58$_{46}$, we therefore know that for every state history $\{X_u = x_u\}$ in $\mathscr{H}$, $P(\bullet \,|\, X_u = x_u)$ is a probability charge on $\mathscr{F}_u$. Whenever each of these probability charges is countably additive, we follow Berti et al. (2002, Section 3) and Lopatatzidis (2017, Definition 6) in calling the jump process $P$ countably additive as well.

**Definition 5.15.** A jump process $P$ is *countably additive* if for all $\{X_u = x_u\}$ in $\mathscr{H}$, the probability charge $P(\bullet \,|\, X_u = x_u)$ on $\mathscr{F}_u$ is countably additive.

Suppose $P$ is a countably additive jump process. Because every probability charge $P(\bullet \,|\, X_u = x_u)$ is then countably additive, we can use the material in Section 5.1$_{216}$ to extend the induced conditional expectation $E_P(\bullet \,|\, X_u = x_u)$. More concretely, for any state history $\{X_u = x_u\}$ in $\mathscr{H}$, we obtain the extended conditional expectation

$$E_P^{\mathrm{D}}(\bullet \,|\, X_u = x_u)\colon \overline{\mathbb{V}}_{\lim}(\mathscr{F}_u) \to \overline{\mathbb{R}}\colon f \mapsto E_P^{\mathrm{D}}(f \,|\, X_u = x_u) \coloneqq E^{\mathrm{D}}_{P(\bullet|X_u=x_u)}(f),$$

where $E^{\mathrm{D}}_{P(\bullet|X_u=x_u)}$ is the Daniell extension of the expectation $E_{P(\bullet|X_u=x_u)}$ corresponding to the countably additive probability charge $P(\bullet \,|\, X_u = x_u)$. This way, we have defined the conditional Daniell extension $E_P^{\mathrm{D}}$ on

$$\mathbb{JD} \coloneqq \big\{(f \,|\, X_u = x_u)\colon u \in \mathscr{U}, x_u \in \mathscr{X}_u, f \in \overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)\big\}. \tag{5.15}$$

For any $\{X_u = x_u\}$ in $\mathscr{H}$ and $A$ in $\mathcal{P}(\Omega)$ such that $(\mathbb{I}_A \,|\, X_u = x_u)$ belongs to $\mathbb{JD}$, it makes sense to think of

$$P^{\mathrm{D}}(A \,|\, X_u = x_u) \coloneqq E_P^{\mathrm{D}}(\mathbb{I}_A \,|\, X_u = x_u)$$

as a conditional probability.

Note that we could have also used the more general set $\mathbb{D}^{\mathrm{D}}_{P(\bullet|X_u=x_u)}$ of D-integrable variables in the definition of $E_P^{\mathrm{D}}(\bullet \,|\, X_u = x_u)$, but we have chosen not to for two reasons. The pragmatic reason is that this smaller domain is all we really need in the remainder, because the idealised variables that we will consider belong to the smaller domain $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$. The second, more important, reason is that due to Theorem 5.12$_{227}$, we are guaranteed that the Daniell extension $E^{\mathrm{D}}_{P(\bullet|X_u=x_u)}$ is defined on $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ for *any* countably additive jump process. In other words, this restriction ensures that the domain $\mathbb{JD}$ of the conditional Daniell expectation $E_P^{\mathrm{D}}$ is the same for any countably additive jump process $P$. This will turn out to be convenient in Section 5.3$_{238}$ further on, where for imprecise jump processes that exclusively consist of countably additive jump processes, this will allow us to extend their lower (and upper) envelope to $\mathbb{JD}$. Due to Corollary C.25$_{474}$ in Appendix C.3$_{470}$, another way to ensure that the extensions $E^{\mathrm{D}}_{P(\bullet|X_u=x_u)}$ have the same domain, is to restrict their domain to those extended real variables that are measurable with respect to the σ-field $\sigma(\mathscr{F}_u)$ generated by $\mathscr{F}_u$ – see Appendices C.1.1$_{464}$ and C.2.1$_{467}$ – and that are either bounded below or bounded above. This would yield a larger domain, but the extra variables in this domain do not have an immediate obvious interpretation as being the point-wise limit of a sequence of simple variables. For this reason, we will not pursue this idea here; instead, we refer the interested reader to (Erreygers & De Bock, 2021, Section 4).

### Uniformly bounded rate

Our extension of $E_P$ from $\mathbb{JS}$ to $\mathbb{JD}$ only works if the jump process $P$ is countably additive. Seeing that the conditions of Definition $5.4_{221}$ are not the most easy to check, it would be nice to have an easier sufficient condition for countable additivity tailored to jump processes. With this in mind, we introduce the notion of uniformly bounded rate, which is a stronger version of the notion of 'bounded rate' in Definition $3.53_{101}$.

**Definition 5.16.** A jump process $P$ has *uniformly bounded rate* if there is some non-negative real number $\lambda$ in $\mathbb{R}_{\geq 0}$ such that for any current time point $t$ in $\mathbb{R}_{\geq 0}$ and any state history $\{X_u = x_u\}$ in $\mathscr{H}$ with $u \prec t$,

$$\limsup_{r \searrow t} \left\| Q_{t,r}^{(X_u=x_u)} \right\| \leq \lambda \text{ and, if } t > 0, \limsup_{s \nearrow t} \left\| Q_{s,t}^{(X_u=x_u)} \right\| \leq \lambda.$$

Whenever this is the case, we call $\lambda$ a *rate bound*.

This condition is not only reminiscent of Definition $3.53_{101}$, but also of Proposition $3.42_{91}$ and Lemma $3.55_{102}$. In fact, it follows immediately from Proposition $3.42_{91}$ that any homogeneous Markovian jump process that is characterised by a rate operator has uniformly bounded rate.

**Corollary 5.17.** *Consider a probability mass function $p_0$ on $\mathscr{X}$ and a rate operator $Q$. Then the corresponding homogeneous and Markovian jump process $P_{p_0,Q}$ has uniformly bounded rate, and $\lambda = \|Q\|_{\mathrm{op}}$ is a rate bound.*

Similarly, any jump process that is consistent with a bounded set of rate operators has uniformly bounded rate due to Lemma $3.55_{102}$.

**Corollary 5.18.** *Consider a non-empty and bounded set $\mathcal{Q}$ of rate operators. Then any jump process $P$ that is consistent with $\mathcal{Q}$ has uniformly bounded rate, and $\lambda = \|\mathcal{Q}\|_{\mathrm{op}}$ is a rate bound.*

In short, all of the jump processes that we are interested in have uniformly bounded rate. This is extremely convenient, because having uniformly bounded rate is sufficient for a jump process to be countably additive.

**Theorem 5.19.** *If a jump process $P$ has uniformly bounded rate, then it is countably additive.*

Our proof takes *a lot* of work, which is why we have relegated it to Appendix 5.C$_{252}$. This does not mean that this result is unimportant, though; quite the contrary, we regard Theorem 5.19 as one of the cornerstones of this chapter.

Rather ironically, our proof of Theorem 5.19 makes use of the more conventional measure-theoretic approach to jump processes, and one cause for the excessive length of our proof is that we need to modify that approach to fit our setting. For this reason, we also discuss the measure-theoretical approach in quite some detail in Appendix 5.B$_{245}$. As far as the difference between the measure-theoretic approach and our approach to modelling jump processes is concerned, it suffices to understand that in the former, one starts from the set $\tilde{\Omega}$ of all (so not necessarily càdlàg) paths and subsequently modifies or adapts the projector variables $(\tilde{X}_t)_{t\in\mathbb{R}_{\geq 0}}$ with respect to this set $\tilde{\Omega}$ in such a way as to end up with càdlàg 'sample paths'; in essence, one modifies the outcomes of the original jump process. Whether or not this makes sense is a question that is often overlooked by advocates of the measure-theoretic approach. Furthermore, seeing that the end goal is to obtain a jump process with càdlàg sample paths, the question arises why one does not immediately start with the set $\Omega$ of càdlàg paths as the possibility space; this appears to be something that is rarely done. In contrast, our approach starts from the càdlàg paths, so we do not need to modify the outcomes.

### 5.2.2 The expected number of jumps

A pivotal argument in our proof of Theorem 5.19 is that the notion of uniformly bounded rate allows us to bound the expected number of jumps over a finite time horizon, albeit in the measure-theoretic framework. The number of jumps is also interesting in our framework; in fact, it is the monotone limit of a non-decreasing sequence of simple variables, which makes it a prime example of an idealised variable.

#### Jumps of càdlàg paths

For any càdlàg path $\omega$ in $\Omega$, we say that a jump occurs at time $t$ in $\mathbb{R}_{>0}$ if

$$\lim_{\Delta\searrow 0}\omega(t-\Delta)\neq\omega(t).$$

Note that the limit on the left always exists because $\omega$ is càdlàg – that is, due to Eq. (3.2)$_{58}$. For any time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s\leq r$ and any càdlàg path $\omega$, the corresponding set

$$\mathscr{J}_{[s,r]}(\omega) := \left\{t\in\,]s,r]\colon\lim_{\Delta\searrow 0}\omega(t-\Delta)\neq\omega(t)\right\} \tag{5.16}$$

of the jump times of $\omega$ in $[s,r]$ is always finite, as the following lemma makes clear. This result is essentially well-known, but we provide a proof for the sake of completeness.

**Lemma 5.20.** *Consider a càdlàg path $\omega$ in $\Omega$ and time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. Then the set $\mathscr{J}_{[s,r]}(\omega)$ of the jump times of $\omega$ in the interval $[s,r]$ is finite.*

*Proof.* As the statement holds trivially whenever $s = r$, we may assume without loss of generality that $s < r$. First, we observe that due to Eq. (3.1)$_{58}$, there is an $s'$ in $]s, +\infty[$ such that $\omega(t) = \omega(s)$ for all $t$ in $[s, s']$. Consequently, $\lim_{\Delta \searrow 0} \omega(t-\Delta) = \omega(t) = \omega(s)$ for all $t$ in $[s, s']$. If $s' \geq r$, then clearly $|\mathscr{J}_{[s,r]}(\omega)| = 0$ so the statement is true.

We continue with the case that $s' < r$. Because $\omega$ is constant on $[s, s']$,

$$\begin{aligned}\mathscr{J}_{[s,r]}(\omega) &= \left\{ t \in\, ]s,r] \colon \lim_{\Delta\searrow 0} \omega(t-\Delta) \neq \omega(t) \right\} \\ &= \left\{ t \in\, ]s',r] \colon \lim_{\Delta\searrow 0} \omega(t-\Delta) \neq \omega(t) \right\}. \end{aligned} \tag{5.17}$$

Fix any $t$ in the closed interval $[s', r]$. Observe that, due to Eqs. (3.1)$_{58}$ and (3.2)$_{58}$, there is a $\delta_t^-$ in $\mathbb{R}_{>0}$ such that $\omega(t') = \omega(t'')$ for all $t', t''$ in $]t-\delta_t^-, t[$ and a $\delta_t^+$ in $\mathbb{R}_{>0}$ such that $\omega(t') = \omega(t)$ for all $t'$ in $]t, t+\delta_t^+[$. Thus, the open interval $]t-\delta_t^-, t+\delta_t^+[$ contains at most a single time point $t'$ where $\lim_{\Delta\searrow 0} \omega(t'-\Delta) \neq \omega(t')$: the point $t$ itself.

Observe that $\left(]t-\delta_t^-, t+\delta_t^+[\right)_{t\in[s',r]}$ is an open cover of $[s', r]$. Because $[s', r]$ is a closed and bounded subset of $\mathbb{R}$, it follows from the Heine-Borel theorem that it is compact. Consequently, there is a finite subcover $\left(]t-\delta_t^-, t+\delta_t^+[\right)_{t\in\mathscr{C}}$ of $[s', r]$, where $\mathscr{C}$ is a finite subset of $[s', r]$. Recall that by construction, the open interval $]t-\delta_t^-, t+\delta_t^+[$ contains at most a single jump point of the path $\omega$ for every $t$ in $\mathscr{C}$, and if it contains one, then it must be equal to $t$. For this reason, and by Eq. (5.17),

$$\mathscr{J}_{[s,r]}(\omega) = \left\{ t \in\, ]s',r] \colon \lim_{\Delta\searrow 0} \omega(t-\Delta) \neq \omega(t) \right\} \subseteq \mathscr{C}.$$

Since the set on the right-hand side of the inclusion is finite, the set on the left-hand side of the inclusion is finite as well. □

Due to the preceding lemma, for all time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, the number of jumps in $[s, r]$, denoted by

$$\eta_{[s,r]} \colon \Omega \to \mathbb{R}_{\geq 0} \colon \omega \mapsto \eta_{[s,r]}(\omega) := \left|\mathscr{J}_{[s,r]}(\omega)\right|, \tag{5.18}$$

is a non-negative real-valued variable. To prove that $\eta_{[s,r]}$ is the point-wise limit of simple variables, we will make use of grids of time points.

### Grids of time points

Consider a closed interval $[s, r]$ of the time axis $\mathbb{R}_{\geq 0}$. A grid over $[s, r]$ is a non-empty sequence of time points that starts in $s$ and ends in $r$, as depicted in Fig. 5.1$_{\curvearrowright}$. More formally, for all $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, a grid $\upsilon = (t_0, \ldots, t_n)$ over $[s, r]$ is a non-empty sequence of time points that starts in $t_0 = s$ and ends in $t_n = r$. Thus, a grid $\upsilon = (t_0, \ldots, t_n)$ over $[s, r]$ divides the interval $[s, r]$ in $n$ subintervals $[t_{k-1}, t_k]$, with $k$ in $\{1, \ldots, n\}$. Note that in

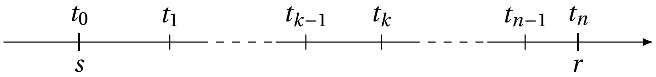



**Figure 5.1** A grid $(t_0,\dots,t_n)$ over the time interval $[s,r]$.

contrast to general sequences of time points, the first time point of a grid has index 0 instead of index 1; we do this for aesthetic reasons, as will become clear later on. We collect all grids over $[s,r]$ in the set

$$\mathscr{U}_{[s,r]} = \big\{(t_0,\dots,t_n) \in \mathscr{U}_{\neq()} \colon\, t_0 = s, t_n = r\big\}; \tag{5.19}$$

note that $\mathscr{U}_{[s,s]}$ is a singleton because there is precisely one sequence of time points that starts and ends in $s$: the monuple $(s)$.

For any grid $v = (t_0,\dots,t_n)$ in $\mathscr{U}_{\neq()}$ with $n \geq 1$, we define the grid width $\Delta(v)$ of $v$ as the width of the largest subinterval:

$$\Delta(v) \coloneqq \max\big\{t_k - t_{k-1} \colon\, k \in \{1,\dots,n\}\big\}. \tag{5.20}$$

As a special case, we set $\Delta(v) = 0$ for the degenerate grid $v = (s)$ over the degenerate interval $[s,s]$. If $v$ and $w$ are grids over $[s,r]$, then we say that $w$ *refines* $v$ whenever every subinterval in $w$ is a subset of some subinterval in $v$, or equivalently, whenever the time points in $w$ include the time points in $v$; hence, we denote this by $w \supseteq v$. Note that in this case, $\Delta(w) \leq \Delta(v)$.

### Expected number of jumps as a limit

Let $s$ and $r$ be time points in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. For any grid $v = (t_0,\dots,t_n)$ over $[s,r]$, the number of jumps along the grid $v$ is the non-negative real variable

$$\eta_v \colon \Omega \to \mathbb{R}_{\geq 0} \colon \omega \mapsto \eta_v(\omega) \coloneqq \Big|\big\{k \in \{1,\dots,n\} \colon\, \omega(t_{k-1}) \neq \omega(t_k)\big\}\Big|. \tag{5.21}$$

It is easy to see that

$$\eta_v = \sum_{k=1}^{n} \mathbb{I}_{\{X_{t_{k-1}} \neq X_{t_k}\}}, \tag{5.22}$$

where we follow the convention that the empty sum is zero and where for any $t$ and $t'$ in $\mathbb{R}_{\geq 0}$, we let

$$\{X_t \neq X_{t'}\} \coloneqq \big\{\omega \in \Omega \colon\, \omega(t) \neq \omega(t')\big\}. \tag{5.23}$$

It is intuitively clear that $\eta_v$ is an $\mathscr{F}_u$-simple variable for any sequence of time points $u$ in $\mathscr{U}$ such that $\max u \leq s$, but we nonetheless establish this formally.

**Lemma 5.21.** *Consider a sequence of time points $u$ in $\mathscr{U}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that* $\max u \leq s \leq r$. *Let $v = (t_0, \dots, t_n)$ be a grid over $[s, r]$. Then for all $k$ in $\{1, \dots, n\}$, $\{X_{t_{k-1}} \neq X_{t_k}\}$ belongs to $\mathscr{F}_u$. Therefore, $\eta_v$ is an $\mathscr{F}_u$-simple variable.*

*Proof.* Note that for all $k$ in $\{1, \dots, n\}$, $(t_{k-1}, t_k) \succcurlyeq u$ and

$$\{X_{t_{k-1}} \neq X_{t_k}\} = \{X_{(t_{k-1}, t_k)} \in B\} \quad \text{with } B := \left\{(x, y) \in \mathscr{X}^2 \colon x \neq y\right\},$$

so $\{X_{t_{k-1}} \neq X_{t_k}\}$ belongs to $\mathscr{F}_u$ due to Eq. (3.16)63. For this reason, it follows immediately from Eq. (5.22) that $\eta_v$ is $\mathscr{F}_u$-simple. □

We now establish some convenient properties of $\eta_v$. First, we observe that we can add the number of jumps over two grids if the last time point of the first coincides with the first time point of the second.

**Lemma 5.22.** *Consider time points $s, t, r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r \leq t$. If $v_1$ is a grid over $[s, t]$ and $v_2$ is a grid over $[t, r]$, then $v := v_1 \cup v_2$ is a grid over $[s, r]$, and*

$$\eta_v = \eta_{v_1} + \eta_{v_2}.$$

*Proof.* Follows immediately from Eq. (5.22). □

Second, we observe that whenever we refine a grid, the number of jumps stays the same or increases in steps of 2.

**Lemma 5.23.** *Consider time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, and two grids $v$ and $w$ over $[s, r]$ such that $w$ refines $v$ – that is, $w \supseteq v$. Then for all $\omega$ in $\Omega$, there is some $k_\omega$ in $\mathbb{Z}_{\geq 0}$ such that*

$$\eta_w(\omega) = \eta_v(\omega) + 2k_\omega;$$

*consequently, $\eta_w \geq \eta_v$.*

Lemma 5.23 follows almost immediately from Lemma 5.22 and the following straightforward observation.

**Lemma 5.24.** *Consider time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $s < r$ and a grid $v = (t_0, \dots, t_n)$ over $[s, r]$ with $n \geq 2$. Then*

$$\eta_v = \eta_{(s,r)} + 2 \sum_{k=1}^{n-1} \mathbb{I}_{\left\{X_{t_{k-1}} \neq X_{t_k} \neq X_r\right\}}.$$

*Proof.* Crucial to our proof is the following observation. Let $w = (s_0, \dots, s_m)$ be any grid over $[s, r]$, and fix some time point $t$ in $]s_{m-1}, s_m[$. Then for all $\omega$ in $\Omega$,

$$\eta_{w \cup (t)}(\omega) = \begin{cases} \eta_w(\omega) + 2 & \text{if } \omega(s_{m-1}) \neq \omega(t) \neq \omega(s_n), \\ \eta_w(\omega) & \text{otherwise.} \end{cases}$$

Hence,

$$\eta_{w\cup(t)} = \eta_w + 2\mathbb{I}_{\{X_{s_{m-1}}\neq X_t \neq X_{s_m}\}}. \tag{5.24}$$

Fix some $\omega$ in $\Omega$, and let $v_0 := (s, r)$. Furthermore, for all $k$ in $\{1, \ldots, n-1\}$, we let $v_k := (t_0, t_1, \ldots, t_k, t_n)$; note that $v_{n-1} = v$. Then it follows from Eq. (5.24) that for all $k$ in $\{1, \ldots, n-1\}$,

$$\eta_{v_k} = \eta_{v_{k-1}} + 2\mathbb{I}_{\{X_{t_{k-1}}\neq X_{t_k}\neq X_r\}}.$$

We repeatedly apply the preceding equality, to yield

$$\eta_v = \eta_{v_{n-1}} = \eta_{v_{n-2}} + 2\mathbb{I}_{\{X_{t_{n-2}}\neq X_{t_{n-1}}\neq X_r\}} = \cdots = \eta_{(s,r)} + 2\sum_{k=1}^{n-1}\mathbb{I}_{\{X_{t_{k-1}}\neq X_{t_k}\neq X_r\}}. \qquad \square$$

*Proof of Lemma 5.23.* The statement is clearly trivial in case $[s, r]$ is a degenerate interval, so we assume without loss of generality that $s < r$. Enumerate the time points in $v$ as $(t_0, \ldots, t_n)$, and note that $n \geq 1$ because $s < r$. For all $\ell$ in $\{1, \ldots, n\}$, we let $w_\ell$ be the sequence of time points that consists of those time points in $w$ that belong to $[t_{\ell-1}, t_\ell]$; because $w$ refines $v$, $w_\ell$ is a grid over $[t_{\ell-1}, t_\ell]$. It follows from repeated application of Lemma 5.22 that

$$\eta_v = \sum_{\ell=1}^{n} \eta_{(t_{\ell-1},t_\ell)} \quad \text{and} \quad \eta_w = \sum_{\ell=1}^{n} \eta_{w_\ell}. \tag{5.25}$$

Fix some $\omega$ in $\Omega$. Then it follows from Lemma 5.24 that for all $\ell$ in $\{1, \ldots, n\}$, there is a non-negative integer $k_{\omega,\ell}$ such that

$$\eta_{w_\ell}(\omega) = \eta_{(t_{\ell-1},t_\ell)}(\omega) + 2k_{\omega,\ell}.$$

It follows immediately from this and Eq. (5.25) that

$$\eta_w(\omega) = \sum_{\ell=1}^{n} \eta_{w_\ell}(\omega) = \sum_{\ell=1}^{n} \left(\eta_{(t_{\ell-1},t_\ell)}(\omega) + 2k_{\omega,\ell}\right) = \eta_v(\omega) + \sum_{\ell=1}^{n} 2k_{\omega,\ell} = \eta_v(\omega) + 2k_\omega,$$

where we let $k_\omega := \sum_{\ell=1}^{n} k_{\omega,\ell}$. $\square$

Third, we establish that for every càdlàg path $\omega$, there is some grid width $\Delta^{\omega}_{[s,r]}$ such that every grid $v$ with a grid width $\Delta(v)$ smaller than $\Delta^{\omega}_{[s,r]}$ captures all jumps of the path $\omega$ in $[s, r]$.

**Lemma 5.25.** *Consider time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. For any càdlàg path $\omega$ in $\Omega$, there is a positive real number $\Delta^{\omega}_{[s,r]}$ such that if $v$ is a grid over $[s, r]$ with $\Delta(v) < \Delta^{\omega}_{[s,r]}$, then*

$$\eta_v(\omega) = \eta_{[s,r]}(\omega).$$

*Proof.* Fix some path $\omega$ in $\Omega$. Recall from Lemma 5.20232 that the set $\mathcal{J}_{[s,r]}(\omega)$ of the jump times of $\omega$ in $[s, r]$ is finite. If this set is empty, then $\eta_{[s,r]}(\omega) = 0$ and $\omega$ is clearly constant over $[s, r]$. Consequently, $\eta_v(\omega) = 0$ for any grid $v$ of time points over $[s, r]$, so the statement holds for any positive real number $\Delta^{\omega}_{[s,r]}$.

If on the other hand the set $\mathcal{J}_{[s,r]}(\omega)$ of jump times is non-empty, then we can order the time points in this finite set. This way, we obtain a grid $(t_0, \ldots, t_n)$ over $[s, r]$

– where we always add $t_0 = s$ and only add $t_n = r$ if it is not a jump time. Note that for all $k$ in $\{1, \dots, n\}$, $\omega(t) = \omega(t_{k-1})$ for all $t$ in $[t_{k-1}, t_k[$, so $\omega$ is constant over $[t_{k-1}, t_k[$. Let

$$\Delta^{\omega}_{[s,r]} := \min\{t_k - t_{k-1} : k \in \{1, \dots, n\}\},$$

and take any grid $v$ over $[s,r]$ such that $\Delta(v) < \Delta^{\omega}_{[s,r]}$. It follows from this condition on $v$ that, for all $k$ in $\{1, \dots, n\}$, $v$ contains at least one time point $t$ in the subinterval $[t_{k-1}, t_k[$ where $\omega$ is constant. For this reason, $\eta_{[s,r]}(\omega) = \eta_v(\omega)$, as required. □

Finally, we can use the three preceding intermediary results to establish that for every $u$ in $\mathscr{U}$ with $\max u \leq s$, $\eta_{[s,r]}$ is the monotone limit of a sequence of $\mathscr{F}_u$-simple variables; hence, we can determine the (conditional) expected number of jumps as the limit of the expectation of these simple variables.

**Theorem 5.26.** *Consider a sequence of time points $u$ in $\mathscr{U}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Let $(v_n)_{n\in\mathbb{N}}$ be a sequence of grids over $[s,r]$ with $\lim_{n\to+\infty} \Delta(v_n) = 0$. Then $(\eta_{v_n})_{n\in\mathbb{N}}$ is a sequence of non-negative $\mathscr{F}_u$-simple variables that converges point-wise to $\eta_{[s,r]}$, so $\eta_{[s,r]}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$. If furthermore $v_n \subseteq v_{n+1}$ for all $n$ in $\mathbb{N}$, then the sequence $(\eta_{v_n})_{n\in\mathbb{N}}$ is non-decreasing. Hence, in that case, for any countably additive jump process $P$ and any $x_u$ in $\mathscr{X}_u$,*

$$E^{\mathrm{D}}_P(\eta_{[s,r]} \,|\, X_u = x_u) = \lim_{n\to+\infty} E_P(\eta_{v_n} \,|\, X_u = x_u).$$

*Proof.* For all $n$ in $\mathbb{N}$, we recall from Lemma 5.21234 that $\eta_{v_n}$ is $\mathscr{F}_u$-simple; that $\eta_{v_n}$ is non-negative follows immediately from Eq. (5.22)233. Because $\lim_{n\to+\infty} \Delta(v_n) = 0$ by assumption, it follows from Lemma 5.25↶ that for every càdlàg path $\omega$, $\lim_{n\to+\infty} \eta_{v_n}(\omega) = \eta_{[s,r]}(\omega)$. Thus, $(\eta_{v_n})_{n\in\mathbb{N}}$ is a sequence of $\mathscr{F}_u$-simple variables that is uniformly bounded below and that converges point-wise to $\eta_{[s,r]}$. This implies that $\eta_{[s,r]}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$.

In case $v_n \subseteq v_{n+1}$ for all $n$ in $\mathbb{N}$, it follows immediately from Lemma 5.23234 that the sequence $(\eta_{v_n})_{n\in\mathbb{N}}$ is non-decreasing. Furthermore, $E^{\mathrm{D}}_P(\eta_{v_1} \,|\, X_u = x_u) \geq 0$ due to (DE3)225 because $\eta_{v_1} \geq 0$. Therefore, it follows from Theorem 5.10226 and (DE1)225 that

$$E^{\mathrm{D}}_P(\eta_{[s,r]} \,|\, X_u = x_u) = \lim_{n\to+\infty} E_P(\eta_{v_n} \,|\, X_u = x_u),$$

which completes our proof. □

### An upper bound for the expected number of jumps

For jump processes that have uniformly bounded rate, any rate bound $\lambda$ naturally bounds the expected number of jumps. More precisely, the following result establishes that $\lambda/2$ is an upper bound on the expected number of jumps per unit of time.

**Theorem 5.27.** *Consider a jump process $P$ that has uniformly bounded rate, with rate bound $\lambda$. Fix some state history $\{X_u = x_u\}$ in $\mathscr{H}$ and time points $s, r$*

*such that* $\max u \leq s \leq r$. *Then for any grid* $v$ *over* $[s,r]$,

$$E_P(\eta_v \mid X_u = x_u) \leq (r-s)\frac{\lambda}{2};$$

*consequently,*

$$E_P^{\mathrm{D}}(\eta_{[s,r]} \mid X_u = x_u) \leq (r-s)\frac{\lambda}{2}.$$

Our proof of Theorem 5.27 relies on the following technical result, the proof of which has been relegated to Appendix 5.A$_{242}$.

**Lemma 5.28.** *Consider a jump process* $P$ *that has uniformly bounded rate with rate bound* $\lambda$. *Fix a state history* $\{X_u = x_u\}$ *in* $\mathscr{H}$ *and time points* $s, r$ *in* $\mathbb{R}_{\geq 0}$ *such that* $\max u \leq s < r$. *Then for any positive real number* $\epsilon$ *in* $\mathbb{R}_{>0}$, *there is a grid* $v = (t_0, \ldots, t_n)$ *over* $[s,r]$ *such that*

$$\big(\forall k \in \{1,\ldots,n\}\big)\; P(X_{t_{k-1}} \neq X_{t_k} \mid X_u = x_u) < (t_k - t_{k-1})\left(\frac{\lambda}{2} + \epsilon\right).$$

*Proof of Theorem 5.27.* Observe that the statement is trivial whenever $s = r$, so we may assume that $s < r$ without loss of generality. Let $v = (t_0, \ldots, t_n)$ be any grid over $[s,r]$. It follows from repeated application of Lemma 5.22$_{234}$ and the additivity of $E_P$ – that is, (ES3)$_{37}$ – that

$$E_P(\eta_v \mid X_u = x_u) = E_P\left(\sum_{k=1}^{n} \eta_{(t_{k-1},t_k)} \,\middle|\, X_u = x_u\right) = \sum_{k=1}^{n} E_P(\eta_{(t_{k-1},t_k)} \mid X_u = x_u). \quad (5.26)$$

Fix an arbitrary positive real number $\epsilon$. For every $k$ in $\{1,\ldots,n\}$, we use Lemma 5.28 to obtain a grid $v_k = (t_{k,0}, \ldots, t_{k,n_k})$ over $[t_{k-1}, t_k]$ such that

$$(\forall \ell \in \{1,\ldots,n_k\})\; P\big(X_{t_{k,\ell-1}} \neq X_{t_{k,\ell}} \,\big|\, X_u = x_u\big) < (t_{k,\ell} - t_{k,\ell-1})\lambda_\epsilon,$$

with $\lambda_\epsilon \coloneqq \epsilon + \lambda/2$. It follows immediately from this, Lemma 5.21$_{234}$ and Eqs. (5.22)$_{233}$ and (2.19)$_{36}$ that for every $k$ in $\{1,\ldots,n\}$,

$$\begin{aligned} E_P(\eta_{v_k} \mid X_u = x_u) &= \sum_{\ell=1}^{n_k} P\big(X_{t_{k,\ell-1}} \neq X_{t_{k,\ell}} \,\big|\, X_u = x_u\big) \\ &< \sum_{\ell=1}^{n_k} (t_{k,\ell} - t_{k,\ell-1})\lambda_\epsilon = (t_{k,n_k} - t_{k,0})\lambda_\epsilon \\ &= (t_k - t_{k-1})\lambda_\epsilon. \end{aligned}$$

Recall from Lemma 5.23$_{234}$ that $\eta_{v_k} \geq \eta_{(t_{k-1},t_k)}$ because $v_k$ refines $(t_{k-1}, t_k)$. It follows from this, the above inequality and the monotonicity of $E_P$ – that is, (ES4)$_{37}$ – that for all $k$ in $\{1,\ldots,n\}$,

$$E_P(\eta_{(t_{k-1},t_k)} \mid X_u = x_u) \leq E_P(\eta_{v_k} \mid X_u = x_u) < (t_k - t_{k-1})\lambda_\epsilon.$$

Because this inequality holds for any positive real number $\epsilon$, we conclude that

$$E_P(\eta_{(t_{k-1},t_k)} \mid X_u = x_u) \leq (t_k - t_{k-1})\frac{\lambda}{2}. \quad (5.27)$$

We now combine Eqs. (5.26)↶ and (5.27)↶, to yield

$$E_P(\eta_v \mid X_u = x_u) \leq \sum_{k=1}^{n} (t_k - t_{k-1}) \frac{\lambda}{2} = (t_n - t_0) \frac{\lambda}{2} = (r - s) \frac{\lambda}{2}, \tag{5.28}$$

establishing the first part of the statement.

The second part of the statement essentially follows from the first part due to Theorem 5.26$_{236}$. Let $(v_n)_{n \in \mathbb{N}}$ be the sequence grids over $[s, r]$ such that for any natural number $n$ in $\mathbb{N}$, $v_n$ is the grid over $[s, r]$ that divides $[s, r]$ in $2^n$ subintervals of equal length. Then by construction, $\lim_{n \to +\infty} \Delta(v_n) = \lim_{n \to +\infty} {}^{(r-s)}/_{2^n} = 0$ and $v_n \subseteq v_{n+1}$ for all $n$ in $\mathbb{N}$. Hence, it follows from Theorem 5.26$_{236}$ that

$$E_P^{\mathrm{D}}(\eta_{[s,r]} \mid X_u = x_u) = \lim_{n \to +\infty} E_P(\eta_{v_n} \mid X_u = x_u),$$

The second inequality of the statement now follows immediately from this limit expression and Eq. (5.28). □

## 5.3 Extending the domain of the lower and upper expectations of an imprecise jump process

If every jump process $P$ in an imprecise jump process $\mathscr{P}$ is countably additive, then every induced conditional expectation operator $E_P$ on $\mathbb{JS}$ can be extended to a conditional expectation operator $E_P^{\mathrm{D}}$ on $\mathbb{JD}$. Instead of taking the lower and upper envelope of the conditional expectations $E_P$ on $\mathbb{JS}$, we can therefore also take the lower and upper envelope of the conditional Daniell extensions $E_P^{\mathrm{D}}$ on $\mathbb{JD}$. Thus, we define the lower envelope $\underline{E}_{\mathscr{P}}^{\mathrm{D}}$ on $\mathbb{JD}$ by

$$\underline{E}_{\mathscr{P}}^{\mathrm{D}}(f \mid X_u = x_u) := \inf\{E_P^{\mathrm{D}}(f \mid X_u = x_u) \colon P \in \mathscr{P}\} \quad \text{for all } (f \mid X_u = x_u) \in \mathbb{JD},$$

and the upper envelope $\overline{E}_{\mathscr{P}}^{\mathrm{D}}$ on $\mathbb{JD}$ by

$$\overline{E}_{\mathscr{P}}^{\mathrm{D}}(f \mid X_u = x_u) := \sup\{E_P^{\mathrm{D}}(f \mid X_u = x_u) \colon P \in \mathscr{P}\} \quad \text{for all } (f \mid X_u = x_u) \in \mathbb{JD},$$

where we will leave the conditioning event $\{X_u = x_u\}$ out of our notation whenever it is the sure event $\Omega$. Due to (DE4)$_{225}$, the lower and upper envelope satisfy the conjugacy relation

$$\overline{E}_{\mathscr{P}}^{\mathrm{D}}(f \mid X_u = x_u) = -\underline{E}_{\mathscr{P}}^{\mathrm{D}}(-f \mid X_u = x_u) \quad \text{for all } (f \mid X_u = x_u) \in \mathbb{JD}. \tag{5.29}$$

We extend our notation for lower and upper probabilities as well, so for all events $A$ in $\mathcal{P}(\Omega)$ and state histories $\{X_u = x_u\}$ in $\mathscr{H}$ such that $(\mathbb{I}_A \mid X_u = x_u)$ belongs to $\mathbb{JD}$, we let

$$\underline{P}_{\mathscr{P}}^{\mathrm{D}}(A \mid X_u = x_u) := \underline{E}_{\mathscr{P}}^{\mathrm{D}}(\mathbb{I}_A \mid X_u = x_u)$$

and

$$\overline{P}_{\mathscr{P}}^{\mathrm{D}}(A \mid X_u = x_u) := \overline{E}_{\mathscr{P}}^{\mathrm{D}}(\mathbb{I}_A \mid X_u = x_u).$$

In particular, we can extend the domain of the lower (and upper) expectations with respect to $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. First, we establish that we can do this for $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$.

**Corollary 5.29.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions and a non-empty set $\mathscr{Q}$ of rate operators. Then every jump process $P$ in $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is countably additive.*

*Proof.* Due to Eq. (3.46)$_{90}$ and Corollary 5.17$_{230}$, every $P$ in $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ has uniformly bounded rate. Therefore, it follows from Theorem 5.19$_{230}$ that every $P$ in $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is countably additive. □

Second, we establish that every jump process in $\mathbb{P}_{\mathscr{Q}}$ is countably additive, given that $\mathscr{Q}$ is bounded.

**Corollary 5.30.** *Consider a non-empty and bounded set $\mathscr{Q}$ of rate operators. Then every imprecise jump process $\mathscr{P}$ such that $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$ only contains countably additive jump processes. In particular, for every non-empty set $\mathscr{M}$ of initial mass functions, the imprecise jump processes $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ only contain countably additive jump processes.*

*Proof.* Follows immediately from Corollary 5.18$_{230}$ and Theorem 5.19$_{230}$. □

Due to the this last result, we can extend the domain of the conditional lower expectations $\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\underline{E}_{\mathscr{M},\mathscr{Q}}$ from $\mathbb{JS}$ to $\mathbb{JD}$ whenever $\mathscr{Q}$ is bounded, and the same holds for the conjugate upper expectations. In order not to needlessly complicate our notation, we will do this implicitly.

### 5.3.1 Continuity properties of the lower and upper envelopes

Suppose $\mathscr{P}$ is an imprecise jump process that consists of countably additive jump processes. Then for every jump process $P$ in $\mathscr{P}$, the (conditional) Daniell expectation $E^{\mathrm{D}}_{P}$ is continuous with respect to point-wise convergence for (some) sequences of idealised variables – monotone ones due to Theorem 5.10$_{226}$ and 'dominated' ones due to Theorem 5.11$_{226}$. Unfortunately, the lower and upper expectations $\underline{E}^{\mathrm{D}}_{\mathscr{P}}$ and $\overline{E}^{\mathrm{D}}_{\mathscr{P}}$ with respect to $\mathscr{P}$ do not necessarily have the same continuity properties.

This potential lack of continuity is not exclusive to the (conditional) lower and upper envelopes with respect to an imprecise jump process. More precisely, Miranda et al. (2017, Section 5.1) establish that for any given set $\mathcal{M}$ of coherent expectations on some domain $\mathcal{G} \subseteq \mathbb{G}(\mathcal{X})$ that are continuous with respect to monotone sequences,[2] the corresponding lower envelope $\underline{E}_{\mathcal{M}}$ is always continuous with respect to monotone non-increasing sequences but may not be continuous with respect to non-decreasing sequences.

[2] More precisely, that satisfy (i) and (ii) of Definition 5.4$_{221}$ on $\mathcal{G}$.

The following two results formally establish the potential lack of continuity in our setting of imprecise jump processes. Because the established properties are essentially well-known, we have relegated their proofs to Appendix $5.\mathrm{D}_{269}$.

**Theorem 5.31.** *Consider an imprecise jump process $\mathscr{P}$ that consists of countably additive jump processes. Fix some $\{X_u = x_u\}$ in $\mathscr{H}$ and $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$, and let $(f_n)_{\in\mathbb{N}}$ be a sequence of variables in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ that converges monotonically to $f$. If $(f_n)_{n\in\mathbb{N}} \nearrow f$ and $\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_1 \,|\, X_u = x_u) > -\infty$, then*

$$\lim_{n\to+\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \leq \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u)$$

*and*

$$\lim_{n\to+\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) = \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u).$$

*Similarly, if $(f_n)_{n\in\mathbb{N}} \searrow f$ and $\overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_1 \,|\, X_u = x_u) < +\infty$, then*

$$\lim_{n\to+\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) = \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u)$$

*and*

$$\lim_{n\to+\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \geq \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u).$$

**Theorem 5.32.** *Consider an imprecise jump process $\mathscr{P}$ that consists of countably additive jump processes. Fix some $\{X_u = x_u\}$ in $\mathscr{H}$ and $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$, and let $(f_n)_{\in\mathbb{N}}$ be a sequence of variables in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ that converges point-wise to $f$. If there is some $g$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ with $\overline{E}^{\mathrm{D}}_{\mathscr{P}}(g \,|\, X_u = x_u) < +\infty$ such that $|f_n| \leq g$ for all $n$ in $\mathbb{N}$, then*

$$\limsup_{n\to+\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \leq \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u) \\ \leq \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u) \leq \liminf_{n\to+\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u).$$

Often, we will invoke Theorems 5.31 and 5.32 for a sequence $(f_n)_{n\in\mathbb{N}}$ of variables that trivially satisfies the additional conditions in the statement. For Theorem 5.31, the condition $\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_1 \,|\, X_u = x_u) > -\infty$ in the first part of the statement is trivially satisfied for a non-decreasing sequence of variables that is uniformly bounded below due to $(\mathrm{DE3})_{225}$, and similarly for the condition in the second part of the statement.

**Corollary 5.33.** *Consider an imprecise jump process $\mathscr{P}$ that consists of countably additive jump processes. Fix some $\{X_u = x_u\}$ in $\mathscr{H}$ and $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$, and let $(f_n)_{\in\mathbb{N}}$ be a sequence of variables in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ that converges monotonically to $f$. If $(f_n)_{n\in\mathbb{N}} \nearrow f$ and $\inf f_1 > -\infty$, then*

$$\lim_{n\to+\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \leq \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u)$$

*and*

$$\lim_{n\to+\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) = \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u).$$

*Similarly, if* $(f_n)_{n\in\mathbb{N}} \searrow f$ *and* $\sup f_1 < +\infty$*, then*

$$\lim_{n\to+\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) = \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u)$$

*and*

$$\lim_{n\to+\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \geq \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u).$$

*Proof.* Follows immediately from Theorem 5.31 and (DE3)$_{225}$. □

For a sequence $(f_n)_{n\in\mathbb{N}}$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ that is uniformly bounded, the condition in Theorem 5.32 concerning $g$ is trivially satisfied due to (DE6)$_{226}$. Hence, we have the following immediate consequence to Theorem 5.32.

**Corollary 5.34.** *Consider an imprecise jump process* $\mathscr{P}$ *that consists of countably additive jump processes. Fix some* $\{X_u = x_u\}$ *in* $\mathscr{H}$ *and* $f$ *in* $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$*, and let* $(f_n)_{\in\mathbb{N}}$ *be a uniformly bounded sequence of variables in* $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ *that converges point-wise to* $f$*. Then*

$$\begin{aligned}\limsup_{n\to+\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) &\leq \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u) \\ &\leq \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u) \leq \liminf_{n\to+\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u).\end{aligned}$$

*Proof.* Because $(f_n)_{n\in\in\mathbb{N}}$ is uniformly bounded by assumption,

$$\beta := \sup\{\sup|f_n| \colon n \in \mathbb{N}\} < +\infty$$

Note that $\overline{E}^{\mathrm{D}}_{\mathscr{P}}(\beta \,|\, X_u = x_u) = \beta < +\infty$ due to Lemma 2.39$_{36}$, (DE1)$_{225}$ and (DE3)$_{225}$. Because furthermore $|f_n| \leq \beta$ for all $n$ in $\mathbb{N}$, the statement follows immediately from Theorem 5.32 with $g = \beta$. □

Unfortunately, the inequalities in these results can be strict. In Joseph's Example 6.17$_{287}$ in Chapter 6$_{273}$ further on, we will construct an example where this is indeed the case.

### An alternative approach, and why we do not follow it

At this point, it is only fair to mention again that Troffaes et al. (2014, Chapter 15) also propose a way to extend a coherent lower expectation $\underline{E}$ from gambles to real variables. They start from a coherent lower expectation $\underline{E}$ on the set $\mathbb{G}(\mathcal{X})$ of all gambles, and then extend the domain of $\underline{E}$ to the 'previsible' real variables, where a real variable $f$ is 'previsible' if there is a sequence $(f_n)_{n\in\mathbb{N}}$ of gambles that converge 'in upper probability' to $f$ and that satisfies a Cauchy-like condition (see Troffaes et al., 2014, Definition 15.6). They subsequently show that their extension is continuous for any sequence of previsible real variables that converges 'in upper probability' and satisfies a Cauchy condition (Troffaes et al., 2014, Theorem 15.16), and they also establish a generalisation of Lebesgue's Dominated Convergence Theorem (Troffaes et al., 2014, Theorem 15.25). As mentioned before in Section 5.1.1$_{217}$, the reason why we do not use their extension is twofold. First and foremost, our setting does not line up well with theirs, because we start from a coherent expectation $E_P(\bullet \mid X_u = x_u)$ on $\mathbb{S}(\mathscr{F}_u)$ – or, alternatively, a coherent lower expectation $\underline{E}_{\mathscr{P}}(\bullet \mid X_u = x_u)$ on $\mathbb{S}(\mathscr{F}_u)$ or even a coherent conditional lower expectation $\underline{E}_{\mathscr{P}}$ on $\mathbb{JS}$. The second reason is that Troffaes et al. (2014) only consider real variables, and we are also interested in extended real variables.

### How to deal with the potential lack of continuity

The (potential) lack of continuity of our approach is perhaps a bit disappointing, but for the most part, it poses no real issue for the idealised variables and imprecise jump processes that we are interested in. In Chapter 6$_{273}$ further on, we investigate the continuity properties of the lower and upper envelopes for four types of idealised variables, including the number of jumps. There, we show that in (most of) these particular cases, the lower and upper envelopes *are* continuous, in the sense that the inequalities in Theorems 5.31$_{240}$ and 5.32$_{240}$ hold with equality.

## 5.A Proof of Lemma 5.28

In the first appendix to this chapter, we prove Lemma 5.28$_{237}$. For the sake of convenience, we repeat the formulation here.

**Lemma 5.28.** *Consider a jump process $P$ that has uniformly bounded rate with rate bound $\lambda$. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Then for any positive real number $\epsilon$ in $\mathbb{R}_{>0}$, there is a grid $v = (t_0, \dots, t_n)$ over $[s, r]$ such that*

$$\big(\forall k \in \{1, \dots, n\}\big)\; P(X_{t_{k-1}} \neq X_{t_k} \mid X_u = x_u) < (t_k - t_{k-1})\left(\frac{\lambda}{2} + \epsilon\right).$$

The following intermediary result will come in handy in our proof.

**Lemma 5.35.** *Consider a jump process $P$ that has uniformly bounded rate with rate bound $\lambda$. Then for any state history $\{X_u = x_u\}$ in $\mathscr{H}$ and any time point $t$ in $\mathbb{R}_{\geq 0}$ such that $t \geq \max u$,*

$$\limsup_{r \searrow t} \frac{P(X_t \neq X_r \mid X_u = x_u)}{r - t} \leq \frac{\lambda}{2}$$

*and, if $t > \max u$,*

$$\limsup_{s \nearrow t} \frac{P(X_s \neq X_t \mid X_u = x_u)}{t - s} \leq \frac{\lambda}{2}.$$

*Proof.* We need to distinguish two cases, and we start with the simplest case that $t \succ u$ – meaning that either $u \neq ()$ and $t > \max u$ or $u = ()$. Fix some time point $r$ in $\mathbb{R}_{\geq 0}$ such that $t < r$. Because $Q_{t,r}^{\{X_u = x_u\}}$ is a rate operator, it follows from (R5)$_{81}$ that

$$\|Q_{t,r}^{\{X_u=x_u\}}\|_{\mathrm{op}} = 2\max\{-Q_{t,r}^{\{X_u=x_u\}}(x,x) \colon x \in \mathscr{X}\}.$$

Observe now that for all $x$ in $\mathscr{X}$,

$$\begin{aligned} Q_{t,r}^{\{X_u=x_u\}}(x,x) &= \frac{1}{r-t}\big(T_{t,r}^{\{X_u=x_u\}}(x,x) - 1\big) = \frac{1}{r-t}\big(P(X_r = x \,|\, X_u = x_u, X_t = x) - 1\big) \\ &= -\frac{1}{r-t} P(X_r \neq x \,|\, X_u = x_u, X_t = x), \end{aligned}$$

where we let $\{X_r \neq x\} := \{X_r = x\}^{\mathrm{c}}$, and where we have used Eq. (3.36)$_{84}$ and (CP7)$_{42}$. Combining the two preceding equalities, we conclude that

$$\|Q_{t,r}^{\{X_u=x_u\}}\|_{\mathrm{op}} = \frac{2}{r-t}\max\{P(X_r \neq x \,|\, X_u = x_u, X_t = x) \colon x \in \mathscr{X}\}. \tag{5.30}$$

Because $(\{X_t = x\})_{x\in\mathscr{X}}$ is a partition of $\Omega$, it follows from (CP3)$_{41}$ and (CP4)$_{41}$ that

$$\begin{aligned} P(X_t \neq X_r \,|\, X_u = x_u) &= P\left(\left(\bigcup_{x\in\mathscr{X}} \{X_t = x\}\right) \cap \{X_t \neq X_r\} \,\middle|\, X_u = x_u\right) \\ &= P\left(\bigcup_{x\in\mathscr{X}} \{X_t = x\} \cap \{X_t \neq X_r\} \,\middle|\, X_u = x_u\right) \\ &= \sum_{x\in\mathscr{X}} P\big(\{X_t = x\} \cap \{X_t \neq X_r\} \,\big|\, X_u = x_u\big) \\ &= \sum_{x\in\mathscr{X}} P\big(X_t \neq X_r \,\big|\, X_u = x_u, X_t = x\big) P(X_t = x \,|\, X_u = x_u). \end{aligned}$$

From the final equality, it follows that

$$\begin{aligned} &P(X_t \neq X_r \,|\, X_u = x_u) \\ &\quad = \sum_{x\in\mathscr{X}} P\big(X_r \neq x \,\big|\, X_u = x_u, X_t = x\big) P(X_t = x \,|\, X_u = x_u) \\ &\quad \leq \max\Big\{P\big(X_r \neq x \,\big|\, X_u = x_u, X_t = x\big) \colon x \in \mathscr{X}\Big\} \sum_{x\in\mathscr{X}} P(X_t = x \,|\, X_u = x_u) \\ &\quad = \max\Big\{P\big(X_r \neq x \,\big|\, X_u = x_u, X_t = x\big) \colon x \in \mathscr{X}\Big\}, \end{aligned}$$

where for the first equality we used (JP1)$_{69}$, the inequality holds due to (CP2)$_{41}$ and the last equality holds due to (CP3)$_{41}$ and (CP1)$_{41}$. From this and Eq. (5.30), it follows that

$$P(X_t \neq X_r \,|\, X_u = x_u) \leq \frac{r-t}{2}\|Q_{t,r}^{\{X_u=x_u\}}\|_{\mathrm{op}}.$$

Because $P$ has rate bound $\lambda$, it follows from the previous inequality that

$$\limsup_{r\searrow t} \frac{P(X_t \neq X_r \,|\, X_u = x_u)}{r-t} \leq \limsup_{r\searrow t} \frac{\|Q_{t,r}^{\{X_u=x_u\}}\|_{\mathrm{op}}}{2} \leq \frac{\lambda}{2},$$

which is the first inequality of the statement. An analoguous argument proves the second inequality of the statement.

We can use the previous to verify the statement for the remaining case that $u \neq ()$ and $t = \max u$. To this end, we let $u' := u \setminus (t)$. Then for any $r > t$,

$$\begin{aligned} P(X_t \neq X_r \,|\, X_u = x_u) &= P(X_t \neq X_r \,|\, X_{u'} = x_{u'}, X_t = x_t) \\ &= P(\{X_t = x_t\} \cap \{X_t \neq X_r\} \,|\, X_{u'} = x_{u'}, X_t = x_t) \\ &= P(X_r \neq x_t \,|\, X_{u'} = x_{u'}, X_t = x_t) \\ &\leq \frac{r-t}{2} \big\| Q_{t,r}^{\{X_{u'}=x_{u'}\}} \big\|_{\mathrm{op}}, \end{aligned}$$

where the second equality follows from (CP1)$_{41}$ and (CP9)$_{42}$ – because, due to the former, $P(X_t = x_t \,|\, X_{u'} = x_{u'}, X_t = x_t) = 1$ – and the inequality follows from Eq. (5.30)↶. Again, the inequality of the statement follows immediately from the preceding inequality and the fact that $\lambda$ is a rate bound. ☐

*Proof of Lemma 5.28$_{237}$.* We use the same standard trick as in the proof by Krak et al. (2017, Lemma F.1). Because $P$ has uniformly bounded rate with bound $\lambda$, we know from Lemma 5.35$_{242}$ that for every $t$ in $[s,r]$ there is a $\delta_t$ in $\mathbb{R}_{>0}$ such that

$$\big(\forall r_t \in \,]t, t+\delta_t[\big)\; P(X_t \neq X_{r_t} \,|\, X_u = x_u) < (r_t - t)\left(\frac{\lambda}{2} + \epsilon\right)$$

and, if $t > s$,

$$\big(\forall s_t \in \,]t-\delta_t, t[\big)\; P(X_{s_t} \neq X_t \,|\, X_u = x_u) < (t - s_t)\left(\frac{\lambda}{2} + \epsilon\right)$$

If $r - \delta_r < s + \delta_s$, then the statement holds trivially for the grid $v := (s,t,r)$ with $t$ any time point in $]r-\delta_r, s+\delta_s[\, \cap \,]s,r[$.

Therefore, without loss of generality we assume that $s + \delta_s \leq r - \delta_r$. Fix some $s'$ in $]s, s+\delta_s[$ and $r'$ in $]r-\delta_r, r[$, and observe that $[s', r']$ is a bounded interval, and therefore a compact set. For this reason, it follows from the Heine-Borel theorem that the open cover

$$\big(]t-\delta_t, t+\delta_t[\big)_{t\in[s',r']}$$

of $[s',r']$ has a finite subcover

$$\big(]t-\delta_t, t+\delta_t[\big)_{t\in\mathscr{C}},$$

where $\mathscr{C}$ is a finite subset of $[s',r']$. Without loss of generality, we may assume that this cover is minimal, in the sense that by removing an element we do not have a subcover any more.

We order the time points in $\mathscr{C}$ from small to large, denote this order by $t'_1, \dots, t'_m$ and set $\delta'_k := \delta_{t'_k}$ for all $k$ in $\{1,\dots,m\}$. It is not difficult but slightly cumbersome to verify (see Krak et al., 2017, Eqn. (F.4)) that by construction, $t'_1 - \delta'_1 < s'$, $r' < t'_m + \delta'_m$ and

$$t'_{k+1} - \delta'_{k+1} < t'_k + \delta'_k \quad \text{for all } k \in \{1,\dots,m-1\}.$$

For this reason, we can define a grid $v = (t_0, \dots, t_n)$ over $[s,r]$, with $n := 2m+2$, as follows. Clearly, we need to set $t_0 := s$ and $t_n = t_{2m+2} := r$. Next, we choose a $t_1$

in $]s, s'[ \cap ]t'_1 - \delta'_1, t'_1[$ and a $t_{n-1} = t_{2m+1}$ in $]t'_m, t'_m + \delta'_m[ \cap ]r', r[$ and let $t_{n-2} = t_{2m} := t'_m$ – note that this is always possible because $t'_1 - \delta'_1 < s' \leq t'_1$ and $t'_m \leq r' < t'_m + \delta'_m$. Finally, for any $k$ in $\{1, \ldots, m-1\}$, we let $t_{2k} := t'_k$ and choose some $t_{2k+1}$ in

$$]t'_k, t'_k + \delta'_k[ \cap ]t'_{k+1} - \delta'_{k+1}, t'_{k+1}[.$$

This way, we have constructed a grid $v = (t_0, \ldots, t_n)$ over $[s, r]$ such that

$$(\forall k \in \{1, \ldots, n\})\ P(X_{t_{k-1}} \neq X_{t_k} \,|\, X_u = x_u) < (t_k - t_{k-1})\left(\frac{\lambda}{2} + \epsilon\right). \qquad \square$$

## 5.B Measure-theoretic jump processes

Stochastic processes in general and jump processes in particular have received a lot of attention in the measure-theoretic framework. A good starting point is the influential work of Doob (1953), but more recent introductions can be found in (Todorovic, 1992, Chapter 1), (Billingsley, 1995, Chapter 7), (Fristedt et al., 1997, Chapter 31) or (Borovkov, 2013, Chapter 18). In this appendix, we summarise the relevant concepts and results regarding measure-theoretic jump processes; for a primer on measure-theoretic probability in general, see Appendix $C_{461}$.

### 5.B.1 The general framework

Recall from Section $2.1.3_{13}$ that a variable is any function on a possibility space $\mathcal{X}$. In measure-theoretical probability theory, a stochastic process with state space $\mathscr{X}$ is simply a family $(Y_t)_{t\in\mathbb{T}}$ of $\mathscr{X}$-valued variables – that is, a family of maps from the possibility space $\mathcal{X}$ to some state space $\mathscr{X}$ – that is indexed by an ordered set of time points $\mathbb{T}$ and satisfies some measurability condition with respect to a field $\mathcal{F}$ over $\mathcal{X}$. In the setting of jump processes, the state space $\mathscr{X}$ is finite and the set of time points $\mathbb{T}$ is the set of non-negative real numbers, so the following measurability condition suffices.

**Definition 5.36.** Consider a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. An *$\mathcal{F}$-measurable jump process (with state space $\mathscr{X}$)* is a family $(Y_t)_{t\in\mathbb{R}_{\geq 0}}$ of $\mathscr{X}$-valued variables such that for all time points $t$ in $\mathbb{R}_{\geq 0}$ and all states $x$ in $\mathscr{X}$,

$$\{Y_t = x\} := \big\{\omega \in \mathcal{X} \colon Y_t(\omega) = x\big\} \in \mathcal{F}. \tag{5.31}$$

Note that in this definition, we have used $\omega$ instead of $x$ to denote a generic outcome in the abstract possibility space $\mathcal{X}$ in order to avoid confusion with states; we adhere to this ‘new’ convention throughout the remainder of this section. It is also important to note that in contrast to Definition $3.12_{65}$, (conditional) probabilities do not play a role in this definition.

Because $\mathcal{F}$ is a field, we immediately obtain that more general events concerning the state of the system at a finite number of time points $t_1, \ldots,$

$t_n$ belong to $\mathcal{F}$ as well. Following the notational conventions of Chapter 353, for any such sequence of time points $u$ in $\mathscr{U}$ and any state instantiation $x_u$ in $\mathscr{X}_u$, we define

$$\{Y_u = x_u\} := \big\{\omega \in \mathcal{X} \colon (\forall t \in u)\ Y_t(\omega) = x_t\big\} = \bigcap_{t \in u} \{Y_t = x_t\}. \tag{5.32}$$

Because $\{Y_u = x_u\}$ is a finite intersection of events in the field $\mathcal{F}$, it also belongs to the field $\mathcal{F}$. Similarly, for any subset $B$ of $\mathscr{X}_u$,

$$\{Y_u \in B\} := \bigcup_{x_u \in B} \{Y_u = x_u\} \tag{5.33}$$

also belongs to the field $\mathcal{F}$ because it is a finite union of events in this field.

**Equivalence and modifications**

Consider two $\mathcal{F}$-measurable jump processes $(Y_t)_{t\in\mathbb{R}_{\geq 0}}$ and $(Z_t)_{t\in\mathbb{R}_{\geq 0}}$. We now ask ourselves when these two processes are, in some sense, equivalent. Clearly, events of the form

$$\{Y_t = Z_t\} := \big\{\omega \in \mathcal{X} \colon Y_t(\omega) = Z_t(\omega)\big\}$$

will play an essential role. Note that

$$\{Y_t = Z_t\} = \bigcup_{x \in \mathscr{X}} \{Y_t = x\} \cap \{Z_t = x\},$$

so these events always belong to the field $\mathcal{F}$ because they consist of finite unions of finite intersections of events in $\mathcal{F}$.

**Definition 5.37.** Consider a probability charge $P$ on a field of events $\mathcal{F}$ over a possibility space $\mathcal{X}$, and two $\mathcal{F}$-measurable jump processes $(Y_t)_{t\in\mathbb{R}_{\geq 0}}$ and $(Z_t)_{t\in\mathbb{R}_{\geq 0}}$ with state space $\mathscr{X}$. Then $(Y_t)_{t\in\mathbb{R}_{\geq 0}}$ and $(Z_t)_{t\in\mathbb{R}_{\geq 0}}$ are *equivalent with respect to P* if

$$P(Y_t = Z_t) = 1 \quad \text{for all } t \in \mathbb{R}_{\geq 0}.$$

Whenever this is the case, we also say that $(Y_t)_{t\in\mathbb{R}_{\geq 0}}$ is a *modification of* $(Z_t)_{t\in\mathbb{R}_{\geq 0}}$ *with respect to P*, and vice versa.

Crucially, whenever the two $\mathcal{F}$-measurable jump processes $(Y_t)_{t\in\mathbb{R}_{\geq 0}}$ and $(Z_t)_{t\in\mathbb{R}_{\geq 0}}$ are equivalent with respect to $P$, they have the same probabilities for the finitary events.

**Lemma 5.38.** *Consider a probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$, and two $\mathcal{F}$-measurable jump processes $(Y_t)_{t\in\mathbb{R}_{\geq 0}}$ and $(Z_t)_{t\in\mathbb{R}_{\geq 0}}$. If $(Y_t)_{t\in\mathbb{R}_{\geq 0}}$ and $(Z_t)_{t\in\mathbb{R}_{\geq 0}}$ are equivalent with respect to $P$, then*

$$P(Y_u \in B) = P(Z_u \in B) \quad \textit{for all } u \in \mathscr{U}, B \subseteq \mathscr{X}_u.$$

*Proof.* If $u$ is the empty time sequence $()$, then the statement holds trivially; for this reason, we assume without loss of generality that $u$ is non-empty. Let

$$\{Y_u = Z_u\} := \bigcap_{t \in u} \{Y_t = Z_t\}.$$

Observe now that because $\{Y_t\}_{t \in \mathbb{R}_{\geq 0}}$ and $\{Z_t\}_{t \in \mathbb{R}_{\geq 0}}$ are equivalent, $P(Y_t = Z_t) = 1$ for all $t$ in $u$. It follows from this and repeated application of (PM8)$_{461}$ that $P(Y_u = Z_u) = 1$. Observe now that

$$P(Y_u \in B) = P(\{Y_u \in B\} \cap \{Y_u = Z_u\}) = P(\{Y_u = Z_u\} \cap \{Z_u \in B\}) = P(Z_u \in B),$$

where for the first and last equality we used (PM8)$_{461}$ again. □

### Sample paths

Let $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$ be some $\mathcal{F}$-measurable jump process with state space $\mathscr{X}$. For any outcome $\omega$ in the possibility space $\mathcal{X}$, the function

$$Y_\bullet(\omega) \colon \mathbb{R}_{\geq 0} \to \mathscr{X} \colon t \mapsto Y_t(\omega)$$

is called a *sample path*. Important to realise is that the set of sample paths

$$\left\{Y_\bullet(\omega) \colon \omega \in \mathcal{X}\right\}$$

is not necessarily equal to the possibility space $\mathcal{X}$. Because every sample path $Y_\bullet(s)$ is a path – that is, a map from $\mathbb{R}_{\geq 0}$ to $\mathscr{X}$ – the set of sample paths is included in – but not necessarily equal to – the set of all paths $\tilde{\Omega}$. Furthermore, a sample path is *not* necessarily càdlàg.

## 5.B.2 Constructing a probability measure for a jump process

In order to define a measure-theoretic jump process, we need to define (i) a possibility space $\mathcal{X}$, (ii) a field of events $\mathcal{F}$ over $\mathcal{X}$, (iii) a family $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$ of $\mathscr{X}$-valued variables that forms an $\mathcal{F}$-measurable jump process, and (iv) a probability charge on $\mathcal{F}$. Note that we have more or less already done this in Section 3.1$_{55}$, albeit we considered there a structure of fields instead of a field of events and a coherent conditional probability instead of a probability charge. In fact, the conventional measure-theoretical approach is largely analoguous to the approach that we followed in Section 3.1$_{55}$, although there is one important difference. In the measure-theoretic approach, it is customary to use as a possibility space the set $\tilde{\Omega}$ of *all* paths instead of the set $\Omega$ of all càdlàg paths; the reason for this difference will become clear in Corollary 5.41$_{249}$ further on.

For all $t$ in $\mathbb{R}_{\geq 0}$, we let $\tilde{X}_t$ be the $\mathscr{X}$-valued variable on $\tilde{\Omega}$ that assumes the value of the path $\omega$ in $\tilde{\Omega}$ at time $t$:

$$\tilde{X}_t \colon \tilde{\Omega} \to \mathscr{X} \colon \omega \mapsto \tilde{X}_t(\omega) := \omega(t).$$

This way, we extend the use of the tilde to distinguish the set $\tilde{\Omega}$ of all paths from the set $\Omega$ of càdlàg paths to the corresponding projector variables: $\tilde{X}_t$ corresponds to $\tilde{\Omega}$ just like $X_t$ – as defined by Eq. (3.3)59 – corresponds to $\Omega$. Because by definition,

$$\tilde{X}_t(\omega) = \omega(t) \quad \text{for all } \omega \in \tilde{\Omega} \text{ and } t \in \mathbb{R}_{\geq 0},$$

the set of sample paths corresponding to $(\tilde{X}_t)_{t \in \mathbb{R}_{\geq 0}}$ is equal to the set of paths:

$$\{\tilde{X}_\bullet(\omega) \colon \omega \in \tilde{\Omega}\} = \tilde{\Omega}.$$

### Finitary events

Note that the finitary events corresponding to $(\tilde{X}_t)_{t \in \mathbb{R}_{\geq 0}}$ are of the form

$$\{\tilde{X}_u \in B\} = \left\{\omega \in \tilde{\Omega} \colon \omega|_u \in B\right\},$$

where $u$ is a sequence of time points in $\mathscr{U}$ and $B$ is a subset of $\mathscr{X}_u$. The corresponding set of finitary events is

$$\tilde{\mathscr{F}} := \left\{\{\tilde{X}_u \in B\} \colon u \in \mathscr{U}, B \subseteq \mathscr{X}_u\right\},$$

and, for all time sequences $u$ in $\mathscr{U}$, the set of finitary events with time points in or succeeding $u$ is

$$\tilde{\mathscr{F}}_u := \left\{\{\tilde{X}_v \in B\} \colon v \in \mathscr{U}_{\succcurlyeq u}, B \subseteq \mathscr{X}_v\right\},$$

with $\tilde{\mathscr{F}}_{()} = \tilde{\mathscr{F}}$. As above, we use a tilde to distinguish the sets of finitary events for all paths from the sets of finitary events for the càdlàg paths; in other words, $\tilde{\mathscr{F}}$ and $\tilde{\mathscr{F}}_u$ are to $\tilde{\Omega}$ what $\mathscr{F}$ and $\mathscr{F}_u$ are to $\Omega$.

Using entirely similar reasoning as in the proof of Lemma 3.1063, we can show that for any sequence of time points $u$, the corresponding set of finitary events $\tilde{\mathscr{F}}_u$ is a field.

**Lemma 5.39.** *For any sequence of time points $u$ in $\mathscr{U}$, the set of finitary events $\tilde{\mathscr{F}}_u$ is a field over the set $\tilde{\Omega}$ of all paths.*

Up to now, we have defined a possibility space $\tilde{\Omega}$, a field of events $\tilde{\mathscr{F}}_u$ and a family $(\tilde{X}_t)_{t \in \mathbb{R}_{\geq 0}}$ of $\mathscr{X}$-valued variables. Observe that by construction of $\tilde{\mathscr{F}}$, the event $\{\tilde{X}_t = x\}$ belongs to $\tilde{\mathscr{F}}$ for all $t$ in $\mathbb{R}_{\geq 0}$ and $x$ in $\mathscr{X}$. Consequently, the family $(\tilde{X}_t)_{t \in \mathbb{R}_{\geq 0}}$ of $\mathscr{X}$-valued variables is an $\tilde{\mathscr{F}}$-measurable jump process. In short, the mathematical foundations for our uncertainty model are in place, and it is time to look at the part of the model that quantifies the uncertainty regarding the jump process.

### Fixing the finite dimensional distribution

In the measure-theoretic approach, one's uncertainty is modelled by a probability charge $\tilde{P}$ on the field of finitary events $\tilde{\mathscr{F}}$. In other words, we fix the probability $\tilde{P}(\tilde{X}_u \in B)$ of every finitary event of the form $\{\tilde{X}_u \in B\}$; in measure-theoretic parlance, we fix the *finite-dimensional distributions* of the process. As before, we use $\tilde{P}$ to emphasise that we consider *all* paths and not only the càdlàg paths.

It is at this point that it becomes essential that we have chosen the set of *all* paths $\tilde{\Omega}$ as a possibility space, because any probability charge $\tilde{P}$ on $\tilde{\mathscr{F}}$ is countably additive. This well-known result is a special case of Theorem 5.44$_{252}$ further on.

**Proposition 5.40.** *Any probability charge $\tilde{P}$ on the field of finitary events $\tilde{\mathscr{F}}$ over the set of all paths $\tilde{\Omega}$ is countably additive.*

*Proof.* Follows immediately from Theorem 5.44$_{252}$ with $u = ()$. □

### Extension to a probability measure

Due to Proposition 5.40, we can always invoke Theorem C.10$_{466}$, Carathéodory's Theorem, to end up with a probability measure $\tilde{P}_\sigma$ on $\sigma(\tilde{\mathscr{F}})$.

**Corollary 5.41.** *For every probability charge $\tilde{P}$ on the field of finitary events $\tilde{\mathscr{F}}$ over the set of all paths $\tilde{\Omega}$, there is a unique probability measure $\tilde{P}_\sigma$ on the generated σ-field $\sigma(\tilde{\mathscr{F}})$ that extends it.*

By Corollary 5.41, any probability charge $\tilde{P}$ on $\tilde{\mathscr{F}}$ has a unique probability measure $\tilde{P}_\sigma$ on $\sigma(\tilde{\mathscr{F}})$ that extends it. For this reason, we can extend the expectation $E_{\tilde{P}}$ of the $\mathcal{F}$-simple variables with respect to $\tilde{P}$ to the expectation $E^{\mathrm{L}}_{\tilde{P}_\sigma}$ on the set $\mathbb{D}^{\mathrm{L}}_{\tilde{P}_\sigma}$ of all L-integrable $\sigma(\mathscr{F})$-measurable variables.

Unfortunately, the generated σ-field $\sigma(\tilde{\mathscr{F}})$ is not very rich because – generally speaking – it only contains events that depend on a countable number of time points, and many inferences depend on the state of the path at an *uncountable* number of time points. In fact, it can be shown – see for instance (Billingsley, 1995, Theorem 36.3) – that for any event $A$ in the generated σ-field $\sigma(\tilde{\mathscr{F}})$, there is a countable set $S_A \subset \mathbb{R}_{\geq 0}$ of time points such that if $\omega$ is a path in $A$, any path $\omega'$ in $\tilde{\Omega}$ that agrees with $\omega$ on $S_A$ – meaning that $\omega(t) = \omega'(t)$ for all $t$ in $S_A$ – belongs to $A$ as well.

For this reason, we are not guaranteed that inferences that depend on the state at all time points – like temporal averages, hitting times or hitting probabilities – are $\sigma(\tilde{\mathscr{F}})$-measurable. Take, for example, the hitting time $\tilde{\tau}^A$ of a subset $A$ of $\mathscr{X}$, defined by

$$\tilde{\tau}^A(\omega) := \inf\{t \in \mathbb{R}_{\geq 0} : \omega(t) \in A\} \quad \text{for all } \omega \in \tilde{\Omega}.$$

Recall from Definition C.12[467] that for $\tilde{\tau}^A$ to be $\sigma(\tilde{\mathscr{F}})$-measurable, we need that the level set $\{\tilde{\tau}^A > \alpha\}$ belongs to $\sigma(\tilde{\mathscr{F}})$ for all $\alpha$ in $\mathbb{R}$. It is easy to see that for a given non-negative real number $\alpha$, this level set depends on all time points $t$ that precede $\alpha$:

$$\{\tilde{\tau}^A > \alpha\} = \big\{\omega \in \tilde{\Omega} \colon \big(\forall t \in [0,\alpha]\big)\, \omega(t) \notin A\big\} = \bigcap_{t\in[0,\alpha]} \{\tilde{X}_t \in A^{\mathrm{c}}\}.$$

That is, we can write the level set $\{\tilde{\tau}^A > \alpha\}$ as an uncountable intersection of finitary events in the field $\tilde{\mathscr{F}}$. However, this does not imply that the level set $\{\tilde{\tau}^A > \alpha\}$ belongs to $\sigma(\tilde{\mathscr{F}})$, because we are only guaranteed that the generated σ-field $\sigma(\tilde{\mathscr{F}})$ includes countable intersections of events in $\tilde{\mathscr{F}}$. In fact, we can easily show that $\{\tilde{\tau}^A > \alpha\}$ does not belong to $\sigma(\tilde{\mathscr{F}})$.

Assume *ex absurdo* that the level set $B := \{\tilde{\tau}^A > \alpha\}$ belongs to $\sigma(\tilde{\mathscr{F}})$ – to avoid any edge cases, we also assume that $\alpha$ is positive and that $A$ is non-empty. Then from the preceding – that is, from (Billingsley, 1995, Theorem 36.3) – we know that there is some countable subset $S_B$ of $\mathbb{R}_{\geq 0}$ such that for any $\omega$ in $B$, any $\omega'$ in $\tilde{\Omega}$ that coincides with $\omega$ on $S_B$ must also belong to $B$. Because $S_B$ is countable, we can fix some time point $s$ in $[0,\alpha] \setminus S_B$. To obtain a contradiction, we simply change the path $\omega$ at the point $s$: we let $\omega'$ be the path that is defined for all $t$ in $\mathbb{R}_{\geq 0}$ by $\omega'(t) := \omega(t)$ if $t \neq s$ and $\omega'(t) := x$ if $t = s$, where $x$ is an arbitrary element of $A$. Note that, by construction, $\omega'$ coincides with $\omega$ on $S_B$, so $\omega'$ should belong to $B$. However, we also see that $\tilde{\tau}^A(\omega') = s \leq \alpha$ by construction, so $\omega'$ *does not* belong to $B = \{\tilde{\tau}^A > \alpha\}$, a clear contradiction.

Hence, we have shown that for positive real numbers $\alpha$, the corresponding level sets $\{\tilde{\tau}^A > \alpha\}$ of $\tilde{\tau}^A$ do not belong to $\sigma(\tilde{\mathscr{F}})$. Therefore, the hitting time $\tilde{\tau}$ is *not* $\sigma(\tilde{\mathscr{F}})$-measurable.

#### Càdlàg modification through separation and continuity

To ensure that the level sets of inferences that depend on the state of the system at an uncountable number of time points do belong to the generated σ-field, it would be helpful if the sample paths of our $\mathcal{F}$-measurable jump process were determined by the states that they assume on some countable subset $D$ of $\mathbb{R}_{\geq 0}$. One way to achieve this, is to require that the jump process should have càdlàg sample paths, provided $D$ is dense in $\mathbb{R}_{\geq 0}$.

**Lemma 5.42.** *Consider a countable dense subset $D$ of $\mathbb{R}_{\geq 0}$, and two càdlàg paths $\omega_1$ and $\omega_2$ in $\Omega$. If $\omega_1(t) = \omega_2(t)$ for all $t$ in $D$, then $\omega_1 = \omega_2$.*

*Proof.* To prove the statement, we fix any arbitrary time point $t$ in $\mathbb{R}_{\geq 0}$ and prove that $\omega_1(t) = \omega_2(t)$. We may assume without loss of generality that $t$ does not belong to $D$. Because $\omega_1$ and $\omega_2$ are both càdlàg, there is a strictly positive real number $\delta$ such that

$$\big(\forall r \in \,]t, t+\delta[\big)\, \omega_1(r) = \omega_1(t) \qquad \text{and} \qquad \big(\forall r \in \,]t, t+\delta[\big)\, \omega_2(r) = \omega_2(t).$$

Because $D$ is dense in $\mathbb{R}_{\geq 0}$, we can choose an $r$ in $D \cap ]t, t+\delta[$. Due to the previous, we are guaranteed that $\omega_1(t) = \omega_1(r)$ and $\omega_2(t) = \omega_2(r)$. Because $\omega_1(r) = \omega_2(r)$ by the assumption of the statement, we can infer that $\omega_1(t) = \omega_2(t)$, as required. □

Due to Lemma 5.42↶, it is customary to define a modification $(\tilde{Y}_t)_{t \in \mathbb{R}_{\geq 0}}$ of $(\tilde{X}_t)_{t \in \mathbb{R}_{\geq 0}}$ that has càdlàg sample paths. Pivotal in obtaining this càdlàg modification are the concepts of 'separability' and 'stochastic continuity'. In essence, an $\mathbb{R}$-valued stochastic process $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$ is 'separable' if its sample paths are fully determined by the states they assume on some countable dense subset; this crucial property was first introduced by Doob (1953), see also (Todorovic, 1992, Section 1.7), (Billingsley, 1995, Section 38) or (Borovkov, 2013, Definition 18.2.1). Additionally, $(Y_t)_{t \in \mathbb{R}_{\geq 0}}$ is 'stochastically continuous' if for all $t$ in $\mathbb{R}_{\geq 0}$ and $\epsilon$ in $\mathbb{R}_{>0}$, the probability of the event $\{|Y_t - Y_r| > \epsilon\}$ goes to zero as $r$ approaches $t$ – see (Todorovic, 1992, Section 1.10) or (Borovkov, 2013, Definition 18.2.2).

Because our state space $\mathscr{X}$ is finite, we follow a slightly different path, but the underlying ideas remain the same. After quite a bit of work – we will discover the details in Sections 5.C.2$_{255}$ and 5.C.3$_{259}$ further on – we obtain the following result.

**Proposition 5.43.** *Consider a probability measure $\tilde{P}_\sigma$ on the σ-field $\sigma(\tilde{\mathscr{F}})$ generated by the finitary events. If there is a non-negative real number $\lambda$ such that*

$$\limsup_{r \to t} \frac{\tilde{P}_\sigma(\tilde{X}_t \neq \tilde{X}_r)}{|r-t|} \leq \lambda \quad \textit{for all } t \in \mathbb{R}_{\geq 0},$$

*then there is a $\sigma(\tilde{\mathscr{F}})$-measurable jump process $(\tilde{Y}_t)_{t \in \mathbb{R}_{\geq 0}}$ with càdlàg sample paths that is a modification of $(\tilde{X}_t)_{t \in \mathbb{R}_{\geq 0}}$ with respect to $\tilde{P}_\sigma$.*

*Proof.* This is an obvious special case of Theorem 5.52$_{261}$ further on. □

It is now customary to use the càdlàg modification $(\tilde{Y}_t)_{t \in \mathbb{R}_{\geq 0}}$ of $(\tilde{X}_t)_{t \in \mathbb{R}_{\geq 0}}$ with respect to $\tilde{P}_\sigma$ as if it were the original jump process $(\tilde{X}_t)_{t \in \mathbb{R}_{\geq 0}}$, because the càdlàg property ensures that all relevant inferences are $\sigma(\tilde{\mathscr{F}})$-measurable. Take for example the hitting time of the subset $A$ of $\mathscr{X}$, which is then defined as

$$\tilde{\tau}^A_{(\tilde{Y}_t)_{t \in \mathbb{R}_{\geq 0}}}(\omega) := \inf\{t \in \mathbb{R}_{\geq 0} \colon \tilde{Y}_t(\omega) \in A\} \quad \text{for all } \omega \in \tilde{\Omega}.$$

It is not difficult to prove that

$$\{\tilde{\tau}^A_{(\tilde{Y}_t)_{t \in \mathbb{R}_{\geq 0}}} > \alpha\} = \bigcap_{t \in [0,\alpha] \cap D} \{\tilde{Y}_t \in A^{\mathrm{c}}\} \quad \text{for all } \alpha \in \mathbb{R}_{\geq 0},$$

where $D$ is any countable dense subset of $\mathbb{R}_{\geq 0}$. Observe that this level set belongs to $\sigma(\tilde{\mathscr{F}})$ because the events $\{\tilde{Y}_t \in A^{\mathrm{c}}\}$ belong to $\sigma(\tilde{\mathscr{F}})$ by construction and the σ-field $\sigma(\tilde{\mathscr{F}})$ is closed under taking countable intersections due to Lemma C.7$_{465}$ (ii). Thus, the hitting time $\tilde{\tau}^A_{(\tilde{Y}_t)_{t \in \mathbb{R}_{\geq 0}}}$ is $\sigma(\tilde{\mathscr{F}})$-measurable.

Note that by using the modification $(\tilde{Y}_t)_{t\in\mathbb{R}_{\geq 0}}$ as if it were the original process $(\tilde{X}_t)_{t\in\mathbb{R}_{\geq 0}}$, one is essentially changing the outcome of the experiment: any outcome $\omega$ in the possibility space $\tilde{\Omega}$ is modified to the sample path $\tilde{Y}_\bullet(\omega)$. As we have previously mentioned at the end of Section 5.2.1228, advocates of the measure-theoretic approach provide little or no motivation as to why working with this modification makes sense, besides the (implicit) pragmatic justification that 'it works'.

## 5.C Countable additivity of a jump process with uniformly bounded rate

In this appendix, we will use the measure-theoretic framework for jump processes – as summarised in Appendix 5.B245 – to prove the following quintessential result that forms the basis of the approach that was outlined in Section 5.2228. Amusingly, the result itself has nothing to do with the measure-theoretic framework.

**Theorem 5.19.** *If a jump process $P$ has uniformly bounded rate, then it is countably additive.*

Our proof for this result is rather lengthy, which is why we have chosen to break it up into parts that are more easily digestible on their own. More concretely, we fix some jump process $P$ with uniformly bounded rate and some state history $\{X_u = x_u\}$ in $\mathscr{H}$, and set out to prove that the probability charge $P(\bullet \mid X_u = x_u)$ is countably additive. We start our dirty detour in Appendix 5.C.1 by moving from the probability charge $P(\bullet \mid X_u = x_u)$ on the field $\mathscr{F}_u$ of events over the set $\Omega$ of càdlàg paths to a probability measure $\tilde{P}_\sigma^{x_u}$ on the σ-field of events $\sigma(\tilde{\mathscr{F}}_u)$ over the set $\tilde{\Omega}$ of *all* paths. Next, for this probability measure $\tilde{P}_\sigma^{x_u}$, we obtain an upper bound on the expected number of jumps in Appendix 5.C.2255. In Appendix 5.C.3259, we construct a modification $(\tilde{Y}_t)_{t\in\mathbb{R}_{\geq 0}}$ of $(\tilde{X}_t)_{t\in\mathbb{R}_{\geq 0}}$ with càdlàg sample paths, essentially following the classical approach towards constructing jump processes as outlined in Appendix 5.B.2247. Finally, we use this modification $(\tilde{Y}_t)_{t\in\mathbb{R}_{\geq 0}}$ and the relation between the probability measure $\tilde{P}_\sigma^{x_u}$ and the probability charge $P(\bullet \mid X_u = x_u)$ to prove that the latter is countably additive in Appendix 5.C.5267.

### 5.C.1 From càdlàg paths to all paths

Our first step towards proving Theorem 5.19230 is to move from the set of càdlàg paths $\Omega$ to the set of all paths $\tilde{\Omega}$ as the possibility space. Our sole reason for doing so is the following result, which is well-known but almost always stated without a formal proof.

**Theorem 5.44.** *Fix some sequence of time points $u$ in $\mathscr{U}$. Then any probability charge $\tilde{P}$ on $\tilde{\mathscr{F}}_u$ is countably additive.*

Our proof of Theorem 5.44 is a straightforward modification of Billingsley's (1995, Theorem 2.3) argument from discrete time to continuous time, and it hinges on the following important lemma.

**Lemma 5.45.** *Consider a sequence of time points $u$ in $\mathscr{U}$. Then for any non-increasing sequence $(\tilde{A}_n)_{n\in\mathbb{N}}$ of non-empty events in $\tilde{\mathscr{F}}_u$,*

$$\bigcap_{n\in\mathbb{N}} \tilde{A}_n \neq \emptyset.$$

*Proof.* In essence, the statement holds because any countable sequence of events in $\tilde{\mathscr{F}}_u$ only fixes the path on a countable sequence of time points, and one can always construct a path that assumes these given values. To prove the statement formally, we let $\mathcal{N}$ be the sequence of natural numbers that is constructed by starting with the initial index $n_1 := 1$, then adding the smallest natural number $n_2$ such that $\tilde{A}_{n_2} \neq \tilde{A}_{n_1}$ whenever this exists, and so on. Note that, by construction, $(\tilde{A}_n)_{n\in\mathcal{N}}$ is decreasing – in the sense that $\tilde{A}_n \supset \tilde{A}_m$ for all $n$ and $m$ in $\mathcal{N}$ such that $n < m$ – and that $\bigcap_{n\in\mathcal{N}} \tilde{A}_n = \emptyset$. Furthermore, we observe that $\mathcal{N}$ is finite if and only if $\tilde{A}_n = \emptyset$ for some $n$ in $\mathbb{N}$; whenever this is the case, the statement is trivially true. For this reason, we may assume without loss of generality that the sequence $(\tilde{A}_n)_{n\in\mathbb{N}}$ is decreasing, in the sense that $\tilde{A}_n \supset \tilde{A}_{n+1}$ for all $n$ in $\mathbb{N}$.

First, we observe that due to the definition of $\tilde{\mathscr{F}}_u$, for every natural number $n$ there is a sequence of time points $v'_n$ in $\mathscr{U}_{\succcurlyeq u}$ and a subset $B'_n$ of $\mathscr{X}_{v'_n}$ such that $\tilde{A}_n = (\tilde{X}_{v'_n} \in B'_n)$. For any $n$ in $\mathbb{N}$, we let $v_n := \bigcup_{k=1}^{n} v'_k$ be the ordered union of the time points in $v'_1, \dots, v'_n$, and furthermore define

$$B_n := \{x_{v_n} \in \mathscr{X}_{v_n} : x_{v'_n} \in B'_n\}.$$

This way, $\tilde{A}_n = \{\tilde{X}_{v_n} \in B_n\}$ and $v_n \subseteq v_{n+1}$ for all $n$ in $\mathbb{N}$.

Observe that because the sequence $(\tilde{A}_n)_{n\in\mathbb{N}} = (\{\tilde{X}_{v_n} \in B_n\})_{n\in\mathbb{N}}$ of non-empty events is decreasing, either $v_n \subset v_{n+1}$, or $v_n = v_{n+1}$ and $B_n \supset B_{n+1} \neq \emptyset$. Because $B_n$ is a subset of the finite set $\mathscr{X}_{v_n}$, we infer from this that for every natural number $n$, there is a smallest natural number $n' > n$ such that $v_n \subset v_{n'}$.

Let $(t_k)_{k\in\mathbb{N}}$ be the sequence of time points in $\mathbb{R}_{\geq 0}$ that is constructed by starting with the time points in $v_1$, subsequently appending the time points in $v_2 \setminus v_1$, then appending those in $v_3 \setminus v_2$, and so on. By construction, there is a non-decreasing sequence $(k_n)_{n\in\mathbb{N}}$ of natural numbers such that for all $n$ in $\mathbb{N}$, $v_n$ is the ordered version of $(t_1, \dots, t_{k_n})$. Conversely, for every natural number $k$, we let $n_k$ be the smallest natural number $n$ such that $t_k$ belongs to $v_n$ and $v_n \subset v_{n+1}$ – recall from before that this last condition is always satisfied for some $n$. Note that $(n_k)_{k\in\mathbb{N}}$ is a non-decreasing sequence of natural numbers because whenever $t_k$ is in $v_n$ then so is $t_{k-1}$, and that $n_{k_n} \geq n$ for every natural number $n$.

For any $k$ in $\mathbb{N}$, we fix a path $\omega_k$ in $\tilde{A}_{n_k}$ – note that this is possible because $\tilde{A}_{n_k}$ is non-empty by the assumption in the statement. Consider the sequence $(\omega_k(t_1))_{k\in\mathbb{N}}$. Because the state space $\mathscr{X}$ is finite, there is at least one state that occurs infinitely often in this sequence; we choose any such state, and denote it by $x^\star_{t_1}$. Additionally, we let $(\ell_{1,i})_{i\in\mathbb{N}}$ denote the increasing sequence of natural numbers $\ell$ such that $\omega_\ell(t_1) = x^\star_{t_1}$.

Next, we consider the sequence $(\omega_{\ell_{1,i}}(t_2))_{i\in\mathbb{N}}$. Again, there is at least one state that occurs infinitely often in this sequence. We choose any such state, denote it by $x^\star_{t_2}$ and let $(\ell_{2,n})_{n\in\mathbb{N}}$ denote the increasing subsequence of $(\ell_{1,i})_{i\in\mathbb{N}}$ such that $\omega_\ell(t_2) = x^\star_{t_2}$ for all $\ell$ in this sequence. Note that due to the previous, $\omega_\ell(t_1) = x^\star_{t_1}$ for all $\ell$ in the sequence $(\ell_{2,i})_{i\in\mathbb{N}}$.

It is clear that if we continue in the same manner, then for every natural number $k$ – and the corresponding time point $t_k$ – we choose a state $x^\star_{t_k}$ and obtain a subsequence $(\ell_{k,i})_{i\in\mathbb{N}}$ such that $\omega_\ell(t_k) = x^\star_{t_k}$ for all $\ell$ in this sequence, and also $\omega_\ell(t_j) = x^\star_{t_j}$ for all $\ell$ in $(\ell_{k,i})_{i\in\mathbb{N}}$ and $j$ in $\{1,\dots,k-1\}$. Observe that $\ell_k := \ell_{k,k} \geq k$ because we always take increasing subsequences.

Fix any natural number $n$. Observe that $n_{\ell_{k_n}} \geq n_{k_n}$ because $\ell_{k_n} \geq k_n$ and the sequence $(n_k)_{k\in\mathbb{N}}$ is non-decreasing. We have seen before that $n_{k_n} \geq n$, so $n_{\ell_{k_n}} \geq n_{k_n} \geq n$. Because $(\tilde{A}_k)_{k\in\mathbb{N}}$ is decreasing, it follows from these inequalities that

$$\omega_{\ell_{k_n}} \in \tilde{A}_{n_{\ell_{k_n}}} \subseteq \tilde{A}_{n_{k_n}} \subseteq \tilde{A}_n = \{\tilde{X}_{v_n} \in B_n\},$$

where we have also used that $\omega_\ell$ belongs to $\tilde{A}_{n_\ell}$ because of how $\omega_\ell$ was chosen. For this reason,

$$\omega_{\ell_{k_n}}\Big|_{v_n} = \Big(\omega_{\ell_{k_n}}(t)\Big)_{t\in v_n} \in B_n. \tag{5.34}$$

On the other hand, we observe that by construction,

$$\omega_{\ell_{k_n}}(t_k) = x^\star_{t_k} \qquad \text{for all } k \in \{1,\dots,k_n\};$$

because $v_n$ is the ordered version of $(t_1,\dots,t_{k_n})$,

$$\omega_{\ell_{k_n}}\Big|_{v_n} = \Big(\omega_{\ell_{k_n}}(t)\Big)_{t\in v_n} = (x^\star_t)_{t\in v_n} = x^\star_{v_n}. \tag{5.35}$$

Combining Eqs. (5.34) and (5.35), we obtain that

$$x^\star_{v_n} = (x^\star_t)_{t\in v_n} \in B_n. \tag{5.36}$$

Finally, we fix some $x^\star$ in $\mathscr{X}$ and define the path

$$\omega^\star\colon \mathbb{R}_{\geq 0} \to \mathscr{X}\colon t \mapsto \omega^\star(t) := \begin{cases} x^\star_{t_n} & \text{if } t = t_n \text{ for some } n \in \mathbb{N}, \\ x^\star & \text{otherwise.} \end{cases}$$

It is obvious that by construction, $\omega^\star|_{v_n} = x^\star_{v_n}$ for any natural number $n$. From this and Eq. (5.36), we infer that

$$\omega^\star \in \{\tilde{X}_{v_n} \in B_n\} = \tilde{A}_n \qquad \text{for all } n \in \mathbb{N}.$$

Of course, this implies that $\bigcap_{n\in\mathbb{N}} \tilde{A}_n \neq \emptyset$, which is precisely the claim that we set out to prove. □

*Proof of Theorem 5.44*$_{252}$. Recall from Lemma C.3$_{463}$ (CA4) that in order to prove that the probability charge $\tilde{P}$ is countably additive, it suffices to verify that $\lim_{n\to+\infty} \tilde{P}(A_n) = 0$ for any non-increasing sequence $(A_n)_{n\in\mathbb{N}}$ of non-empty events in $\tilde{\mathscr{F}}_u$ such that $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$. Because no such sequence exists due to Lemma 5.45$_\curvearrowleft$, the probability charge $\tilde{P}$ is trivially countably additive. □

### 5.C.2 An upper bound on the expected number of jumps

Our next step consists in bounding the expected number of jumps. Our approach is for the most part similar to that in Section 5.2.2<sub>231</sub>, but there are two notable differences. The first is that we consider a probability measure on $\sigma(\tilde{\mathscr{F}}_u)$ instead of a probability charge on $\mathscr{F}_u$. For this reason, we only deal with the number of jumps along some countable dense set $D$, which is the second difference.

#### Number of jumps along a grid

Fix some time sequence $u$ in $\mathscr{U}$. For any time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$, it is clear that the event

$$\{\tilde{X}_s \neq \tilde{X}_r\} := \{\omega \in \tilde{\Omega} \colon \omega(s) \neq \omega(r)\}$$

belongs to $\tilde{\mathscr{F}}_u$, and therefore also to $\sigma(\tilde{\mathscr{F}}_u)$. Consider some grid $v = (t_0, \ldots, t_n)$ in $\mathscr{U}_{\neq ()}$. Similar to how we defined $\eta_v$, the number of jumps of the càdlàg paths over $v$, we define the number of jumps of all paths over $v$ as

$$\tilde{\eta}_v \colon \tilde{\Omega} \to \mathbb{R}_{\geq 0} \colon \omega \mapsto \tilde{\eta}_v(\omega) := \left|\left\{k \in \{1, \ldots, n\} \colon \omega(t_{k-1}) \neq \omega(t_k)\right\}\right|,$$

where we again follow the notational convention of using a tilde to distinguish functions on $\tilde{\Omega}$ from their counterparts on $\Omega$. Because

$$\tilde{\eta}_v = \sum_{k=1}^{n} \mathbb{I}_{\{\tilde{X}_{t_{k-1}} \neq \tilde{X}_{t_k}\}}, \tag{5.37}$$

we see that $\tilde{\eta}_v$ is an $\tilde{\mathscr{F}}_u$-simple variable – and therefore also a $\sigma(\tilde{\mathscr{F}}_u)$-simple variable. Furthermore, we also see that if $v$ and $w$ are grids such that $w$ refines $v$, then $\tilde{\eta}_w \geq \tilde{\eta}_v$.

#### The number of jumps along a countable dense set

Fix some time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. For any countable dense subset $D$ of $\mathbb{R}_{\geq 0}$, we let $\mathscr{U}^D_{[s,r]}$ be the subset of all grids over $[s,r]$ whose interior grid points belong to $D$:

$$\mathscr{U}^D_{[s,r]} := \{(t_0, \ldots, t_n) \in \mathscr{U}_{[s,r]} \colon v \subseteq D \cup \{s, r\}\}.$$

With this notation, we can define the number of jumps over the time interval $[s,r]$ and along the countable dense subset $D$ of $\mathbb{R}_{\geq 0}$ as

$$\tilde{\eta}^D_{[s,r]} \colon \tilde{\Omega} \to \overline{\mathbb{R}} \colon \omega \mapsto \tilde{\eta}^D_{[s,r]}(\omega) := \sup\{\tilde{\eta}_v(\omega) \colon v \in \mathscr{U}^D_{[s,r]}\}. \tag{5.38}$$

Important for the remainder is that this non-negative extended real variable $\tilde{\eta}^D_{[s,r]}$ is a $\sigma(\tilde{\mathscr{F}}_u)$-over variable, and therefore also $\sigma(\tilde{\mathscr{F}}_u)$-measurable.

**Lemma 5.46.** *Consider a sequence of time points $u$ in $\mathscr{U}$ and time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Let $D$ be a countable dense subset of $\mathbb{R}_{\geq 0}$. Then there is a sequence $(v_n)_{n\in\mathbb{N}}$ of grids in $\mathscr{U}^D_{[s,r]}$ such that $(\tilde{\eta}_{v_n})_{n\in\mathbb{N}}$ is a sequence of $\sigma(\tilde{\mathscr{F}}_u)$-simple variables that is non-decreasing and that converges point-wise to $\tilde{\eta}^D_{[s,r]}$. Consequently, $\tilde{\eta}^D_{[s,r]}$ is $\sigma(\tilde{\mathscr{F}}_u)$-measurable.*

*Proof.* First, we prove that $\tilde{\eta}^D_{[s,r]}$ is an $\tilde{\mathscr{F}}_u$-over variable. Because $D$ is countable, there is a sequence $(v_n)_{n\in\mathbb{N}}$ in $\mathscr{U}^D_{[s,r]}$ such that $v_n \subseteq v_{n+1}$ for all $n$ in $\mathbb{N}$ and $\bigcup_{n\in\mathbb{N}} v_n = \{s,r\} \cup \big([s,r]\cap D\big)$. If $s=r$, the only possible such sequence is given by $v_n = (s)$ for all $n$ in $\mathbb{N}$. If on the other hand $s<r$, there are many such sequences; to give just one example, consider the sequence that is recursively defined by letting $v_0 := (s,r)$ and $v_n := v_{n-1} \cup (t_n)$ for all $n$ in $\mathbb{N}$, where $(t_n)_{n\in\mathbb{N}}$ is any enumeration of the time points in the countable set $]s,r[\cap D$.

Observe that $\tilde{\eta}_{v_{n+1}} \geq \tilde{\eta}_{v_n}$ because $v_{n+1}$ refines $v_n$, so $(\tilde{\eta}_{v_n})_{n\in\mathbb{N}}$ is a non-decreasing sequence of $\tilde{\mathscr{F}}_u$-simple variables. Thus, to prove that $\tilde{\eta}^D_{[s,r]}$ is an $\tilde{\mathscr{F}}_u$-over variable, it remains for us to verify that the monotone limit $\text{p-w}\lim_{n\to+\infty}\tilde{\eta}_{v_n}$ is equal to $\tilde{\eta}^D_{[s,r]}$. To this end, we observe that for all $n$ in $\mathbb{N}$, $\tilde{\eta}_{v_n} \leq \tilde{\eta}^D_{[s,r]}$ due to Eq. (5.38)↶. Consequently,

$$\operatorname*{p-w\,lim}_{n\to+\infty}\tilde{\eta}_{v_n}(\omega) \leq \tilde{\eta}^D_{[s,r]}(\omega).$$

On the other hand, we observe that for any grid $v$ in $\mathscr{U}^D_{[s,r]}$, there is an index $n$ in $\mathbb{N}$ such that $v \subseteq v_n$. Therefore,

$$\tilde{\eta}_v(\omega) \leq \operatorname*{p-w\,lim}_{n\to+\infty}\tilde{\eta}_{v_n}(\omega) \quad \text{for all } v \in \mathscr{U}^D_{[s,r]}.$$

From the two previous inequalities, we infer that

$$\tilde{\eta}^D_{[s,r]}(\omega) = \sup\big\{\tilde{\eta}_v(\omega) \colon v \in \mathscr{U}^D_{[s,r]}\big\} = \operatorname*{p-w\,lim}_{n\to+\infty}\tilde{\eta}_{v_n}(\omega) \quad \text{for all } \omega \in \tilde{\Omega}.$$

To summarise, we have shown that $(\tilde{\eta}_{v_n})_{n\in\mathbb{N}}$ is a non-decreasing sequence of $\tilde{\mathscr{F}}_u$-simple variables that converges point-wise to $\tilde{\eta}^D_{[s,r]}$. Because every $\tilde{\mathscr{F}}_u$-simple variable is trivially $\sigma(\tilde{\mathscr{F}}_u)$-simple as well, we conclude that the non-negative extended real variable $\tilde{\eta}^D_{[s,r]}$ is an $\sigma(\tilde{\mathscr{F}}_u)$-over variable, and therefore also a $\sigma(\tilde{\mathscr{F}}_u)$-measurable variable due to Lemma C.20$_{471}$. □

### An upper bound on the expected number of jumps along a countable dense set

Next, we determine an upper bound on the expected number of jumps along the countable dense set $D$. We start with a version of Definition 5.16$_{230}$ – or, more precisely, Lemma 5.35$_{242}$ – that is adapted to our different setting.

**Definition 5.47.** Consider a sequence of time points $u$. A probability measure $\tilde{P}_\sigma$ on $\sigma(\tilde{\mathscr{F}}_u)$ has *uniformly bounded rate* if there is a non-negative real number $\lambda$ such that for any time point $t$ in $[\max u, +\infty[$,

$$\limsup_{r\searrow t}\frac{\tilde{P}_\sigma(\tilde{X}_t \neq \tilde{X}_r)}{r-t} \leq \frac{\lambda}{2} \quad \text{and, if } t > \max u, \quad \limsup_{s\nearrow t}\frac{\tilde{P}_\sigma(\tilde{X}_s \neq \tilde{X}_t)}{t-s} \leq \frac{\lambda}{2}.$$

Whenever this is the case, we call $\lambda$ a *rate bound*.

Because Lemma 5.35$_{242}$ and Definition 5.47 are so similar, we more or less immediately obtain the following intermediary result that is similar to Lemma 5.28$_{237}$.

**Lemma 5.48.** *Consider a sequence of time points $u$ in $\mathscr{U}$ and a probability measure $\tilde{P}_\sigma$ on $\sigma(\tilde{\mathscr{F}}_u)$ that has uniformly bounded rate, with rate bound $\lambda$. Fix any countable dense subset $D$ of $\mathbb{R}_{\geq 0}$ and any two time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Then for all $\epsilon$ in $\mathbb{R}_{>0}$, there is a grid $v = (t_0, \dots, t_n)$ over $[s, r]$ such that*

$$\big(\forall k \in \{1, \dots, n\}\big)\; \tilde{P}_\sigma(\tilde{X}_{t_{k-1}} \neq \tilde{X}_{t_k}) < (t_k - t_{k-1})\left(\frac{\lambda}{2} + \epsilon\right).$$

*Proof.* The proof is almost entirely the same as that of Lemma 5.28$_{237}$; the only real difference is that we need to invoke Definition 5.47 instead of Lemma 5.35$_{242}$. □

With the help of this intermediary result, we can easily prove a version of Theorem 5.27$_{236}$ that is suited for the present setting.

**Lemma 5.49.** *Consider a sequence of time points $u$ in $\mathscr{U}$, time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$ and a probability measure $\tilde{P}_\sigma$ on $\sigma(\tilde{\mathscr{F}}_u)$ that has uniformly bounded rate, with bound $\lambda$. Then for any grid $v$ over $[s, r]$,*

$$E_{\tilde{P}_\sigma}(\tilde{\eta}_v) \leq (r - s)\frac{\lambda}{2};$$

*consequently, for any countable dense subset $D$ of $\mathbb{R}_{\geq 0}$,*

$$E^{\mathrm{D}}_{\tilde{P}_\sigma}\left(\tilde{\eta}^{D}_{[s,r]}\right) \leq (r - s)\frac{\lambda}{2}.$$

*Proof.* The proof of this result is more or less the same as the proof of Theorem 5.27$_{236}$, except for some minor differences. Here too, the statement clearly trivially holds whenever $s = r$, so we may assume that $s < r$ without loss of generality.

To prove the first part of the statement, we let $v = (t_0, \dots, t_n)$ be any grid over $[s, r]$. Then due to Eq. (5.37)$_{255}$,

$$\tilde{\eta}_v = \sum_{k=1}^{n} \tilde{\eta}_{(t_{k-1}, t_k)}.$$

For all $k$ in $\{1, \dots, n\}$, we now use Lemma 5.48 – and not Lemma 5.28$_{237}$ – to construct a grid $v_k = (t_{k,0}, \dots, t_{k,n_k})$ over $[t_{k-1}, t_k]$ such that

$$\big(\forall \ell \in \{1, \dots, n_k\}\big)\; \tilde{P}_\sigma(\tilde{X}_{t_{k,\ell-1}} \neq \tilde{X}_{t_{k,\ell}}) < (t_{k,\ell} - t_{k,\ell-1})\left(\frac{\lambda}{2} + \epsilon\right).$$

Observe that for any $k$ in $\{1, \dots, n\}$,

$$\tilde{\eta}_{(t_{k-1}, t_k)} \leq \tilde{\eta}_{v_k} = \sum_{\ell=1}^{n_k} \mathbb{I}_{\{\tilde{X}_{t_{k,\ell-1}} \neq \tilde{X}_{t_{k,\ell}}\}}.$$

Thus, as in the proof of Theorem 5.27$_{236}$, it follows from the additivity and monotonicity of $E_{\tilde{P}_\sigma}$ – so (ES3)$_{37}$ and (ES4)$_{37}$ – that

$$\begin{aligned} E_{\tilde{P}_\sigma}(\tilde{\eta}_{\,v}) &= \sum_{k=1}^{n} E_{\tilde{P}_\sigma}(\tilde{\eta}_{\,(t_{k-1},t_k)}) \\ &\leq \sum_{k=1}^{n} E_{\tilde{P}_\sigma}(\tilde{\eta}_{\,v_k}) < \sum_{k=1}^{n}\sum_{\ell=1}^{n_k}(t_{k,\ell}-t_{k,\ell-1})\left(\frac{\lambda}{2}+\epsilon\right) = (r-s)\left(\frac{\lambda}{2}+\epsilon\right). \end{aligned}$$

Because this inequality holds for any positive real number $\epsilon$, we conclude that

$$E_{\tilde{P}_\sigma}(\tilde{\eta}_{\,v}) \leq (r-s)\frac{\lambda}{2},$$

which verifies the first part of the statement.

To prove the second part of the statement, we recall from Lemma 5.46$_{256}$ that there is a sequence $(v_n)_{n\in\mathbb{N}}$ of grids over $[s,r]$ such that $(\tilde{\eta}_{\,v_n})_{n\in\mathbb{N}}$ is a non-decreasing sequence of $\sigma(\tilde{\mathscr{F}}_u)$-simple variables that converges point-wise to $\tilde{\eta}^D_{[s,r]}$. It follows from this and (DE2)$_{225}$ that

$$E^{\mathrm{D}}_{\tilde{P}_\sigma}\left(\tilde{\eta}^D_{[s,r]}\right) = \lim_{n\to+\infty} E_{\tilde{P}_\sigma}(\tilde{\eta}_{\,v_n}).$$

The second part of the statement follows immediately from the preceding equality and the first part of the statement. ☐

Finally, the previous result allows us to prove that for a probability measure with uniformly bounded rate, the probability of having a finite number of jumps along a dense set is 1.

**Proposition 5.50.** *Consider a sequence of time points $u$ in $\mathscr{U}$ and a probability measure $\tilde{P}_\sigma$ on $\sigma(\tilde{\mathscr{F}}_u)$ that has uniformly bounded rate. Then for any countable dense subset $D$ of $\mathbb{R}_{\geq 0}$ and any time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that* $\max u \leq s \leq r$,

$$\left\{\tilde{\eta}^D_{[s,r]} \lessdot +\infty\right\} := \left\{\omega\in\Omega\colon \tilde{\eta}^D_{[s,r]}(\omega) < +\infty\right\}$$

*is an event in $\sigma(\tilde{\mathscr{F}}_u)$, and*

$$\tilde{P}_\sigma\left(\left\{\tilde{\eta}^D_{[s,r]} \lessdot +\infty\right\}\right) = 1.$$

*Proof.* For the sake of notational brevity, we let $\tilde{A} := \{\tilde{\eta}^D_{[s,r]} \lessdot +\infty\}$. For any natural number $n$, we furthermore let

$$\tilde{A}_n := \{\tilde{\eta}^D_{[s,r]} \lessdot n\}.$$

Recall from Lemma 5.46$_{256}$ that $\tilde{\eta}^D_{[s,r]}$ is $\sigma(\tilde{\mathscr{F}}_u)$-measurable, so $\tilde{A}_n$ belongs to $\sigma(\tilde{\mathscr{F}}_u)$ by Definition C.12$_{467}$ (iv). Observe that $(\tilde{A}_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence of events in $\sigma(\tilde{\mathscr{F}}_u)$ with

$$\bigcup_{n\in\mathbb{N}} \tilde{A}_n = \{\tilde{\eta}^D_{[s,r]} \lessdot +\infty\} = \tilde{A}.$$

By Lemma C.7[465] (iii), this implies that $\tilde{A}$ belongs to $\sigma(\tilde{\mathscr{F}}_u)$, which proves the first part of the statement.

To prove the second part of the statement, we recall from (PM6)[461] that

$$\tilde{P}_\sigma(\tilde{A}) = 1 - \tilde{P}_\sigma(\tilde{A}^c). \tag{5.39}$$

Observe that $(\tilde{A}_n^c)_{n\in\mathbb{N}} = (\{\tilde{\eta}_{[s,r]}^D \geq n\})_{n\in\mathbb{N}}$ is a non-increasing sequence of events in $\sigma(\tilde{\mathscr{F}}_u)$ with $\bigcap_{n\in\mathbb{N}} \tilde{A}_n^c = \tilde{A}^c$. For this reason, it follows from Lemma C.3[463] (CA3) that

$$\tilde{P}_\sigma(\tilde{A}^c) = \lim_{n\to+\infty} \tilde{P}_\sigma(\tilde{A}_n^c).$$

Observe that for all $n$ in $\mathbb{N}$, because $\tilde{\eta}_{[s,r]}^D \geq 0$

$$n\mathbb{I}_{\tilde{A}_n^c} = n\mathbb{I}_{\{\tilde{\eta}_{[s,r]}^D \geq n\}} \leq \tilde{\eta}_{[s,r]}^D.$$

It follows from this, Eq. (2.19)[36], (DE1)[225], (DE6)[226] and (DE4)[225] – essentially Markov's inequality – that for all $n$ in $\mathbb{N}$,

$$\tilde{P}_\sigma(\tilde{A}_n^c) = \tilde{P}_\sigma\big(\{\tilde{\eta}_{[s,r]}^D \geq n\}\big) = E_{\tilde{P}_\sigma}^{\mathrm{D}}\big(\mathbb{I}_{\{\tilde{\eta}_{[s,r]}^D \geq n\}}\big) \leq E_{\tilde{P}_\sigma}^{\mathrm{D}}\Big(\frac{1}{n}\tilde{\eta}_{[s,r]}^D\Big) = \frac{1}{n}E_{\tilde{P}_\sigma}^{\mathrm{D}}\big(\tilde{\eta}_{[s,r]}^D\big).$$

Because $\tilde{P}_\sigma$ has uniformly bounded rate with bound $\lambda$, it follows from Lemma 5.49[257] and the previous inequality that

$$\tilde{P}_\sigma(\tilde{A}_n^c) \leq \frac{1}{n}E_{\tilde{P}_\sigma}^{\mathrm{D}}(\tilde{\eta}_{[s,r]}^D) \leq \frac{(r-s)\lambda}{2n} \quad \text{for all } n \in \mathbb{N}.$$

Additionally, we recall from (P2)[34] that $\tilde{P}_\sigma(\tilde{A}_n^c) \geq 0$. From these two inequalities, we infer that

$$0 \leq \tilde{P}_\sigma(\tilde{A}^c) = \lim_{n\to+\infty} \tilde{P}_\sigma(\tilde{A}_n^c) \leq \lim_{n\to+\infty} \frac{(r-s)\lambda}{2n} = 0.$$

From this and Eq. (5.39), it follows that $\tilde{P}_\sigma(\tilde{A}) = 1$, which proves the second part of the statement. □

### 5.C.3 A càdlàg modification

We now use Proposition 5.50↶ to obtain a càdlàg modification of $(\tilde{X}_t)_{t\in\mathbb{R}_{\geq 0}}$. Crucial in this step is the following observation inspired by (Le Gall, 2016, Lemma 3.16), which concerns the right-sided limit of a path $\omega$ at a time point $t$ 'coming along D', denoted by $\lim_{D\ni r\searrow t}\omega(r)$. Formally, for any path $\omega$ in $\tilde{\Omega}$ and time point $t$ in $\mathbb{R}_{\geq 0}$, we say that $\lim_{D\ni r\searrow t}\omega(r)$ exists whenever there is some $x$ in $\mathscr{X}$ such that

$$\big(\exists\delta \in \mathbb{R}_{>0}\big)\big(\forall r \in\, ]t, t+\delta[\, \cap D\big)\ \omega(r) = x,$$

and whenever this is the case, we let $\lim_{D\ni r\searrow t}\omega(r) = x$.

**Lemma 5.51.** *Consider any countable dense subset $D$ of $\mathbb{R}_{\geq 0}$, any time point $s$ in $\mathbb{R}_{\geq 0}$ and any path $\omega$ in $\tilde{\Omega}$. If $\tilde{\eta}^{D}_{[s,r]}(\omega) < +\infty$ for all $r$ in $D$ such that $s < r$, then the right limit*

$$\lim_{D \ni t' \searrow t} \omega(t')$$

*exists for every time point $t$ in $[s, +\infty[$. Moreover, the modified path*

$$\omega' \colon [s, +\infty[ \to \mathscr{X} \colon t \mapsto \omega'(t) := \lim_{D \ni t' \searrow t} \omega(t')$$

*is then càdlàg over $[s, +\infty[$, meaning that it is continuous from the right at all $t$ in $[s, +\infty[$ and the limit from the left exists at all $t$ in $]s, +\infty[$.*

*Proof.* Fix any time points $t$ in $\mathbb{R}_{\geq 0}$ and $r$ in $D$ such that $s \leq t < r$. Because $\tilde{\eta}^{D}_{[s,r]}(\omega) < +\infty$ due to the conditions of the statement, there is a non-negative integer $n$ and a grid $v = (t_0, \dots, t_k)$ in $\mathscr{U}^{D}_{[s,r]}$ such that

$$\tilde{\eta}^{D}_{[s,r]}(\omega) = n = \tilde{\eta}_{v}(\omega).$$

Furthermore, for any grid $w$ in $\mathscr{U}^{D}_{[s,r]}$ that refines $v$, $\tilde{\eta}_{w}(\omega) = \tilde{\eta}_{v}(\omega) = n$ because

$$n = \tilde{\eta}_{v}(\omega) \leq \tilde{\eta}_{w}(\omega) \leq \tilde{\eta}^{D}_{[s,r]}(\omega) = n.$$

Thus, we see that no matter how many time points of $]s, r[ \cap D$ we add to the grid $v$, the number of jumps of $\omega$ along the grid will always be the same. This implies that for every $\ell$ in $\{1, \dots, k\}$, the path $\omega$ is piece-wise constant and changes states at most once along $[t_{\ell-1}, t_{\ell}] \cap D$; that is, for all $\ell$ in $\{1, \dots, k\}$, there is a $t^{\star}_{\ell}$ in $[t_{\ell-1}, t_{\ell}]$ such that

$$\omega(t') = \begin{cases} \omega(t_{\ell-1}) & \text{if } t' < t^{\star}_{\ell} \\ \omega(t_{\ell}) & \text{if } t' > t^{\star}_{\ell} \end{cases} \quad \text{for all } t' \in [t_{\ell-1}, t_{\ell}] \cap D.$$

Set $t^{\star}_{0} := s$ and $t^{\star}_{k+1} := r$. Then for every $\ell$ in $\{1, \dots, k+1\}$,

$$\omega(t') = \omega(t'') \quad \text{for all } t', t'' \in \,]t^{\star}_{\ell-1}, t^{\star}_{\ell}[\, \cap D,$$

where we interpret the interval $]t^{\star}_{\ell-1}, t^{\star}_{\ell}[$ as the empty set whenever $t^{\star}_{\ell-1} = t^{\star}_{\ell}$. Because $s \leq t < r$, there is an $\ell_t$ in $\{0, \dots, k\}$ such that $t^{\star}_{\ell_t} \leq t < t^{\star}_{\ell_t+1}$. Therefore, $]t, t^{\star}_{\ell_t+1}[\, \cap D \neq \emptyset$ and

$$\omega(t') = \omega(t'') \quad \text{for all } t', t'' \in \,]t, t^{\star}_{\ell_t+1}[\, \cap D. \tag{5.40}$$

Clearly, this implies that $\lim_{D \ni t' \searrow t} \omega(t')$ exists, which verifies the first part of the statement.

It also follows immediately from Eq. (5.40) that the modified path $\omega'$ is right-continuous at $t$. Furthermore, it is easy to see that the modified path $\omega'$ is constant over the intervals $[t^{\star}_{\ell-1}, t^{\star}_{\ell}[$. For this reason, the left-sided limit $\lim_{\Delta \searrow 0} \omega'(t - \Delta)$ at $t$ exists as well whenever $t > s$. Because $t$ is an arbitrary time point in $[s, +\infty[$, this proves the second part of the statement. ☐

Finally, we combine Proposition 5.50₂₅₈ with Lemma 5.51↶ to construct a càdlàg modification of $(\tilde{X}_t)_{t\in\mathbb{R}_{\geq 0}}$.

**Theorem 5.52.** *Consider a sequence of time points $u$ in $\mathscr{U}$ and a probability measure $\tilde{P}_\sigma$ on $\sigma(\tilde{\mathscr{F}}_u)$ that has uniformly bounded rate. Let $s := \max u$. Then there is a family $(\tilde{Y}_t)_{t\in[s,+\infty[}$ of $\mathscr{X}$-valued variables such that*

(i) *for any path $\omega$ in $\tilde{\Omega}$, the sample path*

$$\tilde{Y}_\bullet(\omega)\colon [s,+\infty[ \to \mathscr{X}\colon t \mapsto \tilde{Y}_t(\omega)$$

*is càdlàg over $[s,+\infty[$;*

(ii) *for any $t$ in $[s,+\infty[$ and $x$ in $\mathscr{X}$, the event*

$$\{\tilde{Y}_t = x\} := \left\{\omega \in \tilde{\Omega}\colon \tilde{Y}_t(\omega) = x\right\}$$

*belongs to $\sigma(\tilde{\mathscr{F}}_u)$;*

(iii) *for any $t$ in $[s,+\infty[$, the event*

$$\{\tilde{X}_t = \tilde{Y}_t\} := \bigcup_{x\in\mathscr{X}} \{\tilde{X}_t = x\} \cap \{\tilde{Y}_t = x\}$$

*belongs to $\sigma(\tilde{\mathscr{F}}_u)$ and $\tilde{P}_\sigma(\tilde{X}_t = \tilde{Y}_t) = 1$;*

(iv) *for any sequence of time points $v$ in $\mathscr{U}$ such that $v \subseteq [s,+\infty[$ and any subset $B$ of $\mathscr{X}_v$, the event*

$$\{\tilde{Y}_v \in B\} := \bigcup_{x_v\in B} \bigcap_{t\in v} \{\tilde{Y}_t = x_t\}$$

*belongs to $\sigma(\tilde{\mathscr{F}}_u)$, and*

$$\tilde{P}_\sigma(\tilde{Y}_v \in B) = \tilde{P}_\sigma(\tilde{X}_v \in B).$$

*Proof.* Fix any countable dense subset $D$ of $\mathbb{R}_{\geq 0}$. Recall from Proposition 5.50₂₅₈ that for any future time point $r$ in $[s,+\infty[$, $\{\tilde{\eta}^D_{[s,r]} \lessdot +\infty\}$ belongs to $\sigma(\tilde{\mathscr{F}}_u)$ and

$$\tilde{P}_\sigma\left(\{\tilde{\eta}^D_{[s,r]} \lessdot +\infty\}\right) = 1. \tag{5.41}$$

Let $(r_n)_{n\in\mathbb{N}}$ be any enumeration of $D \cap [s,+\infty[$. For any natural number $n$, we define the event

$$\tilde{A}_n := \{\tilde{\eta}^D_{[s,r_n]} \lessdot +\infty\}.$$

Recall from (right above) Eq. (5.41) that $\tilde{A}_n$ belongs to $\sigma(\tilde{\mathscr{F}}_u)$, so it follows from Lemma C.7₄₆₅ (ii) that the event

$$\tilde{A}_D := \bigcap_{n\in\mathbb{N}} \tilde{A}_n$$

belongs to $\sigma(\tilde{\mathscr{F}}_u)$. Furthermore, it follows from repeated application of Eq. (5.41) and (PM8)$_{461}$ that for all $k$ in $\mathbb{N}$,

$$\tilde{P}_\sigma\left(\bigcap_{n=1}^{k} \tilde{A}_n\right) = \tilde{P}_\sigma\left(\bigcap_{n=1}^{k} \{\tilde{\eta}^D_{[s,r_n]} < +\infty\}\right) = 1; \tag{5.42}$$

from this equality and Lemma C.3$_{463}$ (CA3), it follows immediately that

$$\tilde{P}_\sigma(\tilde{A}_D) = \lim_{k\to+\infty} \tilde{P}_\sigma\left(\bigcap_{n=1}^{k} \tilde{A}_n\right) = 1.$$

The key point is that by construction, every path $\omega$ in $\tilde{A}_D$ satisfies the conditions of Lemma 5.51$_{260}$. Consequently, for every $t$ in $[s,+\infty[$ we can define the $\mathscr{X}$-valued variable $\tilde{Y}_t$ by

$$\tilde{Y}_t(\omega) := \begin{cases} \lim_{D\ni r\searrow t}\omega(r) & \text{if } \omega\in\tilde{A}_D \\ \omega(s) & \text{otherwise} \end{cases} \quad \text{for all } \omega\in\tilde{\Omega}.$$

Observe that (i) follows immediately from the definition above and Lemma 5.51$_{260}$. Thus, what is left, is to prove the three remaining properties of $(\tilde{Y}_t)_{t\in[s,+\infty[}$.

To prove (ii), we fix a future time point $t$ in $[s,+\infty[$ and a state $x$. Let $(r_n)_{n\in\mathbb{N}}$ be any decreasing sequence in $]t,+\infty[\cap D$ that converges to $t$. Then it follows immediately from the definition of the variable $\tilde{Y}_t$ that for all $\omega$ in $\tilde{\Omega}$,

$$\tilde{Y}_t(\omega) = \begin{cases} \lim_{n\to+\infty}\omega(r_n) & \text{if } \omega\in\tilde{A}_D, \\ \omega(s) & \text{otherwise.} \end{cases} \tag{5.43}$$

Consequently,

$$\{\tilde{Y}_t = x\} = \left(\tilde{A}_D \cap \left(\bigcup_{n\in\mathbb{N}} \bigcap_{k=n}^{+\infty} \{\tilde{X}_{r_k} = x\}\right)\right) \cup \left(\tilde{A}_D^{\mathrm{c}} \cap \{\tilde{X}_s = x\}\right).$$

Because $\tilde{A}_D$ and all $\{\tilde{X}_{r_n} = x\}$ and $\{\tilde{X}_s = x\}$ belong to $\sigma(\tilde{\mathscr{F}}_u)$, it follows from this and the properties of σ-fields that $\{\tilde{Y}_t = x\}$ belongs to $\sigma(\tilde{\mathscr{F}}_u)$.

Next, we prove (iii). Again, we fix any future time point $t$ in $[s,+\infty[$, and let $(r_n)_{n\in\mathbb{N}}$ be any decreasing sequence in $]t,+\infty[\cap D$ that converges to $t$. Recall that for all $x$ in $\mathscr{X}$, $\{\tilde{X}_t = x\}$ belongs to $\sigma(\tilde{\mathscr{F}}_u)$ by construction of $\tilde{\mathscr{F}}_u$ and $\{\tilde{Y}_t = x\}$ belongs to $\sigma(\tilde{\mathscr{F}}_u)$ by (ii). Consequently,

$$\{\tilde{X}_t = \tilde{Y}_t\} = \bigcup_{x\in\mathscr{X}} \{\tilde{X}_t = x\} \cap \{\tilde{Y}_t = x\}$$

belongs to $\sigma(\tilde{\mathscr{F}}_u)$ as well because this σ-field is closed under countable – and hence finite – unions. Consequently, $\{\tilde{X}_t \neq \tilde{Y}_t\} := \{\tilde{X}_t = \tilde{Y}_t\}^{\mathrm{c}}$ belongs to $\sigma(\tilde{\mathscr{F}}_u)$ due to (F2)$_{32}$, and it follows from (PM6)$_{461}$ that

$$\tilde{P}_\sigma(\tilde{X}_t = \tilde{Y}_t) = 1 - \tilde{P}_\sigma(\tilde{X}_t \neq \tilde{Y}_t). \tag{5.44}$$

Observe that

$$\{\tilde{X}_t \neq \tilde{Y}_t\} = \left(\tilde{A}_D \cap \{\tilde{X}_t \neq \tilde{Y}_t\}\right) \cup \left(\tilde{A}_D^{\mathrm{c}} \cap \{\tilde{X}_t \neq \tilde{Y}_t\}\right).$$

From this and Eq. (5.43)↶, it follows that

$$\mathbb{I}_{\{\tilde{X}_t \neq \tilde{Y}_t\}} = \mathbb{I}_{\left(\tilde{A}_D \cap \{\tilde{X}_t \neq \tilde{Y}_t\}\right) \cup \left(\tilde{A}_D^{\mathrm{c}} \cap \{\tilde{X}_t \neq \tilde{Y}_t\}\right)} = \operatorname*{p-w\,lim}_{n \to +\infty} \mathbb{I}_{\left(\tilde{A}_D \cap \{\tilde{X}_t \neq \tilde{X}_{r_n}\}\right) \cup \left(\tilde{A}_D^{\mathrm{c}} \cap \{\tilde{X}_t \neq \tilde{X}_s\}\right)}.$$

Because $\tilde{A}_D$, $\{\tilde{X}_t \neq \tilde{X}_{r_n}\}$ and $\{\tilde{X}_t \neq \tilde{X}_s\}$ belong to $\sigma(\tilde{\mathscr{F}}_u)$, it follows from this, (DE1)$_{225}$ and Theorem 5.11$_{226}$, that

$$\begin{aligned}\tilde{P}_\sigma(\tilde{X}_t \neq \tilde{Y}_t) = E^{\mathrm{D}}_{\tilde{P}_\sigma}(\mathbb{I}_{\{\tilde{X}_t \neq \tilde{Y}_t\}}) &= \lim_{n \to +\infty} E^{\mathrm{D}}_{\tilde{P}_\sigma}\left(\mathbb{I}_{\left(\tilde{A}_D \cap \{\tilde{X}_t \neq \tilde{X}_{r_n}\}\right) \cup \left(\tilde{A}_D^{\mathrm{c}} \cap \{\tilde{X}_t \neq \tilde{X}_s\}\right)}\right) \\ &= \lim_{n \to +\infty} \tilde{P}_\sigma\left(\left(\tilde{A}_D \cap \{\tilde{X}_t \neq \tilde{X}_{r_n}\}\right) \cup \left(\tilde{A}_D^{\mathrm{c}} \cap \{\tilde{X}_t \neq \tilde{X}_s\}\right)\right).\end{aligned}$$

Observe that because $\tilde{P}_\sigma(\tilde{A}_D) = 1$, it follows from (PM8)$_{461}$ that for all $n$ in $\mathbb{N}$,

$$\tilde{P}_\sigma\left(\left(\tilde{A}_D \cap \{\tilde{X}_t \neq \tilde{X}_{r_n}\}\right) \cup \left(\tilde{A}_D^{\mathrm{c}} \cap \{\tilde{X}_t \neq \tilde{X}_s\}\right)\right) = \tilde{P}_\sigma\left(\tilde{A}_D \cap \{\tilde{X}_t \neq \tilde{X}_{r_n}\}\right) = \tilde{P}_\sigma\left(\tilde{X}_t \neq \tilde{X}_{r_n}\right).$$

We substitute this equality in the preceding equality, to yield

$$\tilde{P}_\sigma(\tilde{X}_t \neq \tilde{Y}_t) = \lim_{n \to +\infty} \tilde{P}_\sigma\left(\tilde{X}_t \neq \tilde{X}_{r_n}\right).$$

Finally, because $(r_n)_{n \in \mathbb{N}}$ decreases to $t$ and $\tilde{P}_\sigma$ has uniformly bounded rate, we can conclude that

$$\tilde{P}_\sigma(\tilde{X}_t \neq \tilde{Y}_t) = \lim_{n \to +\infty} \tilde{P}_\sigma(\tilde{X}_t \neq \tilde{X}_{r_n}) = 0.$$

The statement follows immediately from this and Eq. (5.44)↶.

Finally, we prove (iv)$_{261}$. The statement holds trivially in case $v$ is the empty sequence of time points, so we may assume without loss of generality that $v$ is non-empty. Because in the definition

$$\{\tilde{Y}_v \in B\} = \bigcup_{x_v \in B} \bigcap_{t \in v} \{\tilde{Y}_t = x_t\},$$

all events on the right-hand side belong to $\sigma(\tilde{\mathscr{F}})$ due to (ii)$_{261}$, so does $\{\tilde{Y}_v \in B\}$. This verifies the first part of (iv)$_{261}$. To verify the second part, we observe that

$$\{\tilde{X}_v = \tilde{Y}_v\} := \bigcap_{t \in v} \{\tilde{X}_t = \tilde{Y}_t\}$$

belongs to $\sigma(\tilde{\mathscr{F}}_u)$ because all events on the right-hand side belong to $\sigma(\tilde{\mathscr{F}}_u)$, this time by construction of $\tilde{\mathscr{F}}_u$ and due to (iii)$_{261}$. Because furthermore $\tilde{P}_\sigma(\tilde{X}_t = \tilde{Y}_t) = 1$ for all $t$ in $v$ due to (iii)$_{261}$, it follows from (PM8)$_{461}$ that $\tilde{P}_\sigma(\tilde{X}_v = \tilde{Y}_v) = 1$. For this reason, it follows from (PM8)$_{461}$ that

$$\tilde{P}_\sigma(\tilde{Y}_v \in B) = \tilde{P}_\sigma\left(\{\tilde{Y}_v \in B\} \cap \{\tilde{X}_v = \tilde{Y}_v\}\right) = \tilde{P}_\sigma\left(\{\tilde{X}_v = \tilde{Y}_v\} \cap \{\tilde{X}_v \in B\}\right) = \tilde{P}_\sigma(\tilde{X}_v \in B),$$

which verifies the second part of (iv)$_{261}$. □

### 5.C.4 From the set of càdlàg paths to the set of all paths and back

In order to use the results that we have just proven, we need a way to go from finitary events on the set of càdlàg paths to finitary events on the set of all paths and vice versa. The most obvious way to do this is through the following 'projection'. For any sequence of time points $u$ in $\mathscr{U}$, we let

$$\Gamma_u\colon \tilde{\mathscr{F}}_u \mapsto \mathscr{F}_u\colon \tilde{A} \mapsto \Gamma_u(\tilde{A}) := \tilde{A} \cap \Omega. \tag{5.45}$$

To see that the image $\Gamma(\tilde{A})$ of an event $\tilde{A}$ in $\tilde{\mathscr{F}}_u$ indeed belongs to $\mathscr{F}_u$, we recall that by construction, the event $\tilde{A}$ in $\tilde{\mathscr{F}}_u$ has a (non-unique) representation of the form $(\tilde{X}_v \in B)$, with $v$ in $\mathscr{U}_{\succcurlyeq u}$. Observe that

$$\begin{aligned}\Gamma_u(\tilde{A}) = \tilde{A} \cap \Omega = \{\tilde{X}_v \in B\} \cap \Omega &= \big\{\omega \in \tilde{\Omega}\colon \omega|_v \in B\big\} \cap \Omega \\ &= \big\{\omega \in \Omega\colon \omega|_v \in B\big\} = \{X_v \in B\},\end{aligned}$$

from which we infer that $\Gamma_u(\tilde{A})$ belongs to $\mathscr{F}_u$. In fact, it is straightforward to verify that $\Gamma_u$ is a bijection between $\mathscr{F}_u$ and $\tilde{\mathscr{F}}_u$ that has some nice properties.

**Lemma 5.53.** *Consider some sequence of time points $u$ in $\mathscr{U}$. The projection $\Gamma_u$ is a bijection, and*

(i) $\Gamma_u(\tilde{\Omega}) = \Omega$;

(ii) $\Gamma_u(\tilde{A}^{\mathrm{c}}) = \Omega \setminus \Gamma_u(\tilde{A})$ *for all $\tilde{A}$ in $\tilde{\mathscr{F}}_u$;*

(iii) $\Gamma_u(\tilde{A} \cup \tilde{B}) = \Gamma_u(\tilde{A}) \cup \Gamma_u(\tilde{B})$ *for all $\tilde{A}, \tilde{B}$ in $\tilde{\mathscr{F}}_u$, and* $\Gamma_u(\tilde{A}) \cap \Gamma_u(\tilde{B}) = \varnothing$ *whenever* $\tilde{A} \cap \tilde{B} = \varnothing$.

*Proof.* First, we recall that every event $A$ in $\mathscr{F}_u$ has a (non-unique) representation $\{X_v \in B\}$, and similarly for every event $\tilde{A}$ in $\tilde{\mathscr{F}}_u$. To prove that $\Gamma_u$ is a bijection, we verify that it is injective (one-to-one) and subsequently verify that it is surjective (onto).

In order to verify that $\Gamma_u$ is injective, we fix two events $\tilde{A}_1$ and $\tilde{A}_2$ in $\tilde{\mathscr{F}}_u$. Let $\{\tilde{X}_{v_1} \in B_1\}$ and $\{\tilde{X}_{v_2} \in B_2\}$ be non-unique representations of $\tilde{A}_1$ and $\tilde{A}_2$, respectively. Let $v := v_1 \cup v_2$, $B'_1 := \{x_v \in \mathscr{X}_v\colon x_{v_1} \in B_1\}$ and $B'_2 := \{x_v \in \mathscr{X}_v\colon x_{v_2} \in B_2\}$. Then by construction,

$$\tilde{A}_1 = \{\tilde{X}_{v_1} \in B_1\} = \{\tilde{X}_v \in B'_1\} \quad \text{and} \quad \tilde{A}_2 = \{\tilde{X}_{v_2} \in B_2\} = \{\tilde{X}_v \in B'_2\},$$

and therefore

$$\tilde{A}_1 = \tilde{A}_2 \Leftrightarrow B'_1 = B'_2. \tag{5.46}$$

By definition of $\Gamma_u$,

$$\Gamma_u(\tilde{A}_1) = \tilde{A}_1 \cap \Omega = \{\tilde{X}_v \in B'_1\} \cap \Omega = \{X_v \in B'_1\}$$

and

$$\Gamma_u(\tilde{A}_2) = \tilde{A}_2 \cap \Omega = \{\tilde{X}_v \in B'_2\} \cap \Omega = \{X_v \in B'_2\}.$$

It follows from this and Eq. (5.46) that

$$\Gamma_u(\tilde{A}_1) = \Gamma_u(\tilde{A}_2) \Leftrightarrow B'_1 = B'_2 \Leftrightarrow \tilde{A}_1 = \tilde{A}_2,$$

so $\Gamma_u$ is injective.

In order to verify that $\Gamma_u$ is surjective, we fix an arbitrary event $A$ in $\mathscr{F}_u$. Let $\{X_v \in B\}$ be a non-unique representation of the event $A$. It is clear that the event $\tilde{A} := \{\tilde{X}_v \in B\}$ in $\tilde{\mathscr{F}}$ is mapped to $A$ by $\Gamma_u$:

$$\Gamma_u(\tilde{A}) = \{\tilde{X}_v \in B\} \cap \Omega = \{X_v \in B\} = A,$$

so $\Gamma_u$ is indeed onto.

Finally, the three properties of the second part of the statement follow immediately from the definition of $\Gamma_u$. □

Let $P$ be any jump process. Recall from Corollary 2.58<sub>46</sub> that for any state history $\{X_u = x_u\}$ in $\mathscr{H}$, $P(\bullet \mid X_u = x_u)$ is a probability charge on the field $\mathscr{F}_u$. This probability charge induces a probability charge $\tilde{P}^{x_u}$ on $\tilde{\mathscr{F}}_u$ through the projection $\Gamma_u$:

$$\tilde{P}^{x_u} \colon \tilde{\mathscr{F}}_u \to \mathbb{R} \colon \tilde{A} \mapsto \tilde{P}^{x_u}(\tilde{A}) := P(\Gamma_u(\tilde{A}) \mid X_u = x_u).$$

**Corollary 5.54.** *Consider a jump process $P$. Then for any $\{X_u = x_u\}$ in $\mathscr{H}$, $\tilde{P}^{x_u}$ is a probability charge on $\tilde{\mathscr{F}}_u$.*

*Proof.* That $\tilde{P}^{x_u}$ is a probability charge on $\tilde{\mathscr{F}}_u$ follows immediately from Corollary 2.58[46], Lemma 5.53 and Definition 2.36[34]. □

This induced probability charge $\tilde{P}^{x_u}$ is countably additive due to Theorem 5.44[252], so it follows from Theorem C.10[466] that there is a unique probability measure $\tilde{P}^{x_u}_\sigma$ on $\sigma(\tilde{\mathscr{F}}_u)$ that extends it. Even more, this measure has bounded rate whenever $P$ has bounded rate.

**Lemma 5.55.** *Consider a jump process $P$. Then for any state history $\{X_u = x_u\}$ in $\mathscr{H}$, there is a unique probability measure $\tilde{P}^{x_u}_\sigma$ on $\sigma(\tilde{\mathscr{F}}_u)$ that extends $\tilde{P}^{x_u}$. If the jump process $P$ has uniformly bounded rate, then so does the induced probability measure $\tilde{P}^{x_u}_\sigma$.*

*Proof.* Recall from Corollary 5.54 that $\tilde{P}^{x_u}$ is a probability charge on $\tilde{\mathscr{F}}_u$, so it is countably additive due to Theorem 5.44[252]. For this reason, it follows immediately from Theorem C.10[466] that there is a unique probability charge $P^{x_u}_\sigma$ on $\sigma(\tilde{\mathscr{F}}_u)$ that extends it.

To verify the second part of the statement, we assume that $P$ has bounded rate and we fix two time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Observe that $\{X_s \neq X_r\} = \{X_v \in B\}$, with $v := (s, r)$ and $B := \{y_v \in \mathscr{X}_v \colon y_s \neq y_r\}$, and similarly for $\{\tilde{X}_s \neq \tilde{X}_r\}$. Because

$$\Gamma_u\big(\{\tilde{X}_s \neq \tilde{X}_r\}\big) = \Gamma_u\big(\{\tilde{X}_v \in B\}\big) = \{X_v \in B\} = \{X_s \neq X_r\},$$

it follows from the definition of $\tilde{P}^{x_u}$ that

$$P(X_s \neq X_r \mid X_u = x_u) = \tilde{P}^{x_u}(\tilde{X}_s \neq \tilde{X}_r) = \tilde{P}^{x_u}_{\sigma}(\tilde{X}_s \neq \tilde{X}_r).$$

Because this equality holds for any two time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$, it follows from Lemma 5.35$_{242}$ that $\tilde{P}^{x_u}_{\sigma}$ has uniformly bounded rate. □

The final intermediary result that we will need in the proof of Theorem 5.19$_{230}$ deals with non-increasing sequences of events in $\mathscr{F}_u$.

**Lemma 5.56.** *Consider a jump process $P$, a state history $\{X_u = x_u\}$ in $\mathscr{H}$ and a sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathscr{F}_u$, and let $s := \max u$. Then for every $n$ in $\mathbb{N}$, there is a sequence of time points $v_n$ in $\mathscr{U}_{\succcurlyeq(s)}$ with $s \in v_n$ and a subset $B_n$ of $\mathscr{X}_{v_n}$ such that*

$$P(A_n \mid X_u = x_u) = P(X_{v_n} \in B_n \mid X_u = x_u) \quad \textit{for all } n \in \mathbb{N}.$$

*Whenever the sequence $(A_n)_{n\in\mathbb{N}}$ is non-increasing and $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$, the same holds for the sequence $(\{X_{v_n} \in B_n\})_{n\in\mathbb{N}}$.*

*Proof.* From Lemma 3.11$_{64}$, we recall that for every $n$ in $\mathbb{N}$, there is a sequence of time points $v'_n$ in $\mathscr{U}_{>u}$ and a subset $B'_n$ of $\mathscr{X}_{u\cup v'_n}$ such that $A_n = \{X_{u\cup v'_n} \in B'_n\}$. For any natural number $n$, we let $v_n := (s) \cup v'_n$ – note that $v_n \succcurlyeq (s)$ and $s \in v_n$ – and

$$B_n := \{y_{v_n} : y_{u\cup v'_n} \in B''_n\} \quad \text{with } B''_n := \{y_{u\cup v'_n} \in B'_n : y_u = x_u\},$$

and observe that

$$\{X_u = x_u\} \cap A_n = \{X_u = x_u\} \cap \{X_{u\cup v'_n} \in B'_n\} = \{X_u = x_u\} \cap \{X_{v_n} \in B_n\}. \tag{5.47}$$

Recall from (CP1)$_{41}$ that $P(X_u = x_u \mid X_u = x_u) = 1$. For this reason, it follows from Eq. (5.47) and (CP9)$_{42}$ that

$$\begin{aligned} P(A_n \mid X_u = x_u) &= P(\{X_u = x_u\} \cap A_n \mid X_u = x_u) \\ &= P(\{X_u = x_u\} \cap \{X_{v_n} \in B_n\} \mid X_u = x_u) \\ &= P(X_{v_n} \in B_n \mid X_u = x_u). \end{aligned}$$

Because this equality holds for all $n$ in $\mathbb{N}$, we have proven the first part of the statement.

Next, we turn to the second part of the statement; that is, from here on we assume that $(A_n)_{n\in\mathbb{N}}$ is non-increasing with $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$. In order to show that the sequence $(\{X_{v_n} \in B_n\})_{n\in\mathbb{N}}$ is non-increasing, we assume *ex absurdo* that it is not, meaning that there is an $n$ in $\mathbb{N}$ such that $\{X_{v_{n+1}} \in B_{n+1}\} \cap \{X_{v_n} \in B_n\}^{\mathrm{c}} \neq \emptyset$. Fix any càdlàg path $\omega_1$ in $\{X_{v_{n+1}} \in B_{n+1}\} \cap \{X_{v_n} \in B_n\}^{\mathrm{c}}$. Due to Corollary 3.9$_{63}$, there is a càdlàg path $\omega_2$ in $\{X_u = x_u\}$. Note that $\omega_1(s) = x_s$ by definition of $B_{n+1}$; because $\omega_2(s) = x_s$ as well, the path

$$\omega : \mathbb{R}_{\geq 0} \to \mathscr{X} : t \mapsto \begin{cases} \omega_1(t) & \text{if } t \geq s \\ \omega_2(t) & \text{if } t < s \end{cases}$$

is càdlàg. By construction, $\omega$ clearly belongs to $\{X_u = x_u\}$. Furthermore, we recall that $\omega_1 \in \{X_{v_{n+1}} \in B_{n+1}\}$ by assumption and we observe that $\omega|_{v_{n+1}} = \omega_1|_{v_{n+1}}$ because $v_{n+1} \succcurlyeq (s)$; consequently, $\omega \in \{X_{v_{n+1}} \in B_{n+1}\}$ as well. However, $\omega \notin \{X_{v_n} \in B_n\}$, because $\omega|_{v_n} = \omega_1|_{v_n}$ and because $\omega_1 \notin \{X_{v_n} \in B_n\}$ by assumption. Thus,

$$\omega \in \big(\{X_u = x_u\} \cap \{X_{v_{n+1}} \in B_{n+1}\}\big) \setminus \big(\{X_u = x_u\} \cap \{X_{v_n} \in B_n\}\big),$$

so by Eq. (5.47)$_\curvearrowleft$,

$$\omega \in \big(\{X_u = x_u\} \cap A_{n+1}\big) \setminus \big(\{X_u = x_u\} \cap A_n\big). \tag{5.48}$$

However, because $(A_m)_{m\in\mathbb{N}}$ is non-increasing, $A_n \supseteq A_{n+1}$ and therefore

$$\{X_u = x_u\} \cap A_n \supseteq \{X_u = x_u\} \cap A_{n+1},$$

which clearly contradicts Eq. (5.48). Consequently, $(\{X_{v_n} \in B_n\})_{n\in\mathbb{N}}$ is non-increasing, as required.

Finally, we show that $\bigcap_{n\in\mathbb{N}}\{X_{v_n} \in B_n\} = \emptyset$. To this end, we assume *ex absurdo* that $\bigcap_{n\in\mathbb{N}}\{X_{v_n} \in B_n\} \neq \emptyset$. Fix any $\omega_1$ in $\bigcap_{n\in\mathbb{N}}\{X_{v_n} \in B_n\}$ and $\omega_2$ in $\{X_u = x_u\}$. Because $\omega_1(s) = x_s = \omega_2(s)$, the path

$$\omega \colon \mathbb{R}_{\geq 0} \to \mathscr{X} \colon t \mapsto \begin{cases} \omega_1(t) & \text{if } t \geq s \\ \omega_2(t) & \text{if } t < s \end{cases}$$

is càdlàg. Furthermore, $\omega \in \{X_u = x_u\}$ and, for all $n$ in $\mathbb{N}$, $\omega \in \{X_{v_n} \in B_n\}$ – because $\omega_1 \in \{X_{v_n} \in B_n\}$ and $\omega|_{v_n} = \omega_1|_{v_n}$ since $v_n \succcurlyeq (s)$. Consequently,

$$\omega \in \{X_u = x_u\} \cap \left(\bigcap_{n\in\mathbb{N}} \{X_{v_n} \in B_n\}\right). \tag{5.49}$$

In order to obtain the contradiction that we are after, we observe that

$$\begin{aligned}\{X_u = x_u\} \cap \left(\bigcap_{n\in\mathbb{N}} \{X_{v_n} \in B_n\}\right) &= \bigcap_{n\in\mathbb{N}} \big(\{X_u = x_u\} \cap \{X_{v_n} \in B_n\}\big) \\ &= \bigcap_{n\in\mathbb{N}} \big(\{X_u = x_u\} \cap A_n\big) = \{X_u = x_u\} \cap \left(\bigcap_{n\in\mathbb{N}} A_n\right),\end{aligned}$$

where the second equality holds due to Eq. (5.47)$_\curvearrowleft$. Because $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$ by assumption, we infer from this that

$$\{X_u = x_u\} \cap \left(\bigcap_{n\in\mathbb{N}} \{X_{v_n} \in B_n\}\right) = \emptyset,$$

which contradicts Eq. (5.49). Consequently, $\bigcap_{n\in\mathbb{N}}\{X_{v_n} \in B_n\} = \emptyset$. □

### 5.C.5 Assembling the proof

Finally, we can join all the preceding intermediary results in our proof of Theorem 5.19$_{230}$.

**Theorem 5.19.** *If a jump process $P$ has uniformly bounded rate, then it is countably additive.*

*Proof of Theorem 5.19*$_{230}$. We need to verify that for every state history $\{X_u = x_u\}$ in $\mathscr{H}$, $P(\bullet \,|\, X_u = x_u)$ is countably additive. To this end, we fix any state history $\{X_u = x_u\}$ in $\mathscr{H}$. By Lemma C.3$_{463}$ (CA4), it suffices to check that $\lim_{n\to+\infty} P(A_n \,|\, X_u = x_u) = 0$ for any non-increasing sequence $(A_n)_{n\in\mathbb{N}}$ of non-empty events in $\mathscr{F}_u$ such that $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$. Hence, we fix any such sequence.

Let $s := \max u$. By Lemma 5.56$_{266}$, for every $n$ in $\mathbb{N}$, there is a $v_n$ in $\mathscr{U}_{\succcurlyeq(s)}$ with $s \in v_n$ and a subset $B_n$ of $\mathscr{X}_{v_n}$ such that

$$P(A_n \,|\, X_u = x_u) = P(X_{v_n} \in B_n \,|\, X_u = x_u)$$

and such that $(\{X_{v_n} \in B_n\})_{n\in\mathbb{N}}$ is non-increasing and $\bigcap_{n\in\mathbb{N}}\{X_{v_n} \in B_n\} = \emptyset$. Therefore, it is clear that

$$\lim_{n\to+\infty} P(A_n \,|\, X_u = x_u) = \lim_{n\to+\infty} P(X_{v_n} \in B_n \,|\, X_u = x_u). \tag{5.50}$$

We now show that the limit on the right is equal to 0, which of course implies that the limit on the left is equal to 0 as well, as required.

Recall from Lemma 5.55$_{265}$ that $\tilde{P}^{x_u}_{\sigma}$ is the unique probability measure on $\sigma(\tilde{\mathscr{F}}_u)$ that extends the induced probability charge $\tilde{P}^{x_u}$ on $\tilde{\mathscr{F}}_u$, and that $\tilde{P}^{x_u}_{\sigma}$ has uniformly bounded rate. Observe that

$$\tilde{P}^{x_u}_{\sigma}(\tilde{X}_{v_n} \in B_n) = \tilde{P}^{x_u}(\tilde{X}_{v_n} \in B_n) = P(X_{v_n} \in B_n \,|\, X_u = x_u) \quad \text{for all } n \in \mathbb{N}, \tag{5.51}$$

where the final equality holds due to the definition of $\tilde{P}^{x_u}$. Because $\tilde{P}^{x_u}_{\sigma}$ has uniformly bounded rate, we can invoke Theorem 5.52$_{261}$ to obtain a family $(\tilde{Y}_t)_{t\in[s,+\infty[}$ with the four properties as specified in that theorem. More precisely, it follows from Theorem 5.52$_{261}$ (iv) that

$$\tilde{P}^{x_u}_{\sigma}(\tilde{Y}_{v_n} \in B_n) = \tilde{P}^{x_u}_{\sigma}(\tilde{X}_{v_n} \in B_n) \quad \text{for all } n \in \mathbb{N}. \tag{5.52}$$

Recall that $(\{X_{v_n} \in B_n\})_{n\in\mathbb{N}}$ is non-increasing with $\bigcap_{n\in\mathbb{N}}\{X_{v_n} \in B_n\} = \emptyset$. We now set out to prove that this implies that $(\{\tilde{Y}_{v_n} \in B_n\})_{n\in\mathbb{N}}$ is non-increasing with $\bigcap_{n\in\mathbb{N}}\{\tilde{Y}_{v_n} \in B_n\} = \emptyset$.

First, we establish that $(\{\tilde{Y}_{v_n} \in B_n\})_{n\in\mathbb{N}}$ is non-increasing. To this end, we assume *ex absurdo* that there is a natural number $n$ such that $\{\tilde{Y}_{v_n} \in B_n\}^{\mathrm{c}} \cap \{\tilde{Y}_{v_{n+1}} \in B_{n+1}\} \neq \emptyset$, and we fix any path $\tilde{\omega}$ in this non-empty intersection. Then the path

$$\omega^\star \colon \mathbb{R}_{\geq 0} \to \mathscr{X} \colon t \mapsto \omega^\star(t) := \begin{cases} \tilde{Y}_t(\tilde{\omega}) & \text{it } t \geq s, \\ \tilde{Y}_s(\tilde{\omega}) & \text{otherwise} \end{cases}$$

is càdlàg because $(\tilde{Y}_t)_{t\in[s,+\infty[}$ has càdlàg sample paths by Theorem 5.52$_{261}$ (i). Furthermore, it is clear that $\omega^\star|_{v_n} \in B_n^{\mathrm{c}}$ and $\omega^\star|_{v_{n+1}} \in B_{n+1}$ because $\tilde{\omega}$ is an element of $\{\tilde{Y}_{v_n} \in B_n\}^{\mathrm{c}} = \{\tilde{Y}_{v_n} \in B_n^{\mathrm{c}}\}$ and $\{\tilde{Y}_{v_{n+1}} \in B_{n+1}\}$. For this reason, $\omega^\star$ belongs to $\{X_{v_n} \in B_n^{\mathrm{c}}\} = \{X_{v_n} \in B_n\}^{\mathrm{c}}$ and $\{X_{v_{n+1}} \in B_{n+1}\}$; this implies that

$$\{X_{v_n} \in B_n\}^{\mathrm{c}} \cap \{X_{v_{n+1}} \in B_{n+1}\} \neq \emptyset,$$

but this is a contradiction because $(\{X_{v_n} \in B_n\})_{n\in\mathbb{N}}$ is non-increasing.

Second, we establish that $\bigcap_{n\in\mathbb{N}}\{\tilde{Y}_{v_n} \in B_n\} = \emptyset$. To this end, we assume *ex absurdo* that there is a path $\tilde{\omega}$ in $\bigcap_{n\in\mathbb{N}}\{\tilde{Y}_{v_n} \in B_n\}$. Then the path

$$\omega^\star : \mathbb{R}_{\geq 0} \to \mathscr{X} : t \mapsto \omega^\star(t) := \begin{cases} \tilde{Y}_t(\tilde{\omega}) & \text{it } t \geq s, \\ \tilde{Y}_s(\tilde{\omega}) & \text{otherwise} \end{cases}$$

is càdlàg because $(\tilde{Y}_t)_{t\in[s,+\infty[}$ has càdlàg sample paths by Theorem 5.52$_{261}$ (i). Furthermore, because

$$v_n \succcurlyeq (s) \quad \text{and} \quad \tilde{\omega} \in \{\tilde{Y}_{v_n} \in B_n\} \quad \text{for all } n \in \mathbb{N},$$

we see that, by construction,

$$\omega^\star|_{v_n} \in B_n \quad \text{for all } n \in \mathbb{N}.$$

This implies that $\omega^\star \in \bigcap_{n\in\mathbb{N}}\{X_{v_n} \in B_n\}$, but this is a contradiction because $\bigcap_{n\in\mathbb{N}}\{X_{v_n} \in B_n\} = \emptyset$.

Because $(\{\tilde{Y}_{v_n} \in B_n\})_{n\in\mathbb{N}}$ is a non-increasing sequence of events in $\sigma(\tilde{\mathscr{F}}_u)$ with $\bigcap_{n\in\mathbb{N}}\{\tilde{Y}_{v_n} \in B_n\} = \emptyset$, it follows from Lemma C.3$_{463}$ (CA4) that

$$\lim_{n\to+\infty} \tilde{P}_\sigma^{x_u}(\tilde{Y}_{v_n} \in B_n) = 0.$$

Finally, it follows from this and Eqs. (5.50)$_\curvearrowleft$ to (5.52)$_\curvearrowleft$ that

$$\lim_{n\to+\infty} P(A_n \,|\, X_u = x_u) = \lim_{n\to+\infty} \tilde{P}_\sigma^{x_u}(\tilde{Y}_{v_n} \in B_n) = 0,$$

as required. □

## 5.D Continuity properties of the lower and upper envelopes

In the final appendix to this chapter, we prove the two continuity properties of the lower and upper expectations with respect to an imprecise jump process $\mathscr{P}$. First, we investigate the continuity with respect to monotone sequences of limit variables.

**Theorem 5.31.** *Consider an imprecise jump process $\mathscr{P}$ that consists of countably additive jump processes. Fix some $\{X_u = x_u\}$ in $\mathscr{H}$ and $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$, and let $(f_n)_{\in\mathbb{N}}$ be a sequence of variables in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ that converges monotonically to $f$. If $(f_n)_{n\in\mathbb{N}} \nearrow f$ and $\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_1 \,|\, X_u = x_u) > -\infty$, then*

$$\lim_{n\to+\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \leq \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u)$$

*and*

$$\lim_{n\to+\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) = \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u).$$

*Similarly, if $(f_n)_{n\in\mathbb{N}} \searrow f$ and $\overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_1 \,|\, X_u = x_u) < +\infty$, then*

$$\lim_{n\to+\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) = \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u)$$

*and*

$$\lim_{n\to+\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \geq \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u).$$

*Proof.* We only prove the first part of the statement; the proof for the second part is analoguous. Note that because $(f_n)_{n\in\mathbb{N}} \nearrow f$ by assumption, $f_n \leq f_{n+1}$ for all $n$ in $\mathbb{N}$ and $\text{p-w}\lim_{n\to+\infty} f_n = f$. Furthermore, for all $n$ in $\mathbb{N}$, $f_n$ is a D-integrable variable – that is, belongs to $\mathbb{D}^{\mathrm{D}}_{P(\bullet|X_u=x_u)}$ – due to Theorem 5.12$_{227}$.

Fix some $P$ in $\mathscr{P}$. Then it follows from (DE6)$_{226}$ that the sequence $(E^{\mathrm{D}}_P(f_n \,|\, X_u = x_u))_{n\in\mathbb{N}}$ of extended real numbers is non-decreasing. Furthermore, we observe that

$$E^{\mathrm{D}}_P(f_1 \,|\, X_u = x_u) \geq \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_1 \,|\, X_u = x_u) > -\infty,$$

where the strict inequality holds by assumption. Consequently, it follows from Theorem 5.10$_{226}$ that

$$E^{\mathrm{D}}_P(f \,|\, X_u = x_u) = \lim_{n\to+\infty} E^{\mathrm{D}}_P(f_n \,|\, X_u = x_u) = \sup\left\{E^{\mathrm{D}}_P(f_n \,|\, X_u = x_u) \colon n \in \mathbb{N}\right\}, \tag{5.53}$$

where the final equality holds because $(E^{\mathrm{D}}_P(f_n \,|\, X_u = x_u))_{n\in\mathbb{N}}$ is non-decreasing.

Because for every $P$ in $\mathscr{P}$, $(E^{\mathrm{D}}_P(f_n \,|\, X_u = x_u))_{n\in\mathbb{N}}$ is a non-decreasing sequence of extended real numbers, it is easy to see that

$$\left(\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u)\right)_{n\in\mathbb{N}} \quad \text{and} \quad \left(\overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u)\right)_{n\in\mathbb{N}}$$

are non-decreasing sequences of extended real numbers, so the limits of these sequences exist. Because these sequences are non-decreasing,

$$\lim_{n\to+\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) = \sup\left\{\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \colon n \in \mathbb{N}\right\} \tag{5.54}$$

and

$$\lim_{n\to+\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) = \sup\left\{\overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \colon n \in \mathbb{N}\right\} \tag{5.55}$$

To verify the equality in the first part of the statement, we recall that by definition of the upper envelope $\overline{E}^{\mathrm{D}}_{\mathscr{P}}$,

$$\overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u) = \sup\left\{E^{\mathrm{D}}_P(f \,|\, X_u = x_u) \colon P \in \mathscr{P}\right\}.$$

From this and Eq. (5.53), it follows that

$$\begin{aligned}
\overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u) &= \sup\left\{\sup\left\{E^{\mathrm{D}}_P(f_n \,|\, X_u = x_u) \colon n \in \mathbb{N}\right\} \colon P \in \mathscr{P}\right\} \\
&= \sup\left\{\sup\left\{E^{\mathrm{D}}_P(f_n \,|\, X_u = x_u) \colon P \in \mathscr{P}\right\} \colon n \in \mathbb{N}\right\} \\
&= \sup\left\{\overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \colon n \in \mathbb{N}\right\}.
\end{aligned}$$

The equality in the first part of the statement follows immediately from the preceding equality and Eq. (5.55):

$$\overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u) = \sup\left\{\overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u) \colon n \in \mathbb{N}\right\} = \lim_{n\to+\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \,|\, X_u = x_u).$$

The inequality in the first part of the statement follows from a similar argument. By definition of $\underline{E}^{\mathrm{D}}_{\mathscr{P}}$,

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u) = \inf\left\{E^{\mathrm{D}}_P(f \,|\, X_u = x_u) \colon P \in \mathscr{P}\right\}.$$

From this and Eq. (5.53)↶, it follows that

$$\begin{aligned}\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) &= \inf\Big\{\sup\{E^{\mathrm{D}}_P(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\} \colon P \in \mathscr{P}\Big\} \\ &\geq \sup\Big\{\inf\{E^{\mathrm{D}}_P(f_n \mid X_u = x_u) \colon P \in \mathscr{P}\} \colon n \in \mathbb{N}\Big\} \\ &= \sup\{\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\},\end{aligned}$$

where the inequality holds because we have changed the order of the supremum and the infimum (see Troffaes et al., 2014, Lemma 15.18). The inequality in the first part of the statement follows immediately from the preceding inequality and Eq. (5.54)↶:

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) \geq \sup\{\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\} = \lim_{n \to +\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \mid X_u = x_u). \qquad \square$$

Second, we consider uniformly bounded sequences of limit variables that converge point-wise.

**Theorem 5.32.** *Consider an imprecise jump process $\mathscr{P}$ that consists of countably additive jump processes. Fix some $\{X_u = x_u\}$ in $\mathscr{H}$ and $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$, and let $(f_n)_{\in\mathbb{N}}$ be a sequence of variables in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ that converges point-wise to $f$. If there is some $g$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ with $\overline{E}^{\mathrm{D}}_{\mathscr{P}}(g \mid X_u = x_u) < +\infty$ such that $|f_n| \leq g$ for all $n$ in $\mathbb{N}$, then*

$$\begin{aligned}\limsup_{n \to +\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \mid X_u = x_u) &\leq \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) \\ &\leq \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) \leq \liminf_{n \to +\infty} \overline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \mid X_u = x_u).\end{aligned}$$

*Proof.* As in the proof of Theorem 5.31240, it clearly suffices to prove the continuity properties for the conditional lower expectation $\underline{E}^{\mathrm{D}}_{\mathscr{P}}$, because they imply those of the conditional upper expectation $\overline{E}^{\mathrm{D}}_{\mathscr{P}}$ through the conjugacy relation of Eq. (5.29)238.

Observe that by assumption,

$$(\forall P \in \mathscr{P})\ E^{\mathrm{D}}_P(g \mid X_u = x_u) \leq \overline{E}^{\mathrm{D}}_{\mathscr{P}}(g \mid X_u = x_u) =: \beta < +\infty.$$

Because $-g \leq f_n \leq g$ by assumption, it follows from this, (DE6)226 and (DE4)225 that

$$(\forall n \in \mathbb{N})(\forall P \in \mathscr{P})\ -\beta \leq E^{\mathrm{D}}_P(f_n \mid X_u = x_u) \leq \beta. \tag{5.56}$$

Hence, $-\beta \leq \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \mid X_u = x_u) \leq \beta$ for all $n$ in $\mathbb{N}$.

Recall from Theorem 5.12227 that for all $n$ in $\mathbb{N}$, $f_n$ is a D-integrable variable – that is, belongs to $\mathbb{D}^{\mathrm{D}}_{P(\bullet \mid X_u = x_u)}$ – because $f_n$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ by assumption. Because furthermore by assumption $(f_n)_{n\in\mathbb{N}}$ converges point-wise to $f$ and $|f_n| \leq g$, it follows from Theorem 5.11226 that

$$(\forall P \in \mathscr{P})\ E^{\mathrm{D}}_P(f \mid X_u = x_u) = \lim_{n \to +\infty} E^{\mathrm{D}}_P(f_n \mid X_u = x_u).$$

From this equality and Eq. (5.56), we infer that $-\beta \leq E^{\mathrm{D}}_P(f \mid X_u = x_u) \leq \beta$ for all $P$ in $\mathscr{P}$, and therefore also $-\beta \leq \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) \leq \beta$.

Fix some $\epsilon$ in $\mathbb{R}_{>0}$. As $\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u)$ is real-valued, there is some $P$ in $\mathscr{P}$ such that

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) > E^{\mathrm{D}}_{P}(f \mid X_u = x_u) - \epsilon = \lim_{n \to +\infty} E^{\mathrm{D}}_{P}(f_n \mid X_u = x_u) - \epsilon.$$

Note that $E^{\mathrm{D}}_{P}(f_n \mid X_u = x_u) \geq \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \mid X_u = x_u)$ for all $n$ in $\mathbb{N}$, and therefore

$$\lim_{n \to +\infty} E^{\mathrm{D}}_{P}(f_n \mid X_u = x_u) = \limsup_{n \to +\infty} E^{\mathrm{D}}(f_n \mid X_u = x_u) \geq \limsup_{n \to +\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \mid X_u = x_u).$$

It follows immediately from the preceding to inequalities that

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) > \limsup_{n \to +\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \mid X_u = x_u) - \epsilon.$$

Because $\epsilon$ was an arbitrary positive real number, we conclude that

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) \geq \limsup_{n \to +\infty} \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f_n \mid X_u = x_u),$$

and this is what we set out to prove. □

# Computing lower expectations of idealised variables 6

Extending the domain of $\underline{E}_{\mathscr{P}}$ and $\overline{E}_{\mathscr{P}}$ to idealised variables is one thing, but computing the lower and upper expectation for a specific idealised variable is quite another. We have dealt with the former in the previous chapter, and we will deal with the latter in the present chapter. That is to say, here we will argue that we can compute the lower and upper expectation of four types of idealised variables: the number of jumps that we encountered in the previous chapter, but also until events, hitting times and Riemann integrals. By no means is that an exhaustive list of idealised variables that might pop up in practice, but they will go a long way. Crucial to our exposition is that these four types of idealised variables all have in common that they are the point-wise limit of a (monotone or uniformly bounded) sequence of simple variables, and that the lower and upper expectations of these simple variables converge to the lower and upper expectation of the idealised variable.

In Section 6.1, we prove this continuity for the number of jumps, and we generalise part of our argument to (some) 'generic' idealised variables. In Sections $6.2_{281}$ to $6.4_{295}$, we define three other important types of idealised variables – until events, hitting times and Riemann integrals – as point-wise limits of simple variables, and we show that the lower (or upper) expectation of these simple variables converges to the lower (or upper) expectation of the limit variable. Finally, for each of these three types of variables, we propose an intuitive method to (efficiently) compute their lower and upper expectations in Section $6.5_{310}$.

## 6.1 Establishing continuity

In the context of the present chapter, we will always assume that there is a non-empty and bounded set $\mathscr{Q}$ of rate operators such that every jump process $P$ in the jump process $\mathscr{P}$ is consistent with $\mathscr{Q}$ – meaning that $\mathscr{P}$ is contained in $\mathbb{P}_{\mathscr{Q}}$. Our reason for doing so is twofold. First, Corollary $5.30_{239}$ then ensures that every jump process $P$ in $\mathscr{P}$ is countably additive. Second, it then follows from Corollary 5.18<sub>230</sub> that every jump process $P$ in $\mathscr{P}$ has uniformly bounded rate with rate bound $\|\mathscr{Q}\|_{\text{op}}$, and it is this rate bound that essentially allows us to prove the desired continuity properties.

In Section 6.1.1, we will discover how in the particular case of $\eta_{[s,r]}$, the number of jumps in $[s,r]$, we can use this rate bound $\|\mathscr{Q}\|_{\text{op}}$ to establish that the lower (upper) expectation of $\eta_v$ for some grid $v$ over $[s,r]$ converges to the lower (upper) expectation of $\eta_{[s,r]}$ as the grid width $\Delta(v)$ vanishes. We then use these results to establish some important intermediary results in Section 6.1.2[278]; it is these results that we will use in Sections 6.2[281] to 6.4[295] to prove the desired continuity properties for the other three types of idealised variables.

### 6.1.1 The particular case of the number of jumps

Consider an imprecise jump process $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$, a state history $\{X_u = x_u\}$ in $\mathscr{H}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$, and let $(v_n)_{n\in\mathbb{N}}$ be a sequence of grids over $[s,r]$ such that $\lim_{n\to+\infty} \Delta(v_n) = 0$. Then by Theorem 5.26[236], the idealised variable $\eta_{[s,r]}$ is the point-wise limit of the sequence $(\eta_{v_n})_{n\in\mathbb{N}}$ of $\mathscr{F}_u$-simple variables. In order to prove that the lower and upper expectation of $\eta_{v_n}$ converge to the lower and upper expectation of $\eta_{[s,r]}$, it suffices to find an upper bound, say $\epsilon_n$, on

$$\left|E_P^{\mathrm{D}}(\eta_{[s,r]} \,|\, X_u = x_u) - E_P(\eta_{v_n} \,|\, X_u = x_u)\right|$$

that holds for any jump process $P$ in $\mathscr{P}$ and that vanishes as $n$ recedes to $+\infty$. The reason for this is the following important intermediary result.

**Lemma 6.1.** *Consider an imprecise jump process $\mathscr{P}$ that consists of countably additive jump processes. Fix some $\{X_u = x_u\}$ in $\mathscr{H}$, $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ and $g$ in $\mathbb{S}(\mathscr{F}_u)$. If $\epsilon$ is a non-negative real number such that*

$$(\forall P \in \mathscr{P})\ \left|E_P^{\mathrm{D}}(f \,|\, X_u = x_u) - E_P(g \,|\, X_u = x_u)\right| \leq \epsilon,$$

*then*

$$\left|\underline{E}_{\mathscr{P}}^{\mathrm{D}}(f \,|\, X_u = x_u) - \underline{E}_{\mathscr{P}}(g \,|\, X_u = x_u)\right| \leq \epsilon$$

*and*

$$\left|\overline{E}_{\mathscr{P}}^{\mathrm{D}}(f \,|\, X_u = x_u) - \overline{E}_{\mathscr{P}}(g \,|\, X_u = x_u)\right| \leq \epsilon.$$

*Proof.* Here we only prove the inequality of the statement for the lower expectation. The inequality for the upper expectation can be proven in an analoguous manner, but also follows from the one for the lower expectation due to conjugacy.

Note that for all $P$ in $\mathscr{P}$, $\min g \leq E_P(g \,|\, X_u = x_u) \leq \max g$ due to (ES1)[37]. From this, we infer that $\min g \leq \underline{E}_{\mathscr{P}}(g \,|\, X_u = x_u) \leq \max g$; in other words, $\underline{E}_{\mathscr{P}}(g \,|\, X_u = x_u)$

is real valued. Observe that

$$\begin{aligned}\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) = \inf\{E^{\mathrm{D}}_P(f \mid X_u = x_u) \colon P \in \mathscr{P}\} &\leq \inf\{E_P(g \mid X_u = x_u) + \epsilon \colon P \in \mathscr{P}\} \\ &= \inf\{E_P(g \mid X_u = x_u) \colon P \in \mathscr{P}\} + \epsilon \\ &= \underline{E}_{\mathscr{P}}(g \mid X_u = x_u) + \epsilon,\end{aligned}$$

where for the inequality we have used the assumption of the statement. Similarly, we find that

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) \geq \underline{E}_{\mathscr{P}}(g \mid X_u = x_u) - \epsilon.$$

The inequality of the statement follows immediately from these two inequalities because $\underline{E}_{\mathscr{P}}(g \mid X_u = x_u)$ is real valued. □

Thus, we need to determine an upper bound $\epsilon_n$ on the difference between the Daniell expectation of $\eta_{[s,r]}$ and the expectation of $\eta_{v_n}$ that holds for any jump process $P$ in $\mathscr{P}$. The following result establishes such an upper bound. It follows from Theorem 5.27236, but the proof is rather long; for this reason, we have relegated it to Appendix 6.A321.

**Proposition 6.2.** *Consider a jump process $P$ that has uniformly bounded rate, with rate bound $\lambda$. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$ and a grid $v = (t_0, \ldots, t_n)$ over $[s, r]$. Then $\eta_{[s,r]} - \eta_v$ is a non-negative $\mathscr{F}_u$-over variable, and*

$$\begin{aligned}E^{\mathrm{D}}_P(\eta_{[s,r]} - \eta_v \mid X_u = x_u) &= E^{\mathrm{D}}_P(\eta_{[s,r]} \mid X_u = x_u) - E_P(\eta_v \mid X_u = x_u) \\ &\leq \frac{1}{4}\Delta(v)(r-s)\lambda^2.\end{aligned}$$

It is now obvious that Corollary 5.18230, Lemma 6.1↶ and Proposition 6.2 imply the following result, which establishes the upper bound $\epsilon_n$ that we are after. Generally speaking, this result prescribes how fine the grid $v$ over $[s, r]$ should be in order for the lower and upper expectation of $\eta_v$ to be $\epsilon$-close to those of $\eta_{[s,r]}$.

**Proposition 6.3.** *Consider a non-empty and bounded set $\mathscr{Q}$ of rate operators and an imprecise jump process $\mathscr{P}$ such that $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$. Fix some $\{X_u = x_u\}$ in $\mathscr{H}$ and $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Then for any grid $v$ over $[s, r]$,*

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(\eta_{[s,r]} \mid X_u = x_u) - \underline{E}_{\mathscr{P}}(\eta_v \mid X_u = x_u)\right| \leq \frac{1}{4}\Delta(v)(r-s)\|\mathscr{Q}\|^2_{\mathrm{op}}$$

*and*

$$\left|\overline{E}^{\mathrm{D}}_{\mathscr{P}}(\eta_{[s,r]} \mid X_u = x_u) - \overline{E}_{\mathscr{P}}(\eta_v \mid X_u = x_u)\right| \leq \frac{1}{4}\Delta(v)(r-s)\|\mathscr{Q}\|^2_{\mathrm{op}}.$$

*In particular, this holds for $\mathscr{P} = \mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathscr{P} = \mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathscr{P} = \mathbb{P}_{\mathscr{M},\mathscr{Q}}$, with $\mathscr{M}$ a non-empty set of initial mass functions.*

*Proof.* Recall from Corollary 5.18<sub>230</sub> that every jump process $P$ in $\mathscr{P}$ has uniformly bounded rate with rate bound $\|\mathcal{Q}\|_{\mathrm{op}}$. Thus, it follows from Proposition 6.2 that for all $P$ in $\mathscr{P}$,

$$0 \leq E^{\mathrm{D}}_P(\eta_{[s,r]} - \eta_\upsilon \,|\, X_u = x_u) = E^{\mathrm{D}}_P(\eta_{[s,r]} \,|\, X_u = x_u) - E_P(\eta_\upsilon \,|\, X_u = x_u) \\ \leq \frac{1}{4}\Delta(\upsilon)(r-s)\|\mathcal{Q}\|^2_{\mathrm{op}},$$

where for the first inequality we used (DE6)<sub>226</sub> because $\eta_{[s,r]} - \eta_\upsilon$ is non-negative. Because these inequalities hold for all $P$ in $\mathscr{P}$, they imply the inequalities of the statement due to Lemma 6.1<sub>274</sub>. □

Because we have fixed the sequence of grids $(\upsilon_n)_{n\in\mathbb{N}}$ over $[s,r]$ such that $\lim_{n\to+\infty}\Delta(\upsilon) = 0$, it follows almost immediately from Proposition 6.3 that the lower and upper expectation of $\eta_{\upsilon_n}$ converge to the lower and upper Daniell expectation of $\eta_{[s,r]}$, respectively.

**Corollary 6.4.** *Consider a non-empty and bounded set $\mathcal{Q}$ of rate operators and an imprecise jump process $\mathscr{P}$ such that $\mathscr{P} \subseteq \mathbb{P}_{\mathcal{Q}}$. Fix some $\{X_u = x_u\}$ in $\mathscr{H}$ and $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Then for any sequence $(\upsilon_n)_{n\in\mathbb{N}}$ of grids over $[s,r]$ with $\lim_{n\to+\infty}\Delta(\upsilon_n) = 0$,*

$$\lim_{n\to+\infty} \underline{E}_{\mathscr{P}}(\eta_{\upsilon_n} \,|\, X_u = x_u) = \underline{E}^{\mathrm{D}}_{\mathscr{P}}(\eta_{[s,r]} \,|\, X_u = x_u)$$

*and*

$$\lim_{n\to+\infty} \overline{E}_{\mathscr{P}}(\eta_{\upsilon_n} \,|\, X_u = x_u) = \overline{E}^{\mathrm{D}}_{\mathscr{P}}(\eta_{[s,r]} \,|\, X_u = x_u).$$

*In particular, this holds for $\mathscr{P} = \mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathcal{Q}}$, $\mathscr{P} = \mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathcal{Q}}$ and $\mathscr{P} = \mathbb{P}_{\mathscr{M},\mathcal{Q}}$, with $\mathscr{M}$ a non-empty set of initial mass functions.*

*Proof.* Follows immediately from Proposition 6.3. □

Note that Corollary 6.4 is an example where the limits in Theorem 5.32<sub>240</sub> provide tight bounds. It might not be immediately obvious that Theorem 5.32<sub>240</sub> is applicable here, but we can invoke it with $g = \eta_{[s,r]}$ because the upper expectation of $\eta_{[s,r]}$ is different from $+\infty$ due to Theorem 5.27<sub>236</sub>. If we additionally assume that $\upsilon_n \subseteq \upsilon_{n+1}$ for all $n$ in $\mathbb{N}$, then $(\eta_{\upsilon_n})_{n\in\mathbb{N}}$ is non-decreasing due to Theorem 5.26<sub>236</sub>; in this case, Corollary 6.4 is an example where (one of) the inequalities in Corollary 5.33<sub>240</sub> hold(s) with equality!

To conclude our discussion of the continuity for the expected number of jumps, we highlight a second important consequence of Proposition 6.3. Due to this result, we can compute the lower and upper expected number of jumps up to arbitrary precision using Algorithm 4.3<sub>171</sub>, albeit only for $\mathbb{P}_{\mathscr{M},\mathcal{Q}}$ in case $\mathcal{Q}$ is convex and has separately specified rows.[1] Let us illustrate this with our running example.

[1] It is possible to establish a much more efficient computation method in the spirit of the methods in Section 6.5<sub>310</sub> that also works for $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathcal{Q}}$ and that does not require $\mathcal{Q}$ to be convex.

**Joseph's Example 6.5.** Recall from Joseph's Example 4.13171 that Eleanor's beliefs about Joseph's machine are accurately modelled by $\underline{E}_{\mathscr{M},\mathscr{Q}_2}$, with $\mathscr{M} = \{\mathbb{I}_{\mathrm{H}}\}$ as defined in Joseph's Example 4.3161 and

$$\mathscr{Q}_2 = \left\{ \begin{pmatrix} -\lambda_{\mathrm{H}} & \lambda_{\mathrm{H}} \\ \lambda_{\mathrm{T}} & -\lambda_{\mathrm{T}} \end{pmatrix} : \lambda_{\mathrm{H}}, \lambda_{\mathrm{T}} \in [\underline{\lambda}, \overline{\lambda}] \right\}$$

as defined in Joseph's Example 4.4163. Note that $\mathscr{Q}_2$ is non-empty, bounded and convex and has separately specified rows.

Suppose Eleanor is interested in the expected number of times that Joseph's machine jumps – that is, that its display changes from heads to tails or from tails to heads – over the first $r$ time units, with $r$ a positive real number. That is, she is interested in

$$\underline{E}_{\mathscr{M},\mathscr{Q}_2}(\eta_{[0,r]}) \quad \text{and} \quad \overline{E}_{\mathscr{M},\mathscr{Q}_2}(\eta_{[0,r]}).$$

Say we want to approximate the lower and upper expected number of jumps with a maximum error of $\epsilon$, with $\epsilon$ a positive real number. Then by Proposition 6.3275, we need to construct a grid $v = (t_0, \dots, t_n)$ over $[0, r]$ such that

$$\frac{1}{4}\Delta(v) r \|\mathscr{Q}_2\|_{\mathrm{op}}^2 \leq \epsilon.$$

The most obvious way to do so, is to divide the interval $[0, r]$ into subintervals of equal length. From the preceding inequality, it follows immediately that it suffices to use

$$n := \left\lceil \frac{r^2 \|\mathscr{Q}_2\|_{\mathrm{op}}^2}{4\epsilon} \right\rceil \tag{6.1}$$

subintervals to attain the desired accuracy; that is, with $v := (t_0, \dots, t_n)$ the grid over $[0, r]$ such that $t_k := {}^{kr}/_{n}$ for all $k$ in $\{0, \dots, n\}$,

$$\left|\underline{E}_{\mathscr{M},\mathscr{Q}_2}(\eta_{[0,r]}) - \underline{E}_{\mathscr{M},\mathscr{Q}_2}(\eta_v)\right| \leq \epsilon \quad \text{and} \quad \left|\overline{E}_{\mathscr{M},\mathscr{Q}_2}(\eta_{[0,r]}) - \overline{E}_{\mathscr{M},\mathscr{Q}_2}(\eta_v)\right| \leq \epsilon.$$

It follows from Eq. (5.22)233 that $\eta_v = f(X_v)$, where $f$ is the gamble on $\mathscr{X}_v$ defined for all $x_v$ in $\mathscr{X}_v$ by

$$f(x_v) := \sum_{k=1}^{n} \mathbb{I}_A(x_{t_{k-1}}, x_{t_k}), \quad \text{with } A := \mathscr{X}^2 \setminus \{(x, x) : x \in \mathscr{X}\}.$$

Hence, and because $\mathscr{Q}_2$ has separately specified rows and is convex, we can use Algorithm 4.3171 to compute $\underline{E}_{\mathscr{M},\mathscr{Q}_2}(\eta_v)$; due to conjugacy, we can also use Algorithm 4.3171 to compute the conjugate upper expectation.

Proving the validity of this method would require us to establish an 'approximate second-order' sum-product law of iterated lower expectations – that is, the continuous-time counterpart of Definition 2 in (De Bock et al., 2021) – and this would lead us too far. We refer to Section 5.2 in (Erreygers & De Bock, 2021) for some preliminary related results (stated there without proof), and to Section 5.2 in (De Bock et al., 2021) for some results for the discrete-time setting (stated there with proof).

Let us do so for some numerical values. As in Joseph's Example 4.21$_{181}$, we take $\underline{\lambda} := 1$ and $\overline{\lambda} := 3/2$; then

$$\|\mathcal{Q}_2\|_{\mathrm{op}} = \sup\big\{\|Q\|_{\mathrm{op}} \colon Q \in \mathcal{Q}_2\big\} = \sup\big\{2\max\{\lambda_{\mathtt{H}}, \lambda_{\mathtt{T}}\} \colon \lambda_{\mathtt{H}}, \lambda_{\mathtt{T}} \in [\underline{\lambda}, \overline{\lambda}]\big\} = 2\overline{\lambda} = 3,$$

where the first equality is the definition of $\|\mathcal{Q}_2\|_{\mathrm{op}}$ and the second equality holds due to (R5)$_{81}$. Furthermore, we choose the time point $r := 1$ and tolerance $\epsilon := 1 \cdot 10^{-4}$. By Eq. (6.1)$_{\frown}$, $n = 22\,500$ subintervals suffice to obtain the desired maximal error. Executing Algorithm 4.3$_{171}$ for these parameter values, and using Eq. (3.75)$_{115}$ to determine $e^{\Delta \underline{Q}}$, we find that

$$\underline{E}_{\mathscr{M},\mathcal{Q}_2}(\eta_{[0,1]}) \approx 0.9999 \quad \text{and} \quad \overline{E}_{\mathscr{M},\mathcal{Q}_2}(\eta_{[0,1]}) \approx 1.500,$$

up to four significant digits.

Remarkably, the lower and upper expected number of jumps in $[0, r]$ are (approximately) equal to $r\underline{\lambda}$ and $r\overline{\lambda}$, respectively. This is interesting for several reasons.

First, one can show – but we will not do this here – that $r\underline{\lambda}$ is the expected number of jumps in $[0, r]$ for the homogeneous Markovian jump process $P_{\mathbb{I}_{\mathtt{H}},Q}$ in $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathcal{Q}_2}$ with $Q(\mathtt{H},\mathtt{T}) = \underline{\lambda} = Q(\mathtt{T},\mathtt{H})$, and similarly, $r\overline{\lambda}$ is the expected number of jumps in $[0, r]$ for the homogeneous Markovian jump process $P_{\mathbb{I}_{\mathtt{H}},Q}$ in $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathcal{Q}_2}$ with $Q(\mathtt{H},\mathtt{T}) = \overline{\lambda} = Q(\mathtt{T},\mathtt{H})$; so in this particular case, the lower and upper expectations for $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathcal{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathcal{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathcal{Q}}$ coincide.

A second reason is that this shows that the parameters $\underline{\lambda}$ and $\overline{\lambda}$ are lower and upper bounds on the expected number of decay events per time unit. This should come as no surprise though, as it is precisely (the precise version of) this fact that we implicitly used in Joseph's Example 3.38$_{87}$ to construct a homogeneous Markovian jump process to model Joseph's machine.

Thirdly, we know from Theorem 5.27$_{236}$ that the (upper) expected number of jumps in $[0, r]$ is bounded above by $r\|\mathcal{Q}\|_{\mathrm{op}}/2 = r\overline{\lambda}$. Thus, in this case, this upper bound is reached by the upper expectation. This is a coincidence though, and this will *not* be the case in general! ♪

### 6.1.2 The general case

The approach that we used to establish continuity for the number of jumps is, to some extent, also applicable to other limit variables, and this is precisely what we do in Sections 6.2$_{281}$ to 6.4$_{295}$ further on. In each of the three cases that we consider, we consider an idealised variable $f$ that depends on the state of the system at the time points in some time horizon $[s, r]$, and we approximate this variable with a simple variable $g(X_v)$, with $v$ a grid over $[s, r]$. As in Proposition 6.2$_{275}$, we then need to bound the difference between the Daniell expectation of $f$ and the expectation of $g(X_v)$; interestingly, we will always use the number of jumps $\eta_{[s,r]}$ or the 'number of extra jumps' $\eta_{[s,r]} - \eta_v$ to do so. For that reason, the following generalisations of Propositions 6.2$_{275}$

and 6.3[275] will play a critical role in the remainder; the proof of the first result can be found in Appendix 6.B[325].

**Lemma 6.6.** *Consider a jump process $P$ that has uniformly bounded rate, with rate bound $\lambda$. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that* $\max u \leq s \leq r$, *a grid $v$ over $[s, r]$, a limit variable $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ and an $\mathscr{F}_u$-simple variable $g$. If there are non-negative real numbers $\alpha, \beta, \gamma$ such that*

$$|f - g| \leq \alpha\Delta(v) + \beta\big(\eta_{[s,r]} - \eta_v\big) + \gamma\Delta(v)\eta_{[s,r]},$$

*then*

$$\Big|E_P^{\mathrm{D}}(f \,|\, X_u = x_u) - E_P(g \,|\, X_u = x_u)\Big| \leq \Delta(v)\bigg(\alpha + \frac{1}{4}\beta(r-s)\lambda^2 + \frac{1}{2}\gamma(r-s)\lambda\bigg).$$

**Proposition 6.7.** *Consider a non-empty and bounded set $\mathscr{Q}$ of rate operators and an imprecise jump process $\mathscr{P}$ such that $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that* $\max u \leq s \leq r$, *a grid $v$ over $[s, r]$, a limit variable $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ and an $\mathscr{F}_u$-simple variable $g$. If there are non-negative real numbers $\alpha, \beta, \gamma$ such that*

$$|f - g| \leq \alpha\Delta(v) + \beta\big(\eta_{[s,r]} - \eta_v\big) + \gamma\Delta(v)\eta_{[s,r]},$$

*then*

$$\begin{aligned}&\Big|\underline{E}_{\mathscr{P}}^{\mathrm{D}}(f \,|\, X_u = x_u) - \underline{E}_{\mathscr{P}}(g \,|\, X_u = x_u)\Big| \\ &\qquad\qquad \leq \Delta(v)\bigg(\alpha + \frac{1}{4}\beta(r-s)\|\mathscr{Q}\|_{\mathrm{op}}^2 + \frac{1}{2}\gamma(r-s)\|\mathscr{Q}\|_{\mathrm{op}}\bigg),\end{aligned}$$

*and the same upper bound holds for* $\Big|\overline{E}_{\mathscr{P}}^{\mathrm{D}}(f \,|\, X_u = x_u) - \overline{E}_{\mathscr{P}}(g \,|\, X_u = x_u)\Big|$.

*Proof.* Recall from Corollary 5.18[230] that any jump process $P$ in $\mathscr{P}$ has uniformly bounded rate with rate bound $\|\mathscr{Q}\|_{\mathrm{op}}$. Therefore, it follows from Lemma 6.6 that

$$\begin{aligned}(\forall P \in \mathscr{P})\ &\Big|E_P^{\mathrm{D}}(f \,|\, X_u = x_u) - E_P(g \,|\, X_u = x_u)\Big| \\ &\qquad \leq \Delta(v)\bigg(\alpha + \frac{1}{4}\beta(r-s)\|\mathscr{Q}\|_{\mathrm{op}}^2 + \frac{1}{2}\gamma(r-s)\|\mathscr{Q}\|_{\mathrm{op}}\bigg).\end{aligned}$$

The statement follows immediately from this due to Lemma 6.1[274]. □

To establish that the idealised variable $f$ and its approximating simple variable $g(X_v)$ satisfy the condition in Proposition 6.7, the following two intermediary results can come in handy. For example, we will use Lemmas 6.8 and 6.9↷ in the proofs of Lemmas 6.11[282] and 6.20[292] further on.

**Lemma 6.8.** *Consider some time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s < r$. Then for any grid $v$ over $[s, r]$,*

$$\frac{1}{2}\big(\eta_{[s,r]} - \eta_v\big) \geq \mathbb{I}_A \qquad \textit{with } A := \big\{\omega \in \Omega \colon \eta_v(\omega) < \eta_{[s,r]}(\omega)\big\}.$$

*Proof.* Let $(v_\ell)_{\ell\in\mathbb{N}}$ be the sequence of grids as constructed in the proof of Proposition $6.2_{275}$ – see Appendix $6.A_{321}$. Furthermore, for all $\ell$ in $\mathbb{N}$, we let

$$A_\ell := \{\omega \in \Omega\colon \eta_v(\omega) < \eta_{v_\ell}(\omega)\}.$$

Fix some $\ell$ in $\mathbb{N}$. Then because $v_\ell \supseteq v$ by construction, it follows from Lemma $5.23_{234}$ that for any $\omega$ in $\Omega$, $\eta_{v_\ell}(\omega) - \eta_v(\omega)$ is either equal to 0 or greater than or equal to 2. Consequently,

$$\mathbb{I}_{A_\ell} = (\eta_{v_\ell} - \eta_v) \wedge 1 \le \frac{1}{2}(\eta_{v_\ell} - \eta_v). \tag{6.2}$$

Recall from the proof of Proposition $6.2_{275}$ that the sequence $(\eta_{v_\ell} - \eta_v)_{\ell\in\mathbb{N}}$ converges point-wise to $\eta_{[s,r]} - \eta_v$. It follows immediately from this and Eq. (6.2) that $\mathbb{I}_{A_\ell}$ converges point-wise to $\mathbb{I}_A$, and that

$$\mathbb{I}_A = \operatorname*{p\text{-}w\,lim}_{\ell\to+\infty} \mathbb{I}_{A_\ell} \le \operatorname*{p\text{-}w\,lim}_{\ell\to+\infty} \frac{1}{2}(\eta_{v_\ell} - \eta_v) = \frac{1}{2}(\eta_{[s,r]} - \eta_v),$$

as required. □

**Lemma 6.9.** *Consider some time points $s, r$ in $\mathbb{R}_{\ge 0}$ such that $s \le r$ and a grid $v = (t_0, \dots, t_n)$ over $[s,r]$. Fix some $\omega$ in $\Omega$. If there is some $t$ in $[s,r]$ and some $k$ in $\{1,\dots,n\}$ such that $t_{k-1} < t < t_k$ and $\omega(t_{k-1}) \ne \omega(t) \ne \omega(t_k)$, then $\eta_v(\omega) < \eta_{[s,r]}(\omega)$.*

*Proof.* The statement is trivial whenever $s = r$, so we may assume without loss of generality that $s < t$. It follows from repeated application of Lemma $5.22_{234}$ that

$$\eta_v(\omega) = \sum_{\ell=1}^{n} \eta_{(t_{\ell-1},t_\ell)}(\omega).$$

Observe that the sets $\mathscr{J}_{[t_0,t_1]}(\omega), \dots, \mathscr{J}_{[t_{n-1},t_n]}(\omega)$ of jump times for the subintervals of $v$ are pair-wise disjoint. Hence, it follows from Eq. $(5.18)_{232}$ that

$$\eta_{[s,r]}(\omega) = \sum_{\ell=1}^{n} \eta_{[t_{\ell-1},t_\ell]}(\omega).$$

Thus, we see that

$$\eta_{[s,r]}(\omega) - \eta_v(\omega) = \sum_{\ell=1}^{n} \eta_{[t_{\ell-1},t_\ell]}(\omega) - \eta_{(t_{\ell-1},t_\ell)}(\omega).$$

Recall from Proposition $6.2_{275}$ that for all $\ell$ in $\{1,\dots,n\}$, $\eta_{[t_{\ell-1},t_\ell]}(\omega) - \eta_{(t_{\ell-1},t_\ell)}(\omega) \ge 0$. Observe furthermore that $\eta_{[t_{k-1},t_k]} - \eta_{(t_{k-1},t_k)} > 0$ due to the assumption of the statement. From all this we conclude that $\eta_{[s,r]}(\omega) - \eta_v(\omega) > 0$, as required. □

## 6.2 Until events

The first type of limit variables that we will consider are indicators of 'until events'; as we will discover in Section 6.2.3$_{287}$ further on, these include indicators of hitting events. In Joseph's Example 5.1$_{215}$, we somewhat informally defined a hitting event corresponding to a subset $G$ of $\mathscr{X}$ as the event that the path $\omega$ ever 'hits', or belongs to, $G$ at some future time point. An 'until event' is similar, although it concerns two subsets of $\mathscr{X}$ instead of one: a set $G$ of 'goal' states and a set $S$ of 'safe' states. These until events play an important role in the setting of model checking for jump processes (see Baier et al., 2003, 2008; Katoen et al., 2012). Formally, for any two subsets $S, G$ of $\mathscr{X}$ and any subset $\mathscr{R}$ of $\mathbb{R}_{\geq 0}$, Baier et al. (2003, Definition 6) and Katoen et al. (2012, Section 2.2) consider the event

$$H_{\mathscr{R}}^{S,G} := \left\{\omega \in \Omega \colon (\exists t \in \mathscr{R})\ \omega(t) \in G \text{ and } (\forall s \in \mathscr{R}, s < t)\ \omega(s) \in S\right\} \tag{6.3}$$

that the state of the system is in the set $G$ of goal states at some time point $t$ in $\mathscr{R}$, while it is in the set $S$ of safe states at all time points in $\mathscr{R}$ that precede $t$. To simplify our notation, we will denote the indicator of $H_{\mathscr{R}}^{S,G}$ by $h_{\mathscr{R}}^{S,G} := \mathbb{I}_{H_{\mathscr{R}}^{S,G}}$.

We focus on three types of subsets $\mathscr{R}$ of $\mathbb{R}_{\geq 0}$:

1. sequences of time points of the form $v = (t_0, \ldots, t_n)$, in which case we call $H_v^{S,G}$ an *approximating until event*;
2. bounded time intervals of the form $[s, r]$, in which case we call $H_{[s,r]}^{S,G}$ a *time-bounded until event*;
3. unbounded time intervals of the form $[s, +\infty[$, in which case we call $H_{[s,+\infty[}^{S,G}$ an (unbounded) *until event*.

It will prove to be the case that (the indicators of) each of these events belong to our domain: the indicator of an approximating until event is a simple variable, while the indicator of a time-bounded until event is the point-wise limit of the indicators of a sequence of approximating until events, as is the indicator of an until event.

### 6.2.1 From approximating to time-bounded until events

First things first, we establish that an approximating until event is a finitary event. We do so through a bit of a detour: we actually establish that its indicator is a simple variable because it has a sum-product representation.

**Lemma 6.10.** *Consider subsets $S, G$ of $\mathscr{X}$, a sequence of time points $u$ in $\mathscr{U}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$, and fix some grid $v = (t_0, \ldots, t_n)$ over $[s, r]$. Then*

$$h_v^{S,G} = \sum_{k=0}^{n} \mathbb{I}_G(X_{t_k}) \prod_{\ell=0}^{k-1} \mathbb{I}_{S \setminus G}(X_{t_\ell}),$$

*so $h_v^{S,G}$ has a sum-product representation over $v$. Consequently, $h_v^{S,G}$ is $\mathscr{F}_u$-simple, so $H_v^{A,B}$ belongs to $\mathscr{F}_u$.*

*Proof.* That $h_v^{S,G}$ has the sum-product representation over $v$ of the statement is a matter of straightforward verification. Observe that $v \succcurlyeq u$ because, by assumption, $v$ is a grid over $[s,r]$ and $\max u \leq s$. Therefore, it follows from Lemma 4.6$_{164}$ that $h_v^{S,G}$ is an $\mathscr{F}_u$-simple variable. That $H_v^{S,G}$ belongs to $\mathscr{F}_u$ follows from the preceding and Lemma C.4$_{463}$. □

Next, we set out to approximate the time-bounded until event $H_{[s,r]}^{S,G}$ with a sequence $(H_{v_n}^{S,G})_{n\in\mathbb{N}}$ of approximating until events that corresponds to a sequence $(v_n)_{n\in\mathbb{N}}$ of grids over $[s,r]$ with vanishing grid width. To do so, we determine an upper bound on the difference between the indicators of $H_{[s,r]}^{S,R}$ and $H_{v_n}^{S,G}$.

**Lemma 6.11.** *Consider subsets $S,G$ of $\mathscr{X}$ and time points $s,r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. Then for any grid $v$ over $[s,r]$,*

$$\left|h_{[s,r]}^{S,G} - h_v^{S,G}\right| \leq \frac{1}{2}(\eta_{[s,r]} - \eta_v).$$

*Proof fo Lemma 6.11.* We enumerate the time points in $v$ as $(t_0,\ldots,t_n)$ and fix any path $\omega$ in $\Omega$. Because $h_{[s,r]}^{S,G}$ and $h_v^{S,G}$ are indicators, it is clear that $h_\omega := \left|h_{[s,r]}^{S,G}(\omega) - h_v^{S,G}(\omega)\right|$ can only assume one of two values: 0 or 1. We will now show that

$$h_\omega = \left|h_{[s,r]}^{S,G}(\omega) - h_v^{S,G}(\omega)\right| \leq \mathbb{I}_A(\omega) \quad \text{with } A := \left\{\omega' \in \Omega \colon \eta_v(\omega') < \eta_{[s,r]}(\omega')\right\}. \tag{6.4}$$

If $h_\omega = 0$, then this inequality holds trivially. The alternative case $h_\omega = 1$ occurs if and only if $\omega$ belongs to precisely one of the two events $H_{[s,r]}^{S,G}$ and $H_v^{S,G}$.

First, we assume that $\omega$ belongs to $H_{[s,r]}^{S,G} \setminus H_v^{S,G}$. Then there is a time point $t$ in $[s,r]$ such that

$$\omega(t) \in G \quad \text{and} \quad \left(\forall t' \in [s,r], t' < t\right) \omega(t') \in S \setminus G.$$

Note that $t$ cannot belong to $v$ because otherwise $\omega$ would belong to $H_v^{S,G}$; hence, there is a $k$ in $\{1,\ldots,n\}$ such that $t_{k-1} < t < t_k$. Note that $\omega(t_{k-1}) \in S \setminus G$ and $\omega(t) \in G$. Furthermore, $\omega(t_k) \in G^{\mathrm{c}}$, because otherwise $\omega$ would again belong to $H_v^{S,G}$. Hence, we see that $\omega(t_{k-1}) \neq \omega(t) \neq \omega(t_k)$; it follows from this and Lemma 6.9$_{280}$ that $\omega$ belongs to $A$, so $h_\omega = 1 \leq \mathbb{I}_A(\omega)$.

Second, we assume that $\omega$ belongs to $H_v^{S,G} \setminus H_{[s,r]}^{S,G}$. Then there is an index $k$ in $\{0,\ldots,n\}$ such that

$$\omega(t_k) \in G \quad \text{and} \quad \left(\forall i \in \{0,\ldots,k-1\}\right) \omega(t_i) \in S \setminus G.$$

Note that $k \neq 0$, because otherwise $\omega$ would belong to $H_{[s,r]}^{S,G}$. For the same reason, there is at least one time point $t$ in $[t_0,t_k] \setminus v$ such that $\omega(t) \in S^{\mathrm{c}} \cap G^{\mathrm{c}}$. Let $\ell$ be the element of $\{1,\ldots,k\}$ such that $t_{\ell-1} < t < t_\ell$. Then $\omega(t_{\ell-1}) \neq \omega(t) \neq \omega(t_\ell)$ because

$\omega(t_{\ell-1}) \in S \setminus G$, $\omega(t) \in S^c \cap G^c$ and $\omega(t_\ell) \in S \cup G$. Consequently, $\omega$ belongs to $A$ due to Lemma 6.9280, so $h_\omega = 1 \leq \mathbb{I}_A(\omega)$.

Thus, we have shown that Eq. (6.4) holds for all $\omega$ in $\Omega$. The inequality of the statement follows immediately from Eq. (6.4) and Lemma 6.8279. □

It is now a piece of cake to show that $h^{S,G}_{[s,r]}$ is a limit variable, because this follows almost immediately from Lemmas 5.25235, 6.10281 and 6.11.

**Lemma 6.12.** *Consider subsets $S, G$ of $\mathscr{X}$, a sequence of time points $u$ in $\mathscr{U}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Let $(v_n)_{n\in\mathbb{N}}$ be a sequence of grids over $[s,r]$ with $\lim_{n\to+\infty}\Delta(v_n) = 0$. Then $\big(h^{S,G}_{v_n}\big)_{n\in\mathbb{N}}$ is a sequence of $\mathscr{F}_u$-simple variables that is uniformly bounded and that converges point-wise to $h^{S,G}_{[s,r]}$, so $h^{S,G}_{[s,r]}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$.*

*Proof.* For all $n$ in $\mathbb{N}$, $v_n$ is a grid over $[s,r]$ and it therefore follows from Lemma 6.10281 that $h^{S,G}_{v_n}$ is an $\mathscr{F}_u$-simple variable. Thus, $\big(h^{S,G}_{v_n}\big)_{n\in\mathbb{N}}$ is a sequence of $\mathscr{F}_u$-simple variables, and it obvious that this sequence is uniformly bounded. Hence, it remains for us to prove that for all $\omega$ in $\Omega$,

$$\lim_{n\to+\infty} h^{S,G}_{v_n}(\omega) = h^{S,G}_{[s,r]}(\omega). \tag{6.5}$$

To this end, we fix some $\omega$ in $\Omega$. Recall from Lemma 6.11 that for all $n$ in $\mathbb{N}$,

$$\big|h^{S,G}_{[s,r]}(\omega) - h^{S,G}_{v_n}(\omega)\big| \leq \frac{1}{2}\big(\eta_{[s,r]}(\omega) - \eta_{v_n}(\omega)\big).$$

Furthermore, because $\lim_{n\to+\infty}\Delta(v_n) = 0$ by the assumptions of the statement, it follows from Lemma 5.25235 that $\lim_{n\to+\infty}\eta_{v_n}(\omega) = \eta_{[s,r]}(\omega)$. Hence,

$$\lim_{n\to+\infty}\big|h^{S,G}_{[s,r]}(\omega) - h^{S,G}_{v_n}(\omega)\big| \leq \lim_{n\to+\infty}\frac{1}{2}\big(\eta_{[s,r]}(\omega) - \eta_{v_n}(\omega)\big) = 0,$$

which implies Eq. (6.5). □

The upper bound on the difference between the indicators $h^{S,G}_{[s,r]}$ and $h^{S,G}_{v}$ in Lemma 6.11 also gives rise to an upper bound on the difference between the lower and upper probabilities of the corresponding events.

**Proposition 6.13.** *Consider a non-empty and bounded set $\mathscr{Q}$ of rate operators, and an imprecise jump process $\mathscr{P}$ such that $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$. Fix some subsets $S, G$ of $\mathscr{X}$, a state history $\{X_u = x_u\}$ in $\mathscr{H}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Then for any grid $v$ over $[s,r]$,*

$$\big|\underline{P}^{\mathrm{D}}_{\mathscr{P}}\big(H^{S,G}_{[s,r]} \,\big|\, X_u = x_u\big) - \underline{P}_{\mathscr{P}}\big(H^{S,G}_{v} \,\big|\, X_u = x_u\big)\big| \leq \frac{1}{8}\Delta(v)(r-s)\|\mathscr{Q}\|^2_{\mathrm{op}}$$

*and*

$$\big|\overline{P}^{\mathrm{D}}_{\mathscr{P}}\big(H^{S,G}_{[s,r]} \,\big|\, X_u = x_u\big) - \overline{P}_{\mathscr{P}}\big(H^{S,G}_{v} \,\big|\, X_u = x_u\big)\big| \leq \frac{1}{8}\Delta(v)(r-s)\|\mathscr{Q}\|^2_{\mathrm{op}}.$$

*In particular, this holds for $\mathscr{P} = \mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathscr{P} = \mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathscr{P} = \mathbb{P}_{\mathscr{M},\mathscr{Q}}$, with $\mathscr{M}$ a non-empty set of initial mass functions.*

*Proof.* Due to Lemmas 6.10$_{281}$, 6.11$_{282}$ and 6.12$_{\curvearrowleft}$, this follows immediately from Proposition 6.7$_{279}$ with $\alpha = 0$, $\beta = 1/2$ and $\gamma = 0$. □

As we will now see, it follows from Proposition 6.13$_{\curvearrowleft}$ and Theorem 4.9$_{166}$ that we can compute the lower and upper probability of $H^{S,G}_{[s,r]}$ up to arbitrary precision, at least for any imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$ whenever $\mathscr{Q}$ has separately specified rows. Say we want to approximate the lower (or upper) probability of $H^{S,G}_{[s,r]}$ with a maximum error of $\epsilon$. Then we choose a grid $v$ over $[s,r]$ such that

$$\frac{1}{8}\Delta(v)(r-s)\|\mathscr{Q}\|^2_{\mathrm{op}} \leq \epsilon;$$

for example, we can use a grid that consists of

$$n \geq \frac{(r-s)^2\|\mathscr{Q}\|^2_{\mathrm{op}}}{8\epsilon}$$

subintervals of equal length. Then the lower probability of $H^{S,G}_{v}$ is $\epsilon$-close to that of $H^{S,G}_{v}$. Recall from Lemma 6.10$_{281}$ that $h^{S,G}_{v}$ has a sum-product representation over $v$. Because $\mathscr{Q}$ has separately specified rows, we can therefore use the iterative procedure in Algorithm 4.2$_{168}$ to compute the lower (or conjugate upper) probability of $H^{S,G}_{v}$ for any imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. However, this is not the most natural computation method; we argue why in Section 6.5$_{310}$ further on, where we also propose a more intuitive – and more efficient – method to compute lower and upper probabilities of time-bounded until events.

As a second important consequence of Proposition 6.13$_{\curvearrowleft}$, we can strengthen Corollary 5.34$_{241}$ in the particular case of time-bounded until events.

**Corollary 6.14.** *Consider a non-empty and bounded set $\mathscr{Q}$ of rate operators, and an imprecise jump process $\mathscr{P}$ such that $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$. Fix some subsets $S, G$ of $\mathscr{X}$, a state history $\{X_u = x_u\}$ in $\mathscr{H}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that* $\max u \leq s \leq r$. *Then for any sequence $(v_n)_{n\in\mathbb{N}}$ of grids over $[s,r]$ with* $\lim_{n\to+\infty}\Delta(v_n) = 0$,

$$\underline{P}^{\mathrm{D}}_{\mathscr{P}}\big(H^{S,G}_{[s,r]} \,\big|\, X_u = x_u\big) = \lim_{n\to+\infty}\underline{P}_{\mathscr{P}}\big(H^{S,G}_{v_n} \,\big|\, X_u = x_u\big)$$

*and*

$$\overline{P}^{\mathrm{D}}_{\mathscr{P}}\big(H^{S,G}_{[s,r]} \,\big|\, X_u = x_u\big) = \lim_{n\to+\infty}\overline{P}_{\mathscr{P}}\big(H^{S,G}_{v_n} \,\big|\, X_u = x_u\big).$$

*In particular, this holds for $\mathscr{P} = \mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathscr{P} = \mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathscr{P} = \mathbb{P}_{\mathscr{M},\mathscr{Q}}$, with $\mathscr{M}$ a non-empty set of initial mass functions.*

*Proof.* Follows immediately from Proposition 6.13$_{\curvearrowleft}$. □

### 6.2.2 Approximating unbounded until events

Next, we move from time-bounded until events to unbounded until events. As is to be expected, the time-bounded until event $H^{S,G}_{[s,r]}$ approximates the unbounded until event $H^{S,G}_{[s,+\infty[}$ as $r$ recedes to $+\infty$.

**Lemma 6.15.** *Consider subsets $S,G$ of $\mathscr{X}$ and a time point $s$ in $\mathbb{R}_{\geq 0}$. Then for any sequence $(r_n)_{n\in\mathbb{N}}$ in $[s,+\infty[$ with $\lim_{n\to+\infty} r_n = +\infty$,*

$$\operatorname{p-w}\lim_{n\to+\infty} h^{S,G}_{[s,r_n]} = h^{S,G}_{[s,+\infty[}.$$

*Proof.* We need to prove that for all $\omega$ in $\Omega$,

$$\lim_{n\to+\infty} h^{S,G}_{r_n}(\omega) = h^{S,G}_{[s,+\infty[}(\omega). \tag{6.6}$$

To do so, we fix some $\omega$ in $\Omega$ and distinguish two cases.

If $h^{S,G}_{[s,+\infty[}(\omega) = 0$, then it is clear that $h^{S,G}_{[s,r_n]}(\omega) = 0$ for all $n$ in $\mathbb{N}$, so Eq. (6.6) holds in this case.

If on the other hand $h^{S,G}_{[s,+\infty[}(\omega) = 1$, then there is a time point $t$ in $[s,+\infty[$ such that

$$\omega(t) \in G \quad \text{and} \quad \big(\forall t' \in [s,+\infty[, t' < t\big)\ \omega(t') \in S \setminus G.$$

Thus, it is clear that $h^{S,G}_{[s,r_n]}(\omega) = 1$ for all $n$ in $\mathbb{N}$ such that $r_n \geq t$. Because $\lim_{n\to+\infty} r_n = +\infty$ by assumption, we infer from this that Eq. (6.6) also holds in this case. □

Unfortunately, it does not follow from Lemma 6.15 that $h^{S,G}_{[s,+\infty[}$ is a limit variable, because we only include the point-wise limit of sequences of simple variables and $h^{S,G}_{[s,r]}$ is not a simple variable. This is not an issue though, because we can also directly approximate the unbounded until event $H^{S,G}_{[s,+\infty[}$ with a sequence of approximating until events.

**Lemma 6.16.** *Consider subsets $S,G$ of $\mathscr{X}$, a sequence of time points $u$ in $\mathscr{U}$ and a time point $s$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s$. Let $(v_n)_{n\in\mathbb{N}}$ be a sequence in $\mathscr{U}_{\neq ()}$ such that $\min v_n = s$ for all $n$ in $\mathbb{N}$, $\lim_{n\to+\infty} \max v_n = +\infty$ and $\lim_{n\to+\infty} \Delta(v_n) = 0$. Then $(h^{S,G}_{v_n})_{n\in\mathbb{N}}$ is a uniformly bounded sequence of $\mathscr{F}_u$-simple variables that converges point-wise to $h^{S,G}_{[s,+\infty[}$, so $h^{S,G}_{[s,+\infty[}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$.*

*Proof.* Our proof is similar to that of Lemma 6.15. Recall from Lemma 6.10281 that for all $n$ in $\mathbb{N}$, $h^{S,G}_{v_n}$ is an $\mathscr{F}_u$-simple variable. Hence, $(h^{S,G}_{v_n})_{n\in\mathbb{N}}$ is a sequence of $\mathscr{F}_u$-simple variables, and this sequence is trivially uniformly bounded. In order to prove the statement, it therefore remains for us to show that for all $\omega$ in $\Omega$,

$$\lim_{n\to+\infty} h^{S,G}_{v_n}(\omega) = h^{S,G}_{[s,+\infty]}(\omega) =: h_\omega. \tag{6.7}$$

Thus, we fix any $\omega$ in $\Omega$, and note that $h_\omega$ can assume two values: 0 and 1.

Let us start with the case $h_\omega = 0$. Then there is a time point $t$ in $[s, +\infty[$ such that

$$\omega(t) \in S^c \cap G^c \quad \text{and} \quad \big(\forall t' \in [s, +\infty[, t' < t\big)\, \omega(t') \in S \setminus G.$$

Because $\omega$ is càdlàg, it follows from Eq. (3.1)$_{58}$ in Definition 3.4$_{57}$ that there is some $\delta$ in $\mathbb{R}_{>0}$ such that $\omega(r) = \omega(t) \in S^c \cap G^c$ for all $r$ in $[t, t+\delta[$. Because $\lim_{n\to+\infty} \max v_n = +\infty$ and $\lim_{n\to+\infty} \Delta(v_n) = 0$ by assumption, there is a natural number $N$ such that for all $n \geq N$, $\max v_n \geq t$ and $\Delta(v) < \delta$ – this ensures that there is a time point $t_{n,k}$ in $v_n$ that also belongs to $[t, t+\delta[$. It is obvious that then $H^{S,G}_{v_n} = 0$, so this verifies Eq. (6.7)$_\curvearrowleft$ in case $h_\omega = 0$.

Next, we consider the case $h_\omega = 1$. Then there is a time point $t$ in $[s, +\infty[$ such that

$$\omega(t) \in G \quad \text{and} \quad \big(\forall t' \in [s, +\infty[, t' < t\big)\, \omega(t') \in S \setminus G.$$

Because $\omega$ is càdlàg, it follows from Eq. (3.1)$_{58}$ in Definition 3.4$_{57}$ that there is some $\delta$ in $\mathbb{R}_{>0}$ such that $\omega(r) = \omega(t) \in G$ for all $r$ in $[t, t+\delta[$. Because $\lim_{n\to+\infty} \max v_n = +\infty$ and $\lim_{n\to+\infty} \Delta(v_n) = 0$ by assumption, there is a natural number $N$ such that for all $n \geq N$, $\max v_n \geq t$ and $\Delta(v) < \delta$ – this ensures that there is a time point $t_{n,k}$ in $v_n$ that also belongs to $[t, t+\delta[$. It is obvious that then $H^{S,G}_{v_n} = 1$, so this verifies Eq. (6.7)$_\curvearrowleft$ in case $h_\omega = 1$. □

Of course, Lemma 6.16$_\curvearrowleft$ is only useful if there is a sequence $(v_n)_{n\in\mathbb{N}}$ in $\mathscr{U}_{\neq()}$ that satisfies the three conditions given there. One easy way to construct such a sequence is as follows. For all $n$ in $\mathbb{N}$, we let $v_n$ be the grid over $[s, s+n]$ that has $n^2$ subintervals of width $1/n$, so

$$v_n = \left(s, s + \frac{1}{n}, \ldots, s + n^2\frac{1}{n}\right). \tag{6.8}$$

Then by construction, $\min v_n = s$, $\max v_n = s + n$ and $\Delta(v_n) = 1/n$ for all $n$ in $\mathbb{N}$, and therefore $\lim_{n\to+\infty} \max v_n = +\infty$ and $\lim_{n\to+\infty} \Delta(v_n) = 0$.

Now that we know that the extended domain contains (unbounded) until events, we wonder whether we can also approximate their lower and upper probabilities up to arbitrary precision. To the best of our knowledge this is not the case, at least not in general. The best that we can do is combine Lemma 6.15$_\curvearrowleft$ or Lemma 6.16$_\curvearrowleft$ with Corollary 5.34$_{241}$ to obtain bounds on the lower and upper probability. In the former case, the lower and upper probability of $H^{S,G}_{[s,r]}$ provide bounds on the lower and upper probability of $H^{S,G}_{[s,+\infty[}$ as we let $r$ recede to $+\infty$; in the latter case, the lower and upper probability of $H^{S,G}_{v_n}$ provide bounds on the lower and upper probability of $H^{S,G}_{[s,+\infty[}$ as we let $n$ recede to $+\infty$. As Joseph's Example 6.17$_\curvearrowright$ further on illustrates, these bounds need not be tight – at least not for $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ if we only require that $\mathscr{Q}$ should have separately specified rows. However, as we will see at the end of Section 6.2.3$_\curvearrowright$, there might very well be another way to determine the lower and upper probability of specific unbounded until events.

### 6.2.3 Lower and upper hitting probabilities

Let us focus on a special type of until events. Suppose that we are interested in whether or not the state of the system is in some set $G \subseteq \mathscr{X}$ of goal states at some time point $t$ in $\mathscr{R} \subseteq \mathbb{R}_{\geq 0}$, but we do not care which state the system was in before this time point $t$. In Joseph's Example 5.1$_{215}$, we called this event

$$H_{\mathscr{R}}^{G} := H_{\mathscr{R}}^{\mathscr{X},G}$$

a *hitting event*; following Norris (1997, Section 3.3), we call the (lower and upper) probabilities of these type of events *(lower and upper) hitting probabilities.*

Interestingly, we can use lower hitting probabilities to illustrate that the inequalities in Theorems 5.31$_{240}$ and 5.32$_{240}$ and Corollaries 5.33$_{240}$ and 5.34$_{241}$ can be strict. Up to some extent, the following example can be seen as the continuous-time counterpart of Bruno's Example 5.3$_{218}$.

**Joseph's Example 6.17.** For any rate $\lambda$ in $\mathbb{R}_{\geq 0}$, we let $Q_\lambda$ be the rate operator with $Q_\lambda(\mathtt{H},\mathtt{T}) = \lambda$ and $Q_\lambda(\mathtt{T},\mathtt{H}) = 0$, and we let $P_\lambda$ denote the homogeneous Markovian jump process $P_{\mathbb{I}_{\mathtt{H}},Q_\lambda}$ that is defined by the initial probability mass function $\mathbb{I}_{\mathtt{H}}$ and this rate operator $Q_\lambda$. Here, we focus on the imprecise Markovian jump process

$$\mathbb{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}} = \{P_\lambda \colon \lambda \in ]0,1]\},$$

with $\mathscr{M} := \{\mathbb{I}_{\mathtt{H}}\}$ and $\mathscr{Q} := \{Q_\lambda \colon \lambda, \lambda_{\mathtt{T}} \in ]0,1]\}$. Note that the set $\mathscr{Q}$ of rate operators is non-empty and bounded – and also convex with separately specified rows – so $\mathbb{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}$ consists of countably additive jump processes due to Corollary 5.30$_{239}$.

We want to determine the lower probability of 'hitting' tails starting from heads at time $t = 0$. That is, we seek to determine

$$\underline{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}\big(H^{\mathtt{T}} \,\big|\, X_0 = \mathtt{H}\big) = \underline{E}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}(h^{\mathtt{T}} \mid X_0 = \mathtt{H})$$

where $h^{\mathtt{T}}$ denotes the indicator of

$$H^{\mathtt{T}} := H_{[0,+\infty[}^{\{\mathtt{T}\}} = H_{[0,+\infty[}^{\mathscr{X},\{\mathtt{T}\}} = \{\omega \in \Omega \colon (\exists t \in [0,+\infty[)\ \omega(t) = \mathtt{T}\}.$$

For all $n$ in $\mathbb{N}$, we let $v_n$ be the grid over $[0,n]$ as defined in Eq. (6.8)$_{\curvearrowleft}$, so

$$v_n = (0, \Delta_n, \ldots, n^2\Delta_n),$$

where we let $\Delta_n := 1/n$. Furthermore, we let $h_n^{\mathtt{T}} := h_{v_n}^{\{\mathtt{T}\}}$ be the indicator of $H_n^{\mathtt{T}} := H_{v_n}^{\{\mathtt{T}\}} = H_{v_n}^{\mathscr{X},\{\mathtt{T}\}}$.

Let us first compute the lower probability of $H^{\mathtt{T}}$ directly. To this end, we need to determine the probability of $H^{\mathtt{T}}$ for every $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}$. Fix some $\lambda$ and

$\lambda_{\mathrm{T}}$ in $]0,1]$. It follows from Lemma 6.16₂₈₅ and Theorem 5.11₂₂₆ – with the trivial dominating variable $g = 1$ – that

$$P_\lambda^{\mathrm{D}}(H^{\mathrm{T}} \mid X_0 = \mathtt{H}) = E_{P_\lambda}^{\mathrm{D}}(h^{\mathrm{T}} \mid X_0 = \mathtt{H}) \\ = \lim_{n\to+\infty} E_{P_\lambda}(h_n^{\mathrm{T}} \mid X_0 = \mathtt{H}) = \lim_{n\to+\infty} P_\lambda(H_n^{\mathrm{T}} \mid X_0 = \mathtt{H}). \quad (6.9)$$

Fix some $n$ in $\mathbb{N}$. Observe that $H_n$ is the complement of $\bigcap_{k=0}^{n^2}\{X_{k\Delta_n} = \mathtt{H}\}$, so by (CP7)₄₂,

$$P_\lambda(H_n \mid X_0 = \mathtt{H}) = 1 - P_\lambda(X_{0\Delta_n} = \mathtt{H}, \dots, X_{n^2\Delta_n} = \mathtt{H} \mid X_0 = \mathtt{H}).$$

It follows from this, (JP3)₇₀ and (CP1)₄₁ that

$$P_\lambda(H_n \mid X_0 = \mathtt{H}) = 1 - \prod_{k=1}^{n^2} P_\lambda(X_{k\Delta_n} = \mathtt{H} \mid X_{0\Delta_n} = \mathtt{H}, \dots, X_{(k-1)\Delta_n} = \mathtt{H}).$$

We use the Markovianity and homogeneity of $P_\lambda$, to yield

$$P_\lambda(H_n \mid X_0 = \mathtt{H}) = 1 - \prod_{k=1}^{n^2} P_\lambda(X_{k\Delta_n} = \mathtt{H} \mid X_{(k-1)\Delta_n} = \mathtt{H}) \\ = 1 - \big(P_\lambda(X_{\Delta_n} = \mathtt{H} \mid X_0 = \mathtt{H})\big)^{n^2}.$$

Because $P_\lambda = P_{\mathbb{I}_{\mathtt{H}},Q_\lambda}$ by definition, it follows from Theorem 3.37₈₇ and Eq. (3.34)₈₄ that $P_\lambda(X_{\Delta_n} = \mathtt{H} \mid X_0 = \mathtt{H}) = e^{\Delta_n Q_\lambda}(\mathtt{H},\mathtt{H})$. From Joseph's Example 3.32₈₃, we know that $e^{\Delta_n Q_\lambda}(\mathtt{H},\mathtt{H}) = e^{-\Delta_n \lambda}$, so

$$P_\lambda(H_n \mid X_0 = \mathtt{H}) = 1 - \big(e^{-\frac{\lambda}{n}}\big)^{n^2} = 1 - e^{-n\lambda}. \quad (6.10)$$

Substituting this into Eq. (6.9), we find that

$$P_\lambda^{\mathrm{D}}(H^{\mathrm{T}} \mid X_0 = \mathtt{H}) = \lim_{n\to+\infty} P_\lambda(H_n \mid X_0 = \mathtt{H}) = \lim_{n\to+\infty} 1 - e^{-n\lambda} = 1, \quad (6.11)$$

where the final equality holds because $\lambda > 0$.

Due to Eq. (6.11), we can directly determine the lower probability of $H^{\mathrm{T}}$:

$$\underline{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}(H^{\mathrm{T}} \mid X_0 = \mathtt{H}) = \inf\big\{P_\lambda^{\mathrm{D}}(H^{\mathrm{T}} \mid X_0 = \mathtt{H})\colon \lambda \in\, ]0,1]\big\} = 1.$$

On the other hand, it follows from Eq. (6.10) that for all $n$ in $\mathbb{N}$,

$$\underline{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}(H_n \mid X_0 = \mathtt{H}) = \inf\big\{P_\lambda(H_n \mid X_0 = \mathtt{H})\colon \lambda \in\, ]0,1]\big\} = 0.$$

All of this work has not been for nothing, because we see that

$$\lim_{n\to+\infty} \underline{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}(H_n \mid X_0 = \mathtt{H}) = \lim_{n\to+\infty} \underline{E}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}(h_n \mid X_0 = \mathtt{H}) \\ < \underline{E}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}(h^{\mathrm{T}} \mid X_0 = \mathtt{H}) = \underline{P}_{\mathscr{M},\mathscr{Q}}^{\mathrm{HM}}(H^{\mathrm{T}} \mid X_0 = \mathtt{H}).$$

To understand why this is significant, we recall from Lemma 6.16$_{285}$ that $(h_n)_{n\in\mathbb{N}}$ is a uniformly bounded sequence of $\mathscr{F}_{(0)}$-simple variables that converges point-wise to $h_{\mathrm{T}}$, and that the latter belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_{(0)})$. Hence, we have constructed an example where the inequalities in Corollary 5.34$_{241}$ – and therefore also Theorem 5.32$_{240}$ – are strict. Even more, it is not difficult to verify that $(h_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence; thus, this also illustrates that the inequalities in Corollary 5.33$_{240}$ – and therefore also Theorem 5.31$_{240}$ – can be strict! Unfortunately, there is no obvious way to check whether or not in this case the lower probabilities corresponding to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ or $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ are continuous; however, an optimist would say that this does not rule out that – in some general case – the inequalities in Theorem 5.31$_{240}$ actually hold with equality for $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ or $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. ૐ

In this example, we went through more trouble than necessary to determine the hitting probability. Indeed, suppose we have a homogeneous Markovian jump process $P_{p_0,Q}$ and that we are interested in the probability of hitting $G \subset \mathscr{X}$. That is, we want to determine the gamble $h^G$ on $\mathscr{X}$ defined by

$$h^G(x) := P_{p_0,Q}(H^G_{[0,+\infty[} \,|\, X_0 = x) \quad \text{for all } x \in \mathscr{X}.$$

Then it is well-known – see for example (Norris, 1997, Theorem 3.3.1) – that $h^G$ is the minimal and non-negative solution to

$$\mathbb{I}_G h = \mathbb{I}_G + \mathbb{I}_{G^c} Q h,$$

where $h^G$ is minimal in the sense that $h^G \leq h$ for any non-negative gamble $h$ on $\mathscr{X}$ that also satisfies this equality.

It would be interesting to investigate whether or not a similar result holds – but with $\underline{Q}_{\mathscr{Q}}$ and $\overline{Q}_{\mathscr{Q}}$ instead of $Q$ – for the lower and upper probability with respect to any of the three Markovian imprecise jump processes characterised by $\mathscr{Q}$. Krak et al. (2019, Corollary 19) generalise the discrete-time counterpart of this result (see Norris, 1997, Theorem 1.3.2) to imprecise Markov chains, and it is not all too far fetched to assume that their proofs can be adapted to the continuous-time setting. Be that as it may, we elect to leave this path undiscovered, even though the prospect of going down this rabbit hole is alarmingly alluring.

## 6.3 Hitting times

Besides hitting events, we also informally introduced hitting times in Joseph's Example 5.1$_{215}$. Consider a set $G \subseteq \mathscr{X}$ of 'goal' states and any time point $s$ in $\mathbb{R}_{\geq 0}$. Following Norris (1997, Section 3.3), we defined the *hitting time of $G$ after $s$* as the non-negative extended real variable $\tau^G_{[s,+\infty[}$ defined by

$$\tau^G_{[s,+\infty[}(\omega) := \inf\{t \in [s,+\infty[ : \omega(t) \in G\} \quad \text{for all } \omega \in \Omega. \tag{6.12}$$

Furthermore, for any $r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, we define the *truncated hitting time*

$$\tau^G_{[s,r]} := \tau^G_{[s,+\infty[} \wedge r; \tag{6.13}$$

it is easy to see that, for all $\omega$ in $\Omega$,

$$\tau^G_{[s,r]}(\omega) = \min\Big(\big\{t \in [s,r] \colon \omega(t) \in G\big\} \cup \{r\}\Big),$$

where the infimum is reached because $\omega$ is càdlàg, and therefore continuous from the right. Additionally, for any grid $v = (t_0, \dots, t_n)$ over $[s,r]$, we call

$$\tau^G_v \colon \Omega \to [s,r] \colon \omega \mapsto \min\big(\big\{t \in v \colon \omega(t) \in G\big\} \cup \{r\}\big) \tag{6.14}$$

an *approximating hitting time*.

Because hitting times are similar to hitting events, it is reasonable to expect that a hitting time is the point-wise limit of a sequence of approximating hitting times and/or a sequence of truncated hitting times, and that, in turn, a truncated hitting time is the point-wise limit of approximating hitting times. We will prove that this is the case, in Section 6.3.1 for hitting times and in Section 6.3.2292 for truncated hitting times. First, however, we show that the approximating hitting time $\tau^G_v$ is a simple variable.

**Lemma 6.18.** *Consider a subset $G$ of $\mathscr{X}$, a sequence of time points $u$ in $\mathscr{U}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$, and fix some grid $v$ over $[s,r]$. Then*

$$\tau^G_v = s + \sum_{k=1}^{n} (t_k - t_{k-1}) \prod_{\ell=0}^{k-1} \mathbb{I}_{G^c}\big(X_{t_\ell}\big),$$

*so $\tau^G_v$ has a sum-product representation over $v$. Consequently, $\tau^G_v$ is $\mathscr{F}_u$-simple.*

*Proof.* That $\tau^G_v$ has the sum-product representation over $v$ of the statement is a matter of straightforward verification. Observe that $v \succcurlyeq u$ because, by assumption, $v$ is a grid over $[s,r]$ and $\max u \leq s$. For this reason, and because $\tau^G_v$ has a sum-product representation over $v$, it follows from Lemma 4.6164 that $\tau^G_v$ is $\mathscr{F}_u$-simple. □

### 6.3.1 From approximating hitting times to hitting times

Our investigation into hitting times starts with the following counterpart of Lemma 6.16285 for hitting times.

**Lemma 6.19.** *Consider a subset $G$ of $\mathscr{X}$, a sequence of time points $u$ in $\mathscr{U}$ and a time points $s$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s$. Let $(v_n)_{n \in \mathbb{N}}$ be a sequence in $\mathscr{U}_{\neq ()}$ such that $\min v_n = s$ for all $n$ in $\mathbb{N}$, $\lim_{n \to +\infty} \max v_n = +\infty$ and $\lim_{n \to +\infty} \Delta(v_n) = 0$. Then $(\tau^G_{v_n})_{n \in \mathbb{N}}$ is a sequence of $\mathscr{F}_u$-simple variables that is uniformly bounded below and that converges point-wise to $\tau^G_{[s,+\infty[}$, so $\tau^G_{[s,+\infty[}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$.*

*Proof.* For all $n$ in $\mathbb{N}$, $\tau^G_{v_n}$ is $\mathscr{F}_u$-simple due to Lemma 6.18 and bounded below by $s$ due to Eq. (6.14). To prove the statement, we still need to show that the sequence $(\tau^G_{v_n})_{n\in\mathbb{N}}$ converges point-wise to $\tau^G_{[s,+\infty[}$; that is, for all $\omega$ in $\Omega$ we need to show that

$$\lim_{n\to+\infty} \tau^G_{v_n}(\omega) = \tau^G_{[s,+\infty[}(\omega). \tag{6.15}$$

To this end, we fix any path $\omega$ in $\Omega$. Observe that $\tau^G_{[s,+\infty[}(\omega) = +\infty$ if and only if $\omega(t) \notin G$ for all $t$ in $\mathbb{R}_{\geq 0}$. Whenever this is the case, it is clear that $\tau^G_{v_n}(\omega) = \max v_n$ for all $n$ in $\mathbb{N}$. Because $\lim_{n\to+\infty} \max v_n = +\infty$ by assumption, this verifies Eq. (6.15).

What remains is the case $t := \tau^G_{[s,+\infty[}(\omega) < +\infty$; note that then

$$\omega(t) \in G \quad \text{and} \quad \big(\forall t' \in [s,t[\big)\ \omega(t') \notin G. \tag{6.16}$$

Because $\omega$ is càdlàg, it is continuous from the right at $t$; hence, there is a positive real number $r > t$ such that

$$\big(\forall t' \in [t,r[\big)\ \omega(t') = \omega(t) \in G. \tag{6.17}$$

As $\lim_{n\to+\infty} \max v_n = +\infty$ by assumption, there is a natural number $N_t$ such that $\max v_n > t$ for all $n \geq N_t$. Furthermore, for all $\epsilon$ in $\mathbb{R}_{>0}$, there is a natural number $N_\epsilon$ such that $\Delta(v_n) < \epsilon$ because $\lim_{n\to+\infty} \Delta(v_n) = 0$ by assumption. Fix any positive real number $\epsilon$ such that $\epsilon < r - t$. Then for any natural number $n \geq \max\{N_t, N_\epsilon\}$, $\max v_n > t$ and $\Delta(v_n) < \epsilon < r - t$, so there is at least one time point in $v_n$ that belongs to $[t, r[$; let us denote the smallest one by $t_{n,k}$. Observe that $\tau^G_{v_n}(\omega) = t_{n,k}$ due to Eqs. (6.16) and (6.17), so

$$|\tau^G_{[s,+\infty[}(\omega) - \tau^G_{v_n}(\omega)| = |t - t_{n,k}| \leq \Delta(v_n) < \epsilon.$$

Because $\epsilon$ was an arbitrary – yet sufficiently small – positive real number, we infer from this that Eq. (6.15) also holds whenever $\tau^G_{[s,+\infty[}(\omega) < +\infty$, as required. □

Besides establishing that hitting times are limit variables, Lemma 6.19 does not have much useful consequences. Because it establishes that a hitting time is the point-wise limit of a sequence of simple variables that are uniformly bounded below, we could rely on Theorem 5.32240 to determine bounds on the lower and upper expected hitting time. However, we can only do so if there is a limit variable $g$ that (i) has finite upper expectation, and (ii) dominates the approximating hitting times. To our great dismay, there is no obvious or natural candidate for this, so we are left empty handed – for now.

In this regard, a promising avenue of research is to investigate whether the following result regarding expected hitting times for homogeneous Markovian jump processes generalises to imprecise ones. Consider a homogeneous Markovian jump process $P_{p_0,Q}$, let $G$ be a subset of the state space $\mathscr{X}$, and let $h^G$ be the non-negative extended real valued function on $\mathscr{X}$ defined by

$$h^G(x) := E_{p_0,Q}\big(\tau^G_{[0,+\infty[} \,\big|\, X_0 = x\big).$$

Then it is well-known (see Norris, 1997, Theorem 3.3.3) that $h^G$ is the minimal non-negative solution of

$$\mathbb{I}_G h = \mathbb{I}_{G^c}(1 + Qh);$$

note that $h$ is non-negative but can attain $+\infty$, so technically we should extend the domain of $Q$ accordingly. Krak et al. (2019, Corollary 13) establish that the discrete-time counterpart of this result (see Norris, 1997, Theorem 1.3.5) also holds for imprecise Markov chains, and an optimist might be inclined to conjecture that the same is true for continuous time. To protect ourselves from going down a rabbit hole, we do not pursue this attractive avenue of research here; instead, we opt for a detour that involves truncated hitting times.

### 6.3.2 Truncated hitting times

In our study of truncated hitting times, we will follow more or less the same approach as in Section $6.2.1_{281}$, albeit this time for approximating and truncated hitting times instead of approximating and time-bounded until events. First, we establish an upper bound on the difference between the truncated hitting time $\tau^G_{[s,r]}$ and the approximating hitting time $\tau^G_v$ corresponding to any grid $v$ over $[s,r]$.

**Lemma 6.20.** *Consider a subset $G$ of $\mathscr{X}$ and time points $s,r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. Then for any grid $v$ over $[s,r]$,*

$$\left|\tau^G_{[s,r]} - \tau^G_v\right| \leq \Delta(v) + \frac{1}{2}(r-s)(\eta_{[s,r]} - \eta_v).$$

*Proof.* The statement is trivially true whenever $s = r$, so without loss of generality we may assume that $s < r$. We enumerate the time points in $v$ by $(t_0, \dots, t_n)$, and fix an arbitrary path $\omega$ in $\Omega$. Then by Eq. $(6.13)_{290}$, the truncated hitting time $\tau^G_{[s,r]}(\omega)$ assumes a value in $[s,r]$; let us denote this value by $t$. Similarly, it follows from Eq. $(6.14)_{290}$ that $\tau^G_v(\omega) = t_k$ for some $k$ in $\{0,\dots,n\}$. Note that

$$\tau^G_{[s,r]}(\omega) = t \leq t_k = \tau^G_v(\omega).$$

Furthermore, it is clear from Eqs. $(6.13)_{290}$ and $(6.14)_{290}$ that if $t$ belongs to $v$, then $t$ has to be equal to $t_k$; whenever this is the case,

$$\left|\tau^G_{[s,r]}(\omega) - \tau^G_v(\omega)\right| = t_k - t = 0. \tag{6.18}$$

Next, we investigate the remaining case that $t$ does not belong to $v$; observe that in this case $t < t_k$. We let $t_\ell$ denote the first – or smallest – time point in $v$ that succeeds $t$; that is, $\ell$ is the unique index in $\{1,\dots,n\}$ such that $t_{\ell-1} < t < t_\ell$. Note that the path $\omega$ jumps to $G$ somewhere in the interval $]t_{\ell-1}, t_\ell[$. As $t < t_k$ by assumption, it is clear that $t_{\ell-1} < t < t_\ell \leq t_k$ and therefore $0 < \ell \leq k$. In case $\ell = k$,

$$\left|\tau^G_{[s,r]}(\omega) - \tau^G_v(\omega)\right| = t_k - t \leq t_k - t_{k-1}. \tag{6.19}$$

In case $\ell < k$, the path $\omega$ is in $G^c$ at $t_{\ell-1}$, in $G$ at $t$ and in $G^c$ at $t_\ell$; clearly, this implies that $\omega(t_{\ell-1}) \neq \omega(t) \neq \omega(t_\ell)$. For this reason, it follows from Lemma $6.9_{280}$ that $\omega$ belongs to $A := \{\omega' \in \Omega \colon \eta_v(\omega') < \eta_{[s,r]}(\omega')\}$, so

$$\left|\tau^G_{[s,r]}(\omega) - \tau^G_v(\omega)\right| = t_k - t \leq (r-s) = (r-s)\mathbb{I}_A \leq \frac{1}{2}(r-s)\big(\eta_{[s,r]}(\omega) - \eta_v(\omega)\big), \tag{6.20}$$

where the second inequality follows from Lemma 6.8$_{279}$.

To finalise our proof, we recall from Proposition 6.2$_{275}$ that $(\eta_{[s,r]} - \eta_{v})$ is non-negative. Because furthermore $(t_k - t_{k-1}) \leq \Delta(v)$ for all $k$ in $\{1, \ldots, n\}$, the inequality of the statement follows immediately from Eqs. (6.18)↶ to (6.20)↶. □

Next, we use Lemma 6.20↶ to establish that the truncated hitting time $\tau^G_{[s,r]}$ is the point-wise limit of a uniformly bounded sequence of approximating hitting times.

**Lemma 6.21.** *Consider a subset $G$ of $\mathscr{X}$, a sequence of time points $u$ in $\mathscr{U}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Let $(v_n)_{n \in \mathbb{N}}$ be a sequence of grids over $[s,r]$ such that $\lim_{n \to +\infty} \Delta(v_n) = 0$. Then $(\tau^G_{v_n})_{n \in \mathbb{N}}$ is a uniformly bounded sequence of $\mathscr{F}_u$-simple variables that converges point-wise to $\tau^G_{[s,r]}$, so $\tau^G_{[s,r]}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$.*

*Proof.* Fix some $n$ in $\mathbb{N}$. Because $v_n$ is a grid over $[s,r]$, it follows from Lemma 6.18$_{290}$ that $\tau^G_{v_n}$ is an $\mathscr{F}_u$-simple variable. Furthermore, it follows immediately from Eq. (6.14)$_{290}$ that $s \leq \tau^G_{v_n} \leq r$ for all $n$ in $\mathbb{N}$. Thus, we have shown that $(\tau^G_{v_n})_{n \in \mathbb{N}}$ is a uniformly bounded sequence of $\mathscr{F}_u$-simple variables. To verify the statement, it remains for us to prove that this sequence converges point-wise to $\tau^G_{[s,r]}$; that is, we need to show that for all $\omega$ in $\Omega$,

$$\lim_{n \to +\infty} \tau^G_{v_n}(\omega) = \tau^G_{[s,r]}(\omega). \tag{6.21}$$

Fix an arbitrary path $\omega$ in $\Omega$, and recall from Lemma 6.20↶ that for all $n$ in $\mathbb{N}$,

$$\left|\tau^G_{[s,r]}(\omega) - \tau^G_{v_n}(\omega)\right| \leq \Delta(v_n) + \frac{1}{2}(r-s)\big(\eta_{[s,r]}(\omega) - \eta_{v}(\omega)\big).$$

Note that because $\lim_{n \to +\infty} \Delta(v_n) = 0$ by assumption, it follows from Lemma 5.25$_{235}$ that $\lim_{n \to +\infty}\big(\eta_{[s,r]}(\omega) - \eta_{v_n}(\omega)\big) = 0$. Hence, we infer from the preceding inequality that

$$\lim_{n \to +\infty} \left|\tau^G_{[s,r]}(\omega) - \tau^G_{v_n}(\omega)\right| \leq \lim_{n \to +\infty} \Big(\Delta(v_n) + \frac{1}{2}(r-s)\big(\eta_{[s,r]}(\omega) - \eta_{v}(\omega)\big)\Big) = 0,$$

and this implies Eq. (6.21), as required. □

Additionally, Lemma 6.20↶ induces an upper bound on the difference between the lower and upper expectation of the truncated hitting time $\tau^G_{[s,r]}$ on the one hand and the lower and upper expectation of the approximating hitting time $\tau^G_{v}$ corresponding to any grid $v$ over $[s,r]$ on the other hand.

**Proposition 6.22.** *Consider a non-empty and bounded set $\mathscr{Q}$ of rate operators, and an imprecise jump process $\mathscr{P}$ such that $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$, and let $\lambda := \|\mathscr{Q}\|_{\mathrm{op}}$. Fix some subset $G$ of $\mathscr{X}$, a state history $\{X_u = x_u\}$ in $\mathscr{H}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Then for any grid $v$ over $[s,r]$,*

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}\big(\tau^G_{[s,r]} \,\big|\, X_u = x_u\big) - \underline{E}_{\mathscr{P}}\big(\tau^G_{v} \,\big|\, X_u = x_u\big)\right| \leq \Delta(v) + \frac{1}{8}\Delta(v)(r-s)^2 \|\mathscr{Q}\|^2_{\mathrm{op}}$$

*and*

$$\left|\overline{E}^{\mathrm{D}}_{\mathscr{P}}\left(\tau^{G}_{[s,r]} \,\middle|\, X_u = x_u\right) - \overline{E}_{\mathscr{P}}\left(\tau^{G}_{v} \,\middle|\, X_u = x_u\right)\right| \leq \Delta(v) + \frac{1}{8}\Delta(v)(r-s)^2\|\mathcal{Q}\|^2_{\mathrm{op}}.$$

*In particular, this holds for* $\mathscr{P} = \mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$, $\mathscr{P} = \mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ *and* $\mathscr{P} = \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, *with* $\mathcal{M}$ *a non-empty set of initial mass functions.*

*Proof.* Due to Lemmas 6.20<sub></sub>292 and 6.21↶, the statement follows immediately from Proposition 6.7 279 with $\alpha = 1$, $\beta = \frac{1}{2}(r-s)$ and $\gamma = 0$. □

Because of the preceding result, we can approximate lower and upper expected truncated hitting times up to arbitrary precision. Similar comments hold here as those that we made right after Proposition 6.13 283; again, we postpone an in-depth investigation until Section 6.5 310 further on. Proposition 6.22↶ also implies that in the particular case of approximating and truncated hitting times, the bounds in Corollary 5.34 241 are tight.

**Corollary 6.23.** *Consider a non-empty bounded set* $\mathcal{Q}$ *of rate operators, and an imprecise jump process* $\mathscr{P}$ *such that* $\mathscr{P} \subseteq \mathbb{P}_{\mathcal{Q}}$. *Fix a subset* $G$ *of* $\mathscr{X}$, *a state history* $\{X_u = x_u\}$ *in* $\mathscr{H}$ *and time points* $s, r$ *in* $\mathbb{R}_{\geq 0}$ *such that* $\max u \leq s \leq r$. *Then for any sequence* $(v_n)_{n\in\mathbb{N}}$ *of grids over* $[s,r]$ *with* $\lim_{n\to+\infty}\Delta(v_n) = 0$,

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}\left(\tau^{G}_{[s,r]} \,\middle|\, X_u = x_u\right) = \lim_{n\to+\infty} \underline{E}_{\mathscr{P}}\left(\tau^{G}_{v_v} \,\middle|\, X_u = x_u\right)$$

*and*

$$\overline{E}^{\mathrm{D}}_{\mathscr{P}}\left(\tau^{G}_{[s,r]} \,\middle|\, X_u = x_u\right) = \lim_{n\to+\infty} \overline{E}_{\mathscr{P}}\left(\tau^{G}_{v_n} \,\middle|\, X_u = x_u\right).$$

*In particular, this holds for* $\mathscr{P} = \mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$, $\mathscr{P} = \mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ *and* $\mathscr{P} = \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, *with* $\mathcal{M}$ *a non-empty set of initial mass functions.*

*Proof.* Follows almost immediately from Proposition 6.22↶. □

This result is particularly interesting because the truncated hitting time $\tau^{G}_{[s,r]}$ converges monotonically to the hitting time $\tau^{G}_{[s,+\infty[}$ for increasing $r$.

**Lemma 6.24.** *Consider a subset* $G$ *of* $\mathscr{X}$, *a sequence of time points* $u$ *in* $\mathscr{U}$ *and a time point* $s$ *in* $\mathbb{R}_{\geq 0}$ *such that* $\max u \leq s$. *Let* $(r_n)_{n\in\mathbb{N}}$ *be a non-decreasing sequence of time points in* $[s,+\infty[$ *with* $\lim_{n\to+\infty} r_n = +\infty$. *Then the corresponding sequence* $(\tau^{G}_{[s,r_n]})_{n\in\mathbb{N}}$ *of truncated hitting times in* $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ *is non-decreasing and uniformly bounded below, and this sequence converges pointwise to* $\tau^{G}_{[s,+\infty[}$.

*Proof.* Recall from Lemma 6.21↶ that for all $n$ in $\mathbb{N}$, $\tau^{G}_{[s,r_n]}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$. Because $(r_n)_{n\in\mathbb{N}}$ is non-decreasing by assumption, it follows immediately from Eq. (6.13) 290 that $(\tau^{G}_{[s,r_n]})_{n\in\mathbb{N}}$ is a non-decreasing sequence. Furthermore, it is clear that $\inf \tau^G_{v_n} \geq s > -\infty$ for all $n$ in $\mathbb{N}$. Finally, the sequence of truncated hitting times converges point-wise to the hitting time because for all $\omega$ in $\Omega$,

$$\lim_{n\to+\infty} \tau^G_{[s,r_n]}(\omega) = \lim_{n\to+\infty}\Big(\tau^G_{[s,+\infty[}(\omega) \wedge r_n\Big) = \tau^G_{[s,+\infty[}(\omega),$$

where for the first equality we used Eq. (6.13)$_{290}$ and where the second equality holds because $\lim_{n\to+\infty} r_n = +\infty$. □

Due to the preceding result and Corollary 5.33$_{240}$, we can determine (a bound on) the lower and upper expectation of the hitting time $\tau^G_{[s,+\infty[}$ by determining the lower and upper expectation of the truncated hitting time $\tau^G_{[s,r]}$ for increasing values of $r$. Unfortunately, there seems to be no clear cut way to show that the bound for the lower expectation is actually tight. The imprecise jump process in Joseph's Example 6.17$_{287}$ does not provide a counterexample this time around, as the lower and upper expected truncated hitting times do converge to the lower and upper expected hitting times.

## 6.4 Idealised variables in the form of Riemann integrals

Suppose that for some gamble $f$ on $\mathscr{X}$, we are interested in the integral of $f(X_t)$ over the time period $[s,r]$, where $s$ and $r$ are time points in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. For example, 'the length of time that the system is in the state $x$ over $[s,r]$' corresponds to the integral of $\mathbb{I}_x(X_t)$. Alternatively, if $r > s$, 'the average value of $f(X_t)$ over $[s,r]$' corresponds to the integral of $\frac{1}{r-s} f(X_t)$.

Formally, the integral of $f(X_t)$ is a variable on $\Omega$ that we define through Riemann integration; hence, we need that for all $\omega$ in $\Omega$,

$$f \circ \omega \colon [s,r] \to \mathbb{R} \colon t \mapsto [f \circ \omega](t) = f\big(\omega(t)\big)$$

is Riemann integrable. We show in Proposition 6.30$_{298}$ further on that this is always the case. The variable

$$\Omega \to \mathbb{R} \colon \omega \mapsto \int_s^r [f \circ \omega](t)\,\mathrm{d}t = \int_s^r f\big(\omega(t)\big)\,\mathrm{d}t$$

is therefore well-defined, and this variable is what we will call the integral of $f(X_t)$ over $[s,r]$. We treat the particular case of these integrals in Section 6.4.4$_{305}$ further on. First, however, in Section 6.4.1$_{\curvearrowright}$ we recall the definition of the Riemann integral, as well as some of its convenient properties. We subsequently consider more general idealised variables that are defined as the 'Riemann integral of a family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$'. More precisely, we introduce these variables through limit arguments in Section 6.4.2$_{297}$, and investigate if and how these limit arguments carry over to the (conditional) lower and upper expectations in Section 6.4.3$_{301}$.

### 6.4.1 The Riemann integral

Consider two real numbers $s$ and $r$ such that $s \leq r$, and a real-valued function $h$ on $[s, r]$. The Riemann integral of $h$, sometimes also called the Riemann-Darboux integral, has multiple equivalent definitions: it can be defined through an epsilon-delta statement, through 'lower and upper Darboux sums' or as the 'limit of a net' (see Schechter, 1997, Chapter 24). These three equivalent definitions all use '(tagged) partitions of $[s, r]$': a 'partition of $[s, r]$' is a grid $v = (t_0, \dots, t_n)$ over $[s, r]$, and we 'tag' this grid by fixing a point $s_k$ in the subinterval $[t_{k-1}, t_k]$, for all $k$ in $\{1, \dots, n\}$. The epsilon-delta definition of the Riemann integral uses 'tagged partitions' as follows (see Schechter, 1997, Definition 24.3).

**Definition 6.25.** Consider real numbers $s, r$ such that $s \leq r$. A function $h \colon [s, r] \to \mathbb{R}$ is called *Riemann integrable* if there is some real number $\gamma$ such that for any positive real number $\epsilon$, there is some maximum grid width $\delta_\epsilon$ in $\mathbb{R}_{>0}$ such that for any grid $v = (t_0, \dots, t_n)$ over $[s, r]$ with $\Delta(v) < \delta_\epsilon$ and any $s_1, \dots, s_n$ in $[s, r]$ with $t_{k-1} \leq s_k \leq t_k$ for all $k$ in $\{1, \dots, n\}$,

$$\left| \sum_{k=1}^{n} h(s_k)(t_k - t_{k-1}) - \gamma \right| < \epsilon.$$

Whenever such a real number $\gamma$ exists, it is unique; hence, we call it the *Riemann integral of* $h$, and denote it by $\int_s^r h(t)\,\mathrm{d}t$.

It is well-known that the set of Riemann integrable functions is a real vector space that includes the constant functions – much like the simple variables with respect to some field of events. The following result establishes this, as well as that the Riemann integral is homogeneous, additive and monotone – similar to the expectation corresponding to a probability charge.

**Proposition 6.26.** *Consider real numbers $s, r$ such that $s \leq r$, real-valued functions $h, g$ on $[s, r]$ that are Riemann integrable, and a real number $\mu$. Then*

RI1. *$\mu h$ is Riemann integrable, and $\int_s^r [\mu h](t)\,\mathrm{d}t = \mu \int_s^r h(t)\,\mathrm{d}t$;*

RI2. *$h + g$ is Riemann integrable, and $\int_s^r [h+g](t)\,\mathrm{d}t = \int_s^r h(t)\,\mathrm{d}t + \int_s^r g(t)\,\mathrm{d}t$;*

RI3. *the constant function $\mu$ is Riemann integrable, and $\int_s^r \mu\,\mathrm{d}t = (r-s)\mu$.*

*Furthermore,*

RI4. *$\int_s^r h(t)\,\mathrm{d}t \leq \int_s^r g(t)\,\mathrm{d}t$ whenever $h \leq g$;*

RI5. *for every grid $v = (t_0, \dots, t_n)$ over $[s, r]$, the restrictions of $h$ to the subintervals $[t_{k-1}, t_k]$ of $v$ are Riemann integrable, and*

$$\int_s^r h(t)\,\mathrm{d}t = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} h(t)\,\mathrm{d}t$$

*Proof.* These properties are well-known; for their proofs, we refer to (Schechter, 1997, Chapter 24) and/or (Rudin, 1976, Theorem 6.12) □

Lebesgue's integrability criterion (see Schechter, 1997, Theorem 24.46) substantiates a necessary and sufficient condition for the Riemann integrability of a real-valued function $f$ on $[s,r]$. For our objectives, the following sufficient condition for Riemann integrability will suffice (see Schechter, 1997, Definition 19.28 and Section 24.26).

**Definition 6.27.** Consider real numbers $s, r$ such that $s \leq r$. The function $h\colon [s,r] \to \mathbb{R}$ is called *piece-wise continuous* if (i) $h$ has a left-sided limit in $\mathbb{R}$ at all $t$ in $]s,r]$, (ii) $h$ has a right-sided limit in $\mathbb{R}$ at all $t$ in $[s,r[$, and (iii) the set of discontinuity points

$$\left\{t \in \left]s,r\right] : h(t) \neq \lim_{\Delta \searrow 0} h(t-\Delta)\right\} \cup \left\{t \in \left[s,r\right[ : h(t) \neq \lim_{\Delta \searrow 0} h(t+\Delta)\right\}$$

is finite.

**Proposition 6.28.** *Consider real numbers $s$ and $r$ such that $s \leq r$ and a function $h\colon [s,r] \to \mathbb{R}$. If $h$ is piece-wise continuous, then $h$ is Riemann integrable.*

### 6.4.2 The integral of a piece-wise continuous family of gambles

Suppose that instead of a fixed gamble $f$ on $\mathscr{X}$, we have a gamble $f_t$ on $\mathscr{X}$ for every time point $t$ in $[s,r]$. Then we can only define a variable that corresponds to 'the integral of $f_t(X_t)$ over $[s,r]$' if for all $\omega$ in $\Omega$,

$$f_\bullet \circ \omega\colon [s,r] \to \mathbb{R}\colon t \mapsto f_t\big(\omega(t)\big)$$

is Riemann integrable. We could set out to characterise all families $(f_t)_{t\in[s,r]}$ that satisfy this requirement with the help of Lebesgue's criterion for Riemann integrability, but this would lead us rather far astray. Instead, we opt to establish a sufficient condition with the help of Proposition 6.28. For this, we use a notion of piece-wise continuity for families of gambles on $\mathscr{X}$ – that is, piece-wise continuity in the sense of Schechter (1997, Definition 19.28) for maps from $[s,r]$ to $\mathbb{G}(\mathscr{X})$.

**Definition 6.29.** Consider time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. The family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$ is *piece-wise continuous* if (i) $\lim_{\Delta\searrow 0} f_{t-\Delta}$ exists for all $t$ in $]s,r]$, (ii) $\lim_{\Delta\searrow 0} f_{t+\Delta}$ exists for all $t$ in $[s,r[$, and (iii) the set of discontinuity points

$$\left\{t \in \left]s,r\right] : f_t \neq \lim_{\Delta \searrow 0} f_{t-\Delta}\right\} \cup \left\{t \in \left[s,r\right[ : f_t \neq \lim_{\Delta \searrow 0} f_{t+\Delta}\right\}$$

is finite.

Because we have chosen to restrict ourselves to the càdlàg paths, this condition is sufficient for $f_\bullet \circ \omega$ to be Riemann integrable.

**Proposition 6.30.** *Consider time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, and a family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$ that is piece-wise continuous. Then for all $\omega$ in $\Omega$, the function*

$$f_\bullet \circ \omega \colon [s,r] \to \mathbb{R} \colon t \mapsto f_t\big(\omega(t)\big)$$

*is piece-wise continuous, and therefore Riemann integrable.*

*Proof.* Due to Proposition 6.28, it suffices to check whether $f_\bullet \circ \omega$ is piece-wise continuous; that is, we need to check whether the three requirements in Definition 6.27 hold. First, we verify that $f_\bullet \circ \omega$ has a left-sided limit at all $t$ in $]s,r]$. To this end, we fix some $t$ in $]s,r]$. Because $\omega$ is càdlàg, there is a $\delta_1$ in $]0,t]$ such that

$$\big(\forall s_1, s_2 \in\, ]t-\delta_1, t[\big)\ \omega(s_1) = \omega(s_2);$$

that is, the path $\omega$ is constant over $]t-\delta_1, t[$. We denote the value of $\omega$ on $]t-\delta_1, t[$ by $x$; then

$$\big(\forall s' \in\, ]t-\delta_1, t[\big)\ [f_\bullet \circ \omega](s') = f_{s'}\big(\omega(s')\big) = f_{s'}(x). \tag{6.22}$$

Because the family $(f_t)_{t\in[s,r]}$ is piece-wise continuous by assumption, the left-sided limit $\lim_{\Delta\searrow 0} f_{t-\Delta}$ exists. We denote this limit by $f_t^-$, and fix any positive real number $\epsilon$. Then there is a positive real number $\delta_2$ in $]0,t]$ such that

$$\big(\forall s' \in\, ]t-\delta_2, t[\big)\ \|f_{s'} - f_t^-\| < \epsilon. \tag{6.23}$$

Let $\delta := \min\{\delta_1, \delta_2\}$. It follows immediately from Eqs. (6.22) and (6.23) and the definition of the supremum norm $\|\bullet\|$ that

$$\big(\forall s' \in\, ]t-\delta, t[\big)\ \big|[f_\bullet \circ \omega](s') - f_t^-(x)\big| = \big|f_{s'}(x) - f_t^-(x)\big| \leq \|f_{s'} - f_t^-\| < \epsilon.$$

Because $\epsilon$ was an arbitrary positive real number, we infer from this that the left-sided limit of $f_\bullet \circ \omega$ at $t$ is equal to $f_t^-(x)$; because $f_t^- = \lim_{\Delta\searrow 0} f_{t-\Delta}$ and $x = \lim_{\Delta\searrow 0} \omega(t-\Delta)$,

$$\lim_{\Delta\searrow 0} [f_\bullet \circ \omega](t-\Delta) = f_t^-(x) = \left[\lim_{\Delta\searrow 0} f_{t-\Delta}\right]\left(\lim_{\Delta\searrow 0} \omega(t-\Delta)\right). \tag{6.24}$$

An entirely analogous argument shows that $f_\bullet \circ \omega$ has a limit from the right at all $t$ in $[s,r[$. Furthermore,

$$\lim_{\Delta\searrow 0} [f_\bullet \circ \omega](t+\Delta) = \left[\lim_{\Delta\searrow 0} f_{t+\Delta}\right]\left(\lim_{\Delta\searrow 0} \omega(t+\Delta)\right) = \left[\lim_{\Delta\searrow 0} f_{t+\Delta}\right]\big(\omega(t)\big), \tag{6.25}$$

where for the second equality we used that $\omega$ is continuous from the right.

Finally, we verify that the set of discontinuity points

$$\left\{t \in\, ]s,r] \colon [f_\bullet \circ \omega](t) \neq \lim_{\Delta\searrow 0} [f_\bullet \circ \omega](t-\Delta)\right\} \cup \left\{t \in [s,r[\, \colon [f_\bullet \circ \omega](t) \neq \lim_{\Delta\searrow 0} [f_\bullet \circ \omega](t+\Delta)\right\}$$

is finite. To this end, we observe that by Eqs. (6.24) and (6.25), $t$ in $[s,r]$ is a discontinuity point of $f_\bullet \circ \omega$ only if $t$ is a discontinuity point of $(f_t)_{t\in[s,r]}$ and/or $t$ is a jump time of $\omega$. Hence, the set of discontinuity points of $f_\bullet \circ \omega$ is a subset of the union of the set of discontinuity points of $(f_t)_{t\in[s,r]}$ and the set of jump times $\mathscr{J}_{[s,r]}(\omega)$. The former is finite because $(f_t)_{t\in[s,r]}$ is piece-wise continuous by assumption, and the latter is finite due to Lemma 5.20232. Thus, the set of discontinuity points of $f_\bullet \circ \omega$ is finite, as required. □

We cannot emphasise enough that the preceding result only holds because we have restricted ourselves to càdlàg paths. Because of this result, we can formally define the 'Riemann integral of the family $(f_t)_{t\in[s,r]}$' as follows. For any time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$ and any piece-wise continuous family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$, we call

$$\int_s^r f_t(X_t)\,\mathrm{d}t \colon \Omega \to \mathbb{R} \colon \omega \mapsto \int_s^r f_t\big(\omega(t)\big)\,\mathrm{d}t \tag{6.26}$$

the *Riemann integral of* $f_t(X_t)$ *over* $[s,r]$. Note that $\int_s^r f_t(X_t)\,\mathrm{d}t$ is a real variable because every Riemann integral is real-valued. Furthermore, $\int_s^r f_t(X_t)\,\mathrm{d}t$ turns out to be bounded, so it is a gamble. This essentially holds because every piece-wise continuous family $(f_t)_{t\in[s,r]}$ is bounded in the following sense.

**Lemma 6.31.** *Consider time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. Then for any piece-wise continuous family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$,*

$$-\infty < \inf\big\{\min f_t \colon t \in [s,r]\big\} \leq \sup\big\{\max f_t \colon t \in [s,r]\big\} < +\infty.$$

*Proof.* Observe that for all $t$ in $[s,r]$, $-\|f_t\| \leq \min f_t \leq \max f_t \leq \|f_t\|$. These inequalities imply the middle inequality of the statement. The outer inequalities hold because the piece-wise continuous family $(f_t)_{t\in[s,r]}$ is (uniformly) bounded, in the sense that $\sup\big\{\|f_t\| \colon t \in [s,r]\big\} < +\infty$; this well-known property follows immediately from (Schechter, 1997, Definition 19.28 (C)) and (Rudin, 1976, Theorem 4.15). □

That $\int_s^r f_t(X_t)\,\mathrm{d}t$ is bounded follows immediately from the preceding lemma and some properties of the Riemann integral.

**Corollary 6.32.** *Consider time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, and a piece-wise continuous family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$. Then*

$$(r-s)\inf\big\{\min f_t \colon t \in [s,r]\big\} \leq \int_s^r f_t(X_t)\,\mathrm{d}t \leq (r-s)\sup\big\{\max f_t \colon t \in [s,r]\big\},$$

*so $\int_s^r f_t(X_t)\,\mathrm{d}t$ is bounded.*

*Proof.* Fix an arbitrary path $\omega$ in $\Omega$, and let

$$\gamma_- := \inf\big\{\min f_t \colon t \in [s,r]\big\} \quad \text{and} \quad \gamma_+ := \sup\big\{\max f_t \colon t \in [s,r]\big\}.$$

Note that for all $t$ in $[s,r]$,

$$-\infty < \gamma_- \leq f_t\big(\omega(t)\big) \leq \gamma_+ < +\infty,$$

where the outer inequalities hold due to Lemma 6.31. It follows from these inequalities, (RI3)296 and (RI4)296 that

$$-\infty < (r-s)\gamma_- \leq \int_s^r f_t\big(\omega(t)\big)\,\mathrm{d}t \leq (r-s)\gamma_+ < +\infty.$$

As $\omega$ is an arbitrary path in $\Omega$, this proves the statement. □

What is left is to verify that the gamble $\int_s^r f_t(X_t)\,\mathrm{d}t$ is an idealised variable, in the sense that it is the point-wise limit of a uniformly bounded sequence of simple variables. To do so, we turn to approximating Riemann sums.

### Approximating Riemann sum

Recall from Definition 6.25296 that the Riemann integral of $f_\bullet \circ \omega\colon [s,r] \to \mathbb{R}$ is essentially defined as the limit of approximating Riemann sums, where every Riemann sum corresponds to a (tagged) grid $v$ over $[s,r]$. For any time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, any piece-wise continuous family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$ and any grid $v = (t_0,\dots,t_n)$ over $[s,r]$, we define the corresponding *approximating Riemann sum* as

$$\langle f_\bullet\rangle_v := \sum_{k=1}^{n} (t_k - t_{k-1}) f_{t_k}(X_{t_k}). \tag{6.27}$$

In this definition, we could replace $f_{t_k}(X_{t_k})$ with $f_{s_k}(X_{s_k})$ where $s_k$ is any time point in $[t_{k-1}, t_k]$. We have chosen to always use the end point of this interval because this ensures that $\langle f_\bullet\rangle_v$ has a sum-product representation over $v$; note that this would also be the case if we were to use the starting point $t_{k-1}$ instead of the end point $t_k$ of the subinterval $[t_{k-1}, t_k]$.

**Lemma 6.33.** *Consider a sequence of time points $u$ in $\mathscr{U}$, time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$, and a piece-wise continuous family $(f_t)_{t\in\mathbb{R}_{\geq 0}}$ of gambles on $\mathscr{X}$. Then for any grid $v$ over $[s,r]$, the corresponding approximating Riemann sum $\langle f_\bullet\rangle_v$ has a sum-product representation over $v$, so it is an $\mathscr{F}_u$-simple variable; furthermore,*

$$(r-s)\inf\big\{\min f_t\colon t\in[s,r]\big\} \leq \langle f_\bullet\rangle_v \leq (r-s)\sup\big\{\min f_t\colon t\in[s,r]\big\}.$$

*Proof.* It follows immediately from Eq. (6.27) that

$$\langle f_\bullet\rangle_v = \sum_{k=0}^{n} g_k\big(X_{t_k}\big) \prod_{\ell=1}^{k-1} h_\ell\big(X_{t_\ell}\big),$$

with $g_0 := 0$ and, for all $k$ in $\{1,\dots,n\}$, $g_k := (t_k - t_{k-1}) f_{t_k}$ and $h_{k-1} := 1$. Hence, $\langle f_\bullet\rangle_v$ has a sum-product representation over $v$. Observe that $v \succcurlyeq u$ because, by assumption, $v$ is a grid over $[s,r]$ and $\max u \leq s$. Hence, it follows from Lemma 4.6164 that $\langle f_\bullet\rangle_v$ is $\mathscr{F}_u$-simple. That the inequalities of the statement hold is a matter of straightforward verification. □

### The point-wise limit of approximating Riemann sums

Finally, we verify that the integral $\int_s^r f_t(X_t)\,\mathrm{d}t$ of the family $(f_t)_{t\in[s,r]}$ belongs to our extended domain. To do so, we argue that it is the point-wise limit of a uniformly bounded sequence of approximating Riemann sums that corresponds to a sequence of grids over $[s,r]$ with vanishing grid width.

**Lemma 6.34.** *Consider a sequence of time points $u$ in $\mathscr{U}$, time points $s,r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$ and a piece-wise continuous family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$. Let $(v_n)_{n\in\mathbb{N}}$ be a sequence of grids over $[s,r]$ such that $\lim_{n\to+\infty}\Delta(v_n) = 0$. Then $(\langle f_\bullet\rangle_{v_n})_{n\in\mathbb{N}}$ is a uniformly bounded sequence of $\mathscr{F}_u$-simple variables that converges point-wise to $\int_s^r f_t(X_t)\,\mathrm{d}t$, so $\int_s^r f_t(X_t)\,\mathrm{d}t$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$.*

*Proof.* Let $f := \int_s^r f_t(X_t)\,\mathrm{d}t$. Recall from Lemma 6.33 that for all $n$ in $\mathbb{N}$, $\langle f_\bullet\rangle_{v_n}$ is an $\mathscr{F}_u$-simple variable with

$$(r-s)\inf\{\min f_t\colon t\in[s,r]\} \leq \langle f_\bullet\rangle_{v_n} \leq (r-s)\sup\{\min f_t\colon t\in[s,r]\}.$$

Furthermore, because $\lim_{n\to+\infty}\Delta(v_n) = 0$, it follows immediately from Definition 6.25296 that $\text{p-w}\lim_{n\to+\infty}\langle f_\bullet\rangle_{v_n} = f$. Consequently, $\big(\langle f_\bullet\rangle_{v_n}\big)_{n\in\mathbb{N}}$ is a uniformly bounded sequence of $\mathscr{F}_u$-simple variables that converges point-wise to $f$, so $f$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$. □

## 6.4.3 The integral of a Lipschitz continuous family of gambles

So far, we have established that the integral $\int_s^r f_t(X_t)\,\mathrm{d}t$ of the family $(f_t)_{t\in[s,r]}$ is the point-wise limit of a uniformly bounded sequence of simple variables. Next, we investigate whether this limit behaviour carries over to the lower and upper expectations with respect to an imprecise jump process. As will become clear in Corollary 6.39304 further on, it suffices that the family $(f_t)_{t\in[s,r]}$ is Lipschitz continuous – with respect to the supremum norm $\|\bullet\|$ – in the sense of Schechter (1997, Definition 18.2).

**Definition 6.35.** Consider time points $s,r$ in $\mathbb{R}_{\geq 0}$ such that $s\leq r$. The family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$ is *Lipschitz continuous* if there is some $\kappa$ in $\mathbb{R}_{\geq 0}$ such that for all $t_1$ and $t_2$ in $[s,r]$, $\|f_{t_1} - f_{t_2}\| \leq \kappa|t_1 - t_2|$. Whenever this is the case, we call $\kappa$ a *Lipschitz constant.*

Note that Lipschitz continuity implies (ordinary) continuity, and therefore also piece-wise continuity. Hence, it follows from Proposition 6.30298 that the Riemann integral $\int_s^r f_t(X_t)\,\mathrm{d}t$ of a Lipschitz continuous family $(f_t)_{t\in[s,r]}$ is well-defined.

**Corollary 6.36.** *Consider time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s\leq r$, and a family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$ that is Lipschitz continuous. Then the* *corresponding Riemann integral $\int_s^r f_t(X_t)\,\mathrm{d}t$ is well-defined because for all $\omega$ in $\Omega$, the function*

$$f_\bullet \circ \omega\colon [s,r] \to \mathbb{R}\colon t \mapsto f_t\big(\omega(t)\big)$$

*is Riemann integrable.*

*Proof.* The Lipschitz continuous family $(f_t)_{t\in[s,r]}$ is trivially (piece-wise) continuous, so the statement is a special case of Proposition 6.30298. □

Whenever the family $(f_t)_{t\in[s,r]}$ is Lipschitz continuous, we can use the Lipschitz constant $\kappa$ to bound the difference between between the Riemann integral $\int_s^r f_t(X_t)\,\mathrm{d}t$ and the approximating Riemann sum $\langle f_\bullet\rangle_v$ that corresponds to a grid $v$ over $[s,r]$.

**Lemma 6.37.** *Consider time points $s,r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, a Lipschitz continuous family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$ with Lipschitz constant $\kappa$, and let*

$$\gamma := \sup\big\{\max f_t\colon t \in [s,r]\big\} - \inf\big\{\min f_t\colon t\in[s,r]\big\}.$$

*Then for any grid $v$ over $[s,r]$,*

$$\left|\int_s^r f_t(X_t)\,\mathrm{d}t - \langle f_\bullet\rangle_v\right| \leq \Delta(v)\big((r-s)\kappa + \gamma\eta_{[s,r]}\big).$$

*Proof.* We enumerate the time points in $v$ as $(t_0,\dots,t_n)$. Then by definition of $\langle f_\bullet\rangle_v$,

$$\left|\int_s^r f_t(X_t)\,\mathrm{d}t - \langle f_\bullet\rangle_v\right| = \left|\int_s^r f_t(X_t)\,\mathrm{d}t - \sum_{k=1}^n f_{t_k}(X_{t_k})(t_k - t_{k-1})\right|$$

It follows from this equality, (RI5)296 and the triangle inequality that

$$\begin{aligned}\left|\int_s^r f_t(X_t)\,\mathrm{d}t - \langle f_\bullet\rangle_v\right| &= \left|\sum_{k=1}^n \int_{t_{k-1}}^{t_k} f_t(X_t)\,\mathrm{d}t - \sum_{k=1}^n f_{t_k}(X_{t_k})(t_k - t_{k-1})\right| \\ &\leq \sum_{k=1}^n \left|\int_{t_{k-1}}^{t_k} f_t(X_t)\,\mathrm{d}t - f_{t_k}(X_{t_k})(t_k - t_{k-1})\right|. \end{aligned} \tag{6.28}$$

We now investigate the terms of the sum on the right-hand side of this inequality individually. To this end, we fix some $k$ in $\{1,\dots,n\}$ and $\omega$ in $\Omega$. Note that by (RI3)296,

$$\left|\int_{t_{k-1}}^{t_k} f_t\big(\omega(t)\big)\,\mathrm{d}t - f_{t_k}\big(\omega(t_k)\big)(t_{k+1} - t_k)\right| = \left|\int_{t_{k-1}}^{t_k} f_t\big(\omega(t)\big) - f_{t_k}\big(\omega(t_k)\big)\,\mathrm{d}t\right|. \tag{6.29}$$

First, we consider the case that $\omega$ is constant over $[t_{k-1},t_k]$, meaning that $\eta_{[t_{k-1},t_k]}(\omega) = 0$. Let $x_k := \omega(t_k)$; then for all $t$ in $[t_{k-1},t_k]$, $\omega(t) = \omega(t_k) = x_k$, and hence $f_t\big(\omega(t)\big) = f_t(x_k)$. Because $(f_t)_{t\in[s,r]}$ is Lipschitz continuous with Lipschitz constant $\kappa$,

$$\big(\forall t \in [t_{k-1},t_k]\big)\ \|f_t - f_{t_k}\| \leq \kappa(t_k - t) \leq \kappa(t_k - t_{k-1}),$$

and therefore

$$\big(\forall t \in [t_{k-1}, t_k]\big)\ \big|f_t\big(\omega(t)\big) - f_{t_k}\big(\omega(t_k)\big)\big| = \big|f_t(x_k) - f_{t_k}(x_k)\big| \le \|f_t - f_{t_k}\| \\ \le \kappa(t_k - t_{k-1}). \quad (6.30)$$

It follows from Eqs. (6.29) and (6.30) that

$$\left|\int_{t_{k-1}}^{t_k} f_t\big(\omega(t)\big)\,\mathrm{d}t - f_{t_k}\big(\omega(t_k)\big)(t_{k+1} - t_k)\right| \le \kappa(t_k - t_{k-1})^2.$$

where for the inequality we also used (RI4)296 and (RI3)296. Because $(t_k - t_{k-1}) \le \Delta(v)$ and $\gamma\eta_{[t_{k-1},t_k]}(\omega) = 0$, it follows that

$$\left|\int_{t_{k-1}}^{t_k} f_t\big(\omega(t)\big)\,\mathrm{d}t - f_{t_k}\big(\omega(t_k)\big)(t_{k+1} - t_k)\right| \le \Delta(v)\big((t_k - t_{k-1})\kappa + \gamma\eta_{[t_{k-1},t_k]}(\omega)\big).$$

Second, we consider the case that $\omega$ is not constant over $[t_{k-1}, t_k]$, in the sense that $\eta_{[t_{k-1},t_k]}(\omega) > 0$. Observe that

$$\big(\forall t \in [t_{k-1}, t_k]\big)\ \big|f_t\big(\omega(t)\big) - f_{t_k}\big(\omega(t_k)\big)\big| \le \gamma.$$

As before, it follows from Eq. (6.29) and this inequality that

$$\left|\int_{t_{k-1}}^{t_k} f_t\big(\omega(t)\big)\,\mathrm{d}t - f_{t_k}\big(\omega(t_k)\big)(t_{k+1} - t_k)\right| \le (t_k - t_{k-1})\gamma \le \Delta(v)\gamma,$$

where for the second equality we used that $(t_k - t_{k-1}) \le \Delta(v)$. Hence, and because $\eta_{[t_{k-1},t_k]}(\omega) \ge 1$ and $\kappa(t_k - t_{k-1}) \ge 0$, it follows that

$$\left|\int_{t_{k-1}}^{t_k} f_t\big(\omega(t)\big)\,\mathrm{d}t - f_{t_k}\big(\omega(t_k)\big)(t_{k+1} - t_k)\right| \le \Delta(v)\gamma\eta_{[t_{k-1},t_k]}(\omega) \\ \le \Delta(v)\big((t_k - t_{k-1})\kappa + \gamma\eta_{[t_{k-1},t_k]}(\omega)\big).$$

Thus, we have shown that for all $k$ in $\{1, \dots, n\}$,

$$\left|\int_{t_{k-1}}^{t_k} f_t(X_t)\,\mathrm{d}t - f_{t_k}(X_{t_k})(t_k - t_{k-1})\right| \le \Delta(v)\big((t_k - t_{k-1})\kappa + \gamma\eta_{[t_{k-1},t_k]}(\omega)\big).$$

It follows from this and Eq. (6.28) that

$$\left|\int_s^r f_t(X_t)\,\mathrm{d}t - \langle f_\bullet\rangle_v\right| \le \sum_{k=1}^n \Delta(v)\big((t_k - t_{k-1})\kappa + \gamma\eta_{[t_{k-1},t_k]}(\omega)\big) \\ = \Delta(v)\big((r-s)\kappa + \gamma\eta_{[s,r](\omega)}\big),$$

where for the equality we used that $\sum_{k=1}^n (t_k - t_{k-1}) = t_n - t_0 = r - s$ and that $\sum_{k=1}^n \eta_{[t_{k-1},t_k]}(\omega) = \eta_{[s,r](\omega)}$. □

By virtue of Proposition 6.7279, the bound in Lemma 6.37 induces an upper bound on the difference of the lower (or upper) expectation of the Riemann integral $\int_s^r f_t(X_t)\,\mathrm{d}t$ and the lower (or upper) expectation of the approximating Riemann sum $\langle f_\bullet\rangle_v$ for any grid $v$ over $[s, r]$.

**Proposition 6.38.** *Consider a non-empty and bounded set $\mathcal{Q}$ of rate operators, and an imprecise jump process $\mathscr{P}$ such that $\mathscr{P} \subseteq \mathbb{P}_{\mathcal{Q}}$. Fix some state history $\{X_u = x_u\}$ in $\mathscr{H}$, time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$ and a Lipschitz continuous family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$ with Lipschitz constant $\kappa$, and let $f := \int_s^r f_t(X_t)\,\mathrm{d}t$ and*

$$\gamma := \sup\{\max f_t \colon t \in [s,r]\} - \inf\{\min f_t \colon t \in [s,r]\}.$$

*Then for any grid $v$ over $[s,r]$,*

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) - \underline{E}_{\mathscr{P}}\big(\langle f_\bullet\rangle_v \,\big|\, X_u = x_u\big)\right| \leq \Delta(v)(r-s)\Big(\kappa + \frac{\gamma}{2}\|\mathcal{Q}\|_{\mathrm{op}}\Big)$$

*and*

$$\left|\overline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) - \overline{E}_{\mathscr{P}}\big(\langle f_\bullet\rangle_v \,\big|\, X_u = x_u\big)\right| \leq \Delta(v)(r-s)\Big(\kappa + \frac{\gamma}{2}\|\mathcal{Q}\|_{\mathrm{op}}\Big).$$

*In particular, this holds for $\mathscr{P} = \mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathcal{Q}}$, $\mathscr{P} = \mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathcal{Q}}$ and $\mathscr{P} = \mathbb{P}_{\mathscr{M},\mathcal{Q}}$, with $\mathscr{M}$ a non-empty set of initial mass functions.*

*Proof.* Due to Lemmas 6.34$_{301}$ and 6.37$_{302}$, the statement follows immediately from Proposition 6.7$_{279}$ with $\alpha = (r-s)\kappa$ and $\beta = 0$. □

Crucially, the preceding result implies that the lower (and upper) expectation of the Riemann integral $\int_s^r f_t(X_t)\mathrm{d}t$ is equal to the limit of the lower (and upper) expectation of the approximating Riemann sums $\langle f_\bullet\rangle_{v_n}$ that correspond to a sequence $(v_n)_{n\in\mathbb{N}}$ of grids over $[s,r]$ with vanishing grid width. In the particular case of the 'Riemann integral of a Lipschitz continuous family $(f_t)_{t\in[s,r]}$', we can therefore strengthen Corollary 5.34$_{241}$ as follows.

**Corollary 6.39.** *Consider a non-empty bounded set $\mathcal{Q}$ of rate operators, and an imprecise jump process $\mathscr{P}$ such that $\mathscr{P} \subseteq \mathbb{P}_{\mathcal{Q}}$. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$ and a Lipschitz continuous family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$. Then for any sequence $(v_n)_{n\in\mathbb{N}}$ of grids over $[s,r]$ with $\lim_{n\to+\infty}\Delta(v_n) = 0$,*

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}\bigg(\int_s^r f_t(X_t)\,\mathrm{d}t \,\bigg|\, X_u = x_u\bigg) = \lim_{n\to+\infty} \underline{E}_{\mathscr{P}}\big(\langle f_\bullet\rangle_{v_n} \,\big|\, X_u = x_u\big)$$

*and*

$$\overline{E}^{\mathrm{D}}_{\mathscr{P}}\bigg(\int_s^r f_t(X_t)\,\mathrm{d}t \,\bigg|\, X_u = x_u\bigg) = \lim_{n\to+\infty} \overline{E}_{\mathscr{P}}\big(\langle f_\bullet\rangle_{v_n} \,\big|\, X_u = x_u\big).$$

*In particular, this holds for $\mathscr{P} = \mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathcal{Q}}$, $\mathscr{P} = \mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathcal{Q}}$ and $\mathscr{P} = \mathbb{P}_{\mathscr{M},\mathcal{Q}}$, with $\mathscr{M}$ a non-empty set of initial mass functions.*

*Proof.* Follows immediately from Corollary 5.34$_{241}$, Lemma 6.34$_{301}$ and Proposition 6.38. □

### 6.4.4 The special case of temporal averages

To conclude this section on variables that are defined through Riemann integrals, we take a closer look at the special case of a constant family $(f_t)_{t\in[s,r]} = (f)_{t\in[s,r]}$ of gambles on $\mathscr{X}$, so at the integral of $f(X_t)$. As we have already mentioned at the beginning of Section $6.4_{295}$, a common example of such variables are *occupancy times* – see (Kulkarni, 2011, Section 4.5) or (Iosifescu, 1980, Section 8.6). In particular, for any subset $G$ of $\mathscr{X}$ and any time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, $\int_s^r \mathbb{I}_G(X_t)\,\mathrm{d}t$ is the length of time that (the state of) the system occupies $G$. Here, we shall focus on general integrals of $f(X_t)$, albeit in the form of temporal averages.

For any gamble $f$ on $\mathscr{X}$ and any time points $s$ and $r$ in $\mathbb{R}_{\geq 0}$ such that $s < r$, we define the *temporal average of* $f(X_t)$ *over* $[s,r]$ as

$$\llbracket f \rrbracket_{[s,r]} \coloneqq \frac{1}{r-s}\int_s^r f(X_t)\,\mathrm{d}t \colon \Omega \to \mathbb{R} \colon \omega \mapsto \frac{1}{r-s}\int_s^r f\big(\omega(t)\big)\,\mathrm{d}t;$$

additionally, for any grid $v = (t_0, \dots, t_n)$ over $[s,r]$, we let

$$\llbracket f \rrbracket_v \coloneqq \frac{1}{r-s}\sum_{k=1}^{n}(t_k - t_{k-1}) f(X_{t_k}).$$

To verify that the variable $\llbracket f \rrbracket_{[s,r]}$ is well-defined, we let $(f_t)_{t\in[s,r]}$ be the family of gambles on $\mathscr{X}$ that is defined by $f_t \coloneqq f/(r-s)$ for all $t$ in $[s,r]$. Because the family $(f_t)_{t\in[s,r]}$ is constant, it is trivially Lipschitz continuous with Lipschitz constant $\kappa = 0$, so we may specialise all of the results in Section $6.4.3_{301}$. For example, it follows from Corollary $6.36_{301}$ that

$$\llbracket f \rrbracket_{[s,r]} = \int_s^r \frac{1}{r-s} f(X_t)\,\mathrm{d}t = \int_s^r f_t(X_t)\,\mathrm{d}t \tag{6.31}$$

is well-defined. Furthermore, Proposition 6.38↶ specialises as follows.

**Corollary 6.40.** *Consider a non-empty and bounded set $\mathscr{Q}$ of rate operators, and an imprecise jump process $\mathscr{P}$ such that $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{Q}}$. Fix some state history $\{X_u = x_u\}$ in $\mathscr{H}$, time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$ and a gamble $f$ on $\mathscr{X}$. Then for any grid $v$ over $[s,r]$,*

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(\llbracket f \rrbracket_{[s,r]} \,|\, X_u = x_u) - \underline{E}_{\mathscr{P}}\big(\llbracket f \rrbracket_v \,\big|\, X_u = x_u\big)\right| \leq \Delta(v)\|f\|_{\mathrm{c}}\|\mathscr{Q}\|_{\mathrm{op}}$$

*and*

$$\left|\overline{E}^{\mathrm{D}}_{\mathscr{P}}(\llbracket f \rrbracket_{[s,r]} \,|\, X_u = x_u) - \overline{E}_{\mathscr{P}}\big(\llbracket f \rrbracket_v \,\big|\, X_u = x_u\big)\right| \leq \Delta(v)\|f\|_{\mathrm{c}}\|\mathscr{Q}\|_{\mathrm{op}}.$$

*In particular, this holds for $\mathscr{P} = \mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathscr{P} = \mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathscr{P} = \mathbb{P}_{\mathscr{M},\mathscr{Q}}$, with $\mathscr{M}$ a non-empty set of initial mass functions.*

*Proof.* For all $t$ in $[s,r]$, we let $f_t \coloneqq f/(r-s)$. Then the family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$ is Lipschitz continuous with Lipschitz constant $\kappa \coloneqq 0$. Furthermore, we observe that $[\![f]\!]_{[s,r]} = \int_s^r f_t(X_t)\,\mathrm{d}t$ and that, for any grid $v$ over $[s,r]$, $[\![f]\!]_v = \langle f_\bullet \rangle_v$. Finally, it is easy to see that

$$\gamma := \sup\{\max f_t \colon t \in [s,r]\} - \inf\{\min f_t \colon t \in [s,r]\} = \frac{\max f - \min f}{r-s} = \frac{2\|f\|_c}{r-s}.$$

With all this in mind, the statement follows immediately from Proposition 6.38304. □

## Long-term temporal averages

In many applications, for example in Power Network Example 6.51319 and Section 8.2.4417 further on, we are interested in the (lower and upper) expected 'long-term temporal average of $f(X_t)$'. Formally, this means that we are interested in the (lower and upper) expectation of

$$[\![f]\!]_{[s,r]} = \frac{1}{r-s}\int_s^r f(X_t)\,\mathrm{d}t$$

as $r$ recedes to $+\infty$. Instead, one could also try to define 'the long-term temporal average of $f(X_t)$' as a point-wise limit, that is, as the real variable that maps every path $\omega$ in $\Omega$ to the limit of $[\![f]\!]_{[s,r]}$ as $r$ recedes to $+\infty$. Unfortunately, this approach does not work because the point-wise limit may not exist. We will once again use our running example to illustrate this.

**Joseph's Example 6.41.** We are interested in the proportion of time that Joseph's machine displays heads. That is, we want to determine the 'long-term temporal average of $\mathbb{I}_{\mathrm{H}}(X_t)$'. We now set out to explicitly construct a path $\omega$ in $\Omega$ for which

$$[\![\mathbb{I}_{\mathrm{H}}]\!]_{[0,r]}(\omega) = \frac{1}{r}\int_0^r \mathbb{I}_{\mathrm{H}}\big(\omega(t)\big)\,\mathrm{d}t$$

does not converge as $r$ recedes to $+\infty$.

To this end, we let $r_0 := 0$ and $r_1 := 1$, and for any natural number $n$, we let $r_{n+1} := 4 \times 3^{n-1}$. Note that the sequence $(r_n)_{n\in\mathbb{N}}$ is increasing with $\lim_{n\to+\infty} r_n = +\infty$.[2] We denote by $\omega$ the path that starts with heads, jumps to tails at $r_1 = 1$, jumps back to heads at $r_2 = 4$, and so on; that is, we let

$$\omega \colon \mathbb{R}_{\geq 0} \to \mathscr{X} \colon t \mapsto \omega(t) := \begin{cases} \mathrm{H} & \text{if } t \in [r_{2n-2}, r_{2n-1}[ \text{ for some } n \in \mathbb{N}, \\ \mathrm{T} & \text{if } t \in [r_{2n-1}, r_{2n}[ \text{ for some } n \in \mathbb{N}. \end{cases}$$

By construction, the path $\omega$ is càdlàg. Furthermore, it is not all too difficult – but slightly cumbersome – to verify that as a function of $r$,

$$[\![\mathbb{I}_{\mathrm{H}}]\!]_{[0,r]}(\omega) = \frac{1}{r}\int_0^r \mathbb{I}_{\mathrm{H}}\big(\omega(t)\big)\,\mathrm{d}t$$

[2] The sequence $(r_n)_{n\in\mathbb{N}}$ is catalogued as sequence A052156 in The On-Line Encyclopedia of Integer Sequences.

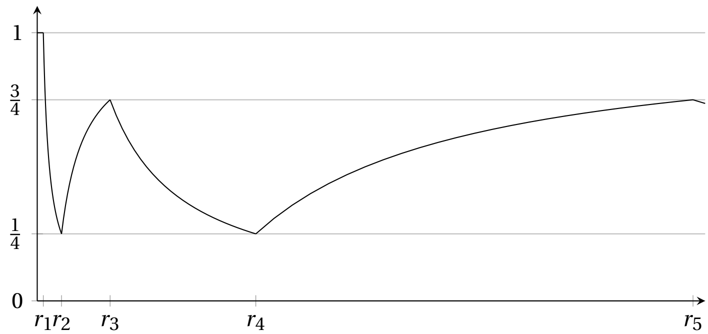



**Figure 6.1** Graph of $\frac{1}{r}\int_0^r \mathbb{I}_{\mathrm{H}}\big(\omega(t)\big)\,\mathrm{d}t$ as function of $r$ for the path $\omega$ as defined in Joseph's Example 6.41

evolves as depicted in Fig. 6.1: $[\![\mathbb{I}_{\mathrm{H}}]\!]_{[0,r]}(\omega)$ is 1 on $]r_0, r_1]$, then decreases on $[r_1, r_2]$ to end up in 1/4 at $r_2$, and subsequently increases on $[r_2, r_3]$ to 3/4 at $r_3$; further along the $r$-axis, this pattern of decreasing to 1/4 and increasing to 3/4 repeats *ad infinitum*. In other words, as $r$ grows, $[\![\mathbb{I}_{\mathrm{H}}]\!]_{[0,r]}(\omega)$ oscillates between 1/4 and 3/4. Consequently,

$$\frac{1}{4} = \liminf_{r\to+\infty} [\![\mathbb{I}_{\mathrm{H}}]\!]_{[0,r]}(\omega) < \limsup_{r\to+\infty} [\![\mathbb{I}_{\mathrm{H}}]\!]_{[0,r]}(\omega) = \frac{3}{4},$$

so $\lim_{r\to+\infty} [\![\mathbb{I}_{\mathrm{H}}]\!]_{[0,r]}(\omega)$ does not exist.

Because 'the long-term temporal average of $f(X_t)$' cannot be expressed as a point-wise limit, we need to go about this in a different manner. This is why, instead of looking at the limit of $[\![f]\!]_{[s,r]}$ directly, we look at the limit of the (lower and upper) expectation of $[\![f]\!]_{[s,r]}$ as $r$ recedes to $+\infty$. This limit has been thoroughly studied for homogeneous and Markovian jump processes, so let us start there.

Consider a homogeneous Markovian jump process $P_{p_0,Q}$ with an ergodic rate operator $Q$. Then it is well-known that for all $f$ in $\mathbb{G}(\mathscr{X})$,

$$\lim_{r\to+\infty} E_{p_0,Q}\left(\frac{1}{r}\int_0^r f(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right) = E_{\lim}(f) \quad \text{for all } x \in \mathscr{X}, \tag{6.32}$$

where $E_{\lim}$ is the so-called limit expectation of $Q$ that maps $f$ to the value of the constant function $\lim_{r\to+\infty} e^{rQ} f$ – see also Section 7.3373 in Chapter 7337 further on. The proper result, known as the *Point-Wise Ergodic Theorem*, is actually stronger; Iosifescu (1980, Section 8.6.6) and Norris (1997, Theorem 3.8.1) formulate it as follows: the event

$$\left\{\omega \in \Omega\colon \liminf_{r\to+\infty} \frac{1}{r}\int_0^r f\big(\omega(t)\big)\,\mathrm{d}t = \limsup_{r\to+\infty} \frac{1}{r}\int_0^r f\big(\omega(t)\big)\,\mathrm{d}t = E_{\lim}(f)\right\}$$

has probability one – for a suitable extension of $P_{p_0,Q}$. Note that the possible lack of convergence of $\frac{1}{r}\int_0^r f(X_t)\,\mathrm{d}t$ is dealt with in this case.

So, do these results generalise to imprecise Markovian jump processes? Before even thinking of trying to answer this question, it is wise to consult the literature regarding the discrete-time setting. De Cooman et al. (2016, Theorem 32) have shown that the discrete-time counterpart of the second result – so the point-wise ergodic theorem for Markov chains (see Norris, 1997, Theorem 1.10.2) – generalises to imprecise Markov chains. However, because they assume the framework of game-theoretic stochastic processes and we do not work in this framework, their result is not immediately relevant to our setting. Later, T'Joens et al. (2021, Theorem 14) have shown that the discrete-time counterpart of Eq. (6.32)$_{\curvearrowleft}$ also generalises to imprecise Markov chains; their result *is* relevant to our setting, because their imprecise Markov chains are defined as sets of 'consistent discrete-time stochastic processes'. One crucial difference is that they only show that the limit of the lower expected temporal average exists and does not depend on the initial state, but *not* that it is equal to the limit lower expectation; their reason for doing so is simple: the limit of the lower expected temporal average can be (strictly) greater than the limit of the lower expectation (see T'Joens et al., 2021, Example 2). They do show, however, that the limit of the lower expected temporal average is the same for the discrete-time counterparts of $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$.

Because nearly all results for imprecise Markov chains translate to imprecise homogeneous Markovian jump processes, we are led to believe that these results regarding the convergence of expected temporal averages might very well generalise to imprecise Markovian jump processes. We have to draw the line somewhere though, so we will not investigate this in any theoretical manner. We will, however, provide anecdotal evidence for our conjecture; of course, we turn to our running example for this.

**Joseph's Example 6.42.** Recall from Joseph's Example 4.13$_{171}$ that Eleanor's beliefs about Joseph's machine are accurately modelled by $\underline{E}_{\mathscr{M},\mathscr{Q}_2}$, with $\mathscr{M} = \{\mathbb{I}_{\mathrm{H}}\}$ as in Joseph's Example 4.3$_{161}$ and

$$\mathscr{Q}_2 = \left\{\begin{pmatrix} -\lambda_{\mathrm{H}} & \lambda_{\mathrm{H}} \\ \lambda_{\mathrm{T}} & -\lambda_{\mathrm{T}} \end{pmatrix} : \lambda_{\mathrm{H}}, \lambda_{\mathrm{T}} \in [\underline{\lambda}, \overline{\lambda}]\right\}$$

as in Joseph's Example 4.4$_{163}$. Note that $\mathscr{Q}_2$ is non-empty, bounded and has separately specified rows, and recall from Joseph's Example 4.30$_{192}$ that $\underline{Q} := \underline{Q}_{\mathscr{Q}}$ is ergodic if and only if $\overline{\lambda} + \underline{\lambda} > 0$.

In this example, we assume that Eleanor is interested in the long-term expected proportion of time that Joseph's machine displays heads. That is, she is interested in the limit of

$$\underline{E}^{\mathrm{D}}_{\mathscr{M},\mathscr{Q}_2}\left(\frac{1}{r}\int_0^r \mathbb{I}_{\mathrm{H}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right) \quad \text{and} \quad \overline{E}^{\mathrm{D}}_{\mathscr{M},\mathscr{Q}_2}\left(\frac{1}{r}\int_0^r \mathbb{I}_{\mathrm{H}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right)$$

as $r$ recedes to $+\infty$, with $x$ in $\mathscr{X}$. By virtue of Corollary 6.40[305] and Algorithm 4.2[168], we can compute these lower and upper expectations up to arbitrary precision – note though that it is arguably more efficient to do this with the recursive method in Theorem 6.50[318] further on. Say we want to approximate the lower and upper expected temporal average with a maximum error of $\epsilon$, with $\epsilon$ a positive real number. Then by Corollary 6.40[305], we need to construct a grid $v = (t_0, \dots, t_n)$ over $[0, r]$ such that

$$\Delta(v) \|\mathbb{I}_H\|_{\mathrm{c}} \|\underline{\mathcal{Q}}_2\|_{\mathrm{op}} \leq \epsilon.$$

As in Joseph's Example 6.5[277], we divide the interval $[0, r]$ into $n$ subintervals of equal length. From the preceding inequality, it follows immediately that it suffices to use

$$n := \left\lceil \frac{r \|\mathbb{I}_H\|_{\mathrm{c}} \|\underline{\mathcal{Q}}_2\|_{\mathrm{op}}}{\epsilon} \right\rceil \tag{6.33}$$

subintervals to attain the desired accuracy; that is, with $v := (t_0, \dots, t_n)$ the grid over $[0, r]$ such that $t_k := {}^{kr}/{}_{n}$ for all $k$ in $\{0, \dots, n\}$,

$$\left| \underline{E}^{\mathrm{D}}_{\mathscr{M}, \mathcal{Q}_2}\left( \frac{1}{r} \int_0^r \mathbb{I}_{\mathrm{H}}(X_t) \, \mathrm{d}t \,\middle|\, X_0 = x \right) - \underline{E}_{\mathscr{M}, \mathcal{Q}_2}(\llbracket \mathbb{I}_{\mathrm{H}} \rrbracket_v \,|\, X_0 = x) \right| \leq \epsilon$$

and

$$\left| \overline{E}^{\mathrm{D}}_{\mathscr{M}, \mathcal{Q}_2}\left( \frac{1}{r} \int_0^r \mathbb{I}_{\mathrm{H}}(X_t) \, \mathrm{d}t \,\middle|\, X_0 = x \right) - \overline{E}_{\mathscr{M}, \mathcal{Q}_2}(\llbracket \mathbb{I}_{\mathrm{H}} \rrbracket_v \,|\, X_0 = x) \right| \leq \epsilon.$$

Crucially, we can compute $\underline{E}_{\mathscr{M}, \mathcal{Q}_2}(\llbracket \mathbb{I}_{\mathrm{H}} \rrbracket_v \,|\, X_0 = x)$ and $\overline{E}_{\mathscr{M}, \mathcal{Q}_2}(\llbracket \mathbb{I}_{\mathrm{H}} \rrbracket_v \,|\, X_0 = x)$ because $\llbracket \mathbb{I}_{\mathrm{H}} \rrbracket_v$ has a sum-product representation over $v$. In particular, since $\mathcal{Q}_2$ has separately specified rows, we can use Algorithm 4.2[168] to compute $\underline{E}_{\mathscr{M}, \mathcal{Q}_2}(\llbracket \mathbb{I}_{\mathrm{H}} \rrbracket_v \,|\, X_0 = x)$. The same is true for $\overline{E}_{\mathscr{M}, \mathcal{Q}_2}(\llbracket \mathbb{I}_{\mathrm{H}} \rrbracket_v \,|\, X_0 = x)$ because of Lemma 4.7[165] and because, by conjugacy, $\overline{E}_{\mathscr{M}, \mathcal{Q}_2}(\llbracket \mathbb{I}_{\mathrm{H}} \rrbracket_v \,|\, X_0 = x) = -\underline{E}_{\mathscr{M}, \mathcal{Q}_2}(-\llbracket \mathbb{I}_{\mathrm{H}} \rrbracket_v \,|\, X_0 = x)$.

As in Joseph's Example 4.21[181], we take $\underline{\lambda} := 1$ and $\overline{\lambda} := {}^{3}/{}_{2}$; then

$$\|\underline{\mathcal{Q}}_2\|_{\mathrm{op}} = \sup\left\{ \|Q\|_{\mathrm{op}} \colon Q \in \mathcal{Q}_2 \right\} = \sup\left\{ 2 \max\{\lambda_{\mathrm{H}}, \lambda_{\mathrm{T}}\} \colon \lambda_{\mathrm{H}}, \lambda_{\mathrm{T}} \in [\underline{\lambda}, \overline{\lambda}] \right\} = 2\overline{\lambda} = 3,$$

where the first equality is the definition of $\|\underline{\mathcal{Q}}_2\|_{\mathrm{op}}$ and the second equality holds due to (R5)[81]. Furthermore, we choose the tolerance $\epsilon := 1 \cdot 10^{-4}$, and we consider a grid of $n = 200$ equally-spaced time points $r$ in the interval $]0, 40]$. We run Algorithm 4.2[168] for these parameter values – using Eq. (3.75)[115] to determine $e^{(t_k - t_{k-1})\underline{Q}}$ – and plot our results in Fig. 6.2[⌒]. The evolution of the lower and upper expected averages of $\mathbb{I}_{\mathrm{H}}(X_t)$ agree with what we conjectured: they converge, and their limit values are the same for all initial states $x$. Furthermore, the lower expected average of $\mathbb{I}_{\mathrm{H}}(X_t)$ seems to converge to ${}^{2}/{}_{5}$, and this is precisely the limit lower expectation: for all $x$ in $\{\mathrm{H}, \mathrm{T}\}$,

$$\lim_{r \to +\infty} [e^{r\underline{Q}} \mathbb{I}_{\mathrm{H}}](x) = \lim_{r \to +\infty} \mathbb{I}_{\mathrm{H}}(x) + \frac{1 - e^{-r(\overline{\lambda} + \underline{\lambda})}}{\overline{\lambda} + \underline{\lambda}} \lambda_x = \frac{\underline{\lambda}}{\overline{\lambda} + \underline{\lambda}} = \frac{2}{5},$$

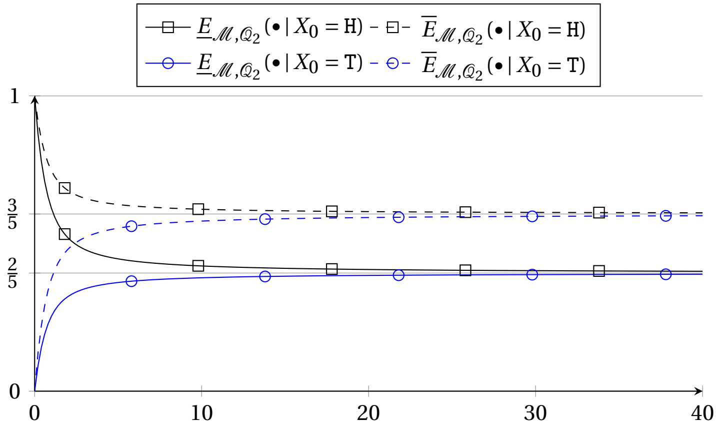



**Figure 6.2** Lower and upper expectation of $[\![\mathbb{I}_{\mathtt{H}}]\!]_{[0,r]}$ as a function of $r$.

where for the first equality we used Eq. $(3.75)_{115}$; as we will see in Example $6.52_{320}$ further on, this is a coincidence. More importantly, 2/5 is also the limit expectation of the rate operator $Q$ in $\mathcal{Q}_2$ with $Q(\mathtt{H},\mathtt{T}) = \overline{\lambda}$ and $Q(\mathtt{T},\mathtt{H}) = \underline{\lambda}$. The reason why this is important is that for the corresponding homogeneous Markovian jump process $P_{\mathbb{I}_{\mathtt{H}},Q}$, this limit expectation of $\mathbb{I}_{\mathtt{H}}$ is equal to the limit of the expected temporal average of $\mathbb{I}_{\mathtt{H}}(X_t)$ due to the Point-Wise Ergodic Theorem. Therefore, and because this homogenous Markovian jump process$P_{\mathbb{I}_{\mathtt{H}},Q}$ belongs to $\mathbb{P}_{\mathcal{M},\mathcal{Q}_2}$, we see that in this case, the limit of the lower expected temporal average of $\mathbb{I}_{\mathtt{H}}(X_t)$ is actually reached by a homogeneous and Markovian jump process. This strengthens our belief that the limit of the lower expected temporal average is the same for $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Similar observations hold for the limit of the upper expected temporal average. ♞

## 6.5 Computing lower and upper expectations of idealised variables

To conclude this chapter, we propose intuitive methods to compute lower and upper expectations of the three types of idealised variables that we have investigated in Sections $6.2_{281}$ to $6.4_{295}$. All of these variables have one thing in common: they are the point-wise limit of simple variables with a sum-product representation. In Section 6.5.1↷, we make two important observations for such variables. We use these observations in Section $6.5.2_{313}$, where we will propose and test our computation methods.

### 6.5.1 Approximating idealised variables

Fix some non-empty set $\mathscr{M}$ of initial probability mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Suppose that we want to determine the lower expectation of the limit variable $f$ – for example, the indicator of a time-bounded until event, a truncated hitting time or a temporal average – up to some maximal error $\epsilon$ in $\mathbb{R}_{>0}$, and that we can do this by determining the lower expectation of some simple variable – for example, the indicator of an approximating until event, an approximating hitting time or an approximating Riemann sum – of the form

$$g := \sum_{k=0}^{n} g_k(X_{t_k}) \prod_{\ell=0}^{k-1} h_\ell(X_{t_\ell}),$$

with $v = (t_0, \dots, t_n)$ a grid over $[s, r]$, $g_0, \dots, g_n$ gambles on $\mathscr{X}$ and $h_0, \dots, h_{n-1}$ non-negative gambles on $\mathscr{X}$ – note that this is the case in all considered instances.

Because $f$ has a sum-product representation over $v$, we can use Algorithm 4.2~168~ to determine its lower expectation conditional on any $\{X_u = x_u\}$ in $\mathscr{H}$ such that $\max u \leq s = \min v$. There is one main drawback to the recursive approximation procedure in Algorithm 4.2~168~ though: for all $k$ in $\{0, \dots, n-1\}$, we need to determine $e^{(t_{k+1}-t_k)\underline{Q}_{\mathscr{Q}}} f_{k+1}$. In general, we can only do so by means of the numerical methods in Section 4.2~173~. That is, for all $k$ in $\{0, \dots, n-1\}$, we need to construct a grid $v_k = (t_{k,0}, \dots, t_{k,n_k})$ over $[t_k, t_{k+1}]$ and then evaluate $(I + (t_{k,\ell} - t_{k,\ell-1})\underline{Q}_{\mathscr{Q}})$ as $\ell$ decreases from $n_k$ to 1. Intuitively, however, we expect that whenever the grid $v$ is sufficiently fine, we can get away with a 'one-step estimate', because then

$$e^{(t_{k+1}-t_k)\underline{Q}_{\mathscr{Q}}} f_{k+1} \approx (I + (t_{k+1} - t_k)\underline{Q}_{\mathscr{Q}}) f_{k+1}$$

due to Lemma 4.16~178~ and (LT14)~179~. The following result bounds the error that is made by relying on this one-step estimate; its proof is straightforward, and can be found in Appendix 6.C~327~.

**Lemma 6.43.** *Consider a lower rate operator $\underline{Q}$ and a sequence of time points $v = (t_0, \dots, t_n)$ in $\mathscr{U}_{\neq()}$ such that $\Delta(v)\|\underline{Q}\|_{\mathrm{op}} \leq 2$. Fix gambles $g_0, \dots, g_n$ and $h_0, \dots, h_{n-1}$ in $\mathbb{G}(\mathscr{X})$ such that $0 \leq h_k \leq 1$ for all $k$ in $\{0, \dots, n-1\}$. Let $f_0$ and $\tilde{f}_0$ be the gambles on $\mathscr{X}$ defined by the initial condition $f_n := g_n =: \tilde{f}_n$ and, for all $k$ in $\{0, \dots, n-1\}$, by the recursive relations*

$$f_k := g_k + h_k e^{\Delta_{k+1}\underline{Q}} f_{k+1} \quad \text{and} \quad \tilde{f}_k := g_k + h_k (I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1},$$

*where for all $k$ in $\{0, \dots, n-1\}$, we let $\Delta_{k+1} := (t_{k+1} - t_k)$. Then*

$$\|f_0 - \tilde{f}_0\| \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=1}^{n} \Delta_k^2 \|\tilde{f}_k\|_{\mathrm{c}}.$$

In Section 6.5.2$_{\curvearrowright}$ further on, we use this idea to arrive at methods to compute the lower and upper expectations of time-bounded until events, truncated hitting times and Riemann integrals of a Lipschitz continuous family. These methods will only consider conditioning events of the form $\{X_s = x\}$, with $s$ the first time point of the interval on which these variables depend. In the remainder of this section, we explain why this can be done without any loss of generality.

All three of the types of idealised variables that we will consider have in common that they are defined over a time interval $[s, r]$, and that for any desired maximal error $\epsilon$ in $\mathbb{R}_{>0}$, we can use the bounds in Propositions 6.13$_{283}$, 6.22$_{293}$ or 6.38$_{304}$ to determine a grid $v$ over $[s, r]$ such that the lower and upper expectations of the corresponding approximating sum-product variable $f_v$ are $\epsilon$-close to the lower and upper Daniell expectations of the idealised variable $f_{[s,r]}$. More formally, they clearly satisfy the condition of the following result. We state this result for the lower expectation only, but it is clear that a similar statement holds for the upper expectation due to conjugacy; in order not to disrupt the main text too much, we have relegated our proof to Appendix 6.C$_{327}$.

**Lemma 6.44.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathcal{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathcal{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathcal{Q}}$. Fix some time points $s, r$ in $\mathbb{R}_{\geq 0}$ and an idealised variable $f$ such that for all $u$ in $\mathscr{U}$ with $\max u \leq s$, $f$ belongs to $\overline{\mathbb{V}}_{\mathrm{lim}}(\mathscr{F}_u)$. If for all $\epsilon$ in $\mathbb{R}_{>0}$, there is some grid $v$ over $[s, r]$ and some real variable $g$ in $\mathbb{V}(\Omega)$ with a sum-product representation over $v$ such that, for all $\{X_u = x_u\}$ in $\mathscr{H}$ with $\max u \leq s$,*

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) - \underline{E}_{\mathscr{P}}(g \mid X_u = x_u)\right| \leq \epsilon,$$

*then the map $f_s$ on $\mathscr{X}$ defined by*

$$f_s(x) := \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_s = x) \quad \text{for all } x \in \mathscr{X}$$

*is a gamble on $\mathscr{X}$, and for all $\{X_u = x_u\}$ in $\mathscr{H}$ with $\max u \leq s$,*

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\left(\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_s) \,\middle|\, X_u = x_u\right) = \underline{E}_{\mathscr{P}}(f_s(X_s) \mid X_u = x_u).$$

Note that $f_s(X_s)$ in this result is a real variable that trivially has a sum-product representation over $(s)$, so we can compute $\underline{E}_{\mathscr{P}}(f_s(X_s) \mid X_u = x_u)$ with Theorem 4.9$_{166}$. Therefore, we can henceforth indeed focus on computing

$$f_s(x) = \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_s = x) \quad \text{for all } x \in \mathscr{X},$$

without loss of generality.

### 6.5.2 Recursive computational methods

Lemma 6.43$_{311}$ is not the only result that we need to establish our 'intuitive' approximation methods. For lower and upper probabilities of time-bounded until events, we also need Proposition 6.13$_{283}$, for lower and upper expect truncated hitting times we invoke Proposition 6.22$_{293}$, and for lower and upper expectations of Riemann integrals we rely on Proposition 6.38$_{304}$. For each of these three cases, we consider the limit variable $f$ corresponding to $[s,r]$, with $s,r$ in $\mathbb{R}_{\geq 0}$ such that $s<r$, and construct a sequence $(\tilde{f}_{n,0}(x))_{n\in\mathbb{N}}$ of approximations of the lower expectation of this limit variable $f$ conditional on $\{X_s = x\}$.

Our usual running example is just a tad too basic to illustrate our methods. The slogan 'four legs good, two legs bad' is applicable here: we increase the size of the state space from 2 to 4.

**Power Network Example 6.45.** Troffaes et al. (2015) use an imprecise jump process to asses the reliability of a power network. They follow up on their earlier work (Troffaes et al., 2013) and consider a power network that consists of two parallel power lines, called A and B, so the network is up and running as long as at least one of the two power lines is in operation. Thus, an independent failure of one of the two power lines is not that much of an issue, because it does not cause a power outage as long as the other power line is in operation. If both power lines fail due to the same cause, this does result in a power outage; whenever this occurs, we speak of a *common cause failure.*

Troffaes et al. (2015, Sections 2.2 and 3.4) model this power network with an imprecise jump process as follows. The state space of the network is $\mathscr{X} := \{\mathtt{AB},\mathtt{A},\mathtt{B},\mathtt{F}\}$, where the state F corresponds to a failure of both power lines and where the other state labels indicate the power lines that are working. The set $\mathcal{Q}$ of rate operators is specified through lower and upper bounds on the off-diagonal components of corresponding matrices:

$$\mathcal{Q} := \left\{Q \in \mathfrak{Q} \colon (\forall x,y \in \mathscr{X}, x \neq y)\ Q_{\mathrm{L}}(x,y) \leq Q(x,y) \leq Q_{\mathrm{U}}(x,y)\right\},$$

where the matrices

$$Q_{\mathrm{L}} := \begin{pmatrix} * & 0.32 & 0.32 & 0.19 \\ 730 & * & 0 & 0.51 \\ 730 & 0 & * & 0.51 \\ 0 & 730 & 730 & * \end{pmatrix} \quad\text{and}\quad Q_{\mathrm{U}} := \begin{pmatrix} * & 0.37 & 0.37 & 0.24 \\ 1460 & * & 0 & 0.61 \\ 1460 & 0 & * & 0.61 \\ 0 & 1460 & 1460 & * \end{pmatrix}$$

collect the bounds on the off-diagonal components – so $Q_{\mathrm{L}}(\mathtt{A},\mathtt{B}) = 0$ and $Q_{\mathrm{U}}(\mathtt{B},\mathtt{F}) = 0.61$. Because every rate matrix has rows that sum to zero, the constraints on the diagonal elements of $Q$ are implied by the others. Note that $\mathcal{Q}$ is non-empty and bounded, and that it has separately specified rows by construction. It is also easy to see that $\mathcal{Q}$ is convex, so $\mathcal{Q}$ is equal to the set $\mathcal{Q}_{\underline{Q}}$ of rate operators that dominate its lower envelope $\underline{Q} := \underline{Q}_{\mathcal{Q}}$. Furthermore, evaluating the lower envelope $\underline{Q} := \underline{Q}_{\mathscr{Q}}$ of $\mathscr{Q}$ is almost trivial because of the specific structure of $\mathscr{Q}$. For example, we immediately find that

$$\|\underline{Q}\|_{\mathrm{op}} = 2\max\{-[\underline{Q}\mathbb{I}_x](x) \colon x \in \mathscr{X}\} = -2[\underline{Q}\mathbb{I}_{\mathtt{F}}](\mathtt{F}) = -2(-2 \times 1460) = 5840,$$

where the first equality is (LR7)$_{111}$.

Troffaes et al. (2015) do not explicitly specify a set $\mathscr{M}$ of initial probability mass functions, and use (an informal version of) the imprecise jump process $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ to model the power network; because $\mathscr{M}$ does not play a role in our analysis, we take it to be equal to $\Sigma_{\mathscr{X}}$. £

### Time-bounded until events

First, we use Proposition 6.13$_{283}$, Theorem 4.9$_{166}$ and Lemma 6.43$_{311}$ to establish a method to recursively compute the lower and upper probability of a bounded until event. Our proof is relatively straightforward, so we have relegated it to Appendix 6.D$_{330}$.

**Theorem 6.46.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Fix subsets $S, G$ of $\mathscr{X}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $s < r$. For all $n$ in $\mathbb{N}$, we let $\Delta_n := {}^{(r-s)}/_{n}$ and let $\tilde{f}_{n,0}$ be the gamble on $\mathscr{X}$ that is defined by the initial condition $\tilde{f}_{n,n} := \mathbb{I}_G$ and, for all $k$ in $\{0,\dots n-1\}$, by the recursive relation*

$$\tilde{f}_{n,k} := \mathbb{I}_G + \mathbb{I}_{S\setminus G}\big(I + \Delta_n \underline{Q}_{\mathscr{Q}}\big)\tilde{f}_{n,k+1}. \tag{6.34}$$

*Then for all $x$ in $\mathscr{X}$ and $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}_{\mathscr{Q}}\|_{\mathrm{op}} \leq 2n$,*

$$\big|\underline{P}^{\mathrm{D}}_{\mathscr{P}}\big(H^{S,G}_{[s,r]} \,\big|\, X_s = x\big) - \tilde{f}_{n,0}(x)\big| \leq \frac{3}{8}\frac{(r-s)^2}{n}\|\underline{Q}_{\mathscr{Q}}\|^2_{\mathrm{op}},$$

*and therefore*

$$\underline{P}^{\mathrm{D}}_{\mathscr{P}}\big(H^{S,G}_{[s,r]} \,\big|\, X_s = x\big) = \lim_{n\to+\infty} \tilde{f}_{n,0}(x).$$

*The same holds for $\overline{P}^{\mathrm{D}}_{\mathscr{P}}$ if in Eq. (6.34) we replace $\underline{Q}_{\mathscr{Q}}$ by $\overline{Q}_{\mathscr{Q}}$.*

Let us test the recursive method in Theorem 6.46 in the setting of our new running example.

**Power Network Example 6.47.** Troffaes et al. (2015, Sections 2.3 and 3.5) look at the network over a period of 10 years, and we will do the same. Here, we are interested in the event that over this 10 year period, there is a common cause failure first, in the sense that both power lines never fail separately before the first time that the network is down. In our model, this corresponds to a jump from `AB` to `F`, without visiting other states; hence, the event ‘there is a common cause failure first’ corresponds to the time-bounded until event $H^{\mathtt{AB},\mathtt{F}}_{[0,10]} := H^{S,G}_{[0,10]}$ with $S := \{\mathtt{AB}\}$ and $G := \{\mathtt{F}\}$ – ironically, in this instance the set of 'goal' states contains the only state that we do not want our system to be in. We make the natural assumption that both lines of the network are currently working, so we need to determine

$$\underline{P}_{\mathscr{M},\mathscr{Q}}\big(H^{\mathtt{AB},\mathtt{F}}_{[0,10]} \,\big|\, X_0 = \mathtt{AB}\big) \quad \text{and} \quad \overline{P}_{\mathscr{M},\mathscr{Q}}\big(H^{\mathtt{AB},\mathtt{F}}_{[0,10]} \,\big|\, X_0 = \mathtt{AB}\big).$$

Needless to say, we use Theorem 6.46 to do so. The minimum number of iterations that Theorem 6.46 dictates is

$$n_{\min} := \left\lceil \frac{10\|\underline{Q}\|_{\mathrm{op}}}{2} \right\rceil = \left\lceil \frac{10 \times 5840}{2} \right\rceil = 29\,200.$$

With the recursive procedure in Theorem 6.46 for $n = n_{\min}$, we find that

$$\underline{P}_{\mathscr{M},\mathscr{Q}}\big(H^{\mathtt{AB},\mathtt{F}}_{[0,10]} \,\big|\, X_0 = \mathtt{AB}\big) \approx 0.2043 \quad \text{and} \quad \overline{P}_{\mathscr{M},\mathscr{Q}}\big(H^{\mathtt{AB},\mathtt{F}}_{[0,10]} \,\big|\, X_0 = \mathtt{AB}\big) \approx 0.2726.$$

Quite remarkably, increasing the number of iterations does not change the first four significant digits of these approximations.

For both the lower and the upper probability, we determine $\tilde{f}_{n,0}(x_s)$ for increasing values of $n$, starting from the minimum number of iterations $n = n_{\min}$ and then repeatedly doubling the number of iterations until $n = 2^{10} n_{\min}$. In Fig. 6.3, we report the relative difference

$$\frac{\tilde{f}_{2^k n_{\min},0}(x_s) - \tilde{f}_{2^{10} n_{\min},0}(x_s)}{\tilde{f}_{2^{10} n_{\min},0}(x_s)}$$

between the $k$-th estimate and the last estimate, with $k$ ranging from 0 to 9. The relative difference for $k = 0$ is less than $1 \cdot 10^{-6}$ and then decreases as the number of iterations grows; the decrease is linear on the log-log graph, so the error decreases exponentially as the number of iterations increases. For $n = 2^k n_{\min}$, the theoretical upper bound on the error in Theorem 6.46 is

$$\frac{3}{8} \frac{10^2}{2^k n_{\min}} \|\underline{Q}\|^2_{\mathrm{op}} = \frac{43\,800}{2^k}.$$

While this bound does decrease exponentially, it is clearly overly conservative; it overestimates the error by about ten orders of magnitude!

Finally, we use Theorem 6.46 to determine the lower and upper probability of $H^{\mathtt{AB},\mathtt{F}}_{[0,r]}$ for various values of $r$ and plot these in Fig. 6.4. From this graph, it is clear that these lower and upper probabilities converge; these limit values are approximately equal to the lower and upper probability of $H^{\mathtt{AB},\mathtt{F}}_{[0,10]}$. £

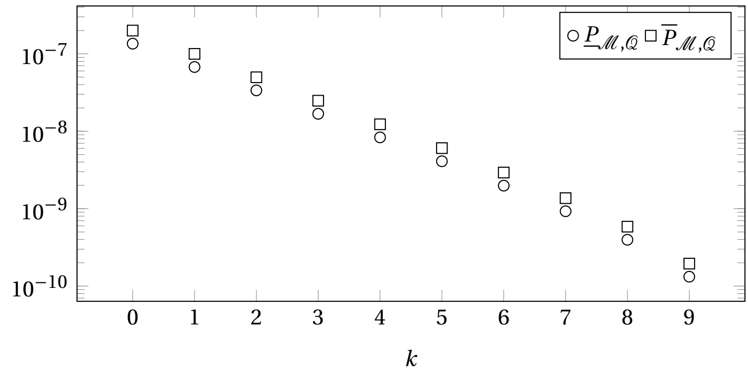



**Figure 6.3** Relative difference between the approximation for $n = 2^k n_{\min}$ and that for $n = 2^{10} n_{\min}$ for the lower and upper probability of $H^{\mathtt{AB},\mathtt{F}}_{[0,10]}$.

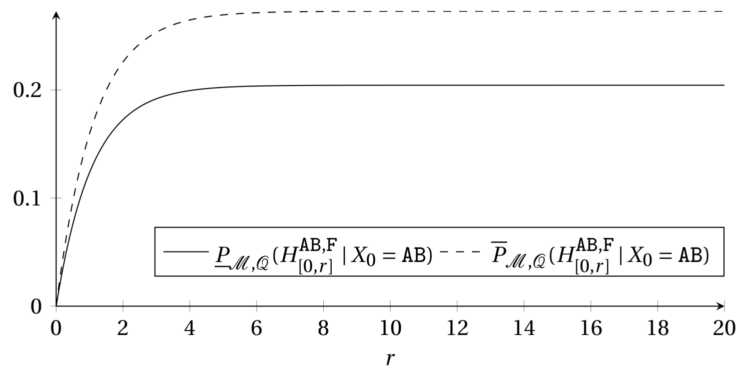



**Figure 6.4** Lower and upper probability of $H^{\mathtt{AB},\mathtt{F}}_{[0,r]}$ as a function of $r$.

### Truncated hitting times

Next, we propose a recursive method to iteratively compute the lower and upper expectation of a truncated hitting time. The proof is similar to the proof of Theorem 6.46<sub>314</sub>, although we now rely on Proposition 6.22<sub>293</sub> instead of on Proposition 6.13<sub>283</sub>; for the details, see Appendix 6.D<sub>330</sub>.

**Theorem 6.48.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Fix some subset $G$ of $\mathscr{X}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $s < r$. For all $n$ in $\mathbb{N}$, we let $\Delta_n := {}^{(r-s)}/{}_{n}$ and let $\tilde{f}_{n,0}$ be the gamble on $\mathscr{X}$ that is defined by the initial* *condition* $\tilde{f}_{n,n} \coloneqq \Delta_n$ *and, for all k in* $\{0,\dots n-1\}$*, by the recursive relation*

$$\tilde{f}_{n,k} \coloneqq \begin{cases} \Delta_n + \mathbb{I}_{G^c}\big(I + \Delta_n \underline{Q}_{\mathcal{Q}}\big)\tilde{f}_{n,k+1} & \text{if } k \geq 1, \\ s + \mathbb{I}_{G^c}\big(I + \Delta_n \underline{Q}_{\mathcal{Q}}\big)\tilde{f}_{n,k+1} & \text{if } k = 0. \end{cases} \tag{6.35}$$

*Then for all x in* $\mathcal{X}$ *and all n in* $\mathbb{N}$ *such that* $(r-s)\|\underline{Q}_{\mathcal{Q}}\|_{\text{op}} \leq 2$*,*

$$\big|\underline{E}^{\text{D}}_{\mathscr{P}}\big(\tau^G_{[s,r]} \,\big|\, X_s = x\big) - \tilde{f}_{n,0}(x)\big| \leq \frac{r-s}{n} + \frac{1}{8}\frac{(r-s)^3}{n}\frac{2n+1}{n}\|\underline{Q}_{\mathcal{Q}}\|^2_{\text{op}},$$

*and therefore*

$$\underline{E}^{\text{D}}_{\mathscr{P}}\big(\tau^G_{[s,r]} \,\big|\, X_s = x\big) = \lim_{n\to+\infty} \tilde{f}_{n,0}(x).$$

*The same holds for* $\overline{E}^{\text{D}}_{\mathscr{P}}$ *if in Eq.* (6.35) *we replace* $\underline{Q}_{\mathcal{Q}}$ *by* $\overline{Q}_{\mathcal{Q}}$*.*

Here too, we use our running example to put our new recursive computation method to the test.

**Power Network Example 6.49.** In our model, the time until a failure occurs corresponds to $\tau^{\text{F}} \coloneqq \tau^{\{\text{F}\}}_{[0,+\infty[}$, so this is a prime example of a hitting time. Due to Lemma 6.24<sub>294</sub> and Corollary 5.33<sub>240</sub>,

$$\underline{E}_{\mathscr{M},\mathcal{Q}}\big(\tau^{\text{F}}_{[0,r]} \,\big|\, X_0 = \texttt{AB}\big) \quad \text{and} \quad \overline{E}_{\mathscr{M},\mathcal{Q}}\big(\tau^{\text{F}}_{[0,r]} \,\big|\, X_0 = \texttt{AB}\big)$$

converge to a conservative lower bound on the lower expected hitting time and to the exact value of the upper expected hitting time, respectively.

Let us again start with $r = 10$. The minimum number of iterations that Theorem 6.48↶ dictates is $n_{\text{min}} = 29\,200$. With the recursive procedure in Theorem 6.48↶ for $n = n_{\text{min}}$, we find that

$$\underline{E}_{\mathscr{M},\mathcal{Q}}\big(\tau^{\text{F}}[0,10] \,\big|\, X_0 = \texttt{AB}\big) \approx 3.784 \quad \text{and} \quad \overline{E}_{\mathscr{M},\mathcal{Q}}\big(\tau^{\text{F}}[0,10] \,\big|\, X_0 = \texttt{AB}\big) \approx 4.474.$$

As in Power Network Example 6.47<sub>314</sub>, increasing the number of iterations does not change the first four significant digits of these approximations. In Fig. 6.5↷, we report the relative difference between the $k$-th estimate corresponding to $n = 2^k n_{\text{min}}$ and the estimate corresponding to $n = 2^{10} n_{\text{min}}$, with $k$ ranging from 0 to 9. The relative difference for $k = 0$ is less than $1 \cdot 10^{-4}$ and then decreases as the number of iterations grows; as in Fig. 6.3↶, the error decreases exponentially.

Let us get back to determining the lower and upper expected hitting time. To this end, we determine the lower and upper expectation of $\tau^{\text{F}}_{[0,r]}$ for various values of $r$ and plot these in Fig. 6.6↷. The lower and upper expected truncated hitting time converges, as required. Up to four significant digits, the limit values are 4.160 and 5.2590; about twice as large as the lower and upper expectation of $\tau^{\text{F}}_{[0,10]}$. £

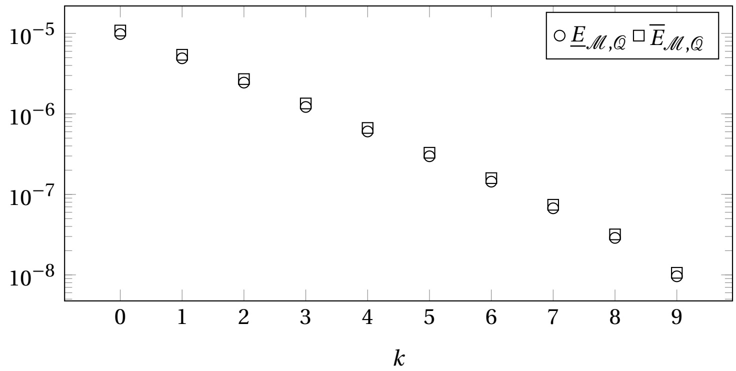



**Figure 6.5** Relative difference between the approximation corresponding to $n = 2^k n_{\min}$ and the one corresponding to $n = 2^{10} n_{\min}$ for the lower and upper expectation of $\tau^{\mathtt{F}}_{[0,10]}$.

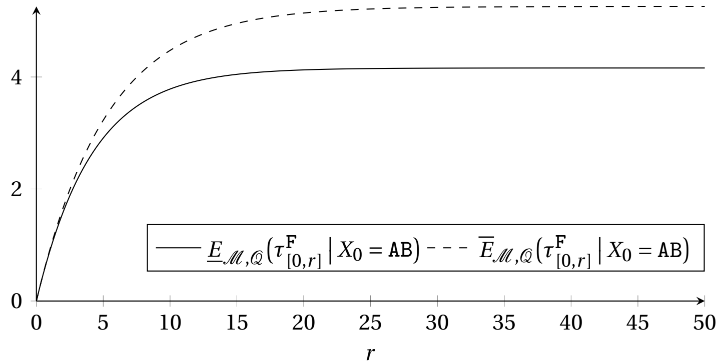



**Figure 6.6** Lower and upper expectation of $\tau^{\mathtt{F}}_{[0,r]}$ as a function of $r$.

### Riemann integrals

Finally, we propose a method to compute the lower and upper expectation of an idealised variable in the form of a Riemann integral.

**Theorem 6.50.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Fix some time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $s < r$ and a Lipschitz continuous family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$ with Lipschitz constant $\kappa$, and let*

$$\gamma := \sup\{\max f_t \colon t \in [s,r]\} - \inf\{\min f_t \colon t \in [s,r]\}.$$

*For all $n$ in $\mathbb{N}$, we let $\Delta_n := {}^{(r-s)}/_n$ and let $\tilde{f}_{n,0}$ be the gamble on $\mathscr{X}$ that is defined by the initial condition $\tilde{f}_{n,n} := \Delta_n f_r$ and, for all $k$ in $\{0,\dots n-1\}$, by the recursive relation*

$$\tilde{f}_{n,k} := \begin{cases} \Delta_n f_{s+k\Delta_n} + \left(I + \Delta_n \underline{Q}_{\mathscr{Q}}\right)\tilde{f}_{n,k+1} & \text{if } k \geq 1, \\ \left(I + \Delta_n \underline{Q}_{\mathscr{Q}}\right)\tilde{f}_{n,k+1} & \text{if } k = 0. \end{cases} \tag{6.36}$$

*Then for all $x$ in $\mathscr{X}$ and $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}_{\mathscr{Q}}\|_{\mathrm{op}} \leq 2n$,*

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}\left(\int_s^r f_t(X_t)\,\mathrm{d}t \,\middle|\, X_s = x\right) - \tilde{f}_{n,0}(x)\right| \leq \frac{(r-s)^2}{n}\left(\kappa + \frac{\gamma}{2}\|\underline{Q}\|_{\mathrm{op}} + \frac{1}{8}(r-s)\frac{n+1}{n}\|\underline{Q}\|^2_{\mathrm{op}}\right),$$

*and therefore*

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}\left(\int_s^r f_t(X_t)\,\mathrm{d}t \,\middle|\, X_s = x\right) = \lim_{n\to+\infty} \tilde{f}_{n,0}(x).$$

*The same holds for $\overline{E}^{\mathrm{D}}_{\mathscr{P}}$ if in Eq. (6.36) we replace $\underline{Q}_{\mathscr{Q}}$ by $\overline{Q}_{\mathscr{Q}}$.*

Let us use the recursive method in Theorem 6.50 to compute a temporal average for our running example.

**Power Network Example 6.51.** Troffaes et al. (2015) assess the reliability of the power network by looking at two inferences: the number of failures and the length of time that the network is down, both over a ten year period. In their model, the latter corresponds to the time that the system spends in the state F, so to $\int_0^{10} \mathbb{I}_{\mathtt{F}}(X_t)\,\mathrm{d}t$. Thus, we should determine

$$\underline{E}_{\mathscr{M},\mathscr{Q}}\left(\int_0^{10} \mathbb{I}_{\mathtt{F}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = \mathtt{AB}\right) \quad \text{and} \quad \overline{E}_{\mathscr{M},\mathscr{Q}}\left(\int_0^{10} \mathbb{I}_{\mathtt{F}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = \mathtt{AB}\right),$$

the lower and upper expected downtime of the system over a ten year period. Of course, we do so with the method in Theorem 6.50. As in Power Network Examples $6.47_{314}$ and $6.49_{317}$, the minimum number of iterations is $n_{\min} = 29\,200$. With $n = n_{\min}$ in Theorem 6.50, we find the approximations $6.512 \cdot 10^{-4}$ and $1.647 \cdot 10^{-3}$ for the lower and upper expected down time, respectively. In contrast to Power Network Examples $6.47_{314}$ and $6.49_{317}$, increasing the number of iterations does change the fourth significant digit of these approximations. In Fig. 6.7, we report the relative difference between the $k$-th estimate corresponding to $n = 2^k n_{\min}$ and the estimate corresponding to $n = 2^{10} n_{\min}$, with $k$ ranging from 0 to 9. The relative difference for $k = 0$ is still well below $1 \cdot 10^{-4}$, and then decreases as the number of iterations grows; here too, the error decreases exponentially.

Troffaes et al. (2015) approximate the lower (and upper) expected downtime heuristically. In essence, they use the limit lower expectation

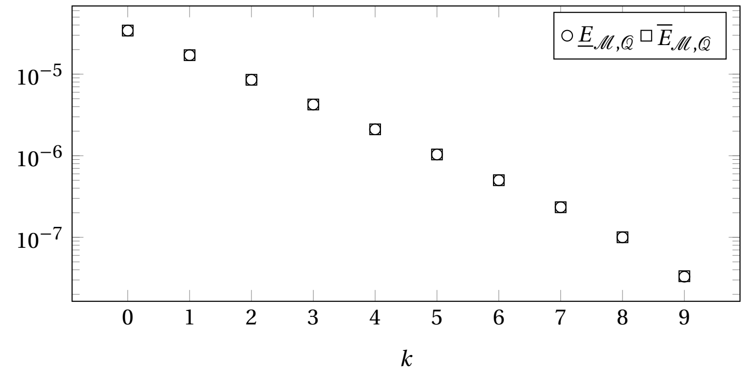



**Figure 6.7** Relative difference between the approximation corresponding to $n = 2^k n_{\text{min}}$ and the one corresponding to $n = 2^{10} n_{\text{min}}$ for the lower and upper expectation of $\int_0^{10} \mathbb{I}_{\mathtt{F}}(X_t)\,\mathrm{d}t$.

$\lim_{t\to+\infty} e^{t\underline{Q}}\mathbb{I}_F$ as an approximation for the limit of the lower expected temporal average, and then multiply this approximation by 10, the length of the time period. In this way, they find approximations for the lower and upper expected downtime that are virtually the same as ours: $6.513 \cdot 10^{-4}$ and $1.647 \cdot 10^{-3}$, respectively. Hence, this is a second example where the lower and upper expected temporal average seem to converge to the lower and upper limit expectation. This *need not* always be the case, as Example 6.52 further on illustrates, so the heuristic of Troffaes et al. (2015) can sometimes yield overly conservative bounds.

Another issue with this heuristic method, is that the resulting approximation is the same for every initial state $x$. This is not necessarily the case for the actual value though! For example, we find that conditional on $\{X_0 = \mathtt{F}\}$, the lower and upper expected downtime is $9.938 \cdot 10^{-4}$ and $2.332 \cdot 10^{-3}$, respectively. To explain this difference, we plot the lower and upper expected temporal average of $\mathbb{I}_{\mathtt{F}}$ as function of the time horizon in Fig. 6.8↷. We see that when starting from the initial state AB, the expected temporal average converges almost instantaneously compared to the expected temporal average starting from F. It is this extensive transient phase that causes our values to be so different. £

To conclude this chapter on computing lower and upper expectations for idealised variables, we give an example where the upper expected temporal average *does not* converge to the upper limit expectation.

**Example 6.52.** We consider a ternary state space $\mathscr{X} := \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the set of rate

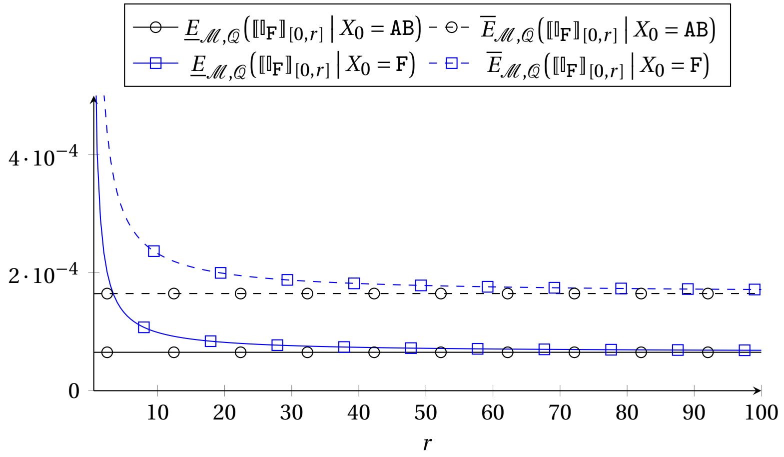



**Figure 6.8** Lower and upper expectation of $\llbracket \mathbb{I}_{\mathtt{F}} \rrbracket_{[0,r]} = \frac{1}{r} \int_0^r \mathbb{I}_{\mathtt{F}}(X_t)\,\mathrm{d}t$ as a function of $r$.

operators

$$\mathscr{Q} := \left\{ \begin{pmatrix} -\lambda_{\mathtt{a}} & \lambda_{\mathtt{a}} & 0 \\ \mu_{\mathtt{b}} & -\mu_{\mathtt{b}} - \lambda_{\mathtt{b}} & \lambda_{\mathtt{b}} \\ 0 & \mu_{\mathtt{c}} & -\mu_{\mathtt{c}} \end{pmatrix} : \begin{array}{l} \lambda_{\mathtt{a}} = 1, \mu_{\mathtt{b}} = 10, \\ \lambda_b \in [1, 100], \\ \mu_{\mathtt{c}} \in [1, 100] \end{array} \right\}$$

and an arbitrary non-empty set $\mathscr{M} \subseteq \Sigma_{\mathscr{X}}$ of initial probability mass functions. For the Markovian imprecise jump process $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, we are after the upper expected fraction of time that the system is in state $\mathtt{b}$. For each initial state $x$ in $\mathscr{X}$ and for increasing values of $r$ in $\mathbb{R}_{\geq 0}$, we determine

$$\overline{E}_{\mathscr{M},\mathscr{Q}}\left( \frac{1}{r} \int_0^r \mathbb{I}_{\mathtt{b}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x \right) = -\underline{E}_{\mathscr{M},\mathscr{Q}}\left( \frac{1}{r} \int_0^r [-\mathbb{I}_{\mathtt{b}}](X_t)\,\mathrm{d}t \,\middle|\, X_0 = x \right)$$

with Theorem 6.50318. For large enough $r$, we obtain 0.091 – or 9.1 % – and this value does not depend on the initial state $x$. To compare with the heuristic of Troffaes et al. (2015), we determine the upper limit expectation $\lim_{t \to +\infty} -e^{t\underline{Q}_{\mathscr{Q}}}(-\mathbb{I}_{\mathtt{b}})$. Remarkably, this upper expectation is 0.703 – or 70.3 % – so this shows that the upper expected temporal average need not converge to the upper limit expectation, and that it can converge to a *significantly* lower value. ◇

## 6.A Proof of Proposition 6.2

This appendix is devoted to the proof of Proposition 6.2275. Before we can get around to our proof, we need to establish the following upper bound on the probability of the event of having a jump over two consecutive subintervals.

**Lemma 6.53.** *Consider a jump process $P$ that has uniformly bounded rate, with rate bound $\lambda$. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$ and time points $s, t, r$ in $\mathbb{R}_{\geq 0}$ such that* $\max u \leq s < t < r$. *Then*

$$\{X_s \neq X_t \neq X_r\} := \big\{\omega \in \Omega \colon \omega(s) \neq \omega(t) \neq \omega(r)\big\}$$

*belongs to $\mathscr{F}_u$, and*

$$P(X_s \neq X_t \neq X_r \,|\, X_u = x_u) \leq \frac{1}{4}(t-s)(r-t)\lambda^2.$$

*Proof.* To verify the first part of the statement, we observe that

$$A := \{X_s \neq X_t \neq X_r\} = \{X_v \in B\},$$

where we let $v := (s, t, r)$ and $B := \big\{(x, y, z) \in \mathscr{X}_{(s,r,t)} \colon x \neq y \neq z\big\}$. Note that $v \succcurlyeq u$ by the assumption of the statement, so it follows from Eq. (3.16)$_{63}$ that $\{X_s \neq X_t \neq X_r\}$ belongs to $\mathscr{F}_u$.

To verify the second part of the statement, we let

$$f \colon \mathscr{X}_{(s,t,r)} \to \mathbb{R} \colon (x, z, y) \mapsto \begin{cases} 1 & \text{if } x \neq y \neq z \\ 0 & \text{otherwise} \end{cases}$$

and

$$g \colon \mathscr{X}^2 \to \mathbb{R} \colon (x, y) \mapsto \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{otherwise,} \end{cases}$$

and observe that

$$\mathbb{I}_A = f(X_s, X_t, X_r) = g(X_s, X_t)g(X_t, X_r). \tag{6.37}$$

Consequently,

$$P(A \,|\, X_u = x_u) = E_P(\mathbb{I}_A \,|\, X_u = x_u) = E_P(f(X_s, X_t, X_r) \,|\, X_u = x_u). \tag{6.38}$$

We will now use this equality, the law of iterated expectations and Theorem 5.27$_{236}$ to prove the inequality of the statement. To make our lives a little easier, we assume that $\max u < s$; the proof for the case that $\max u = s$ is analoguous, the only difference being that the first and last step are not needed then.

It follows from Theorem 3.19$_{72}$ that

$$E_P(f(X_s, X_t, X_r) \,|\, X_u = x_u) = E_P\big(E_P(f(X_s, X_t, X_r) \,|\, X_v) \,\big|\, X_u = x_u\big), \tag{6.39}$$

with $v := u \cup (s)$. We take a closer look at the innermost conditional expectation. To this end, we fix any $y_v$ in $\mathscr{X}_v$. Then by Theorem 3.19$_{72}$,

$$E_P\big(f(X_s, X_t, X_r) \,\big|\, X_v = y_v\big) = E_P\big(E_P\big(f(X_s, X_t, X_r) \,\big|\, X_w\big) \,\big|\, X_v = y_v\big), \tag{6.40}$$

with $w := u \cup (s, t)$.

Again, we focus on the innermost conditional expectation. To this end, we fix any $z_w$ in $\mathscr{X}_w$. Then it follows from Corollary 3.18$_{71}$, Eq. (6.37)$_{\curvearrowleft}$ and (ES2)$_{37}$ that

$$\begin{aligned} E_P\big(f(X_s,X_t,X_r)\,\big|\,X_w=z_w\big) &= E_P\big(f(z_s,z_t,X_r)\,\big|\,X_w=z_w\big) \\ &= E_P\big(g(z_s,z_t)g(z_t,X_r)\,\big|\,X_w=z_w\big) \\ &= g(z_s,z_t)E_P\big(g(z_t,X_r)\,\big|\,X_w=z_w\big) \\ &= g(z_s,z_t)E_P\big(g(X_t,X_r)\,\big|\,X_w=z_w\big). \end{aligned}$$

By definition, $g(X_t,X_r)=\eta_{(t,r)}$, so it follows from Theorem 5.27$_{236}$ – with $v=(t,r)$ – that

$$E_P\big(f(X_s,X_t,X_r)\,\big|\,X_w=z_w\big) \le g(z_s,z_t)(r-t)\frac{\lambda}{2}.$$

Because this inequality holds for all $z_w$ in $\mathscr{X}_w$, we conclude that

$$E_P\big(f(X_s,X_t,X_r)\,\big|\,X_w\big) \le g(X_s,X_t)\frac{1}{2}(r-t)\lambda.$$

It follows from this inequality, Eq. (6.40)$_{\curvearrowleft}$, (ES2)$_{37}$ and (ES4)$_{37}$ that

$$E_P\big(f(X_s,X_t,X_r)\,\big|\,X_v=y_v\big) \le \frac{1}{2}(r-t)\lambda E_P\big(g(X_s,X_t)\,\big|\,X_v=y_v\big).$$

We now apply the same trick as before, to obtain that for all $y_v$ in $\mathscr{X}_v$,

$$E_P\big(f(X_s,X_t,X_r)\,\big|\,X_v=y_v\big) \le \frac{1}{4}(t-s)(r-t)\lambda^2.$$

It follows from this inequality, Eq. (6.39)$_{\curvearrowleft}$ and (ES1)$_{37}$ that

$$E_P\big(f(X_s,X_t,X_r)\,\big|\,X_u=x_u\big) \le \frac{1}{4}(t-s)(r-t)\lambda^2.$$

Finally, it follows from the preceding inequality and Eq. (6.38)$_{\curvearrowleft}$ that

$$P(X_s\neq X_t\neq X_r\,|\,X_u=x_u) \le \frac{1}{4}(t-s)(r-t)\lambda^2,$$

which is the inequality of the second part of the statement. □

Even with Lemma 6.53$_{\curvearrowleft}$, our proof for Proposition 6.2$_{275}$ is pretty long.

**Proposition 6.2.** *Consider a jump process $P$ that has uniformly bounded rate, with rate bound $\lambda$. Fix a state history $\{X_u=x_u\}$ in $\mathscr{H}$, time points $s,r$ in $\mathbb{R}_{\geq 0}$ such that $\max u\le s<r$ and a grid $v=(t_0,\dots,t_n)$ over $[s,r]$. Then $\eta_{[s,r]}-\eta_v$ is a non-negative $\mathscr{F}_u$-over variable, and*

$$\begin{aligned} E_P^{\mathrm{D}}(\eta_{[s,r]}-\eta_v\,|\,X_u=x_u) &= E_P^{\mathrm{D}}(\eta_{[s,r]}\,|\,X_u=x_u)-E_P(\eta_v\,|\,X_u=x_u) \\ &\le \frac{1}{4}\Delta(v)(r-s)\lambda^2. \end{aligned}$$

*Proof.* For every $\ell$ in $\mathbb{N}$ and $k$ in $\{1,\dots,n\}$, we let $v_{\ell,k}$ be the grid over $[t_{k-1},t_k]$ that divides this subinterval in $2^\ell$ subintervals of equal length. That is, for all $\ell$ in $\mathbb{N}$ and $k$ in $\{1,\dots,n\}$, we let $v_{\ell,k} := (t_{\ell,k,0},\dots,t_{\ell,k,2^\ell})$ where for all $i$ in $\{0,\dots,2^\ell\}$,

$$t_{\ell,k,i} := t_{k-1} + (t_k - t_{k-1})\frac{i}{2^\ell}.$$

Next, for all $\ell$ in $\mathbb{N}$, we let $v_\ell$ be the (ordered) union of $v_{\ell,1}$, …, $v_{\ell,n}$; this way, $v_\ell$ is a grid over $[s,r]$ with $\Delta(v_\ell) = \Delta(v)2^{-\ell}$ such that $v \subseteq v_\ell \subseteq v_{\ell+1}$. Recall from Lemma 5.21<sub>234</sub> that $\eta_v$ and, for all $\ell$ in $\mathbb{N}$, $\eta_{v_\ell}$ are $\mathscr{F}_u$-simple variables. Therefore, it follows immediately from Lemma 2.39[36] that for all $\ell$ in $\mathbb{N}$, $(\eta_{v_\ell}-\eta_v)$ is an $\mathscr{F}_u$-simple variable. Furthermore, for all $\ell$ in $\mathbb{N}$, it follows immediately from Lemma 5.23[234] that $\eta_{v_{\ell+1}} \geq \eta_{v_\ell} \geq \eta_v$ because $v_{\ell+1} \supseteq v_\ell \supseteq v$ by construction. Thus, we have shown that $(\eta_{v_\ell}-\eta_v)_{\ell\in\mathbb{N}}$ is a non-decreasing sequence of non-negative $\mathscr{F}_u$-simple variables; that this sequence converges point-wise to $\eta_{[s,r]}-\eta_v$ follows immediately from Theorem 5.26[236]. Hence, $\eta_{[s,r]}-\eta_v$ is a non-negative $\mathscr{F}_u$-over variable, and it follows from (DE1)[225], (DE3)[225] and Theorem 5.10[226] that

$$E_P^{\mathrm{D}}(\eta_{[s,r]}-\eta_v \mid X_u = x_u) = \lim_{\ell\to+\infty} E_P(\eta_{v_\ell}-\eta_v \mid X_u = x_u). \tag{6.41}$$

In order to verify the inequality of the statement, we investigate the expectations on the right-hand side of the preceding equality. To this end, we fix any $\ell$ in $\mathbb{N}$. It follows from (repeated application of) Lemma 5.22[234] that

$$\eta_{v_\ell}-\eta_v = \sum_{k=1}^n \eta_{v_{\ell,k}} - \sum_{k=1}^n \eta_{(t_{k-1},t_k)} = \sum_{k=1}^n \big(\eta_{v_{\ell,k}} - \eta_{(t_{k-1},t_k)}\big). \tag{6.42}$$

Recall from Lemma 5.24[234] that, for all $k$ in $\{1,\dots,n\}$,

$$\eta_{v_{\ell,k}} = \eta_{(t_{k-1},t_k)} + 2\sum_{i=1}^{2^\ell-1} \mathbb{I}_{\{X_{t_{\ell,k,i-1}}\neq X_{t_{\ell,k,i}}\neq X_{t_{\ell,k,2^\ell}}\}}.$$

We substitute the preceding equality in Eq. (6.42), to yield

$$\eta_{v_\ell}-\eta_v = 2\sum_{k=1}^n\sum_{i=1}^{2^\ell-1} \mathbb{I}_{\{X_{t_{\ell,k,i-1}}\neq X_{t_{\ell,k,i}}\neq X_{t_{\ell,k,2^\ell}}\}};$$

from this equality, it follows that

$$\begin{aligned}E_P(\eta_{v_\ell}-\eta_v \mid X_u = x_u) &= E_P\left(2\sum_{k=1}^n\sum_{i=1}^{2^\ell-1} \mathbb{I}_{\{X_{t_{\ell,k,i-1}}\neq X_{t_{\ell,k,i}}\neq X_{t_{\ell,k,2^\ell}}\}} \,\middle|\, X_u = x_u\right)\\ &= 2\sum_{k=1}^n\sum_{i=1}^{2^\ell-1} P\big(X_{t_{\ell,k,i-1}}\neq X_{t_{\ell,k,i}}\neq X_{t_{\ell,k,2^\ell}} \mid X_u = x_u\big),\end{aligned}$$

where for the second equality we used Eq. (2.19)$_{36}$. We replace the probabilities on the right-hand side of the equality by the upper bound in Lemma 6.53$_{322}$, to yield

$$\begin{aligned}
E_P(\eta_{v_\ell} - \eta_v \,|\, X_u = x_u) &\leq 2 \sum_{k=1}^{n} \sum_{i=1}^{2^\ell - 1} \frac{1}{4} \big(t_{\ell,k,i} - t_{\ell,k,i-1}\big)\big(t_{\ell,k,2^\ell} - t_{\ell,k,i}\big)\lambda^2 \\
&= 2 \sum_{k=1}^{n} \sum_{i=1}^{2^\ell - 1} \frac{1}{4} \frac{t_k - t_{k-1}}{2^\ell} \frac{(t_k - t_{k-1})(2^\ell - i)}{2^\ell} \lambda^2 \\
&= \frac{1}{2} \lambda^2 \sum_{k=1}^{n} (t_k - t_{k-1})^2 \frac{1}{2^\ell} \sum_{i=1}^{2^\ell - 1} \frac{2^\ell - i}{2^\ell},
\end{aligned}$$

where the two equalities follow after some straightforward manipulations. Because

$$\sum_{i=1}^{2^\ell - 1} \frac{2^\ell - i}{2^\ell} = \frac{1}{2^\ell} \sum_{i=1}^{2^\ell - 1} (2^\ell - i) = \frac{1}{2^\ell} \sum_{i=1}^{2^\ell - 1} i = \frac{1}{2^\ell} \frac{(2^\ell - 1)2^\ell}{2} = \frac{2^\ell - 1}{2},$$

it follows from this inequality that

$$\begin{aligned}
E_P(\eta_{v_\ell} - \eta_v \,|\, X_u = x_u) &\leq \frac{1}{2} \lambda^2 \sum_{k=1}^{n} (t_k - t_{k-1})^2 \frac{1}{2^\ell} \frac{2^\ell - 1}{2} \\
&= \frac{1}{4} \lambda^2 \frac{2^\ell - 1}{2^\ell} \sum_{k=1}^{n} (t_k - t_{k-1})^2 \\
&\leq \frac{1}{4} \Delta(v)(r - s)\lambda^2 \frac{2^\ell - 1}{2^\ell},
\end{aligned}$$

where for the last inequality we used that $(t_k - t_{k-1}) \leq \Delta(v)$ for all $k$ in $\{1, \ldots, n\}$ and that $\sum_{k=1}^{n} (t_k - t_{k-1}) = (r - s)$.

It follows from the preceding inequality and Eq. (6.41)$_{\curvearrowleft}$ that

$$E_P^{\mathrm{D}}(\eta_{[s,r]} - \eta_v \,|\, X_u = x_u) \leq \lim_{\ell \to +\infty} \frac{1}{4} \Delta(v)(r - s)\lambda^2 \frac{2^\ell - 1}{2^\ell} = \frac{1}{4} \Delta(v)(r - s)\lambda^2,$$

establishing the inequality in the statement. Furthermore, because $\eta_{[s,r]}$ and $\eta_{[s,r]} - \eta_v$ are non-negative $\mathscr{F}_u$-over variables – see Theorem 5.26$_{236}$ for the former – and because $\eta_v$ is an $\mathscr{F}_u$-simple variable (and hence bounded), it follows from (DE1)$_{225}$, (DE2)$_{225}$, (DE3)$_{225}$ and (DE5)$_{225}$ that

$$\begin{aligned}
E_P^{\mathrm{D}}(\eta_{[s,r]} - \eta_v \,|\, X_u = x_u) &= E_P^{\mathrm{D}}(\eta_{[s,r]} \,|\, X_u = x_u) - E_P^{\mathrm{D}}(\eta_v \,|\, X_u = x_u) \\
&= E_P^{\mathrm{D}}(\eta_{[s,r]} \,|\, X_u = x_u) - E_P(\eta_v \,|\, X_u = x_u),
\end{aligned}$$

and this proves the equality in the statement. □

## 6.B Proof of Lemma 6.6

**Lemma 6.6.** *Consider a jump process $P$ that has uniformly bounded rate, with rate bound $\lambda$. Fix a state history $\{X_u = x_u\}$ in $\mathscr{H}$, time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$, a grid $v$ over $[s, r]$, a limit variable $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ and* *an $\mathscr{F}_u$-simple variable $g$. If there are non-negative real numbers $\alpha, \beta, \gamma$ such that*

$$|f-g| \leq \alpha\Delta(v) + \beta\big(\eta_{[s,r]} - \eta_v\big) + \gamma\Delta(v)\eta_{[s,r]},$$

*then*

$$\left|E_P^{\mathrm{D}}(f \,|\, X_u = x_u) - E_P(g \,|\, X_u = x_u)\right| \leq \Delta(v)\left(\alpha + \frac{1}{4}\beta(r-s)\lambda^2 + \frac{1}{2}\gamma(r-s)\lambda\right).$$

*Proof.* To simplify our notation, we let $P^{x_u} := P(\bullet \,|\, X_u = x_u)$. Note that $P^{x_u}$ is a countably additive probability charge on $\mathscr{F}_u$ because $P$ is a countably additive jump process. Thus,

$$E_P^{\mathrm{D}}(f \,|\, X_u = x_u) - E_P(g \,|\, X_u = x_u) = E_{P^{x_u}}^{\mathrm{D}}(f) - E_{P^{x_u}}(g) = E_{P^{x_u}}^{\mathrm{D}}(f) - E_{P^{x_u}}^{\mathrm{D}}(g),$$

where the second equality follows from (DE1)$_{225}$ and where $E_{P^{x_u}}^{\mathrm{D}}$ is the Daniell extension of the Dunford integral $E_{P^{x_u}}$ corresponding to $P^{x_u}$. Because $g$ is an $\mathscr{F}_u$-simple variable by assumption, $-\infty < \min g \leq \max g < +\infty$; hence, $-\infty < E_{P^{x_u}}^{\mathrm{D}}(g) < +\infty$ due to (DE3)$_{225}$. For this reason, it follows from (DE5)$_{225}$ that $f-g$ belongs to $\mathbb{D}_{P^{x_u}}^{\mathrm{D}}$ and that

$$E_P^{\mathrm{D}}(f \,|\, X_u = x_u) - E_P(g \,|\, X_u = x_u) = E_{P^{x_u}}^{\mathrm{D}}(f) - E_{P^{x_u}}^{\mathrm{D}}(g) = E_{P^{x_u}}^{\mathrm{D}}(f-g). \tag{6.43}$$

Next, we set out to show using (DE5)$_{225}$ that

$$h := \alpha\Delta(v) + \beta\big(\eta_{[s,r]} - \eta_v\big) + \gamma\Delta(v)\eta_{[s,r]}$$

is D-integrable. For the first term, we recall from Lemma 2.39$_{36}$ that $\alpha\Delta(v)$ belongs to $\mathbb{S}(\mathscr{F}_u)$ and from (ES1)$_{37}$ that $E_{P^{x_u}}(\alpha\Delta(v)) = \alpha\Delta(v)$; hence, it follows from (DE1)$_{225}$ that $\alpha\Delta(v)$ belongs to $\mathbb{D}_{P^{x_u}}^{\mathrm{D}}$, and that $E_{P^{x_u}}^{\mathrm{D}}(\alpha\Delta(v)) = \alpha\Delta(v)$. For the second term, we recall from Proposition 6.2$_{275}$ that $\big(\eta_{[s,r]} - \eta_v\big)$ is a non-negative $\mathscr{F}_u$-over variable; hence, it follows from (DE2)$_{225}$, (DE4)$_{225}$ and Proposition 6.2$_{275}$ that $\beta\Delta(v)\big(\eta_{[s,r]} - \eta_v\big)$ is D-integrable, with

$$E_{P^{x_u}}^{\mathrm{D}}\big(\beta\big(\eta_{[s,r]} - \eta_v\big)\big) = \beta E_{P^{x_u}}^{\mathrm{D}}\big(\eta_{[s,r]} - \eta_v\big) \leq \frac{1}{4}\beta\Delta(v)(r-s)\lambda^2.$$

For the third term, we recall from Theorem 5.26$_{236}$ – and Theorem 5.12$_{227}$ – that $\eta_{[s,r]}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u) \subseteq \mathbb{D}_{P^{x_u}}^{\mathrm{D}}$; consequently, it follows from (DE4)$_{225}$ and Theorem 5.27$_{236}$ that $\gamma\Delta(v)\eta_{[s,r]}$ is D-integrable, with

$$E_{P^{x_u}}^{\mathrm{D}}\big(\gamma\Delta(v)\eta_{[s,r]}\big) = \gamma\Delta(v)E^{\mathrm{D}}\big(\eta_{[s,r]}\big) \leq \frac{1}{2}\gamma\Delta(v)(r-s)\lambda.$$

It follows from all this and (DE5)$_{225}$ that $h$ is D-integrable, with

$$E_{P^{x_u}}^{\mathrm{D}}(h) \leq \Delta(v)\left(\alpha + \frac{1}{4}\beta(r-s)\lambda^2 + \frac{1}{2}\gamma(r-s)\lambda\right). \tag{6.44}$$

Furthermore, it follows from this and (DE4)$_{225}$ that $-h$ is D-integrable with

$$E_{P^{x_u}}^{\mathrm{D}}(-h) = -E_{P^{x_u}}^{\mathrm{D}}(h). \tag{6.45}$$

As $-h \leq f - g \leq h$ by assumption, it follows from (DE6)$_{226}$ that

$$E^{\mathrm{D}}_{Px_u}(-h) \leq E^{\mathrm{D}}_{Px_u}(f-g) \leq E^{\mathrm{D}}_{Px_u}(h);$$

we substitute Eqs. (6.44) and (6.45) in the preceding expression, to yield

$$\left|E^{\mathrm{D}}_{Px_u}(f-g)\right| \leq \Delta(v)\left(\alpha + \frac{1}{4}\beta(r-s)\lambda^2 + \frac{1}{2}\gamma(r-s)\lambda\right),$$

where we also used that $\alpha, \beta, \gamma$ are non-negative by assumption. Finally, the inequality in the statement follows from this inequality and Eq. (6.43). □

## 6.C Proof of Lemmas 6.43 and 6.44

In this appendix, we prove the two results in Section 6.5.1$_{311}$. The first one is Lemma 6.43$_{311}$.

**Lemma 6.43.** *Consider a lower rate operator* $\underline{Q}$ *and a sequence of time points* $v = (t_0, \ldots, t_n)$ *in* $\mathscr{U}_{\neq()}$ *such that* $\Delta(v)\|\underline{Q}\|_{\mathrm{op}} \leq 2$*. Fix gambles* $g_0, \ldots, g_n$ *and* $h_0, \ldots, h_{n-1}$ *in* $\mathbb{G}(\mathscr{X})$ *such that* $0 \leq h_k \leq 1$ *for all* $k$ *in* $\{0, \ldots, n-1\}$*. Let* $f_0$ *and* $\tilde{f}_0$ *be the gambles on* $\mathscr{X}$ *defined by the initial condition* $f_n := g_n =: \tilde{f}_n$ *and, for all* $k$ *in* $\{0, \ldots, n-1\}$*, by the recursive relations*

$$f_k := g_k + h_k e^{\Delta_{k+1}\underline{Q}} f_{k+1} \quad \text{and} \quad \tilde{f}_k := g_k + h_k(I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1},$$

*where for all* $k$ *in* $\{0, \ldots, n-1\}$*, we let* $\Delta_{k+1} := (t_{k+1} - t_k)$*. Then*

$$\left\|f_0 - \tilde{f}_0\right\| \leq \frac{1}{2}\|\underline{Q}\|^2_{\mathrm{op}} \sum_{k=1}^{n} \Delta_k^2 \|\tilde{f}_k\|_{\mathrm{c}}.$$

*Proof.* Fix some $k$ in $\{0, \ldots, n-1\}$. Then by definition of $f_k$ and $\tilde{f}_k$,

$$\begin{aligned}\left\|f_k - \tilde{f}_k\right\| &= \left\|g_k + h_k e^{\Delta_{k+1}\underline{Q}} f_{k+1} - g_k - h_k(I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1}\right\| \\ &= \left\|h_k e^{\Delta_{k+1}\underline{Q}} f_{k+1} - h_k(I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1}\right\|.\end{aligned}$$

Note that for all $x$ in $\mathscr{X}$,

$$\begin{aligned}&\left|h_k(x)\left[e^{\Delta_{k+1}\underline{Q}} f_{k+1}\right](x) - h_k(x)\left[(I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1}\right](x)\right| \\ &\quad = |h_k(x)|\left|\left[e^{\Delta_{k+1}\underline{Q}} f_{k+1}\right](x) - \left[(I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1}\right](x)\right| \\ &\quad \leq \|h_k\|\left\|e^{\Delta_{k+1}\underline{Q}} f_{k+1} - (I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1}\right\|.\end{aligned}$$

Consequently,

$$\begin{aligned}\left\|f_k - \tilde{f}_k\right\| &\leq \|h_k\|\left\|e^{\Delta_{k+1}\underline{Q}} f_{k+1} - (I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1}\right\| \\ &\leq \left\|e^{\Delta_{k+1}\underline{Q}} f_{k+1} - (I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1}\right\|,\end{aligned}$$

where for the second inequality we used that $\|h_k\| \leq 1$ because $0 \leq h_k \leq 1$ by assumption. As we have done many times before in Chapter 4$_{157}$, we execute the classic trick of adding and subtracting something – here $e^{\Delta_{k+1}\underline{Q}}\tilde{f}_{k+1}$ – to subsequently invoke the triangle inequality (N2)$_{76}$:

$$\|f_k - \tilde{f}_k\| \leq \|e^{\Delta_{k+1}\underline{Q}} f_{k+1} - e^{\Delta_{k+1}\underline{Q}}\tilde{f}_{k+1}\| + \|e^{\Delta_{k+1}\underline{Q}}\tilde{f}_{k+1} - (I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1}\|. \quad (6.46)$$

We can bound the two terms on the right-hand side of this inequality with the help of some earlier results. For the first term, we recall from Proposition 3.74$_{114}$ that $e^{\Delta_{k+1}\underline{Q}}$ is a lower transition operator, so it follows from (LT8)$_{108}$ that

$$\|e^{\Delta_{k+1}\underline{Q}} f_{k+1} - e^{\Delta_{k+1}\underline{Q}}\tilde{f}_{k+1}\| \leq \|f_{k+1} - \tilde{f}_{k+1}\|.$$

For the second term, we observe that $\Delta_{k+1}\|\underline{Q}\|_{\mathrm{op}} \leq 2$, because $\Delta(v)\|\underline{Q}\|_{\mathrm{op}} \leq 2$ by assumption and $\Delta_{k+1} \leq \Delta(v)$. Hence, it follows from Lemma 4.19$_{180}$ that

$$\|e^{\Delta_{k+1}\underline{Q}}\tilde{f}_{k+1} - (I + \Delta_{k+1}\underline{Q})\tilde{f}_{k+1}\| \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \Delta_{k+1}^2 \|\tilde{f}_{k+1}\|_{\mathrm{c}}.$$

We use the two preceding inequalities to bound the two terms on the right-hand side of the inequality in Eq. (6.46), to yield

$$\|f_k - \tilde{f}_k\| \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \Delta_{k+1}^2 \|\tilde{f}_{k+1}\|_{\mathrm{c}} + \|f_{k+1} - \tilde{f}_{k+1}\|. \quad (6.47)$$

To obtain the inequality in the statement, we apply Eq. (6.47) for $k$ ranging from 0 to $n-1$:

$$\|f_0 - \tilde{f}_0\| \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \Delta_1^2 \|\tilde{f}_1\|_{\mathrm{c}} + \|f_1 - \tilde{f}_1\| \leq \cdots \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \sum_{k=1}^{n} \Delta_k^2 \|\tilde{f}_k\|_{\mathrm{c}}. \qquad \square$$

The second result in Section 6.5.1$_{311}$ is Lemma 6.44$_{312}$; as we will see now, its proof is fairly straightforward.

**Lemma 6.44.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Fix some time points $s, r$ in $\mathbb{R}_{\geq 0}$ and an idealised variable $f$ such that for all $u$ in $\mathscr{U}$ with* $\max u \leq s$*, $f$ belongs to $\overline{\mathbb{V}}_{\mathrm{lim}}(\mathscr{F}_u)$. If for all $\epsilon$ in $\mathbb{R}_{>0}$, there is some grid $v$ over $[s, r]$ and some real variable $g$ in $\mathbb{V}(\Omega)$ with a sum-product representation over $v$ such that, for all $\{X_u = x_u\}$ in $\mathscr{H}$ with* $\max u \leq s$,

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u) - \underline{E}_{\mathscr{P}}(g \,|\, X_u = x_u)\right| \leq \epsilon,$$

*then the map $f_s$ on $\mathscr{X}$ defined by*

$$f_s(x) := \underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_s = x) \quad \text{for all } x \in \mathscr{X}$$

*is a gamble on $\mathscr{X}$, and for all $\{X_u = x_u\}$ in $\mathscr{H}$ with* $\max u \leq s$,

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_u = x_u) = \underline{E}_{\mathscr{P}}\big(\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \,|\, X_s) \,\big|\, X_u = x_u\big) = \underline{E}_{\mathscr{P}}(f_s(X_s) \,|\, X_u = x_u).$$

*Proof.* First, we prove that $f_s$ is a gamble on $\mathscr{X}$. Fix any $\epsilon$ in $\mathbb{R}_{>0}$. Then by the condition in the statement, there is some grid $v$ over $[s,r]$ and some real variable $g$ in $\mathbb{V}(\Omega)$ with a sum-product representation over $v$ such that

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) - \underline{E}_{\mathscr{P}}(g \mid X_u = x_u)\right| \leq \epsilon \tag{6.48}$$

and, for all $x$ in $\mathscr{X}$,

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_s = x) - \underline{E}_{\mathscr{P}}(g \mid X_s = x)\right| \leq \epsilon. \tag{6.49}$$

Because $g$ has a sum-product representation over $v$, it is an $\mathscr{F}_{(s)}$-simple variable – and therefore a bounded one – due to Lemma 4.6$_{164}$. Hence, it follows from (ES1)$_{37}$ that $\underline{E}_{\mathscr{P}}(g \mid X_s = x)$ is real for all $x$ in $\mathscr{X}$; due to Eq. (6.49), this implies that $f_s$ is a gamble. The same argument shows that $\underline{E}_{\mathscr{P}}(g \mid X_u = x_u)$ is real-valued, and therefore, due to Eq. (6.48), also $\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u)$

Second, we fix any $\{X_u = x_u\}$ in $\mathscr{H}$ with $\max u \leq s$, and set out to prove that

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\big(\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_s) \,\big|\, X_u = x_u\big) = \underline{E}_{\mathscr{P}}(f_s(X_s) \mid X_u = x_u). \tag{6.50}$$

Again, we fix any $\epsilon$ in $\mathbb{R}_{>0}$; then by assumption, there is some grid $v$ over $[s,r]$ and some real variable $g$ in $\mathbb{V}(\Omega)$ with a sum-product representation over $v$ such that

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) - \underline{E}_{\mathscr{P}}(g \mid X_u = x_u)\right| \leq \frac{1}{2}\epsilon \tag{6.51}$$

and

$$|f_s(X_s) - g_s(X_s)| = \left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_s) - \underline{E}_{\mathscr{P}}(g \mid X_s)\right| \leq \frac{1}{2}\epsilon, \tag{6.52}$$

where we let $g_s$ be the gamble on $\mathscr{X}$ defined by $g_s(x) := \underline{E}_{\mathscr{P}}(g \mid X_s = x)$. Because $g$ has a sum-product representation over $v$ and because $\max u \leq s = \min v$ by assumption, it follows from Theorem 4.9$_{166}$ that

$$\underline{E}_{\mathscr{P}}(g \mid X_u = x_u) = \underline{E}_{\mathscr{P}}\big(\underline{E}_{\mathscr{P}}(g \mid X_s) \,\big|\, X_u = x_u\big) = \underline{E}_{\mathscr{P}}\big(g_s(X_s) \,\big|\, X_u = x_u\big). \tag{6.53}$$

Recall from Eq. (6.52) that $|f_s(X_s) - g_s(X_s)| \leq \frac{1}{2}\epsilon$. Hence, for all $P$ in $\mathscr{P}$, it follows from Lemma 2.39$_{36}$, (E5)$_{22}$ and (E6)$_{22}$ – which we may use due to Proposition 2.43$_{38}$ – that

$$E_P\big(f_s(X_s) \,\big|\, X_u = x_u\big) - \frac{1}{2}\epsilon \leq E_P\big(g_s(X_s) \,\big|\, X_u = x_u\big) \leq E_P\big(f_s(X_s) \,\big|\, X_u = x_u\big) + \frac{1}{2}\epsilon.$$

Because $f_s(X_s)$ and $g_s(X_s)$ are trivially $\mathscr{F}_u$-simple, it follows from this inequality, Lemma 6.1$_{274}$ and (DE1)$_{225}$ that

$$\left|\underline{E}_{\mathscr{P}}\big(g_s(X_s) \,\big|\, X_u = x_u\big) - \underline{E}_{\mathscr{P}}\big(f_s(X_s) \,\big|\, X_u = x_u\big)\right| \leq \frac{1}{2}\epsilon. \tag{6.54}$$

Observe that – because $\underline{E}_{\mathscr{P}}(f_s(X_s) \mid X_u = x_u)$ and $\underline{E}_{\mathscr{P}}(g \mid X_u = x_u)$ are real-valued –

$$\begin{aligned}
&\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) - \underline{E}_{\mathscr{P}}(f_s(X_s) \mid X_u = x_u)\right| \\
&\qquad \leq \left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) - \underline{E}_{\mathscr{P}}(g \mid X_u = x_u)\right| \\
&\qquad\qquad + \left|\underline{E}_{\mathscr{P}}(g \mid X_u = x_u) - \underline{E}_{\mathscr{P}}(f_s(X_s) \mid X_u = x_u)\right|.
\end{aligned}$$

We substitute Eqs. (6.51), (6.53) and (6.54), to yield

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}(f \mid X_u = x_u) - \underline{E}_{\mathscr{P}}(f_s(X_s) \mid X_u = x_u)\right| \leq \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon.$$

This inequality holds for arbitrary $\epsilon$ in $\mathbb{R}_{>0}$, so this clearly implies Eq. (6.50), as required. □

## 6.D Proof of Theorems 6.46 to 6.50

In the last appendix to this chapter, we prove the three results in Section $6.5.2_{313}$. First, we prove Theorem $6.46_{314}$.

**Theorem 6.46.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Fix subsets $S, G$ of $\mathscr{X}$ and time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $s < r$. For all $n$ in $\mathbb{N}$, we let $\Delta_n := {}^{(r-s)}/_{n}$ and let $\tilde{f}_{n,0}$ be the gamble on $\mathscr{X}$ that is defined by the initial condition $\tilde{f}_{n,n} := \mathbb{I}_G$ and, for all $k$ in $\{0, \dots n-1\}$, by the recursive relation*

$$\tilde{f}_{n,k} := \mathbb{I}_G + \mathbb{I}_{S\setminus G}\big(I + \Delta_n \underline{Q}_{\mathscr{Q}}\big)\tilde{f}_{n,k+1}. \tag{6.34}$$

*Then for all $x$ in $\mathscr{X}$ and $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}_{\mathscr{Q}}\|_{\mathrm{op}} \leq 2n$,*

$$\big|\underline{P}^{\mathrm{D}}_{\mathscr{P}}\big(H^{S,G}_{[s,r]} \,\big|\, X_s = x\big) - \tilde{f}_{n,0}(x)\big| \leq \frac{3}{8}\frac{(r-s)^2}{n}\|\underline{Q}_{\mathscr{Q}}\|^2_{\mathrm{op}},$$

*and therefore*

$$\underline{P}^{\mathrm{D}}_{\mathscr{P}}\big(H^{S,G}_{[s,r]} \,\big|\, X_s = x\big) = \lim_{n\to+\infty} \tilde{f}_{n,0}(x).$$

*The same holds for $\overline{P}^{\mathrm{D}}_{\mathscr{P}}$ if in Eq. $(6.34)_{314}$ we replace $\underline{Q}_{\mathscr{Q}}$ by $\overline{Q}_{\mathscr{Q}}$.*

*Proof.* Let $\underline{Q} := \underline{Q}_{\mathscr{Q}}$. Because every rate operator $Q$ in $\mathscr{Q}$ dominates $\underline{Q}$, it follows immediately from $(\mathrm{LR7})_{111}$ that $\|\mathscr{Q}\|_{\mathrm{op}} \leq \|\underline{Q}\|_{\mathrm{op}}$.

Fix some $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}\|_{\mathrm{op}} \leq 2n$, and let $v$ be the grid over $[s,r]$ with $n$ subintervals of length $\Delta_n$ – that is, we let $v := (s, s+\Delta_n, \dots, s+n\Delta_n)$. Then by Proposition $6.13_{283}$,

$$\begin{aligned}\Big|\underline{P}^{\mathrm{D}}_{\mathscr{P}}\big(H^{S,G}_{[s,r]} \,\big|\, X_s = x\big) - \underline{E}_{\mathscr{P}}\big(h^{S,G}_{v} \,\big|\, X_s = x\big)\Big| &\leq \frac{1}{8}\Delta(v)(r-s)\|\mathscr{Q}\|^2_{\mathrm{op}}\\ &\leq \frac{1}{8}\frac{(r-s)^2}{n}\|\underline{Q}\|^2_{\mathrm{op}},\end{aligned} \tag{6.55}$$

where for the second inequality we used that $\Delta(v) = {}^{(r-s)}/_{n}$ and that $\|\mathscr{Q}\|_{\mathrm{op}} \leq \|\underline{Q}\|_{\mathrm{op}}$.

Recall from Lemma $6.10_{281}$ that $h^{S,G}_v$ has a sum-product representation over $v$:

$$h^{S,G}_v = \sum_{k=0}^{n} g_k\big(X_{s+k\Delta_n}\big)\prod_{\ell=0}^{k-1} h_\ell\big(X_{s+\ell\Delta_n}\big),$$

with $g_k := \mathbb{I}_G$ for all $k$ in $\{0, \dots, n\}$ and $h_\ell := \mathbb{I}_{S\setminus G}$ for all $\ell$ in $\{0, \dots, n-1\}$. For this reason, it follows from Theorem $4.9_{166}$ that

$$\underline{E}_{\mathscr{P}}\big(h^{S,G}_v \,\big|\, X_s = x\big) = f_{n,0}(x), \tag{6.56}$$

where $f_{n,0}$ is the gamble on $\mathscr{X}$ that is defined by the initial condition $f_{n,n} := g_n = \mathbb{I}_G$ and, for all $k$ in $\{0, \dots, n-1\}$, by the recursive relation

$$f_{n,k} := \mathbb{I}_G + \mathbb{I}_{S\setminus G}e^{\Delta_n \underline{Q}} f_{n,k+1}.$$

Furthermore, it follows from Lemma 6.43<sub>311</sub> that

$$\left|f_{n,0}(x)-\tilde{f}_{n,0}(x)\right| \leq \left\|f_{n,0}-\tilde{f}_{n,0}\right\|_{\mathrm{op}} \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \frac{(r-s)^2}{n^2} \sum_{k=1}^{n} \|\tilde{f}_{n,k}\|_{\mathrm{c}}. \tag{6.57}$$

We now claim that for all $k$ in $\{1,\ldots,n\}$, $\min \tilde{f}_{n,k} \geq 0$ and $\max \tilde{f}_{n,k} \leq 1$, and therefore $\|\tilde{f}_{n,k}\|_{\mathrm{c}} \leq \frac{1}{2}$. Our proof will be one by induction. For the base case $k=n$, this is obvious because $\tilde{f}_{n,n} = \mathbb{I}_G$ by definition. For the inductive step, we fix some $k$ in $\{1,\ldots,n-1\}$ and assume that $\min \tilde{f}_{n,k+1} \geq 0$ and $\max \tilde{f}_{n,k+1} \leq 1$. Because $\Delta_n\|\underline{Q}\|_{\mathrm{op}} \leq 2$, $(I+\Delta_n\underline{Q})$ is a lower transition operator due to Lemma $3.72_{112}$. Hence, it follows from the induction hypothesis and (LT4)$_{108}$ that

$$0 \leq \min \tilde{f}_{n,k+1} \leq (I+\Delta_n\underline{Q})\tilde{f}_{n,k+1} \leq \max \tilde{f}_{n,k+1} \leq 1.$$

For this reason, and because $\tilde{f}_{n,k} = \mathbb{I}_G + \mathbb{I}_{S\setminus G}(I+\Delta_n\underline{Q})\tilde{f}_{n,k+1}$ by definition, we see that $\min \tilde{f}_{n,k} \geq 0$ and $\max \tilde{f}_{n,k} \leq 1$, as required.

Because $\|\tilde{f}_{n,k}\|_{\mathrm{c}} \leq 1/2$ for all $k$ in $\{1,\ldots,n\}$, it follows from Eq. (6.57) that

$$\left|f_{n,0}(x)-\tilde{f}_{n,0}(x)\right| \leq \frac{1}{4}\|\underline{Q}\|_{\mathrm{op}}^2 \frac{(r-s)^2}{n}. \tag{6.58}$$

Finally, it follows from Eqs. (6.55)$_{\curvearrowleft}$, (6.56)$_{\curvearrowleft}$ and (6.58) and the triangle inequality that

$$\begin{aligned}\left|\underline{P}_{\mathscr{P}}^{\mathrm{D}}\left(H_{[s,r]}^{S,G} \,\middle|\, X_s = x\right) - \tilde{f}_{n,0}(x_s)\right| &\leq \frac{1}{8}\frac{(r-s)^2}{n}\|\underline{Q}\|_{\mathrm{op}}^2 + \frac{1}{4}\|\underline{Q}\|_{\mathrm{op}}^2\frac{(r-s)^2}{n} \\ &= \frac{3}{8}\frac{(r-s)^2}{n}\|\underline{Q}\|_{\mathrm{op}}^2. \end{aligned} \tag{6.59}$$

Because Eq. (6.59) holds for all $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}\|_{\mathrm{op}} \leq 2n$ and because the right-hand side of the inequality vanishes as $n$ recedes to $+\infty$, we have proven the limit statement for $\underline{P}_{\mathscr{P}}^{\mathrm{D}}$.

The statement for $\overline{P}_{\mathscr{P}}^{\mathrm{D}}$ essentially follows from conjugacy. More precisely, the argument is almost exactly the same as the argument in the first part of this proof. We do need a couple of extra steps though. First, we use that

$$\overline{E}_{\mathscr{P}}\left(h_v^{S,G} \,\middle|\, X_s = x\right) = -\underline{E}_{\mathscr{P}}\left(-h_v^{S,G} \,\middle|\, X_s = x\right),$$

for any grid $v$ over $[s,r]$. Second, we use that $-h_v^{S,G}$ also has a sum-product representation over $v$: by Lemma $4.7_{165}$,

$$-h_v^{S,G} = \sum_{k=0}^{n} [-\mathbb{I}_G]\left(X_{s+k\Delta_n}\right) \prod_{\ell=0}^{k-1} \mathbb{I}_{S\setminus G}\left(X_{s+\ell\Delta_n}\right)$$

Third, we again use Lemmas $4.9_{166}$ and $6.43_{311}$, but this time to approximate $\underline{E}_{\mathscr{P}}\left(-h_v^{S,G} \,\middle|\, X_s = x\right)$ instead of $\underline{E}_{\mathscr{P}}\left(h_v^{S,G} \,\middle|\, X_s = x\right)$. This way, we find that

$$\left|\underline{E}_{\mathscr{P}}(-h_v^{S,G} \mid X_s = x) - \check{f}_{n,0}(x)\right| \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \frac{(r-s)^2}{n^2} \sum_{k=1}^{n} \|\check{f}_{n,k}\|_{\mathrm{c}},$$

where $\check{f}_{n,0}(x)$ is recursively defined by the initial condition $\check{f}_{n,n} := -\mathbb{I}_G$ and, for all $k$ in $\{0,\dots,n-1\}$, by the recursive relation

$$\check{f}_{n,k} := -\mathbb{I}_G + \mathbb{I}_{S\setminus G}\big(I + \Delta_n \underline{Q}\big)\check{f}_{n,k+1}.$$

Obviously, $\tilde{f}_{n,n} = -\check{f}_{n,n}$. Furthermore, it is easy to verify that, for all $k$ in $\{0,\dots,n-1\}$, $\|\check{f}_{n,k}\|_{\mathrm{c}} \le 1/2$ and that, by conjugacy,

$$\tilde{f}_{n,k} = \mathbb{I}_G + \mathbb{I}_{S\setminus G}\big(I + \Delta_n \overline{Q}\big)\tilde{f}_{n,k} = -\Big(-\mathbb{I}_G + \mathbb{I}_{S\setminus G}\big(I + \Delta_n \underline{Q}\big)(-\tilde{f}_{n,k+1})\Big) = -\check{f}_{n,k}.$$

Therefore, and because $\overline{E}_{\mathscr{P}}\big(h_v^{S,G} \,\big|\, X_s = x\big) = -\underline{E}_{\mathscr{P}}\big(-h_v^{S,G} \,\big|\, X_s = x\big)$,

$$\big|\overline{E}_{\mathscr{P}}(h_v^{S,G} \,|\, X_s = x) - \tilde{f}_{n,0}(x)\big| \le \frac{1}{4}\|\underline{Q}\|_{\mathrm{op}}^2 \frac{(r-s)^2}{n}.$$

The remainder of the proof is again similar to the first part of the proof. □

Next, we prove Theorem 6.48~316~. Note that the proof is largely similar to the proof of Theorem 6.46~314~.

**Theorem 6.48.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Fix some subset $G$ of $\mathscr{X}$ and time points $s, r$ in $\mathbb{R}_{\ge 0}$ such that $s < r$. For all $n$ in $\mathbb{N}$, we let $\Delta_n := (r-s)/n$ and let $\tilde{f}_{n,0}$ be the gamble on $\mathscr{X}$ that is defined by the initial condition $\tilde{f}_{n,n} := \Delta_n$ and, for all $k$ in $\{0,\dots n-1\}$, by the recursive relation*

$$\tilde{f}_{n,k} := \begin{cases} \Delta_n + \mathbb{I}_{G^{\mathrm{c}}}\big(I + \Delta_n \underline{Q}_{\mathscr{Q}}\big)\tilde{f}_{n,k+1} & \text{if } k \ge 1, \\ s + \mathbb{I}_{G^{\mathrm{c}}}\big(I + \Delta_n \underline{Q}_{\mathscr{Q}}\big)\tilde{f}_{n,k+1} & \text{if } k = 0. \end{cases} \tag{6.35}$$

*Then for all $x$ in $\mathscr{X}$ and all $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}_{\mathscr{Q}}\|_{\mathrm{op}} \le 2$,*

$$\big|\underline{E}^{\mathrm{D}}_{\mathscr{P}}\big(\tau^G_{[s,r]} \,\big|\, X_s = x\big) - \tilde{f}_{n,0}(x)\big| \le \frac{r-s}{n} + \frac{1}{8}\frac{(r-s)^3}{n}\frac{2n+1}{n}\|\underline{Q}_{\mathscr{Q}}\|_{\mathrm{op}}^2,$$

*and therefore*

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}\big(\tau^G_{[s,r]} \,\big|\, X_s = x\big) = \lim_{n\to+\infty} \tilde{f}_{n,0}(x).$$

*The same holds for $\overline{E}^{\mathrm{D}}_{\mathscr{P}}$ if in Eq. (6.35)~317~ we replace $\underline{Q}_{\mathscr{Q}}$ by $\overline{Q}_{\mathscr{Q}}$.*

*Proof.* Let $\underline{Q} := \underline{Q}_{\mathscr{Q}}$. Because every rate operator $Q$ in $\mathscr{Q}$ dominates $\underline{Q}$, it follows immediately from (LR7)~111~ that $\|\mathscr{Q}\|_{\mathrm{op}} \le \|\underline{Q}\|_{\mathrm{op}}$.

Fix some $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}\|_{\mathrm{op}} \le 2n$, and let $v$ be the grid over $[s,r]$ with $n$ subintervals of length $\Delta_n$ – that is, we let $v := (s, s+\Delta_n, \dots, s+n\Delta_n)$. Then by Proposition 6.22~293~,

$$\begin{aligned}\Big|\underline{E}^{\mathrm{D}}_{\mathscr{P}}\big(\tau^G_{[s,r]} \,\big|\, X_s = x\big) - \underline{E}_{\mathscr{P}}\big(\tau^G_v \,\big|\, X_s = x\big)\Big| &\le \Delta(v) + \frac{1}{8}\Delta(v)(r-s)^2\|\mathscr{Q}\|_{\mathrm{op}}^2 \\ &\le \frac{r-s}{n} + \frac{1}{8}\frac{(r-s)^3}{n}\|\underline{Q}\|_{\mathrm{op}}^2,\end{aligned} \tag{6.60}$$

where for the second inequality we used that $\Delta(v) = {}^{(r-s)}/n$ and that $\|\mathcal{Q}\|_{\mathrm{op}} \leq \|\underline{Q}\|_{\mathrm{op}}$.

Recall from Lemma 6.18<sub></sub> that $\tau_v^G$ has a sum-product representation over $v$:

$$\tau_v^G = \sum_{k=0}^{n} g_k(X_{s+k\Delta}) \prod_{\ell=0}^{k-1} h_\ell(X_{s+\ell\Delta}),$$

with $g_0 \coloneqq s$ and, for all $k$ in $\{1,\dots,n\}$, $g_k \coloneqq \Delta_n$ and $h_{k-1} \coloneqq \mathbb{I}_{G^c}$. For this reason, it follows from Theorem $4.9_{166}$ that

$$\underline{E}_{\mathscr{P}}(\tau_v^G \mid X_s = x) = f_{n,0}(x), \tag{6.61}$$

where $f_{n,0}$ is the gamble on $\mathscr{X}$ that is defined by the initial condition $f_{n,n} \coloneqq g_n$ and, for all $k$ in $\{0,\dots,n-1\}$, by the recursive relation

$$f_{n,k} \coloneqq g_k + h_{k-1} e^{\Delta_n \underline{Q}} f_{n,k+1}.$$

Furthermore, it follows from Lemma $6.43_{311}$ that

$$\left|f_{n,0}(x) - \tilde{f}_{n,0}(x)\right| \leq \left\|f_{n,0} - \tilde{f}_{n,0}\right\|_{\mathrm{op}} \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \frac{(r-s)^2}{n^2} \sum_{k=1}^{n} \|\tilde{f}_{n,k}\|_{\mathrm{c}}. \tag{6.62}$$

We now claim that for all $k$ in $\{1,\dots,n\}$, $\min \tilde{f}_{n,k} \geq 0$ and $\max \tilde{f}_{n,k} \leq (n-k+1)\Delta_n$, and therefore $\|\tilde{f}_{n,k}\|_{\mathrm{c}} \leq \frac{(n-k+1)\Delta_n}{2}$. Our proof will be one by induction. For the base case $k = n$, this is obvious because $\tilde{f}_{n,n} = \Delta_n$ by definition. For the inductive step, we fix some $k$ in $\{1,\dots,n-1\}$ and assume that $\min \tilde{f}_{n,k+1} \geq 0$ and $\max \tilde{f}_{n,k+1} \leq (n-k)\Delta_n$. Because $\Delta_n\|\underline{Q}\|_{\mathrm{op}} \leq 2$, $(I + \Delta_n \underline{Q})$ is a lower transition operator due to Lemma $3.72_{112}$. Hence, it follows from the induction hypothesis and (LT4)$_{108}$ that

$$0 \leq \min \tilde{f}_{n,k+1} \leq (I + \Delta_n \underline{Q})\tilde{f}_{n,k+1} \leq \max \tilde{f}_{n,k+1} \leq (n-k)\Delta_n.$$

For this reason, and because $\tilde{f}_{n,k} = \Delta_n + \mathbb{I}_{G^c}(I + \Delta_n \underline{Q})\tilde{f}_{n,k+1}$ by definition, we see that $\min \tilde{f}_{n,k} \geq 0$ and $\max \tilde{f}_{n,k} \leq (n-k+1)\Delta_n$, as required.

Because $\|\tilde{f}_{n,k}\|_{\mathrm{c}} \leq \frac{(n-k+1)\Delta_n}{2}$ for all $k$ in $\{1,\dots,n\}$, it follows from Eq. (6.62) that

$$\begin{aligned} \left|f_{n,0}(x) - \tilde{f}_{n,0}(x)\right| &\leq \frac{1}{4}\|\underline{Q}\|_{\mathrm{op}}^2 \frac{(r-s)^3}{n^3} \sum_{k=1}^{n} (n-k+1) \\ &= \frac{1}{4}\|\underline{Q}\|_{\mathrm{op}}^2 \frac{(r-s)^3}{n^3} \frac{n(n+1)}{2} \\ &= \frac{1}{8}\|\underline{Q}\|_{\mathrm{op}}^2 \frac{(r-s)^3}{n} \frac{n+1}{n}, \end{aligned} \tag{6.63}$$

where for the inequality we also used that $\Delta_n = {}^{(r-s)}/n$. Finally, it follows from Eqs. (6.60)↶, (6.61) and (6.63) and the triangle inequality that

$$\begin{aligned} \left|\underline{E}_{\mathscr{P}}^{\mathrm{D}}(\tau_{[s,r]}^G \mid X_s = x) - \tilde{f}_{n,0}(x)\right| &\leq \frac{r-s}{n} + \frac{1}{8}\frac{(r-s)^3}{n}\|\underline{Q}\|_{\mathrm{op}}^2 + \frac{1}{8}\frac{(r-s)^3}{n}\frac{n+1}{n}\|\underline{Q}\|_{\mathrm{op}}^2 \\ &= \frac{r-s}{n} + \frac{1}{8}\frac{(r-s)^3}{n}\frac{2n+1}{n}\|\underline{Q}\|_{\mathrm{op}}^2. \end{aligned} \tag{6.64}$$

Because Eq. (6.64)↶ holds for all $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}\|_{\mathrm{op}} \leq 2n$ and because the right-hand side of the inequality vanishes as $n$ recedes to $+\infty$, we have proven the limit statement for $\underline{E}^{\mathrm{D}}_{\mathscr{P}}$.

The statement for $\overline{E}^{\mathrm{D}}_{\mathscr{P}}$ essentially follows from conjugacy; as in the proof of Theorem 6.46<sub>314</sub>, we need some obvious extra/different steps. □

Finally, we prove Theorem 6.50$_{318}$. The proof is similar to the proofs of Theorems 6.46$_{314}$ and 6.48$_{316}$, although this time around we need to invoke Proposition 6.38$_{304}$.

**Theorem 6.50.** *Consider a non-empty set $\mathscr{M}$ of initial mass functions, a non-empty and bounded set $\mathscr{Q}$ of rate operators that has separately specified rows and an imprecise jump process $\mathscr{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \subseteq \mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Fix some time points $s, r$ in $\mathbb{R}_{\geq 0}$ such that $s < r$ and a Lipschitz continuous family $(f_t)_{t\in[s,r]}$ of gambles on $\mathscr{X}$ with Lipschitz constant $\kappa$, and let*

$$\gamma := \sup\{\max f_t \colon t \in [s,r]\} - \inf\{\min f_t \colon t \in [s,r]\}.$$

*For all $n$ in $\mathbb{N}$, we let $\Delta_n := {}^{(r-s)}/_{n}$ and let $\tilde{f}_{n,0}$ be the gamble on $\mathscr{X}$ that is defined by the initial condition $\tilde{f}_{n,n} := \Delta_n f_r$ and, for all $k$ in $\{0,\dots n-1\}$, by the recursive relation*

$$\tilde{f}_{n,k} := \begin{cases} \Delta_n f_{s+k\Delta_n} + \big(I + \Delta_n \underline{Q}_{\mathscr{Q}}\big)\tilde{f}_{n,k+1} & \text{if } k \geq 1, \\ \big(I + \Delta_n \underline{Q}_{\mathscr{Q}}\big)\tilde{f}_{n,k+1} & \text{if } k = 0. \end{cases} \tag{6.36}$$

*Then for all $x$ in $\mathscr{X}$ and $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}_{\mathscr{Q}}\|_{\mathrm{op}} \leq 2n$,*

$$\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}\left(\int_s^r f_t(X_t)\,\mathrm{d}t \,\middle|\, X_s = x\right) - \tilde{f}_{n,0}(x)\right| \leq \frac{(r-s)^2}{n}\left(\kappa + \frac{\gamma}{2}\|\underline{Q}\|_{\mathrm{op}} + \frac{1}{8}(r-s)\frac{n+1}{n}\|\underline{Q}\|^2_{\mathrm{op}}\right),$$

*and therefore*

$$\underline{E}^{\mathrm{D}}_{\mathscr{P}}\left(\int_s^r f_t(X_t)\,\mathrm{d}t \,\middle|\, X_s = x\right) = \lim_{n\to+\infty} \tilde{f}_{n,0}(x).$$

*The same holds for $\overline{E}^{\mathrm{D}}_{\mathscr{P}}$ if in Eq. (6.36)$_{319}$ we replace $\underline{Q}_{\mathscr{Q}}$ by $\overline{Q}_{\mathscr{Q}}$.*

*Proof.* Let $\underline{Q} := \underline{Q}_{\mathscr{Q}}$. Because every rate operator $Q$ in $\mathscr{Q}$ dominates $\underline{Q}$, it follows immediately from (LR7)$_{111}$ that $\|\mathscr{Q}\|_{\mathrm{op}} \leq \|\underline{Q}\|_{\mathrm{op}}$.

Fix some $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}\|_{\mathrm{op}} \leq 2n$, and let $v$ be the grid over $[s,r]$ with $n$ subintervals of length $\Delta_n$ – that is, we let $v := (s, s+\Delta_n, \dots, s+n\Delta_n)$. Then by Proposition 6.38$_{304}$,

$$\begin{aligned}\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}\left(\int_s^r f(X_t)\,\mathrm{d}t \,\middle|\, X_s = x\right) - \underline{E}_{\mathscr{P}}(\langle f_\bullet\rangle_v \,|\, X_s = x)\right| &\leq \Delta(v)(r-s)\left(\kappa + \frac{\gamma}{2}\|\mathscr{Q}\|_{\mathrm{op}}\right) \\ &\leq \frac{(r-s)^2}{n}\left(\kappa + \frac{\gamma}{2}\|\underline{Q}\|_{\mathrm{op}}\right),\end{aligned} \tag{6.65}$$

where for the second inequality we used that $\Delta(v) = (r-s)/n$ and that $\|\mathscr{Q}\|_{\mathrm{op}} \leq \|\underline{Q}_{\mathscr{Q}}\|_{\mathrm{op}}$.

Recall from the proof of Lemma 6.33300 that $\langle f_\bullet\rangle_v$ has a sum-product representation over $v$:

$$\langle f_\bullet\rangle_v = \sum_{k=0}^{n} g_k\big(X_{s+k\Delta}\big)\prod_{\ell=0}^{k-1} h_\ell\big(X_{s+\ell\Delta}\big),$$

with $g_0 := 0$ and, for all $k$ in $\{1,\dots,n\}$, $g_k := \Delta_n f_{t_k}$ and $h_{k-1} := 1$. For this reason, it follows from Theorem 4.9166 that

$$\underline{E}_{\mathscr{P}}\big(\langle f_\bullet\rangle_v \,\big|\, X_s = x\big) = f_{n,0}(x), \tag{6.66}$$

where $f_{n,0}$ is the gamble on $\mathscr{X}$ that is defined by the initial condition $f_{n,n} := g_n = \Delta_n f_{t_n} = \Delta_n f_r$ and, for all $k$ in $\{0,\dots,n-1\}$, by the recursive relation

$$f_{n,k} := g_k + e^{\Delta_n \underline{Q}} f_{n,k+1}.$$

Furthermore, it follows from Lemma 6.43311 that

$$\big|f_{n,0}(x) - \tilde{f}_{n,0}(x)\big| \leq \big\|f_{n,0} - \tilde{f}_{n,0}\big\|_{\mathrm{op}} \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}}^2 \frac{(r-s)^2}{n^2}\sum_{k=1}^{n}\|\tilde{f}_{n,k}\|_{\mathrm{c}}. \tag{6.67}$$

Let $\gamma^+ := \sup\{\max f_t\colon t \in [s,r]\}$ and $\gamma_- := \inf\{\min f_t\colon t \in [s,r]\}$. We now claim that for all $k$ in $\{1,\dots,n\}$, $\min \tilde{f}_{n,k} \geq (n-k+1)\Delta_n\gamma_-$ and $\max \tilde{f}_{n,k} \leq (n-k+1)\Delta_n\gamma^+$, and therefore $\|\tilde{f}_{n,k}\|_{\mathrm{c}} \leq \frac{(n-k+1)\Delta_n\gamma}{2}$. Our proof will be one by induction. For the base case $k = n$, this is obvious because $\tilde{f}_{n,n} = \Delta_n f_r$. For the inductive step, we fix some $k$ in $\{1,\dots,n-1\}$ and assume that $\min \tilde{f}_{n,k+1} \geq (n-k)\Delta_n\gamma^-$ and $\max \tilde{f}_{n,k+1} \leq (n-k)\Delta_n\gamma^+$. Because $\Delta_n\|\underline{Q}\|_{\mathrm{op}} \leq 2$, $(I + \Delta_n\underline{Q})$ is a lower transition operator due to Lemma 3.72112. Hence, it follows from the induction hypothesis and (LT4)108 that

$$(n-k)\Delta_n\gamma_- \leq \min \tilde{f}_{n,k+1} \leq (I + \Delta_n\underline{Q})\tilde{f}_{n,k+1} \leq \max \tilde{f}_{n,k+1} \leq (n-k)\Delta_n\gamma_+.$$

For this reason, and because $\tilde{f}_{n,k} = \Delta_n f_{t_k} + (I + \Delta_n\underline{Q})\tilde{f}_{n,k+1}$ by definition, we see that $\min \tilde{f}_{n,k} \geq (n-k+1)\Delta_n\gamma_-$ and $\max \tilde{f}_{n,k} \leq (n-k+1)\Delta_n\gamma^+$, as required.

Because $\|\tilde{f}_{n,k}\|_{\mathrm{c}} \leq \frac{(n-k+1)\Delta_n\gamma}{2}$ for all $k$ in $\{1,\dots,n\}$, it follows from Eq. (6.67) that

$$\begin{aligned}\big|f_{n,0}(x) - \tilde{f}_{n,0}(x)\big| &\leq \frac{1}{4}\|\underline{Q}\|_{\mathrm{op}}^2\gamma\frac{(r-s)^3}{n^3}\sum_{k=1}^{n}(n-k+1)\\ &= \frac{1}{4}\|\underline{Q}\|_{\mathrm{op}}^2\gamma\frac{(r-s)^3}{n^3}\frac{n(n+1)}{2}\\ &= \frac{1}{8}\|\underline{Q}\|_{\mathrm{op}}^2\gamma\frac{(r-s)^3}{n}\frac{n+1}{n},\end{aligned} \tag{6.68}$$

where for the inequality we also used that $\Delta_n = (r-s)/n$. Finally, it follows from Eqs. (6.65)↶, (6.66) and (6.68) and the triangle inequality that

$$\begin{aligned}&\left|\underline{E}^{\mathrm{D}}_{\mathscr{P}}\left(\int_s^r f(X_t)\,\mathrm{d}t \,\middle|\, X_s = x\right) - \tilde{f}_{n,0}(x)\right|\\ &\qquad\leq \frac{(r-s)^2}{n}\left(\kappa + \frac{\gamma}{2}\|\underline{Q}\|_{\mathrm{op}}\right) + \frac{1}{8}\frac{(r-s)^3}{n}\gamma\frac{n+1}{n}\|\underline{Q}\|_{\mathrm{op}}^2\\ &\qquad= \frac{(r-s)^2}{n}\left(\kappa + \frac{\gamma}{2}\|\underline{Q}\|_{\mathrm{op}} + \frac{1}{8}(r-s)\frac{n+1}{n}\|\underline{Q}\|_{\mathrm{op}}^2\right).\end{aligned} \tag{6.69}$$

Because Eq. (6.69)↶ holds for all $n$ in $\mathbb{N}$ such that $(r-s)\|\underline{Q}\|_{\mathrm{op}} \leq 2n$ and because the right-hand side of the inequality vanishes as $n$ recedes to $+\infty$, we have proven the limit statement for $\underline{E}^{\mathrm{D}}_{\mathscr{P}}$. As in the proof of Theorem 6.48$_{316}$, the statement for $\overline{E}^{\mathrm{D}}_{\mathscr{P}}$ essentially follows from conjugacy. □

# Lumping 7

State space explosion, or the polynomial or even exponential dependency of the cardinality of the finite state space on the parameters that govern the system's dimensions, is a frequently encountered inconvenience when constructing mathematical models of systems. In the setting of jump processes, this rapidly growing number of states has as a consequence that using the model to perform inferences about large-scale systems becomes computationally intractable. In many cases, however, we can formalise the inferences we would like to make in a higher-level state description, allowing for a reduced state space with considerably fewer states. The catch is that the low-level description and its corresponding larger state space are necessary in order to precisely or exactly characterise the system's dynamics. This is where imprecise jump processes come in handy, because they provide a way to deal with partially specified dynamics.

The procedure of going from a low-level to a higher-level state description is called *lumping*. It was – to the best of our knowledge – first proposed by Burke et al. (1958). These authors exploit the relation between the original state space, corresponding to the low-level state description, and the lumped state space, corresponding to the higher-level state description, to obtain a *lumped jump process* from the original jump process. Rather unfortunately, this lumped jump process need not be a (homogeneous) Markovian one. In fact, Burke et al. (1958) provide a very stringent (necessary and) sufficient condition on (the rate operator of) the original homogeneous Markovian jump process under which the lumped jump process is homogeneous and Markovian too. That this condition is not trivially satisfied is quite unfortunate, because if the lumped jump process is not a homogeneous Markovian jump process, then using it to make inferences about the system is not feasible in practice.

Subsequent research on the lumping of homogeneous Markovian jump processes centred around two separate topics. On the one hand, several authors generalised the aforementioned (necessary and) sufficient conditions to other settings (Hachigian, 1963; Ball et al., 1993; Rubino et al., 1993; Hartfiel, 1994) or devised algorithms to determine the smallest lumped state space for which the lumped process is still a homogeneous Markovian jump process (Derisavi et al., 2003; Valmari et al., 2010). Franceschinis et al. (1994) and Buchholz (2005), on the other hand, proposed methods based on the lumping procedure to bound limit expectations with respect to the original homogeneous Markovian jump process. Furthermore, the lumping procedure has also been used by Katoen et al. (2012) in the context of model checking and bisimulation.

In this chapter, we more or less follow the historical evolution of the previously mentioned research: we start with a theoretical study of the lumped jump process and then propose methods based on the lumping procedure to bound expectations with respect to homogeneous Markovian (imprecise) jump processes. In Section 7.1, we propose a formal procedure to lump a single jump process – previous studies, for instance that of Burke et al. (1958), always ignore some technicalities in the construction of the lumped jump process. In Section $7.2_{352}$, we look at the particular case of lumping a jump processes $P$ that is consistent with a set $\mathscr{M}$ of initial probability mass functions and a set $\mathscr{Q}$ of rate operators. Specifically, we argue that 'the' lumped jump process is consistent with a set of probability mass functions on the lumped state space induced by $\mathscr{M}$ and a set of rate operators induced by $\mathscr{Q}$. More importantly, we show that we can use the homogeneous and Markovian imprecise jump process that is characterised by these induced sets to obtain bounds on expectations with respect to any original jump processes $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ and any imprecise jump process $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$. In Section $7.3_{373}$, we propose two lumping-based methods to approximate the limit expectation of ergodic homogeneous Markovian jump processes.

The present chapter is largely based on (Erreygers & De Bock, 2019a), although many of the results in Section $7.2_{352}$ are new.

## 7.1 Lumping a jump process

We have seen in Chapters $4_{157}$ and $5_{215}$ that computing expectations for homogeneous Markovian (imprecise) jump processes requires numerical methods. Quite obviously, these numerical methods are computationally intractable when the state space becomes too large. For example, for the model in Chapter $8_{403}$ further, the cardinality of the state space grows polynomially with the parameters that describe the system's dimensions, and already for moderate dimensions the cardinality of the state space leads to tractability issues. Throughout this chapter, we will use the following more basic queueing system as an example, following Franceschinis et al. (1994) and Buchholz (2005).

**Queuing Network Example 7.1.** Franceschinis et al. (1994) consider the

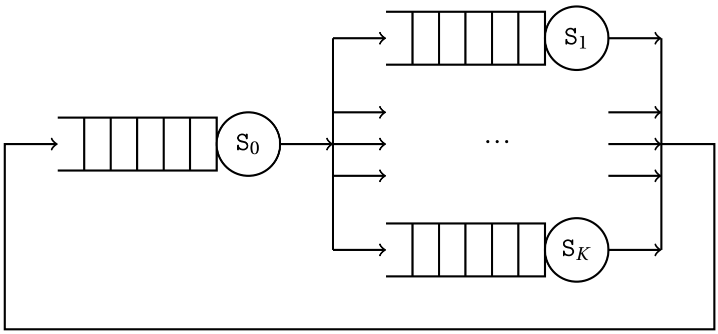



**Figure 7.1** The closed queueing network

closed queueing network that is depicted in Fig. 7.1: it consist of a single server $S_0$ in series with $K$ parallel servers $S_1, \dots, S_K$, with $K$ a natural number. The network is populated by $N$ customers, with $N$ a natural number, and these customers cycle between the server $S_0$ and one of the parallel servers $S_1, \dots, S_K$. Each of the servers can only service one customer at a time, so customers are not always immediately serviced by a server.

One obvious way to describe this closed queueing network is to use states of the form $(n_0, n_1, \dots, n_K)$, where $n_k$ is the number of customers in the server $S_k$. This state description yields the state space

$$\mathscr{X} := \left\{ (n_0, n_1, \dots, n_K) \in \{0, \dots, N\}^{K+1} \colon \sum_{k=0}^{K} n_k = N \right\}.$$

Using Feller's (1968, Chapter II, Section 5) 'stars and bars' method, we see that this state space $\mathscr{X}$ contains

$$|\mathscr{X}| = \binom{K+N}{K} = \binom{K+N}{N} = \frac{(K+N)!}{K!N!}$$

states. Due to the standard bounds on the binomial coefficient,

$$\frac{(N+1)^K}{K!} \leq |\mathscr{X}| \leq \frac{(K+N)^K}{K!} \quad \text{and} \quad \frac{(K+1)^N}{N!} \leq |\mathscr{X}| \leq \frac{(K+N)^N}{N!}.$$

In Fig. 7.2, we plot the cardinality of $\mathscr{X}$ for different values of $N$ and $K$. This plot verifies that the number of states grows polynomially with $N$ and $K$, so we have state space explosion. Even in the moderate case of $K = 8$ parallel servers and $N = 10$ customers, the state space $\mathscr{X}$ contains a whopping 43 758 states.

Fortunately, the state space $\mathscr{X}$ is often unnecessarily detailed, at least from the point of view of the inferences that one would like to make. Indeed,

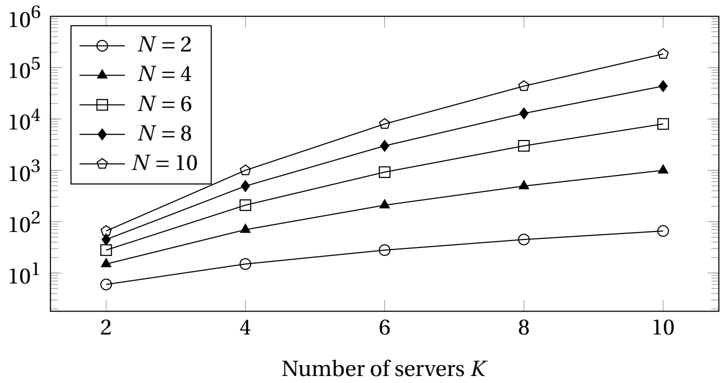



**Figure 7.2** Number of states in the state space $\mathscr{X}$ as a function of the number of parallel servers $K$ and the number of customers $N$. Note that the vertical scale is logarithmic.

many interesting inferences can usually still be unambiguously defined on a less detailed state space $\hat{\mathscr{X}}$ that corresponds to a higher-level description of the system.

**Queuing Network Example 7.2.** As a higher-level state description, Franceschinis et al. (1994) propose to use $(n_0, m_0, \ldots, m_N)$, where $n_0$ is again the number of customers in the single server $\mathtt{S}_0$ and where $m_\ell$ is the number of parallel servers that have $\ell$ customers, with $\ell$ a non-negative integer that is at most $N$. Thus, the less detailed state space is

$$\hat{\mathscr{X}} := \Big\{(n_0, m_0, \ldots, m_N) \in \{0, \ldots, N\} \times \{0, \ldots, K\}^{N+1} : \\ n_0 + \sum_{\ell=1}^{N} \ell m_\ell = N, \sum_{\ell=0}^{N} m_\ell = K\Big\}.$$

Comparing Fig. 7.3↷ to Fig. 7.2, it is clear that the lumped state space is significantly smaller than the original state space $\mathscr{X}$. ⅏

Of course, the rationale for using the detailed state space $\mathscr{X}$ instead of the less detailed state space $\hat{\mathscr{X}}$ in the first place is that this allows one to model the system. For example, in our running example, but also in Chapter 8403 further on, the detailed state space $\mathscr{X}$ and ‘standard’ queueing theory assumptions ensure that the system can be modelled by a homogeneous Markovian jump process.

**Queuing Network Example 7.3.** The time that a customer is being serviced by a server is called the *service time*, and we are uncertain about the length of these service times. As is typically done in (basic) queueing theory, Frances-

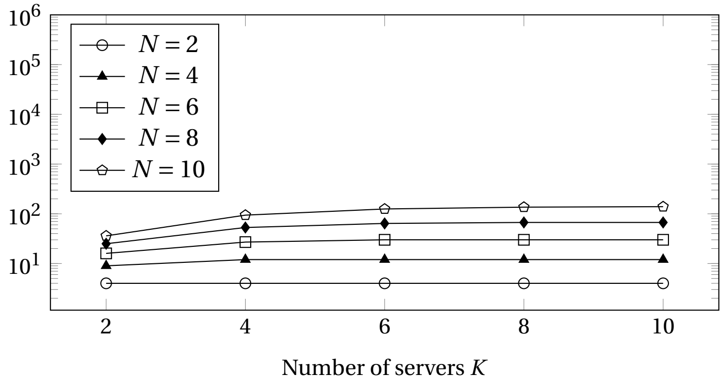



**Figure 7.3** Number of states in the state space $\hat{\mathscr{X}}$ as a function of the number of parallel servers $K$ and the number of customers $N$. The vertical scale is the same as the one in Fig. 7.2.

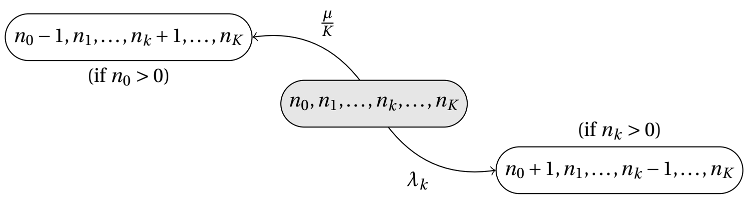



**Figure 7.4** State transition diagram for the rate operator $Q$ defined by Eq. (7.1) in Queuing Network Example 7.3

chinis et al. (1994) assume that these service times are all 'exponentially distributed' and 'independent'; the service time for $\mathtt{S}_0$ has rate $\mu$, and the service time for the $k$-th parallel server $\mathtt{S}_k$ has rate $\lambda_k$, with $k$ in $\{1, \ldots, K\}$. Furthermore, Franceschinis et al. (1994) assume that after being serviced by $\mathtt{S}_0$, the customer is assigned to one of the parallel servers 'at random' – so with uniform probability. Franceschinis et al. (1994) are only interested in limit expectations, so they discuss the dynamics but not the initial condition; here, we make the assumption that our uncertainty about the initial state of the system is accurately modelled by some initial probability mass function $p_0$ on the detailed state space $\mathscr{X}$ as defined in Queuing Network Example 7.1338. Under these assumptions, the queueing network can be modelled by a homogeneous Markovian jump process $P_{p_0,Q}$ on the detailed state space $\mathscr{X}$. The rate operator $Q$ is completely defined by the rate parameters $\mu$ and $\lambda_1$, …, $\lambda_K$. For a given state $(n_0, \ldots, n_K)$, the non-zero off-diagonal components of the matrix representation of $Q$ are schematically depicted in Fig. 7.4. For the formal definition of $Q$, we fix some $f$ in $\mathbb{G}(\mathscr{X})$ and $x = (n_0, \ldots, n_K)$ in $\mathscr{X}$. For all $k$ in $\{1, \ldots, K\}$, we let $x_k^+ := (n_0 - 1, n_1, \ldots, n_k + 1, \ldots, n_K)$ if $n_0 > 0$ and $x_k^- := (n_0 + 1, n_1, \ldots, n_k - 1, \ldots, n_K)$ if $n_k > 0$. Then

$$[Qf](x) = \sum_{k=1}^{K} \frac{\mu}{K}\big(f(x_k^+) - f(x)\big) + \sum_{k \in \mathscr{K}_x} \lambda_k\big(f(x_k^-) - f(x)\big), \qquad (7.1)$$

where the first summation is only added if $n_0 > 0$ and where $\mathscr{K}_x$ is the set of indices $k$ in $\{1, \ldots, K\}$ such that $n_k > 0$. 𝕿

### 7.1.1 The lumped state space and the lumping map

Usually, the less detailed state space $\hat{\mathscr{X}}$ is obtained by *lumping* – sometimes also called grouping or aggregating, see (Burke et al., 1958; Ball et al., 1993) – states in $\mathscr{X}$, so $1 \leq |\hat{\mathscr{X}}| \leq |\mathscr{X}|$. In this case, we call $\hat{\mathscr{X}}$ the *lumped state space*, and we formalise this lumping through a *lumping map* $\Lambda\colon \mathscr{X} \to \hat{\mathscr{X}}$ that maps every state $x$ in $\mathscr{X}$ to a *lumped state* $\Lambda(x)$ in $\hat{\mathscr{X}}$.

**Queuing Network Example 7.4.** Consider a state $(n_0, \ldots, n_K)$ in the detailed state space $\mathscr{X}$ as defined in Queuing Network Example 7.1338. Then for all $k$ in $\{0, \ldots, N\}$, the number of parallel servers with $k$ customers is

$$m_k := \big|\{\ell \in \{1, \ldots, K\}\colon n_\ell = k\}\big|.$$

Hence, the state $(n_0, \ldots, n_K)$ corresponds to the lumped state $(n_0, m_0, \ldots, m_N)$ in the lumped state space $\hat{\mathscr{X}}$ as defined in Queuing Network Example 7.2340. Note that permuting the number of customers in the parallel servers yields a state in $\mathscr{X}$, and this permuted state corresponds to the same lumped state $(n_0, m_0, \ldots, m_N)$. In other words, different states in $\mathscr{X}$ can correspond to the same lumped state in $\hat{\mathscr{X}}$. 𝕿

We assume that the lumping map $\Lambda$ is surjective or onto, meaning that for every lumped state $\hat{x}$ in $\hat{\mathscr{X}}$ there is at least one state $x$ in $\mathscr{X}$ such that $\Lambda(x) = \hat{x}$. We will frequently need the obvious component-wise extension of the lumping map $\Lambda$ to tuples of states, so to $\mathscr{X}_u$ for some $u$ in $\mathscr{U}$; because the domain is always clear from the context, we will also denote this extension by $\Lambda$. Furthermore, the *inverse lumping map* $\Lambda^{-1}$ is the (set-valued) inverse of $\Lambda$: the empty tuple $\hat{x}_{()} = \diamond$ is mapped to $\Lambda^{-1}(\hat{x}_{()}) := x_{()} = \diamond$ and, for all $u$ in $\mathscr{U}_{\neq()}$ and $\hat{x}_u$ in $\hat{\mathscr{X}}_u$,

$$\Lambda^{-1}(\hat{x}_u) := \big\{x_u \in \mathscr{X}_u\colon \Lambda(x_u) = \hat{x}_u\big\} = \big\{x_u \in \mathscr{X}_u\colon (\forall t \in u)\ \Lambda(x_t) = \hat{x}_t\big\}. \qquad (7.2)$$

In order to lighten our notation taking sums over $\Lambda^{-1}(\hat{x}_u)$, we will frequently shorten '$x_u \in \Lambda^{-1}(\hat{x}_u)$' to '$x_u \in \hat{x}_u$'. Finally, we adhere to the standard notational convention that, for all $u$ in $\mathscr{U}$, $B \subseteq \mathscr{X}_u$ and $\hat{B} \subseteq \hat{\mathscr{X}}_u$,

$$\Lambda(B) := \big\{\Lambda(x_u)\colon x_u \in B\big\} \quad \text{and} \quad \Lambda^{-1}(\hat{B}) := \big\{x_u \in \mathscr{X}_u\colon \Lambda(x_u) \in \hat{B}\big\}; \qquad (7.3)$$

note that $\hat{B} = \Lambda\big(\Lambda^{-1}(\hat{B})\big)$ but $B \subseteq \Lambda^{-1}\big(\Lambda(B)\big)$, at least in general.

It is customary to let the lumped state space $\hat{\mathscr{X}}$ correspond to a natural higher-level description of the state of the system. This is done by Franceschinis et al. (1994), Buchholz (2005), Ganguly et al. (2014), and Kim, Yan, et al. (2015); following these authors, we do so in the running example in the present chapter and in Chapter $8_{403}$ further on. That said, we would like to emphasise that as far as our theoretical results are concerned, it does not matter how the states are lumped.

### 7.1.2 A lumped jump process

The idea behind lumping is to describe the jump process $P$ in terms of the lumped state space $\hat{\mathscr{X}}$ instead of the original state space $\mathscr{X}$. To that end, we will 'lump' the original jump process $P$ to obtain a 'lumped jump process' $\hat{P}$ with state space $\hat{\mathscr{X}}$.

So what does it mean to 'lump' a jump process $P$ with state space $\mathscr{X}$? Formally, we lump the state $X_t$ of the system at all time points $t$ in $\mathbb{R}_{\geq 0}$: every path $\omega$ in the set $\Omega$ of càdlàg paths is lumped to the path

$$\Lambda \circ \omega\colon \mathbb{R}_{\geq 0} \to \hat{\mathscr{X}}\colon t \mapsto [\Lambda \circ \omega](t) = \Lambda\big(\omega(t)\big).$$

Hence, the possibility space $\Omega$ becomes the lumped possibility space

$$\hat{\Omega} := \big\{\Lambda \circ \omega\colon \omega \in \Omega\big\}; \tag{7.4}$$

this set of lumped paths $\hat{\Omega}$ is equal to the set $\Omega_{\hat{\mathscr{X}}}$ of càdlàg paths for the lumped state space $\hat{\mathscr{X}}$.

**Lemma 7.5.** *For any path $\omega$ in $\Omega$, the corresponding lumped path $\hat{\omega} := \Lambda \circ \omega$ is càdlàg. Furthermore, $\hat{\Omega}$ as defined in Eq. (7.4) is equal to the set $\Omega_{\hat{\mathscr{X}}}$ of all càdlàg paths with state space $\hat{\mathscr{X}}$.*

*Proof.* The first part of the statement is almost trivial. By Definition $3.4_{57}$, $\hat{\omega}$ is càdlàg if for all $t$ in $\mathbb{R}_{\geq 0}$, $\hat{\omega}$ is continuous from the right and, if $t > 0$, has a left-sided limit. Fix any $t$ in $\mathbb{R}_{\geq 0}$; then because $\omega$ is càdlàg by assumption, it follows from Eq. $(3.1)_{58}$ in Definition $3.4_{57}$ that $\lim_{r \searrow t} \omega(t)$ exists and is equal to $\omega(t)$. Consequently, $\lim_{r \searrow t} \hat{\omega}(t) = \lim_{r \searrow t} [\Lambda \circ \omega](t)$ exists and is equal to $[\Lambda \circ \omega](t) = \hat{\omega}(t)$, so $\hat{\omega}$ is continuous from the right. If $t > 0$, a similar argument but with Eq. $(3.2)_{58}$ in the role of Eq. $(3.1)_{58}$ shows that $\hat{\omega}$ has a left-sided limit at $t$. Because $t$ is an arbitrary time point in $\mathbb{R}_{\geq 0}$, this proves that $\hat{\omega}$ is càdlàg.

The second part of the statement is almost trivial as well. To prove it, it remains to show that for any arbitrary path $\hat{\omega}$ in $\Omega_{\hat{\mathscr{X}}}$, there is some path $\omega$ in $\Omega = \Omega_{\mathscr{X}}$ such that $\Lambda \circ \omega = \hat{\omega}$. Fix any $\hat{\omega}$ in $\Omega_{\hat{\mathscr{X}}}$. For all $\hat{x}$ in $\hat{\mathscr{X}}$, we choose any $x_{\hat{x}}$ in $\Lambda^{-1}(\hat{x})$. Then the path

$$\omega^\star\colon \mathbb{R}_{\geq 0} \to \mathscr{X}\colon t \mapsto \omega^\star(t) := x_{\hat{\omega}(t)}.$$

is càdlàg because $\hat{\omega}$ is càdlàg – this follows from essentially the same arguments and considerations as before. Furthermore, it is clear that $\Lambda \circ \omega^\star = \hat{\omega}$ because, by construction, $[\Lambda\circ\omega^\star](t)=\Lambda\big(x_{\hat{\omega}(t)}\big)=\hat{\omega}(t)$ for all $t$ in $\mathbb{R}_{\geq 0}$. This proves the second part of the statement. □

We denote the projector variables corresponding to $\hat{\Omega}$ as defined in Eq. (3.3)₅₉ by $(\hat{X}_t)_{t\in\mathbb{R}_{\geq 0}}$. In the same manner, we extend the rest of the notation regarding events as introduced in Section 3.1.2₅₈. For all $v$ in $\mathscr{U}$, $\hat{x}_v$ in $\hat{\mathscr{X}}_v$ and $\hat{B}\subseteq\hat{\mathscr{X}}_v$, we let

$$\{\hat{X}_v=\hat{x}_v\}:=\{\hat{\omega}\in\hat{\Omega}\colon \hat{\omega}|_v=\hat{x}_v\}$$

and

$$\{\hat{X}_v\in\hat{B}\}:=\{\hat{\omega}\in\hat{\Omega}\colon \hat{\omega}|_v\in\hat{B}\}=\bigcup_{\hat{x}_v\in\hat{B}}\{\hat{X}_v=\hat{x}_v\}. \tag{7.5}$$

Then the set of lumped state histories is

$$\hat{\mathscr{H}}:=\big\{\{\hat{X}_u=\hat{x}_u\}\colon u\in\mathscr{U},\hat{x}_u\in\hat{\mathscr{X}}_u\big\}, \tag{7.6}$$

and for all $u$ in $\mathscr{U}$, the corresponding set of lumped cylinder events is

$$\hat{\mathscr{F}}_u:=\big\{\{\hat{X}_v\in\hat{B}\}\colon v\in\mathscr{U}_{\succcurlyeq u},\hat{B}\subseteq\hat{\mathscr{X}}_v\big\}. \tag{7.7}$$

Thus, a jump process with state space $\hat{\mathscr{X}}$ should have domain

$$\hat{\mathscr{D}}:=\mathscr{D}_{\hat{\mathscr{X}}}=\big\{(\hat{A}\,|\,\hat{X}_u=\hat{x}_u)\colon u\in\mathscr{U},\hat{x}_u\in\hat{\mathscr{X}}_u,\hat{A}\in\hat{\mathscr{F}}_u\big\}. \tag{7.8}$$

Recall from Eq. (7.4)⌒ that $\hat{\Omega}$ is derived from $\Omega$ by means of the lumping map $\Lambda$: for all $t$ in $\mathbb{R}_{\geq 0}$, $\hat{X}_t$ corresponds to $\Lambda\circ X_t$, so the event

$$\{\hat{X}_t=\hat{x}\}=\big\{\hat{\omega}\in\hat{\Omega}\colon \hat{\omega}(t)=\hat{x}\big\}$$

naturally corresponds to the event

$$\begin{aligned}\big\{\Lambda\circ X_t=\hat{x}\big\}&=\big\{\omega\in\Omega\colon [\Lambda\circ\omega](t)=\hat{x}\big\}=\big\{\omega\in\Omega\colon \Lambda\circ\omega\in\Lambda^{-1}(\hat{x})\big\}\\&=\big\{X_t\in\Lambda^{-1}(\hat{x})\big\}.\end{aligned}$$

More generally, any event $\hat{A}$ in $\mathcal{P}(\hat{\Omega})$ corresponds to the event

$$\Lambda_\Omega^{-1}(\hat{A}):=\big\{\omega\in\Omega\colon \Lambda\circ\omega\in\hat{A}\big\} \tag{7.9}$$

in $\mathcal{P}(\Omega)$; note that $\Lambda_\Omega^{-1}(\hat{A})$ is the inverse image of the set $\hat{A}$.[1] For all $\hat{A}$ in $\mathcal{P}(\hat{\Omega})$, it follows from Eqs. (7.4)⌒ and (7.9) that

$$\big\{\Lambda\circ\omega\colon \omega\in\Lambda_\Omega^{-1}(\hat{A})\big\}=\hat{A}. \tag{7.10}$$

Furthermore, $\Lambda_\Omega^{-1}$ maps cylinder events in $\hat{\mathscr{F}}_u$ to cylinder events in $\mathscr{F}_u$.

[1] It would arguably make more sense to denote this inverse image by $\Lambda_{\mathcal{P}(\Omega)}^{-1}(\hat{A})$ instead of by $\Lambda_\Omega^{-1}(\hat{A})$, but we use the latter because it is shorter.

**Lemma 7.6.** *Consider a sequence of time points $u$ in $\mathscr{U}$. Then for all $v$ in $\mathscr{U}_{\succcurlyeq u}$ and $\hat{B} \subseteq \hat{\mathscr{X}}_v$,*

$$\Lambda_\Omega^{-1}\big(\{\hat{X}_v \in \hat{B}\}\big) = \big\{X_v \in \Lambda^{-1}(\hat{B})\big\}. \tag{7.11}$$

*Consequently, for all $\hat{A}$ in $\hat{\mathscr{F}}_u$, $\Lambda_\Omega^{-1}(\hat{A})$ belongs to $\mathscr{F}_u$.*

*Proof.* To prove Eq. (7.11), we recall from Eq. (7.5)↶ that

$$\{\hat{X}_v \in \hat{B}\} = \{\hat{\omega} \in \hat{\Omega} : \hat{\omega}|_v \in \hat{B}\}.$$

Hence, it follows from Eq. (7.9)↶ that

$$\begin{aligned}\Lambda_\Omega^{-1}\big(\{\hat{X}_v \in \hat{B}\}\big) &= \big\{\omega \in \Omega : \Lambda \circ \omega \in \{\hat{\omega} \in \hat{\Omega} : \hat{\omega}|_v \in \hat{B}\}\big\} = \big\{\omega \in \Omega : (\Lambda \circ \omega)|_v \in \hat{B}\big\} \\ &= \big\{\omega \in \Omega : \Lambda(\omega|_v) \in \hat{B}\big\}.\end{aligned}$$

By Eq. (7.3)342, $\Lambda^{-1}(\hat{B})$ contains all $x_v$ in $\mathscr{X}_v$ such that $\Lambda(x_v)$ belongs to $\hat{B}$. Consequently,

$$\Lambda_\Omega^{-1}\big(\{\hat{X}_v \in \hat{B}\}\big) = \big\{X_v \in \Lambda^{-1}(\hat{B})\big\},$$

as required.

The second part of the statement follows immediately from the first part and Eqs. (3.16)63 and (7.7)↶. ☐

The correspondence between lumped cylinder events in $\hat{\mathscr{F}}_u$ and cylinder events in $\mathscr{F}_u$ allows us to derive from a jump process $P$ with state space $\mathscr{X}$ a lumped jump process $\hat{P}$ with state space $\hat{\mathscr{X}}$. First, let us consider the unconditional probabilities. For all $\{\hat{X}_v \in \hat{B}\}$ in $\hat{\mathscr{F}}_{()}$, the corresponding event

$$\Lambda_\Omega^{-1}\big(\hat{X}_v \in \hat{B}\big) = \big\{X_v \in \Lambda^{-1}(\hat{B})\big\}$$

belongs to $\mathscr{F}_{()}$ due to Lemma 7.6, so it makes sense to let

$$\hat{P}(\hat{X}_v \in \hat{B}) = P\big(\Lambda_\Omega^{-1}\big(\hat{X}_v \in \hat{B}\big)\big) = P\big(X_v \in \Lambda^{-1}(\hat{B})\big). \tag{7.12}$$

The conditional probabilities then follow from the unconditional ones and Bayes's rule: for all $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$ such that $u \neq ()$ and $\hat{P}(\hat{X}_u = \hat{x}_u) > 0$ and all $\hat{A}$ in $\mathscr{F}_u$,

$$\hat{P}(\hat{A} \,|\, \hat{X}_u = \hat{x}_u) = \frac{\hat{P}\big(\{\hat{X}_u = \hat{x}_u\} \cap \hat{A}\big)}{\hat{P}(\hat{X}_u = \hat{x}_u)}.$$

There is one issue with this natural definition: it may not define $\hat{P}$ on its entire domain $\hat{\mathscr{D}}$, because we cannot invoke Bayes's rule for lumped state histories $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$ with probability 0. One way to circumvent this, is to start from a homogeneous Markovian jump process $P_{p_0,Q}$ with a positive initial probability mass function $p_0 > 0$ and an irreducible rate operator $Q$ – see Section 7.3.1374 further on for a definition. In that case, any lumped state history $\{\hat{X}_u = \hat{x}_u\}$ has non-zero probability, so the lumped jump process is uniquely defined; we refer to (Erreygers & De Bock, 2018b, Appendix D.1) for a more detailed explanation. Here, we will deal with the issue of conditioning on events with probability zero through coherence.

Let us try to define the conditional probabilities for the lumped jump process $\hat{P}$ directly instead of through Bayes's rule. In Eq. (7.12)↶, we implicitly used the notational conventions $\hat{P}(\bullet) = \hat{P}(\bullet \,|\, \hat{\Omega})$ and $P(\bullet) = P(\bullet \,|\, \Omega)$. Because furthermore $\{\hat{X}_{()} = \hat{x}_{()}\} = \hat{\Omega}$ and

$$\Lambda_\Omega^{-1}\big(\{\hat{X}_{()} = \hat{x}_{()}\}\big) = \Lambda_\Omega^{-1}(\hat{\Omega}) = \Omega = \{X_{()} = x_{()}\} = \{X_{()} \in \Lambda^{-1}(\hat{x}_{()})\},$$

we see that Eq. (7.12)↶ actually states that for all $(\hat{A} \,|\, \hat{X}_{()} = \hat{x}_{()})$ in $\hat{\mathscr{D}}$,

$$\hat{P}(\hat{A} \,|\, \hat{X}_{()} = \hat{x}_{()}) = P\big(\Lambda_\Omega^{-1}(\hat{A}) \,\big|\, X_{()} \in \Lambda^{-1}(\hat{x}_{()})\big). \tag{7.13}$$

At first sight, it seems that we can generalise this to unconditional probabilities, but this is *not* the case. Fix some $(\hat{A} \,|\, \hat{X}_u = \hat{x}_u)$ in $\hat{\mathscr{D}}$. Then by Eq. (7.11)↶ in Lemma 7.6↶, the corresponding conditioning event is

$$\Lambda_\Omega^{-1}\big(\{\hat{X}_u = \hat{x}_u\}\big) = \big\{X_u \in \Lambda^{-1}(\hat{x}_u)\big\}.$$

Note that $\Lambda^{-1}(\hat{x}_u)$ need not be a singleton, and this is what causes problems: if $\Lambda^{-1}(\hat{x}_u)$ is not a singleton, then the conditioning event $\{X_u \in \Lambda^{-1}(\hat{x}_u)\}$ does not belong to $\mathscr{H}$, so

$$\big(\Lambda_\Omega^{-1}(\hat{A}) \,\big|\, X_u \in \Lambda^{-1}(\hat{x}_u)\big)$$

does not belong to the domain $\mathscr{D}$ of $P$.

Fortunately, we can always extend $P$ to a larger domain because, by definition, $P$ is a coherent conditional probability on $\mathscr{D} \subseteq \mathcal{P}(\Omega) \times \mathcal{P}(\Omega)_{\supset\emptyset}$. By Theorem 2.54₄₅, $P$ can therefore be extended to a coherent conditional probability $P^\star$ on any domain $\mathscr{D}^\star$ such that $\mathscr{D} \subseteq \mathscr{D}^\star \subseteq \mathcal{P}(\Omega) \times \mathcal{P}(\Omega)_{\supset\emptyset}$. In the setting of lumping, we get by with the following domain:

$$\mathscr{D}^\star := \big\{(A \,|\, X_u \in B) \colon u \in \mathscr{U}, A \in \mathscr{F}_u, \emptyset \neq B \subseteq \mathscr{X}_u\big\}, \tag{7.14}$$

which is a structure of fields – see Definition 2.57₄₆. Note that for any $u$ in $\mathscr{U}$ and any non-empty subset $B$ of $\mathscr{X}_u$, $\{X_u \in B\} = \bigcup_{x_u \in B}\{X_u = x_u\}$, so this event is non-empty by Lemma 3.5₅₈. Because furthermore $\{X_u = x_u\} = \{X_u \in \{x_u\}\}$ for all $\{X_u = x_u\}$ in $\mathscr{H}$, we see that $\mathscr{D} \subseteq \mathscr{D}^\star \subseteq \mathcal{P}(\Omega) \times \mathcal{P}(\Omega)_{\supset\emptyset}$. The following result establishes that the domain $\mathscr{D}^\star$ is sufficiently large for our purposes.

**Lemma 7.7.** *For all $(\hat{A} \,|\, \hat{X}_u = \hat{x}_u)$ in $\hat{\mathscr{D}}$,*

$$\big(\Lambda_\Omega^{-1}(\hat{A}) \,\big|\, \Lambda_\Omega^{-1}\big(\{\hat{X}_u = \hat{x}_u\}\big)\big) = \big(\Lambda_\Omega^{-1}(\hat{A}) \,\big|\, X_u \in \Lambda^{-1}(\hat{x}_u)\big)$$

*belongs to $\mathscr{D}^\star$.*

*Proof.* By definition of $\hat{\mathscr{D}}$ – see Eq. (7.8)₃₄₄ – the event $\hat{A}$ belongs to $\hat{\mathscr{F}}_u$. From Lemma 7.6↶, we know that the event $\Lambda_\Omega^{-1}(\hat{A})$ belongs to $\mathscr{F}_u$ and that

$$\Lambda_\Omega^{-1}\big(\{\hat{X}_u = \hat{x}_u\}\big) = \big\{X_u \in \Lambda^{-1}(\hat{x}_u)\big\}.$$

Because $\big(\Lambda_\Omega^{-1}(\hat{A}) \,\big|\, X_u \in \Lambda^{-1}(\hat{x}_u)\big)$ belongs to $\mathscr{D}^\star$ by definition, this proves the statement. □

In order to lighten our notation, we will henceforth write $\{X_u \in \hat{x}_u\}$ instead of $\{X_u \in \Lambda^{-1}(\hat{x}_u)\}$.

Consider a jump process $P$ with state space $\mathscr{X}$ and let $P^\star$ be a coherent extension of $P$ to $\mathscr{D}^\star$. Then the corresponding *lumped jump process* $\hat{P}$ is the real-valued map on $\hat{\mathscr{D}}$ defined for all $(\hat{A} \,|\, \hat{X}_u = \hat{x}_u)$ in $\hat{\mathscr{D}}$ by

$$\hat{P}(\hat{A} \,|\, \hat{X}_u = \hat{x}_u) := P^\star\big(\Lambda_\Omega^{-1}(\hat{A}) \,\big|\, X_u \in \hat{x}_u\big). \tag{7.15}$$

In contrast to our previous 'natural' definition of $\hat{P}$, this *is* a proper definition due to Lemma 7.7. Note that the lumped jump process is not necessarily unique because the coherent extension $P^\star$ of the jump process $P$ to $\mathscr{D}^\star$ need not be unique; any such coherent extension does yield a jump process though.

**Theorem 7.8.** *Consider a jump process $P$ with state space $\mathscr{X}$ and any coherent extension $P^\star$ of $P$ to $\mathscr{D}^\star$. Then the corresponding real-valued map $\hat{P}$ on $\hat{\mathscr{D}}$, as defined by Eq.* (7.15)*, is a jump process with state space $\hat{\mathscr{X}}$.*

The following intermediary result comes in handy in the proof of Theorem 7.8, and also in that of Lemma 7.13350 further on.

**Lemma 7.9.** *For all $\hat{A}$ in $\mathcal{P}(\hat{\Omega})$, $\mathbb{I}_{\hat{A}} \circ \Lambda = \mathbb{I}_{\Lambda^{-1}(\hat{A})}$, in the sense that for all $\omega$ in $\Omega$,*

$$\mathbb{I}_{\hat{A}}(\Lambda \circ \omega) = \mathbb{I}_{\Lambda_\Omega^{-1}(\hat{A})}(\omega).$$

*Proof.* It follows immediately from Eq. (7.9)344 that

$$\mathbb{I}_{\hat{A}}(\Lambda \circ \omega) = \begin{cases} 1 & \text{if } \Lambda \circ \omega \in \hat{A} \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 1 & \text{if } \omega \in \Lambda_\Omega^{-1}(\hat{A}) \\ 0 & \text{otherwise} \end{cases} = \mathbb{I}_{\Lambda_\Omega^{-1}(\hat{A})}(\omega). \qquad \square$$

*Proof of Theorem 7.8.* By Definition 3.1265, we need to show that the real-valued map $\hat{P}$ on $\hat{\mathscr{D}} = \mathscr{D}_{\hat{\mathscr{X}}}$ is a coherent conditional probability. To verify the condition in Definition 2.5144, we fix some $n$ in $\mathbb{N}$, $(\hat{A}_1, \hat{C}_1), \dots, (\hat{A}_n, \hat{C}_n)$ in $\hat{\mathscr{D}}$ and $\mu_1, \dots, \mu_n$ in $\mathbb{R}$. We let

$$\mathcal{S} := \left\{ \sum_{k=1}^n \mu_k \mathbb{I}_{\hat{C}_k}(\hat{\omega}) \Big( \hat{P}(\hat{A}_k \,|\, \hat{C}_k) - \mathbb{I}_{\hat{A}_k}(\hat{\omega}) \Big) \colon \hat{\omega} \in \bigcup_{k=1}^n \hat{C}_k \right\}, \tag{7.16}$$

and set out to show that $\max \mathcal{S} \geq 0$.

By Eq. (7.15), the definition of $\hat{P}$,

$$\mathcal{S} = \left\{ \sum_{k=1}^n \mu_k \mathbb{I}_{\hat{C}_k}(\hat{\omega}) \Big( P^\star\big(\Lambda_\Omega^{-1}(\hat{A}_k) \,|\, \Lambda_\Omega^{-1}(\hat{C}_k)\big) - \mathbb{I}_{\hat{A}_k}(\hat{\omega}) \Big) \colon \hat{\omega} \in \bigcup_{k=1}^n \hat{C}_k \right\}.$$

Recall from Eq. (7.10)344 that for all $\hat{B}$ in $\mathcal{P}(\hat{\Omega})$, $\hat{B} = \big\{\Lambda \circ \omega \colon \omega \in \Lambda_\Omega^{-1}(\hat{B})\big\}$. Hence,

$$\mathcal{S} = \left\{ \sum_{k=1}^n \mu_k \mathbb{I}_{\hat{C}_k}(\Lambda \circ \omega) \Big( P^\star\big(\Lambda_\Omega^{-1}(\hat{A}_k) \,|\, \Lambda_\Omega^{-1}(\hat{C}_k)\big) - \mathbb{I}_{\hat{A}_k}(\Lambda \circ \omega) \Big) \colon \omega \in \bigcup_{k=1}^n \Lambda_\Omega^{-1}\big(\hat{C}_k\big) \right\}.$$

Next, we recall from Lemma 7.9↶ that $\mathbb{I}_{\hat{B}}(\Lambda \circ \omega) = \mathbb{I}_{\Lambda_\Omega^{-1}(\hat{B})}(\omega)$ for all $\hat{B}$ in $\mathcal{P}(\hat{\Omega})$ and $\omega$ in $\Omega$, so

$$\mathcal{S} = \left\{ \sum_{k=1}^{n} \mu_k \mathbb{I}_{C_k}(\omega)\big(P^\star(A_k \,|\, C_k) - \mathbb{I}_{A_k}(\omega)\big) \colon \omega \in \bigcup_{k=1}^{n} C_k \right\},$$

where for all $k$ in $\{1,\ldots,n\}$ we let $A_k := \Lambda_\Omega^{-1}(\hat{A}_k)$ and $C_k := \Lambda_\Omega^{-1}(\hat{C}_k)$. By Lemma 7.7[346], $(A_k \,|\, C_k)$ belongs to $\mathscr{D}^\star$ for all $k$ in $\{1,\ldots,n\}$. Because $P^\star$ is a coherent conditional probability on $\mathscr{D}^\star$ by assumption, it follows from the preceding equality and Definition 2.51[44] that $\max \mathcal{S} \geq 0$, as required. □

Theorem 7.8↶ validates our use of the term 'lumped jump process' for $\hat{P}$. Henceforth, we will usually not mention the coherent extension $P^\star$ of $P$ to $\mathscr{D}^\star$ that is used to define a lumped jump process $\hat{P}$. Whenever we talk about 'a lumped jump process $\hat{P}$ corresponding to a jump process $P$', we implicitly assume that $\hat{P}$ is the lumped jump process that corresponds to some coherent extension $P^\star$ of $P$ to $\mathscr{D}^\star$ through Eq. (7.15)↶.

Even though there is no unique lumped jump process, every lumped jump process $\hat{P}$ that we consider agrees with the 'natural' definition of $\hat{P}$.

**Corollary 7.10.** *Consider a jump process $P$ with state space $\mathcal{X}$ and any corresponding lumped jump process $\hat{P}$. Then for all $\{\hat{X}_v \in \hat{B}\}$ in $\hat{\mathscr{F}}_{()}$,*

$$\hat{P}\big(\hat{X}_v \in \hat{B}\big) = P\big(X_v \in \Lambda^{-1}(\hat{B})\big).$$

*Furthermore, for all $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$ with $\hat{P}(\hat{X}_u = \hat{x}_u) > 0$ and all $\hat{A}$ in $\hat{\mathscr{F}}_u$,*

$$\hat{P}(\hat{A} \,|\, \hat{X}_u = \hat{x}_u) = \frac{\hat{P}\big(\{\hat{X}_u = \hat{x}_u\} \cap \hat{A}\big)}{\hat{P}(\hat{X}_u = \hat{x}_u)}$$

*Proof.* Let $P^\star$ be the coherent extension of $P$ to $\mathscr{D}^\star$ that defines $\hat{P}$. First, we prove the first equality in the statement. By Eqs. (7.4)[343] and (7.9)[344], $\Lambda_\Omega^{-1}(\hat{\Omega}) = \Omega$. Hence, it follows from Eq. (7.15)↶ and Lemma 7.6[345] that

$$\hat{P}(\hat{X}_v \in \hat{B}) = \hat{P}(\hat{X}_v \in \hat{B} \,|\, \hat{\Omega}) = P^\star\big(\Lambda_\Omega^{-1}\big(\hat{X}_v \in \hat{B}\big) \,\big|\, \Omega\big) = P^\star(X_v \in \Lambda^{-1}(\hat{B}) \,|\, \Omega).$$

Observe that $(X_v \in \Lambda^{-1}(\hat{B}) \,|\, \Omega)$ belongs to the domain $\mathscr{D}$ of $P$; because $P^\star$ coincides with $P$ on $\mathscr{D}$, it follows from the preceding equality that

$$\hat{P}(\hat{X}_v \in \hat{B}) = P(X_v \in \Lambda^{-1}(\hat{B}) \,|\, \Omega) = P(X_v \in \Lambda^{-1}(\hat{B})),$$

as required.

The second equality in the statement follows immediately from (CP4)[41] because $\hat{P}$ is a coherent conditional probability by Theorem 7.8↶. □

Thus, the only 'degrees of freedom' for a lumped jump process $\hat{P}$ are the probabilities conditional on events with zero probability. A lumped jump process is a coherent conditional probability on $\hat{\mathscr{D}}$, so the coherence condition puts a constraint on these 'undetermined' probabilities. Even more, because any lumped jump process $\hat{P}$ is derived from a coherent extension $P^\star$ of $P$ to $\mathscr{D}^\star$, we can relate the 'undetermined' probabilities to the original jump process $P$: it turns out that $\hat{P}(\bullet \mid \hat{X}_u = \hat{x}_u)$ is some convex combination of $P(\bullet \mid X_u = x_u)$ for $x_u$ in $\Lambda^{-1}(\hat{x}_u)$.

**Lemma 7.11.** *Consider a jump process $P$ with state space $\mathscr{X}$ and any coherent extension $P^\star$ of $P$ to $\mathscr{D}^\star$, and let $\hat{P}$ be the corresponding lumped jump process. Then for all $u$ in $\mathscr{U}$, $\hat{x}_u$ in $\hat{\mathscr{X}}_u$ and $\hat{A}$ in $\hat{\mathscr{F}}_u$,*

$$\hat{P}(\hat{A} \mid \hat{X}_u = \hat{x}_u) = \sum_{x_u \in \hat{x}_u} P\big(\Lambda_\Omega^{-1}(\hat{A}) \,\big|\, X_u = x_u\big) P^\star\big(X_u = x_u \,\big|\, X_u \in \hat{x}_u\big), \tag{7.17}$$

*where $P^\star(X_u = x_u \mid X_u \in \hat{x}_u) \geq 0$ for all $x_u$ in $\mathscr{X}_u$ and*

$$\sum_{x_u \in \hat{x}_u.} P^\star(X_u = x_u \mid X_u \in \hat{x}_u) = 1.$$

*Proof.* To ease our notation, we let $A := \Lambda_\Omega^{-1}(\hat{A})$. Then by Eq. (7.15)$_{347}$,

$$\hat{P}(\hat{A} \mid \hat{X}_u = \hat{x}_u) = P^\star(A \mid X_u \in \hat{x}_u). \tag{7.18}$$

For all $x_u$ in $\Lambda^{-1}(\hat{x}_u)$, $P^\star(X_u = x_u \mid X_u \in \hat{x}_u)$ is non-negative by (CP2)$_{41}$. Furthermore, $\{X_u \in \hat{x}_u\} = \bigcup_{x_u \in \hat{x}_u} \{X_u = x_u\}$, and for all $x_u, y_u$ in $\Lambda^{-1}(\hat{x}_u)$, the events $\{X_u = x_u\}$ and $\{X_u = y_u\}$ are disjoint whenever $x_u \neq y_u$. Consequently,

$$\sum_{x_u \in \hat{x}_u} P^\star(X_u = x_u \mid X_u \in \hat{x}_u) = P^\star(X_u \in \hat{x}_u \mid X_u \in \hat{x}_u) = 1 \tag{7.19}$$

due to (CP3)$_{41}$ and (CP1)$_{41}$. Hence, it follows from (CP9)$_{42}$ that

$$P^\star\big(A \,\big|\, X_u \in \hat{x}_u\big) = P^\star\big(\{X_u \in \hat{x}_u\} \cap A \,\big|\, X_u \in \hat{x}_u\big), \tag{7.20}$$

and from (CP3)$_{41}$ that

$$\begin{aligned} P^\star\big(\{X_u \in \hat{x}_u\} \cap A \,\big|\, X_u \in \hat{x}_u\big) &= P^\star\left(\left(\bigcup_{x_u \in \hat{x}_u} \{X_u = x_u\}\right) \cap A \,\middle|\, X_u \in \hat{x}_u\right) \\ &= P^\star\left(\bigcup_{x_u \in \hat{x}_u} \big(\{X_u = x_u\} \cap A\big) \,\middle|\, X_u \in \hat{x}_u\right) \\ &= \sum_{x_u \in \hat{x}_u} P^\star\big(\{X_u = x_u\} \cap A \,\big|\, X_u \in \hat{x}_u\big). \end{aligned} \tag{7.21}$$

Let us investigate the terms in this sum. For all $x_u$ in $\Lambda^{-1}(\hat{x}_u)$,

$$P^\star\big(\{X_u = x_u\} \cap A \,\big|\, X_u \in \hat{x}_u\big) = P^\star\big(A \,\big|\, \{X_u = x_u\} \cap \{X_u \in \hat{x}_u\}\big) P^\star(X_u = x_u \mid X_u \in \hat{x}_u)$$

due to (CP4)$_{41}$; because $\{X_u = x_u\} \cap \{X_u \in \hat{x}_u\} = \{X_u = x_u\}$ and because $P^\star$ coincides with $P$ on $\mathscr{D}$, this implies that

$$P^\star\big(\{X_u = x_u\} \cap A \,\big|\, X_u \in \hat{x}_u\big) = P(A \mid X_u = x_u) P^\star(X_u = x_u \mid X_u \in \hat{x}_u). \tag{7.22}$$

Equation (7.17) follows immediately from Eq. (7.18) and Eqs. (7.20) to (7.22). □

**Corollary 7.12.** *Consider a jump process $P$ with state space $\mathscr{X}$ and any corresponding lumped jump process $\hat{P}$. Then for all $u$ in $\mathscr{U}$, $\hat{x}_u$ in $\hat{\mathscr{X}}_u$ and $\hat{A}$ in $\hat{\mathscr{F}}_u$,*

$$\min\left\{P\left(\Lambda_\Omega^{-1}(\hat{A}) \,\middle|\, X_u = x_u\right) : x_u \in \hat{x}_u\right\} \leq \hat{P}(\hat{A} \,|\, \hat{X}_u = \hat{x}_u) \leq \max\left\{P\left(\Lambda_\Omega^{-1}(\hat{A}) \,\middle|\, X_u = x_u\right) : x_u \in \hat{x}_u\right\}.$$

*Proof.* Follows immediately from Lemma 7.11↶ because a convex combination of terms is bounded by the minimum and the maximum of these terms. □

### 7.1.3 Expectation corresponding to a lumped jump process

Let us now look at the expectation with respect to a lumped jump process $\hat{P}$, and how this is related to the expectation with respect to the original jump process $P$. To do so, we relate variables on $\hat{\Omega}$ to variables on $\Omega$ through the relation between $\hat{\Omega}$ and $\Omega$. By Lemma 7.5$_{343}$, any path $\omega$ in $\Omega$ corresponds to a lumped path $\Lambda \circ \omega$ in $\hat{\Omega}$. Hence, any extended real variable $\hat{f}$ in $\overline{\mathbb{V}}(\hat{\Omega})$ is related to the *cylindrical extension* $\hat{f}^{\uparrow\Omega}$ of $\hat{f}$ to $\Omega$:

$$\hat{f}^{\uparrow\Omega} : \Omega \to \overline{\mathbb{R}} : \omega \mapsto \hat{f}^{\uparrow\Omega}(\omega) := \hat{f}(\Lambda \circ \omega). \tag{7.23}$$

This cylindrical extension generalises the inverse $\Lambda_\Omega^{-1}$: by Lemma 7.9$_{347}$, the cylindrical extension $(\mathbb{I}_{\hat{A}})^{\uparrow\Omega}$ of the indicator $\mathbb{I}_{\hat{A}}$ of an event $\hat{A}$ in $\mathcal{P}(\hat{\Omega})$ is equal to the indicator $\mathbb{I}_{\Lambda_\Omega^{-1}(\hat{A})}$ of $\Lambda_\Omega^{-1}(\hat{A})$. As simple variables are linear combinations of indicators, this essentially implies that the cylindrical extension of an $\hat{\mathscr{F}}_u$-simple variable is an $\mathscr{F}_u$-simple variable.

**Lemma 7.13.** *Consider a sequence of time points $u$ in $\mathscr{U}$. Then for any $\hat{\mathscr{F}}_u$-simple variable $\hat{f}$, the corresponding cylindrical extension $\hat{f}^{\uparrow\Omega}$ is an $\mathscr{F}_u$-simple variable.*

*Proof.* Because $\hat{f}$ is $\hat{\mathscr{F}}_u$-simple by assumption, there are a natural number $n$, real numbers $a_1, \dots, a_n$ and events $A_1, \dots \hat{A}_n$ in $\hat{\mathscr{F}}_u$ such that

$$\hat{f} = \sum_{k=1}^{n} a_k \mathbb{I}_{\hat{A}_k}$$

due to Definition 2.38$_{36}$. Recall from Lemma 7.9$_{347}$ that for all $k$ in $\{1, \dots, n\}$ and $\omega$ in $\Omega$, $\mathbb{I}_{\hat{A}_k}(\Lambda \circ \omega) = \mathbb{I}_{\Lambda_\Omega^{-1}(\hat{A}_k)}(\omega)$. Thus, we see that

$$\hat{f}^{\uparrow\Omega} = \sum_{k=1}^{n} a_k \mathbb{I}_{\Lambda_\Omega^{-1}(\hat{A}_k)}.$$

Because $\Lambda_\Omega^{-1}(\hat{A}_k)$ belongs to $\mathscr{F}_u$ for all $k$ in $\{1, \dots, n\}$ by Lemma 7.6$_{345}$, this equality shows that $f$ is $\mathscr{F}_u$-simple. □

Combining Lemmas 7.11$_{349}$ and 7.13$_{\curvearrowleft}$, we unearth the relation between the expectation $E_{\hat{P}}$ corresponding to a lumped jump process $\hat{P}$ and the expectation $E_P$ corresponding to the original jump process $E_P$.

**Lemma 7.14.** *Consider a jump process $P$ with state space $\mathscr{X}$ and any corresponding lumped jump process $\hat{P}$. Then for all $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$ and $\hat{f}$ in $\mathbb{S}(\hat{\mathscr{F}}_u)$,*

$$\min\{E_P(\hat{f}^{\uparrow\Omega} \mid X_u = x_u) : x_u \in \hat{x}_u\} \leq E_{\hat{P}}(\hat{f} \mid \hat{X}_u = \hat{x}_u) \\ \leq \max\{E_P(\hat{f}^{\uparrow\Omega} \mid X_u = x_u) : x_u \in \hat{x}_u\}.$$

*Proof.* Let $P^\star$ be the coherent extension of $P$ to $\mathscr{D}^\star$ that defines $\hat{P}$. Recall from the proof of Lemma 7.13$_{\curvearrowleft}$ that there is a natural number $n$, real numbers $a_1, \dots, a_n$ and events $\hat{A}_1, \dots, \hat{A}_n$ in $\hat{\mathscr{F}}_u$ such that

$$\hat{f} = \sum_{k=1}^{n} a_k \mathbb{I}_{\hat{A}_k} \quad \text{and} \quad \hat{f}^{\uparrow\Omega} = \sum_{k=1}^{n} a_k \mathbb{I}_{\Lambda_\Omega^{-1}(\hat{A}_k)},$$

and that $A_1 := \Lambda_\Omega^{-1}(\hat{A}_1), \dots, A_n := \Lambda_\Omega^{-1}(\hat{A}_n)$ are events in $\mathscr{F}_u$. Then by Eq. (2.19)$_{36}$ and Eq. (7.17)$_{349}$ in Lemma 7.11$_{349}$,

$$\begin{aligned} E_{\hat{P}}(\hat{f} \mid X_u = x_u) &= \sum_{k=1}^{n} a_k \hat{P}(\hat{A}_k \mid \hat{X}_u = \hat{x}_u) \\ &= \sum_{k=1}^{n} a_k \sum_{x_u \in \hat{x}_u} P(A_k \mid X_u = x_u) P^\star(X_u = x_u \mid X_u \in \hat{x}_u). \end{aligned} \tag{7.24}$$

We change the order of the summations and use Eq. (2.19)$_{36}$, to yield

$$\begin{aligned} E_{\hat{P}}(\hat{f} \mid X_u = x_u) &= \sum_{x_u \in \hat{x}_u} P^\star(X_u = x_u \mid X_u \in \hat{x}_u) \sum_{k=1}^{n} a_k P(A_k \mid X_u = x_u) \\ &= \sum_{x_u \in \hat{x}_u} P^\star(X_u = x_u \mid X_u \in \hat{x}_u) E_P(\hat{f}^{\uparrow\Omega} \mid X_u = x_u). \end{aligned}$$

By Lemma 7.11$_{349}$, the sum on the right-hand side of the last equality is a convex combination of $E_P(\hat{f}^{\uparrow\Omega} \mid X_u = x_u)$ with $x_u$ in $\Lambda^{-1}(\hat{x}_u)$, so this implies the inequalities in the statement. □

Let us now move from $\hat{\mathscr{F}}_u$-simple variables in $\mathbb{S}(\hat{\mathscr{F}}_u)$ to limit variables in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$. Of course, this is only useful for a countably additive lumped jump process $\hat{P}$. The following result establishes that given a countably additive jump process $P$, any corresponding lumped jump process is countably additive. In essence, this holds because by Lemma 7.11$_{349}$, any lumped jump process $\hat{P}$ corresponding to $P$ is a 'convex combination' of $P$. Formally, we prove this with the help of Lemma 7.14; because this result is rather intuitive, we have relegated our formal proof to Appendix 7.A$_{384}$.

**Lemma 7.15.** *Consider a jump process $P$ with state space $\mathscr{X}$. If $P$ is countably additive, then any corresponding lumped jump process $\hat{P}$ is countably additive too.*

By definition, any limit variable $\hat{f}$ in $\hat{\mathscr{F}}_u$ is the point-wise limit of a sequence $(\hat{f}_n)_{n\in\mathbb{N}}$ of $\hat{\mathscr{F}}_u$ simple variables. Due to Lemma 7.13<sub>350</sub>, for every $\hat{\mathscr{F}}_u$-simple variable $\hat{f}_n$ in this sequence, its cylindrical extension $\hat{f}_n^{\uparrow\Omega}$ is an $\mathscr{F}_u$-simple variable. It is not difficult to verify that this sequence of cylindrical extensions converges to the cylindrical extension $\hat{f}^{\uparrow\Omega}$ of $\hat{f}$, and this basically proves that this cylindrical extension $\hat{f}^{\uparrow\Omega}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$. A formal proof of this result can be found in Appendix 7.A<sub>384</sub>.

**Lemma 7.16.** *Consider a sequence of time points $u$ in $\mathscr{U}$. Then for any limit variable $\hat{f}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$, the corresponding cylindrical extension $\hat{f}^{\uparrow\Omega}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$.*

Finally, we generalise Lemma 7.14 to limit variables: we bound the Daniell expectation $E^{\mathrm{D}}_{\hat{P}}(\hat{f}\,|\,\hat{X}_u=\hat{x}_u)$ of a limit variable $\hat{f}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$ with the Daniell expectation of its cylindrical extension $\hat{f}^{\uparrow\Omega}$ with respect to the original jump process $P$. In essence, the following result follows from Lemma 7.14 and the definition of the Daniell expectation; the formal proof is a bit long, so we have relegated it to Appendix 7.A<sub>384</sub>.

**Lemma 7.17.** *Consider a countably additive jump process $P$ with state space $\mathscr{X}$ and any corresponding lumped jump process $\hat{P}$. Then for all $\{\hat{X}_u=\hat{x}_u\}$ in $\hat{\mathscr{H}}$ and $\hat{f}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$,*

$$\min\left\{E^{\mathrm{D}}_{P}(\hat{f}^{\uparrow\Omega}\,|\,X_u=x_u)\colon x_u\in\hat{x}_u\right\}\leq E^{\mathrm{D}}_{\hat{P}}(\hat{f}\,|\,\hat{X}_u=\hat{x}_u)\leq\max\left\{E^{\mathrm{D}}_{P}(\hat{f}^{\uparrow\Omega}\,|\,X_u=x_u)\colon x_u\in\hat{x}_u\right\}.$$

## 7.2 Lumping and consistency

All of our results in Section 7.1.3<sub>350</sub> put bounds on a lumped jump process $\hat{P}$ in terms of the original jump process $P$. This is rather unfortunate, because this is precisely the opposite of what we wanted to achieve: our intention was to use 'a' jump process $\hat{P}$ with state space $\hat{\mathscr{X}}$ to describe the original jump process $P$ and not the other way around.

Part of the reason why we have not succeeded in our goal yet is that until now, we have considered *any* jump process $P$. This is far from the original setting of a single homogeneous and Markovian jump process $P_{p_0,Q}$ that is considered in most if not all existing research on lumping. Here, we will assume a setting that lies in between these two extremes. Instead of considering any generic jump process or a single homogeneous and Markovian jump process, we consider a jump process $P$ that is consistent with a given set $\mathscr{M}$ of initial probability mass functions on $\mathscr{X}$ and a given bounded set $\mathscr{Q}$ of rate operators on $\mathbb{G}(\mathscr{X})$. Recall from Eq. (3.66)<sub>104</sub> that if $\mathscr{M}$ is the singleton $\{p_0\}$ and $\mathscr{Q}$ the singleton $\{Q\}$, then the only jump process that is consistent with $\mathscr{M}$ and $\mathscr{Q}$ is $P_{p_0,Q}$; hence, our setting includes the particular case of a single homogeneous and Markovian jump process. In any case, as we will see, we can use these sets $\mathscr{M}$ and $\mathscr{Q}$ to describe any lumped jump process $\hat{P}$ corresponding to $P$, at least in terms of consistency. In Section 7.2.1, we derive from $\mathscr{M}$ a subset $\hat{\mathscr{M}}$ of $\Sigma_{\hat{\mathscr{X}}}$ such that any lumped jump process $\hat{P}$ corresponding to $P$ is consistent with $\hat{\mathscr{M}}$, and we do the same in Section 7.2.2$_{355}$ but for a subset $\hat{\mathscr{Q}}$ of $\mathfrak{Q}_{\hat{\mathscr{X}}}$. In Section 7.2.3$_{361}$, we detail why and how we can use lower and upper expectations with respect to the lumped imprecise jump process $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ to bound expectations with respect to the original jump process $P$, and we do the same for an imprecise jump process $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$ in Section 7.2.4$_{368}$. Finally, in Section 7.2.5$_{371}$ we return to the original setting in which lumping was first used.

Throughout this section, we illustrate our findings by means of the following setting in our running example.

**Queuing Network Example 7.18.** Franceschinis et al. (1994) are particularly interested in the case where the number of servers $K$ is even, and where half of the parallel servers have service rate $\underline{\lambda}$ and the other half have service rate $\overline{\lambda}$. Henceforth, we assume that $K$ is even. For now, we assume furthermore that for all $k$ in $\{1,\dots,K\}$, $\lambda_k := \underline{\lambda}$ if $k \leq K/2$ and $\lambda_k := \overline{\lambda}$ otherwise. It is only fair to mention that under these assumptions, the state space $\mathscr{X}$ as defined in Queuing Network Example 7.1$_{338}$ is too detailed. In this particular case, we could get away with keeping track of the number of servers that have $k$ customers, in the spirit of the state description that lead to the lumped state space $\hat{\mathscr{X}}$, but separately for the servers with service rate $\underline{\lambda}$ and those with service rate $\overline{\lambda}$; we refer to Franceschinis et al. (1994, p. 226) for additional details. This 'intermediate' state description results in a significantly reduced state space, but is only applicable to the arguably artificial case where there are two classes of servers. ♖

### 7.2.1 The set of lumped initial probability mass functions

In order to determine the set $\hat{\mathscr{M}}$, we need to determine the initial probability mass function of any lumped jump process $\hat{P}$ corresponding to $P$. Of course, the initial probability mass function of $\hat{P}$ is related to that of $P$ through Corollary 7.10$_{348}$.

**Corollary 7.19.** *Consider a jump process $P$ and any corresponding lumped jump process $\hat{P}$, and denote the initial probability mass functions of $P$ and $\hat{P}$ by $p_0$ and $\hat{p}_0$, respectively. Then for all $\hat{x}$ in $\hat{\mathscr{X}}$,*

$$\hat{p}_0(\hat{x}) = \hat{P}(\hat{X}_0 = \hat{x}) = \sum_{x \in \hat{x}} P(X_0 = x) = \sum_{x \in \hat{x}} p_0(x).$$

*Proof.* By Corollary 7.10$_{348}$ with $\hat{B} = \{\hat{x}\}$,

$$\hat{P}(\hat{X}_0 = \hat{x}) = P\big(X_0 \in \Lambda^{-1}(\hat{x})\big).$$

Observe that $\{X_0 \in \Lambda^{-1}(\hat{x})\} = \bigcup_{x \in \hat{x}} \{X_0 = x\}$. Hence, the equalities of the statement follow from the preceding equality and (JP2)$_{69}$. □

By Corollary 7.19↶, the mass $\hat{p}_0(\hat{x})$ of a lumped state $\hat{x}$ in $\hat{\mathscr{X}}$ can be obtained from the probability mass function $p_0$ of the original jump process $P$ by summing the masses $p_0(x)$ of the states $x$ in the lump $\Lambda^{-1}(\hat{x})$. We generalise this to arbitrary probability mass functions: for any probability mass function $p$ on $\mathscr{X}$, we let $\hat{p}_p$ be the corresponding real-valued map on $\hat{\mathscr{X}}$ defined by

$$\hat{p}_p(\hat{x}) := \sum_{x \in \hat{x}} p(x) \quad \text{for all } \hat{x} \in \hat{\mathscr{X}}. \tag{7.25}$$

It is clear that $\hat{p}_p$ is a probability mass function on $\hat{\mathscr{X}}$, so we call $\hat{p}_p$ the *lumped probability mass function* corresponding to $p$. For any non-empty set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, we gather the corresponding lumped probability mass functions in

$$\hat{\mathscr{M}}_{\mathscr{M}} := \{\hat{p}_p : p \in \mathscr{M}\}. \tag{7.26}$$

In order not to burden our notation, we will simply denote $\hat{\mathscr{M}}_{\mathscr{M}}$ by $\hat{\mathscr{M}}$ whenever it is clear from the context to which original set $\mathscr{M}$ it corresponds.

The definition of $\hat{\mathscr{M}}_{\mathscr{M}}$ is inspired by Corollary 7.19↶, so it is obvious that for a given jump process $P$ that is consistent with $\mathscr{M}$, any corresponding lumped jump process $\hat{P}$ is consistent with $\hat{\mathscr{M}}_{\mathscr{M}}$.

**Lemma 7.20.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$, and a jump process $P$ that is consistent with $\mathscr{M}$. Then any lumped jump process $\hat{P}$ corresponding to $P$ is consistent with $\hat{\mathscr{M}}$.*

*Proof.* For all $\hat{x}$ in $\hat{\mathscr{X}}$, it follows from Corollary 7.19↶ that

$$\hat{P}(\hat{X}_0 = \hat{x}) = \sum_{x \in \hat{x}} p_0(x) = \hat{p}_{p_0}(\hat{x}),$$

where $p_0$ denotes the initial probability mass function of $P$. Because this equality holds for every lumped state $\hat{x}$ in $\hat{\mathscr{X}}$, we conclude that $\hat{P}(X_0 = \bullet) = \hat{p}_{p_0}$. Observe that $p_0$ belongs to $\mathscr{M}$ because $P$ is consistent with $\mathscr{M}$ by assumption. Hence, the lumped probability mass function $\hat{p}_{p_0}$ belongs to $\hat{\mathscr{M}}$ by definition, and this proves the statement. □

In the particular case that the lumped state space $\hat{\mathscr{X}}$ corresponds to some higher-order state description, we can often determine $\underline{E}_{\hat{\mathscr{M}}}$ directly, without having to explicitly construct $\hat{\mathscr{M}}$ or even determine $\underline{E}_{\mathscr{M}}$. In general, the lower envelope of the set $\hat{\mathscr{M}}$ of lumped probability mass functions is related to the lower envelope of the original set $\mathscr{M}$ in the following way.

**Lemma 7.21.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$. Then for all $\hat{f}$ in $\mathbb{G}(\hat{\mathscr{X}})$,*

$$\underline{E}_{\hat{\mathscr{M}}}(\hat{f}) = \underline{E}_{\mathscr{M}}(\hat{f} \circ \Lambda)$$

*Proof.* Fix any probability mass function $p$ in $\mathscr{M}$, and let $\hat{p} \coloneqq \hat{p}_p$. Then by Eq. $(2.8)_{23}$ and Eq. $(7.25)_{\curvearrowleft}$,

$$E_{\hat{p}}(\hat{f}) = \sum_{\hat{x}\in\hat{\mathscr{X}}} \hat{f}(\hat{x})\hat{p}_p(\hat{x}) = \sum_{\hat{x}\in\hat{\mathscr{X}}} \hat{f}(\hat{x}) \sum_{x\in\hat{x}} p(x) = \sum_{\hat{x}\in\hat{\mathscr{X}}} \sum_{x\in\hat{x}} \hat{f}(\hat{x}) p(x).$$

Observe that, for all $\hat{x}$ in $\hat{\mathscr{X}}$ and $x$ in $\Lambda^{-1}(\hat{x})$, $\hat{f}(\hat{x}) = \hat{f}(\Lambda(x)) = [\hat{f}\circ\Lambda](x)$. Hence, it follows from the preceding and Eq. $(2.8)_{23}$ that

$$E_{\hat{p}}(\hat{f}) = \sum_{\hat{x}\in\hat{\mathscr{X}}} \sum_{x\in\hat{x}} [\hat{f}\circ\Lambda](x)p(x) = \sum_{x\in\mathscr{X}} [\hat{f}\circ\Lambda](x)p(x) = E_p(\hat{f}\circ\Lambda).$$

Due to Eq. $(7.26)_{\curvearrowleft}$, the equality in the statement follows from the preceding equality and Eq. $(3.76)_{117}$:

$$\underline{E}_{\hat{\mathscr{M}}}(\hat{f}) = \inf\{E_{\hat{p}_p}(\hat{f}) \colon p \in \mathscr{M}\} = \inf\{E_p(\hat{f}\circ\Lambda) \colon p\in\mathscr{M}\} = \underline{E}_{\mathscr{M}}(\hat{f}\circ\Lambda). \qquad \square$$

### 7.2.2 The set of lumped rate operators

Next, we set out to derive from $\mathscr{Q}$ a set $\hat{\mathscr{Q}}$ of rate operators on $\mathbb{G}(\hat{\mathscr{X}})$ such that any lumped jump process $\hat{P}$ corresponding to $P$ is consistent with $\hat{\mathscr{Q}}$. As consistency with a set of rate operators goes through the transition probabilities by Definition $3.50_{99}$, we should take a look at the transition probabilities for the lumped jump process $\hat{P}$.

**Corollary 7.22.** *Consider a jump process $P$ with state space $\mathscr{X}$ and any coherent extension $P^\star$ of $P$ to $\mathscr{D}^\star$, and let $\hat{P}$ be the corresponding lumped jump process. Then for all $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$, $t, r$ in $\mathbb{R}_{\geq 0}$ such that $u \prec t < r$ and $\hat{x}$ in $\hat{\mathscr{X}}$, there is a probability mass function $p^\star$ on $\Lambda^{-1}(\hat{x}_u) \times \Lambda^{-1}(\hat{x})$ such that*

$$\hat{P}(\hat{X}_r = \hat{y} \,|\, \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) \sum_{x_u\in\hat{x}_u} \sum_{x\in\hat{x}} p^\star(x_u, x) \sum_{y\in\hat{y}} P(X_r = y \,|\, X_u = x_u, X_t = x).$$

*Proof.* Follows almost immediately from Lemmas $7.6_{345}$ and $7.11_{349}$ and $(\text{JP2})_{69}$. $\square$

So, the transition probabilities of any lumped jump process $\hat{P}$ are some convex combination of sums of transition probabilities of $P$. This is perhaps more easily understood if we think about the matrix representation of the history-dependent transition operator $\hat{T}_{t,r}^{\{\hat{X}_u=\hat{x}_u\}}$ of the lumped jump process $\hat{P}$. For the row $\hat{T}_{t,r}^{\{\hat{X}_u=\hat{x}_u\}}(\hat{x}, \bullet)$, we essentially take some convex combination of the rows $T_{t,r}^{\{X_u=x_u\}}(x,\bullet)$ for $x$ in $\Lambda^{-1}(\hat{x})$ and for $x_u$ in $\Lambda^{-1}(\hat{x}_u)$ of the history-dependent transition operator, and then sum some columns. Translating this to the set of rate operators is a bit of a mess; here, we will take a detour via the lower envelopes, as this allows for a more elegant treatment.

For any non-empty and bounded set $\mathscr{Q}$ of rate operators, we define the operator $\underline{\hat{Q}}_{\mathscr{Q}} \colon \mathbb{G}(\hat{\mathscr{X}}) \to \mathbb{G}(\hat{\mathscr{X}})$ that maps a gamble $\hat{f}$ in $\mathbb{G}(\hat{\mathscr{X}})$ to the gamble $\underline{\hat{Q}}_{\mathscr{Q}}\hat{f}$ on $\hat{\mathscr{X}}$, defined for all $\hat{x}$ in $\hat{\mathscr{X}}$ by

$$[\underline{\hat{Q}}_{\mathscr{Q}}\hat{f}](\hat{x}) \coloneqq \min\Big\{[\underline{Q}_{\mathscr{Q}}(\hat{f}\circ\Lambda)](x) \colon x\in\hat{x}\Big\}. \tag{7.27}$$

In order to lighten our notation, we let $\underline{\hat{Q}}_Q := \underline{\hat{Q}}_{\{Q\}}$ for any rate operator $Q$ in $\mathfrak{Q}_{\mathscr{X}}$. Due to the following result, we call the operator $\underline{\hat{Q}}_{\mathscr{Q}}$ the *lumped lower rate operator* corresponding to $\mathscr{Q}$, and similarly for $\underline{\hat{Q}}_Q$.

**Lemma 7.23.** *Consider a non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$. Then the corresponding operator $\underline{\hat{Q}}_{\mathscr{Q}}$ is a lower rate operator on $\mathbb{G}(\hat{\mathscr{X}})$.*

*Proof.* By Definition 3.63$_{109}$, we need to verify that $\underline{\hat{Q}}_{\mathscr{Q}}$ satisfies (LR1)$_{109}$–(LR4)$_{109}$. Before doing so, we let $\underline{\hat{Q}} := \underline{\hat{Q}}_{\mathscr{Q}}$, and we recall from Proposition 3.65$_{110}$ that $\underline{Q} := \underline{Q}_{\mathscr{Q}}$ is a lower rate operator.

(LR1) Fix some $\hat{x}$ in $\hat{\mathscr{X}}$. Observe that $\mathbb{I}_{\hat{\mathscr{X}}} \circ \Lambda = \mathbb{I}_{\mathscr{X}}$, so by Eq. (7.27)$_\frown$,

$$\big[\underline{\hat{Q}}\mathbb{I}_{\hat{\mathscr{X}}}\big](\hat{x}) = \min\Big\{\big[\underline{Q}\mathbb{I}_{\mathscr{X}}\big](x) \colon x \in \hat{x}\Big\} = 0,$$

where the final equality follows from (LR1)$_{109}$ because $\underline{Q}$ is a lower rate operator.

(LR2) Fix some $\hat{x}, \hat{y}$ in $\hat{\mathscr{X}}$ such that $\hat{x} \neq \hat{y}$. Observe that $\mathbb{I}_{\hat{y}} \circ \Lambda = \sum_{y \in \hat{y}} \mathbb{I}_y$, so by Eq. (7.27)$_\frown$,

$$\big[\underline{\hat{Q}}\mathbb{I}_{\hat{y}}\big](\hat{x}) = \min\left\{\left[\underline{Q}\left(\sum_{y \in \hat{y}} \mathbb{I}_y\right)\right](x) \colon x \in \hat{x}\right\}.$$

For all $x$ in $\Lambda^{-1}(\hat{x})$,

$$\left[\underline{Q}\left(\sum_{y \in \hat{y}} \mathbb{I}_y\right)\right](x) \geq \sum_{y \in \hat{y}} [\underline{Q}\mathbb{I}_y](x) \geq 0,$$

where the inequalities follow from (LR4)$_{109}$ and (LR2)$_{109}$ because $\underline{Q}$ is a lower rate operator and $x \neq y$ for all $y$ in $\Lambda^{-1}(\hat{y})$. From the preceding equality and inequality, it follows immediately that $[\underline{\hat{Q}}\mathbb{I}_{\hat{y}}](\hat{x}) \geq 0$, as required.

(LR3) Fix some $\hat{x}$ in $\hat{\mathscr{X}}$, $\hat{f}$ in $\mathbb{G}(\hat{\mathscr{X}})$ and $\mu$ in $\mathbb{R}_{\geq 0}$. Observe that $(\mu\hat{f}) \circ \Lambda = \mu(\hat{f} \circ \Lambda)$, so by Eq. (7.27)$_\frown$,

$$\begin{aligned}\big[\underline{\hat{Q}}(\mu\hat{f})\big](\hat{x}) = \min\Big\{\big[\underline{Q}(\mu(\hat{f} \circ \Lambda))\big](x) \colon x \in \hat{x}\Big\} &= \min\Big\{\mu\big[\underline{Q}(\hat{f} \circ \Lambda)\big](x) \colon x \in \hat{x}\Big\} \\ &= \mu\big[\underline{\hat{Q}}\hat{f}\big](\hat{x}),\end{aligned}$$

where for the second equality we used (LR3)$_{109}$ for $\underline{Q}$ and for the third equality we used that $\mu$ is non-negative by assumption.

(LR4) Fix some $\hat{x}$ in $\hat{\mathscr{X}}$, and some $\hat{f}, \hat{g}$ in $\mathbb{G}(\hat{\mathscr{X}})$. Observe that $(\hat{f} + \hat{g}) \circ \Lambda = \hat{f} \circ \Lambda + \hat{g} \circ \Lambda$, so

$$\big[\underline{\hat{Q}}(\hat{f} + \hat{g})\big](\hat{x}) = \min\Big\{\big[\underline{Q}(\hat{f} \circ \Lambda + \hat{g} \circ \Lambda)\big](x) \colon x \in \hat{x}\Big\}.$$

Because $\underline{Q}$ satisfies (LR4)$_{109}$ and the minimum operator is super-additive, it follows from this that

$$\begin{aligned}\big[\underline{\hat{Q}}(\hat{f} + \hat{g})\big](\hat{x}) &\geq \min\Big\{\big[\underline{Q}(\hat{f} \circ \Lambda)\big](x) + \big[\underline{Q}(\hat{g} \circ \Lambda)\big](x) \colon x \in \hat{x}\Big\} \\ &\geq \min\Big\{\big[\underline{Q}(\hat{f} \circ \Lambda)\big](x) \colon x \in \hat{x}\Big\} + \min\Big\{\big[\underline{Q}(\hat{g} \circ \Lambda)\big](x) \colon x \in \hat{x}\Big\} \\ &= \big[\underline{\hat{Q}}\hat{f}\big](\hat{x}) + \big[\underline{\hat{Q}}\hat{g}\big](\hat{x}),\end{aligned}$$

as required. □

Important to mention here is that whenever the lumped state space corresponds to some higher-order state description, the optimisation in Eq. (7.27)355 tends to be fairly straightforward. For example, in Chapter 8403 further on, determining $\underline{\hat{Q}}_{\mathcal{Q}}\hat{f}$ reduces to computing the minimum over a small number of cases, and this is also the case in the setting of our running example; see also Queuing Network Example 7.36368 further on.

**Queuing Network Example 7.24.** Let us determine $\underline{\hat{Q}}_{Q}$ as defined by Eq. (7.27)355 for the rate operator $Q$ defined in Queueing Network Examples 7.3340 and 7.18353. As an intermediate step, we determine $[Q(\hat{f}\circ\Lambda)](x)$ for any rate operator $Q$ as defined by Eq. (7.1)342 in Queuing Network Example 7.3340.

Fix any $\hat{f}$ in $\mathbb{G}(\hat{\mathscr{X}})$ and $\hat{x}=(n_0,m_0,\dots,m_N)$ in $\hat{\mathscr{X}}$. Then by Eq. (7.1)342 in Queuing Network Example 7.3340, for all $x=(x_0,x_1,\dots,x_K)$ in $\Lambda^{-1}(\hat{x})$,

$$\begin{aligned}
&[Q(\hat{f}\circ\Lambda)](x)\\
&\quad=\sum_{k=1}^{K}\frac{\mu}{K}\big([\hat{f}\circ\Lambda](x_k^+)-[\hat{f}\circ\Lambda](x)\big)+\sum_{k\in\mathscr{K}_x}\lambda_k\big([\hat{f}\circ\Lambda](x_k^-)-[\hat{f}\circ\Lambda](x)\big)\\
&\quad=\sum_{k=1}^{K}\frac{\mu}{K}\big([\hat{f}\circ\Lambda](x_k^+)-\hat{f}(\hat{x})\big)+\sum_{k\in\mathscr{K}_x}\lambda_k\big([\hat{f}\circ\Lambda](x_k^-)-\hat{f}(\hat{x})\big),
\end{aligned}\tag{7.28}$$

where the first summation is only added if $x_0>0$ and where $\mathscr{K}_x$ is the set of indices $k$ in $\{1,\dots,K\}$ such that $x_k>0$.

Fix any $x=(x_0,x_1,\dots,x_K)$ in $\Lambda^{-1}(\hat{x})$. By our discussion in Queuing Network Example 7.4342, $x_0=n_0$ and for all $\ell$ in $\{1,\dots,N\}$, there are $m_{\ell-1}$ indices $k$ in $\{1,\dots,K\}$ such that $x_k=\ell-1$. The first summation in Eq. (7.28) is only added if $x_0=n_0>0$. In this case, for all $k$ in $\{1,\dots,K\}$, $\Lambda(x_k^+)$ corresponds to the lumped state $(n_0-1,m_0,\dots,m_{\ell-1}-1,m_\ell+1,\dots,m_N)$ with $\ell-1=x_k$. Because there are $m_{\ell-1}$ indices $k$ in $\{1,\dots,K\}$ such that $x_k=\ell-1$, we see that

$$\sum_{k=1}^{K}\frac{\mu}{K}\big([\hat{f}\circ\Lambda](x_k^+)-\hat{f}(\hat{x})\big)=\sum_{\ell\in\mathscr{L}_{\hat{x}}^+}\frac{\mu}{K}m_{\ell-1}\big(\hat{f}(\hat{x}_\ell^+)-\hat{f}(\hat{x})\big),\tag{7.29}$$

where $\mathscr{L}_{\hat{x}}^+$ is the set of indices $\ell$ in $\{1,\dots,N\}$ such that $m_{\ell-1}>0$ and where for all $\ell$ in $\mathscr{L}_{\hat{x}}^+$, we let $\hat{x}_\ell^+:=(n_0-1,m_0,\dots,m_{\ell-1}-1,m_\ell+1,\dots,m_N)$.

The second summation in Eq. (7.28) is a bit more involved. There are $\sum_{\ell=1}^{N}m_\ell$ indices $k$ in $\{1,\dots,K\}$ such that $x_k>0$, and for each of these indices $k$, $\Lambda(x_k^-)$ is the lumped state $(n_0+1,m_0,\dots,m_{\ell-1}+1,m_\ell-1,\dots,m_N)$ with $\ell=x_k$. We let $\mathscr{L}_{\hat{x}}^-$ be the set of indices $\ell$ in $\{1,\dots,N\}$ such that $m_\ell>0$, and for each $\ell$ in $\mathscr{L}_{\hat{x}}^-$, we let $\hat{x}_\ell^-:=(n_0,m_0,\dots,m_{\ell-1}+1,m_\ell-1,\dots,m_N)$ and we let $\mathscr{K}_{x,\ell}$ be those $m_\ell$ indices in $\{1,\dots,K\}$ such that $x_k=\ell$. Then

$$\sum_{k\in\mathscr{K}_x}\lambda_k\big([\hat{f}\circ\Lambda](x_k^-)-\hat{f}(\hat{x})\big)=\sum_{\ell\in\mathscr{L}_{\hat{x}}^-}\left(\sum_{k\in\mathscr{K}_{x,\ell}}\lambda_k\right)\big(\hat{f}(\hat{x}_\ell^-)-\hat{f}(\hat{x})\big).\tag{7.30}$$

For all $x = (x_0, \dots, x_K)$ in $\Lambda^{-1}(\hat{x})$, it follows from Eqs. (7.28) to (7.30) that

$$[Q(\hat{f} \circ \Lambda)](x) = \sum_{\ell \in \mathscr{L}^+_{\hat{x}}} \frac{\mu}{K} m_\ell \big(\hat{f}(\hat{x}^+_\ell) - \hat{f}(\hat{x})\big) + \sum_{\ell \in \mathscr{L}^-_{\hat{x}}} \left( \sum_{k \in \mathscr{K}_{x,\ell}} \lambda_k \right) \big(\hat{f}(\hat{x}^-_\ell) - \hat{f}(\hat{x})\big). \quad (7.31)$$

Finally, we return to the setting in Queuing Network Example $7.18_{353}$, so $K$ is even and we take $\lambda_k = \underline{\lambda}$ if $k \le K/2$ and $\lambda_k = \overline{\lambda}$ otherwise. By Eq. $(7.27)_{355}$, $[\underline{\hat{Q}}_Q \hat{f}](\hat{x})$ is the minimum of $[Q(\hat{f} \circ \Lambda)](x)$ for $x$ in $\Lambda^{-1}(\hat{x})$. Basically, we run over all possible configurations of customers of the $K$ parallel servers, with the only constraint that, for all $\ell$ in $\{1, \dots, N\}$, exactly $m_\ell$ servers have $\ell$ customers. Hence, it follows from Eq. (7.31) and the definition of $\mathscr{K}_{x,\ell}$ that

$$[\underline{\hat{Q}}_Q \hat{f}](\hat{x}) = \sum_{\ell \in \mathscr{L}^+_{\hat{x}}} \frac{\mu}{K} m_\ell \big(\hat{f}(\hat{x}^+_\ell) - \hat{f}(\hat{x})\big) + \sum_{\ell \in \mathscr{L}^-_{\hat{x}}} (\lambda^\star_{\ell,1} + \dots + \lambda^\star_{\ell,m_\ell}) \big(\hat{f}(\hat{x}^-_\ell) - \hat{f}(\hat{x})\big), \quad (7.32)$$

where the $\lambda^\star_{\ell,i}$'s are equal to $\underline{\lambda}$ or $\overline{\lambda}$ and are chosen in such a way that they minimise the expression, under the constraint that, in total, at most $K/2$ can have the same value. One rather obvious way to solve this minimisation problem is to order the values of $\hat{f}(\hat{x}^-_\ell) - \hat{f}(\hat{x})$ for the indices $\ell$ in $\mathscr{L}^-_{\hat{x}}$, and to subsequently assign as much $\overline{\lambda}$'s as possible to the negative values and as much $\underline{\lambda}$'s as possible to positive values. 𝕰

Because $\underline{\hat{Q}}_{\mathcal{Q}}$ is a lower rate operator, we know from Lemma $3.66_{110}$ that it induces a set of dominating transition rate matrices. In the present setting of lumping, we call

$$\hat{\mathcal{Q}}_{\mathcal{Q}} := \left\{ \hat{Q} \in \mathfrak{Q}_{\hat{\mathcal{X}}} : \big(\forall \hat{f} \in \mathbb{G}(\hat{\mathcal{X}})\big)\ \hat{Q}\hat{f} \ge \underline{\hat{Q}}_{\mathcal{Q}} \hat{f} \right\} \quad (7.33)$$

the set of *lumped rate operators*. Following our notational convention for $\hat{\mathscr{M}}$, we will often drop the subscript of $\hat{\mathcal{Q}}_{\mathcal{Q}}$ whenever the original set $\mathcal{Q}$ is clear from the context. Similarly, in line with our notational convention for $\underline{\hat{Q}}_Q$, we let $\hat{\mathcal{Q}}_Q := \hat{\mathcal{Q}}_{\{Q\}}$ for all $Q$ in $\mathfrak{Q}_{\mathcal{X}}$.

It follows immediately from Lemmas $3.66_{110}$ and $3.69_{111}$ that the set $\hat{\mathcal{Q}}$ has the following nice properties.

**Corollary 7.25.** *Consider a non-empty and bounded set $\mathcal{Q}$ of rate operators. Then the corresponding set $\hat{\mathcal{Q}}_{\mathcal{Q}}$ of lumped rate operators is non-empty, bounded, closed and convex and has separately specified rows, and its lower envelope is $\underline{\hat{Q}}_{\mathcal{Q}}$.*

*Proof.* Because $\underline{\hat{Q}}_{\mathcal{Q}}$ is a lower rate operator by Lemma 7.23$_{356}$ and the set $\hat{\mathcal{Q}}_{\mathcal{Q}}$ is defined in Eq. (7.33)$_{\curvearrowleft}$ as the set of rate operators that dominate $\underline{\hat{Q}}_{\mathcal{Q}}$, the statement follows immediately from Lemmas 3.66$_{110}$ and 3.69$_{111}$. □

Suppose $P$ is a jump process that is consistent with some non-empty and bounded subset $\mathcal{Q}$ of $\mathfrak{Q}_{\mathcal{X}}$. Our definition of the set of lumped rate operators $\hat{\mathcal{Q}}$ is inspired by Lemmas 7.21$_{354}$ and 7.14$_{351}$, and the latter implies that any lumped jump process $\hat{P}$ corresponding to $P$ is consistent with $\hat{\mathcal{Q}}$. We would like to invoke Proposition 3.57$_{104}$ for our proof, so first we establish that $\hat{P}$ has bounded rate.

**Lemma 7.26.** *Consider a jump process $P$ that has bounded rate. Then any corresponding lumped jump process $\hat{P}$ has bounded rate too.*

*Proof.* By Lemma 3.54$_{102}$, we need to show that for all $t$ in $\mathbb{R}_{\geq 0}$, $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$ such that $u \prec t$ and $\hat{x}$ in $\hat{\mathcal{X}}$,

$$\limsup_{r \searrow t} \frac{1}{r-t}\big(1 - \hat{P}(\hat{X}_r = \hat{x} \,|\, \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x})\big) < +\infty \tag{7.34}$$

and, if $t > 0$,

$$\limsup_{s \nearrow t} \frac{1}{t-s}\big(1 - P(\hat{X}_t = \hat{x} \,|\, \hat{X}_u = \hat{x}_u, \hat{X}_s = \hat{x})\big) < +\infty. \tag{7.35}$$

Thus, we fix some $t$ in $\mathbb{R}_{\geq 0}$, $\{\hat{X}_u = \hat{x}_u\}$ in $\mathscr{H}$ such that $u \prec t$ and $\hat{x}$ in $\hat{\mathcal{X}}$.

Here, we will only prove Eq. (7.34); the proof for Eq. (7.35) is analoguous. Fix some $r$ in $]t, +\infty[$. Then by Corollary 7.12$_{350}$ and Lemma 7.6$_{345}$,

$$\hat{P}(\hat{X}_r = \hat{x} \,|\, \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) \geq \min\big\{P(X_r \in \hat{x} \,|\, X_u = x_u, X_t = x) \colon x_u \in \hat{x}_u, x \in \hat{x}\big\}.$$

For all $x_u$ in $\Lambda^{-1}(\hat{x}_u)$ and $x$ in $\Lambda^{-1}(\hat{x})$, it follows from (CP8)$_{42}$ that

$$P(X_r \in \hat{x} \,|\, X_u = x_u, X_t = x) \geq P(X_r = x \,|\, X_u = x_u, X_t = x).$$

Hence, by the previous two inequalities,

$$\hat{P}(\hat{X}_r = \hat{x} \,|\, \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x}) \geq \min\big\{P(X_r = x \,|\, X_u = x_u, X_t = x) \colon x_u \in \hat{x}_u, x \in \hat{x}\big\},$$

and therefore

$$\begin{aligned}\frac{1}{r-t}&\big(1 - \hat{P}(\hat{X}_r = \hat{x} \,|\, \hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x})\big) \\ &\leq \frac{1}{r-t}\big(1 - \min\big\{P(X_r = x \,|\, X_u = x_u, X_t = x) \colon x_u \in \hat{x}_u, x \in \hat{x}\big\}\big) \\ &= \max\Big\{\frac{1}{r-t}\big(1 - P(X_r = x \,|\, X_u = x_u, X_t = x)\big) \colon x_u \in \hat{x}_u, x \in \hat{x}\Big\}.\end{aligned}$$

Because $P$ has bounded rate by assumption and $\Lambda^{-1}(\hat{x}_u) \times \Lambda^{-1}(\hat{x})$ is finite, Eq. (7.34) follows from this inequality and Lemma 3.54$_{102}$ for $P$. □

Due to the preceding result and Proposition 3.57$_{104}$, the result that we are after follows from Lemma 7.14$_{351}$ and the definition of $\hat{\mathcal{Q}}_{\mathcal{Q}}$.

**Lemma 7.27.** *Consider a non-empty and bounded subset $\mathcal{Q}$ of $\mathfrak{Q}_{\mathcal{X}}$, and a jump process $P$ that is consistent with $\mathcal{Q}$. Then any lumped jump process $\hat{P}$ corresponding to $P$ is consistent with $\hat{\mathcal{Q}}_{\mathcal{Q}}$.*

In our proof, we need Lemma 7.14$_{351}$ in the following form.

**Corollary 7.28.** *Consider a jump process $P$ with state space $\mathcal{X}$, and let $\hat{P}$ be any corresponding lumped jump process. Then for all $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$, $t, r$ in $\mathbb{R}_{\geq 0}$ such that $u < t \leq r$, $\hat{x}$ in $\hat{\mathcal{X}}$ and $\hat{f}$ in $\mathbb{G}(\hat{\mathcal{X}})$,*

$$\begin{aligned}\min\{E_P\big([\hat{f}\circ\Lambda](X_r)\,\big|\,X_u = x_u, X_t = x\big) \colon x_u \in \hat{x}_u, x \in \hat{x}\}\\ \leq E_{\hat{P}}(\hat{f}(\hat{X}_r)\,|\,\hat{X}_u = \hat{x}_u, \hat{X}_t = \hat{x})\\ \leq \max\{E_P\big([\hat{f}\circ\Lambda](X_r)\,\big|\,X_u = x_u, X_t = x\big) \colon x_u \in \hat{x}_u, x \in \hat{x}\}.\end{aligned}$$

*Proof.* Observe that for all $\omega$ in $\Omega$,

$$[\hat{f}\circ\Lambda]\big(\omega(r)\big) = \hat{f}\big(\Lambda\big(\omega(r)\big)\big) = \hat{f}\big([\Lambda\circ\omega](r)\big),$$

so $\big(\hat{f}(\hat{X}_r)\big)^{\uparrow\Omega} = [\hat{f}\circ\Lambda](X_r)$. For this reason, the inequalities of the statement follow immediately from Lemma 7.14$_{351}$. □

*Proof of Lemma 7.27.* By Lemma 3.55$_{102}$, $P$ has bounded rate because by assumption $P$ is consistent with $\mathcal{Q}$ and $\mathcal{Q}$ is bounded; consequently, $\hat{P}$ has bounded rate too by Lemma 7.26$_{\frown}$. We adhere to the same notation as in Lemma 7.26$_{\frown}$: for all time points $t, r$ in $\mathbb{R}_{\geq 0}$ and all state histories $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$ such that $u < t \leq r$, we let $\hat{T}_{t,r}^{\{\hat{X}_u=\hat{x}_u\}}$ denote the history-dependent transition operator corresponding to $\hat{P}$ as defined by Eq. (3.35)$_{84}$.

By Definition 3.50$_{99}$, we need to show that for all $t$ in $\mathbb{R}_{\geq 0}$ and $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$ such that $u < t$, $\hat{T}_{t,t}^{\{\hat{X}_u=\hat{x}_u\}}$ is $d_{\hat{\mathfrak{Q}}}$-differentiable – where we let $\hat{\mathfrak{Q}} := \mathfrak{Q}_{\hat{\mathcal{X}}}$ – with

$$\partial_+ \hat{T}_{t,t}^{\{\hat{X}_u=\hat{x}_u\}} \subseteq \hat{\mathcal{Q}}_{\mathcal{Q}} \quad \text{and, if } t > 0, \quad \partial_- \hat{T}_{t,t}^{\{\hat{X}_u=\hat{x}_u\}} \subseteq \hat{\mathcal{Q}}_{\mathcal{Q}}.$$

We will only prove this for the right-sided $d_{\hat{\mathfrak{Q}}}$-derivative, the proof for the left-sided one is analogous.

Fix some $t$ in $\mathbb{R}_{\geq 0}$ and $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$ such that $u < t$. As $\hat{P}$ has bounded rate, it follows from Proposition 3.57$_{104}$ that $\hat{T}_{t,t}^{\{\hat{X}_u=\hat{x}_u\}}$ is $d_{\hat{\mathfrak{Q}}}$-differentiable and that

$$\partial_+ \hat{T}_{t,t}^{\{\hat{X}_u=\hat{x}_u\}} = \left\{\hat{Q} \in \mathfrak{Q}_{\hat{\mathcal{X}}} \colon \big(\exists (r_n)_{n\in\mathbb{N}} \searrow t\big) \lim_{n\to+\infty} \frac{\hat{T}_{t,r_n}^{\{\hat{X}_u=\hat{x}_u\}} - I}{r_n - t} = \hat{Q}\right\}.$$

We fix any $\hat{Q}$ in $\partial_+ \hat{T}_{t,t}^{\{\hat{X}_u=\hat{x}_u\}}$, and let $(r_n)_{n\in\mathbb{N}}$ be a decreasing sequence in $]t, +\infty[$ with $\lim_{n\to+\infty} r_n = t$ such that

$$\lim_{n\to+\infty} \frac{\hat{T}_{t,r_n}^{\{\hat{X}_u=\hat{x}_u\}} - I}{r_n - t} = \hat{Q}. \tag{7.36}$$

Fix some $n$ in $\mathbb{N}$, $\hat{f}$ in $\mathbb{G}(\hat{\mathscr{X}})$ and $\hat{x}$ in $\hat{\mathscr{X}}$. Then by Corollary 7.28,

$$\big[\hat{T}_{t,r_n}^{\{\hat{X}_u=\hat{x}_u\}}\hat{f}\big](\hat{x}) \geq \min\Big\{\big[T_{t,r_n}^{\{X_u=x_u\}}(\hat{f}\circ\Lambda)\big](x) \colon x_u \in \hat{x}_u, x \in \hat{x}\Big\}.$$

Recall from Proposition $3.80_{117}$ that because $P$ is consistent with the bounded set $\mathscr{Q}$ of rate operators,

$$\big[T_{t,r_n}^{\{X_u=x_u\}}(\hat{f}\circ\Lambda)\big](x) \geq \Big[e^{(r_n-t)\underline{Q}_{\mathscr{Q}}}(\hat{f}\circ\Lambda)\Big](x) \quad \text{for all } x_u \in \hat{x}_u, x \in \hat{x}.$$

From the two preceding inequalities, we infer that

$$\big[\hat{T}_{t,r_n}^{\{\hat{X}_u=\hat{x}_u\}}\hat{f}\big](\hat{x}) \geq \min\big\{\big[e^{(r_n-t)\underline{Q}_{\mathscr{Q}}}(\hat{f}\circ\Lambda)\big](x) \colon x \in \hat{x}\big\}.$$

We subtract $\hat{f}(\hat{x})$ from both sides of this inequality and divide both sides by $r_n - t$, to yield

$$\frac{\big[\hat{T}_{t,r_n}^{\{\hat{X}_u=\hat{x}_u\}}\hat{f}\big](\hat{x}) - \hat{f}(\hat{x})}{r_n - t} \geq \min\left\{\frac{\big[e^{(r_n-t)\underline{Q}_{\mathscr{Q}}}(\hat{f}\circ\Lambda)\big](x) - [\hat{f}\circ\Lambda](x)}{r_n-t} \colon x \in \hat{x}\right\}. \tag{7.37}$$

Because $\lim_{n\to+\infty} r_n = t$, it follows from Proposition $3.78_{115}$ that

$$\lim_{n\to+\infty} \frac{\big[e^{(r_n-t)\underline{Q}_{\mathscr{Q}}}(\hat{f}\circ\Lambda)\big](x) - [\hat{f}\circ\Lambda](x)}{r_n-t} = \big[\underline{Q}_{\mathscr{Q}}(\hat{f}\circ\Lambda)\big](x) \quad \text{for all } x \in \hat{x},$$

and therefore

$$\lim_{n\to+\infty} \min\left\{\frac{\big[e^{(r_n-t)\underline{Q}_{\mathscr{Q}}}(\hat{f}\circ\Lambda)\big](x) - [\hat{f}\circ\Lambda](x)}{r_n-t} \colon x \in \hat{x}\right\} = \min\big\{\big[\underline{Q}_{\mathscr{Q}}(\hat{f}\circ\Lambda)\big](x) \colon x \in \hat{x}\big\}.$$

Hence, it follows from Eqs. (7.36) and (7.37) and the preceding equality that

$$[\hat{Q}\hat{f}](\hat{x}) \geq \min\big\{\big[\underline{Q}_{\mathscr{Q}}(\hat{f}\circ\Lambda)\big](x) \colon x \in \hat{x}\big\} = [\underline{\hat{Q}}_{\mathscr{Q}}\hat{f}](\hat{x}),$$

where for the equality we used Eq. $(7.27)_{355}$. Because $\hat{x}$ and $\hat{f}$ are arbitrary elements of $\hat{\mathscr{X}}$ and $\mathbb{G}(\hat{\mathscr{X}})$, respectively, we infer from this inequality and Eq. $(7.33)_{358}$ that $\hat{Q}$ belongs to $\hat{\mathscr{Q}}_{\mathscr{Q}}$. Seeing that $\hat{Q}$ is an arbitrary element of $\partial_+\hat{T}_{t,t}^{\{\hat{X}_u=\hat{x}_u\}}$, we can finally conclude that $\partial_+\hat{T}_{t,t}^{\{\hat{X}_u=\hat{x}_u\}} \subseteq \hat{\mathscr{Q}}_{\mathscr{Q}}$, which is what we needed to prove. □

### 7.2.3 Describing a consistent jump process with the lumped imprecise jump process

Suppose we have a jump process $P$ that is consistent with the set $\mathscr{M}$ of initial probability mass functions on $\mathscr{X}$ and the non-empty and bounded set $\mathscr{Q}$ of rate operators on $\mathbb{G}(\mathscr{X})$. Then any lumped jump process $\hat{P}$ corresponding to $P$ is consistent with $\hat{\mathscr{M}}$ due to Lemma $7.20_{354}$ and consistent with $\hat{\mathscr{Q}}$ due to Lemma 7.27.

**Theorem 7.29.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$ and a non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$. Fix some jump process $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Then every corresponding lumped jump process $\hat{P}$ is consistent with $\hat{\mathscr{M}}_{\mathscr{M}}$ and $\hat{\mathscr{Q}}_{\mathscr{Q}}$.*

*Proof.* Follows immediately from Lemmas 7.20<sub>354</sub> and 7.27<sub>360</sub> □

This is a nice theoretical result, but it is not entirely the result that we are after: Theorem 7.29 characterises any lumped jump process $\hat{P}$ corresponding to $P$ in terms of consistency, but it does *not* make clear how we can use this lumped jump process $\hat{P}$ to describe the original jump process $P$. In particular, we would like to say something about expectations of the form $E_P^{\mathrm{D}}(f \mid X_u = x_u)$, with $\{X_u = x_u\}$ in $\mathscr{H}$ and $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$, and it is not immediately obvious how we can do so using Theorem 7.29 only.

Let us start with the unconditional case, so with $u = ()$ and therefore

$$\{X_u = x_u\} = \{X_{()} = x_{()}\} = \Omega.$$

Note that $\Lambda^{-1}(\Lambda(x_{()}))$ is equal to the singleton $\{x_{()}\}$. Consequently, it follows from Lemma 7.17<sub>352</sub> that the expectation with respect to any lumped jump process $\hat{P}$ corresponding to $P$ essentially coincides with the expectation with respect to $P$.

**Proposition 7.30.** *Consider a jump process $P$ and a corresponding lumped jump process $\hat{P}$. Then for all $\hat{f}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_{()})$,*

$$E_{\hat{P}}^{\mathrm{D}}(\hat{f}) = E_P^{\mathrm{D}}(\hat{f}^{\uparrow\Omega}).$$

*Proof.* Observe that $\Omega = \{X_{()} = x_{()}\}$ and that $\hat{\Omega} = \{\hat{X}_{()} = \hat{x}_{()}\}$. Because $\Lambda^{-1}(\hat{x}_{()}) = \{x_{()}\}$, it follows from Lemma 7.17<sub>352</sub> that

$$E_{\hat{P}}^{\mathrm{D}}(\hat{f}) = E_{\hat{P}}^{\mathrm{D}}(\hat{f} \mid \hat{X}_{()} = \hat{x}_{()}) = E_P^{\mathrm{D}}(\hat{f}^{\uparrow\Omega} \mid X_{()} = x_{()}) = E_P^{\mathrm{D}}(\hat{f}^{\uparrow\Omega}).$$ □

We combine this with Theorem 7.29 to prove the result that we are after.

**Corollary 7.31.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$ and a non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$. Then for all $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_{()})$ and $\hat{g}, \hat{h}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_{()})$ such that $\hat{g}^{\uparrow\Omega} \leq f \leq \hat{h}^{\uparrow\Omega}$,*

$$\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}(\hat{g}) \leq E_P^{\mathrm{D}}(f) \leq \overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}(\hat{h}).$$

*Proof.* Let $\hat{P}$ be any lumped jump process corresponding to $P$. Then by Proposition 7.30,

$$E_{\hat{P}}^{\mathrm{D}}(\hat{g}) = E_P^{\mathrm{D}}(\hat{g}^{\uparrow\Omega}) \quad \text{and} \quad E_{\hat{P}}^{\mathrm{D}}(\hat{h}) = E_P^{\mathrm{D}}(\hat{h}^{\uparrow\Omega}).$$

Because furthermore $\hat{g}^{\uparrow\Omega} \leq f \leq \hat{h}^{\uparrow\Omega}$ by assumption,

$$E_{\hat{P}}^{\mathrm{D}}(\hat{g}) = E_P^{\mathrm{D}}(\hat{g}^{\uparrow\Omega}) \leq E_P^{\mathrm{D}}(f) \leq E_P^{\mathrm{D}}(\hat{h}^{\uparrow\Omega}) = E_{\hat{P}}^{\mathrm{D}}(\hat{h}),$$

due to (DE6)<sub>226</sub>. The lumped jump process $\hat{P}$ belongs to $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ by Theorem 7.29, so the inequalities in the statement follow immediately from these inequalities. □

In the conditional case, so with $u \neq ()$, things are way more complicated. We cannot put to use the same argument as in the unconditional case, because the inequalities do not reduce to equalities. Specifically, the inequalities in Lemma 7.17<sub>352</sub> are facing the wrong way: Lemma 7.17<sub>352</sub> gives bounds on the expectation with respect to any lumped jump process $\hat{P}$ in terms of the original jump process $P$, but we want to use a – possibly imprecise – jump process with state space $\hat{\mathscr{X}}$ to bound the expectation with respect to the original jump process $P$. Nevertheless, we can extend Corollary 7.31 to the conditional case, and this is made possible due to the following result; note that in this result, unlike in Proposition 7.30, the jump process $\hat{P}$ is *not* guaranteed to be a lumped jump process corresponding to the original jump process $P$. The statement of this result is straightforward, but looks can be deceiving; the entirety of Appendix 7.B<sub>387</sub> is devoted to its proof.

**Theorem 7.32.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$ and a non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$. Fix any $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\{X_u = x_u\}$ in $\mathscr{H}$ and $\hat{f}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}})_u$. Then there is a jump process $\hat{P}$ in $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ such that*

$$E^{\mathrm{D}}_{\hat{P}}\big(\hat{f} \,\big|\, \hat{X}_u = \Lambda(x_u)\big) = E^{\mathrm{D}}_{P}\big(\hat{f}^{\uparrow\Omega} \,\big|\, X_u = x_u\big).$$

Because we have Theorem 7.32, it is child's play to prove the result that we are after.

**Theorem 7.33.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$ and a non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$. Then for all $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\{X_u = x_u\}$ in $\mathscr{H}$, $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ and $\hat{g}, \hat{h}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$ such that $\hat{g}^{\uparrow\Omega} \leq f \leq \hat{h}^{\uparrow\Omega}$,*

$$\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}\big(\hat{g} \,\big|\, \hat{X}_u = \Lambda(x_u)\big) \leq E^{\mathrm{D}}_{P}\big(f \,\big|\, X_u = x_u\big) \leq \overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}\big(\hat{h} \,\big|\, \hat{X}_u = \Lambda(x_u)\big).$$

*Proof.* By Theorem 7.32, there is a jump process $\hat{P}$ in $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ such that

$$E^{\mathrm{D}}_{\hat{P}}\big(\hat{g} \,\big|\, \hat{X}_u = \Lambda(x_u)\big) = E^{\mathrm{D}}_{P}\big(\hat{g}^{\uparrow\Omega} \,\big|\, X_u = x_u\big) \leq E^{\mathrm{D}}_{P}\big(f \,\big|\, X_u = x_u\big),$$

where for the inequality we used (DE6)<sub>226</sub>. Because $\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ is the lower envelope of the expectations with respect to the jump processes in $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$, this inequality implies the first inequality in the statement.

The second inequality is implied by the first equality: note that $-\hat{h} \leq -f$, so

$$\begin{aligned} E^{\mathrm{D}}_{P}\big(f \,\big|\, X_u = x_u\big) &= -E^{\mathrm{D}}_{P}\big(-f \,\big|\, X_u = x_u\big) \\ &\leq -\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}\big(-\hat{h} \,\big|\, \hat{X}_u = \Lambda(x_u)\big) = \overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}\big(\hat{h} \,\big|\, \hat{X}_u = \Lambda(x_u)\big), \end{aligned}$$

where for the first equality we used (DE4)<sub>225</sub>, for the inequality we used the first inequality in the statement and for the second equality we used conjugacy. □

At this point, it is only fair to mention that Katoen et al. (2012) obtained a result that is – in some sense – similar to Theorem 7.33, but *much* less general.

They state their results for what they call an ‘abstract continuous-time Markov chain’ (see Katoen et al., 2012, Definition 7), but comparing their results to Theorem 7.33 only makes sense for the particular case of a homogeneous Markovian jump process $P_{p_0,Q}$. Instead of lumping $P_{p_0,Q}$ directly, they use the uniformisation method – also known as Jensen’s (1953) method, see (Diener et al., 1995) or (Stewart, 2009, Section 10.7.2) – to obtain the so-called embedded Markov chain, and subsequently lump this embedded Markov chain. Katoen et al. (2012, Sections 4.3 and 5) show how to use this lumped embedded Markov chain to compute the lower and upper probability of until events – at least for subsets $S, G$ of $\mathscr{X}$ such that $S = \mathscr{X}$ and $G = \Lambda^{-1}(\Lambda(G))$. Although they do not appear to mention this explicitly, a similar approach can be used to determine lower and upper bounds on the expectation of simple variables.

**Lumpable variables**

Theorem 7.33 is only relevant from a practical point of view if given an ‘interesting’ $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$, we can easily find variables $\hat{g}, \hat{h}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$ such that $\hat{g}^{\uparrow\Omega} \leq f \leq \hat{h}^{\uparrow\Omega}$. Whenever the lumped state space $\hat{\mathscr{X}}$ corresponds to a higher-level state description, this is usually not a problem; as mentioned in the introduction to this chapter, ‘in many cases, we can formalise the inferences we would like to make in a higher-level state description’. Formally, a variable $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ ‘can be formalised in the higher-level state description’ if there is a variable $\hat{f}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$ such that $\hat{f}^{\uparrow\Omega} = f$; whenever this is the case, we call the variable $f$ *lumpable*. Note that $\hat{f}$ then takes the role of $\hat{g}$ *and* $\hat{h}$. Let us investigate for some particular types of variables whether they are lumpable or not.

We start with the $\mathscr{F}_u$-simple variables. Recall from (our discussion before) Lemma 3.1568 that any $\mathscr{F}_u$-simple variable is of the form $f(X_v)$, with $v$ in $\mathscr{U}_{\succcurlyeq u}$ and $f$ in $\mathbb{G}(\mathscr{X}_v)$. Thus, we fix some $u$ in $\mathscr{U}$, $v$ in $\mathscr{U}_{\succcurlyeq u}$ and $f$ in $\mathbb{G}(\mathscr{X}_v)$. We draw inspiration from Franceschinis et al. (1994, p. 232) and let

$$f^{\downarrow\min}\colon \hat{\mathscr{X}}_v \to \mathbb{R}\colon \hat{x}_v \mapsto f^{\downarrow\min}(\hat{x}_v) := \min\{f(x_v)\colon x_v \in \hat{x}_v\} \tag{7.38}$$

and

$$f^{\downarrow\max}\colon \hat{\mathscr{X}}_v \to \mathbb{R}\colon \hat{x}_v \mapsto f^{\downarrow\max}(\hat{x}_v) := \max\{f(x_v)\colon x_v \in \hat{x}_v\}. \tag{7.39}$$

Note that, in general,

$$f^{\downarrow\min}\big(\Lambda(x_v)\big) \leq f(x_v) \leq f^{\downarrow\max}\big(\Lambda(x_v)\big) \quad \text{for all } x_v \in \mathscr{X}_v. \tag{7.40}$$

Obviously, these inequalities hold with equality if and only if $f$ is constant on the lumps, in the sense that $f(x_v) = f(y_v)$ for all $x_v, y_v$ in $\mathscr{X}_v$ such that $\Lambda(x_v) = \Lambda(y_v)$; in this case, we call $f$ *lumpable*. It follows immediately from Eq. (7.40) that

$$\big(f^{\downarrow\min}(\hat{X}_v)\big)^{\uparrow\Omega} \leq f(X_v) \leq \big(f^{\downarrow\max}(\hat{X}_v)\big)^{\uparrow\Omega}, \tag{7.41}$$

and it is obvious that $f(X_v)$ is lumpable if and only if $f$ is lumpable.

**Queuing Network Example 7.34.** Franceschinis et al. (1994) are interested in the 'population' at $\mathrm{S}_0$: the number of customers that are being serviced by or waiting to be serviced by $\mathrm{S}_0$. At time $t$ in $\mathbb{R}_{\geq 0}$, the population at $\mathrm{S}_0$ is $f(X_t)$, with

$$f\colon \mathscr{X} \to \mathbb{R}\colon (n_0, n_1, \dots, n_K) \mapsto n_0. \tag{7.42}$$

Note that $f^{\downarrow\min} = f^{\downarrow\max} = \hat{f}$ with

$$\hat{f}\colon \hat{\mathscr{X}} \to \mathbb{R}\colon (n_0, m_0, \dots, m_N) \mapsto n_0, \tag{7.43}$$

so $f(X_t)$ is a lumpable variable because $(\hat{f}(\hat{X}_t))^{\uparrow\Omega} = f(X_t)$.

Let $p_0$ and $Q$ be as defined in Queueing Network Examples $7.3_{340}$ and $7.18_{353}$, and let $\hat{\mathscr{M}} := \hat{\mathscr{M}}_{\{p_0\}}$ and $\hat{\mathscr{Q}} := \hat{\mathscr{Q}}_Q$. For all $x$ in $\mathscr{X}$, it follows from Theorem $7.33_{363}$ that, with $\hat{x} := \Lambda(x)$,

$$\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}(\hat{f}(\hat{X}_t) \,|\, X_0 = \hat{x}) \leq E_{p_0,Q}(f(X_t) \,|\, X_0 = x) \leq \overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}(\hat{f}(\hat{X}_t) \,|\, X_0 = \hat{x}).$$

To determine these (lower and upper) expectations, we recall from Theorem $3.37_{87}$ that $T_{0,t} = e^{tQ}$, where $T_{0,t}$ is the transition operator from 0 to $t$ corresponding to $P_{p_0,Q}$. Hence, it follows from Eq. $(3.35)_{84}$ that

$$E_{p_0,Q}(f(X_t) \,|\, X_0 = x) = \left[e^{tQ} f\right](x).$$

Because the set $\hat{\mathscr{Q}}$ has separately specified rows due to Corollary $7.25_{358}$ and has $\underline{Q}_{\hat{\mathscr{Q}}} = \underline{\hat{Q}}_Q$ as lower envelope, it follows from Proposition $3.81_{117}$ that

$$\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}(\hat{f}(\hat{X}_t) \,|\, X_0 = \hat{x}) = \left[e^{t\underline{Q}_{\hat{\mathscr{Q}}}} \hat{f}\right](\hat{x}) = \left[e^{t\underline{\hat{Q}}_Q} \hat{f}\right](\hat{x}),$$

and similarly for the upper expectation due to conjugacy. We can approximate these with the methods discussed in Section $4.2_{173}$. When doing so, we use Eq. $(7.32)_{358}$ in Queuing Network Example $7.24_{357}$ to determine $\underline{\hat{Q}}_Q$.

For a numerical example, we are interested in two 'extreme' initial states: the state $x_{\mathrm{e}} := (0, N, 0, \dots, 0)$ where the queue at $\mathrm{S}_0$ is empty and all customers are at $\mathrm{S}_1$, and a state $x_{\mathrm{f}} := (N, 0, \dots, 0)$ where all customers are at $\mathrm{S}_0$. Observe that $\hat{x}_{\mathrm{e}} := \Lambda(x_{\mathrm{e}}) = (0, K-1, 0, \dots, 1)$ and $\hat{x}_{\mathrm{f}} := \Lambda(x_{\mathrm{f}}) = (N, K, 0, \dots, 0)$. We use the same values for the system parameters as Franceschinis et al. (1994): $K = 4$, $N = 5$, $\mu = 1$, $\underline{\lambda} = 1$ and $\overline{\lambda} = 1.01$. We report the expected population at time $t = 10$ conditional on $\{X_0 = x\}$, as well as the lower and upper bounds, in Table 7.1$_{\curvearrowright}$. The lower and upper bounds obtained with $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ are quite good: they only differ from the actual value from the third significant digit on. That said, we do observe that, for these values of $K$ and $N$, computing the lower and upper bounds with $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ takes *longer* than computing the actual value of the expectation, which is rather unfortunate. This is a consequence of the relatively small dimensions of the system; for larger values of $K$ and $N$,

**Table 7.1** Lower and upper bounds on the expected population at $\mathrm{S}_0$ for $t = 10$ conditional on $\{X_0 = x\}$. Parameters: $K = 4$, $N = 5$, $\mu = 1$, $\underline{\lambda} = 1$ and $\overline{\lambda} = 1.01$.

| $x$ | $\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ | $E_{p_0,Q}$ | $\overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ |
|---|---|---|---|
| $x_\mathrm{e}$ | 3.604 | 3.610 | 3.623 |
| $x_\mathrm{f}$ | 3.726 | 3.733 | 3.740 |

computing the lower and upper bounds with $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ is actually (considerably) faster than computing the actual value of the expectation. We do not compare the computation times in detail here, as the observations regarding computation times in Queuing Network Example 7.39<sub>370</sub> further on hold in this case too. 𝔗

Next, we investigate the lumpability of variables $f$ in $\overline{\mathbb{V}}_{\mathrm{lim}}(\mathscr{F}_u)$ that are not $\mathscr{F}_u$-simple. Recall from Eq. (5.14)$_{227}$ that such a limit variable $f$ is defined as the point-wise limit of a sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathscr{F}_u$-simple variables; note that this sequence need not be unique. Intuitively, we expect that $f$ is lumpable whenever there is some defining sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathscr{F}_u$-simple variables such that $f_n$ is lumpable for all $n$ in $\mathbb{N}$. While we could deal with this in detail, it makes more sense to consider the three particular types of limit variables for which we have computational methods. For these three types, the lumpability is almost trivial to check and, perhaps more importantly, the lumped variables are of the same type as the original variable.

Let us start with the indicator $h^{S,G}_{[s,r]}$ of an until event, with $S,G$ subsets of $\mathscr{X}$ and $s,r$ time points in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Intuitively, we expect that this indicator $h^{S,G}_{[s,r]}$ is lumpable if the sets $S$ and $G$ contain entire lumps. In general, we say that a subset $B$ of $\mathscr{X}$ is *lumpable* whenever

$$B = \Lambda^{-1}\big(\Lambda(B)\big) = \big\{x \in \mathscr{X} \colon \Lambda^{-1}\big(\Lambda(x)\big) \subseteq B\big\};$$

obviously, this is the case if and only if

$$(\mathbb{I}_B)^{\downarrow\min} = \mathbb{I}_{\hat{B}} = (\mathbb{I}_B)^{\downarrow\max} \quad \text{with } \hat{B} := \Lambda(B) = \{\Lambda(x) \colon x \in B\}.$$

If $S$ and $G$ are lumpable, then clearly

$$\big(h^{\hat{S},\hat{G}}_{[s,r]}\big)^{\uparrow\Omega} = h^{S,G}_{[s,r]} \quad \text{with } \hat{S} := \Lambda(S) \text{ and } \hat{G} := \Lambda(G),$$

so the indicator $h^{S,G}_{[s,r]}$ is lumpable. For the hitting time $\tau^G_{[s,r]}$, we find something similar: if $G$ is lumpable, then

$$\big(\tau^{\hat{G}}_{[s,r]}\big)^{\uparrow\Omega} = \tau^G_{[s,r]} \quad \text{with } \hat{G} := \Lambda(G).$$

Finally, we come to limit variables of the form $\int_s^r f_t(X_t)\,\mathrm{d}t$. Then obviously,

$$\left(\int_s^r [f_t^{\downarrow\min}](\hat{X}_t)\,\mathrm{d}t\right)^{\uparrow\Omega} \leq \int_s^r f_t(X_t)\,\mathrm{d}t \leq \left(\int_s^r [f_t^{\downarrow\max}](\hat{X}_t)\,\mathrm{d}t\right)^{\uparrow\Omega}. \tag{7.44}$$

Hence, we see that if $f_t$ is lumpable for all $t$ in $[s,r]$, then $\int_s^r f(X_t)\,\mathrm{d}t$ is lumpable.

**Queuing Network Example 7.35.** As in Queuing Network Example 7.34$_{365}$, we are interested in the population at $\mathrm{S}_0$, and in particular in the temporal average of this population over $[0,r]$. That is, we seek to determine the expectation of $[\![f]\!]_{[0,r]} = \frac{1}{r}\int_0^r f(X_t)\,\mathrm{d}t$, with $f$ as defined in Eq. (7.42)$_{365}$. Recall from Queuing Network Example 7.34$_{365}$ that $f^{\downarrow\min} = \hat{f} = f^{\downarrow\max}$, with $\hat{f}$ as defined in Eq. (7.43)$_{365}$, so

$$\left([\![\hat{f}]\!]_{[0,r]}\right)^{\uparrow\Omega} = [\![f]\!]_{[0,r]} \quad \text{with} \quad [\![\hat{f}]\!]_{[0,r]} = \frac{1}{r}\int_0^r \hat{f}(\hat{X}_t)\,\mathrm{d}t$$

due to Eq. (7.44). For this reason, it follows from Theorem 7.33$_{363}$ that, for all $x$ in $\mathscr{X}$ and with $\hat{x} := \Lambda(x)$,

$$\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}([\![\hat{f}]\!]_{[0,r]} \mid X_0 = \hat{x}) \leq E_{p_0,Q}([\![f]\!]_{[0,r]} \mid X_0 = x) \leq \overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}([\![\hat{f}]\!]_{[0,r]} \mid X_0 = \hat{x}).$$

To compute these (lower and upper) expectations, we resort to the iterative procedure in Theorem 6.50$_{318}$, following essentially the same method as in Power Network Example 6.51$_{319}$.

We use the same parameter values as in Queuing Network Example 7.34$_{365}$: $K = 4$, $N = 5$, $\mu = 1$, $\underline{\lambda} = 1$ and $\overline{\lambda} = 1.01$. To determine the expectation with respect to $P_{p_0,Q}$, we determine $\tilde{f}_{n,0}$ as defined by the recursive scheme in Eq. (6.36)$_{319}$, first for $n = n_{\min} := r\|Q\|_{\mathrm{op}}$, then for $n = 2n_{\min}$, and so on until we observe 'convergence' up to four significant digits; we follow the same strategy for the lower and upper expectations with respect to $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$. In Table 7.2, we report (lower and upper bounds on) the expected temporal average of the population at $\mathrm{S}_0$ over $[0,10]$, where – as in Queuing Network Example 7.34$_{365}$ – we consider the two 'extreme' initial states $x_{\mathrm{e}}$ and $x_{\mathrm{f}}$. As in Queuing Network Example 7.34$_{365}$, we observe that the bounds ob-

**Table 7.2** Lower and upper bounds on the expected temporal average of the population at $\mathrm{S}_0$ over $[0,10]$ conditional on $\{X_0 = x\}$. Parameters: $K = 4$, $N = 5$, $\mu = 1$, $\underline{\lambda} = 1$ and $\overline{\lambda} = 1.01$.

| $x$ | $\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ | $E_{p_0,Q}$ | $\overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ |
|---|---|---|---|
| $x_{\mathrm{e}}$ | 2.467 | 2.470 | 2.487 |
| $x_{\mathrm{f}}$ | 3.920 | 3.925 | 3.929 |

tained with $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ are decent, but that for these dimensions, the computations for $P_{p_0,Q}$ take less time than those for $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ – see also Table 7.5371 further on. The plot of the temporal evolution of the expected temporal average and its bounds in Fig. 7.5 confirm the accuracy of the bounds; we only plot the lower bound for $x_{\mathrm{e}}$ because this is indistinguishable from the upper bound, and similarly for the upper bound for $x_{\mathrm{f}}$.

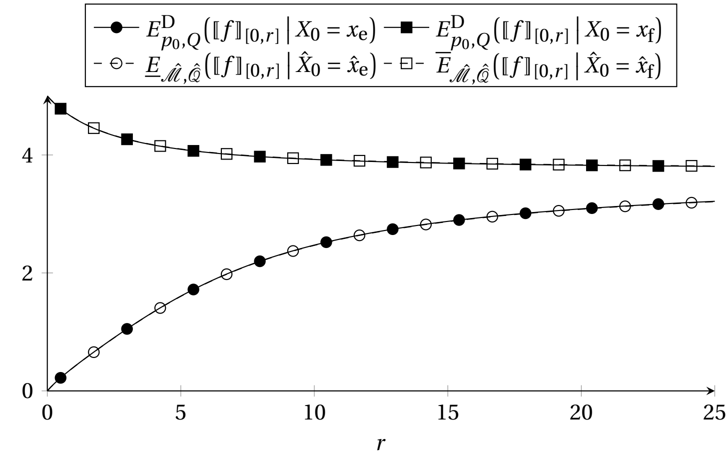



**Figure 7.5** Expected temporal average of the population at $S_0$ over $[0,r]$ as a function of $r$. Parameters: $K=4$, $N=5$, $\mu=1$, $\underline{\lambda}=1$ and $\overline{\lambda}=1.01$.

### 7.2.4 Describing an imprecise jump process with the lumped imprecise jump process

In Section 7.2.3361, we looked at a single jump process $P$ that is consistent with $\mathscr{M}$ and $\mathscr{Q}$. Of course, we might instead be interested in a set of such processes, so in an imprecise jump process $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$.

**Queuing Network Example 7.36.** In Queuing Network Example 7.18353, we made the somewhat arbitrary assumption that half of the parallel servers have service rate $\lambda_k = \underline{\lambda}$ and the other half have service rate $\lambda_k = \overline{\lambda}$. Alternatively, we can also model the system by means of an imprecise jump process model, so that we do not need to specify the service rates in a precise or exact manner. Henceforth, we will consider the homogeneous and Markovian imprecise jump process $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, with $\mathscr{M}$ and $\mathscr{Q}$ defined as follows.

The set $\mathscr{M}$ of initial probability mass functions does not play a role in our analysis, so it can be any non-empty subset of $\Sigma_{\mathscr{X}}$. Here, we assume that we have no knowledge about the initial state, so we let $\mathscr{M} := \Sigma_{\mathscr{X}}$.

Instead of considering the single rate operator $Q$ as defined in Eq. (7.1)$_{342}$, here we consider the set $\mathscr{Q}$ of rate operators $Q'$ that are defined as by Eq. (7.1)$_{342}$ in Queuing Network Example 7.3$_{340}$, but where the rates $\lambda_1$, ..., $\lambda_K$ are only required to belong to $[\underline{\lambda}, \overline{\lambda}]$ and where this choice of rates need not even be the same for all $x$ in $\mathscr{X}$. Note that, by construction, the rate operator $Q$ as defined in Queuing Network Example 7.18$_{353}$ – so the one where $\lambda_k = \underline{\lambda}$ if $k \leq K/2$ and $\lambda_k = \overline{\lambda}$ otherwise – belongs to $\mathscr{Q}$.

It is not difficult to verify that, by construction, $\mathscr{Q}$ is non-empty, bounded and has separately specified rows, so the lower envelope $\underline{Q}_{\mathscr{Q}}$ is well-defined. Because we can vary the $\lambda_k$'s separately, it follows immediately from Eq. (7.1)$_{342}$ that, for all $f$ in $\mathbb{G}(\mathscr{X})$ and $x = (n_0, \ldots, n_K)$ in $\mathscr{X}$,

$$\begin{aligned}\left[\underline{Q}_{\mathscr{Q}} f\right](x) = \sum_{k=1}^{K} \frac{\mu}{K}\big(f(x_k^+) - f(x)\big) \\ + \sum_{k \in \mathscr{K}_x} \min\big\{\lambda_k\big(f(x_k^-) - f(x)\big) \colon \lambda_k \in \{\underline{\lambda}, \overline{\lambda}\}\big\}, \quad (7.45)\end{aligned}$$

where the first summation is only added if $n_0 > 0$.

Let $\hat{\mathscr{Q}}_2 := \hat{\mathscr{Q}}_{\mathscr{Q}}$. By definition – see Eq. (7.33)$_{358}$ – the lower envelope of this set is $\underline{\hat{Q}}_{\mathscr{Q}}$, as defined by Eq. (7.27)$_{355}$. For all $\hat{f}$ in $\mathbb{G}(\hat{\mathscr{X}})$ and $\hat{x} = (n_0, m_0, \ldots, m_N)$ in $\hat{\mathscr{X}}$, we find after a bit of work – and using Eq. (7.27)$_{355}$ and Eq. (7.31)$_{358}$ in Queuing Network Example 7.24$_{357}$ – that

$$\begin{aligned}\left[\underline{\hat{Q}}_{\mathscr{Q}} \hat{f}\right](\hat{x}) = \sum_{\ell \in \mathscr{L}_{\hat{x}}^+} \frac{\mu}{K} m_\ell\big(\hat{f}(\hat{x}_\ell^+) - \hat{f}(\hat{x})\big) \\ + \sum_{\ell \in \mathscr{L}_{\hat{x}}^-} \min\big\{m_\ell \lambda_\ell\big(\hat{f}(\hat{x}_\ell^-) - \hat{f}(\hat{x})\big) \colon \lambda_\ell \in \{\underline{\lambda}, \overline{\lambda}\}\big\}. \quad (7.46)\end{aligned}$$

Note that this optimisation problem is easier than the one in Eq. (7.32)$_{358}$ for $\underline{\hat{Q}}_Q$. ℧

Even if we are interested in an imprecise jump process $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$ instead of in a single jump process $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, we can still use Theorem 7.33$_{363}$ to bound the lower and upper expectations with respect to $\mathscr{P}$. The following corollary makes clear how exactly.

**Corollary 7.37.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathcal{X}}$, a non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathcal{X}}$, and let $\hat{\mathscr{M}} := \hat{\mathscr{M}}_{\mathscr{M}}$ and $\hat{\mathscr{Q}} := \hat{\mathscr{Q}}_{\mathscr{Q}}$. Then for any imprecise jump process $\mathscr{P} \subseteq \mathbb{P}_{\mathscr{M},\mathscr{Q}}$, and any $\{X_u = x_u\}$ in $\mathscr{H}$, $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ and $\hat{g}, \hat{h}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$ such that $\hat{g}^{\uparrow\Omega} \leq f \leq \hat{h}^{\uparrow\Omega}$,*

$$\begin{aligned}\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}\big(\hat{g} \,\big|\, \hat{X}_u = \Lambda(x_u)\big) \leq \underline{E}_{\mathscr{P}}\big(f \,\big|\, X_u = x_u\big) \\ \leq \overline{E}_{\mathscr{P}}\big(f \,\big|\, X_u = x_u\big) \leq \overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}\big(\hat{h} \,\big|\, \hat{X}_u = \Lambda(x_u)\big).\end{aligned}$$

*Proof.* Follows immediately from Theorem 7.33$_{363}$. □

**Table 7.3** Lower and upper bounds on the (lower and upper) expected population at $\mathrm{S}_0$ for $t = 10$ conditional on $\{X_0 = x\}$. Parameters: $K = 4$, $N = 5$, $\mu = 1$, $\underline{\lambda} = 1$ and $\overline{\lambda} = 1.01$.

| $x$ | $\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_2}$ | $\underline{E}_{\mathscr{M},\mathscr{Q}}$ | $E_{p_0,Q}$ | $\overline{E}_{\mathscr{M},\mathscr{Q}}$ | $\overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_2}$ |
|---|---|---|---|---|---|
| $x_{\mathrm{e}}$ | 3.603 | 3.603 | 3.610 | 3.623 | 3.623 |
| $x_{\mathrm{f}}$ | 3.726 | 3.726 | 3.733 | 3.740 | 3.740 |

Let us put Corollary 7.37↶ to work in the setting of Queuing Network Example 7.36$_{368}$. First, we do so for a simple variable.

**Queuing Network Example 7.38.** As in Queuing Network Example 7.34$_{365}$, we determine the expected 'population' at $\mathrm{S}_0$. This time around, we model the system with the homogeneous and Markovian imprecise jump process $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, with $\mathscr{M}$ and $\mathscr{Q}$ as defined in Queuing Network Example 7.36$_{368}$.

Then for all $x$ in $\mathscr{X}$, it follows from Theorem 7.32$_{363}$ that, with $\hat{x} := \Lambda(x)$,

$$\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_2}\big(\hat{f}(\hat{X}_t)\,\big|\,\hat{X}_0 = \hat{x}\big) \leq \underline{E}_{\mathscr{M},\mathscr{Q}}\big(f(X_t)\,\big|\,X_0 = x\big) \leq \overline{E}_{\mathscr{M},\mathscr{Q}}\big(f(X_t)\,\big|\,X_0 = x\big) \leq \overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_2}\big(\hat{f}(\hat{X}_t)\,\big|\,\hat{X}_0 = \hat{x}\big).$$

Because $\mathscr{Q}$ and $\hat{\mathscr{Q}}_2$ have separately specified rows, we can use the methods discussed in Section 4.2$_{173}$ to approximate these lower and upper bounds. When doing so, we use the expression in Eq. (7.45)↶ to determine $\underline{Q}_{\mathscr{Q}}$ and the expression in Eq. (7.46)↶ to determine $\underline{Q}_{\hat{\mathscr{Q}}_2} = \hat{\underline{Q}}_{\mathscr{Q}}$.

We use the same parameters as in Queuing Network Example 7.34$_{365}$, and report the lower and upper expected population at time $t = 10$ conditional on $\{X_0 = x\}$, as well as the lower and upper bounds, in Table 7.3. The lower and upper bounds obtained with $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_2}$ are good, and they are (almost) equal to those obtained with $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$. However, we observe that computing the former is (considerably) faster than computing the latter; we do not report these computation times because they follow the same trends as those in Table 7.5↷ further on. 𝔗𝔅

Finally, we give an example where we use Corollary 7.37↶ for a lumpable limit variable.

**Queuing Network Example 7.39.** As in Queuing Network Example 7.35$_{367}$, we seek to determine the (lower and upper) expected average population at $\mathrm{S}_0$, so the lower and upper expectation of $[\![f]\!]_{[0,r]} = \frac{1}{r}\int_0^r f(X_t)\,\mathrm{d}t$ with $f$ as defined in Eq. (7.42)$_{365}$. By Corollary 7.37↶, for all $x$ in $\mathscr{X}$ and with $\hat{x} := \Lambda(x)$,

$$\underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_2}([\![\hat{f}]\!]_{[0,r]}\,|\,X_0 = \hat{x}) \leq \underline{E}_{\mathscr{M},\mathscr{Q}}([\![f]\!]_{[0,r]}\,|\,X_0 = x) \leq \overline{E}_{\mathscr{M},\mathscr{Q}}([\![f]\!]_{[0,r]}\,|\,X_0 = x) \leq \overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_2}([\![\hat{f}]\!]_{[0,r]}\,|\,X_0 = \hat{x}).$$

We follow the method in Queuing Network Example 7.35$_{367}$ to determine these lower and upper bounds, using the same parameters. In Table 7.4↷,

**Table 7.4** Lower and upper bounds on the (lower and upper) expected temporal average of the population at $\mathrm{S}_0$ over $[0,10]$ conditional on $\{X_0 = x\}$. Parameters: $K = 4$, $N = 5$, $\mu = 1$, $\underline{\lambda} = 1$ and $\overline{\lambda} = 1.01$.

| $x$ | $\underline{E}_{\hat{\mathscr{M}},\hat{\mathbb{Q}}_2}$ | $\underline{E}_{\mathscr{M},\mathbb{Q}}$ | $E_{p_0,Q}$ | $\overline{E}_{\mathscr{M},\mathbb{Q}}$ | $\overline{E}_{\hat{\mathscr{M}},\hat{\mathbb{Q}}_2}$ |
|---|---|---|---|---|---|
| $x_\mathrm{e}$ | 2.466 | 2.466 | 2.470 | 2.487 | 2.487 |
| $x_\mathrm{f}$ | 3.919 | 3.919 | 3.925 | 3.930 | 3.930 |

we report the lower and upper expected temporal averages of the population at $\mathrm{S}_0$ over $[0,10]$ for the two 'extreme' initial states $x_\mathrm{e}$ and $x_\mathrm{f}$. As in Queuing Network Example 7.38, we observe that the bounds obtained with $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathbb{Q}}_2}$ are quite good; they are also pretty close to those obtained with $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathbb{Q}}}$.

However, there is a big difference between computation times for $P_{p_0,Q}$, $\mathbb{P}_{\mathscr{M},\mathbb{Q}}$, $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathbb{Q}}}$ and $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathbb{Q}}_2}$. To investigate these differences a bit more, we run the same numerical example for different values of $N$ and $K$, and report the computation times in Table 7.5. The difference in runtime that we observe is

**Table 7.5** Duration of computations (in seconds) to determine (lower and upper bounds on) the expected temporal average of the population at $\mathrm{S}_0$ over $[0,10]$ (average over 100 runs). Parameters: $\mu = 1$, $\underline{\lambda} = 1$ and $\overline{\lambda} = 1.01$.

| $K$ | $N$ | $\lvert\mathscr{X}\rvert$ | $\lvert\hat{\mathscr{X}}\rvert$ | $P_{p_0,Q}$ | $\mathbb{P}_{\mathscr{M},\mathbb{Q}}$ | $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathbb{Q}}}$ | $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathbb{Q}}_2}$ |
|---|---|---|---|---|---|---|---|
| 4 | 6 | 210 | 27 | 0.01308 | 0.1606 | 1.119 | 0.3061 |
| 6 | 8 | 3003 | 64 | 0.6983 | 4.192 | 2.008 | 1.096 |
| 8 | 10 | 43758 | 136 | 17.42 | 47.46 | 5.935 | 1.551 |
| 10 | 12 | 646646 | 269 | 793.2 | 1012 | 30.48 | 8.108 |

explained by the size of the state space: as is to be expected, the runtimes are more or less proportional to the cardinality of (lumped) state space. Observe also that the runtimes for $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathbb{Q}}}$ are longer than those for $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathbb{Q}}_2}$, even though they both have $\hat{\mathscr{X}}$ as state space; this is in line with our observation in Queuing Network Example 7.36368 that the optimisation problem for $\underline{\hat{Q}}_Q$ is harder than that for $\underline{\hat{Q}}_{\mathbb{Q}}$. ⊛

### 7.2.5 Lumpability of a homogeneous Markovian jump process

To conclude this section, we return to the original setting in which the lumping procedure was first proposed: we suppose we have a homogeneous and Markovian jump process $P_{p_0,Q}$, where $p_0$ is a probability mass function on $\mathscr{X}$ and $Q$ a rate operator on $\mathbb{G}(\mathscr{X})$.

Burke et al. (1958, Theorem 4) were the first to investigate whether 'the' corresponding lumped jump process $\hat{P}$ is Markovian, and they found that the following condition is necessary and sufficient.

**Definition 7.40.** A rate operator $Q$ in $\mathfrak{Q}_{\mathcal{X}}$ is *lumpable* if there is some $\hat{Q}$ in $\mathfrak{Q}_{\hat{\mathcal{X}}}$ such that

$$(\forall \hat{x}, \hat{y} \in \hat{\mathcal{X}})(\forall x \in \hat{x}) \sum_{y \in \hat{y}} Q(x,y) = \hat{Q}(\hat{x}, \hat{y}). \tag{7.47}$$

Let us investigate what the implications of lumpability are for our 'alternative' approach. First, we take a look at the set $\hat{\mathcal{Q}}_Q$ of lumped rate operators corresponding to a lumpable rate operator $Q$.

**Lemma 7.41.** *Consider a rate operator $Q$ in $\mathfrak{Q}_{\mathcal{X}}$, and let $\underline{\hat{Q}} := \underline{\hat{Q}}_Q$ and $\hat{\mathcal{Q}} := \hat{\mathcal{Q}}_Q$. Then $Q$ is lumpable if and only if $\underline{\hat{Q}}$ is linear; whenever this is the case, $\hat{\mathcal{Q}} = \{\hat{Q}\}$ and $\underline{\hat{Q}} = \hat{Q}$, where $\hat{Q}$ is the unique rate operator that satisfies Eq. (7.47).*

*Proof.* Recall from Eq. (7.27)355 that, for all $\hat{f}$ in $\mathbb{G}(\hat{\mathcal{X}})$ and $\hat{x}$ in $\hat{\mathcal{X}}$,

$$\begin{aligned} [\underline{\hat{Q}}\hat{f}](\hat{x}) &= \min\left\{[Q(\hat{f}\circ\Lambda)](x) \colon x \in \hat{x}\right\} = \min\left\{\sum_{y\in\mathcal{X}} Q(x,y)[\hat{f}\circ\Lambda](y) \colon x\in\hat{x}\right\} \\ &= \min\left\{\sum_{\hat{y}\in\mathcal{X}} \hat{f}(\hat{y}) \sum_{y\in\hat{y}} Q(x,y) \colon x\in\hat{x}\right\}. \end{aligned} \tag{7.48}$$

To prove the direct implication, we assume that $Q$ is lumpable. Then it follows immediately from Definition 7.40 and Eq. (7.48) that $\underline{\hat{Q}} = \hat{Q}$, with $\hat{Q}$ the – obviously unique – rate operator that satisfies Eq. (7.47). This proves that $\underline{\hat{Q}}$ is linear, and it is easy to see that then $\hat{Q}$ is the unique rate operator that dominates $\underline{\hat{Q}}$, so $\hat{\mathcal{Q}} = \{\hat{Q}\}$.

Next, we prove the converse implication, so we assume that $\underline{Q}$ is linear. Fix some $\hat{x}, \hat{y}$ in $\hat{\mathcal{X}}$. Then by Eq. (7.48) for $\hat{f} = \mathbb{I}_{\hat{y}}$ and $\hat{f} = -\mathbb{I}_{\hat{y}}$ and the linearity of $\underline{\hat{Q}}$,

$$\min\left\{\sum_{y\in\hat{y}} Q(x,y) \colon x\in\hat{x}\right\} = [\underline{\hat{Q}}\mathbb{I}_{\hat{y}}](\hat{x}) = -[\underline{\hat{Q}}(-\mathbb{I}_{\hat{y}})](\hat{x}) = \max\left\{\sum_{y\in\hat{y}} Q(x,y) \colon x\in\hat{x}\right\}.$$

Hence, $Q$ satisfies the condition in Definition 7.40 with $\hat{Q} = \underline{\hat{Q}}$, so $Q$ is lumpable. □

Second, we have the following corollary of the preceding result and Theorem 7.29361, which verifies the sufficiency of lumpability in (Burke et al., 1958, Theorem 4).

**Corollary 7.42.** *Consider a probability mass function $p_0$ in $\Sigma_{\mathcal{X}}$ and a rate operator $Q$ in $\mathfrak{Q}_{\mathcal{X}}$. If $Q$ is lumpable, then the lumped jump process $\hat{P}$ corresponding to $P_{p_0,Q}$ is uniquely defined and equal to $P_{\hat{p}_0,\hat{Q}}$, with $\hat{p}_0 := \hat{p}_{p_0}$ and $\hat{Q} = \underline{\hat{Q}}_Q$.*

*Proof.* Let $\mathscr{M} := \{p_0\}$ and $\mathscr{Q} := \{Q\}$, and recall from Eq. (3.66)$_{104}$ that

$$\mathbb{P}_{\mathscr{M},\mathscr{Q}} = \{P_{p_0,Q}\}.$$

Because $Q$ is lumpable by assumption, we know from Lemma 7.41 that the corresponding set $\hat{\mathscr{Q}}_Q = \hat{\mathscr{Q}}_{\{Q\}} = \hat{\mathscr{Q}}_{\mathscr{Q}}$ of lumped rate operators is the singleton $\{\hat{Q}\}$. Furthermore, $\hat{\mathscr{M}}_{\mathscr{M}} = \{\hat{p}_0\}$ by Eq. (7.26)$_{354}$. Hence, it follows from Eq. (3.66)$_{104}$ that

$$\mathbb{P}_{\hat{\mathscr{M}}_{\mathscr{M}},\hat{\mathscr{Q}}_{\mathscr{Q}}} = \{P_{\hat{p}_0,\hat{Q}}\}. \tag{7.49}$$

The statement follows from this equality because, by Theorem 7.29$_{361}$, any lumped jump process $\hat{P}$ corresponding to $P_{p_0,Q}$ belongs to $\mathbb{P}_{\hat{\mathscr{M}}_{\mathscr{M}},\hat{\mathscr{Q}}_{\mathscr{Q}}}$. □

Burke et al. (1958, Theorem 4) find that lumpability is also necessary for 'the' lumped jump process $\hat{P}$ corresponding to $P_{p_0,Q}$ to be Markovian; to obtain this result, they have to demand that 'the' lumped jump process is a homogeneous Markovian jump process for *every* initial probability mass function $p_0$ on $\mathscr{X}$ such that $p_0 > 0$. Given this extra assumption, this result would hold in our framework as well.

Particularly interesting about Corollary 7.42 is that it implies that we can specialise Theorem 7.33$_{363}$.

**Corollary 7.43.** *Consider an initial probability mass function $p_0$ in $\Sigma_{\mathscr{X}}$ and a lumpable rate operator $Q$ in $\mathfrak{Q}_{\mathscr{X}}$, and let $\hat{p}_0 := \hat{p}_{p_0}$ and $\hat{Q} := \underline{\hat{Q}}_Q$. Then for all $\{X_u = x_u\}$ in $\mathscr{H}$, $f$ in $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$ and $\hat{g}, \hat{h}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$ with $\hat{g}^{\uparrow\Omega} \leq f \leq \hat{h}^{\uparrow\Omega}$,*

$$E^{\mathrm{D}}_{\hat{p}_0,\hat{Q}}\big(\hat{g} \,\big|\, \hat{X}_u = \Lambda(x_u)\big) \leq E^{\mathrm{D}}_{p_0,Q}\big(f \,\big|\, X_u = x_u\big) E^{\mathrm{D}}_{\hat{p}_0,\hat{Q}}\big(\hat{h} \,\big|\, \hat{X}_u = \Lambda(x_u)\big).$$

*Proof.* Follows immediately from Theorem 7.33$_{363}$ and (Eq. (7.49) in the proof of) Corollary 7.42. □

## 7.3 Bounding limit expectations

In many practical applications, see for instance Chapter 8$_{403}$ further on or (Franceschinis et al., 1994; Buchholz, 2005; Ganguly et al., 2014; Troffaes et al., 2015), the system is modelled as a homogeneous and Markovian jump process $P_{p_0,Q}$, and one is interested in the expected long-term temporal average of $f(X_t)$, so in the expectation of

$$[\![f]\!]_{[0,r]} = \frac{1}{r}\int_0^r f(X_t)\,\mathrm{d}t$$

for large values of $r$. Recall from Section 6.4.4$_{305}$ that whenever the rate operator $Q$ is ergodic, this expected long-term temporal average is equal to the limit expectation $E_{\lim}(f)$: the value of the constant function $\lim_{t\to+\infty} e^{tQ}f$. This same limit expectation can also be used to directly determine $\lim_{t\to+\infty} E_{p_0,Q}(f(X_t))$ for some gamble $f$ on $\mathscr{X}$, because

$$\lim_{t\to+\infty} E_{p_0,Q}(f(X_t)) = \lim_{t\to+\infty} E_{p_0}(e^{tQ}f) = E_{\lim}(f).$$

For these reasons, methods to efficiently determine the limit expectation $E_{\lim}$ corresponding to an ergodic rate operator $Q$ are of tremendous practical interest.

There are plenty of methods available to determine this limit expectation; see for example (Stewart, 2009, Section 10) for an overview. However, it is well-known that these methods become intractable as the size of the state space increases (see Franceschinis et al., 1994; Buchholz, 2005). For this reason, we set out to obtain bounds on this limit expectation using the lumped lower rate operator $\underline{\hat{Q}}_Q$. Our hope is that given that the lumped state space is sufficiently small, these bounds *can* be tractably computed. If this is the case, then we can bound inferences that we could not tractably compute using the precise methods for the original model.

In Sections 7.3.1 and 7.3.2$_{376}$, we will adapt two well-known methods to determine the limit expectation precisely. More specifically, we will argue how these methods can be made computationally tractable using the lumped rate operator $\underline{\hat{Q}}_Q$ at the cost of imprecision, provided of course that $\underline{\hat{Q}}_Q \hat{f}$ can be determined much more efficiently than $Qf$. First, however, we start with some general theory concerning ergodic homogeneous Markovian jump processes.

### 7.3.1 Ergodicity, irreducibility and the limit expectation

When investigating the limit expectation, it is customary to focus on ergodic rate operators whose top class $\mathscr{X}_Q$ is equal to the entire state space; following Norris (1997, Sections 1.2 and 3.2), and as mentioned before in Section 6.4.4$_{305}$, we call such rate operators *irreducible*.

**Corollary 7.44.** *A rate operator $Q$ is irreducible if and only if*

$$\mathscr{X}_Q := \{x \in \mathscr{X} : (\forall y \in \mathscr{X})\ y \hookrightarrow x\} = \mathscr{X}.$$

*Proof.* This follows immediately from Proposition 4.33$_{193}$. □

One can get away with investigating irreducible rate operators because as far as the limit expectation is concerned, one can 'reduce' the state space to the top class $\mathscr{X}_Q$. The following result translates Seneta's (1981, Theorem 4.7) result from ergodic transition operators to ergodic rate operators.

**Proposition 7.45.** *Let $Q$ be an ergodic rate operator. Then the linear operator $Q' : \mathbb{G}(\mathscr{X}_Q) \to \mathbb{G}(\mathscr{X}_Q)$, defined by*

$$Q'(x, y) := Q(x, y) \quad \text{for all } x, y \in \mathscr{X}_Q,$$

*is an irreducible rate operator. Furthermore, for all $f$ in $\mathbb{G}(\mathscr{X})$, $E_{\lim}(f) = E'_{\lim}(f')$, where $E'_{\lim}$ denotes the limit expectation corresponding to $Q'$ and $f'$ is the restriction of $f$ to $\mathscr{X}_Q$.*

*Proof.* Our proof hinges on the following claim:

$$Q(x,y) = 0 \quad \text{for all } x \in \mathscr{X}_Q, y \in \mathscr{X} \setminus \mathscr{X}_Q \tag{7.50}$$

To verify this claim, we fix any such $x$ and $y$, and assume *ex absurdo* that $Q(x,y) \neq 0$. Note that $Q(x,y) = \big[Q\mathbb{I}_y\big](x) \geq 0$ because $x \neq y$ and $Q$ is a rate operator, so this implies that $Q(x,y) > 0$. Fix any arbitrary $z$ in $\mathscr{X}$. As $x$ is a state in the top class $\mathscr{X}_Q$, we know that $z \hookrightarrow x$, or equivalently, that there is a sequence $(x_0,\dots,x_n)$ with $n$ in $\mathbb{Z}_{\geq 0}$, $x_0 = z$ and $x_n = x$ such that $Q(x_{k-1},x_k) > 0$ for all $k$ in $\{1,\dots,n\}$. If we let $x_{n+1} := y$, then clearly $Q(x_{k-1},x_k) > 0$ for all $k$ in $\{1,\dots,n+1\}$, so $z \hookrightarrow y$. As $z$ was an arbitrary state, this implies that $y$ is a state in the top class $\mathscr{X}_Q$, which contradicts our initial assumption.

We use Eq. (7.50) to verify that $Q'$ is an irreducible rate operator. That $Q'$ is a rate operator – that is, that it has non-negative off-diagonal elements and rows that sum up to zero – follows almost immediately from Eq. (7.50) because $Q$ is a rate operator. Hence, we focus on verifying that $Q'$ is irreducible: we need to show that for any arbitrary $x,y$ in $\mathscr{X}_Q$, there is a sequence $x_0,\dots,x_n$ in $\mathscr{X}_Q$ with $n$ in $\mathbb{Z}_{\geq 0}$, $x_0 = x$ and $x_n = y$ such that $Q'(x_{k-1},x_k) > 0$ for all $k$ in $\{1,\dots,n\}$. Fix any arbitrary $x$ and $y$ in $\mathscr{X}_Q$. Because $y$ belongs to the top class $\mathscr{X}_Q$, there is a sequence $(x_0,\dots,x_n)$ in $\mathscr{X}$ with $n$ in $\mathbb{Z}_{\geq 0}$, $x_0 = x$ and $x_n = y$ such that $Q(x_{k-1},x_k) > 0$ for all $k$ in $\{1,\dots,n\}$. As $x_0$ is in the top class and $Q(x_0,x_1) > 0$, it follows from Eq. (7.50) that $x_1$ belongs to the top class $\mathscr{X}_Q$. Repeating this argument, we obtain that the entire sequence $(x_1,\dots,x_n)$ belongs to the top class $\mathscr{X}_Q$. Consequently, $Q'(x_{k-1},x_k) = Q(x_{k-1},x_k) > 0$ for all $k$ in $\{1,\dots,n\}$, as required.

Finally, we prove the second part of the statement. To this end, we fix some $g$ in $\mathbb{G}(\mathscr{X})$ and $g'$ in $\mathbb{G}(\mathscr{X}_Q)$ such that $g(y) = g'(y)$ for all $y$ in $\mathscr{X}_Q$. Observe that, by definition of $Q'$ and due to Eq. (7.50), for all $x$ in $\mathscr{X}_Q$,

$$[Qg](x) = \sum_{y\in\mathscr{X}} Q(x,y)g(y) = \sum_{y\in\mathscr{X}_Q} Q(x,y)g(y) = \sum_{y\in\mathscr{X}_Q} Q'(x,y)g'(y) = [Q'g'](x),$$

and therefore

$$(\forall \Delta \in \mathbb{R}_{\geq 0})\ \big[(I+\Delta Q)g\big](x) = \big[(I+\Delta Q')g'\big](x). \tag{7.51}$$

For all $\Delta$ in $\mathbb{R}_{\geq 0}$, $h := (I+\Delta Q)g$ and $h' := (I+\Delta Q')g'$ again satisfy $h(x) = h'(x)$ for all $x$ in $\mathscr{X}_Q$, so we can again use Eq. (7.51). Repeating the same argument, we see that for all $x$ in $\mathscr{X}_Q$,

$$(\forall n \in \mathbb{N})(\forall \Delta \in \mathbb{R}_{\geq 0})\ \big[(I+\Delta Q)g\big]^n(x) = \big[(I+\Delta Q')g'\big]^n(x). \tag{7.52}$$

Finally, it follows from Eq. (7.52) with $g = f$ and $g' = f'$ that for all for all $t$ in $\mathbb{R}_{\geq 0}$ and $x$ in $\mathscr{X}_Q$

$$[e^{tQ}f](x) = \lim_{n\to+\infty}\left[\left(I+\frac{t}{n}Q\right)^n f\right](x) = \lim_{n\to+\infty}\left[\left(I+\frac{t}{n}Q'\right)^n f'\right](x) = [e^{tQ'}f'](x),$$

as required. □

In order to use this result, one has to explicitly determine the top class $\mathscr{X}_Q$, as defined in Proposition 4.33193. If this top class $\mathscr{X}_Q$ can be obtained easily, then reducing the state space to this top class makes sense because it will speed up all methods to determine the limit expectation $E_{\lim}(f)$. However, this is not always the case, as it might occur that checking whether $Q$ is ergodic is straightforward,[2] while explicitly determining the top class is not. It is for this reason that we will consider general ergodic rate operators instead of only irreducible rate operators.

### 7.3.2 Bounding limit expectations with a linear program

The most common or basic method to determine the limit expectation $E_{\lim}(f)$ is to determine the limit expectation operator $E_{\lim}$ corresponding to $Q$ explicitly. Let $p_{\lim}$ be the probability mass function on $\mathscr{X}$ that is in one-to-one correspondence with the limit expectation $E_{\lim}$, so $p_{\lim}(x) = E_{\lim}(\mathbb{I}_x)$ for all $x$ in $\mathscr{X}$. Then for all $f$ in $\mathbb{G}(\mathscr{X})$,

$$E_{\lim}(f) = E_{p_{\lim}}(f) = \sum_{x\in\mathscr{X}} p_{\lim}(x) f(x) = \langle p_{\lim}, f\rangle. \tag{7.53}$$

It is well-known (see Tornambè, 1995, Theorem 4.12) that whenever the rate operator $Q$ is ergodic, this probability mass function $p_{\lim}$ is the unique probability mass function on $\mathscr{X}$ that satisfies the equilibrium condition

$$(\forall y \in \mathscr{X}) \sum_{x\in\mathscr{X}} p_{\lim}(x) Q(x,y) = 0. \tag{7.54}$$

As explained by (Stewart, 2009, Section 10.2), we can determine $p_{\lim}$ by solving the linear system of $|\mathscr{X}|$ equations in Eq. (7.54), but this is computationally infeasible for large state spaces. For this reason, we combine the equilibrium conditions for all states $x$ in the same lump $\Lambda^{-1}(\hat{x})$; additional manipulation of the resulting expressions yields the following result.

**Proposition 7.46.** *Consider an ergodic rate operator $Q$ in $\mathfrak{Q}_{\mathscr{X}}$, and fix some $f$ in $\mathbb{G}(\mathscr{X})$. Then for any $\hat{g}, \hat{h}$ in $\mathbb{G}(\hat{\mathscr{X}})$ such that $\hat{g}\circ\Lambda \leq f \leq \hat{h}\circ\Lambda$ and any subset $\mathscr{G}$ of $\mathbb{G}(\hat{\mathscr{X}})$,*

$$\inf\{\langle\hat{p},\hat{g}\rangle : \hat{p}\in\hat{\Sigma}_{\mathscr{G}}\} \leq E_{\lim}(f) \leq \sup\{\langle\hat{p},\hat{h}\rangle : \hat{p}\in\hat{\Sigma}_{\mathscr{G}}\},$$

*where we let*

$$\hat{\Sigma}_{\mathscr{G}} := \left\{\hat{p}\in\Sigma_{\hat{\mathscr{X}}} : (\forall \hat{f}\in\mathscr{G})\ \left\langle \hat{p}, \underline{\hat{Q}}_Q \hat{f}\right\rangle \leq 0\right\}.$$

*Proof.* Let $\underline{\hat{Q}} := \underline{\hat{Q}}_Q$. By Corollary $2.17_{23}$, $E_{\lim} = E_{p_{\lim}}$ is a coherent expectation on $\mathbb{G}(\mathscr{X})$, Note that $(-\hat{h})\circ\Lambda \leq -f$ because $f \leq \hat{f}\circ\Lambda$ by assumption. For this reason, and because $E_{\lim}$ is linear due to $(E2)_{22}$ and $(E3)_{22}$, the second inequality in the statement follows from the first one with $\hat{g} = -\hat{h}$ and $f$ replaced by $-f$, so we only need to prove the first inequality.

[2] If a rate operator $Q$ satisfies the first condition in Proposition $4.33_{193}$, then the second condition in Proposition $4.33_{193}$ is trivially satisfied – this is almost trivial, but we leave a formal proof as an exercise to the reader. Hence, it suffices to check whether $\mathscr{X}_Q$ is non-empty.

By assumption, $\hat{g} \circ \Lambda \leq f$; because the limit expectation $E_{\text{lim}}$ is monotone by (E6)$_{22}$, this implies that

$$E_{\text{lim}}(\hat{g} \circ \Lambda) \leq E_{\text{lim}}(f).$$

Let $p_{\text{lim}}$ be the unique probability mass function on $\mathscr{X}$ that corresponds to $E_{\text{lim}}$, so $p_{\text{lim}}(x) = E_{\text{lim}}(\mathbb{I}_x)$ for all $x$ in $\mathscr{X}$. Then by Eq. (7.53),

$$\begin{aligned} E_{\text{lim}}(\hat{g} \circ \Lambda) = \langle p_{\text{lim}}, \hat{g} \circ \Lambda \rangle &= \sum_{x \in \mathscr{X}} p_{\text{lim}}(x) \hat{g}(\Lambda(x)) \\ &= \sum_{\hat{x} \in \hat{\mathscr{X}}} \hat{g}(\hat{x}) \sum_{x \in \hat{x}} p_{\text{lim}}(x) = \sum_{\hat{x} \in \hat{\mathscr{X}}} \hat{g}(\hat{x}) \hat{p}_{\text{lim}}(\hat{x}) \\ &= \langle \hat{p}_{\text{lim}}, \hat{g} \rangle, \end{aligned}$$

where we let $\hat{p}_{\text{lim}} \coloneqq \hat{p}_{p_{\text{lim}}}$. We substitute this equality in the previously obtained inequality, to yield

$$\langle \hat{p}_{\text{lim}}, \hat{g} \rangle \leq E_{\text{lim}}(f).$$

Hence, to prove the first inequality in the statement, it suffices to verify that $\hat{p}_{\text{lim}}$ belongs to $\hat{\Sigma}_{\mathscr{G}}$.

By definition, the probability mass function $\hat{p}_{\text{lim}}$ belongs to $\hat{\Sigma}_{\mathscr{G}}$ if for all $\hat{f}$ in $\mathscr{G}$, $\langle \hat{p}_{\text{lim}}, \underline{\hat{Q}}_Q \hat{f} \rangle \leq 0$. Thus, we fix some $\hat{f}$ in $\mathscr{G}$. By Eq. (7.54),

$$\begin{aligned} \sum_{x \in \mathscr{X}} p_{\text{lim}}(x) \big[Q(\hat{f} \circ \Lambda)\big](x) &= \sum_{x \in \mathscr{X}} p_{\text{lim}}(x) \sum_{y \in \mathscr{X}} Q(x,y)[\hat{f} \circ \Lambda](y) \\ &= \sum_{y \in \mathscr{X}} [\hat{f} \circ \Lambda](y) \sum_{x \in \mathscr{X}} p_{\text{lim}}(x) Q(x,y) = 0. \end{aligned}$$

Because $Q(\hat{f} \circ \Lambda) \geq (\underline{\hat{Q}} \hat{f}) \circ \Lambda$ by Eq. (7.27)$_{355}$ and $p_{\text{lim}} \geq 0$ by (MF1)$_{23}$, it follows from the preceding equality that

$$\begin{aligned} 0 \geq \sum_{x \in \mathscr{X}} p_{\text{lim}}(x) \big[(\underline{\hat{Q}} \hat{f}) \circ \Lambda\big](x) = \sum_{\hat{x} \in \hat{\mathscr{X}}} [\underline{\hat{Q}} \hat{f}](\hat{x}) \sum_{x \in \hat{x}} p_{\text{lim}}(x) &= \sum_{\hat{x} \in \hat{\mathscr{X}}} \hat{p}_{\text{lim}}(\hat{x}) [\underline{\hat{Q}} \hat{f}](\hat{x}) \\ &= \langle \hat{p}_{\text{lim}}, \underline{\hat{Q}} \hat{f} \rangle. \end{aligned}$$

Because $\hat{f}$ was an arbitrary gamble in $\mathscr{G}$, this proves that $\hat{p}_{\text{lim}}$ belongs to $\hat{\Sigma}_{\mathscr{G}}$, as required. □

It might not be immediately obvious, but the optimisations in Proposition 7.46 can be solved through linear programming (see, for example, Bertsimas et al., 1997), at least in case $\mathscr{G}$ is a finite subset of $\mathbb{G}(\hat{\mathscr{X}})$. In this linear programming problem, the decision variables are the components of $\hat{p}$, the objective function is $\langle \hat{p}, \hat{g} \rangle$ or $\langle \hat{p}, \hat{h} \rangle$ and the feasible region is $\hat{\Sigma}_{\mathscr{G}}$. As $\hat{p}$ is a probability mass function on $\hat{\mathscr{X}}$, there are $|\hat{\mathscr{X}}|$ decision variables and already $|\hat{\mathscr{X}}| + 2$ inequality constraints: $\hat{p}(\hat{x}) \geq 0$ for all $\hat{x}$ in $\hat{\mathscr{X}}$, and two inequality constraints include the equality $\sum_{\hat{x} \in \hat{\mathscr{X}}} \hat{p}(\hat{x}) = 1$. Some $|\mathscr{G}|$ additional constraints are induced by the requirement that $\hat{p}$ is an element of $\hat{\Sigma}_{\mathscr{G}}$: for all $\hat{f}$ in $\mathscr{G}$, $\langle \hat{p}, \underline{\hat{Q}} \hat{f} \rangle \leq 0$, where we let $\underline{\hat{Q}} \coloneqq \underline{\hat{Q}}_Q$.

An obvious issue when applying the method in Proposition 7.46 is how to choose the subset $\mathscr{G}$ of $\mathbb{G}(\hat{\mathscr{X}})$. One idea is to consider the indicators $\mathbb{I}_{\hat{A}}$

of subsets $\hat{A}$ of the lumped state space $\hat{\mathscr{X}}$. Observe that, because $\underline{\hat{Q}}$ is a lower rate operator by Lemma 7.23$_{356}$, $\underline{\hat{Q}}\mathbb{I}_{\emptyset} = 0$ and $\underline{\hat{Q}}\mathbb{I}_{\hat{\mathscr{X}}} = 0$, so the condition $\langle \hat{p}, \underline{\hat{Q}}\mathbb{I}_{\hat{A}} \rangle \leq 0$ is always satisfied whenever $\hat{A}$ is equal to $\emptyset$ or $\hat{\mathscr{X}}$. Knowing this, one obvious choice for $\mathscr{G}$ is

$$\mathscr{G} = \{\mathbb{I}_{\hat{A}} \colon \hat{A} \in \mathcal{P}(\hat{\mathscr{X}}) \setminus \{\emptyset, \hat{\mathscr{X}}\}\},$$

which leads to a linear program with $|\hat{\mathscr{X}}| + 2^{|\hat{\mathscr{X}}|}$ inequality constraints. As the number of constraints scales exponentially with the number of lumps, this is computationally intractable for large lumped state spaces $\hat{\mathscr{X}}$. An alternative choice for $\mathscr{G}$ that is not exponential in the number of lumps is to consider the indicators of all singletons:

$$\mathscr{G} = \{\mathbb{I}_{\hat{x}} \colon \hat{x} \in \hat{\mathscr{X}}\}.$$

This choice results in $2|\hat{\mathscr{X}}| + 2$ inequality constraints for the linear programming problem, which is certainly tractable. Depending on the way this method is implemented, it can make sense to also add the indicators of the complements of the singletons to $\mathscr{G}$: if the function – or array – $\mathbb{I}_{\hat{x}}$ is explicitly generated, then it is trivial to generate $-\mathbb{I}_{\hat{x}}$ at the same time, and this is relevant because by (LR6)$_{111}$,

$$\underline{\hat{Q}}\mathbb{I}_{\mathscr{X} \setminus \{\hat{x}\}} = \underline{\hat{Q}}(1 - \mathbb{I}_{\hat{x}}) = \underline{\hat{Q}}(-\mathbb{I}_{\hat{x}}).$$

If we add not only the indicators of the singletons but also the indicators of their complements, then the linear programming problem in Proposition 7.46$_{376}$ has $3|\hat{\mathscr{X}}| + 2$ inequality constraints. Of course, there is a trade-off between tractability and tightness: the obtained bound will be tighter for a more constrained feasible set; we leave a thorough assessment of this trade-off for future research.

**Queuing Network Example 7.47.** Franceschinis et al. (1994) compute bounds on the limit expectation of the population at $S_0$, so on $E_{\lim}(f)$ with $f$ as defined in Eq. (7.42)$_{365}$, for $K = 4$ parallel servers and $N = 5$ customers. These parameters yield a state space $\mathscr{X}$ with 126 states and a lumped state space $\hat{\mathscr{X}}$ with 18 states. For the service time distributions, they use the parameters $\mu = 1$, $\underline{\lambda} = 1$ and $\overline{\lambda} = 1.01$. In Table 7.6$_{\curvearrowright}$, we report the bounds on this limit expectation that we obtain with Proposition 7.46$_{376}$, taking $\hat{g} = \hat{f} = \hat{h}$ – with $\hat{f}$ as defined in Eq. (7.43)$_{365}$ – and two of the previously mentioned choices for $\mathscr{G}$: the set $\mathscr{G}_1$ of the indicators of the singletons and the set $\mathscr{G}_2$ of the indicators of the singletons and their complements; we do not use the indicators of the power set because we believe that using $|\hat{\mathscr{X}}| + 2^{|\hat{\mathscr{X}}|} = 262\,162$ inequality constraints for 18 variables is a bit excessive For all cases, our bounds are tighter than those of Franceschinis et al. (1994),[3] and the bounds are noticeably tighter if we add the complements of the singletons. 𝔗𝔅

[3] Franceschinis et al. (1994) report the limit expectation of $N - n_0$ instead of $n_0$, but we have transformed their bounds to correspond to our setting.

**Table 7.6** Comparison of the bounds on the limit expectation of the population at $S_0$ obtained by Franceschinis et al. (1994, Table 2) to those obtained with Propositions 7.46$_{376}$ and 7.51$_{382}$. Model parameters: $K = 4$, $N = 5$, $\mu = 1$, $\underline{\lambda} = 1$, $\overline{\lambda} = 1.01$. Computation parameters: $\mathscr{G}_1$ consists of the indicators of all the singletons and $\mathscr{G}_2$ consists of the indicators of the singletons and their complements.

| | | | Proposition 7.46$_{376}$ | | | | Proposition 7.51$_{382}$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\mathscr{G}_1$ | | $\mathscr{G}_2$ | | $\Delta = 1.8/\lVert Q\rVert_{\mathrm{op}}$ | | $\Delta = 0.9/\lVert Q\rVert_{\mathrm{op}}$ | |
| Exact | Lower | Upper | Lower | Upper | Lower | Upper | Lower | Upper | Lower | Upper |
| 3.727 | 3.614 | 3.749 | 3.616 | 3.749 | 3.711 | 3.741 | 3.720 | 3.734 | 3.720 | 3.734 |

### 7.3.3 Bounding limit expectations iteratively

Next, we look at a method to determine the limit expectation $E_{\mathrm{lim}}(f)$ that is based on a well-known link between homogeneous Markovian jump processes and homogeneous Markov chains. Crucial to this method is that for all $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\lVert Q\rVert_{\mathrm{op}} < 2$, $(I + \Delta Q)$ is a transition operator due to Lemma 3.29$_{82}$ and $(I + \Delta Q)$ is ergodic if $Q$ is ergodic due to Theorems 4.28$_{191}$ and 4.36$_{194}$. Interestingly, in this case, the limit expectation $E_{\mathrm{lim}}$ corresponding to the rate operator $Q$ is equal to the 'limit expectation' corresponding to the transition operator $(I + \Delta Q)$ (see Stewart, 2009, Sections 10.1.1 and 10.3.1); the following result establishes this link in the form that we will need it.

**Lemma 7.48.** *Consider an ergodic rate operator $Q$. Then for all $f$ in $\mathbb{G}(\mathscr{X})$, $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\lVert Q\rVert_{\mathrm{op}} < 2$ and $n$ in $\mathbb{N}$,*

$$\min(I + \Delta Q)^n f \leq E_{\mathrm{lim}}(f) \leq \max(I + \Delta Q)^n f.$$

*Furthermore, the lower and upper bounds in this expression become monotonously tighter with increasing $n$, and they converge to $E_{\mathrm{lim}}(f)$ as $n$ recedes to $+\infty$.*

*Proof.* Let $T := (I + \Delta Q)$. Because $\Delta\lVert Q\rVert_{\mathrm{op}} < 2$ by assumption, $T$ is a transition operator due to Lemma 3.29$_{82}$. Hence, by (LT4)$_{108}$,

$$\min g \leq \min Tg \leq \max Tg \leq \max g \quad \text{for all } g \in \mathbb{G}(\mathscr{X}),$$

and in particular

$$\min T^n f \leq \min T^{n+1} f \leq \max T^{n+1} f \leq T^n f.$$

This confirms that the bounds in the statement become monotonously tighter for increasing $n$.

Next, we observe that due to Theorems 4.28$_{191}$ and 4.36$_{194}$, the transition operator $T$ is ergodic, so $\lim_{n\to+\infty} T^n f$ converges to a constant function. By (Tornambè, 1995, Definition 4.7 and Theorem 4.5), the value of this constant function is

$E_p(f) = \langle p, f\rangle$, where $p$ is the unique probability mass function on $\mathscr{X}$ that satisfies the equilibrium condition

$$(\forall y \in \mathscr{X})\ p(y) = \sum_{x\in\mathscr{X}} p(x)T(x,y).$$

Thus, by definition of $T$,

$$(\forall y \in \mathscr{X})\ p(y) = \sum_{x\in\mathscr{X}} p(x)\big(I(x,y) + \Delta Q(x,y)\big) = p(y) + \Delta \sum_{x\in\mathscr{X}} p(x)Q(x,y).$$

Clearly, the probability mass function $p$ satisfies Eq. (7.54)$_{376}$, so it must be equal to the limit probability mass function $p_{\text{lim}}$ corresponding to $E_{\text{lim}}$. Consequently, $E_{\text{lim}}(f) = E_p(f)$, so the sequence $(T^n f)_{n\in\mathbb{N}}$ converges to a constant function with value $E_{\text{lim}}(f)$, and this proves the inequalities in the statement. □

The step size $\Delta$ in Lemma 7.48 can be any positive real number such that $\Delta\|Q\|_{\text{op}} < 2$. Empirically, we observe that the convergence of the bounds is faster – in the sense that we need smaller $n$ – for larger values of $\Delta$. This method is also computationally intractable for large state spaces $\mathscr{X}$: to keep track of the approximation $(I+\Delta Q)^n f$, we need to store an $|\mathscr{X}|$-dimensional array of floating point numbers. One way to make the computations tractable is to 'replace' the rate operator $Q$ with the lumped lower rate operator $\hat{\underline{Q}}_Q$. This follows almost immediately from the following more general result.

**Proposition 7.49.** *Consider a non-empty and bounded set $\mathscr{Q}$ of ergodic rate operators on $\mathbb{G}(\mathscr{X})$, and fix some $f$ in $\mathbb{G}(\mathscr{X})$. Then for all $\hat{g}, \hat{h}$ in $\mathbb{G}(\hat{\mathscr{X}})$ such that $\hat{g}\circ\Lambda \le f \le \hat{h}\circ\Lambda$, $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\big\|\underline{Q}_{\mathscr{Q}}\big\|_{\text{op}} \le 2$ and $n$ in $\mathbb{N}$,*

$$\begin{aligned}\min\big(I + \Delta\hat{\underline{Q}}_{\mathscr{Q}}\big)^n \hat{g} &\le \inf\{E^Q_{\text{lim}}(f) : Q \in \mathscr{Q}\}\\ &\le \sup\{E^Q_{\text{lim}}(f) : Q \in \mathscr{Q}\} \le -\min\big(I + \Delta\hat{\underline{Q}}_{\mathscr{Q}}\big)^n(-\hat{h}),\end{aligned}$$

*where for every $Q$ in $\mathscr{Q}$, we let $E^Q_{\text{lim}}$ denote the corresponding limit expectation. Moreover, for fixed $\Delta$, these bounds become monotonously tighter with increasing $n$.*

In our proof for this result, we use the following intermediary result.

**Lemma 7.50.** *For any non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$,*

$$\big\|\hat{\underline{Q}}_{\mathscr{Q}}\big\|_{\text{op}} \le \big\|\underline{Q}_{\mathscr{Q}}\big\|_{\text{op}}.$$

*Proof.* Let $\hat{\underline{Q}} := \hat{\underline{Q}}_{\mathscr{Q}}$ and $\underline{Q} := \underline{Q}_{\mathscr{Q}}$. By (LR7)$_{111}$ and Eq. (7.27)$_{355}$,

$$\begin{aligned}\|\hat{\underline{Q}}\|_{\text{op}} &= 2\max\{-[\hat{\underline{Q}}\mathbb{I}_{\hat{x}}](\hat{x}) : \hat{x} \in \hat{\mathscr{X}}\} = 2\max\Big\{-\min\{[\underline{Q}(\mathbb{I}_{\hat{x}}\circ\Lambda)](x) : x \in \hat{x}\} : \hat{x}\in\hat{\mathscr{X}}\Big\}\\ &= 2\max\Big\{\max\{-[\underline{Q}(\mathbb{I}_{\hat{x}}\circ\Lambda)](x) : x\in\hat{x}\} : \hat{x}\in\hat{\mathscr{X}}\Big\}.\end{aligned}$$

For all $\hat{x}$ in $\hat{\mathscr{X}}$, $\mathbb{I}_{\hat{x}} \circ \Lambda = \sum_{y\in\hat{x}} \mathbb{I}_y$, so for all $x$ in $\Lambda^{-1}(\hat{x})$,

$$-[\underline{Q}(\mathbb{I}_{\hat{x}} \circ \Lambda)](x) \leq -\sum_{y\in\hat{x}} [\underline{Q}\mathbb{I}_y](x) \leq -[\underline{Q}\mathbb{I}_x](x) \leq \frac{1}{2}\|\underline{Q}\|_{\mathrm{op}},$$

where we used (LR4)$_{109}$, (LR2)$_{109}$ and (LR7)$_{111}$. We infer from the preceding that

$$\|\underline{\hat{Q}}\|_{\mathrm{op}} \leq 2\max\left\{\max\left\{\frac{1}{2}\|\underline{Q}\|_{\mathrm{op}} : x \in \hat{x}\right\} : \hat{x} \in \hat{\mathscr{X}}\right\} = \|\underline{Q}\|_{\mathrm{op}},$$

as required. □

*Proof of Proposition 7.49*. Let $\underline{\hat{Q}} := \underline{\hat{Q}}_Q$. Fix some $Q$ in $\mathscr{Q}$. Because $\hat{g} \circ \Lambda \leq f$ by assumption and the limit expectation $E_{\mathrm{lim}}$ is monotonous by (LE6)$_{30}$,

$$E^Q_{\mathrm{lim}}(f) \geq E^Q_{\mathrm{lim}}(\hat{g} \circ \Lambda) \geq \min(I + \Delta Q)^n(\hat{g} \circ \Lambda), \tag{7.55}$$

where for the second inequality we used Lemma 7.48$_{379}$.

To prove the first inequality in the statement, we observe that for all $\hat{f}$ in $\mathbb{G}(\hat{\mathscr{X}})$, $Q(\hat{f} \circ \Lambda) \geq (\underline{\hat{Q}}\hat{f}) \circ \Lambda$ due to Eq. (7.27)$_{355}$, and therefore

$$(I + \Delta Q)(\hat{f} \circ \Lambda) = \hat{f} \circ \Lambda + \Delta Q(\hat{f} \circ \Lambda) \geq \hat{f} \circ \Lambda + \Delta(\underline{\hat{Q}}\hat{f}) \circ \Lambda = \big((I + \Delta\underline{\hat{Q}})\hat{f}\big) \circ \Lambda. \tag{7.56}$$

Because $\Delta\|\underline{Q}_{\mathscr{Q}}\|_{\mathrm{op}} \leq 2$ by assumption and $\|Q\|_{\mathrm{op}} \leq \|\underline{Q}_{\mathscr{Q}}\|_{\mathrm{op}}$ by (LR7)$_{111}$, $(I + \Delta Q)$ is a transition operator due to Lemma 3.29$_{82}$. Hence, it follows from Eq. (7.56) with $\hat{f} = \hat{g}$, (LT11)$_{108}$ and (LT6)$_{108}$ that

$$(I + \Delta Q)^n(\hat{g} \circ \Lambda) = (I + \Delta Q)^{n-1}(I + \Delta Q)(\hat{g} \circ \Lambda) \geq (I + \Delta Q)^{n-1}\Big(\big((I + \Delta\underline{\hat{Q}})\hat{g}\big) \circ \Lambda\Big).$$

We apply the same trick $(n-1)$-times more, to yield

$$(I + \Delta Q)^n(\hat{g} \circ \Lambda) \geq \big((I + \Delta\underline{\hat{Q}})^n\hat{g}\big) \circ \Lambda,$$

so clearly

$$\min(I + \Delta Q)^n(\hat{g} \circ \Lambda) \geq \min(I + \Delta\underline{\hat{Q}})^n\hat{g}.$$

Together with Eq. (7.55), this inequality proves the inequality in the statement.

In order to verify the second part of the statement, we set out to show that the sequence $(\min(I + \Delta\underline{\hat{Q}})^m\hat{g})_{m\in\mathbb{N}}$ is non-decreasing. Recall from Lemma 7.50 that $\|\underline{\hat{Q}}\|_{\mathrm{op}} \leq \|\underline{Q}\|_{\mathrm{op}}$; because $\Delta\|\underline{Q}\|_{\mathrm{op}} \leq 2$ by assumption, this implies that $\Delta\|\underline{\hat{Q}}\|_{\mathrm{op}} \leq 2$. For this reason, it follows from Lemmas 3.72$_{112}$ and 7.23$_{356}$ that $(I + \Delta\underline{\hat{Q}})$ is a lower transition operator. Consequently, it follows immediately from (repeated application of) (LT4)$_{108}$ that, for all $m$ in $\mathbb{N}$,

$$\min(I + \Delta\underline{\hat{Q}})^m\hat{g} \leq \min(I + \Delta\underline{\hat{Q}})(I + \Delta\underline{\hat{Q}})^m\hat{g} = \min(I + \Delta\underline{\hat{Q}})^{m+1}\hat{g}.$$

Thus, the sequence $(\min(I + \Delta\underline{\hat{Q}})^m\hat{g})_{m\in\mathbb{N}}$ is non-decreasing. □

**Proposition 7.51.** *Consider an ergodic rate operator $Q$ in $\mathfrak{Q}_{\mathscr{X}}$, and fix some $f$ in $\mathbb{G}(\mathscr{X})$. Then for all $\hat{g}, \hat{h}$ in $\mathbb{G}(\hat{\mathscr{X}})$ such that $\hat{g} \circ \Lambda \leq f \leq \hat{h} \circ \Lambda$, $\Delta$ in $\mathbb{R}_{>0}$ with $\Delta \|Q\|_{\mathrm{op}} \leq 2$ and $n$ in $\mathbb{N}$,*

$$\min\big(I + \Delta \underline{\hat{Q}}_Q\big)^n \hat{g} \leq E_{\mathrm{lim}}(f) \leq -\min\big(I + \Delta \underline{\hat{Q}}_Q\big)^n (-\hat{h}).$$

*Moreover, for fixed $\Delta$, the lower and upper bounds in this expression become monotonously tighter with increasing $n$.*

*Proof.* Follows immediately from Proposition 7.49$_{380}$. □

Empirically, we observe that larger step sizes $\Delta$ in Proposition 7.51 result in faster convergence, in the sense that fewer iterations are required before convergence. The influence of the step size $\Delta$ on the tightness of the bounds is something that we have not yet properly investigated. Our limited experiments suggest that a smaller step size $\Delta$ results in tighter bounds, although after some threshold – that depends on the specific model being used and that can be rather large – the tightness no longer seems to change any more.

**Queuing Network Example 7.52.** Let us return to the setting in Queuing Network Example 7.47$_{378}$. In Table 7.6$_{379}$, we report the bounds on the limit expectation of the population at $\mathrm{S}_0$ that we compute with Proposition 7.51, taking $\hat{g} = \hat{f} = \hat{h}$ – with $\hat{f}$ as defined in Eq. (7.43)$_{365}$ – and $\Delta$ equal to $1.8/\|Q\|_{\mathrm{op}}$ and $0.9/\|Q\|_{\mathrm{op}}$. Our bounds are tighter than those of Franceschinis et al. (1994), and also tighter than the bounds that we computed with the linear programming method in Proposition 7.46$_{376}$. Observe that the bounds are the same – at least up to the reported precision – for both choices of step sizes; in this case, halving the step size results in a doubling of the number of iterations required to reach empirical convergence. ℭ

### 7.3.4 Numerical assessment

Buchholz (2005) improved on the method of Franceschinis et al. (1994), so it is only fair that we should compare our two methods to his. He considers a slightly changed version of the system that we have been studying: instead of assuming that the service time of $\mathrm{S}_0$ is exponentially distributed, he assumes an Erlang-2 distribution with mean service time $1/\mu$. We can still model this system as a homogeneous Markovian jump process: it is as if we replace $\mathrm{S}_0$ with two servers $\mathrm{S}_{0,1}$ and $\mathrm{S}_{0,2}$ in series (see Stewart, 2009, Section 12.1). Given that the second server $\mathrm{S}_{0,2}$ is empty, a customer in the first server $\mathrm{S}_{0,1}$ can transition to the second one with rate $2\mu$; if the second server $\mathrm{S}_{0,2}$ contains a customer, then there are no transitions from $\mathrm{S}_{0,1}$ to $\mathrm{S}_{0,2}$, and there is a transition from $\mathrm{S}_{0,2}$ to $\mathrm{S}_k$ with rate ${}^{2\mu}/_{K}$. Clearly, this requires us to replace the component $n_0$ by two components in both state descriptions. We will not go into details here, but it suffices to understand that the rate operator $Q$ is similar to the one in Eq. (7.1)$_{342}$, and that the corresponding lumped lower rate operator $\underline{\hat{Q}}_Q$ is similar to the one in Eq. (7.32)$_{358}$.

Buchholz (2005, Figure 3) considers several combinations of the number of parallel servers $K$ and the number of customers $N$. For the service time distributions, he uses the parameters $\mu = 5$, $\underline{\lambda} = 1$ and $\overline{\lambda} = \underline{\lambda} + \epsilon$, with $\epsilon$ equal to 0.1 or 0.01. He reports bounds on the limit expectation of two inferences: the population at $\mathrm{S}_0$ and the 'throughput', which assumes the value $\mu$ if there are customers at $\mathrm{S}_0$ and 0 otherwise. The limit expectation of the population that we obtain does not lie in the intervals reported by Buchholz (2005, Fig. 3), and we have not managed to clarify whether or not this is due to an error on our part;[4] interestingly, we have observed that for all but one value, the limit expectation of 'half of the population' lies in the reported bounds. Since this prevents a proper comparison, we have chosen not to report any bounds on the population.

In Table 7.6$_{379}$, we report the bounds on the limit expectation of the throughput that we have computed with the methods in Propositions 7.46$_{376}$ and 7.51↶, and we compare these with Buchholz's (2005). First and foremost,

**Table 7.7** Bounds on the limit expectation of the throughput. Computation parameters: $\mathscr{G}_1$ consists of the indicators of the singletons and $\mathscr{G}_2$ consists of the indicators of the singletons and their complements.

| | | | | | | | Proposition 7.46$_{376}$ | | | | Proposition 7.51↶ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | (Buchholz, 2005) | | $\mathscr{G}_1$ | | $\mathscr{G}_2$ | | $\Delta = 1.8/\lVert Q\rVert_{\mathrm{op}}$ | | $\Delta = 0.9/\lVert Q\rVert_{\mathrm{op}}$ | |
| $K$ | $N$ | $\lvert\mathscr{X}\rvert$ | $\lvert\hat{\mathscr{X}}\rvert$ | Exact | Lower | Upper | Lower | Upper | Lower | Upper | Lower | Upper | Lower | Upper |
| | | | | | | | $\epsilon = 0.1$ | | | | | | | |
| 4 | 6 | 336 | 45 | 2.611 | 2.509 | 2.73 | 1.928 | 3.181 | 2.401 | 2.82 | 2.522 | 2.714 | 2.522 | 2.713 |
| 4 | 8 | 825 | 91 | 2.892 | 2.784 | 3.028 | 1.982 | 3.633 | 2.582 | 3.171 | 2.799 | 3.004 | 2.800 | 3.004 |
| 4 | 10 | 1716 | 165 | 3.090 | 2.973 | 3.239 | 2.002 | 3.961 | 2.67 | 3.435 | 2.995 | 3.208 | 2.995 | 3.208 |
| 6 | 8 | 4719 | 108 | 3.486 | 3.365 | 3.624 | 2.068 | 4.188 | 3.069 | 3.846 | 3.372 | 3.614 | 3.372 | 3.613 |
| 6 | 10 | 13013 | 215 | 3.802 | 3.675 | 3.984 | 2.083 | 4.515 | 3.191 | 4.244 | 3.689 | 3.931 | 3.689 | 3.931 |
| 8 | 10 | 68068 | 232 | 4.202 | 4.087 | 4.327 | 2.111 | 4.736 | 3.542 | 4.595 | 4.093 | 4.320 | 4.093 | 4.320 |
| | | | | | | | $\epsilon = 0.01$ | | | | | | | |
| 4 | 6 | 336 | 45 | 2.520 | 2.509 | 2.532 | 2.428 | 2.591 | 2.500 | 2.541 | 2.511 | 2.530 | 2.511 | 2.530 |
| 4 | 8 | 825 | 91 | 2.793 | 2.780 | 2.806 | 2.655 | 2.894 | 2.764 | 2.821 | 2.783 | 2.803 | 2.783 | 2.803 |
| 4 | 10 | 1716 | 165 | 2.984 | 2.971 | 2.998 | 2.802 | 3.116 | 2.948 | 3.020 | 2.974 | 2.996 | 2.974 | 2.996 |
| 6 | 8 | 4719 | 108 | 3.378 | 3.365 | 3.392 | 3.121 | 3.488 | 3.340 | 3.416 | 3.366 | 3.391 | 3.366 | 3.391 |
| 6 | 10 | 13013 | 215 | 3.689 | 3.675 | 3.704 | 3.336 | 3.821 | 3.639 | 3.738 | 3.677 | 3.702 | 3.677 | 3.702 |
| 8 | 10 | 68068 | 232 | 4.100 | 4.087 | 4.113 | 3.609 | 4.213 | 4.047 | 4.151 | 4.088 | 4.113 | 4.088 | 4.113 |

we observe that the bounds obtained with the iterative method in Proposition 7.51↶ are tighter than Buchholz's (2005), which in turn are tighter than those obtained with the linear programming method in Proposition 7.46$_{376}$; hence, as was also the case in Queuing Network Example 7.52↶, our iterative method outperforms the linear programming method. Second, we observe that halving the step size $\Delta$ in the iterative method does increase the tightness of the bounds, be it only marginally. Adding the complements to the collec-

[4] Buchholz was so kind to point out some errors in my earlier implementation of his model, but he never responded to my inquiry about the throughput.

tion $\mathscr{G}$ in the linear programming method clearly results in tighter bounds, as was also the case in Queuing Network Example $7.47_{378}$.

Buchholz (2005, Section 5.3) mentions that he assumes 'service rates between $\underline{\lambda}$ and $\underline{\lambda}+\epsilon$', but it is unclear to us if he additionally assumes that half of the servers have rate $\underline{\lambda}$ and the other half of the servers have rate $\underline{\lambda}+\epsilon$. If he does not make this additional assumption, his bounds hold for the more general setting that we considered in Queuing Network Example $7.36_{368}$ – so the setting where we do not require that half of the rates should be equal to $\underline{\lambda}$ and the other half should be equal to $\overline{\lambda}$. Note that we can deal with this case too using Proposition $7.49_{380}$, but we will not pursue this here.

We have chosen to limit the scenarios for our numerical experiments in this chapter to those scenarios that were also considered by Franceschinis et al. (1994) and Buchholz (2005). Our reason for this is two-fold. First, this allow us to compare our bounds with those obtained by Franceschinis et al. (1994) and Buchholz (2005) without needing to implement their methods ourselves. Second, the state space $\mathscr{X}$ for these scenarios is not too large, so this allows us to compare our bounds with the exact result. However, because the limit expectations in our numerical experiments can all be computed in a tractable manner, these scenarios do not correspond to the intended range of applications. Fear not, a more realistic test of our methods is just around the corner in Chapter $8_{403}$!

## 7.A Daniell expectation with respect to a lumped jump process

In this appendix, we gather the proofs for those results in Section $7.1.3_{350}$ regarding the countable additivity of $\hat{P}$ and the Daniell expectation with respect to $\hat{P}$. First, we prove Lemma $7.15_{351}$.

**Lemma 7.15.** *Consider a jump process $P$ with state space $\mathscr{X}$. If $P$ is countably additive, then any corresponding lumped jump process $\hat{P}$ is countably additive too.*

In our proof, we use the following intermediary result. This result will also come in handy in our proof for Lemma $7.16_{352}$ further on, hence the separate statement.

**Lemma 7.53.** *Consider a sequence of time points $u$ in $\mathscr{U}$, and a sequence $(\hat{f}_n)_{n\in\mathbb{N}}$ of $\hat{\mathscr{F}}_u$ simple variables that converges point-wise to the limit $\hat{f} := \text{p-w}\lim_{n\to+\infty} \hat{f}_n$. Then $(\hat{f}_n^{\uparrow\Omega})_{n\in\mathbb{N}}$ is a sequence of $\mathscr{F}_u$ simple variables that converges point-wise to $\hat{f}^{\uparrow\Omega}$. Furthermore, if $(\hat{f}_n)_{n\in\mathbb{N}}$ is non-decreasing, non-increasing, uniformly bounded above and/or uniformly bounded below, then so is $(\hat{f}_n^{\uparrow\Omega})_{n\in\mathbb{N}}$.*

*Proof.* For all $n$ in $\mathbb{N}$, $\hat{f}_n$ is $\hat{\mathscr{F}}_u$-simple by assumption; hence, $\hat{f}_n^{\uparrow\Omega}$ is $\mathscr{F}_u$-simple due to Lemma 7.13~350~. This proves that $(\hat{f}_n^{\uparrow\Omega})_{n\in\mathbb{N}}$ is a sequence of $\mathscr{F}_u$ simple variables. Next, we prove that this sequence converges point-wise to $\hat{f}^{\uparrow\Omega}$. Recall from Eq. (7.23)~350~ that for all $\omega$ in $\Omega$ and $n$ in $\mathbb{N}$,

$$\hat{f}^{\uparrow\Omega}(\omega) = \hat{f}(\Lambda\circ\omega) \quad\text{and}\quad \hat{f}_n^{\uparrow\Omega}(\omega) = \hat{f}_n(\Lambda\circ\omega).$$

Because $\Lambda\circ\omega$ belongs to $\hat{\Omega}$ for all $\omega$ in $\Omega$ by Lemma 7.5~343~ and because $(\hat{f}_n)_{n\in\mathbb{N}}$ converges point-wise to $\hat{f}$ by assumption, we infer from this that the sequence $(\hat{f}_n^{\uparrow\Omega})_{n\in\mathbb{N}}$ converges point-wise to $\hat{f}^{\uparrow\Omega}$. Finally, it also follows from these equalities that the non-decreasing, non-increasing, uniformly bounded below and/or uniformly bounded above nature of $(\hat{f}_n)_{n\in\mathbb{N}}$ carries over to the corresponding sequence of cylindrical extensions. □

*Proof of Lemma 7.15~351~.* By Definition 5.15~229~, we need to show that, for all $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$, the probability charge $\hat{P}(\bullet\,|\,\hat{X}_u = \hat{x}_u)$ on $\hat{\mathscr{F}}_u$ is countably additive. To this end, we fix some $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$. We show that the probability charge $\hat{P}(\bullet\,|\,\hat{X}_u = \hat{x}_u)$ is countably additive by verifying that it satisfies Definition 5.4~221~ (iii). To this end, we fix any non-increasing sequence $(\hat{f}_n)_{n\in\mathbb{N}}$ of $\hat{\mathscr{F}}_u$-simple variables that converges point-wise to 0. Then by Lemma 7.53~↶~, $(\hat{f}_n^{\uparrow\Omega})_{n\in\mathbb{N}}$ is a sequence of $\mathscr{F}_u$-simple variables that converges point-wise to 0. For all $x_u$ in $\Lambda^{-1}(\hat{x}_u)$, $P(\bullet\,|\,X_u = x_u)$ is countably additive by assumption, so it follows from Definition 5.4~221~ (iii) for $P(\bullet\,|\,X_u = x_u)$ that

$$\lim_{n\to+\infty} E_P(\hat{f}_n^{\uparrow\Omega}\,|\,X_u = x_u) = 0. \tag{7.57}$$

Recall from Lemma 7.14~351~ that, for all $n$ in $\mathbb{N}$,

$$\begin{aligned}\min\{E_P(\hat{f}_n^{\uparrow\Omega}\,|\,X_u = \hat{x}_u) x_u \in \hat{x}_u\} &\le E_{\hat{P}}(\hat{f}_n\,|\,\hat{X}_u = \hat{x}_u)\\ &\le \max\{E_P(\hat{f}_n^{\uparrow\Omega}\,|\,X_u = \hat{x}_u) x_u \in \hat{x}_u\}.\end{aligned} \tag{7.58}$$

Because $\Lambda^{-1}(\hat{x}_x)$ is finite, the limit of the minimum (maximum) is the minimum (maximum) of the limits $\lim_{n\to+\infty} E_P(\hat{f}_n^{\uparrow\Omega}\,|\,X_u = x_u)$, so it follows from Eqs. (7.57) and (7.58) that

$$\lim_{n\to+\infty} E_{\hat{P}}(\hat{f}_n\,|\,\hat{X}_u = \hat{x}_u) = 0.$$

This equality holds for any non-increasing sequence $(\hat{f}_n)_{n\in\mathbb{N}}$ that converges point-wise to 0, so $E_{\hat{P}}(\bullet\,|\,\hat{X}_u = \hat{x}_u)$ is countably additive due to Definition 5.4~221~ (iii). □

Next, we prove Lemma 7.16~352~ with the help of Lemma 7.53~↶~.

**Lemma 7.16.** *Consider a sequence of time points $u$ in $\mathscr{U}$. Then for any limit variable $\hat{f}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$, the corresponding cylindrical extension $\hat{f}^{\uparrow\Omega}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$.*

*Proof.* By definition of $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$, there is a sequence $(\hat{f}_n)_{n\in\mathbb{N}}$ of $\hat{\mathscr{F}}_u$-simple variables that is uniformly bounded below (or uniformly bounded above) and that converges point-wise to $\hat{f}$. Due to Lemma 7.53~↶~, the corresponding sequence $(\hat{f}_n^{\uparrow\Omega})_{n\in\mathbb{N}}$ of cylindrical extensions is a sequence of $\mathscr{F}_u$-simple variables that is uniformly bounded below (or uniformly bounded above) and converges point-wise to $\hat{f}^{\uparrow\Omega}$. Consequently, $\hat{f}^{\uparrow\Omega}$ belongs to $\overline{\mathbb{V}}_{\lim}(\mathscr{F}_u)$. □

Finally, we prove Lemma 7.17$_{352}$.

**Lemma 7.17.** *Consider a countably additive jump process $P$ with state space $\mathscr{X}$ and any corresponding lumped jump process $\hat{P}$. Then for all $\{\hat{X}_u = \hat{x}_u\}$ in $\hat{\mathscr{H}}$ and $\hat{f}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}}_u)$,*

$$\min\left\{E_P^{\mathrm{D}}(\hat{f}^{\uparrow\Omega} \,|\, X_u = x_u) \colon x_u \in \hat{x}_u\right\} \leq E_{\hat{P}}^{\mathrm{D}}(\hat{f} \,|\, \hat{X}_u = \hat{x}_u) \leq \max\left\{E_P^{\mathrm{D}}(\hat{f}^{\uparrow\Omega} \,|\, X_u = x_u) \colon x_u \in \hat{x}_u\right\}.$$

In our proof, and also in that for Theorem 7.32$_{363}$ further on, we need the following obvious intermediary result.

**Lemma 7.54.** *Consider a sequence of time points $u$ in $\mathscr{U}$. Then for any $\hat{\mathscr{F}}_u$-over variable $\hat{f}$, its cylindrical extension $\hat{f}^{\uparrow\Omega}$ is an $\mathscr{F}_u$-over variable; similarly, for any $\hat{\mathscr{F}}_u$-under variable $\hat{f}$, its cylindrical extension $\hat{f}^{\uparrow\Omega}$ is an $\mathscr{F}_u$-under variable. Hence,*

$$\left\{\hat{f}^{\uparrow\Omega} \colon \hat{f} \in \overline{\mathbb{V}}^{\mathrm{o}}(\hat{\mathscr{F}}_u)\right\} \subseteq \overline{\mathbb{V}}^{\mathrm{o}}(\mathscr{F}_u) \quad \text{and} \quad \left\{\hat{f}^{\uparrow\Omega} \colon \hat{f} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u)\right\} \subseteq \overline{\mathbb{V}}_{\mathrm{u}}(\mathscr{F}_u). \tag{7.59}$$

*Proof.* Fix any $\hat{\mathscr{F}}_u$-over variable $\hat{f}$. Then by definition, there is a non-decreasing sequence $(\hat{f}_n)_{n\in\mathbb{N}}$ of $\hat{\mathscr{F}}_u$-simple variables that converges point-wise to $\hat{f}$. By Lemma 7.53$_{384}$, $(\hat{f}_n^{\uparrow\Omega})_{n\in\mathbb{N}}$ is a non-decreasing sequence of $\mathscr{F}_u$-simple variables that converges point-wise to $\hat{f}^{\uparrow\Omega}$, so $\hat{f}^{\uparrow\Omega}$ is an $\mathscr{F}_u$-over variable, as required.

The statement for the $\hat{\mathscr{F}}_u$-under variable $\hat{f}$ follows from the preceding because $-\hat{f}$ is then an $\hat{\mathscr{F}}_u$-over variable and $-\hat{f}^{\uparrow\Omega}$ an $\mathscr{F}_u$-over variable. Clearly, Eq. (7.59) follows immediately from the first part of the statement. □

*Proof of Lemma 7.17$_{352}$.* For all $x_u$ in $\mathscr{X}_u$, we let $E_{x_u}^{\mathrm{mc}}$ denote the extension of $E_P(\bullet \,|\, X_u = x_u)$ to $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathscr{F}_u)$ as defined by Eq. (5.6)$_{223}$: for all $g$ in $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathscr{F}_u)$,

$$E_{x_u}^{\mathrm{mc}}(g) = \lim_{n\to+\infty} E_P(g_n \,|\, X_u = x_u), \tag{7.60}$$

where $(g_n)_{n\in\mathbb{N}}$ is any monotone sequence of $\mathscr{F}_u$ simple variables that converges point-wise to $g$. Similarly, $\hat{E}^{\mathrm{mc}}$ denotes the extension of $E_{\hat{P}}(\bullet \,|\, \hat{X}_u = \hat{x}_u)$ to $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\hat{\mathscr{F}}_u)$: for all $\hat{g}$ in $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\hat{\mathscr{F}}_u)$,

$$\hat{E}^{\mathrm{mc}}(\hat{g}) = \lim_{n\to+\infty} E_{\hat{P}}(\hat{g}_n \,|\, \hat{X}_u = \hat{x}_u), \tag{7.61}$$

where $(\hat{g}_n)_{n\in\mathbb{N}}$ is any monotone sequence of $\hat{\mathscr{F}}_u$ simple variables that converges point-wise to $\hat{g}$.

Due to Lemmas 7.14$_{351}$ and 7.53$_{384}$, these extensions are related. Fix any $\hat{g}$ in $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathscr{F}_u)$. Then by definition of $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathscr{F}_u)$, there is a monotone sequence $(\hat{g}_n)_{n\in\mathbb{N}}$ of $\hat{\mathscr{F}}_u$ simple variables that converges point-wise to $\hat{g}$. Furthermore, $(\hat{g}_n^{\uparrow\Omega})_{n\in\mathbb{N}}$ is a monotone sequence of $\mathscr{F}_u$-simple variables that converges point-wise to $\hat{g}^{\uparrow\Omega}$ by Lemma 7.53$_{384}$, and $\hat{g}^{\uparrow\Omega}$ belongs to $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathscr{F}_u)$ by Lemma 7.54. By Lemma 7.14$_{351}$, for all $n$ in $\mathbb{N}$,

$$\min\left\{E_P(\hat{g}_n^{\uparrow\Omega} \,|\, X_u = x_u) \colon x_u \in \hat{x}_u\right\} \leq E_{\hat{P}}(\hat{g}_n \,|\, \hat{X}_u = \hat{x}_u) \leq \max\left\{E_P(\hat{g}_n^{\uparrow\Omega} \,|\, X_u = x_u) \colon x_u \in \hat{x}_u\right\}.$$

We take the limit for $n$ going to $+\infty$ and use Eqs. (7.60) and (7.61), to yield

$$\min\{E^{\mathrm{mc}}_{x_u}(\hat{g}^{\uparrow\Omega})\colon x_u \in \hat{x}_u\} \leq \hat{E}^{\mathrm{mc}}(\hat{g}) \leq \max\{E^{\mathrm{mc}}_{x_u}(\hat{g}^{\uparrow\Omega})\colon x_u \in \hat{x}_u\}. \tag{7.62}$$

We are almost done. By definition of $E^{\mathrm{D}}_{P}$, for all $x_u$ in $\mathscr{X}_u$,

$$E^{\mathrm{D}}_{P}(\hat{f}^{\uparrow\Omega} \mid X_u = x_u) = \sup\{E^{\mathrm{mc}}_{x_u}(g)\colon g \in \overline{\mathbb{V}}_{\mathrm{u}}(\mathscr{F}_u), g \leq \hat{f}^{\uparrow\Omega}\} \tag{7.63}$$

$$= \inf\{E^{\mathrm{mc}}_{x_u}(g)\colon g \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathscr{F}_u), g \geq \hat{f}^{\uparrow\Omega}\}; \tag{7.64}$$

similarly, by definition of $E^{\mathrm{D}}_{\hat{P}}$,

$$E^{\mathrm{D}}_{\hat{P}}(\hat{f} \mid \hat{X}_u = \hat{x}_u) = \sup\{\hat{E}^{\mathrm{mc}}(\hat{g})\colon \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u), \hat{g} \leq \hat{f}\} \tag{7.65}$$

$$= \inf\{\hat{E}^{\mathrm{mc}}(\hat{g})\colon \hat{g} \in \overline{\mathbb{V}}^{\mathrm{o}}(\hat{\mathscr{F}}_u), \hat{g} \geq \hat{f}\}. \tag{7.66}$$

By Eqs. (7.65) and (7.62),

$$\begin{aligned}
E^{\mathrm{D}}_{\hat{P}}(\hat{f} \mid \hat{X}_u = \hat{x}_u) &= \sup\{\hat{E}^{\mathrm{mc}}(\hat{g})\colon \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u), \hat{g} \leq \hat{f}\} \\
&\leq \sup\Big\{\max\{E^{\mathrm{mc}}_{x_u}(\hat{g}^{\uparrow\Omega})\colon x_u \in \hat{x}_u\}\colon \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u), \hat{g} \leq \hat{f}\Big\} \\
&= \max\Big\{\sup\{E^{\mathrm{mc}}_{x_u}(\hat{g}^{\uparrow\Omega})\colon \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u), \hat{g} \leq \hat{f}\}\colon x_u \in \hat{x}_u\Big\} \\
&= \max\Big\{\sup\{E^{\mathrm{mc}}_{x_u}(\hat{g}^{\uparrow\Omega})\colon \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u), \hat{g}^{\uparrow\Omega} \leq \hat{f}^{\uparrow\Omega}\}\colon x_u \in \hat{x}_u\Big\},
\end{aligned}$$

where for the last equality we used that $\hat{g} \leq \hat{f}$ if and only if $\hat{g}^{\uparrow\Omega} \leq \hat{f}^{\uparrow\Omega}$. By Lemma 7.54, $\{\hat{g}^{\uparrow\Omega}\colon \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u)\} \subseteq \overline{\mathbb{V}}_{\mathrm{u}}(\mathscr{F}_u)$; consequently,

$$\begin{aligned}
E^{\mathrm{D}}_{\hat{P}}(\hat{f} \mid \hat{X}_u = \hat{x}_u) &\leq \max\Big\{\sup\{E^{\mathrm{mc}}_{x_u}(g)\colon g \in \overline{\mathbb{V}}_{\mathrm{u}}(\mathscr{F}_u), g \leq \hat{f}^{\uparrow\Omega}\}\colon x_u \in \hat{x}_u\Big\} \\
&= \max\{E^{\mathrm{D}}_{P}(\hat{f}^{\uparrow\Omega} \mid X_u = x_u)\colon x_u \in \hat{x}_u\},
\end{aligned}$$

where for the final equality we used Eq. (7.63). Similarly, it follows from Eqs. (7.66) and (7.62), Lemma 7.54 and Eq. (7.64) that

$$E^{\mathrm{D}}_{\hat{P}}(\hat{f} \mid \hat{X}_u = \hat{x}_u) \geq \min\{E^{\mathrm{D}}_{P}(\hat{f}^{\uparrow\Omega} \mid X_u = x_u)\colon x_u \in \hat{x}_u\}. \qquad \square$$

## 7.B Proof of Theorem 7.32

Our proof for Theorem 7.32363 is rather long. What makes it so lengthy is the construction of the jump process $\hat{P}$: we obtain this jump process not by lumping $P$ directly, but by lumping a jump process $P'$ that is derived from $P$. By construction, this derived jump process $P'$ need not be consistent with $\mathscr{M}$ and $\mathscr{Q}$. However, we do construct it in such a way that it is consistent with $\mathscr{M}^\star$ and $\mathscr{Q}^\star$, where $\mathscr{M}^\star$ is a superset of $\mathscr{M}$ such that $\hat{\mathscr{M}}_{\mathscr{M}} = \hat{\mathscr{M}} = \hat{\mathscr{M}}_{\mathscr{M}^\star}$ and $\mathscr{Q}^\star$ is a superset of $\mathscr{Q}$ such that $\hat{\mathscr{Q}}_{\mathscr{Q}} = \hat{\mathscr{Q}} = \hat{\mathscr{Q}}_{\mathscr{Q}^\star}$. We determine these supersets in Appendices 7.B.1 and 7.B.2389, and then get around to proving Theorem 7.32363 in Appendix 7.B.3395.

### 7.B.1 The set $\mathscr{M}^\star$ of permuted probability mass functions

First, we determine a superset $\mathscr{M}^\star$ of $\mathscr{M}$ such that $\hat{\mathscr{M}}_{\mathscr{M}} = \hat{\mathscr{M}}_{\mathscr{M}^\star}$. Our starting point is the following observation: for any probability mass function $p$ in $\mathscr{M}$, we can redistribute the mass of any lump $\Lambda^{-1}(\hat{x})$ over the states in this lump, and the resulting probability mass function will still lump to $\hat{p}_p$. To facilitate this redistribution of the masses, we use permutations of the states. More precisely, we will permute the states in the state space $\mathscr{X}$ lump per lump; formally, we use bijective maps $\pi$ from $\mathscr{X}$ to $\mathscr{X}$ that map every state $x$ in $\mathscr{X}$ to a state $\pi(x)$ in the lump $\Lambda^{-1}(\Lambda(x))$.

**Definition 7.55.** A permutation $\pi\colon \mathscr{X} \to \mathscr{X}$ is a bijective map, and it is called *lump-preserving* whenever $\Lambda = \Lambda \circ \pi$. We denote the set of all lump-preserving permutations by $\Pi_{\mathscr{X}}$.

Consider a lump-preserving permutation $\pi$ in $\Pi_{\mathscr{X}}$. Following De Cooman et al. (2012, Section 4.1), we lift this permutation $\pi$ to a permutation $\pi^{\mathrm{t}}$ on $\mathbb{G}(\mathscr{X})$: for all $f$ in $\mathbb{G}(\mathscr{X})$, we let $\pi^{\mathrm{t}} f := f \circ \pi$. We also extend the domain of $\pi$ in the same way as we did for the lumping map $\Lambda$: for all $u$ in $\mathscr{U}$, $x_u$ in $\mathscr{X}_u$ and $B \subseteq \mathscr{X}_u$, $\pi(x_u) := (\pi(x_t))_{t\in u}$ and $\pi(B) := \{\pi(z_u)\colon z_u \in B\}$. The following obvious properties of lump-preserving permutations will come in handy in the remainder.

**Lemma 7.56.** *For all $\pi$ in $\Pi_{\mathscr{X}}$, $\hat{x}$ in $\hat{\mathscr{X}}$ and $y$ in $\mathscr{X}$,*

(i) *the identity map* $\mathrm{id}\colon \mathscr{X} \to \mathscr{X}\colon x \mapsto x$ *is a lump-preserving permutation;*

(ii) $\pi^{-1}$ *is a lump-preserving permutation;*

(iii) $\{\pi(x)\colon x \in \hat{x}\} = \Lambda^{-1}(\hat{x})$*;*

(iv) $\{\pi^{-1}(y)\colon \pi \in \Pi_{\mathscr{X}}\} = \Lambda^{-1}(\Lambda(y))$*;*

(v) $\{\pi(x)\colon x \in \mathscr{X}\} = \mathscr{X}$*;*

(vi) $\{\pi^{\mathrm{t}} f\colon f \in \mathbb{G}(\mathscr{X}), \|f\| = 1\} = \{f \in \mathbb{G}(\mathscr{X})\colon \|f\| = 1\}$*;*

(vii) $\pi^{\mathrm{t}} \mathbb{I}_y = \mathbb{I}_{\pi^{-1}(y)}$.

*Proof.* Trivial. □

Permuting the masses of a probability mass function $p$ on $\mathscr{X}$ over the states in a lump corresponds to assigning to each state $x$ in $\mathscr{X}$ the mass of the state $\pi(x)$, with $\pi(x)$ a state in the lump $\Lambda^{-1}(\Lambda(x))$. That is, the 'permuted' probability mass function is $\pi^{\mathrm{t}} p = p \circ \pi$, with $\pi$ a lump-preserving permutation on $\mathscr{X}$.

**Lemma 7.57.** *For any probability mass function $p$ on $\mathscr{X}$ and any lump-preserving permutation $\pi$ on $\mathscr{X}$, $\pi^{\mathrm{t}} p$ is a probability mass function on $\mathscr{X}$.*

*Proof.* Observe that $\pi^{\mathrm{t}} p$ is a real-valued function on $\mathscr{X}$. By Definition 2.16₂₃, we need to show that $\pi^{\mathrm{t}} p$ satisfies (MF1)$_{23}$ and (MF2)$_{23}$. For all $x$ in $\mathscr{X}$, it follows from (MF1)$_{23}$ for $p$ that

$$[\pi^{\mathrm{t}} p](x) = [p \circ \pi](x) = p\big(\pi(x)\big) \geq 0,$$

so $\pi^{\mathrm{t}} p$ satisfies (MF1)$_{23}$. Furthermore, because $\pi$ is a permutation on $\mathscr{X}$ by assumption, it follows from (MF2)$_{23}$ for $p$ and Lemma 7.56 (iii) that

$$\sum_{x \in \mathscr{X}} [\pi^{\mathrm{t}} p](x) = \sum_{x \in \mathscr{X}} [p \circ \pi](x) = \sum_{x \in \mathscr{X}} p\big(\pi(x)\big) = \sum_{x \in \mathscr{X}} p(x) = 1,$$

so $\pi^{\mathrm{t}} p$ satisfies (MF2)$_{23}$. □

Furthermore, it is easy to see that the permuted probability mass function $\pi^{\mathrm{t}} p$ corresponds to the same lumped probability mass function as the original probability mass function $p$.

**Lemma 7.58.** *For any probability mass function $p$ on $\mathscr{X}$ and any lump-preserving permutation $\pi$ on $\mathscr{X}$, $\hat{p}_p = \hat{p}_{\pi^{\mathrm{t}} p}$.*

*Proof.* Fix some $\hat{x}$ in $\hat{\mathscr{X}}$. Then because $\pi$ is a lump-preserving permutation by assumption,

$$\sum_{x \in \hat{x}} p(x) = \sum_{x \in \hat{x}} p\big(\pi(x)\big) = \sum_{x \in \hat{x}} [p \circ \pi](x) = \sum_{x \in \hat{x}} [\pi^{\mathrm{t}} p](x),$$

where the first equality holds due to Lemma 7.56 (iii). The equality in the statement follows immediately from the preceding equality and Eq. (7.25)$_{354}$. □

For any set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$, we denote the corresponding set of permuted probability mass functions by

$$\mathscr{M}^{\star} := \{\pi^{\mathrm{t}} p \colon p \in \mathscr{M}, \pi \in \Pi_{\mathscr{X}}\}. \tag{7.67}$$

It follows immediately from Lemma 7.56 (i) that $\mathscr{M}^{\star}$ is a superset of $\mathscr{M}$, and from Lemma 7.58 that $\hat{\mathscr{M}}_{\mathscr{M}} = \hat{\mathscr{M}}_{\mathscr{M}^{\star}}$.

**Lemma 7.59.** *Consider any non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$. Then $\mathscr{M} \subseteq \mathscr{M}^{\star}$ and $\hat{\mathscr{M}}_{\mathscr{M}} = \hat{\mathscr{M}}_{\mathscr{M}^{\star}}$, with $\mathscr{M}^{\star}$ as defined by Eq.* (7.67).

*Proof.* Follows immediately from Lemma 7.56 (i) and Lemma 7.58. □

### 7.B.2 Enlarging the set $\mathscr{Q}$ of rate operators

Next, we determine a superset $\mathscr{Q}^{\star}$ of $\mathscr{Q}$ such that $\hat{\mathscr{Q}}_{\mathscr{Q}} = \hat{\mathscr{Q}}_{\mathscr{Q}^{\star}}$. More exactly, we set out to determine such a superset $\mathscr{Q}^{\star}$ that has 'all the nice properties': one that is bounded and convex and that has separately specified rows; the reason for this will become clear in Proposition 7.69$_{395}$. We go about this by first 'closing' the set $\mathscr{Q}$ of rate operators under taking lump-preserving permutations, then taking the lower envelope of this set and finally considering the set of all rate operators that dominate this lower envelope.

First, let us define what we mean by permuting a rate operator with a lump-preserving permutation. Fix some lump-preserving permutation $\pi$ in $\Pi_{\mathscr{X}}$. Then we lift the permutation $\pi$ to the set $\mathfrak{Q}_{\mathscr{X}}$ of rate operators on $\mathbb{G}(\mathscr{X})$ as follows: for any rate operator $Q$ in $\mathfrak{Q}_{\mathscr{X}}$, we let $\pi Q$ be the operator on $\mathbb{G}(\mathscr{X})$ that maps any $f$ in $\mathbb{G}(\mathscr{X})$ to the gamble $(\pi Q)f$ on $\mathscr{X}$ defined by

$$\big[(\pi Q)f\big](x) := \big[Q(\pi^{\mathrm{t}} f)\big]\big(\pi^{-1}(x)\big) = \big[Q(f \circ \pi)\big]\big(\pi^{-1}(x)\big) \quad \text{for all } x \in \mathscr{X}. \tag{7.68}$$

The definition above does not immediately provide much intuition about what is going on. As the permuted rate operator $\pi Q$ is a linear operator – see Lemma 7.60 further on – it helps to look at the components of $\pi Q$: for all $x, y$ in $\mathscr{X}$,

$$\begin{aligned}[\pi Q](x,y) = [(\pi Q)\mathbb{I}_y](x) &= \big[Q(\mathbb{I}_y \circ \pi)\big]\big(\pi^{-1}(x)\big) = \big[Q\mathbb{I}_{\pi^{-1}(y)}\big]\big(\pi^{-1}(x)\big) \\ &= Q\big(\pi^{-1}(x), \pi^{-1}(y)\big),\end{aligned}$$

where for the penultimate equality we used Lemma $7.56_{388}$ (vii). Thus, we can think of $\pi Q$ as permuting the rows and the columns of the matrix representation of $Q$. The diagonal components remain diagonal components and the row sums are left unchanged, so $\pi Q$ is again a rate operator.

**Lemma 7.60.** *For any rate operator $Q$ in $\mathfrak{Q}_{\mathscr{X}}$ and any lump-preserving permutation $\pi$ in $\Pi_{\mathscr{X}}$, the corresponding operator $\pi Q$ as defined by Eq.* (7.68) *is a rate operator.*

*Proof.* By Definition $3.27_{81}$, we need to verify that $\pi Q$ satisfies $(\mathrm{R1})_{81}$–$(\mathrm{R4})_{81}$. To check $(\mathrm{R1})_{81}$, we fix any $x$ in $\mathscr{X}$. Observe that $\mathbb{I}_{\mathscr{X}} \circ \pi = \mathbb{I}_{\mathscr{X}}$, so

$$\big[(\pi Q)\mathbb{I}_{\mathscr{X}}\big](x) = \big[Q\big(\mathbb{I}_{\mathscr{X}} \circ \pi\big)\big]\big(\pi^{-1}(x)\big) = \big[Q\mathbb{I}_{\mathscr{X}}\big]\big(\pi^{-1}(x)\big) = 0,$$

where the final equality follows from $(\mathrm{R1})_{81}$ for $Q$.

Next, we check $(\mathrm{R2})_{81}$. Fix any $x, y$ in $\mathscr{X}$ such that $x \neq y$. By Lemma $7.56_{388}$ (vii), $\mathbb{I}_y \circ \pi = \mathbb{I}_{\pi^{-1}(y)}$ and therefore

$$\big[(\pi Q)\mathbb{I}_y\big](x) = \big[Q\big(\mathbb{I}_y \circ \pi\big)\big]\big(\pi^{-1}(x)\big) = \big[Q\mathbb{I}_{\pi^{-1}(y)}\big]\big(\pi^{-1}(x)\big) = 0,$$

where the last equality follows from $(\mathrm{R4})_{81}$ for $Q$ because $\pi$ is a permutation and therefore $\pi^{-1}(x) \neq \pi^{-1}(y)$.

To check $(\mathrm{R3})_{81}$, we fix any $x$ in $\mathscr{X}$, $f$ in $\mathbb{G}(\mathscr{X})$ and $\mu$ in $\mathbb{R}_{\geq 0}$. Observe that $(\mu f) \circ \pi = \mu(f \circ \pi)$, so

$$\begin{aligned}\big[(\pi Q)(\mu f)\big](x) &= \big[Q\big((\mu f) \circ \pi\big)\big]\big(\pi^{-1}(x)\big) = \big[Q\big(\mu(f \circ \pi)\big)\big]\big(\pi^{-1}(x)\big) \\ &= \mu\big[Q(f \circ \pi)\big]\big(\pi^{-1}(x)\big) \\ &= \mu\big[(\pi Q)f\big](x),\end{aligned}$$

where the third equality holds due to $(\mathrm{R3})_{81}$ for $Q$ because $\mu$ is non-negative by assumption.

Finally, we verify that $\pi Q$ satisfies (R4)$_{81}$. To this end, we fix any $x$ in $\mathscr{X}$ and $f, g$ in $\mathbb{G}(\mathscr{X})$. Observe that $(f+g)\circ\pi = f\circ\pi + g\circ\pi$, so

$$\begin{aligned}\big[(\pi Q)(f+g)\big](x) &= \big[Q\big((f+g)\circ\pi\big)\big]\big(\pi^{-1}(x)\big) = \big[Q(f\circ\pi + g\circ\pi)\big]\big(\pi^{-1}(x)\big)\\ &= \big[Q(f\circ\pi)\big]\big(\pi^{-1}(x)\big) + \big[Q(g\circ\pi)\big]\big(\pi^{-1}(x)\big)\\ &= \big[(\pi Q)f\big](x) + \big[(\pi Q)g\big](x),\end{aligned}$$

where for the third equality we used (R4)$_{81}$ for $Q$. □

As can be expected, permuting the permuted rate operator $\pi Q$ with the inverse permutation $\pi^{-1}$ yields the original rate operator $Q$.

**Lemma 7.61.** *For all $Q$ in $\mathfrak{Q}_{\mathscr{X}}$ and $\pi$ in $\Pi_{\mathscr{X}}$, $\pi^{-1}(\pi Q) = Q$.*

*Proof.* Let $Q' := \pi Q$. Then by Eq. (7.68)$_{\curvearrowleft}$, for all $f$ in $\mathbb{G}(\mathscr{X})$ and $x$ in $\mathscr{X}$,

$$\big[Q' f\big](x)\big[(\pi Q)f\big](x) = \big[Q(f\circ\pi)\big]\big(\pi^{-1}(x)\big). \tag{7.69}$$

In particular, because $Q'$ is a rate operator – so a linear operator – due to Lemma 7.60$_{\curvearrowleft}$, we find that for all $x, y$ in $\mathscr{X}$,

$$\begin{aligned}Q'(x,y) = \big[Q'\mathbb{I}_y\big](x) = \big[Q(\mathbb{I}_y\circ\pi)\big]\big(\pi^{-1}(x)\big), = \big[Q\mathbb{I}_{\pi^{-1}(y)}\big]\big(\pi^{-1}(x)\big)\\ = Q\big(\pi^{-1}(x), \pi^{-1}(y)\big).\end{aligned} \tag{7.70}$$

Fix any $f$ in $\mathbb{G}(\mathscr{X})$ and $x$ in $\mathscr{X}$. Then by Eq. (7.68)$_{\curvearrowleft}$,

$$\big[\pi^{-1}(\pi Q)f\big](x) = \big[\pi^{-1}Q' f\big](x) = \big[Q'(f\circ\pi^{-1})\big]\big(\pi(x)\big),$$

where we used that $(\pi^{-1})^{-1} = \pi$. Again, $\pi^{-1}Q'$ is a rate operator by Lemma 7.60$_{\curvearrowleft}$, and therefore

$$\big[\pi^{-1}(\pi Q)f\big](x) = \sum_{y\in\mathscr{X}} Q'(\pi(x), y)[f\circ\pi^{-1}](y) = \sum_{y\in\mathscr{X}} Q'(\pi(x), y) f\big(\pi^{-1}(y)\big).$$

We substitute Eq. (7.70), to yield

$$\begin{aligned}\big[\pi^{-1}(\pi Q)f\big](x) &= \sum_{y\in\mathscr{X}} Q\Big(\pi\big(\pi^{-1}\big(\pi(x)\big)\big), \pi^{-1}(y)\Big) f\big(\pi^{-1}(y)\big)\\ &= \sum_{y\in\mathscr{X}} Q\big(x, \pi^{-1}(y)\big) f\big(\pi^{-1}(y)\big)\\ &= \sum_{y\in\mathscr{X}} Q\big(x, y\big) f\big(y\big)\\ &= \big[Qf\big](y),\end{aligned}$$

where for the second equality we used that $\pi^{-1}\circ\pi = \mathrm{id}$ and for the third equality we used Lemma 7.56$_{388}$ (v). Because this equality holds for arbitrary $x$ in $\mathscr{X}$ and $f$ in $\mathbb{G}(\mathscr{X})$, we conclude that $\pi^{-1}(\pi Q) = Q$. □

For any set $\mathcal{Q}$ of rate operators on $\mathbb{G}(\mathcal{X})$, we denote its *closure under lump-preserving permutations*[5] by

$$\mathrm{pm}(\mathcal{Q}) := \{\pi Q \colon Q \in \mathcal{Q}, \pi \in \Pi_{\mathcal{X}}\}. \tag{7.71}$$

This 'closure' $\mathrm{pm}(\mathcal{Q})$ is a set of rate operators due to Lemma 7.60$_{390}$, and it trivially includes the original set $\mathcal{Q}$.

**Lemma 7.62.** *Consider a subset $\mathcal{Q}$ of $\mathfrak{Q}_{\mathcal{X}}$. Then the corresponding set $\mathrm{pm}(\mathcal{Q})$ as defined by Eq.* (7.71) *is a subset of $\mathfrak{Q}_{\mathcal{X}}$ that includes $\mathcal{Q}$.*

*Proof.* That $\mathrm{pm}(\mathcal{Q})$ is a subset of $\mathfrak{Q}_{\mathcal{X}}$ follows immediately fro Lemma 7.60$_{390}$. Because the identity permutation id belongs to $\Pi_{\mathcal{X}}$ due to Lemma 7.56$_{388}$ (i), it follows immediately from Eq. (7.71) that $\mathrm{pm}(\mathcal{Q})$ includes $\mathcal{Q}$. □

Another important property of the closure $\mathrm{pm}(\mathcal{Q})$ of $\mathcal{Q}$ under lump-preserving permutations is that this closure $\mathrm{pm}(\mathcal{Q})$ is bounded whenever $\mathcal{Q}$ is bounded.

**Lemma 7.63.** *Consider a non-empty subset $\mathcal{Q}$ of $\mathfrak{Q}_{\mathcal{X}}$. Then*

$$\|\mathcal{Q}\|_{\mathrm{op}} = \|\mathrm{pm}(\mathcal{Q})\|_{\mathrm{op}},$$

*so* $\mathrm{pm}(\mathcal{Q})$ *is bounded if and only if $\mathcal{Q}$ is bounded.*

The statement in Lemma 7.63 follows immediately from the following intermediary results; we also need these result in the proof of Lemma 7.70$_{395}$ further on, so we provide separate statements.

**Lemma 7.64.** *For all $Q_1, Q_2$ in $\mathfrak{Q}_{\mathcal{X}}$ and $\pi$ in $\Pi_{\mathcal{X}}$,*

$$\|\pi Q_1 - \pi Q_2\|_{\mathrm{op}} = \|Q_1 - Q_2\|_{\mathrm{op}}.$$

*Proof.* Let $\mathbb{G}_1 := \{f \in \mathbb{G} \colon \|f\| = 1\}$. Then by definition of $\|\bullet\|_{\mathrm{op}}$, $\|\bullet\|$, $\pi Q_1$ and $\pi Q_2$,

$$\begin{aligned}
&\|\pi Q_1 - \pi Q_2\|_{\mathrm{op}} \\
&\quad = \sup\{\|(\pi Q_1) f - (\pi Q_2) f\| \colon f \in \mathbb{G}_1\} \\
&\quad = \sup\{\max\{|[(\pi Q_1) f](x) - [(\pi Q_2) f](x)| \colon x \in \mathcal{X}\} \colon f \in \mathbb{G}_1\} \\
&\quad = \sup\{\max\{|[Q_1(f \circ \pi)](\pi^{-1}(x)) - [Q_2(f \circ \pi)](\pi^{-1}(x))| \colon x \in \mathcal{X}\} \colon f \in \mathbb{G}_1\}.
\end{aligned}$$

[5] Our usage of the term closure is justified because for all $Q$ in $\mathrm{pm}(\mathcal{Q})$ and $\sigma$ in $\Pi_{\mathcal{X}}$, $\sigma Q$ belongs to $\mathrm{pm}(\mathcal{Q})$. This is true because for all $Q$ in $\mathcal{Q}$ and $\pi, \sigma$ in $\Pi_{\mathcal{X}}$, $\sigma \circ \pi$ is a lump-preserving permutation and $\sigma(\pi Q) = (\sigma \circ \pi) Q$. Because we do not need this property in the remainder, we do not provide a formal proof.

Recall from Lemma 7.56₃₈₈ that $\{f \circ \pi \colon f \in \mathbb{G}_1\} = \mathbb{G}_1$ and $\{\pi^{-1}(x) \colon x \in \mathscr{X}\} = \mathscr{X}$. Consequently,

$$\begin{aligned}\|\pi Q_1 - \pi Q_2\|_{\mathrm{op}} &= \sup\{\max\{\big|[Q_1 f](\pi^{-1}(x)) - [Q_2 f](\pi^{-1}(x))\big| \colon x \in \mathscr{X}\} \colon f \in \mathbb{G}_1\}\\ &= \sup\{\max\{\big|[Q_1 f](x) - [Q_2 f](x)\big| \colon x \in \mathscr{X}\} \colon f \in \mathbb{G}_1\}\\ &= \sup\{\|Q_1 f - Q_2 f\| \colon f \in \mathbb{G}_1\}\\ &= \|Q_1 - Q_2\|_{\mathrm{op}},\end{aligned}$$

where last two equalities follow from the definition of $\|\bullet\|$ and $\|\bullet\|_{\mathrm{op}}$, respectively. □

**Corollary 7.65.** *For all $Q$ in $\mathfrak{Q}_{\mathscr{X}}$ and $\pi$ in $\Pi_{\mathscr{X}}$, $\|\pi Q\|_{\mathrm{op}} = \|Q\|_{\mathrm{op}}$.*

*Proof.* The zero operator $0 \colon \mathbb{G}(\mathscr{X}) \to \mathbb{G}(\mathscr{X}) \colon f \mapsto 0$ is a rate operator. Obviously, $\pi 0 = 0$, $\|Q\|_{\mathrm{op}} = \|Q - 0\|_{\mathrm{op}}$ and $\|\pi Q\|_{\mathrm{op}} = \|\pi Q - \pi 0\|_{\mathrm{op}}$. Hence, the statement follows immediately from Lemma 7.64 with $Q_1 = Q$ and $Q_2 = 0$. □

*Proof of Lemma 7.63.* It follows immediately from Eq. (3.63)₁₀₀, Eq. (7.71), Corollary 7.65 and again Eq. (3.63)₁₀₀ that

$$\|\mathrm{pm}(\mathscr{Q})\|_{\mathrm{op}} = \sup\{\|\pi Q\|_{\mathrm{op}} \colon Q \in \mathscr{Q}, \pi \in \Pi_{\mathscr{X}}\} = \sup\{\|Q\|_{\mathrm{op}} \colon Q \in \mathscr{Q}\} = \|\mathscr{Q}\|_{\mathrm{op}}. \qquad \square$$

Fix a non-empty and bounded set $\mathscr{Q}$ of rate operators on $\mathbb{G}(\mathscr{X})$. Because the set $\mathrm{pm}(\mathscr{Q})$ is bounded by Lemma 7.63, its lower envelope as defined by Eq. (3.71)₁₀₉ is a lower rate operator due to Proposition 3.65₁₁₀. We denote this lower envelope by $\underline{Q}^{\star}_{\mathscr{Q}}$: for all $f$ in $\mathbb{G}(\mathscr{X})$ and $x$ in $\mathscr{X}$,

$$\big[\underline{Q}^{\star}_{\mathscr{Q}} f\big](x) := \big[\underline{Q}_{\mathrm{pm}(\mathscr{Q})} f\big](x) = \inf\{\big[(\pi Q) f\big](x) \colon Q \in \mathscr{Q}, \pi \in \Pi_{\mathscr{X}}\}. \tag{7.72}$$

**Corollary 7.66.** *For any non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$, the corresponding operator $\underline{Q}^{\star}_{\mathscr{Q}}$ is a lower rate operator.*

*Proof.* Due to Lemma 7.63, this follows immediately from Proposition 3.65₁₁₀. □

Finally, we are ready to determine our superset $\mathscr{Q}^{\star}$ of the non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$: because $\underline{Q}^{\star}_{\mathscr{Q}}$ is a lower rate operator, the set

$$\mathscr{Q}^{\star} := \mathscr{Q}_{\underline{Q}^{\star}_{\mathscr{Q}}} = \{Q \in \mathfrak{Q}_{\mathscr{X}} \colon (\forall f \in \mathbb{G}(\mathscr{X}))\ Qf \geq \underline{Q}^{\star}_{\mathscr{Q}} f\} \tag{7.73}$$

of dominating rate operators is well-defined.

**Corollary 7.67.** *For any non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$, the corresponding set $\mathscr{Q}^{\star}$ as defined by Eq. (7.73) is non-empty, bounded and convex and has separately specified rows, and its lower envelope is $\underline{Q}^{\star}_{\mathscr{Q}}$. Furthermore, $\mathscr{Q}^{\star}$ includes* $\mathrm{pm}(\mathscr{Q})$.

*Proof.* Due to Corollary 7.66, it follows immediately from Lemma $3.66_{110}$ that $\mathcal{Q}^\star$ is non-empty and bounded and that this set has lower envelope $\underline{Q}^\star_{\mathcal{Q}}$, and from Lemma $3.69_{111}$ that this set $\mathcal{Q}^\star$ is convex and has separately specified rows. The set $\mathcal{Q}^\star$ trivially includes $\mathrm{pm}(\mathcal{Q})$ because $\underline{Q}^\star_{\mathcal{Q}}$ is defined as the lower envelope of $\mathrm{pm}(\mathcal{Q})$. □

So, $\mathcal{Q}^\star$ is a superset of $\mathcal{Q}$ that has 'all the nice properties', but we have not yet verified that $\mathcal{Q}$ and $\mathcal{Q}^\star$ lump to the same set of rate operators, in the sense that $\hat{\mathcal{Q}}_{\mathcal{Q}} = \hat{\mathcal{Q}}_{\mathcal{Q}^\star}$. We better get to it.

**Lemma 7.68.** *Consider a non-empty and bounded subset $\mathcal{Q}$ of $\mathfrak{Q}_{\mathcal{X}}$. Then $\hat{\mathcal{Q}}_{\mathcal{Q}} = \hat{\mathcal{Q}}_{\mathcal{Q}^\star}$.*

*Proof.* By Eq. $(7.33)_{358}$, it suffices to prove that $\underline{\hat{Q}}_{\mathcal{Q}} = \underline{\hat{Q}}_{\mathcal{Q}^\star}$. To this end, we fix any $\hat{f}$ in $\mathbb{G}(\hat{\mathcal{X}})$ and $\hat{x}$ in $\hat{\mathcal{X}}$. Then by definition of $\underline{\hat{Q}}_{\mathcal{Q}^\star}$ and $\underline{Q}^\star_{\mathcal{Q}}$ – so Eqs. $(7.27)_{355}$ and (7.72) – and Eq. $(7.68)_{390}$,

$$
\begin{aligned}
\big[\underline{\hat{Q}}_{\mathcal{Q}^\star}\hat{f}\big](\hat{x}) &= \min\big\{\big[\underline{Q}_{\mathcal{Q}^\star}(\hat{f}\circ\Lambda)\big](x)\colon x\in\Lambda^{-1}(\hat{x})\big\}\\
&= \min\big\{\big[\underline{Q}^\star_{\mathcal{Q}}(\hat{f}\circ\Lambda)\big](x)\colon x\in\Lambda^{-1}(\hat{x})\big\}\\
&= \min\big\{\inf\big\{\big[(\pi Q)(\hat{f}\circ\Lambda)\big](x)\colon Q\in\mathcal{Q},\pi\in\Pi_{\mathcal{X}}\big\}\colon x\in\Lambda^{-1}(\hat{x})\big\}\\
&= \min\big\{\inf\big\{\big[Q\big((\hat{f}\circ\Lambda)\circ\pi\big)\big]\big(\pi^{-1}(x)\big)\colon Q\in\mathcal{Q},\pi\in\Pi_{\mathcal{X}}\big\}\colon x\in\Lambda^{-1}(\hat{x})\big\}.
\end{aligned}
$$

For any $\pi$ in $\Pi_{\mathcal{X}}$, $\Lambda\circ\pi=\Lambda$, and therefore

$$(\hat{f}\circ\Lambda)\circ\pi = \hat{f}\circ\Lambda\circ\pi = \hat{f}\circ\Lambda.$$

We substitute this equality in the preceding one, to yield

$$\big[\underline{\hat{Q}}_{\mathcal{Q}^\star}\hat{f}\big](\hat{x}) = \min\big\{\inf\big\{\big[Q(\hat{f}\circ\Lambda)\big]\big(\pi^{-1}(x)\big)\colon Q\in\mathcal{Q},\pi\in\Pi_{\mathcal{X}}\big\}\colon x\in\Lambda^{-1}(\hat{x})\big\}.$$

Recall from Lemma $7.56_{388}$ (iv) that, for all $x$ in $\Lambda^{-1}(\hat{x})$, $\{\pi^{-1}(x)\colon \pi\in\Pi_{\mathcal{X}}\} = \Lambda^{-1}(\hat{x})$. Hence, it follows immediately from the preceding equality and Eq. $(7.27)_{355}$ for $\underline{\hat{Q}}_{\mathcal{Q}}$ that

$$
\begin{aligned}
\big[\underline{\hat{Q}}_{\mathcal{Q}^\star}\hat{f}\big](\hat{x}) &= \min\big\{\inf\big\{\big[Q(\hat{f}\circ\Lambda)\big](y)\colon Q\in\mathcal{Q},y\in\Lambda^{-1}(\hat{x})\big\}\colon x\in\Lambda^{-1}(\hat{x})\big\}\\
&= \inf\big\{\big[Q(\hat{f}\circ\Lambda)\big](y)\colon Q\in\mathcal{Q},y\in\Lambda^{-1}(\hat{x})\big\}\\
&= \min\big\{\inf\big\{\big[Q(\hat{f}\circ\Lambda)\big](x)\colon Q\in\mathcal{Q}\big\}\colon x\in\Lambda^{-1}(\hat{x})\big\}\\
&= \min\big\{\big[\underline{Q}_{\mathcal{Q}}(\hat{f}\circ\Lambda)\big](x)\colon x\in\Lambda^{-1}(\hat{x})\big\}\\
&= \big[\underline{\hat{Q}}_{\mathcal{Q}}\hat{f}\big](\hat{x}).
\end{aligned}
$$

Because this equality holds for arbitrary $\hat{x}$ in $\hat{\mathcal{X}}$ and $\hat{f}$ in $\mathbb{G}(\hat{\mathcal{X}})$, we have shown that $\underline{\hat{Q}}_{\mathcal{Q}} = \underline{\hat{Q}}_{\mathcal{Q}^\star}$, as required. □

### 7.B.3 Putting our proof together

Finally, we set out to prove Theorem $7.32_{363}$.

**Theorem 7.32.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$ and a non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$. Fix any $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\{X_u = x_u\}$ in $\mathscr{H}$ and $\hat{f}$ in $\overline{\mathbb{V}}_{\lim}(\hat{\mathscr{F}})_u$. Then there is a jump process $\hat{P}$ in $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ such that*

$$E^{\mathrm{D}}_{\hat{P}}\big(\hat{f}\,\big|\,\hat{X}_u = \Lambda(x_u)\big) = E^{\mathrm{D}}_{P}\big(\hat{f}^{\uparrow\Omega}\,\big|\,X_u = x_u\big).$$

Most of the heavy lifting in our proof occurs in the following 'intermediary' result.

**Proposition 7.69.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$ and a non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$, and let $\mathscr{M}^\star$ be as defined by Eq. $(7.67)_{389}$ and $\mathscr{Q}^\star$ be as defined by Eq. $(7.73)_{393}$. Fix any $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ and $\{X_u = x_u\}$ in $\mathscr{H}$. Then there is a jump process $P'$ in $\mathbb{P}_{\mathscr{M}^\star,\mathscr{Q}^\star}$ such that for any $y_u$ in $\mathscr{X}_u$ with $\Lambda(y_u) = \Lambda(x_u)$ and any $\hat{A}$ in $\hat{\mathscr{F}}_u$,*

$$P'\big(\Lambda_\Omega^{-1}(\hat{A})\,\big|\,X_u = y_u\big) = P\big(\Lambda_\Omega^{-1}(\hat{A})\,\big|\,X_u = x_u\big).$$

Our proof for Proposition 7.69 is rather long. To make it more easily digestible, we establish the following intermediary result first. It is for this result that we closed $\mathscr{M}$ and $\mathscr{Q}$ under lump-preserving permutations.

**Lemma 7.70.** *Consider a non-empty subset $\mathscr{M}$ of $\Sigma_{\mathscr{X}}$ and a non-empty and bounded subset $\mathscr{Q}$ of $\mathfrak{Q}_{\mathscr{X}}$, and let $\mathscr{M}^\star$ be as defined by Eq. $(7.67)_{389}$ and $\mathscr{Q}^\star$ be as defined by Eq. $(7.73)_{393}$. Fix any $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ and $\pi$ in $\Pi_{\mathscr{X}}$. Then there is a jump process $P_\pi$ in $\mathbb{P}_{\mathscr{M}^\star,\mathscr{Q}^\star}$ such that for any $\{X_u = x_u\}$ in $\mathscr{H}$, $v$ in $\mathscr{U}_{\succcurlyeq u}$ and $\hat{B} \subseteq \hat{\mathscr{X}}_v$,*

$$P_\pi\big(X_v \in \Lambda^{-1}(\hat{B})\,\big|\,X_u = x_u\big) = P\big(X_v \in \Lambda^{-1}(\hat{B})\,\big|\,X_u = \pi(x_u)\big).$$

*Proof.* Our proof consists of three parts: (i) we define a jump process $P_\pi$ that satisfies the equality in the statement; (ii) we show that $P_\pi$ is consistent with $\mathscr{M}^\star$; and (iii) we show that $P_\pi$ is consistent with $\mathscr{Q}^\star$.

First, we define the jump process $P_\pi$. Let $\sigma := \pi^{-1}$; because $\pi$ is a lump-preserving permutation by assumption, $\sigma$ is a lump-preserving permutation due to Lemma $7.56_{388}$ (ii); hence, $\sigma^{-1} = \pi$. We interpret the lump-preserving permutation $\sigma$ as a lumping map from $\mathscr{X}$ to the 'lumped' state space $\mathscr{X}$. Thus, we fix any coherent extension $P^\star$ of $P$ to $\mathscr{D}^\star$, and let $P_\pi$ be the 'lumped' jump process corresponding to $P^\star$ as defined by Eq. $(7.15)_{347}$: for all $(A\,|\,X_u = x_u)$ in $\mathscr{D}$,

$$\begin{aligned} P_\pi(A\,|\,X_u = x_u) &:= P^\star\big(\sigma_\Omega^{-1}(A)\,\big|\,X_u \in \sigma^{-1}(x_u)\big) = P^\star\big(\sigma_\Omega^{-1}(A)\,\big|\,X_u = \pi(x_u)\big) \\ &= P\big(\sigma_\Omega^{-1}(A)\,\big|\,X_u = \pi(x_u)\big), \end{aligned} \tag{7.74}$$

where for the second equality we used that $\sigma^{-1}(x_u) = \pi(x_u)$ is a 'singleton' – it is an element of $\mathscr{X}_u$, to be exact – and for the third equality we used that $P^\star$ coincides with $P$ on $\mathscr{D}$ and that $(\sigma_\Omega^{-1}(A) \mid X_u = \pi(x_u))$ belongs to $\mathscr{D}$. By Theorem 7.8$_{347}$, $P_\pi$ is a jump process.

Next, we prove that this jump process $P_\pi$ satisfies the equality in the statement. To this end, we fix some $\{X_u = x_u\}$ in $\mathscr{H}$, $v$ in $\mathscr{U}_{\succcurlyeq u}$ and $\hat{B} \subseteq \hat{\mathscr{X}}_v$. Let $B := \Lambda^{-1}(\hat{B})$. By Lemma 7.6$_{345}$ for the lumping map $\sigma$,

$$\sigma_\Omega^{-1}\big(\{X_v \in B\}\big) = \big\{X_v \in \sigma^{-1}(B)\big\}. \tag{7.75}$$

Recall from Eq. (7.3)$_{342}$ that, by definition, $B = \Lambda^{-1}(\hat{B})$ contains those $y_v$ in $\mathscr{X}_v$ such that $\Lambda(y_v)$ belongs to $\hat{B}$. Furthermore, because $\Lambda \circ \pi = \Lambda$ by assumption – $\pi$ is a lump-preserving permutation – $\Lambda(y_v) = \Lambda\big(\pi(y_v)\big)$ for all $y_v$ in $\mathscr{X}_v$. Hence, $y_v$ belongs to $B$ if and only if $\pi(y_v)$ belongs to $B$, and therefore

$$\sigma^{-1}(B) = \pi(B) = \big\{\pi(y_v) \colon y_v \in B\big\} = B.$$

We substitute this equality in Eq. (7.75), to yield

$$\sigma_\Omega^{-1}\big(\big\{X_v \in \Lambda^{-1}(\hat{B})\big\}\big) = \sigma_\Omega^{-1}\big(\{X_v \in B\}\big) = \big\{X_v \in B\big\} = \big\{X_v \in \Lambda^{-1}(\hat{B})\big\}.$$

It follows from this and Eq. (7.74)$_\curvearrowleft$ that

$$P_\pi\big(X_v \in \Lambda^{-1}(\hat{B}) \,\big|\, X_u = x_u\big) = P\big(X_v \in \Lambda^{-1}(\hat{B}) \,\big|\, X_u = \pi(x_u)\big),$$

and this proves that the jump process $P_\pi$ satisfies the equality in the statement.

In the second part of this proof, we show that $P_\pi$ is consistent with $\mathscr{M}^\star$. Denote the initial probability mass function of $P$ by $p$ and that of $P_\pi$ by $p_\pi$: for all $x$ in $\mathscr{X}$,

$$p(x) = P(X_0 = x \mid X_{()} = x_{()}) \quad \text{and} \quad p_\pi(x) = P_\pi(X_0 = x \mid X_{()} = x_{()}).$$

For all $x$ in $\mathscr{X}$, it follows from Lemma 7.6$_{345}$ for the lumping map $\sigma$ – and with $v = (0)$ and $\hat{B} = \{x\}$ – that

$$\sigma_\Omega^{-1}\big(\{X_0 = x\}\big) = \big\{X_0 = \sigma^{-1}(x)\big\} = \big\{X_0 = \pi(x)\big\};$$

Because furthermore $\pi(x_{()}) = x_{()}$, it follows from the preceding and Eq. (7.74)$_\curvearrowleft$ that

$$p_\pi(x) = P_\pi(X_0 = x \mid X_{()} = x_{()}) = P(X_0 = \pi(x) \mid X_{()} = x_{()}) = p\big(\pi(x)\big).$$

This equality holds for arbitrary $x$ in $\mathscr{X}$, so we conclude from this that $p_\pi = p \circ \pi = \pi^{\mathrm{t}} p$. The jump process $P$ is consistent with $\mathscr{M}$ by assumption, so its initial probability mass function $p$ belongs to $\mathscr{M}$. By definition of $\mathscr{M}^\star$, this implies that $p_\pi = \pi^{\mathrm{t}} p$ belongs to $\mathscr{M}^\star$, so $P_\pi$ is consistent with $\mathscr{M}^\star$, as required.

In the third and final part of this proof, we show that $P_\pi$ is consistent with $\mathscr{Q}^\star$. For all $\{X_w = y_w\}$ in $\mathscr{H}$ and $t, r$ in $\mathbb{R}_{\geq 0}$ such that $w < t < r$, we denote the corresponding history-dependent transition operator of $P$ as defined by Eq. (3.35)$_{84}$ by $T_{t,r}^{\{X_w = y_w\}}$, and that of $P_\pi$ by $S_{t,r}^{\{X_w = y_w\}}$; furthermore, we let

$$Q_{t,r}^{\{X_w = y_w\}} := \frac{T_{t,r}^{\{X_w = y_w\}} - I}{r - t} \quad \text{and} \quad R_{t,r}^{\{X_w = y_w\}} := \frac{S_{t,r}^{\{X_w = y_w\}} - I}{r - t}.$$

Because of how we defined $P_\pi$, we expect that these rate operators are connected through $\pi$. To establish this connection, we fix some $\{X_w = y_w\}$ in $\mathscr{H}$, $t, r$ in $\mathbb{R}_{\geq 0}$ such that $w \prec t < r$, $f$ in $\mathbb{G}(\mathscr{X})$ and $x$ in $\mathscr{X}$. By Eq. (3.35)$_{84}$ and Eq. (2.19)$_{36}$,

$$\begin{aligned}\left[S_{t,r}^{\{X_w=y_w\}} f\right](x) &= E_{P_\pi}(f(X_r) \,|\, X_w = y_w, X_t = x) \\ &= \sum_{y\in\mathscr{X}} f(y) P_\pi(X_r = y \,|\, X_w = y_w, X_t = x).\end{aligned}$$

It follows from Lemma 7.6$_{345}$ for the lumping map $\sigma$ – and with $v = (r)$ and $\hat{B} = \{y\}$ – that $\sigma_\Omega^{-1}\big(\{X_r = y\}\big) = \{X_r = \pi(y)\}$. Hence, by Eq. (7.74)$_{395}$,

$$\begin{aligned}\left[S_{t,r}^{\{X_w=y_w\}} f\right](x) &= \sum_{y\in\mathscr{X}} f(y) P\big(X_r = \pi(y) \,\big|\, X_w = \pi(y_w), X_t = \pi(x)\big) \\ &= \sum_{y\in\mathscr{X}} f(y) P\big(X_r = \sigma^{-1}(y) \,\big|\, X_w = z_w, X_t = \sigma^{-1}(x)\big),\end{aligned}$$

where for the second equality we used that $\pi = \sigma^{-1}$ and we let $z_w := \pi(y_w)$. We execute some straightforward manipulations, to find that

$$\begin{aligned}\left[S_{t,r}^{\{X_w=y_w\}} f\right](x) &= \sum_{y\in\mathscr{X}} f\Big(\sigma\big(\sigma^{-1}(y)\big)\Big) P\big(X_r = \sigma^{-1}(y) \,\big|\, X_w = z_w, X_t = \sigma^{-1}(x)\big) \\ &= \sum_{y\in\mathscr{X}} [f\circ\sigma]\big(\sigma^{-1}(y)\big) P\big(X_r = \sigma^{-1}(y) \,\big|\, X_w = z_w, X_t = \sigma^{-1}(x)\big) \\ &= \sum_{y\in\mathscr{X}} [f\circ\sigma](y) P\big(X_r = y \,\big|\, X_w = z_w, X_t = \sigma^{-1}(x)\big) \\ &= \left[T_{t,r}^{\{X_w=z_w\}}(f\circ\sigma)\right]\big(\sigma^{-1}(x)\big),\end{aligned}$$

where for the first equality we used that $\sigma\circ\sigma^{-1} = \mathrm{id}$ and for the third equality we used that $\{\sigma^{-1}(y) \colon y \in \mathscr{X}\} = \mathscr{X}$ – see Lemma 7.56$_{388}$ (v). Because $f(x) = [f\circ\sigma]\big(\sigma^{-1}(x)\big)$, it follows from the preceding equality that

$$\begin{aligned}\left[R_{t,r}^{\{X_w=y_w\}} f\right](x) &= \frac{\left[S_{t,r}^{\{X_w=y_w\}} f\right](x) - f(x)}{r-t} \\ &= \frac{\left[T_{t,r}^{\{X_w=z_w\}}(f\circ\sigma)\right]\big(\sigma^{-1}(x)\big) - [f\circ\sigma]\big(\sigma^{-1}(x)\big)}{r-t} \\ &= \left[Q_{t,r}^{\{X_w=z_w\}}(f\circ\sigma)\right]\big(\sigma^{-1}(x)\big).\end{aligned}$$

This equality holds for arbitrary $x$ in $\mathscr{X}$ and $f$ in $\mathbb{G}(\mathscr{X})$, so we conclude that

$$R_{t,r}^{\{X_w=y_w\}} = \sigma Q_{t,r}^{\{X_w=z_w\}}, \tag{7.76}$$

where $\sigma Q_{t,r}^{\{X_w=z_w\}}$ is the permuted rate operator as defined by Eq. (7.68)$_{390}$. Because $\sigma = \pi^{-1}$, it follows from Eq. (7.76) and Lemma 7.61$_{391}$ that

$$\pi R_{t,r}^{\{X_w=y_w\}} = Q_{t,r}^{\{X_w=z_w\}}. \tag{7.77}$$

As an intermediary step towards proving that $P_\pi$ is consistent with $\mathscr{Q}^\star$, we establish that $P_\pi$ has bounded rate. By Definition 3.53<sub>101</sub>, we need to show that for all $t$ in $\mathbb{R}_{\geq 0}$ and $\{X_w = y_w\}$ in $\mathscr{H}$ such that $w \prec t$,

$$\limsup_{r \searrow t} \left\| R_{t,r}^{\{X_w=y_w\}} \right\|_{\mathrm{op}} < +\infty \quad \text{and, if } t > 0, \quad \limsup_{s \nearrow t} \left\| R_{s,t}^{\{X_w=y_w\}} \right\|_{\mathrm{op}} < +\infty. \tag{7.78}$$

Thus, we fix any such $t$ in $\mathbb{R}_{\geq 0}$ and $\{X_w = y_w\}$ in $\mathscr{H}$, and let $z_w := \pi(y_w)$. Then for all $r$ in $]t, +\infty[$ and $s$ in $[0, t[$, it follows from Eq. (7.76)↶ and Corollary 7.65<sub>393</sub> that

$$\left\| R_{t,r}^{\{X_w=y_w\}} \right\|_{\mathrm{op}} = \left\| \sigma Q_{t,r}^{\{X_w=z_w\}} \right\|_{\mathrm{op}} = \left\| Q_{t,r}^{\{X_w=z_w\}} \right\|_{\mathrm{op}}$$

and, if $t > 0$

$$\left\| R_{s,t}^{\{X_w=y_w\}} \right\|_{\mathrm{op}} = \left\| \sigma Q_{s,t}^{\{X_w=z_w\}} \right\|_{\mathrm{op}} = \left\| Q_{s,t}^{\{X_w=z_w\}} \right\|_{\mathrm{op}}.$$

By assumption, $P$ is consistent with $\mathscr{Q}$ and this set is bounded, so $P$ has bounded rate by Lemma 3.55<sub>102</sub>. Hence, it follows from Definition 3.53<sub>101</sub> for $P$ and the preceding equalities that Eq. (7.78) is satisfied, as required.

Finally, we prove that $P_\pi$ is consistent with $\mathscr{Q}^\star$. In order to verify the condition in Definition 3.50<sub>99</sub>, we fix some $t$ in $\mathbb{R}_{\geq 0}$ and $\{X_w = y_w\}$ in $\mathscr{H}$ such that $w \prec t$. We need to show that $S_{t,t}^{\{X_w=y_w\}}$ is $d_{\mathfrak{Q}}$-differentiable, and that

$$\partial_+ S_{t,t}^{\{X_w=y_w\}} \subseteq \mathscr{Q}^\star \quad \text{and, if } t > 0, \quad \partial_- S_{t,t}^{\{X_w=y_w\}} \subseteq \mathscr{Q}^\star.$$

Because $P_\pi$ has bounded rate, $S_{t,t}^{\{X_w=y_w\}}$ is $d_{\mathfrak{Q}}$-differentiable due to Proposition 3.57<sub>104</sub>. Moreover, we can use this result to determine $\partial_+ S_{t,t}^{\{X_w=y_w\}}$ and, if $t > 0$, $\partial_- S_{t,t}^{\{X_w=y_w\}}$. Fix any rate operator $Q$, and let $(r_n)_{n\in\mathbb{N}}$ be any sequence in $]t, +\infty[$ such that $t < r_{n+1} < r_n$ for all $n$ in $\mathbb{N}$ and $\lim_{n\to+\infty} r_n = t$. Then by Lemma 7.64<sub>392</sub> and Eq. (7.76)↶,

$$(\forall n \in \mathbb{N})\ \left\| Q_{t,r}^{\{X_w=z_w\}} - Q \right\|_{\mathrm{op}} = \left\| \sigma Q_{t,r}^{\{X_w=z_w\}} - \sigma Q \right\|_{\mathrm{op}} = \left\| R_{t,r}^{\{X_w=y_w\}} - \sigma Q \right\|_{\mathrm{op}},$$

where we let $z_w := \pi(y_w)$; similarly, by Lemma 7.64<sub>392</sub> and Eq. (7.77)↶,

$$(\forall n \in \mathbb{N})\ \left\| R_{t,r}^{\{X_w=y_w\}} - Q \right\|_{\mathrm{op}} = \left\| \pi R_{t,r}^{\{X_w=y_w\}} - \pi Q \right\|_{\mathrm{op}} = \left\| Q_{t,r}^{\{X_w=z_w\}} - \pi Q \right\|_{\mathrm{op}}.$$

Hence, it follows from Proposition 3.57<sub>104</sub> that

$$\partial_+ S_{t,t}^{\{X_w=y_w\}} \supseteq \left\{ \sigma Q \colon Q \in \partial_+ T_{t,t}^{\{X_w=z_w\}} \right\} \text{ and } \partial_+ T_{t,t}^{\{X_w=z_w\}} \supseteq \left\{ \pi Q \colon Q \in \partial_+ S_{t,t}^{\{X_w=y_w\}} \right\}.$$

Due to Lemma 7.61<sub>391</sub>, this implies that

$$\partial_+ S_{t,t}^{\{X_w=y_w\}} = \left\{ \sigma Q \colon Q \in \partial_+ T_{t,t}^{\{X_w=z_w\}} \right\}.$$

If $t > 0$, then a similar argument shows that

$$\partial_- S_{t,t}^{\{X_w=y_w\}} = \left\{ \sigma Q \colon Q \in \partial_- T_{t,t}^{\{X_w=z_w\}} \right\}.$$

Because $P$ is consistent with $\mathcal{Q}$, $\mathcal{Q}$ includes $\partial_+ T_{t,t}^{\{X_w=z_w\}}$ and, if $t>0$, $\partial_- T_{t,t}^{\{X_w=z_w\}}$. For this reason, and because $\sigma$ is a lump-preserving permutation, it follows from the preceding two equalities and Corollary 7.67$_{393}$ that

$$\partial_+ S_{t,t}^{\{X_w=y_w\}} \subseteq \mathrm{pm}(\mathcal{Q}) \subseteq \mathcal{Q}^\star \quad \text{and, if } t>0, \quad \partial_- S_{t,t}^{\{X_w=y_w\}} \subseteq \mathrm{pm}(\mathcal{Q}) \subseteq \mathcal{Q}^\star,$$

as required. □

*Proof of Proposition 7.69$_{395}$.* Before we start with our proper proof, we recall from Lemma 7.59$_{389}$ that $\mathcal{M}^\star$ includes $\mathcal{M}$, and from Lemma 7.62$_{392}$ and Corollary 7.67$_{393}$ that $\mathcal{Q}^\star$ is a non-empty, bounded and convex set of rate operators that has separately specified rows and that includes $\mathcal{Q}$. The statement is trivially true for $P' := P$ whenever $u = ()$, so we may assume without loss of generality that $u \neq ()$.

Let $t := \max u$, $x := x_t$, $\hat{x} := \Lambda(x)$ and $\hat{x}_u := \Lambda(x_u)$. For all $z$ in $\Lambda^{-1}(\hat{x}) \setminus \{x\}$, we fix some lump-preserving permutation $\pi_z$ in $\Pi_{\mathcal{X}}$ such that $\pi_z^{-1}(x) = z$ – for example, the permutation that maps $z$ to $x$ and $x$ to $z$, and that maps every other $y$ in $\mathcal{X} \setminus \{x,z\}$ to itself. For all $z$ in $\Lambda^{-1}(\hat{x}) \setminus \{x\}$, Lemma 7.70$_{395}$ guarantees that there is a jump process $P_{\pi_z}$ in $\mathbb{P}_{\mathcal{M}^\star,\mathcal{Q}^\star}$ such that for all $v$ in $\mathcal{U}_{\succcurlyeq u}$ and $\hat{B} \subseteq \hat{\mathcal{X}}_v$,

$$\begin{aligned} P_{\pi_z}\big(X_v \in \Lambda^{-1}(\hat{B}) \,\big|\, X_u = \pi_z^{-1}(x_u)\big) &= P\big(X_v \in \Lambda^{-1}(\hat{B}) \,\big|\, X_u = \pi_z(\pi_z^{-1}(x_u))\big) \\ &= P\big(X_v \in \Lambda^{-1}(\hat{B}) \,\big|\, X_u = x_u\big). \end{aligned} \tag{7.79}$$

For all $z_u$ in $\Lambda^{-1}(\hat{x}_u)$ such that $z_t \neq x = x_t$, we let $P_{z_u} := P_{\pi_{z_t}}$ and $y_u^{z_u} := \pi_{z_t}^{-1}(x_u)$; then by construction, $y_t^{z_u} = \pi_{z_t}^{-1}(x) = z_t$. For all $z_u$ in $\Lambda^{-1}(\hat{x}_u)$ such that $z_t = x = x_t$, we let $P_{z_u} := P$ and $y_u^{z_u} := x_u$; here too, $y_t^{z_u} = x_t = z_t$ by construction. Finally, for all $z_u$ in $\mathcal{X}_u \setminus \Lambda^{-1}(\hat{x}_u)$, we let $P_{z_u} := P$ and $y_u^{z_u} := z_u$ such that $y_t^{z_u} = z_t$.

Let $P_0 := P$. Because $\mathcal{M} \subseteq \mathcal{M}^\star$ and $\mathcal{Q} \subseteq \mathcal{Q}^\star$, $\mathbb{P}_{\mathcal{M},\mathcal{Q}} \subseteq \mathbb{P}_{\mathcal{M}^\star,\mathcal{Q}^\star}$. Hence, $P$ belongs to $\mathbb{P}_{\mathcal{M}^\star,\mathcal{Q}^\star}$, and therefore so do $P_0$ and, for all $z_u$ in $\mathcal{X}_u$, $P_{z_u}$. Because $\mathcal{Q}^\star$ is non-empty, bounded and convex and has separately specified rows, it follows from Theorem 3.103$_{139}$ that there is jump process $P'$ in $\mathbb{P}_{\mathcal{M}^\star,\mathcal{Q}^\star}$ such that for all $z_u$ in $\Lambda^{-1}(\hat{x}_u)$, $v$ in $\mathcal{U}_{>u}$ and $B \subseteq \mathcal{X}_v$,

$$P'\big(X_v \in B \,\big|\, X_u = z_u\big) = P_{z_u}\big(X_v \in B \,\big|\, y_u^{z_u}\big);$$

note that, by construction,

$$P'\big(X_v \in B \,\big|\, X_u = z_u\big) = \begin{cases} P_{\pi_{z_t}}(X_v \in B \,|\, X_u = \pi_{z_t}^{-1}(x_u)) & \text{if } z_t \neq x, \\ P(X_v \in B \,|\, X_u = x_u) & \text{if } z_t = x. \end{cases} \tag{7.80}$$

Our definition of $P'$ is perhaps a bit unorthodox, but it ensures the following quintessential property: for all $z_u$ in $\Lambda^{-1}(\hat{x}_u)$, $v$ in $\mathcal{U}_{>u}$ and $\hat{B} \subseteq \hat{\mathcal{X}}_v$, it follows from Eqs. (7.80) and (7.79) that

$$P'\big(X_v \in \Lambda^{-1}(\hat{B}) \,\big|\, X_u = z_u\big) = P\big(X_v \in \Lambda^{-1}(\hat{B}) \,\big|\, X_u = x_u\big). \tag{7.81}$$

It still remains for us to prove that $P'$ satisfies the equality in the statement. To this end, we fix any $y_u$ in $\mathcal{X}_u$ such that $\Lambda(y_u) = \Lambda(x_u) = \hat{x}_u$ – note that $y_u$ belongs to $\Lambda^{-1}(\hat{x}_u)$ – and any $\hat{A}$ in $\hat{\mathscr{F}}_u$. By definition of $\hat{\mathscr{F}}_u$, there is some $v$ in $\mathscr{U}_{\succcurlyeq u}$ and some subset $\hat{B}$ of $\hat{\mathscr{X}}_v$ such that $\hat{A} = \{\hat{X}_v \in \hat{B}\}$. Hence, by Eq. (7.11)345 in Lemma 7.6345,

$$\Lambda_\Omega^{-1}(\hat{A}) = \Lambda_\Omega^{-1}(\{\hat{X}_v \in \hat{B}\}) = \{X_v \in \Lambda^{-1}(\hat{B})\}. \tag{7.82}$$

If $v \succ u$, then the equality in the statement follows immediately from the preceding equality and Eq. (7.81)↶. To prove the equality in the statement in the other case, we resort to (JP1)69. Let $w := v \setminus u$, and let

$$B_{y_u} := \{y_w \in \mathscr{X}_w : (\exists z_v \in \Lambda^{-1}(\hat{B}))\ z_{u\cap v} = y_{u\cap v}, z_w = y_w\},$$
$$B_{x_u} := \{x_w \in \mathscr{X}_w : (\exists z_v \in \Lambda^{-1}(\hat{B}))\ z_{u\cap v} = x_{u\cap v}, z_w = x_w\}$$

and

$$\hat{C} := \{\hat{x}_w \in \hat{\mathscr{X}}_w : (\exists \hat{z}_v \in \hat{B})\ \hat{z}_{u\cap v} = \hat{x}_{u\cap v}, \hat{z}_w = \hat{x}_w\}.$$

By definition, $\Lambda^{-1}(\hat{B})$ contains those $z_v$ in $\mathscr{X}_v$ such that $\Lambda(z_v)$ belongs to $\hat{B}$. Thus, any $y_w$ in $\mathscr{X}_w$ belongs to $B_{y_u}$ if and only if $\Lambda(y_{u\cap v}, y_w)$ belongs to $\hat{B}$, and similarly, any $x_w$ in $\mathscr{X}_w$ belongs to $B_{x_u}$ if and only if $\Lambda(x_{u\cap v}, x_w)$ belongs to $\hat{B}$. Because $\Lambda(y_{u\cap v}) = \Lambda(x_{u\cap v}) = \hat{x}_{u\cap v}$, we infer from this that $B_{x_u} = B_{y_u} = \Lambda^{-1}(\hat{C})$. For this reason, it follows from Eq. (7.82) and (JP1)69 that

$$\begin{aligned} P'(\Lambda_\Omega^{-1}(\hat{A}) \,|\, X_u = y_u) &= P'(X_v \in \Lambda^{-1}(\hat{B}) \,|\, X_u = y_u) = P'(X_w \in B_{y_u} \,|\, X_u = y_u) \\ &= P'(X_w \in \Lambda^{-1}(\hat{C}) \,|\, X_u = y_u), \end{aligned} \tag{7.83}$$

and similarly,

$$P(\Lambda_\Omega^{-1}(\hat{A}) \,|\, X_u = x_u) = P(X_w \in \Lambda^{-1}(\hat{C}) \,|\, X_u = x_u). \tag{7.84}$$

If $w = ()$, then $\{X_w \in \Lambda^{-1}(\hat{C})\} = \Omega$, so the equality in the statement follows immediately from Eqs. (7.83) and (7.84) and (CP1)41. If $w \neq ()$, then $w \succ u$, so the equality in the statement follows immediately from Eqs. (7.83) and (7.84) and Eq. (7.81)↶. □

Finally, we can get around to proving Theorem 7.32363. Most of the work is done by Proposition 7.69395, but we still need an argument similar to the one in our proof for Lemma 7.17352.

*Proof of Theorem 7.32363.* Let $\mathscr{M}^\star$ and $\mathscr{Q}^\star$ be as defined by Eqs. (7.67)389 and (7.73)393, respectively. To ease our notation, we let $\hat{x}_u := \Lambda(x_u)$. By Proposition 7.69395, there is a jump process $P'$ in $\mathbb{P}_{\mathscr{M}^\star,\mathscr{Q}^\star}$ such that for all $y_u$ in $\Lambda^{-1}(\hat{x}_u)$ and $\hat{A}$ in $\hat{\mathscr{F}}_u$,

$$P'(\Lambda_\Omega^{-1}(\hat{A}) \,|\, X_u = y_u) = P(\Lambda_\Omega^{-1}(\hat{A}) \,|\, X_u = x_u). \tag{7.85}$$

Fix any coherent extension $P^\star$ of $P'$ to $\mathscr{D}^\star$, and let $\hat{P}$ be the corresponding lumped jump process. Recall from Lemma 7.59389 that $\hat{\mathscr{M}}_{\mathscr{M}^\star} = \hat{\mathscr{M}}$ and from Lemma 7.68394 that $\hat{\mathscr{Q}}_{\mathscr{Q}^\star} = \hat{\mathscr{Q}}$. Hence, because $P'$ belongs to $\mathbb{P}_{\mathscr{M}^\star,\mathscr{Q}^\star}$, the lumped jump process $\hat{P}$ belongs to $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ due to Theorem 7.29361.

Let us verify that $\hat{P}$ satisfies the equality in the statement. Crucial to this is that we have constructed the lumped jump process $\hat{P}$ in such a way that $\hat{P}(\bullet \,|\, \hat{X}_u = \hat{x}_u)$ is uniquely defined: for all $\hat{A}$ in $\hat{\mathscr{F}}_u$, it follows from Corollary 7.12350 and Eq. (7.85) that

$$\hat{P}(\hat{A} \,|\, \hat{X}_u = \hat{x}_u) = P(\Lambda_\Omega^{-1}(\hat{A}) \,|\, X_u = x_u). \tag{7.86}$$

Fix any $\hat{\mathscr{F}}_u$-simple variable $\hat{g}$. As in the proof of Lemma 7.14~351~, we recall from the proof of Lemma 7.13~350~ that there is a natural number $n$, real numbers $a_1, \dots, a_n$ and events $\hat{A}_1, \dots, \hat{A}_n$ in $\hat{\mathscr{F}}_u$ such that

$$\hat{g} = \sum_{k=1}^{n} a_k \mathbb{I}_{\hat{A}_k} \quad \text{and} \quad \hat{g}^{\uparrow\Omega} = \sum_{k=1}^{n} a_k \mathbb{I}_{\Lambda_\Omega^{-1}(\hat{A}_k)}.$$

Then by Eq. (2.19)~36~, Eq. (7.86)~↶~ and again Eq. (2.19)~36~,

$$\begin{aligned} E_{\hat{P}}(\hat{g} \,|\, \hat{X}_u = \hat{x}_u) &= \sum_{k=1}^{n} a_k \hat{P}(\hat{A}_k \,|\, \hat{X}_u = \hat{x}_u) = \sum_{k=1}^{n} a_k P\big(\Lambda_\Omega^{-1}(\hat{A}_k) \,\big|\, X_u = x_u\big) \\ &= E_P(\hat{g}^{\uparrow\Omega} \,|\, X_u = x_u). \end{aligned} \tag{7.87}$$

Next, alter the argument in our proof for Lemma 7.17~352~. We let $E^{\mathrm{mc}}$ denote the extension of $E_P(\bullet \,|\, X_u = x_u)$ to $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathscr{F}_u)$ as defined by Eq. (5.6)~223~: for all $g$ in $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathscr{F}_u)$,

$$E^{\mathrm{mc}}(g) = \lim_{n\to+\infty} E_P(g_n \,|\, X_u = x_u), \tag{7.88}$$

where $(g_n)_{n\in\mathbb{N}}$ is any monotone sequence of $\mathscr{F}_u$ simple variables that converges point-wise to $g$. Similarly, $\hat{E}^{\mathrm{mc}}$ denotes the extension of $E_{\hat{P}}(\bullet \,|\, \hat{X}_u = \hat{x}_u)$ to $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\hat{\mathscr{F}}_u)$: for all $\hat{g}$ in $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\hat{\mathscr{F}}_u)$,

$$\hat{E}^{\mathrm{mc}}(\hat{g}) = \lim_{n\to+\infty} E_{\hat{P}}(\hat{g}_n \,|\, \hat{X}_u = \hat{x}_u), \tag{7.89}$$

where $(\hat{g}_n)_{n\in\mathbb{N}}$ is any monotone sequence of $\hat{\mathscr{F}}_u$ simple variables that converges point-wise to $\hat{g}$.

The extension $E^{\mathrm{mc}}$ 'coincides' with $\hat{E}^{\mathrm{mc}}$ on a large part of its domain. To prove this, we fix any $\hat{g}$ in $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\hat{\mathscr{F}}_u)$. Then by definition of $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\hat{\mathscr{F}}_u)$, there is a monotone sequence $(\hat{g}_n)_{n\in\mathbb{N}}$ of $\hat{\mathscr{F}}_u$ simple variables that converges point-wise to $\hat{g}$. Furthermore, $(\hat{g}_n^{\uparrow\Omega})_{n\in\mathbb{N}}$ is a monotone sequence of $\mathscr{F}_u$-simple variables that converges point-wise to $\hat{g}^{\uparrow\Omega}$ by Lemma 7.53~384~, and $\hat{g}^{\uparrow\Omega}$ belongs to $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathscr{F}_u)$ by Lemma 7.54~386~. Hence, it follows from Eqs. (7.89), (7.87) and (7.88) that

$$\hat{E}^{\mathrm{mc}}(\hat{g}) = \lim_{n\to+\infty} E_{\hat{P}}(\hat{g}_n \,|\, \hat{X}_u = \hat{x}_u) = \lim_{n\to+\infty} E_P(\hat{g}_n^{\uparrow\Omega} \,|\, X_u = x_u) = E^{\mathrm{mc}}(\hat{g}^{\uparrow\Omega}) \tag{7.90}$$

We are almost done. By definition of $E_P^{\mathrm{D}}$,

$$\begin{aligned} E_P^{\mathrm{D}}(\hat{f}^{\uparrow\Omega} \,|\, X_u = x_u) &= \sup\{E^{\mathrm{mc}}(g) \colon g \in \overline{\mathbb{V}}_{\mathrm{u}}(\mathscr{F}_u), g \le \hat{f}^{\uparrow\Omega}\} && (7.91) \\ &= \inf\{E^{\mathrm{mc}}(g) \colon g \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathscr{F}_u), g \ge \hat{f}^{\uparrow\Omega}\}; && (7.92) \end{aligned}$$

similarly, by definition of $E_{\hat{P}}^{\mathrm{D}}$,

$$\begin{aligned} E_{\hat{P}}^{\mathrm{D}}(\hat{f} \,|\, \hat{X}_u = \hat{x}_u) &= \sup\{\hat{E}^{\mathrm{mc}}(\hat{g}) \colon \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u), \hat{g} \le \hat{f}\} && (7.93) \\ &= \inf\{\hat{E}^{\mathrm{mc}}(\hat{g}) \colon \hat{g} \in \overline{\mathbb{V}}^{\mathrm{o}}(\hat{\mathscr{F}}_u), \hat{g} \ge \hat{f}\}. && (7.94) \end{aligned}$$

By Eqs. (7.93) and (7.90),

$$\begin{aligned} E_{\hat{P}}^{\mathrm{D}}(\hat{f} \,|\, \hat{X}_u = \hat{x}_u) &= \sup\{\hat{E}^{\mathrm{mc}}(\hat{g}) \colon \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u), \hat{g} \le \hat{f}\} \\ &= \sup\{E^{\mathrm{mc}}(\hat{g}^{\uparrow\Omega}) \colon \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u), \hat{g} \le \hat{f}\} \\ &= \sup\{E^{\mathrm{mc}}(\hat{g}^{\uparrow\Omega}) \colon \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u), \hat{g}^{\uparrow\Omega} \le \hat{f}^{\uparrow\Omega}\}, \end{aligned}$$

where for the last equality we used that $\hat{g} \leq \hat{f}$ if and only if $\hat{g}^{\uparrow\Omega} \leq \hat{f}^{\uparrow\Omega}$. Because $\{\hat{g}^{\uparrow\Omega} : \hat{g} \in \overline{\mathbb{V}}_{\mathrm{u}}(\hat{\mathscr{F}}_u)\} \subseteq \overline{\mathbb{V}}_{\mathrm{u}}(\mathscr{F}_u)$ by Eq. (7.59)<sub>386</sub> in Lemma 7.54<sub>386</sub>,

$$E^{\mathrm{D}}_{\hat{P}}(\hat{f} \mid \hat{X}_u = \hat{x}_u) \leq \sup\left\{E^{\mathrm{mc}}(g) : g \in \overline{\mathbb{V}}_{\mathrm{u}}(\mathscr{F}_u), g \leq \hat{f}^{\uparrow\Omega}\right\} = E^{\mathrm{D}}_{P}(\hat{f}^{\uparrow\Omega} \mid X_u = x_u),$$

where for the equality we used Eq. (7.91)↶. Similarly, it follows from Eqs. (7.94)↶ and (7.90)↶, Lemma 7.54<sub>386</sub> and Eq. (7.92)↶ that

$$E^{\mathrm{D}}_{\hat{P}}(\hat{f} \mid \hat{X}_u = \hat{x}_u) \geq E^{\mathrm{D}}_{P}(\hat{f}^{\uparrow\Omega} \mid X_u = x_u).$$

Finally, it follows from the preceding two inequalities that

$$E^{\mathrm{D}}_{\hat{P}}(\hat{f} \mid \hat{X}_u = \hat{x}_u) = E^{\mathrm{D}}_{P}(\hat{f}^{\uparrow\Omega} \mid X_u = x_u). \qquad \square$$

# 8 Spectrum allocation in an optical link

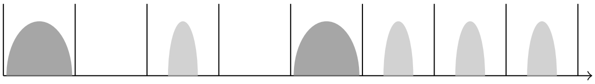
In this penultimate chapter, we put Markovian imprecise jump processes – and in particular our lumping-inspired methods of the preceding chapter – to the test, and we do so with a 'real-world' example from telecommunications engineering. In Section 8.1↷, we look at a single optical link in an optical network, and we provide some background on the problem of spectrum allocation. Under some standard assumptions, we can model (the spectrum of) a single optical link with some allocation policy by means of an exact homogeneous and Markovian jump process; we construct such a model in Section 8.2 (p. 407), and we subsequently use this model to determine performance indicators for the system: the expected number of blocked requests and blocking ratios. We also consider a policy-independent Markovian imprecise jump process model, and use this model to compute policy-independent lower and upper bounds on the performance indicators. These models suffer from state space explosion, so determining the (lower and upper) expected number of blocked requests and the (lower and upper) blocking ratio becomes infeasible for large systems; of course, this is why lumping comes into play in Section 8.3 (p. 421). After having introduced the lumped state space in Section 8.3.1 (p. 422), we introduce and extend the approximate policy-dependent model of Kim, Yan, et al. (2015). As we will see in Section 8.3.2 (p. 426), these models are homogeneous and Markovian jump processes, so we can tractably compute an approximation for the performance indicators with these approximate models; however, the accuracy of these approximations is not guaranteed. Our beloved lumping-based methods, on the other hand, give policy-dependent and policy-independent guaranteed lower and upper bounds on the performance indicators, as we will see in Sections 8.3.4 (p. 431) and 8.3.5 (p. 433).

The present chapter is largely based on (Rottondi et al., 2017; Erreygers, Rottondi, et al., 2018b), but the material in Sections 8.2.3 (p. 412) and 8.3.4 (p. 431) is new.

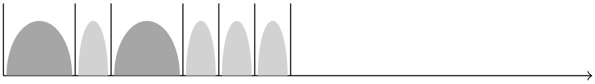
## 8.1 Flexi-grid optical networks

Over the last couple of years – and especially over the last year – our use of the internet has changed significantly: the lengthy e-mail chains of yore have made way for group video meetings, short audio and/or video clips are becoming more popular than lengthy news articles, and so on. This has led to a significant increase in internet traffic, so internet service providers need to add new capacity or, even better, make more efficient use of their existing capacity. One way to achieve the latter is by means of flexi-grid optical networks.

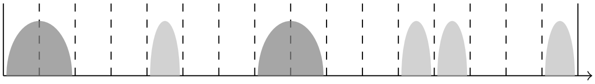
At its core, the internet is essentially a large network of optical fibres, which are known as *optical links*. Sending information from point A to point B is as 'simple' as sending light through these optical links, a bit like one would use a flashlight to signal Morse code. Putting these optical links in place is very costly, so a bunch of clever engineers have found multiple ways to increase the capacity of existing optical links, including *Dense Wavelength Division Multiplexing* (DWDM). In our Morse code analogy, DWDM corresponds to sending multiple messages at the same time – or in parallel, if you will – by using flashlights with different colours. Of course, we can only do this if we can sufficiently distinguish between the different colours of the flashlights, and this is no different with DWDM in optical links: the available (frequency) spectrum – so the range of possible 'colours' – is divided in spectrum slots, typically of width 50 GHz, as depicted in Fig. 8.1(a).



(a) Fixed grid



(b) Flexible grid



(c) Semi-flexible grid

**Figure 8.1** Example of grids. The width of the traffic flows indicates their bitrate, after (Gerstel et al., 2012, Figure 1).

### 8.1.1 Flexible spectrum allocation

Each of the slots in Fig. 8.1(a) can carry a traffic flow with a certain bitrate, and this bitrate has a maximum due to the fixed nature of the slots. In an elastic or flexi-grid optical network (see Gerstel et al., 2012), the width of the slots and their location along the spectrum need not be fixed; as can be seen in Fig. 8.1(b), the idea is to change the slots according to the bitrate requirements of the traffic flows. More precisely, spectral resources are divided in small slots – the width of these slots can be 12.5 GHz, 25 GHz, 50 GHz or 100 GHz, according to the standard by the ITU-T (2020) – and groups of contiguous slots are adaptively grouped into so-called *superchannels* and assigned to different traffic flows according to many factors, including the bitrate that is required by the traffic flow. As can be seen in Fig. 8.1(b), flexi-grid networks use less of the spectrum with respect to traditional MWDM systems – up to 30 % less in practice (Jinno et al., 2009) – but this comes at the cost of more advanced and costly optical devices (Feuer et al., 2012). For a survey of past work on flexi-grid optical networks, we refer to (Azodolmolky et al., 2009; Tomkos et al., 2012; Zhang et al., 2013).

The flexible spectrum allocation techniques enabled by flexi-grid networks typically induce spectrum fragmentation: often, groups of contiguous slots are formed that cannot be assigned to incoming traffic flows because the superchannel they would form is too narrow. Whenever an incoming traffic flow cannot be allocated, we say that it is blocked; the *blocking ratio* or blocking probability is the proportion of incoming traffic flows that are blocked, and this is an important performance indicator: the lower the blocking ratio the better. The issues of spectrum fragmentation becomes even more important for networks of optical links, because a traffic flow should get the same spectrum portion along all the optical links that it traverses.

In order to alleviate the spectrum fragmentation issue and limit the costs of equipment installation without renouncing the benefits of flexi-grid networks in terms of spectrum usage, alternative semi-flexible approaches have been proposed. One approach is to make part of the spectrum a fixed grid for high bitrate signals only (Shen, Hasegawa & Sato, 2014), or to use a dedicated fixed grid for each type of signal (Shen, Hasegawa, Sato, et al., 2014). Alternatively, as depicted in Fig. 8.1(c), one can define a set of superchannels with predefined widths, and allocate traffic flows to the smallest superchannel they fit in, at the price of leaving some spectrum slots unused (Comellas et al., 2015). This second approach allows for more scalable 'analytical' models: Kim, Yan, et al. (2015) and Kim et al. (2016) use an approximate homogeneous and Markovian jump process to determine blocking ratios in a semi-flexible optical link with a random spectrum allocation policy under the assumption that traffic demands are categorised in two types according to their bitrate – high or low, respectively.

Several authors have investigated the blocking ratios in semi-flexible optical networks. For example, Yin et al. (2013) have identified fragmentation-aware spectrum allocation policies and assessed their performance by determining the blocking ratios for traffic flows with different bitrate requirements, and Yu et al. (2013) have proposed homogeneous and Markovian jump processes to model this. For realistic scenarios, these models need a very large number of states to correctly capture the degrees of freedom offered by the flexible grid, thus introducing scalability limitations. To overcome these scalability issues, Kim, Yan, et al. (2015) use an approximate homogeneous and Markovian jump process to approximate blocking ratios of two types of traffic requests along a single optical link, assuming that spectrum allocation is performed at random. The model in (Kim, Yan, et al., 2015) has been refined by Wang et al. (2014), Kim, Wang, et al. (2015), and Kim et al. (2016) to include the case where part of the spectrum is exclusively reserved for each of the two types of traffic flows.

### 8.1.2 A single optical link with a flexible grid

Following Kim, Yan, et al. (2015), we consider a single optical link whose spectrum is partitioned in slots of equal width, for a total number of $m_1$ slots. The optical link is used to transmit traffic flows with varying bitrate demands; here, we limit ourselves to flows of two types for the sake of clarity, but our approach is applicable to optical links with more than two types of flows as well. Flows of *type 1* require 1 slot, whereas flows of *type 2* require $n_2 > 1$ slots. Type 1 and type 2 flows arrive according to Poisson processes with arrival rates $\lambda_1$ and $\lambda_2$, respectively; if assigned to some free slot(s) in the spectrum, they cease after a holding time that is exponentially distributed with average service time $1/\mu_1$ or $1/\mu_2$, respectively. We are not interested in edge cases, so throughout this chapter we assume that $\lambda_1$, $\lambda_2$, $\mu_1$ and $\mu_2$ are all positive real numbers. The spectrum is divided into two overlapping grids of different

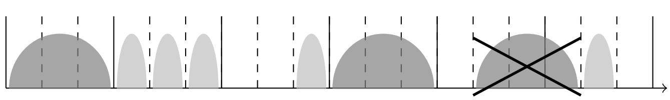


**Figure 8.2** A semi-flexible optical link with $m_1 = 18$ slots and fixed superchannels that consist of $n_2 = 3$ slots.

granularity, as depicted Fig. 8.2: the first grid consists of the $m_1$ frequency slots, and the second grid consists of a sequence of adjacent superchannels of $n_2$ slots. Thus, there are $m_2 := m_1/n_2$ superchannels, where, for the sake of convenience, we assume that $m_1$ is an integer multiple of $n_2$.

An incoming type 1 traffic flow, in the remainder referred to as a *type 1 request*, must be assigned to a free slot, whereas an incoming type 2 traffic flow, in the remainder referred to as a *type 2 request*, must be allocated to a free superchannel that consists of $n_2$ slots. If there is only a single free (super)channel, we always allocate the request to this free (super)channel. If there are multiple free (super)channels, we allocate the request to one of the free (super)channels according to some spectrum allocation policy. For type 2 requests, it will become evident in the remainder that the allocation policy is of no importance to our analysis. For type 1 requests, we only consider allocation policies that depend on the number of type 1 flows currently occupying each superchannel, but not on the order of the superchannels along the grid nor on the order of the flows in a superchannel.

In particular, we consider three such allocation policies. The *Random Allocation* (RA) policy assigns a type 1 request to a randomly selected free slot if possible, where every free slot has the same probability of being selected (Kim, Yan, et al., 2015; Kim et al., 2016). Alternatively, we can assign the type 1 request to one of the free slots in a partially occupied – that is, non-empty and non-full – superchannel that contains either the lowest number of type 1 flows or the highest number of type 1 flows; the former is called the *Least-Filled* (LF) policy, while the latter is called the *Most-Filled* (MF) policy. For both of these policies, if all superchannels are empty or full, the type 1 request is assigned to one of the slots in an empty superchannel – provided they are not all full of course. Throughout the remainder, we use AP to denote a generic Allocation Policy.

If no (super)channel of the required size is available, the incoming traffic request is blocked. For each of the traffic flows, we are interested in the number of traffic requests that are blocked, and in the (long-term) proportion of the traffic requests that are blocked. Determining these key performance indicators is our main objective in this chapter.

## 8.2 Exact models

Kim, Yan, et al. (2015) construct an exact model for the random allocation policy, and we extend their model to the least-filled and most-filled policies: we introduce the exact state description in Section 8.2.1, and define a homogeneous and Markovian jump process model of the optical link for each of the allocation policies in Section 8.2.2↷. In Sections 8.2.3[412] and 8.2.4[417], we explain how we can use these models to compute the (limit) blocking ratios.

### 8.2.1 Exact state space

The state description should allow us to (i) determine whether or not an incoming flow is blocked or allowed into the system, and (ii) accurately model the allocation of the flows and the completion of their holding times. Due to the memorylessness of the exponential distribution, and because the three particular allocation policies that we consider only use the number of type 1 flows in each superchannel – and not the order of the superchannels along the grid or the order of the flows in a superchannel – a sufficiently detailed state description is $(a_0, a_1, \dots, a_{n_2})$, where $a_k$ counts the number of superchannels occupied with $k$ type 1 flows and no type 2 flows. Observe that the number of superchannels that are occupied by a type 2 flow is $b := m_2 - \sum_{k=0}^{n_2} a_k$. Because there are $m_2$ superchannels in total, the state $(a_0, a_1, \dots, a_{n_2})$ should satisfy $\sum_{k=0}^{n_2} a_k \le m_2$, so the detailed state description yields the state space

$$\mathcal{X} := \left\{ (a_0, \dots, a_{n_2}) \in \mathbb{Z}_{\ge 0}^{n_2+1} : \sum_{k=0}^{n_2} a_k \le m_2 \right\}.$$

The total number of states exhibits an $O(m_2^{n_2}) = O((m_1/n_2))^{n_2})$ dependency on the total number of slots $m_1$ and on the number of slots contained in a superchannel $n_2$, so the detailed state description suffers from state space explosion. To verify this, we plot the number $|\mathcal{X}|$ of states in $\mathcal{X}$ as a function of $m_1$ and $n_2$ in Fig. 8.3.

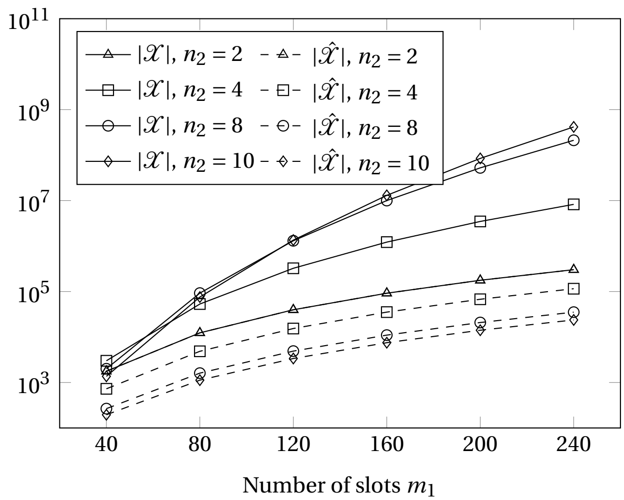



**Figure 8.3** The number of states in the detailed state space $\mathcal{X}$ and the reduced state space $\hat{\mathcal{X}}$ as a function of $m_1$ and $n_2$.

### 8.2.2 Exact models

Because we assume Poisson arrival processes and exponentially distributed service times, and also because the allocation policies that we consider only use the number of allocated slots per superchannel, we can accurately model the optical link with a homogeneous and Markovian jump process (see Norris, 1997, Section 2.6). To determine the defining rate operator $Q_{\text{AP}}$, we suppose that the system is in the state $x := (a_0, \dots, a_{n_2})$ in $\mathcal{X}$. The state of the system can change for four reasons: (i) an arrival of a traffic flow of type 1; (ii) a departure of a traffic flow of type 1; (iii) an arrival of a traffic flow of type 2; (vi) a departure of a traffic flow of type 2. Due to our assumptions regarding the arrival processes and the service times, the probability that these occur at the same time is zero, so we can focus on each of these separately.

**Arrival of a type 1 traffic flow** If all slots are occupied, so if $a_k = 0$ for all $k$ in $\{0, \dots, n_2 - 1\}$, then the request is blocked. Alternatively, the request is allocated to one of the free slots according to the allocation policy. The random allocation policy allocates the request at random to one of the free slots, so the state after this allocation is

$$x_{1,+}^k := (a_0, \dots, a_k - 1, a_{k+1} + 1, \dots, a_{n_2}),$$

where $k$ is one of the indices in $\mathcal{K}_x^+ := \{k \in \{0, \dots, n_2 - 1\} \colon a_k > 0\}$. The slots are chosen at random and with uniform probability, so it follows from the properties of the Poisson process that each of these possible transitions occurs with rate

$$\lambda_{\mathrm{RA}}^k := \lambda_1 \frac{a_k (n_2 - k)}{\sum_{\ell=0}^{n_2-1} a_\ell (n_2 - \ell)}.$$

Under the least-filled and most-filled allocation policies, the new state is $x_{1,+}^{k_{\mathrm{AP}}}$, where the index $k_{\mathrm{AP}}$ in $\mathcal{K}_x^+$ is fixed but depends on the policy $AP$. If all the superchannels are either completely free or completely occupied, so if $a_k = 0$ for all $k$ in $\{1, \dots, n_2 - 1\}$, then both policies assign the request to an empty superchannel: $k_{\mathrm{LF}} := 0 =: k_{\mathrm{MF}}$; if at least one of the superchannels is partially occupied, then the policies allocate the request in the superchannel that is the least filled or most-filled, so

$$k_{\mathrm{LF}} := \min\bigl\{k \in \{1, \dots, n_2 - 1\} \colon a_k > 0\bigr\}$$

and

$$k_{\mathrm{MF}} := \max\bigl\{k \in \{1, \dots, n_2 - 1\} \colon a_k > 0\bigr\}.$$

Hence, with AP equal to LF or MF, for all $k$ in $\mathcal{K}_x^+$, we let

$$\lambda_{\mathrm{AP}}^k := \begin{cases} \lambda_1 & \text{if } k = k_{\mathrm{AP}}. \\ 0 & \text{otherwise.} \end{cases}$$

Note that in case $n_2 = 2$, $k_{\mathrm{LF}} = k_{\mathrm{MF}}$.

**Departure of a type 1 traffic flow** The holding time of each of the allocated type 1 traffic flows can expire, so the state $x$ can change to

$$x_{1,-}^k := (a_0, \dots, a_{k-1} + 1, a_k - 1, \dots, a_{n_2}),$$

where $k$ is an index in $\mathscr{K}_x^- := \{k \in \{1,\dots,n_2\}\colon a_k > 0\}$. There are $k$ allocated flows in each of the $a_k$ superchannels, and the holding time of each of these flows is (independently) distributed with rate $\mu_1$, so it follows from the properties of the exponential distribution that the transition to $x_{1,-}^k$ occurs with rate $ka_k\mu_1$.

**Arrival of a type 2 request** If all superchannels are (partially) occupied, so if $a_0 = 0$, then the incoming request is blocked. Alternatively, the request is allocated to one of the free superchannels, so the state $x$ changes to

$$x_{2,+} := (a_0 - 1, a_1, \dots, a_{n_2}).$$

As type 2 flows arrive according to a Poisson process with rate $\lambda_2$, this transition occurs with the same rate.

**Departure of a type 2 traffic flow** There are $b_x := m_2 - \sum_{k=0}^{n_2} a_k$ allocated type 2 flows, and the holding time of each of the allocated type 2 traffic lows can expire. Hence, if $b_x > 0$, the state $x$ can change to

$$x_{2,-} := (a_0 + 1, a_1, \dots, a_{n_2});$$

the holding time of each of these flows is (independently) exponentially distributed with rate $\mu_2$, so it follows from the properties of the exponential distribution that the transition to $x_{2,-}$ occurs with rate $b_x\mu_2$.

**Policy-dependent homogeneous and Markovian jump process models**

For each of the three allocation policies RA, LF and MF, we have enumerated the non-zero off-diagonal components of the (matrix representation of) the rate operator $Q_{\mathrm{AP}}$; see also the schematic depiction in Fig. 8.4. Thus, we have fully defined the rate operator $Q_{\mathrm{AP}}$: for all $f$ in $\mathbb{G}(\mathscr{X})$ and $x = (a_0, \dots, a_{n_2})$ in $\mathscr{X}$,

$$\begin{aligned}\left[Q_{\mathrm{AP}} f\right](x) &= \lambda_2\big(f(x_{2,+}) - f(x)\big) + b_x\mu_2\big(f(x_{2,-}) - f(x)\big) \\ &\quad + \sum_{k\in\mathscr{K}_x^+} \lambda_{\mathrm{AP}}^k\big(f(x_{1,+}^k) - f(x)\big) + \sum_{k\in\mathscr{K}_x^-} a_k\mu_1\big(f(x_{1,-}^k) - f(x)\big),\end{aligned} \tag{8.1}$$

where the first term is only added if $a_0 > 0$ and the second term is only added if $b_x = m_2 - \sum_{k=1}^{n_2} a_k > 0$. The sum of the $\lambda_{\mathrm{AP}}^k$'s is always $\lambda_1$, so the diagonal components $Q_{\mathrm{AP}}(x,x)$ are equal for every allocation policy; due to (R5)$_{81}$, this implies that $\|Q_{\mathrm{AP}}\|_{\mathrm{op}}$ is the same for every allocation policy.

To end up with a homogeneous and Markovian jump process model, we should also specify an initial probability mass function $p_0$ on $\mathscr{X}$. The initial probability mass function does not play a role in our analysis further on, so we can fix any arbitrary $p_0$ in $\Sigma_{\mathscr{X}}$. Thus, we have defined three homogeneous and Markovian jump process models: $P_{\mathrm{RA}} := P_{p_0,Q_{\mathrm{RA}}}$, $P_{\mathrm{LF}} := P_{p_0,Q_{\mathrm{LF}}}$ and $P_{\mathrm{MF}} := P_{p_0,Q_{\mathrm{MF}}}$.

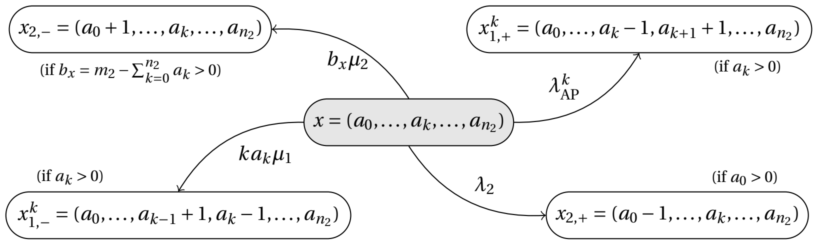



**Figure 8.4** State transition diagram for the rate operator $Q_{\mathrm{AP}}$ corresponding to an allocation policy AP.

### A policy-independent homogeneous and Markovian imprecise jump process model

While this is already nice, we can go even further. Henceforth, we let $\mathscr{Q}$ be the set of rate operators $Q_{\mathrm{AP}}$ in $\mathfrak{Q}_{\mathscr{X}}$ that are of the form in Eq. (8.1)↶, where the rates $\lambda^k_{\mathrm{AP}}$ are non-negative and sum to $\lambda_1$ but may differ for different states $x$. Note that $\mathscr{Q}$ is bounded and convex and has separately specified rows. Here too, our uncertainty regarding the initial state does not influence our analysis further on; hence, we let $\mathscr{M} := \Sigma_{\mathscr{X}}$.

For each of the three allocation policies RA, LF and MF, the rate operator $Q_{\mathrm{AP}}$ obviously belongs to $\mathscr{Q}$, so the homogeneous and Markovian imprecise jump process $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ contains the corresponding homogeneous and Markovian jump process model $P_{\mathrm{AP}}$. Furthermore, $\mathscr{Q}$ includes a rate operator $Q_{\mathrm{AP}}$ for *any* allocation policy AP that depends on the number of type 1 flows per superchannel but not on the order of the superchannels along the grid, so $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ contains a jump process model for any such allocation policy. Even more, $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ includes a – possibly lumped – jump process model for every allocation policy AP that allocates incoming traffic whenever there is a free slot or superchannel in the grid, including time-dependent policies or policies that depend on the order of the allocated flows along the grid!

In the remainder, we will compute lower and upper expectations corresponding to $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, so we need the lower envelope $\underline{Q} := \underline{Q}_{\mathscr{Q}}$ of $\mathscr{Q}$ as defined by Eq. (8.2). It follows immediately from our definition of $\mathscr{Q}$ and Eq. (8.1)↶ that, for all $f$ in $\mathbb{G}(\mathscr{X})$ and $x = (a_0,\dots,a_{n_2})$ in $\mathscr{X}$,

$$\begin{aligned}
[\underline{Q}f](x) &= \inf\{[Q_{\mathrm{AP}}f](x) : Q_{\mathrm{AP}} \in \mathscr{Q}\} \\
&= \lambda_2\big(f(x_{2,+}) - f(x)\big) + b_x\mu_2\big(f(x_{2,-}) - f(x)\big) \\
&\quad + \min\{\lambda_1\big(f(x^k_{1,+}) - f(x)\big) : k \in \mathscr{K}^+_x\} \\
&\quad + \sum_{k\in\mathscr{K}^-_x} a_k\mu_1\big(f(x^k_{1,-}) - f(x)\big),
\end{aligned} \tag{8.2}$$

where the first term is only added if $a_0 > 0$ and the second term is only added if $b_x = m_2 - \sum_{k=1}^{n_2} a_k > 0$. Note that $[\underline{Q}\mathbb{I}_x](x) = Q_{\mathrm{AP}}(x,x)$ for any $x$ in $\mathscr{X}$ and any allocation policy AP; hence, it follows from (R5)$_{81}$ and (LR7)$_{111}$ that $\|\underline{Q}\|_{\mathrm{op}} = \|Q_{\mathrm{AP}}\|_{\mathrm{op}}$, with $AP$ equal to RA, LF or MF.

### 8.2.3 The expected number of blocked requests

Recall from Section 8.1.2$_{406}$ that an important key performance indicator for the allocation policies is the number of blocked requests. This number is uncertain, of course, so it corresponds to a variable on $\Omega$. Let us formally define this variable so that we can determine the *expected* number of blocked requests.

It is imperative for our formal definition that we have assumed Poisson arrival processes. The reason for this is the incredible and well-known property of (measure-theoretic) jump processes that Poisson Arrivals See Time Averages (PASTA) (Wolff, 1982). Fix a jump process $P$ in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, an initial state $x$ in $\mathscr{X}$, a subset $B$ of $\mathscr{X}$, a time point $r$ in $\mathbb{R}_{>0}$ and a type $k$ in $\{1,2\}$. Wolff (1982, Eqn. (3)) proves that the expected number of type $k$ requests that arrive to the system during the time period $[0,r]$ when its state is in $B$ is

$$\lambda_k E_P^{\mathrm{D}}\left(\int_0^r \mathbb{I}_B(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right) = \lambda_k r E_P^{\mathrm{D}}\left(\frac{1}{r}\int_0^r \mathbb{I}_B(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right).$$

Note that, again due to the assumption of Poisson arrival processes, $\lambda_k r$ is the expected number of type $k$ requests over the time period $[0,r]$.

A type $k$ request is blocked if it arrives when the system's state is in $B_k$, where we denote the set of all states that have no free slot by

$$B_1 := \left\{(a_0,\dots,a_{n_2}) \in \mathscr{X} \colon \big(\forall k \in \{0,\dots,n_2-1\}\big)\ a_k = 0\right\}$$

and the set of all states that have no free superchannel by

$$B_2 := \left\{(a_0,\dots,a_{n_2}) \in \mathscr{X} \colon a_0 = 0\right\}.$$

Hence, the expected number of type $k$ requests that are blocked over $[0,r]$ conditional on $\{X_0 = x\}$ is

$$\lambda_k r E_P^{\mathrm{D}}\left(\frac{1}{r}\int_0^r \mathbb{I}_{B_k}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right) = \lambda_k r E_P^{\mathrm{D}}\big(\llbracket \mathbb{I}_{B_k}\rrbracket_{[0,r]} \,\big|\, X_0 = x\big).$$

In particular, we are interested in the expected number of blocked requests for the random, least-filled and most-filled allocation policies. That is, for all $x$ in $\mathscr{X}$ and $r$ in $\mathbb{R}_{>0}$ and with AP equal to RA, LF or MF and $k$ equal to 1 or 2, we can determine

$$\beta_{[0,r]}^{k,\mathrm{AP}}(x) := \lambda_k r E_{P_{\mathrm{AP}}}^{\mathrm{D}}\big(\llbracket \mathbb{I}_{B_k}\rrbracket_{[0,r]} \,\big|\, X_0 = x\big). \tag{8.3}$$

More generally, we can determine policy-independent bounds on the expected number of blocked requests, or more exactly, bounds on the expected number of blocked requests that hold for *any* allocation policy: for all $x$ in $\mathscr{X}$, $r$ in $\mathbb{R}_{>0}$ and $k$ in $\{1,2\}$, the bounds on the expected number of type $k$ requests that are blocked over $[0,r]$ conditional on $\{X_0 = x\}$ are

$$\underline{\beta}^{k}_{[0,r]}(x) := \lambda_k r \underline{E}_{\mathscr{M},\mathscr{Q}}\big(\llbracket \mathbb{I}_{B_k} \rrbracket_{[0,r]} \,\big|\, X_0 = x\big)$$

and

$$\overline{\beta}^{k}_{[0,r]}(x) := \lambda_k r \overline{E}_{\mathscr{M},\mathscr{Q}}\big(\llbracket \mathbb{I}_{B_k} \rrbracket_{[0,r]} \,\big|\, X_0 = x\big).$$

### Computing the expected number of blocked requests

To determine the expected number of blocked requests, we essentially use the iterative method in Theorem 6.50$_{318}$. In contrast to what we did in Power Network Example 6.51$_{319}$ and Queueing Network Examples 7.35$_{367}$ and 7.39$_{370}$, we will not double the number of computations $n$ until we observe convergence. Instead, to compute the expected number of blocked requests $\beta^{k,\mathrm{AP}}_{[0,r]}(x)$, we simply use $2^5 = 32$ times the minimal number of iterations

$$n_{\min} = \left\lceil \frac{r \|Q_{\mathrm{AP}}\|_{\mathrm{op}}}{2} \right\rceil,$$

and similarly for the lower and upper expected number of blocked requests $\underline{\beta}^{k}_{[0,r]}(x)$ and $\overline{\beta}^{k}_{[0,r]}(x)$ – now with $\|\underline{Q}\|_{\mathrm{op}}$ instead of $\|Q_{\mathrm{AP}}\|_{\mathrm{op}}$, but since these coincide, that makes no difference. This appears to suffice for convergence for any initial state $x$ in $\mathscr{X}$ in most of the scenarios that we consider.

### Numerical experiments

We use the aforementioned method to compute the (lower and upper) expected number of blocked requests for a set of scenarios. First, we select four distinct systems by specifying $m_1$ and setting $n_2 = 4$. Each combination of $m_1$ and $n_2$ results in a state space $\mathscr{X}$, the size of which is reported in Table 8.1$_{\curvearrowright}$. For every system defined this way, we let $\mu_1 = 1$, $\mu_2 = 1$, $\lambda_1 = \rho\mu_1$ and $\lambda_2 = \rho\mu_2$, where the positive real number $\rho$ is the so-called *traffic load*. The performance of the aforementioned method – and of all the other computational methods that we will introduce further on – varies with the traffic load $\rho$, which is why for every system we consider a low, medium and high traffic load, as listed in Table 8.1$_{\curvearrowright}$. Note that the traffic loads are proportional to the number of slots $m_1$ to account for the extra capacity in the larger systems. Whenever we need an initial state, we will use one of the following three 'extreme' options: the state $x_{\mathrm{e}} := (m_2, 0, 0, 0, 0)$ that corresponds to a completely empty grid, the state $x_{\mathrm{f},1} := (0, 0, 0, 0, m_2)$ that corresponds to a grid where all superchannels are fully occupied by type 1 flows and the state $x_{\mathrm{f},2} := (0, 0, 0, 0, 0)$ that corresponds to a grid where all superchannels are occupied by type 2 flows.

In a first experiment, we determine the expected number of blocked requests for both types of traffic flows over the time period $[0,r]$ with $r =$

**Table 8.1** Parameters used in the numerical experiments.

| $m_1$ | $n_2$ | $\lvert\mathscr{X}\rvert$ | $\lvert\hat{\mathscr{X}}\rvert$ | $\rho_{\text{low}}$ | $\rho_{\text{med}}$ | $\rho_{\text{hi}}$ |
|---|---|---|---|---|---|---|
| 40 | 4 | 3 003 | 726 | 2 | 10 | 50 |
| 80 | 4 | 53 130 | 4 851 | 4 | 20 | 100 |
| 120 | 4 | 324 632 | 15 376 | 6 | 30 | 150 |
| 160 | 4 | 1 221 759 | 35 301 | 8 | 40 | 200 |

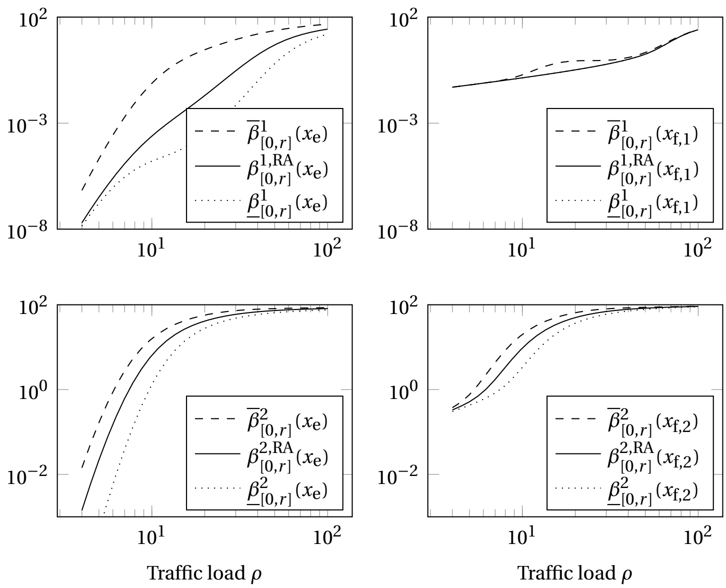



**Figure 8.5** Expected number of blocked requests for $m_1 = 80$ and $n_2 = 4$ for the random allocation policy, and policy-independent lower and upper bounds.

$^{100}/_{\rho}$; this value for $r$ is chosen such that for both types of traffic flows the expected number of requests during this period is 100. Let us take the system with $m_1 = 80$ and the random allocation policy as an example. In Fig. 8.5, we plot $\beta^{k,\text{RA}}_{[0,r]}(x)$, $\underline{\beta}^k_{[0,r]}(x)$ and $\overline{\beta}^k_{[0,r]}(x)$ as a function of $\rho$ over the range $[\rho_{\text{low}}, \rho_{\text{hi}}]$; for type 1 requests we consider the initial states $x_e$ and $x_{f,1}$, and for type 2 requests we consider the initial states $x_e$ and $x_{f,2}$. From these plots, we learn that the initial state heavily influences the expected number of blocked requests, and that for large traffic loads and type 2 traffic flows, the expected number of blocked requests is close to the expected total number of requests. We could do the same for the least-filled and most-filled allocation policies or for different numbers $m_1$ of slots in the grid, but we choose not do this here;

the reason for this is simple: our aim is not to analyse the system in great detail, but rather to show the modelling power of Markovian imprecise jump processes, and in particular to illustrate that our lumping-based methods are useful when studying large-scale models.

For a follow-up experiment, we investigate the time required for the computations. We implement the aforementioned computation method in Julia, a high-performance programming language similar to Python or MatLab that is specifically suited to technical computing (Bezanson et al., 2012) – we do the same for all but one of the other computational methods that we will investigate further on. That is, for every system in Table 8.1, we measure the execution time of the computations that are required to determine the expected number of blocked type 1 and type 2 requests over $[0, r]$ with $r = {}^{100}/_{\rho}$: we fix a value for $m_1$ and a traffic load $\rho$, and time how long it takes to determine the expected number of blocked requests for all three allocations policies, and we do the same for the upper expected number of blocked requests for the policy-independent model $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. We report the median of the execution time over five consequent runs in Table 8.2; in order to facilitate comparing these execution times to those for the methods and models further on, we actually report the median execution time divided by the number of distinct models (here equal to 3). For larger systems, the

**Table 8.2** Median execution time (in seconds) to determine (an upper bound on) the expected number of blocked requests.

| | $P_{\mathrm{AP}}$ | | | $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ | | |
|---|---|---|---|---|---|---|
| $m_1$ | $\rho_{\mathrm{low}}$ | $\rho_{\mathrm{med}}$ | $\rho_{\mathrm{hi}}$ | $\rho_{\mathrm{low}}$ | $\rho_{\mathrm{med}}$ | $\rho_{\mathrm{hi}}$ |
| 40 | 5.56 | 1.51 | 0.78 | 66.98 | 19.78 | 10.12 |
| 80 | 147.24 | 39.60 | 14.95 | / | / | 196.68 |
| 120 | / | / | 108.62 | / | / | / |
| 160 | / | / | / | / | / | / |

execution time is rather large; whenever the 'average' (median) execution time is longer than 200 seconds, we say that the computations have timed out and denote this with a forward slash. For all the models, the execution time decreases for increasing traffic load, and they are (approximately) proportional to the number of states $|\mathscr{X}|$. Note that the median execution time for the policy-dependent models is significantly lower than the execution time for the policy-independent model. Since $\|Q_{\mathrm{AP}}\|_{\mathrm{op}} = \|\underline{Q}\|_{\mathrm{op}}$, the only explanation for this is that determining $\underline{Q}\tilde{f}_{n,k}$ takes longer than determining $Q_{\mathrm{AP}}\tilde{f}_{n,k}$ due to the optimisation.

For a final experiment, we recall from Eq. (8.3)$_{412}$ that the expected number of blocked type k requests in $[0, r]$ is

$$\beta^{k,\mathrm{AP}}_{[0,r]}(x) = \lambda_k r E^{\mathrm{D}}_{P_{\mathrm{AP}}}\big([\![\mathbb{I}_{B_k}]\!]_{[0,r]} \,\big|\, X_0 = x\big),$$

where $E^{\mathrm{D}}_{P_{\mathrm{AP}}}\big([\![\mathbb{I}_{B_k}]\!]_{[0,r]} \,\big|\, X_0 = x\big)$ is the expected temporal average of $\mathbb{I}_{B_k}(X_t)$ over $[0, r]$. We have seen before that expected temporal averages have the tendency to converge, and the present setting forms no exception. As before,

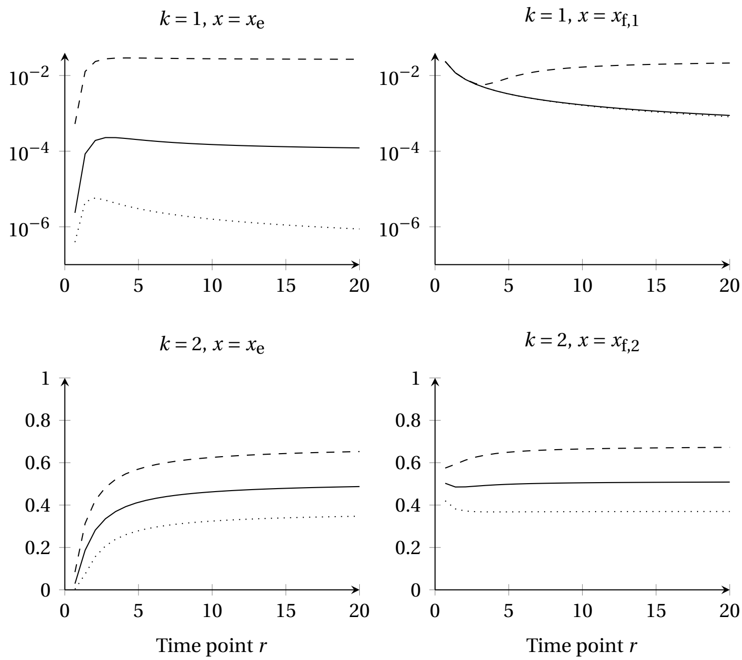



**Figure 8.6** Expectation of $[\![\mathbb{I}_{B_k}]\!]_{[0,r]}$ for $m_1 = 80$, $n_2 = 4$ and $\rho = \rho_{\mathrm{med}} = 20$ conditional on $\{X_0 = x\}$ for the random allocation policy. Dotted lines indicate the lower expectation corresponding to $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, and dashed lines the upper expectation.

we consider the system with $m_1 = 80$ and the random allocation policy as an example. In Fig. 8.6, we plot the expected temporal average

$$E^{\mathrm{D}}_{P_{\mathrm{RA}}}\big([\![\mathbb{I}_{B_k}]\!]_{[0,r]} \,\big|\, X_0 = x\big)$$

for $k$ equal to 1 or 2 and $x$ equal to one of the three 'extreme' initial states, and we also plot the lower and upper expectation of the same variable for the policy-independent model $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. As expected, the (lower and upper) expected temporal average converges, and the limit value does not depend on the initial state $x$. However, the rate of convergence varies in all four plots. Generally speaking, the convergence for type 2 flows is (considerably) faster than that for type 1 flows – one possible explanation for this is that the expected temporal average is much larger in the former case – and the convergence is faster for the initial state $x_{\mathrm{f},k}$ that corresponds to a fully-occupied grid than the initial state $x_{\mathrm{e}}$ that corresponds to an empty grid.

### 8.2.4 Blocking ratios

Why is it relevant that the expected temporal average of $\mathbb{I}_{B_k}(X_t)$ converges? The answer to this question is the Point-Wise Ergodic Theorem (Norris, 1997, Theorem 3.8.1). In Section 6.4.4<sub>305</sub>, we saw that this theorem implies that if the rate operator $Q_{\mathrm{AP}}$ is ergodic, then the long-term expected temporal average of $\mathbb{I}_{B_k}(X_t)$ converges to the limit expectation $E^{Q_{\mathrm{AP}}}_{\lim}(\mathbb{I}_{B_k})$ corresponding to $Q_{\mathrm{AP}}$:

$$\lim_{r\to+\infty} E^{\mathrm{D}}_{P_{\mathrm{AP}}}\big([\![\mathbb{I}_{B_k}]\!]_{[0,r]} \,\big|\, X_0 = x\big) = E^{Q_{\mathrm{AP}}}_{\lim}(\mathbb{I}_{B_k}) \quad \text{for all } x \in \mathscr{X}.$$

This means that *if* the rate operator $Q_{\mathrm{AP}}$ is ergodic, then we can approximate the expected number of blocked type $k$ requests over $[0,r]$ for sufficiently large $r$ by

$$\beta^{k,\mathrm{AP}}_{[0,r]}(x) \approx \lambda_k r E^{Q_{\mathrm{AP}}}_{\lim}(\mathbb{I}_{B_k}) \quad \text{for all } x \in \mathscr{X}. \tag{8.4}$$

As we have seen in Power Network Example 6.51<sub>319</sub>, this is similar to the heuristic that Troffaes et al. (2015) use to determine the expected downtime.

Of course, this is only useful if we can easily determine the corresponding limit expectation $E^{Q_{\mathrm{AP}}}_{\lim}(\mathbb{I}_{B_k})$. As is to be expected from Fig. 8.6↶, the rate operator $Q_{\mathrm{RA}}$ is irreducible – that is, ergodic with top class $\mathscr{X}_{Q_{\mathrm{RA}}} = \mathscr{X}$ – and the same holds for the rate operators $Q_{\mathrm{LF}}$ and $Q_{\mathrm{MF}}$. Proving this is not difficult, but it is *very* laborious. Most 'applied' researchers do not bother to (explicitly) check this, and we follow their example here; we refer the reader who is interested in a formal proof to (Erreygers, Rottondi, et al., 2018a, Appendix A). In fact, using a similar argument, one can show that every rate operator $Q$ in $\mathscr{Q}$ is irreducible.

But there is more! As explained previously in Section 6.4.4<sub>305</sub>, the Point-Wise Ergodic Theorem (Norris, 1997, Theorem 3.8.1) actually states that if $Q_{\mathrm{AP}}$ is irreducible (or even ergodic), then the event

$$\left\{\omega \in \Omega \colon \liminf_{r\to+\infty} \frac{1}{r}\int_0^r \mathbb{I}_{B_k}\big(\omega(t)\big)\,\mathrm{d}t = \limsup_{r\to+\infty} \frac{1}{r}\int_0^r \mathbb{I}_{B_k}\big(\omega(t)\big)\,\mathrm{d}t = E^{Q_{\mathrm{AP}}}_{\lim}(\mathbb{I}_{B_k})\right\}$$

has probability one – for a suitable extension of $P_{\mathrm{AP}}$. For this reason, it follows from (Wolff, 1982, Theorem 1) that for $k$ equal to 1 or 2, the (long-term) ratio[1]

[1] That is, as $r$ recedes to $+\infty$, the ratio of the number of blocked type $k$ requests that arrive over $[0,r]$ and the total number of type $k$ requests that arrive over $[0,r]$.

of type $k$ requests that is blocked is almost surely – or with probability one – equal to $E_{\lim}^{Q_{AP}}(\mathbb{I}_{B_k})$. Hence, for AP equal to RA, LF or MF, we call

$$\beta_{\lim}^{1,AP} := E_{\lim}^{Q_{AP}}(\mathbb{I}_{B_1}) \quad \text{and} \quad \beta_{\lim}^{2,AP} := E_{\lim}^{Q_{AP}}(\mathbb{I}_{B_2}) \tag{8.5}$$

the *blocking ratios*.

### Computing limit expectations

All of this is only relevant because we know how to compute limit expectations. Recall that in Section 7.3$_{373}$, we discussed two ways to compute limit expectations. Here, we look at these two ways from a more practical point of view.

**Solving the equilibrium condition** Recall from Section 7.3.2$_{376}$ that for any $f$ in $\mathbb{G}(\mathcal{X})$, $E_{\lim}(f) = \langle p_{\lim}, f\rangle$ with $p_{\lim}$ the unique probability mass function that corresponds to the limit expectation $E_{\lim}$. There, we explained that we can determine $p_{\lim}$ by solving a system of $|\mathcal{X}|$ linear equations corresponding to the equilibrium condition in Eq. (7.54)$_{376}$. Stewart (2009, Section 10.2) discusses several methods to solve this system of equations; for large state spaces with sparse rate matrices, he recommends iterative solvers. In our implementation, we solve this system of equations using one of the iterative methods that are implemented in the `IterativeSolvers.jl`[2] package for Julia and that are applicable to our setting: GMRES (Saad et al., 1986), BiCGStab($\ell$) (Sleijpen et al., 1993) and IDR($s$) (Sonneveld et al., 2009); the last of these three gives the best results in our application, so we will only use this one.

**Iteratively determining the limit expectation** As an alternative to solving the equilibrium condition, we use the iterative method in Algorithm 8.1$_{\curvearrowright}$, which is inspired by Lemma 7.48$_{379}$. This method iteratively determines the limit expectation $E_{\lim}(f)$ up to some relative tolerance $\phi$; that is, it determines lower and upper bounds on the limit expectation that are 'sufficiently tight', in the sense that the relative 'error'

$$\frac{\max g_n - \min g_n}{\frac{1}{2}|\max g_n + \min g_n|} = 2\frac{\max g_n - \min g_n}{|\max g_n + \min g_n|} = 2\frac{\|g_n\|_v}{|\max g_n + \min g_n|}$$

of the approximation is smaller than the tolerance $\phi$. Our implementation of Algorithm 8.1$_{\curvearrowright}$ in Julia uses a data format that is specifically suited for sparse matrices, which speeds up the computations significantly. For the parameters in Algorithm 8.1$_{\curvearrowright}$, we use $\Delta = 1.9/\|Q\|_{op}$,[3] $\phi = 1\cdot 10^{-3}$ and $n_{\max} = 1\cdot 10^6$, unless stated otherwise.

[2] https://github.com/JuliaLinearAlgebra/IterativeSolvers.jl

[3] Recall from (R5)$_{81}$ that the norm $\|Q\|_{op}$ of a rate operator $Q$ is equal to 2 times the maximum of the absolute value of the diagonal components.

**Algorithm 8.1:** Iteratively compute the limit expectation $E_{\text{lim}}(f)$

**Input:** An ergodic rate operator $Q$, a gamble $f$ on $\mathscr{X}$, a time step $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\|Q\|_{\text{op}} < 2$, a relative tolerance $\phi$ in $\mathbb{R}_{>0}$ and a maximum number of iterations $n_{\max}$ in $\mathbb{N}$.

**Output:** Lower and upper bounds on $E_{\text{lim}}(f)$

$n := 0$
$g_0 := f$
$\epsilon_{\text{rel}} := 2\|g_0\|_{\text{v}}/|\max g_0 + \min g_0|$
**while** $\epsilon_{\text{rel}} > \phi$ **and** $n < n_{\max}$ **do**
  $n := n+1$
  $g_n := g_{n-1} + \Delta Q g_{n-1}$
  $\epsilon_{\text{rel}} := 2\|g_n\|_{\text{v}}/|\max g_n + \min g_n|$
**return** $\min g_n, \max g_n$

**Generating a sample path** For large systems, it can be more efficient to determine an estimate for the blocking ratios through simulation – and the same could be said for the expected number of blocked requests. That is, it might be more interesting to 'generate' (part of) a sample path $\omega$ until some sufficiently large time point $r$, and then approximate the blocking ratios as

$$\beta_{\text{lim}}^{k,\text{AP}} \approx \frac{1}{r}\int_0^r \mathbb{I}_{B_k}\big(\omega(t)\big)\,\mathrm{d}t.$$

Note that the convergence of the estimate is only almost surely and that the quality of the estimate is contingent on the quality of the used random number generator. Here, we generate a sample path using the well-known *Stochastic Simulation Algorithm* (SSA) (Gillespie, 1977), and assess the accuracy of the approximation by means of the batch mean estimation method (Pawlikowski, 1990). By the PASTA property, we only have to observe the system at the arrival epochs of a Poisson process instead of keeping track of all the times in between jumps, so we can simulate the system by generating a sample path of the embedded jump chain (Norris, 1997, Section 2.2). Whenever the selected transition in this chain corresponds to the arrival of a type 1 or type 2 flow, we observe whether one of the two request types would be blocked at that instant. In our C implementation of the combination of the Gillespie algorithm and the batch mean estimation method, we use batches of $4\cdot 10^7$ arrivals and Rule 1 in (Pawlikowski, 1990) as a rule to determine the end of the burn-in period. After that, we simulate at least 5 and at most 50 batches, where a batch is complete after $4\cdot 10^7$ incoming requests. If the number of batches is in between the minimum and maximum number of batches, we pre-emptively stop the simulation if the relative (estimated) error – taken to be the width of the 95%-confidence interval divided by the mean, as proposed by Pawlikowski (1990, Eq. (12)) – is smaller than the tol-

**Table 8.3** Blocking ratios for the random allocation policy.

| | Algorithm 8.1↶ | | | IDR($s$) | | | SSA | | |
|---|---|---|---|---|---|---|---|---|---|
| $m_1$ | $\rho_{\text{low}}$ | $\rho_{\text{med}}$ | $\rho_{\text{hi}}$ | $\rho_{\text{low}}$ | $\rho_{\text{med}}$ | $\rho_{\text{hi}}$ | $\rho_{\text{low}}$ | $\rho_{\text{med}}$ | $\rho_{\text{hi}}$ |
| | | | | Type 1 | | | | | |
| 40 | $5.64 \cdot 10^{-6}$ | $2.16 \cdot 10^{-3}$ | $2.53 \cdot 10^{-1}$ | $5.64 \cdot 10^{-6}$ | $2.16 \cdot 10^{-3}$ | $2.53 \cdot 10^{-1}$ | $5.75 \cdot 10^{-6}$ | $2.17 \cdot 10^{-3}$ | $2.53 \cdot 10^{-1}$ |
| 80 | $2.29 \cdot 10^{-10}$ | $9.40 \cdot 10^{-5}$ | $2.30 \cdot 10^{-1}$ | $2.29 \cdot 10^{-10}$ | $9.41 \cdot 10^{-5}$ | $2.30 \cdot 10^{-1}$ | 0.0 | $9.44 \cdot 10^{-5}$ | $2.30 \cdot 10^{-1}$ |
| 120 | $1.18 \cdot 10^{-14}$ | $7.74 \cdot 10^{-6}$ | $2.22 \cdot 10^{-1}$ | $1.18 \cdot 10^{-14}$ | $7.75 \cdot 10^{-6}$ | $2.22 \cdot 10^{-1}$ | 0.0 | $7.54 \cdot 10^{-6}$ | $2.22 \cdot 10^{-1}$ |
| 160 | $3.62 \cdot 10^{-23}$ | $9.01 \cdot 10^{-7}$ | $2.17 \cdot 10^{-1}$ | $7.16 \cdot 10^{-19}$ | $9.02 \cdot 10^{-7}$ | $2.17 \cdot 10^{-1}$ | 0.0 | $8.96 \cdot 10^{-7}$ | $2.17 \cdot 10^{-1}$ |
| | | | | Type 2 | | | | | |
| 40 | $2.03 \cdot 10^{-3}$ | $5.63 \cdot 10^{-1}$ | $9.98 \cdot 10^{-1}$ | $2.03 \cdot 10^{-3}$ | $5.63 \cdot 10^{-1}$ | $9.99 \cdot 10^{-1}$ | $2.03 \cdot 10^{-3}$ | $5.63 \cdot 10^{-1}$ | $9.99 \cdot 10^{-1}$ |
| 80 | $1.63 \cdot 10^{-5}$ | $5.12 \cdot 10^{-1}$ | $9.99 \cdot 10^{-1}$ | $1.63 \cdot 10^{-5}$ | $5.12 \cdot 10^{-1}$ | 1.00 | $1.63 \cdot 10^{-5}$ | $5.12 \cdot 10^{-1}$ | 1.00 |
| 120 | $1.49 \cdot 10^{-7}$ | $4.85 \cdot 10^{-1}$ | $9.99 \cdot 10^{-1}$ | $1.49 \cdot 10^{-7}$ | $4.85 \cdot 10^{-1}$ | 1.00 | $1.48 \cdot 10^{-7}$ | $4.85 \cdot 10^{-1}$ | 1.00 |
| 160 | $1.44 \cdot 10^{-9}$ | $4.68 \cdot 10^{-1}$ | $9.99 \cdot 10^{-1}$ | $1.44 \cdot 10^{-9}$ | $4.68 \cdot 10^{-1}$ | 1.00 | 0.0 | $4.68 \cdot 10^{-1}$ | 1.00 |

erance $\phi = 1 \cdot 10^{-3}$. Note that because we use batches of $4 \cdot 10^7$ arrivals, this method will not give good estimates whenever the blocking ratio is smaller than $1 \cdot 10^{-7}$.

### Numerical experiments

Let us use the scenarios in Table 8.1$_{414}$ to assess the performance of these three computational methods. First, we use these methods to determine the blocking ratios for the random allocation policy and report the obtained values in Table 8.3. We obtain similar values with all three methods for the medium and high traffic loads, but not for the low traffic load. As we have predicted, the SSA method fails whenever the blocking ratio is small. Furthermore, for the largest two systems and for type 1 requests, the values obtained with Algorithm 8.1↶ do not agree with those obtained with the IDR($s$) solver; we suspect this has to do with numerical instabilities and/or round-off errors.

Next, we measure the execution time of the computations that are required to determine the blocking ratios for the scenarios in Table 8.1$_{414}$: we fix a numerical method, a value for $m_1$ and a traffic load $\rho$, and time how long it takes to determine the blocking ratios for all three allocations policies and the three traffic loads. We report the median of the execution time over five consequent runs in Table 8.4↷ where, as before, we report the median execution time divided by the number of distinct models (here equal to 3). Here too, we say that the computations have timed out whenever the 'average' (median) execution time for the other methods is longer than 200 seconds, and we denote this in Table 8.4↷ with a forward slash. We observe that for Algorithm 8.1↶, the execution time is roughly proportional with the number of states; this is as expected: because $Q_{\text{AP}}$ is sparse, one step for the iterative scheme is linear in the number of states. The execution times for the IDR($s$) method are also proportional to the number of states, but the exact values should be taken with a large grain of salt. For some reason that is not entirely clear to me, the IDR($s$) implementation uses all of the available cores auto-

**Table 8.4** Median execution time (in seconds) to determine exact blocking ratios.

| | Algorithm 8.1$_{419}$ | | | IDR($s$) | | | SSA | | |
|---|---|---|---|---|---|---|---|---|---|
| $m_1$ | $\rho_{\text{low}}$ | $\rho_{\text{med}}$ | $\rho_{\text{hi}}$ | $\rho_{\text{low}}$ | $\rho_{\text{med}}$ | $\rho_{\text{hi}}$ | $\rho_{\text{low}}$ | $\rho_{\text{med}}$ | $\rho_{\text{hi}}$ |
| 40 | 0.04 | 0.06 | 0.30 | 0.01 | 0.01 | 0.01 | 176.91 | 151.13 | 24.35 |
| 80 | 2.07 | 3.14 | 15.09 | 0.19 | 0.43 | 0.79 | 179.43 | 165.39 | 39.97 |
| 120 | 19.65 | 40.35 | 170.01 | 1.54 | 4.05 | 8.58 | 78.89 | 185.46 | 54.70 |
| 160 | 64.05 | / | / | 7.22 | 20.84 | 50.87 | 23.16 | 165.97 | 40.21 |

matically. Given that we run the computations on a CPU with 8 cores, and that – in our implementation – the other methods all use a single thread (or CPU core), it would make sense to multiply the median execution times for the IDR($s$) method by 8; doing so would result in execution times that have the same order of magnitude as those for Algorithm 8.1$_{419}$. The execution times for the two iterative methods (Algorithm 8.1$_{419}$ and IDR($s$)) increase with increasing traffic load, while the execution time for the SSA method is larger for low traffic loads than for high traffic loads. For the two largest systems and the low traffic load, the execution times for the SSA method do not follow this pattern; the cause for this is that the method stops pre-emptively because (almost) none of the $4 \cdot 10^7$ simulated requests in a batch are blocked. Compared to the SSA method, the two iterative methods are considerably faster for small values of $m_1$, but the SSA method has an execution time that does not change significantly with the model size, making it more suited for larger systems. However, the SSA method has two clear drawbacks. First and foremost, one has to choose the number of requests in a batch large enough such that blocking can occur, but this requires at least an estimate of the blocking ratio. Second, the resulting values for the blocking ratios are always an estimate, and the accuracy of this estimate can only be quantified with a confidence interval.

## 8.3 Lumping the exact models

The main point that we take away from Section 8.2$_{407}$ is that for large numbers of slots $m_1$, determining the expected number of blocked requests or the blocking ratio becomes infeasible because the state space $\mathscr{X}$ explodes. The sole exception to this is the SSA method, because this does scale for large traffic loads; however, that method only yields estimates, and no exact values. As we know from Chapter 7$_{337}$, one way to counteract state space explosion is to lump the states, or equivalently, to adopt a higher-level – or less informative – state description. This is precisely what Kim, Yan, et al. (2015) did for the homogeneous Markovian jump process model for the random allocation policy, although they never mention explicitly that they use a lumping strategy. We have studied lumping in great detail in Chapter 7$_{337}$, and the methods that we introduced there will allow us to formalise and adapt their intuitive approach. In Section 8.3.1, we introduce the lumped state space. Next, we explain how Kim, Yan, et al. (2015) obtain an 'approximate' lumped jump process in Section 8.3.2↷; there, we also extend their approach to the least filled and most filled allocation policies, and observe how this approximate lumped jump process reduces the computation time at the cost of accuracy. To mitigate the inaccuracy of the approximate models, we turn to lumped imprecise jump processes to describe the exact models in Section 8.3.3₄₂₈. We use these models to compute policy-dependent and policy-independent bounds on the expected number of blocked requests in Section 8.3.4₄₃₁, and we do the same for blocking ratios in Section 8.3.5₄₃₃.

### 8.3.1 The lumped state space

A higher-level or coarser state space description that still allows us to determine whether or not a request is blocked, is the triplet $(a,b,c)$. In this triplet, $a$ and $b$ count the number of type 1 and type 2 flows that are currently allocated, and $c$ counts the number of unoccupied superchannels. Thus, $a$, $b$ and $c$ should be non-negative integers such that $m_2 \leq a+b+c$ and $a+n_2(b+c) \leq m_1$. Note that the first inequality is not mentioned by Kim, Yan, et al. (2015, Section III. A.), but is nevertheless required to ensure that all superchannels are accounted for. This way, we obtain the lumped state space

$$\hat{\mathscr{X}} \coloneqq \left\{(a,b,c) \in \mathbb{Z}^3_{\geq 0} \colon m_2 \leq a+b+c, a+n_2(b+c) \leq m_1\right\}.$$

The number of lumps has a $O(m_1 m_2^2) = O(m_1 (m_1/n_2)^2)$ dependency on the number of slots $m_1$ and the number of slots that form a superchannel $n_2$, so while $|\mathscr{X}|$ increases with $n_2$, $|\hat{\mathscr{X}}|$ actually decreases with $n_2$! We can confirm this by looking at Fig. 8.3₄₀₈: the number of states $|\mathscr{X}|$ grows rapidly with increasing $m_1$, whereas the number of lumped states $|\hat{\mathscr{X}}|$ is at least an order of magnitude smaller whenever $n_2 > 2$. Moreover, the cardinality of the lumped state space $\hat{\mathscr{X}}$ grows at a lower rate than the cardinality of the detailed state space $\mathscr{X}$ for $n_2 > 2$. Finally, we observe that in case $n_2 = 2$, the number of states in $\mathscr{X}$ is equal to the number of lumped states in $\hat{\mathscr{X}}$; to be more exact, there is a bijection from $\mathscr{X}$ to $\hat{\mathscr{X}}$: the state $x = (a_0, a_1, a_2)$ in $\mathscr{X}$ corresponds to the lumped state $(a_1 + 2a_2, m_2 - a_0 - a_1 - a_2, a_0)$ in $\hat{x}$, and conversely, a lumped state $\hat{x} = (a,b,c)$ in $\hat{\mathscr{X}}$ corresponds to the state $(c, 2(m_2 - b - c) - a, a+b+c-m_2)$ in $\mathscr{X}$.

Recall from Section 7.1.1₃₄₂ that the lumping map $\Lambda\colon \mathscr{X} \to \hat{\mathscr{X}}$ relates the detailed state description to the higher-level state description. For any state $(a_0, \ldots, a_{n_2})$ in $\mathscr{X}$, $a_k$ counts the number of superchannels that contain $k$ type 1 flows and no type 2 flows. Hence, the lumping map $\Lambda$ maps the state $(a_0, \dots, a_{n_2})$ in $\mathscr{X}$ to the lumped state

$$\Lambda(a_0, \dots, a_{n_2}) := \left( \sum_{k=1}^{n_2} k a_k, m_2 - \sum_{k=0}^{n_2} a_k, a_0 \right)$$

in $\hat{\mathscr{X}}$. As observed above, the lumping map $\Lambda$ is invertible whenever $n_2 = 2$: in that case

$$\Lambda^{-1}(a, b, c) = \{(c, 2(m_2 - b - c) - a, a + b + c - m_2)\} \quad \text{for all } (a, b, c) \in \hat{\mathscr{X}}.$$

### 8.3.2 Approximate models

In general, the coarser state description in the lumped state space $\hat{\mathscr{X}}$ is *not* sufficiently detailed to model the system as a homogeneous and Markovian jump process with state space $\hat{\mathscr{X}}$ – the case $n_2 = 2$ is the sole exception. Nonetheless, for the random allocation policy, Kim, Yan, et al. (2015) construct an *approximate* model with state space $\hat{\mathscr{X}}$. We repeat their construction here, and extend it to the least-filled and most-filled allocation policies.

To determine the defining rate operator $\tilde{Q}_{\mathrm{AP}}$, we focus on an arbitrary lumped state $\hat{x} := (a, b, c)$ in $\hat{\mathscr{X}}$. As in Section 8.2.2$_{408}$, we investigate what happens for the four possible changes in the system.

**Arrival of a type 1 traffic flow** If all slots are occupied, so if $a + n_2 b = m_1$, then the request is blocked. Alternatively, the request is allocated to one of the free slots according to the allocation policy. Note that a type 1 request can be assigned to a slot in either a non-full superchannel that already contains some type 1 flows or in a completely free superchannel, so the lumped state can change to

$$\hat{x}_{1,+}^{=} := (a + 1, b, c) \quad \text{or} \quad \hat{x}_{1,+}^{-} := (a + 1, b, c - 1).$$

For this reason, we need to distinguish three subcases. If there is no empty superchannel, so if $c = 0$, then regardless of the spectrum allocation policy, the type 1 request is assigned to a free slot in a non-empty superchannel. This corresponds to a transition to $\hat{x}_{1,+}^{=}$, which in this case has rate $\lambda_{\mathrm{AP}}^{=} = \lambda_1$; note that in this case $\hat{x}_{1,+}^{-}$ does not belong to $\hat{\mathscr{X}}$. Conversely, if every superchannel that is neither empty nor occupied by a type 2 flow is completely occupied by type 1 flows, so if $a + n_2(b + c) = m_1$, then a type 1 request is always assigned to an empty superchannel. This corresponds to a transition to $\hat{x}_{1,+}^{-}$, which in this case has rate $\lambda_{\mathrm{AP}}^{-} = \lambda_1$; note that in this case, $\hat{x}_{1,+}^{=}$ does not belong to $\hat{\mathscr{X}}$. Finally, we have the remaining case that $c > 0$ and $a + n_2(b + c) < m_1$. The two lumped states $\hat{x}_{1,+}^{=}$ and $\hat{x}_{1,+}^{-}$ are then both feasible, and the rates $\lambda_{\mathrm{AP}}^{=}$ and $\lambda_{\mathrm{AP}}^{-}$ of the transitions to these states depend on the used allocation policy. Regardless of the policy, however, both of these rates should be non-negative and their sum should equal $\lambda_1$; for this reason, we can focus on determining $\lambda^-_{AP}$ because $\lambda^=_{AP} = \lambda_1 - \lambda^-_{AP}$. The random allocation policy allocates the request at random to one of the $m_1 - a - n_2 b$ free slots with equal probability, so

$$\lambda^-_{RA} := \lambda_1 \frac{n_2 c}{m_1 - a - n_2 b};$$

see also (Kim, Yan, et al., 2015). The least-filled and most-filled policies only assign a type 1 request to an empty superchannel if all non-empty superchannels containing type 1 requests are completely occupied; this is not the case here, so $\lambda^-_{LF} := 0 =: \lambda^-_{MF}$.

**Departure of a type 1 traffic flow** If $a > 0$, then the holding time of each of the allocated type 1 traffic flows can expire. This might might free up a superchannel or not, depending on whether or not it was the sole type 1 flow in its superchannel; that is, $\hat{x}$ can change to

$$\hat{x}^=_{1,-} := (a-1, b, c) \quad \text{or} \quad \hat{x}^+_{1,-} := (a-1, b, c+1).$$

Again, we need to distinguish three subcases. If every superchannel that is neither empty nor occupied by a type 2 flow only contains a single type 1 flow, so if $a + b + c = m_2$, then the departure of a type 1 flow always frees a superchannel. This corresponds to a transition to $\hat{x}^+_{1,-}$, which in this case has rate $\mu^+ := a\mu_1$; note that in this case $\hat{x}^=_{1,-}$ is not a feasible lumped state. Conversely, if every superchannel that is neither empty nor occupied by a type 2 flow contains at least two type 1 flows, so if $a \geq n_2(m_2 - b - c - 1) + 2$, then the departure of a type 1 flow will never free a superchannel. This corresponds to a transition to $\hat{x}^=_{1,-}$, which in this case has rate $\mu^= := a\mu_1$; note that in this case $\hat{x}^+_{1,-}$ is not a feasible lumped state. Unfortunately, in the remaining case that $m_2 - b - c < a < n_2(m_2 - b - c - 1) + 2$, the coarser state description is not sufficiently informative to capture the distribution of the allocated type 1 flows across the superchannels, at least not in general. For this reason, we cannot determine the rate $\mu^=$ of the transition to $\hat{x}^=_{1,-}$ and the rate $\mu^+$ of the transition to $\hat{x}^+_{1,-}$. In Section 8.3.3428 further on, we deal with this indeterminacy through lumping. For now, we stick to the solution put forward by Kim, Yan, et al. (2015): they use an estimate $\tilde{\mu}^+$ for the rate $\mu^+$, which is based on the assumption – or, more precisely, on the approximation – that all the possible distributions of the $a$ type 1 flows over the $m_2 - b - c$ non-empty superchannels that contain type 1 flows are 'equally probable'. After some combinatorics, they find that

$$\tilde{\mu}^+ := \mu_1 n_2 (m_2 - b - e) \frac{C(m_2 - b - c - 1, a - 1)}{C(m_2 - b - c, a)},$$

where, for all natural numbers $n, k$ such that $n \leq k \leq n_2 n$,

$$C(n,k) := \sum_{i=0}^{n - \lceil k/n_2 \rceil} (-1)^i \binom{n}{i} \binom{(n-i)n_2}{k}$$

is the number of $n$-tuples $(a_1, \ldots, a_n)$ of natural numbers such that $a_j \le n_2$ for all $j$ in $\{1, \ldots, n\}$ and $\sum_{j=1}^{n} a_j = k$. The corresponding estimate for $\mu^=$ is $a\mu_1 - \tilde{\mu}^+$.

**Arrival of a type 2 request** If all superchannels are (partially) occupied, so if $c = 0$, then the incoming request is blocked. Alternatively, the request is allocated to one of the free superchannels, so the state $\hat{x}$ changes to

$$\hat{x}_{2,+} := (a, b+1, c-1),$$

and this transition occurs with rate $\lambda_2$.

**Departure of a type 2 traffic flow** If $b > 0$, then the expiration of the holding time of an allocated type 2 flow corresponds to a transition to

$$\hat{x}_{2,-} := (a, b-1, c+1),$$

and this transition occurs with rate $b\mu_2$.

### Policy-dependent approximate homogeneous and Markovian jump process models

We schematically depict the non-zero off-diagonal components (of the matrix representation) of the rate operator $\tilde{Q}_{\mathrm{AP}}$ in Fig. 8.7. These components fully

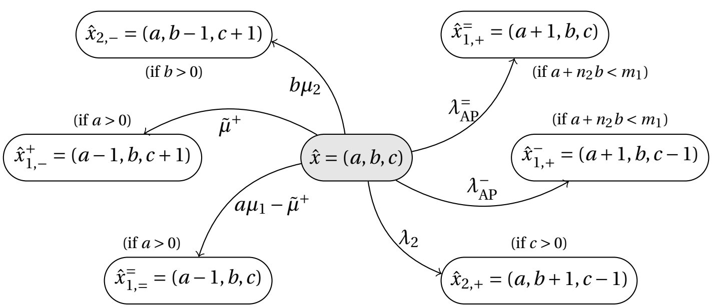



**Figure 8.7** State transition diagram for the approximate rate operator $\tilde{Q}_{\mathrm{AP}}$

define the rate operator $\tilde{Q}_{\mathrm{AP}}$: for any $\hat{f}$ in $\mathbb{G}(\hat{\mathscr{X}})$ and $\hat{x}$ in $\mathscr{X}$,

$$\begin{aligned}\left[\tilde{Q}_{\mathrm{AP}}\hat{f}\right](\hat{x}) = \lambda_2\big(f(\hat{x}_{2,+}) - \hat{f}(\hat{x})\big) + b\mu_2\big(\hat{f}(\hat{x}_{2,-}) - \hat{f}(\hat{x})\big)\\ + \lambda^-_{\mathrm{AP}}\big(\hat{f}(\hat{x}^-_{1,+}) - \hat{f}(x)\big) + \lambda^=_{\mathrm{AP}}\big(\hat{f}(\hat{x}^=_{1,+}) - \hat{f}(x)\big)\\ + \tilde{\mu}^+\big(\hat{f}(\hat{x}^+_{1,-}) - \hat{f}(x)\big) + (a\mu_1 - \tilde{\mu}^+)\big(\hat{f}(\hat{x}^=_{1,-}) - \hat{f}(x)\big),\end{aligned} \tag{8.6}$$

where each of the terms in this expression is only added if the relevant lump belongs to $\hat{\mathscr{X}}$. Note that $\lambda_{\mathrm{LF}}^{-} = \lambda_{\mathrm{MF}}^{-}$, and therefore $\tilde{Q}_{\mathrm{LF}} = \tilde{Q}_{\mathrm{MF}} =: \tilde{Q}_{\mathrm{LM}}$.

The initial probability mass function plays no role in our analysis further on, so we could fix any probability mass function $\tilde{p}$ on $\hat{\mathscr{X}}$. However, we have already 'fixed' a probability mass function $p_0$ on $\mathscr{X}$ before, so we may as well use the corresponding lumped probability mass function $\hat{p} := \hat{p}_{p_0}$. This way, we have defined an approximate homogeneous and Markovian jump process model for each of the three allocation policies; to ease our notation, we denote these models by $\tilde{P}_{\mathrm{RA}} := P_{\hat{p},\tilde{Q}_{\mathrm{RA}}}$ and $\tilde{P}_{\mathrm{LM}} := P_{\hat{p},\tilde{Q}_{\mathrm{LM}}}$.

The idea is to use these approximate models to, well, approximate the expected number of blocked requests and the blocking ratios; because the lumped state space $\hat{\mathscr{X}}$ is at least an order of magnitude smaller than the detailed state space $\mathscr{X}$, we expect that computing these approximations will be tractable for large values of $m_1$. To make this a bit more concrete, we observe that

$$\hat{B}_1 := \Lambda(B_1) = \left\{(a,b,c) \in \hat{\mathscr{X}} \colon a + n_2 b = m_1\right\}$$

and

$$\hat{B}_2 := \Lambda(B_2) = \left\{(a,b,c) \in \hat{\mathscr{X}} \colon c = 0\right\}.$$

The hope is now that because $\tilde{P}_{\mathrm{AP}}$ approximates 'the' lumped jump process corresponding to $P$, we can approximate the expected number of blocked type $k$ requests by

$$\lambda_k r E^{\mathrm{D}}_{\tilde{P}_{\mathrm{AP}}}\left(\llbracket \mathbb{I}_{\hat{B}_k} \rrbracket_{[0,r]} \,\middle|\, \hat{X}_0 = \Lambda(x)\right) \approx \beta^{k,\mathrm{AP}}_{[0,r]} \quad \text{for all } x \in \mathscr{X}.$$

Similarly, and under the assumption that $\tilde{Q}_{\mathrm{AP}}$ is irreducible, the hope is that we can approximate the blocking ratio for type $k$ requests by

$$E^{\tilde{Q}_{\mathrm{AP}}}_{\mathrm{lim}}(\mathbb{I}_{\hat{B}_k}) \approx \beta^{k,\mathrm{AP}}_{\mathrm{lim}}.$$

As we will now see, this approximation is pretty good for the random allocation policy, but less so for the least-filled and most-filled allocation policies. For the sake of brevity, we will only investigate the accuracy of the approximate blocking ratios.

### Approximate blocking ratios

Both approximate rate operators $\tilde{Q}_{\mathrm{RA}}$ and $\tilde{Q}_{\mathrm{LM}}$ are irreducible. As before, we will not prove this in any formal manner – one way to prove this is to start from the irreducibility of $Q_{\mathrm{RA}}$ and $Q_{\mathrm{LF}}$, see (Erreygers, Rottondi, et al., 2018a, Appendix A.C) for more details. With AP equal to RA or LM, we denote the corresponding approximate blocking ratios by

$$\tilde{\beta}^{1,\mathrm{AP}}_{\mathrm{lim}} := E^{\tilde{Q}_{\mathrm{AP}}}_{\mathrm{lim}}(\mathbb{I}_{\hat{B}_1}) \quad \text{and} \quad \tilde{\beta}^{2,\mathrm{AP}}_{\mathrm{lim}} := E^{\tilde{Q}_{\mathrm{AP}}}_{\mathrm{lim}}(\mathbb{I}_{\hat{B}_2})$$

In order to assess the accuracy of these approximations, we run the same computations as in Section 8.2.4$_{417}$, but now using the approximating rate operators $\tilde{Q}_{\mathrm{RA}}$ and $\tilde{Q}_{\mathrm{LM}}$ instead of the exact rate operators $Q_{\mathrm{RA}}$, $Q_{\mathrm{LF}}$ and $Q_{\mathrm{MF}}$: for every combination of $m_1$ and $\rho$ in Table 8.1$_{414}$, we report in Table 8.5 the median of the execution time over five consecutive runs divided by the number of models, in this case 2. By comparing the results in Table 8.5 to

**Table 8.5** Median execution time (in seconds) to determine approximate blocking ratios.

| | Algorithm 8.1$_{419}$ | | | IDR($s$) | | |
|---|---|---|---|---|---|---|
| $m_1$ | $\rho_{\mathrm{low}}$ | $\rho_{\mathrm{med}}$ | $\rho_{\mathrm{hi}}$ | $\rho_{\mathrm{low}}$ | $\rho_{\mathrm{med}}$ | $\rho_{\mathrm{hi}}$ |
| 40 | 0.012 | 0.012 | 0.065 | 0.001 | 0.001 | 0.002 |
| 80 | 0.138 | 0.176 | 0.959 | 0.009 | 0.015 | 0.027 |
| 120 | 0.671 | 0.917 | 4.901 | 0.063 | 0.121 | 0.227 |
| 160 | 1.649 | 3.003 | 15.530 | 0.208 | 0.405 | 0.816 |

those in Table 8.4$_{421}$, we observe that the decrease in the number of states – see Table 8.1$_{414}$ – indeed results in a substantial decrease of the required computational time: as expected, the reduction in execution time is more or less proportional to the reduction in the number of states. For the rest, the execution times in Table 8.5 follow the same trends as those in Table 8.4$_{421}$.

So computing the approximating blocking ratios is much faster, but are these approximations any good? To answer this question, we compare the approximating blocking ratios to the actual ones for a small system with $m_1 = 80$ and $n_2 = 4$ for traffic loads $\rho$ in $[\rho_{\mathrm{low}}, \rho_{\mathrm{hi}}]$. Looking at the plots in Fig. 8.8$_{\curvearrowright}$, we would be inclined to conclude that for the random allocation policy, the blocking ratios obtained with the approximate rate operator $\tilde{Q}_{\mathrm{RA}}$ are in good accordance with those obtained with the exact rate operator $Q_{\mathrm{RA}}$. This is precisely what Kim, Yan, et al. (2015) have done, although they compare the approximate blocking ratios to those obtained through simulation; they not only compare these values graphically, but also through the logarithm of the ratio of these values (see also Yan et al., 2014, Eqn. (19)). For the least-filled and most-filled allocation policies, the approximation is less good. One reason for this is that these two policies lead to different blocking ratios, while the approximate model yields identical approximations for the two policies. From the plot in Fig. 8.8$_{\curvearrowright}$, we deduce that the approximate blocking ratio for type 2 flows is in good accordance with the actual value, and that there is a range of traffic loads for which the approximation for the blocking ratio of type 1 flows is not very good.

To better assess the accuracy of the approximations, we take a closer look

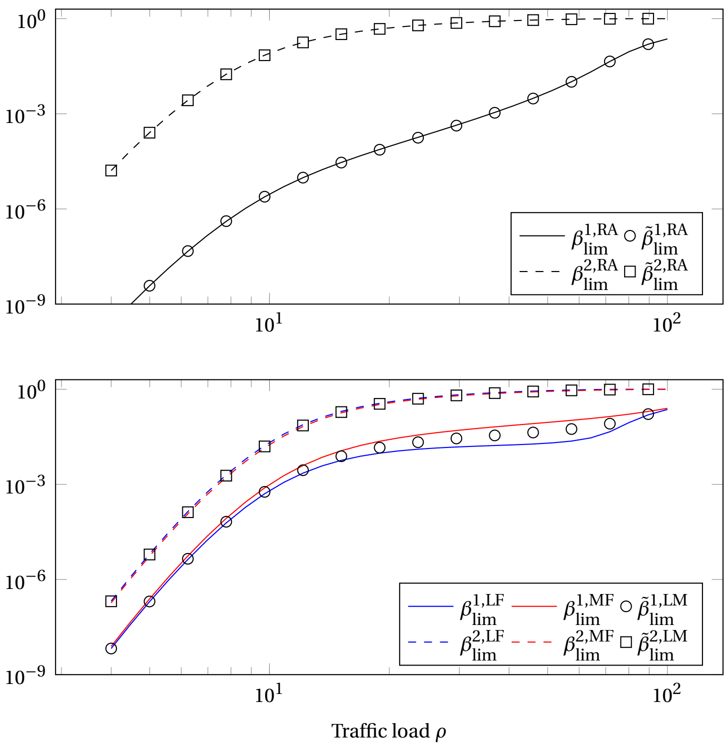



**Figure 8.8** Blocking ratios for $m_1 = 80$ and $n_2 = 4$. Note that both axes are logarithmic.

at the absolute error $|\beta_{\lim}^{k,\mathrm{AP}} - \tilde{\beta}_{\lim}^{k,\mathrm{AP}}|$ and at the relative error

$$\frac{|\beta_{\lim}^{k,\mathrm{AP}} - \tilde{\beta}_{\lim}^{k,\mathrm{AP}}|}{\beta_{\lim}^{k,\mathrm{AP}}}.$$

To this end, we use Algorithm 8.1419 with $\phi = 1 \cdot 10^{-5}$, and plot the absolute and relative error for different values of the traffic load $\rho$ and for the three allocation policies in Fig. 8.9↷. While these errors are acceptable for the random allocation policy, they are less acceptable for the least-filled and most-filled allocation policies.

### 8.3.3 Lumping to imprecise jump process models

Fortunately, we already know how to deal with the indeterminacy that arises due to lumping. To conclude this chapter, we apply the material in Chapter 7337 to the present setting. In particular, we will use Theorem 7.33363

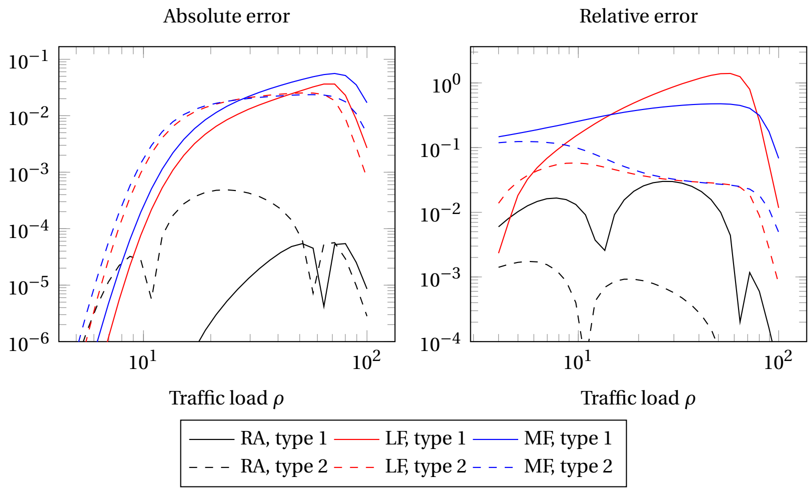



**Figure 8.9** Absolute and relative error for the approximate blocking ratios.

and the iterative method in Theorem $6.50_{318}$ to determine policy-dependent and policy-independent lower and upper bounds on the expected number of blocked requests, and the iterative methods in Propositions $7.49_{380}$ and $7.51_{382}$ to determine policy-dependent and policy-independent lower and upper bounds on the blocking ratios.

For each of the three allocation policies AP, we let $\underline{\hat{Q}}_{\mathrm{AP}} := \underline{\hat{Q}}_{Q_{\mathrm{AP}}}$ be the corresponding lumped lower rate operator, as defined by Eq. $(7.27)_{355}$; similarly, we denote the corresponding set of lumped rate operators, as defined by Eq. $(7.33)_{358}$, by $\hat{\mathscr{Q}}_{\mathrm{AP}} := \hat{\mathscr{Q}}_{Q_{\mathrm{AP}}}$. Furthermore, we also let $\underline{\hat{Q}} := \underline{\hat{Q}}_{\mathscr{Q}}$, $\hat{\mathscr{Q}} := \hat{\mathscr{Q}}_{\mathscr{Q}}$ and $\hat{\mathscr{M}} := \hat{\mathscr{M}}_{\mathscr{M}}$.

Due to Theorem $7.33_{363}$, we can use the policy-dependent lumped jump process $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_{\mathrm{AP}}}$ and the policy-independent lumped jump process $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ to determine lower and upper bounds on the expected number of blocked requests, and we will do so in Section $8.3.4_{431}$ further on. Similarly, in Section $8.3.5_{433}$ we will determine lower and upper bounds on blocking ratios with Propositions $7.49_{380}$ and $7.51_{382}$; we do so for the three allocation policies, but also for *any* allocation policy that depends on the number of type 1 flows per superchannel but not on the order of the superchannels along the grid – more exactly, we determine lower and upper bounds on the blocking ratios for all homogeneous and Markovian jump processes in $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$.

First, however, we determine an expression for $\underline{\hat{Q}}_{\mathrm{AP}}\hat{f}$ and $\underline{\hat{Q}}\hat{f}$ with Eqs. $(8.1)_{410}$ and $(7.27)_{355}$. For $n_2 = 2$ this is trivial due to the one-to-one correspondence between the detailed state description and the higher-level state description: the lower rate operator $\underline{\hat{Q}}_{\mathrm{AP}}$ is then linear and equal to the approximate rate operator $\tilde{Q}_{\mathrm{AP}}$ – this is intuitive and is pretty easy to verify formally. For $n_2 > 2$, however, using Eqs. (8.1)$_{410}$ and (7.27)$_{355}$ becomes a bit tedious, so we will not give the formal derivation here; instead, we will use an intuitive argument akin to our earlier discussion in Section 8.3.2$_{423}$.

In Section 8.3.2$_{423}$, we argued that the indeterminacy arises because we cannot precisely determine the rates of the transitions that correspond to a departure of a type 1 flow. Assume that the current lumped state $\hat{x} = (a, b, c)$ is such that $a > 0$ and $m_2 - b - c < a < n_2(m_2 - b - c - 1) + 2$. Then the departure of a type 1 flow can free up a superchannel or not, so both $\hat{x}^{+}_{1,-}$ and $\hat{x}^{=}_{1,-}$ are feasible lumped states. The rates corresponding to these transitions cannot be determined precisely, but that is not to say that we are completely in the dark, quite the contrary: we are sure that these rates are non-negative and that their sum is equal to $a\mu_1$. Furthermore, as we are about to explain, there are at least $a_{\min} := \max\{0, 2(m_2 - b - c) - a\}$ allocated type 1 flows that are alone in their superchannel, and at most

$$a_{\max} := \left\lfloor \frac{n_2(m_2 - b - c) - a}{n_2 - 1} \right\rfloor$$

such type 1 flows, and these bounds are reached. Hence, the rate $\mu^{+}$ to $\hat{x}^{+}_{1,-}$ is at least $a_{\min}\mu_1 := a_{\min}\mu_1$ and at most $a_{\max}\mu_1 := a_{\max}\mu_1$ – and these bounds are reached.

Let us see why the states in $\Lambda^{-1}(\hat{x})$ satisfy these bounds and reach them both. Note that the only freedom we have in $\Lambda^{-1}(\hat{x})$ is to distribute $a$ type 1 flows over $m_2 - b - c < a$ superchannels, ensuring that each of these superchannels contains at least 1 type 1 flow and at most $n_2$. Thus, we are only free to distribute the $a - (m_2 - b - c)$ remaining type 1 flows. First, we do so such as to minimise the number of allocated type 1 flows that are alone in their superchannel. Note that if the number $a - (m_2 - b - c)$ of type 1 flows that we are free to distribute is larger than or equal to the number $m_2 - b - c$ of superchannels that we need to populate – so if $2(m_2 - b - c) - a \leq 0$ – then we can do this in such a way that every superchannel contains at least 2 type 1 flows; in this case, the rate $\mu^{+}$ to $\hat{x}^{+}_{1,-}$ is $0 = a_{\min}\mu_1$. If on the other hand $2(m_2 - b - c) - a > 0$, then we can populate $a - (m_2 - b - c)$ superchannels with 2 type 1 flows, leaving us with at most $2(m_2 - b - c) - a = a_{\min}$ type 1 flows that are alone in their superchannel; thus, in this case $\mu^{+} = a_{\min}\mu_1$ too. Next, we maximise the number of allocated type 1 flows that are alone in their superchannel. Clearly, we can use the $a - (m_2 - b - c)$ remaining type 1 flows to completely fill up one super channel, then the next, and so on. This way, we put more than 1 type 1 flow in $\lceil (a-(m_2-b-c))/(n_2-1) \rceil$ superchannels, and this leaves $a_{\max}$ superchannels with a single type 1 flow; hence, the rate to $\hat{x}^{+}_{1,-}$ is $\mu^{+} = a_{\max}\mu_1$.

For this reason, we see that for all $\hat{f}$ in $\mathbb{G}(\hat{\mathscr{X}})$ and $\hat{x}$ in $\hat{\mathscr{X}}$,

$$\begin{aligned}[\underline{\hat{Q}}_{\mathrm{AP}}\hat{f}](\hat{x}) = \lambda_2\big(f(\hat{x}_{2,+}) - \hat{f}(\hat{x})\big) + b\mu_2\big(\hat{f}(\hat{x}_{2,-}) - \hat{f}(\hat{x})\big) - a\mu_1\hat{f}(\hat{x}) \\ + \lambda_{\mathrm{AP}}^{-}\big(\hat{f}(\hat{x}_{1,+}^{-}) - \hat{f}(x)\big) + \lambda_{\mathrm{AP}}^{=}\big(\hat{f}(\hat{x}_{1,+}^{=}) - \hat{f}(x)\big) \\ + \min\big\{\mu^{+}\hat{f}(\hat{x}_{1,-}^{+}) + (a\mu_1 - \mu^{+})\hat{f}(\hat{x}_{1,-}^{=}) \colon \mu^{+} \in \{a_{\min}\mu_1, a_{\max}\mu_1\}\big\},\end{aligned}$$

where each of the terms in this expression is only added if the relevant lumped state belongs to $\hat{\mathscr{X}}$. Recall from Section 8.3.2$_{423}$ that $\lambda_{\mathrm{LF}}^{-} = \lambda_{\mathrm{MF}}^{-}$ and $\lambda_{\mathrm{LF}}^{=} = \lambda_{\mathrm{MF}}^{=}$, so $\underline{\hat{Q}}_{\mathrm{LF}} = \underline{\hat{Q}}_{\mathrm{MF}} =: \underline{\hat{Q}}_{\mathrm{LM}}$. Determining $\underline{\hat{Q}}_{\mathrm{AP}}\hat{f}$ is linear in the number of lumped states: for all $\hat{x}$ in $\hat{\mathscr{X}}$, we only need to optimise a single term. Furthermore, $\tilde{Q}_{\mathrm{AP}}$ dominates $\underline{\hat{Q}}_{\mathrm{AP}}$ by construction, so $\tilde{Q}_{\mathrm{AP}}$ belongs to $\mathscr{Q}_{\mathrm{AP}}$.

In a similar manner, we can determine an expression for $\underline{\hat{Q}}$. In this case, we can also freely choose the rates $\lambda_{\mathrm{AP}}^{-}$ and $\lambda_{\mathrm{AP}}^{-}$ whenever $c > 0$ and $a + n_2(b + c) < m_1$ – this corresponds to the case that there is an empty superchannel and at least one superchannel that is neither empty nor fully occupied. Hence, for all $\hat{f}$ in $\mathbb{G}(\hat{\mathscr{X}})$ and $\hat{x}$ in $\hat{\mathscr{X}}$,

$$\begin{aligned}[\underline{\hat{Q}}\hat{f}](\hat{x}) = \lambda_2\big(f(\hat{x}_{2,+}) - \hat{f}(\hat{x})\big) + b\mu_2\big(\hat{f}(\hat{x}_{2,-}) - \hat{f}(\hat{x})\big) - a\mu_1\hat{f}(\hat{x}) \\ + \min\big\{\lambda^{-}\hat{f}(\hat{x}_{1,+}^{-}) + (\lambda_1 - \lambda^{-})\hat{f}(\hat{x}_{1,+}^{=}) - \lambda_1\hat{f}(\hat{x}) \colon \lambda^{-} \in \{0, \lambda_1\}\big\} \\ + \min\big\{\mu^{+}\hat{f}(\hat{x}_{1,-}^{+}) + (a\mu_1 - \mu^{+})\hat{f}(\hat{x}_{1,-}^{=}) \colon \mu^{+} \in \{a_{\min}\mu_1, a_{\max}\mu_1\}\big\},\end{aligned}$$

where each of the terms in this expression is only added if the relevant lumped state belongs to $\hat{\mathscr{X}}$ – the minimisation for $\lambda^{-}$ is only necessary if $c > 0$ and $a + n_2(b + c) < m_1$ as the value of $\lambda^{-}$ is uniquely determined otherwise, and the minimisation for $\mu^{+}$ is only necessary if $a > 0$ and $m_2 - b - c < a < n_2(m_2 - b - c - 1) + 2$ as the value of $\mu^{+}$ is uniquely determined otherwise.

### 8.3.4 Bounding the expected number of blocked requests

Finally, we can put the induced lumped imprecise models to the test. To this end, we observe that the set $B_1$ and $B_2$ are lumpable, in the sense that $B_1 = \Lambda^{-1}(\hat{B}_1)$ and $B_2 = \Lambda^{-1}(\hat{B}_2)$. Hence, $\mathbb{I}_{B_1} = \mathbb{I}_{\hat{B}_1} \circ \Lambda$ and $\mathbb{I}_{B_2} = \mathbb{I}_{\hat{B}_2} \circ \Lambda$, and this implies that the corresponding temporal averages are lumpable too: for all $k$ in $\{1, 2\}$ and $r$ in $\mathbb{R}_{>0}$, it follows from Eq. (7.44)$_{367}$ that

$$\big(\llbracket \mathbb{I}_{\hat{B}_k} \rrbracket_{[0,r]}\big)^{\uparrow\Omega} = \left(\frac{1}{r}\int_0^r \mathbb{I}_{\hat{B}_k}(\hat{X}_t)\,\mathrm{d}t\right)^{\uparrow\Omega} = \frac{1}{r}\int_0^r \mathbb{I}_{B_k}(X_t)\,\mathrm{d}t = \llbracket \mathbb{I}_{B_k} \rrbracket_{[0,r]}.$$

Consequently, it follows from Theorem 7.33$_{363}$ that, for all $k$ in $\{1, 2\}$, $r$ in $\mathbb{R}_{>0}$ and $x$ in $\mathscr{X}$, and with AP equal to RA, LF or MF,

$$\lambda_k r \underline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_{\mathrm{AP}}}\big(\llbracket \mathbb{I}_{\hat{B}_k} \rrbracket_{[0,r]} \,\big|\, \hat{X}_0 = \Lambda(x)\big) \leq \beta_{[0,r]}^{k,\mathrm{AP}} \leq \lambda_k r \overline{E}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_{\mathrm{AP}}}[\,](\llbracket \mathbb{I}_{\hat{B}_k} \rrbracket_{[0,r]} \,|\, \hat{X}_0 = \Lambda(x)),$$

and similarly for $\underline{\beta}_{[0,r]}^{k}$ and $\underline{\beta}_{[0,r]}^{k}$ but with $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ instead of $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_{\mathrm{AP}}}$. To compute these lower and upper bounds, we use the recursive method in Theorem 6.50$_{318}$ in the same way as in Section 8.2.3$_{412}$.

### Numerical experiments

Let us repeat the three experiments in Section 8.2.3 412 to asses whether or not lumping results in more tractable computations. First, we consider the system with $m_1 = 80$ and the random allocation policy. In Fig. 8.10, we depict the lower and upper bounds on the expected number of blocked requests over $[0, r]$ with $r = {}^{100}/_{\rho}$, so we basically add some information to the plots in Fig. 8.5 414. We observe that the bounds obtained with the policy-

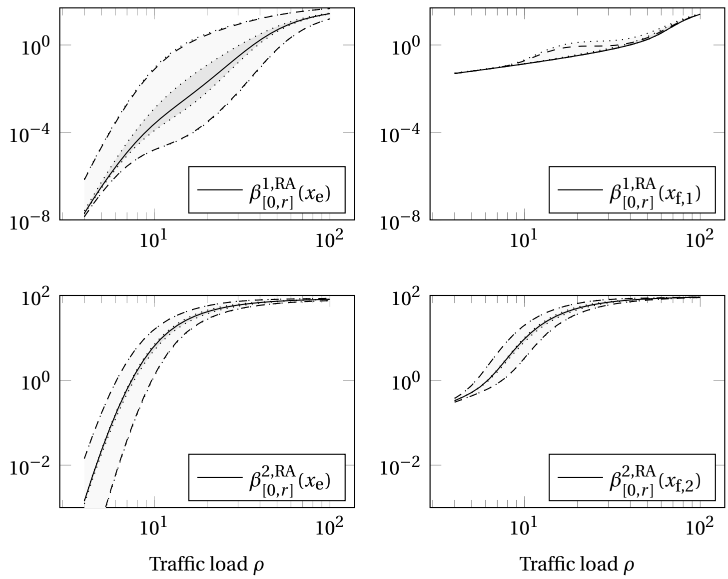



**Figure 8.10** Lower and upper bounds on the expected number of blocked requests for $m_1 = 80$ and $n_2 = 4$ for the random allocation policy. The dashed lines correspond to the policy-independent model $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, the dotted lines with darker shading to the policy-dependent lumped model $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_{\mathrm{RA}}}$ and the dotted lines with the lighter shading to the policy-independent lumped model $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$.

dependent lumped model $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathrm{Q}}_{\mathrm{AP}}}$ are fairly tight, and that those obtained with the policy-independent lumped model $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathrm{Q}}}$ are almost indistinguishable from those obtained with $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$.

Next, we compare the time required to compute $\beta^{k,\mathrm{AP}}_{[0,r]}$ to the time required to compute the lower and upper bounds. We run the same experiment as for Table 8.2 415, only this time around we compute the upper bound on the expected number of blocked requests for the policy-dependent lumped models $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_{\mathrm{RA}}}$ and $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_{\mathrm{LM}}}$ and for the policy-independent lumped model $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$; here too, we report the median execution time for the two policy-dependent models divided by the number of models. Note that computing the upper

**Table 8.6** Median execution time (in seconds) to determine an upper bound on the expected number of blocked requests.

| | $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_{\mathrm{AP}}}$ | | | $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$ | | |
|---|---|---|---|---|---|---|
| $m_1$ | $\rho_{\mathrm{low}}$ | $\rho_{\mathrm{med}}$ | $\rho_{\mathrm{hi}}$ | $\rho_{\mathrm{low}}$ | $\rho_{\mathrm{med}}$ | $\rho_{\mathrm{hi}}$ |
| 40 | 1.25 | 0.34 | 0.17 | 1.32 | 0.35 | 0.18 |
| 80 | 8.20 | 2.30 | 1.08 | 8.74 | 2.41 | 1.17 |
| 120 | 33.03 | 9.62 | 4.22 | 35.32 | 9.80 | 4.56 |
| 160 | 100.72 | 27.56 | 13.10 | 103.48 | 28.21 | 13.44 |

bounds with the lumped models is *much* faster than computing the exact values with the exact models. For the rest, we observe more or less the same trends as in Table 8.2$_{415}$: the execution time decreases as the traffic load increases and is roughly proportional to the number of lumped states, and the execution times for the lumped policy-independent model are slightly longer because determining $\underline{\hat{Q}}\hat{f}$ is slightly more expensive than determining $\underline{\hat{Q}}_{\mathrm{AP}}\hat{f}$.

Finally, we look at how the expected temporal average of $\mathbb{I}_{B_k}(X_t)$ evolves over time. While we looked at a system with $m_1 = 80$ slots in Section 8.3.4$_{431}$, we now consider a system with $m_1 = 160$ slots, and we take the random allocation policy as an example. In Fig. 8.11$_{\curvearrowright}$, we plot lower and upper bounds on the expected temporal average obtained with the policy-dependent lumped model $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}_{\mathrm{RA}}}$ for the random allocation policy and the policy-independent lumped model $\mathbb{P}_{\hat{\mathscr{M}},\hat{\mathscr{Q}}}$. These plots are similar to those in Fig. 8.6$_{416}$, so our observations regarding Fig. 8.6$_{416}$ hold here too. Note that we do not plot the exact values because computing them takes too long.

### 8.3.5 Bounding blocking ratios

Finally, we use Propositions 7.49$_{380}$ and 7.51$_{382}$ to determine lower and upper bounds on the blocking ratios with $\underline{\hat{Q}}_{\mathrm{RA}}$, $\underline{\hat{Q}}_{\mathrm{LM}}$ and $\underline{\hat{Q}}$. Note that we could also use Proposition 7.46$_{376}$ for this, but we will not do this here.

We compute lower and upper bounds on the blocking ratio with the obvious counterpart(s) of Algorithm 8.1$_{419}$: Algorithm 8.2$_{435}$ computes a guaranteed upper bound, and to compute a lower bound we change the first line in Algorithm 8.2$_{435}$ to $\hat{g}_0 := \hat{f}$ and the last line to $\min \hat{g}_n$.

In general, the step size $\Delta$ in Algorithm 8.2$_{435}$ influences the tightness of the bound in two ways: directly because $-\min(I + \Delta\underline{\hat{Q}})^n(-\hat{f})$ can converge to a different value, and indirectly because a smaller step size typically results in more iterations to reach convergence. To determine bounds on the blocking ratios, we always use the step size $\Delta = 1.9/\|\underline{\hat{Q}}_{\hat{\mathscr{Q}}}\|_{\mathrm{op}}$; in our scenarios, using the

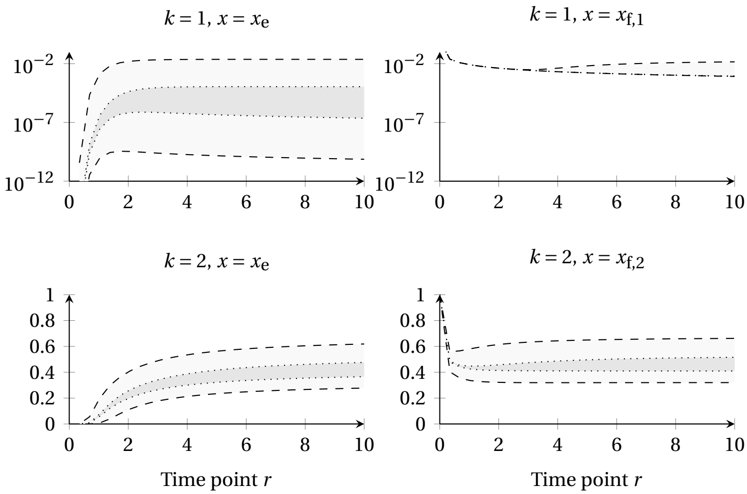



**Figure 8.11** Bounds on the expectation of $[\![\mathbb{I}_{B_k}]\!]_{[0,r]}$ for $m_1 = 160$, $n_2 = 4$ and $\rho = \rho_{\text{med}} = 40$ conditional on $\{X_0 = x\}$ for the random allocation policy. Dotted lines indicate bounds corresponding to $\mathbb{P}_{\hat{\mathscr{M}},\underline{\hat{Q}}_{\text{RA}}}$, and dashed lines the bounds corresponding to $\mathbb{P}_{\hat{\mathscr{M}},\underline{\hat{Q}}}$.

smaller step size $\Delta = 1/2\|\underline{\hat{Q}}_{\mathscr{Q}}\|_{\text{op}}$ only changes some of the bounds from the third significant digit on.

We run the same experiment as in Section 8.2.4~417~, although we use Algorithm 8.2 to determine upper bounds on the blocking ratios using $\underline{\hat{Q}}_{\text{RA}}$, $\underline{\hat{Q}}_{\text{LM}}$, and $\underline{\hat{Q}}$ with parameters $\phi = 10^{-3}$ and $n_{\max} = 10^6$. We report the median execution time in Table 8.7; for the two policy-dependent imprecise models, we report the median execution time over five consecutive runs divided by the number of models (so 2 in this case), whereas for the policy-independent model, we simply report the median execution time over five consecutive runs.

Note that computing an upper bound on the blocking ratios using the lumped models is *considerably* faster than computing the blocking ratios using the exact models. In line with what we have seen before, the execution times are roughly proportional to the number of lumped states, and they increase as the traffic load increases. For the low traffic load, computing the policy-dependent bounds is (marginally) faster than computing the policy-independent bounds; the obvious explanation for this is that determining $\underline{\hat{Q}}\hat{f}$ requires more optimisations that determining $\underline{\hat{Q}}_{\text{AP}}\hat{f}$. For the medium and high traffic load, this is the other way around. The reason for this is that the number of iterations until convergence is similar for $\underline{\hat{Q}}_{RA}$ and $\underline{\hat{Q}}$ but much larger for $\underline{\hat{Q}}_{\text{LM}}$; unfortunately, we have no intuitive explanation for this

**Algorithm 8.2:** Iteratively compute an upper bound on the limit expectation $E^{Q}_{\mathrm{lim}}(f)$ of a lumpable $f$ in $\mathbb{G}(\mathscr{X})$ that holds for every rate operator $Q$ in a set $\mathcal{Q}$ of ergodic rate operators

**Input:** A non-empty set $\mathcal{Q}$ of ergodic rate operators on $\mathbb{G}(\mathscr{X})$, a gamble $f$ in $\mathbb{G}(\mathscr{X})$ such that $f^{\downarrow\min} = f^{\downarrow\max} =: \hat{f}$, a time step $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\|\underline{\hat{Q}}_{\mathcal{Q}}\|_{\mathrm{op}} < 2$, a relative tolerance $\phi$ in $\mathbb{R}_{>0}$ and a maximum number of iterations $n_{\max}$ in $\mathbb{N}$.

**Output:** An upper bound on $E^{Q}_{\mathrm{lim}}(f)$ that holds for all $Q$ in $\mathcal{Q}$

$n := 0$
$\hat{g}_0 := -\hat{f}$
$\epsilon_{\mathrm{rel}} := 2\|\hat{g}_0\|_{\mathrm{v}}/|\max \hat{g}_0 + \min \hat{g}_0|$
**while** $\epsilon_{\mathrm{rel}} > \phi$ **and** $n < n_{\max}$ **do**
    $n := n+1$
    $\hat{g}_n := \hat{g}_{n-1} + \Delta\underline{\hat{Q}}_{\mathcal{Q}}\hat{g}_{n-1}$
    $\epsilon_{\mathrm{rel}} := 2\|\hat{g}_n\|_{\mathrm{v}}/|\max \hat{g}_n + \min \hat{g}_n|$
**return** $-\min \hat{g}_n$

**Table 8.7** Median execution time (in seconds) to determine upper bounds on the blocking ratios with Algorithm 8.2.

| | $\underline{\hat{Q}}_{\mathrm{AP}}$ | | | $\underline{\hat{Q}}$ | | |
|---|---|---|---|---|---|---|
| $m_1$ | $\rho_{\mathrm{low}}$ | $\rho_{\mathrm{med}}$ | $\rho_{\mathrm{hi}}$ | $\rho_{\mathrm{low}}$ | $\rho_{\mathrm{med}}$ | $\rho_{\mathrm{hi}}$ |
| 40 | 0.02 | 0.02 | 0.10 | 0.02 | 0.02 | 0.09 |
| 80 | 0.28 | 0.41 | 1.49 | 0.28 | 0.31 | 1.46 |
| 120 | 1.79 | 3.46 | 9.40 | 1.83 | 1.90 | 9.03 |
| 160 | 4.72 | 15.50 | 25.31 | 5.14 | 6.06 | 19.84 |

behaviour.

To conclude our investigation, we compare the actual values of the blocking ratios to the lower and upper bounds. In Fig. 8.12↷, we depict the blocking ratios and the lower and upper bounds on the blocking ratios for each traffic type for a system with $m_1 = 80$, $n_2 = 4$, $\mu_1 = 1 = \mu_2$ and $\lambda_1 = \rho = \lambda_2$. Note that these graphs all have a double logarithmic scale, so our discussion of the graphs – and our use of terms such as tight, wide, close to or in the middle – should be interpreted in a logarithmic sense. For type 2 flows, the bounds calculated with the policy-dependent model are very tight and show that evaluating the performance with the lumped model yields accurate results for all the considered parameters and traffic loads. However, the calculated policy-dependent lower and upper bounds for type 1 flows are relatively wide, especially for intermediate loads.

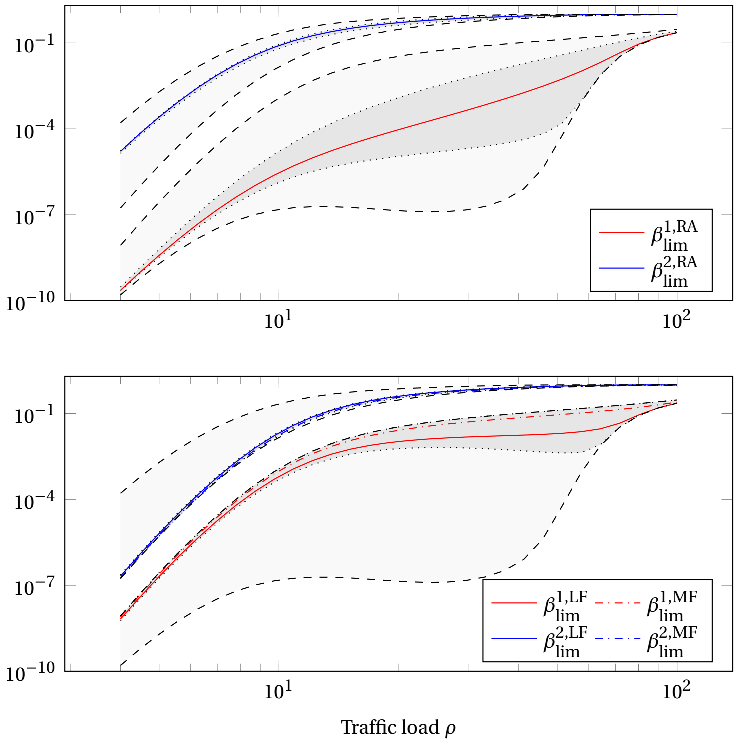



**Figure 8.12** Blocking ratios for $m_1 = 80$ and $n_2 = 4$. The lower and upper bounds for the policy-dependent lumped models are displayed as dotted lines, those for the policy-independent lumped model are displayed as dashed lines.

The policy-dependent and policy-independent lower and upper blocking ratios allow us to asses the performance of the allocation policies. For this analysis, we also take the graphs in Fig. 8.13 into account, where we plot the lower and upper blocking ratios for a system with $m_1 = 160$ slots and $n_2 = 4$ slots per superchannel; computing the exact blocking ratios over the whole range of traffic loads is infeasible for this system, so we plot the approximate blocking ratios that we obtain with the approximate rate operators $\tilde{Q}_{\text{RA}}$ and $\tilde{Q}_{\text{LM}}$. In Figs. 8.12 and 8.13, we can see that the least-filled and most-filled allocation policies yield a blocking ratio very close to the policy-independent lower bound for type 2 flows and very close to the policy-independent upper bound for type 1 flows. This demonstrates that these policies favour type 2 flows. On the other hand, the lower and upper blocking ratios for the random allocation policy are situated more or less in the middle of the policy-

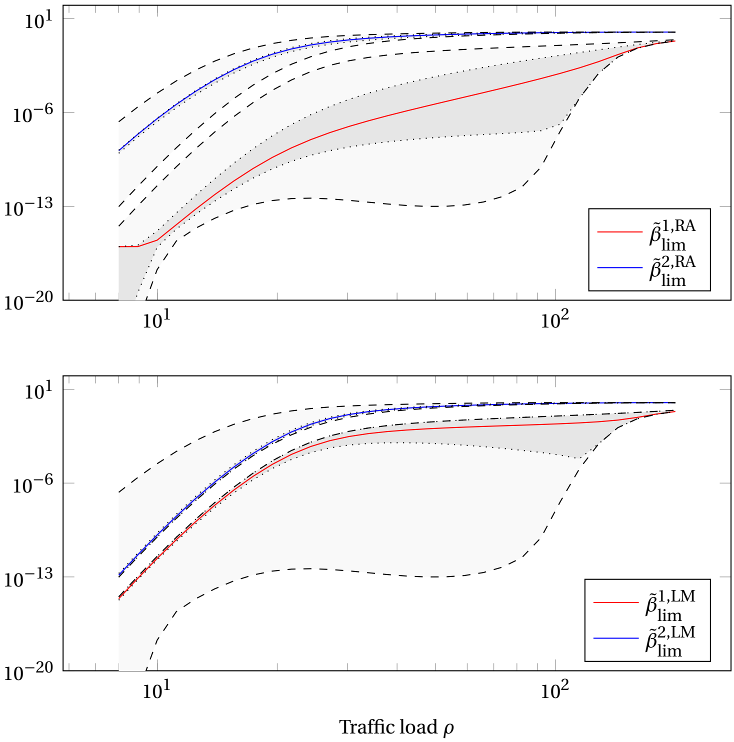



**Figure 8.13** Blocking ratios for $m_1 = 160$ and $n_2 = 4$. The lower and upper bounds for for the policy-dependent lumped models are displayed as dotted lines, those for the policy-independent lumped model are displayed as dashed lines.

independent bounds, and this for both connection types, so this policy shows no preference for one type of flow over the other.

# Conclusions 9

The main conclusion of this dissertation is that Markovian imprecise jump processes generalise 'standard' Markovian jump processes in two ways: they allow for partially specified parameters, and they do not require homogeneity or even Markovianity. Moreover, they do so in such a way that many things that are possible in the precise case, are possible in the imprecise case as well. Let us elaborate this a bit.

From Chapters $3_{53}$ and $4_{157}$, we take away that the law of iterated lower expectations allows us to compute the lower and upper expectations of any simple variable, and that the weaker sum-product law of iterated lower expectations suffices for simple variables that have a sum-product representation. Furthermore, despite the added generality of Markovian imprecise jump processes, the resulting computational methods do not add a large computational burden compared to homogeneous Markovian jump processes.

The main conclusion from Chapter $5_{215}$ is that we can extend the domain of Markovian imprecise jump processes from simple variables to idealised variables, so from variables that depend on the state of the system at a finite number of time points to variables that depend on the state of the system at all time points in an (unbounded) time interval. A second important conclusion is that the resulting extended lower and upper expectations satisfy generalisations of the Monotone Convergence Theorem and Lebesgue's Dominated Convergence Theorem, but that the limit lower and upper bounds in these theorems need not be tight. On a more didactic note, our approach in Chapter $5_{215}$ shows that the conventional route followed in the theory of measure-theoretic jump processes is a bit convoluted. While it has become a tradition to start off with the set of all paths and to subsequently modify the projector variables in such a way that they have càdlag sample paths, we simply consider the set of càdlàg paths from the very beginning – to be fair, we use the measure-theoretical 'modification trick' in one of our proofs. This is also a good time to mention that Nendel (2020, 2021) has generalised (measure-theoretical) homogeneous Markovian jump processes to the setting of convex expectations (Peng, 2005), hence also obtaining – what could be considered – an imprecise Markovian jump process as a special case. In particular, he uses a Kolmogorov-type extension theorem by Denk et al. (2018) to obtain a convex expectation on the set of all bounded variables that are measurable with respect to the sigma algebra generated by the cylinder events. However, he does not modify the projector variables in such a way that they have càdlàg sample paths, and this means that in his framework, most non-trivial idealised variables – including the three types that we consider – are not measurable with respect to the generated sigma algebra, as explained in Appendix $5.B.2_{247}$. We consider this to be an important advantage of our approach.

From Chapter $6_{273}$, we conclude that the limit lower and upper bounds in the aforementioned theorems are actually tight for several types of idealised variables, and even more, that we can compute their lower and upper expectations up to arbitrary precision. More importantly, the work in Chapters $5_{215}$ and $6_{273}$ gives rise to many new research topics in the theory of Markovian imprecise jump process, simply because the domain has been increased from simple variables to idealised ones. We will come back to this further on.

Chapter $7_{337}$ shows that Markovian imprecise jump processes are elegant tools for dealing with the parameter indeterminacy that arises when lumping a homogenenous Markovian (imprecise) jump process. Specifically, they allow us to compute bounds on inferences that we could not have determined otherwise – at least not within a reasonable time span. Chapter $8_{403}$ illustrates this nicely, and this chapter also demonstrates the general modelling power and practical feasibility of our approach.

Following an honourable tradition, we conclude this dissertation with a brief discussion of possible avenues for future research. In Chapter $6_{273}$, we already put forward three such avenues. The first avenue is to determine efficient methods to compute the lower and upper expected number of jumps efficiently; we mentioned this in Footnote $1_{276}$, where we referred to the method that we present (without proof) in (Erreygers & De Bock, 2021, Section 5.2). The two other avenues were put forward at the end of Sections $6.2.3_{287}$ and $6.3.1_{290}$, respectively: (i) investigate whether lower and upper hitting probabilities and/or expected hitting times can be computed in a similar manner as for homogeneous Markovian jump processes; and (ii) investigate the limit behaviour of the lower and upper expected temporal average, and specifically the link with ergodicity. The discrete-time counterparts of these investigations have led to positive results, but the arguments there do not necessarily translate immediately to our continuous-time setting. One reason for this is that many of the arguments for imprecise Markov chains use the game-theoretic framework put forward by Shafer et al. (2001, 2019). To the best of our knowledge, (imprecise) jump processes have yet to be studied in this framework, so this is an interesting – but challenging – avenue for further research. Another ambitious research project could be to develop a theory of imprecise jump processes for countable – or, why not, uncountable – state spaces. Some guidance might be found in my preliminary work on Poisson processes (Erreygers & De Bock, 2019b). This work uses the framework of sets of coherent conditional probabilities, but I suspect that the game-theoretic approach is actually more straightforward for Poisson processes, or even for counting processes in general; hence, as far as the game-theoretic framework is concerned, it might be a good idea to start with Poisson processes instead of jump processes. Finally, while we are on the topic of Poisson processes, a generalisation of the PASTA property – Poisson Arrivals See Temporal-Averages – to (Markovian) imprecise jump processes would be of tremendous practical interest. No shortage of ideas then, so let's get on with it!

# Compactness A

In this appendix, we establish some compactness properties of sets of operators. More precisely, we explain that the set of coherent lower expectations on a finite possibility space is compact in Appendix A.1. In Appendix A.2[446], we subsequently establish that the set of lower transition operators is compact. Finally, we argue that this implies that a bounded set of lower rate operators is compact in Appendix A.3[448].

## A.1 Compactness of the set of coherent lower expectations

Let $\underline{\mathbb{E}}_{\mathscr{X}}$ denote the set of all coherent lower expectations on $\mathbb{G}(\mathscr{X})$, and let $\mathbb{G}_1(\mathscr{X}) := \{f \in \mathbb{G}(\mathscr{X}) : 0 \leq f \leq 1\}$. As shown by Škulj et al. (2013),

$$d : \underline{\mathbb{E}}_{\mathscr{X}} \times \underline{\mathbb{E}}_{\mathscr{X}} \to \mathbb{R}_{\geq 0} : (\underline{E}, \underline{E}') \mapsto \max\{|\underline{E}(f) - \underline{E}'(f)| : f \in \mathbb{G}_1(\mathscr{X})\} \tag{A.1}$$

is a metric on $\underline{\mathbb{E}}_{\mathscr{X}}$. We will need the following convenient property of this metric in the proof of Proposition 4.38[197], so it makes sense to state it here. Our proof is essentially based on (Škulj et al., 2013, Eqn. (11)).

**Lemma A.1.** *Let $\underline{E}_1$ and $\underline{E}_2$ be two coherent lower expectations on $\mathbb{G}(\mathscr{X})$. Then*

$$d(\underline{E}_1, \underline{E}_2) \leq \max\{\overline{E}_1(\mathbb{I}_A) - \underline{E}_2(\mathbb{I}_A) : A \in \mathcal{P}(\mathscr{X}), \emptyset \neq A \neq \mathscr{X}\}.$$

*Proof.* Recall from (LE4)[30] that for all $k$ in $\{1,2\}$ and $f$ in $\mathbb{G}_1(\mathscr{X})$, $0 \leq \underline{E}_k(f) \leq \overline{E}_k(f) \leq 1$. From this, it follows immediately that for any $f$ in $\mathbb{G}_1(\mathscr{X})$,

$$\underline{E}_1(f) - \underline{E}_2(f) \leq \overline{E}_1(f) - \underline{E}_2(f) \quad \text{and} \quad \overline{E}_1(f) - \overline{E}_2(f) \leq \overline{E}_1(f) - \underline{E}_2(f). \tag{A.2}$$

Let $f^\star$ be any gamble in $\mathbb{G}_1(\mathscr{X})$ such that

$$d(\underline{E}_1, \underline{E}_2) = \left|\underline{E}_1(f^\star) - \underline{E}_2(f^\star)\right|.$$

Note that either $|\underline{E}_1(f^\star) - \underline{E}_2(f^\star)| = \underline{E}_1(f^\star) - \underline{E}_2(f^\star)$ or $|\underline{E}_1(f^\star) - \underline{E}_2(f^\star)| = \underline{E}_2(f^\star) - \underline{E}_1(f^\star)$. In the first case,

$$d(\underline{E}_1, \underline{E}_2) = \left|\underline{E}_1(f^\star) - \underline{E}_2(f^\star)\right| = \underline{E}_1(f^\star) - \underline{E}_2(f^\star) \leq \overline{E}_1(f^\star) - \underline{E}_2(f^\star), \tag{A.3}$$

where the inequality follows from Eq. (A.2)$_\curvearrowleft$. In the second case,

$$d(\underline{E}_1,\underline{E}_2) = \left|\underline{E}_1(f^\star) - \underline{E}_2(f^\star)\right| = \underline{E}_2(f^\star) - \underline{E}_1(f^\star).$$

Note that

$$\underline{E}_2(f^\star) - \underline{E}_1(f^\star) = -\overline{E}_2(-f^\star) + \overline{E}_1(-f^\star) = -\overline{E}_2(1-f^\star) + \overline{E}_1(1-f^\star).$$

where the first equality holds due to conjugacy and the second equality holds due to (LT5)$_{108}$. Observe that $1 - f^\star$ belongs to $\mathbb{G}_1(\mathscr{X})$ because $f^\star$ does. Consequently, it follows from the preceding equalities and Eq. (A.2)$_\curvearrowleft$ that

$$d(\underline{E}_1,\underline{E}_2) = \overline{E}_1(1-f^\star) - \overline{E}_2(1-f^\star) \leq \overline{E}_1(1-f^\star) - \underline{E}_2(1-f^\star). \tag{A.4}$$

Observe that $\mathbb{G}_1(\mathscr{X}) = \{1 - f \colon f \in \mathbb{G}_1(\mathscr{X})\}$. For this reason, it follows from Eqs. (A.3)$_\curvearrowleft$ and (A.4) that

$$d(\underline{E}_1,\underline{E}_2) \leq \sup\left\{\overline{E}_1(f) - \underline{E}_2(f) \colon f \in \mathbb{G}_1(\mathscr{X})\right\}. \tag{A.5}$$

Fix any $f$ in $\mathbb{G}(\mathscr{X})$. Recall from Theorem 2.28$_{30}$ that for any $k$ in $\{1,2\}$, $\underline{E}_k(f) = \min\{E_k(f) \colon E_k \in \mathcal{M}_k\}$, where $\mathcal{M}_k := \mathcal{M}_{\underline{E}_k}$ is the set of coherent expectations on $\mathbb{G}(\mathscr{X})$ that dominate $\underline{E}_k$. Consequently,

$$\begin{aligned}\overline{E}_1(f) - \underline{E}_2(f) &= \max\left\{E_1(f) \colon E_1 \in \mathcal{M}_1\right\} - \min\left\{E_2(f) \colon E_2 \in \mathcal{M}_2\right\} \\ &= \max\Big\{\max\left\{E_1(f) - E_2(f) \colon E_2 \in \mathcal{M}_2\right\} \colon E_1 \in \mathcal{M}_1\Big\}.\end{aligned}$$

We use the previous equality to rewrite the right-hand side of Eq. (A.5):

$$\begin{aligned}&\sup\left\{\overline{E}_1(f) - \underline{E}_2(f) \colon f \in \mathbb{G}_1(\mathscr{X})\right\} \\ &\quad = \sup\Big\{\max\Big\{\max\left\{E_1(f) - E_2(f) \colon E_2 \in \mathcal{M}_2\right\} \colon E_1 \in \mathcal{M}_1\Big\} \colon f \in \mathbb{G}_1(\mathscr{X})\Big\} \\ &\quad = \max\Big\{\max\Big\{\sup\left\{E_1(f) - E_2(f) \colon f \in \mathbb{G}_1(\mathscr{X})\right\} \colon E_2 \in \mathcal{M}_2\Big\} \colon E_1 \in \mathcal{M}_1\Big\}.\end{aligned} \tag{A.6}$$

Škulj et al. (2013, Proposition 1)[1] prove that for any two coherent expectations $E_1$ and $E_2$ on $\mathbb{G}(\mathscr{X})$,

$$\sup\left\{|E_1(f) - E_2(f)| \colon f \in \mathbb{G}_1(\mathscr{X})\right\} = \max\left\{|E_1(\mathbb{I}_A) - E_2(\mathbb{I}_A)| \colon A \in \mathcal{P}(\mathscr{X}), \varnothing \neq A \neq \mathscr{X}\right\}.$$

Because

$$E_1(f) - E_2(f) = -\big(E_1(1-f) - E_2(1-f)\big) \quad \text{for all } f \in \mathbb{G}_1(\mathscr{X}),$$

and in particular,

$$E_1(\mathbb{I}_A) - E_2(\mathbb{I}_A) = -\big(E_1(\mathbb{I}_{A^c}) - E_2(\mathbb{I}_{A^c})\big) \quad \text{for all } A \in \mathcal{P}(\mathscr{X}), \varnothing \neq A \neq \mathscr{X},$$

[1] Technically, they prove this for expectations corresponding to probability mass functions, but this is not an issue because by Corollary 2.17$_{23}$ and Proposition 2.18$_{23}$ in Section 2.2.2$_{19}$, there is a one-to-one correspondence between coherent expectations $E$ on $\mathbb{G}(\mathscr{X})$ and probability mass functions on $\mathscr{X}$.

this implies that

$$\sup\{E_1(f)-E_2(f)\colon f\in\mathbb{G}_1(\mathscr{X})\}=\max\{E_1(\mathbb{I}_A)-E_2(\mathbb{I}_A)\colon A\in\mathcal{P}(\mathscr{X}),\varnothing\neq A\neq\mathscr{X}\}.$$

Substituting the preceding equality in Eq. (A.6), we obtain that

$$\begin{aligned}
&\sup\{\overline{E}_1(f)-\underline{E}_2(f)\colon f\in\mathbb{G}_1(\mathscr{X})\}\\
&=\max\Big\{\max\Big\{\max\{E_1(\mathbb{I}_A)-E_2(\mathbb{I}_A)\colon A\in\mathcal{P}(\mathscr{X}),\varnothing\neq A\neq\mathscr{X}\}\colon E_2\in\mathcal{M}_2\Big\}\colon E_1\in\mathcal{M}_1\Big\}\\
&=\max\Big\{\max\Big\{\max\{E_1(\mathbb{I}_A)-E_2(\mathbb{I}_A)\colon E_2\in\mathcal{M}_2\}\colon E_1\in\mathcal{M}_1\Big\}\colon A\in\mathcal{P}(\mathscr{X}),\varnothing\neq A\neq\mathscr{X}\Big\}\\
&\qquad=\max\{\overline{E}_1(\mathbb{I}_A)-\underline{E}_2(\mathbb{I}_A)\colon A\in\mathcal{P}(\mathscr{X}),\varnothing\neq A\neq\mathscr{X}\}.
\end{aligned}$$

The inequality of the statement now follows immediately from the previous equality and Eq. (A.5). □

Importantly, De Bock et al. (2015, Appendix A) prove that $\underline{\mathbb{E}}_{\mathscr{X}}$ is compact with respect to the metric $d$ because the state space $\mathscr{X}$ is finite.

**Lemma A.2.** *The metric space* $(\underline{\mathbb{E}}_{\mathscr{X}},d)$ *is compact.*

We need to take one more step in order to be able to use the preceding result. Recall that in the definition of the operator norm, we consider all gambles $f$ on $\mathscr{X}$ with $\|f\|=1$ instead of $0\leq f\leq 1$. However, we now establish that this makes no real difference.

**Lemma A.3.** *For any two coherent lower previsions* $\underline{E}$ *and* $\underline{E}'$ *on* $\mathbb{G}(\mathscr{X})$*,*

$$d(\underline{E},\underline{E}')=\frac{1}{2}\sup\{|\underline{E}(f)-\underline{E}'(f)|\colon f\in\mathbb{G}(\mathscr{X}),\|f\|=1\}.$$

*Proof.* The statement holds trivially in case $\underline{E}=\underline{E}'$. Thus, from here on we assume that $\underline{E}\neq\underline{E}'$, whence $d(\underline{E},\underline{E}')>0$. Note that for any gamble $f$ on $\mathscr{X}$, $0\leq f\leq 1$ if and only if $-1\leq 2f-1\leq 1$ or, equivalently, $\|2f-1\|\leq 1$. For this reason,

$$\begin{aligned}
d(\underline{E},\underline{E}')&=\max\{|\underline{E}(f)-\underline{E}'(f)|\colon f\in\mathbb{G}(\mathscr{X}),0\leq f\leq 1\}\\
&=\max\{|\underline{E}(f)-\underline{E}'(f)|\colon f\in\mathbb{G}(\mathscr{X}),\|2f-1\|\}\\
&=\max\left\{\left|\frac{1}{2}\underline{E}(2f-1)-\frac{1}{2}\underline{E}'(2f-1)\right|\colon f\in\mathbb{G}(\mathscr{X}),\|2f-1\|\right\}\\
&=\frac{1}{2}\max\{|\underline{E}(2f-1)-\underline{E}'(2f-1)|\colon f\in\mathbb{G}(\mathscr{X}),\|2f-1\|\leq 1\}\\
&=\frac{1}{2}\max\{|\underline{E}(f)-\underline{E}'(f)|\colon f\in\mathbb{G}(\mathscr{X}),\|f\|\leq 1\},
\end{aligned}$$

where for the third equality we have used (LE2)$_{30}$ and (LE5)$_{30}$. To verify that this agrees with the equality of the statement, we assume *ex absurdo* that the maximum in the last equality is not reached for some $f$ in $\mathbb{G}(\mathscr{X})$ with $\|f\|=1$. This implies that there is some $f$ in $\mathbb{G}(\mathscr{X})$ such that $\|f\|<1$ and

$$0<d(\underline{E},\underline{E}')=\frac{1}{2}|\underline{E}(f)-\underline{E}'(f)|.$$

Note that $\|f\| > 0$ because

$$|\underline{E}(0) - \underline{E}'(0)| = |0 - 0| = 0 < 2d(\underline{E}, \underline{E}') = |\underline{E}(f) - \underline{E}'(f)|,$$

where the first equality follows from (LE2)$_{30}$.

Let $\tilde{f} := f / \|f\|$ such that $\|\tilde{f}\| = 1$, and observe that

$$\begin{aligned} 2d(\underline{E}, \underline{E}') = |\underline{E}(f) - \underline{E}'(f)| = \left| \|f\| \underline{E}\left(\frac{f}{\|f\|}\right) - \|f\| \underline{E}'\left(\frac{f}{\|f\|}\right) \right| = \|f\| |\underline{E}(\tilde{f}) - \underline{E}(\tilde{f})| \\ < |\underline{E}(\tilde{f}) - \underline{E}(\tilde{f})|, \end{aligned}$$

where for the second equality we have used (LE2)$_{30}$. However, because $\|\tilde{f}\| \leq 1$,

$$|\underline{E}(\tilde{f}) - \underline{E}(\tilde{f})| \leq 2d(\underline{E}, \underline{E}').$$

This inequality clearly contradict the previous inequality, and this proves the statement. □

**Corollary A.4.** *The function*

$$d' : \underline{\mathbb{E}}_{\mathscr{X}} \times \underline{\mathbb{E}}_{\mathscr{X}} \to \mathbb{R}_{\geq 0} : (\underline{E}, \underline{E}') \mapsto \sup\{|\underline{E}(f) - \underline{E}'(f)| : f \in \mathbb{G}(\mathscr{X}), \|f\| = 1\}$$

*is a metric, and* $(\underline{\mathbb{E}}_{\mathscr{X}}, d')$ *is a compact metric space.*

*Proof.* That $d'$ is a metric follows immediately from Lemma A.3↶. Even more, it is clear from Lemma A.3↶ that the metrics $d$ and $d'$ generate the same topology. Because $(\underline{\mathbb{E}}_{\mathscr{X}}, d)$ is compact due to Lemma A.2↶, we conclude that the metric space $(\underline{\mathbb{E}}_{\mathscr{X}}, d')$ is compact as well. □

## A.2 Compactness of the set of lower transition operators

Because the set of coherent lower expectations is compact, so is the set of lower transition operators.

**Proposition A.5.** *The space* $\underline{\mathfrak{T}}$ *of lower transition operators with the metric induced by the operator norm* $\|\bullet\|_{\text{op}}$ *is (sequentially) compact.*

*Proof.* Recall that a metric space is sequentially compact if and only if every sequence has a convergent subsequence (see Schechter, 1997, Section 17.26(c)), and that a metric space is compact if and only if it is sequentially compact (see Schechter, 1997, Theorem 17.33). To prove that $\underline{\mathfrak{T}}$ is (sequentially) compact, we fix an arbitrary sequence $(\underline{T}_n)_{n\in\mathbb{N}}$ in $\underline{\mathfrak{T}}$, and show that this sequence has a convergent subsequence. To that end, we recall from Corollary 3.61$_{107}$ that for every natural number $n$ and every state $x$ in $\mathscr{X}$, $[\underline{T}_n \bullet](x)$ is a coherent lower expectation on $\mathbb{G}(\mathscr{X})$. Because the state space $\mathscr{X}$ is finite, we can fix an ordering $x_1, \ldots, x_k$ of $\mathscr{X}$, with $k := |\mathscr{X}|$.

By Corollary A.4, the metric space $(\underline{\mathbb{E}}_{\mathscr{X}}, d')$ is compact. Hence, the sequence $([\underline{T}_n \bullet](x_1))_{n\in\mathbb{N}}$ has a convergent subsequence. More precisely, there is a subsequence $\big([\underline{T}_{n_{1,\ell}}\bullet](x_1)\big)_{\ell\in\mathbb{N}}$ – with $(n_{1,\ell})_{\ell\in\mathbb{N}}$ an increasing sequence of natural numbers – and a coherent lower expectation $\underline{E}_{x_1}$ on $\mathbb{G}(\mathscr{X})$ such that

$$\lim_{\ell\to+\infty} d'\Big([\underline{T}_{n_{1,\ell}}\bullet](x_1), \underline{E}_{x_1}\Big) = 0.$$

With the same argument, we obtain a convergent subsequence $\big([\underline{T}_{n_{2,\ell}}\bullet](x_2)\big)_{\ell\in\mathbb{N}}$ of $\big([\underline{T}_{n_{1,\ell}}\bullet](x_2)\big)_{\ell\in\mathbb{N}}$ – with $(n_{2,\ell})_{\ell\in\mathbb{N}}$ an increasing subsequence of $(n_{1,\ell})_{\ell\in\mathbb{N}}$ – and a coherent lower expectation $\underline{E}_{x_2}$ on $\mathbb{G}(\mathscr{X})$ such that

$$\lim_{\ell\to+\infty} d'\Big([\underline{T}_{n_{2,\ell}}\bullet](x_2), \underline{E}_{x_2}\Big) = 0.$$

Note that, because $(n_{2,\ell})_{\ell\in\mathbb{N}}$ is an increasing subsequence of $(n_{1,\ell})_{\ell\in\mathbb{N}}$,

$$\lim_{\ell\to+\infty} d'\Big([\underline{T}_{n_{2,\ell}}\bullet](x_1), \underline{E}_{x_1}\Big) = 0.$$

It is clear that we can use the same argument for the remaining states $x_3, \ldots, x_k$. Eventually, we obtain the coherent lower expectations $\underline{E}_{x_1}, \ldots, \underline{E}_{x_k}$ and an increasing sequence $(n_{k,\ell})_{\ell\in\mathbb{N}}$ of natural numbers such that

$$\lim_{\ell\to+\infty} d'\Big([\underline{T}_{n_{k,\ell}}\bullet](x), \underline{E}_{x}\Big) = 0 \quad \text{for all } x \in \mathscr{X}.$$

We take the maximum over all $x$ in $\mathscr{X}$ and change the order of the maximum and the limit – which is allowed because we take the maximum of a finite number of real-valued limits – to yield

$$\lim_{\ell\to+\infty} \max\Big\{d'\Big([\underline{T}_{n_{k,\ell}}\bullet](x), \underline{E}_{x}\Big) \colon x \in \mathscr{X}\Big\} = 0. \tag{A.7}$$

Let $\underline{T}$ be the operator on $\mathbb{G}(\mathscr{X})$ defined by

$$[\underline{T}f](x) := \underline{E}_x(f) \quad \text{for all } f \in \mathbb{G}(\mathscr{X}) \text{ and } x \in \mathscr{X}.$$

It is a matter of straightforward verification that $\underline{T}$ is a lower transition operator. We now set out to verify that $(\underline{T}_{n_{k,\ell}})_{\ell\in\mathbb{N}}$ converges to $\underline{T}$. To this end, we fix any natural number $\ell$ and observe that

$$\begin{aligned}
\|\underline{T}_{n_{k,\ell}} - \underline{T}\|_{\mathrm{op}} &= \sup\Big\{\max\Big\{\Big|[\underline{T}_{n_{k,\ell}}f](x) - [\underline{T}f](x)\Big| \colon x \in \mathscr{X}\Big\} \colon f \in \mathbb{G}(\mathscr{X}), \|f\| = 1\Big\} \\
&= \max\Big\{\sup\Big\{\Big|[\underline{T}_{n_{k,\ell}}f](x) - [\underline{T}f](x)\Big| \colon f \in \mathbb{G}(\mathscr{X}), \|f\| = 1\Big\} \colon x \in \mathscr{X}\Big\} \\
&= \max\Big\{\sup\Big\{\Big|[\underline{T}_{n_{k,\ell}}f](x) - \underline{E}_x(f)\Big| \colon f \in \mathbb{G}(\mathscr{X}), \|f\| = 1\Big\} \colon x \in \mathscr{X}\Big\} \\
&= \max\Big\{d'\Big([\underline{T}_{n_{k,\ell}}\bullet](x), \underline{E}_x\Big) \colon x \in \mathscr{X}\Big\}.
\end{aligned}$$

From this and Eq. (A.7), we infer that

$$\lim_{\ell\to+\infty} \|\underline{T}_{n_{k,\ell}} - \underline{T}\|_{\mathrm{op}} = 0,$$

so the subsequence $\big(\underline{T}_{n_{k,\ell}}\big)_{\ell\in\mathbb{N}}$ converges to the lower transition operator $\underline{T}$, as we set out to verify.

To summarise, we have shown that the sequence $(\underline{T}_n)_{n\in\mathbb{N}}$ has a convergent subsequence. Because the sequence $(\underline{T}_n)_{n\in\mathbb{N}}$ was arbitrary, we conclude that the space $\underline{\mathfrak{T}}$ is (sequentially) compact, as we set out to prove. □

## A.3 Compactness of a bounded set of lower rate operators

Finally, we are ready to establish that for a given upper bound $\beta$ in $\mathbb{R}_{\geq 0}$, the set of all lower rate operators with norm not greater than the bound $\beta$ is compact.

**Proposition A.6.** *Consider any non-negative real number $\beta$. The space*

$$\underline{\mathfrak{Q}}^{\beta} := \{\underline{Q} \in \underline{\mathfrak{Q}} \colon \|\underline{Q}\|_{\mathrm{op}} \leq \beta\}$$

*with the metric induced by the operator norm $\|\bullet\|_{\mathrm{op}}$ is (sequentially) compact.*

*Proof.* Again, because $\underline{\mathfrak{Q}}^{\beta}$ is a metric space, it suffices to show that $\underline{\mathfrak{Q}}^{\beta}$ is sequentially compact. To that end, we fix any sequence $(\underline{Q}_n)_{n\in\mathbb{N}}$ in $\underline{\mathfrak{Q}}^{\beta}$ and any $\Delta$ in $\mathbb{R}_{>0}$ such that $\Delta\beta \leq 2$. Observe that, for all $n$ in $\mathbb{N}$, $\Delta\|\underline{Q}_n\| \leq \Delta\beta \leq 2$, so

$$\underline{T}_n := I + \Delta\underline{Q}_n$$

is a lower transition operator due to Lemma 3.72<sub>112</sub>. Because $\underline{\mathfrak{T}}$ is a (sequentially) compact metric space due to Proposition A.5<sub>446</sub>, the sequence $(\underline{T}_n)_{n\in\mathbb{N}}$ has a convergent subsequence. More precisely, there is an increasing sequence $(n_k)_{k\in\mathbb{N}}$ of natural numbers and a lower transition operator $\underline{T}$ such that $(\underline{T}_{n_k})_{k\in\mathbb{N}}$ converges to $\underline{T}$, in the sense that $\lim_{k\to+\infty}\|\underline{T}_{n_k} - \underline{T}\| = 0$.

Set $\underline{Q} := (\underline{T} - I)/\Delta$. Recall from Lemma 3.73<sub>113</sub> that $\underline{Q}$ is a lower rate operator. Observe that, for any natural number $n$,

$$\|\underline{Q}_n - \underline{Q}\|_{\mathrm{op}} = \frac{1}{\Delta}\|\Delta\underline{Q}_n - \Delta\underline{Q}\|_{\mathrm{op}} = \frac{1}{\Delta}\|(I + \Delta\underline{Q}_n) - (I + \Delta\underline{Q})\|_{\mathrm{op}} = \frac{1}{\Delta}\|\underline{T}_n - \underline{T}\|_{\mathrm{op}}.$$

Therefore,

$$\lim_{k\to+\infty}\|\underline{Q}_{n_k} - \underline{Q}\|_{\mathrm{op}} = \lim_{k\to+\infty}\frac{1}{\Delta}\|\underline{T}_{n_k} - \underline{T}\|_{\mathrm{op}} = 0,$$

so $(\underline{Q}_{n_k})_{k\in\mathbb{N}}$ converges to the lower rate operator $\underline{Q}$. It remains for us to show that $\underline{Q}$ belongs to $\underline{\mathfrak{Q}}^{\beta}$, that is, that $\|\underline{Q}\|_{\mathrm{op}} \leq \beta$. To that end, we fix a positive real number $\epsilon$. By definition of the operator norm, there is a gamble $f$ on $\mathscr{X}$ such that $\|f\| = 1$ and

$$\|\underline{Q}f\| > \|\underline{Q}\|_{\mathrm{op}} - \frac{\epsilon}{2},$$

and, since $\lim_{k\to+\infty}\|\underline{Q}_{n_k} - \underline{Q}\| = 0$, a natural number $k$ such that

$$\|\underline{Q}_{n_k} - \underline{Q}\|_{\mathrm{op}} < \frac{\epsilon}{2}.$$

Observe also that

$$\|\underline{Q}f\| = \|\underline{Q}f - \underline{Q}_{n_k}f + \underline{Q}_{n_k}f\| \leq \|\underline{Q}_{n_k}f\| + \|\underline{Q}_{n_k}f - \underline{Q}f\| \leq \|\underline{Q}_{n_k}\|_{\mathrm{op}} + \|\underline{Q}_{n_k} - \underline{Q}\|_{\mathrm{op}},$$

where for the last inequality we have used that $\|f\| = 1$. Combining the three preceding inequalities, we see that

$$\|\underline{Q}\|_{\mathrm{op}} < \|\underline{Q}f\| + \frac{\epsilon}{2} \leq \|\underline{Q}_{n_k}\|_{\mathrm{op}} + \|\underline{Q}_{n_k} - \underline{Q}\|_{\mathrm{op}} + \frac{\epsilon}{2} < B + \frac{\epsilon}{2} + \frac{\epsilon}{2} = B + \epsilon.$$

Because $\epsilon$ was an arbitrary positive real number, we infer from this that $\|\underline{Q}\|_{\mathrm{op}} \leq \beta$, which completes our proof. □

As a more or less immediate corollary, we also obtain that the similar set of rate operators whose norm is not greater than a given upper bound $\beta$, is (sequentially) compact.

**Corollary A.7.** *Consider any non-negative real number $\beta$. The space*

$$\mathfrak{Q}^{\beta} := \{Q \in \mathfrak{Q} : \|Q\|_{\mathrm{op}} \leq \beta\}$$

*with the metric induced by the operator norm $\|\bullet\|_{\mathrm{op}}$ is (sequentially) compact.*

*Proof.* Because every rate operator is a lower rate operator, it is clear that $\mathfrak{Q}^{\beta}$ is a subset of $\underline{\mathfrak{Q}}^{\beta}$. Thus, it follows from Proposition A.6↶ that any sequence $(Q_n)_{n\in\mathbb{N}}$ in $\mathfrak{Q}^{\beta}$ has a subsequence that converges to some lower rate operator in $\underline{\mathfrak{Q}}^{\beta}$. However, because the properties (R1)$_{81}$–(R4)$_{81}$ of a rate operator are preserved when taking limits, it is clear that the limit of any convergent subsequence of $(Q_n)_{n\in\mathbb{N}}$ is a rate operator, and therefore belongs to $\mathfrak{Q}^{\beta}$. This proves that $\mathfrak{Q}^{\beta}$ is (sequentially) compact. □

# Daniell integration B

In this appendix, we prove those results of Section $5.1_{216}$ that were stated without proof. More specifically, we focus on the Daniell extension in Appendix B.1 and on the link with coherence in Appendix $B.2_{458}$.

## B.1 Daniell extension

First, we establish that the expectation $E_P$ with respect to a countably additive probability charge $P$ is a so-called 'elementary integral' in Appendix B.1.1. Next, we take a look at some general properties of the Daniell extension $E_P^{\mathrm{D}}$ in Appendix $B.1.2_{453}$, and at the two convergence theorems in particular in Appendix $B.1.3_{458}$.

### B.1.1 Elementary integrals

In order to use the machinery of the Daniell extension, the expectation $E_P$ on $\mathbb{S}(\mathcal{F})$ should satisfy two crucial requirements. First, its domain $\mathbb{S}(\mathcal{F})$ should be a vector lattice, meaning that it is a real vector space that is closed under taking point-wise minima and maxima. Second, and most importantly, $E_P$ should be an elementary integral, meaning that it is an additive, homogeneous and monotone functional that is continuous at zero (see Taylor, 1985, Section 6-1).

**Definition B.1.** Consider a possibility space $\mathcal{X}$. Let $E$ be a real-valued functional on a non-empty subset $\mathcal{G}$ of $\mathbb{V}(\mathcal{X})$. Then $\mathcal{G}$ is a called a *vector lattice* if

LL1. $\mu f \in \mathcal{G}$ for all $\mu \in \mathbb{R}$ and $f \in \mathcal{G}$,

LL2. $f + g \in \mathcal{G}$ for all $f, g \in \mathcal{G}$,

LL3. $f \wedge g \in \mathcal{G}$ for all $f, g \in \mathcal{G}$,

LL4. $f \vee g \in \mathcal{G}$ for all $f, g \in \mathcal{G}$.

Whenever this is the case, $E$ is called an *elementary integral* if furthermore

EI1. $E(\mu f) = \mu E(f)$ for all $\mu \in \mathbb{R}$ and $f \in \mathcal{G}$,

EI2. $E(f+g) = E(f) + E(g)$ for all $f, g \in \mathcal{G}$,

EI3. $E(f) \leq E(g)$ for all $f, g \in \mathcal{G}$ such that $f \leq g$,

EI4. $\lim_{n\to+\infty} E(f_n) = 0$ for any sequence $(f_n)_{n\in\mathbb{N}} \searrow 0$ in $\mathcal{G}$.

To verify that $E_P$ is an elementary integral, we need to check whether its domain $\mathbb{S}(\mathcal{F})$ is a vector lattice. Recall from Lemma $2.39_{36}$ that the set $\mathbb{S}(\mathcal{F})$ of $\mathcal{F}$-simple variables is a vector space – that is, satisfies (LL1)↶ and (LL2)↶. Hence, it remains for us to verify that $\mathbb{S}(\mathcal{F})$ is closed under taking point-wise minima and maxima – that is, satisfies (LL3)↶ and (LL4)↶. This result is essentially well-known; see, for example, D. Williams's (1991, Section 5.1) work.

**Lemma B.2.** *Consider a field of events $\mathcal{F}$ over a possibility space $\mathcal{X}$. Then for all $\mathcal{F}$-simple variables $f$ and $g$ in $\mathbb{S}(\mathcal{F})$, $f \wedge g$ and $f \vee g$ are also $\mathcal{F}$-simple variables.*

Next, we verify that $E_P$ is an elementary integral. To that end, we recall from Proposition $2.42_{37}$ that $E_P$ always satisfies (EI1)–(EI3). Second, we recall from Definition $5.4_{221}$ that, by definition, $E_P$ furthermore satisfies (EI4) if and only if $E_P$ is countably additive. Here, we prove that the three conditions in Definition $5.4_{221}$ are equivalent.

**Definition 5.4.** Consider a probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then the following three conditions are equivalent. Whenever $P$ satisfies one (and hence all) of them, we call $P$ *countably additive*.

(i) For any $\mathcal{F}$-simple variable $f$ and any sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables such that $(f_n)_{n\in\mathbb{N}} \nearrow f$, $E_P(f) = \lim_{n\to+\infty} E_P(f_n)$.

(ii) For any $\mathcal{F}$-simple variable $f$ and any sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables such that$(f_n)_{n\in\mathbb{N}} \searrow f$, $E_P(f) = \lim_{n\to+\infty} E_P(f_n)$.

(iii) For any sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables such that $(f_n)_{n\in\mathbb{N}} \searrow 0$, $\lim_{n\to+\infty} E_P(f_n) = 0$.

*Proof.* First, we prove that $(i)_{221}$ is equivalent to $(ii)_{222}$. To that end, we observe that $(f_n)_{n\in\mathbb{N}} \searrow f$ if and only if $(-f_n)_{n\in\mathbb{N}} \nearrow -f$. Thus the equivalence of $(i)_{221}$ and $(ii)_{222}$ follows immediately from Lemma $2.39_{36}$ and $(ES2)_{37}$.

Next, since $E_P(0) = 0$ because of $(ES1)_{37}$, it is clear that $(ii)_{222}$ implies $(iii)_{222}$. Thus, it remains for us to prove that $(iii)_{222}$ implies either $(i)_{221}$ or $(ii)_{222}$; we will prove the former implication. To that end, we fix some $f$ in $\mathbb{S}(\mathcal{F})$ and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{S}(\mathcal{F})$ such that $(f_n)_{n\in\mathbb{N}} \nearrow f$. For any natural number $n$, we let $f'_n := f - f_n$. Observe that $f'_n$ is an $\mathcal{F}$-simple variable due to Lemma $2.39_{36}$, and that by construction

$$f'_n = f - f_n \geq f - f_{n+1} = f'_{n+1} \quad \text{for all } n \in \mathbb{N}$$

and $\lim_{n\to+\infty} f'_n(x) = \lim_{n\to+\infty} f(x) - f'_n(x) = 0$ for all $x$ in $\mathcal{X}$. Consequently, $(f'_n)_{n\in\mathbb{N}}$ is a non-increasing sequence of $\mathcal{F}$-simple variables that converges point-wise to 0. Therefore, it follows from (iii)$_{222}$ that $\lim_{n\to+\infty} E_P(f'_n) = 0$. Furthermore,

$$E_P(f_n) = E_P(f - f'_n) = E_P(f) - E_P(f'_n) \quad \text{for all } n \in \mathbb{N},$$

where the last equality follows from (ES3)$_{37}$ and (ES2)$_{37}$. Because all of the terms in these equalities are real-valued due to (ES1)$_{37}$, we infer from this all that

$$\lim_{n\to+\infty} E_P(f_n) = \lim_{n\to+\infty} \big(E_P(f) - E_P(f'_n)\big) = E_P(f) - \lim_{n\to+\infty} E_P(f'_n) = E_P(f),$$

as required. □

Combining the foregoing results, we see that the expectation $E_P$ with respect to a countably additive probability charge is an elementary integral. The following result formally establishes that our setting falls squarely within the scope of Taylor's (1985) discussion, and it is for this reason that we may (implicitly) use his results.

**Corollary B.3.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then $\mathbb{S}(\mathcal{F})$ is a vector lattice and $E_P$ is an elementary integral on $\mathbb{S}(\mathcal{F})$.*

*Proof.* That $\mathbb{S}(\mathcal{F})$ is a vector lattice follows immediately from Lemmas 2.39$_{36}$ and B.2$_{\curvearrowleft}$; that $E_P$ is an elementary integral on $\mathbb{S}(\mathcal{F})$ follows immediately from Proposition 2.42$_{37}$ and Definition 5.4$_{221}$. □

### B.1.2 Properties of the (inner and outer) Daniell integral

In Eq. (5.5)$_{222}$, we claimed that $\overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F}) = -\overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})$. The following result establishes this equality, as well as some other convenient properties of $\mathcal{F}$-over and $\mathcal{F}$-under variables; for a proof, we refer to (Taylor, 1985, Section 6-2).

**Lemma B.4.** *Consider a field $\mathcal{F}$ of events over some possibility space $\mathcal{X}$. Then for all $\mathcal{F}$-over variables $f$ and $g$ in $\overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})$,*

(i) *$-f \in \overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$;*

(ii) *$\lambda f \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})$ for all $\lambda \in \mathbb{R}_{\geq 0}$;*

(iii) *$f + g$ in $\overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})$.*

*Similar properties hold for the $\mathcal{F}$-under variables.*

Next, we prove that the limit in the definition of $E_P^{\mathrm{mc}}$ does not depend on the defining sequence $(f_n)_{n\in\mathbb{N}}$.

**Lemma 5.6.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$, and some $f$ in $\overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathcal{F})$. If $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ are monotone sequences of $\mathcal{F}$-simple variables that both converge point-wise to $f$, then*

$$\lim_{n\to+\infty} E_P(f_n) = \lim_{n\to+\infty} E_P(g_n).$$

*Proof.* For the case that $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ are non-decreasing, the equality of the statement is proven by Taylor (1985, Lemma 6-2 II). For the case that $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ are non-increasing, the equality of the statement follows from the previous. More precisely, in this case $(-f_n)_{n\in\mathbb{N}}$ and $(-g_n)_{n\in\mathbb{N}}$ are non-decreasing sequences of $\mathcal{F}$-simple variables that converge to the $\mathcal{F}$-over variable $-f$. Because $E_P(f_n) = -E_P(-f_n)$ and $E_P(g_n) = -E_P(-g_n)$ due to (ES2)$_{37}$, it now follows from the previous that

$$\lim_{n\to+\infty} E_P(f_n) = -\lim_{n\to+\infty} E_P(-f_n) = -\lim_{n\to+\infty} E_P(-g_n) = \lim_{n\to+\infty} E_P(g_n),$$

as required.

What remains is the case that $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ are not both non-decreasing or non-increasing; without loss of generality, we assume that $(f_n)_{n\in\mathbb{N}}$ is non-decreasing and that $(g_n)_{n\in\mathbb{N}}$ is non-increasing. Observe that $f_n \leq f_{n+1} \leq f \leq g_{n+1} \leq g_n$ for all $n$ in $\mathbb{N}$. Consequently, it follows from (ES1)$_{37}$ that $\lim_{n\to+\infty} E_P(f_n)$ and $\lim_{n\to+\infty} E_P(g_n)$ are real-valued. Furthermore, from these inequalities and Lemma 2.39$_{36}$, we infer that $(g_n - f_n)_{n\in\mathbb{N}}$ is a non-increasing sequence of non-negative $\mathcal{F}$-simple variables. Observe that $\lim_{n\to+\infty}(g_n - f_n) = 0$ because $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ converge to the same limit $f$. For this reason, it follows from Definition 5.4$_{221}$ (iii) that

$$\lim_{n\to+\infty} E_P(g_n - f_n) = 0.$$

Observe now that by (ES3)$_{37}$ and (ES2)$_{37}$,

$$E_P(g_n - f_n) = E_P(g_n) - E_P(f_n) \quad \text{for all } n \in \mathbb{N}.$$

Taking the limit for $n$ going to $+\infty$ on both sides of the equality, we see that

$$0 = \lim_{n\to+\infty}\big(E_P(g_n) - E_P(f_n)\big) = \lim_{n\to+\infty} E_P(g_n) - \lim_{n\to+\infty} E_P(f_n),$$

where the second equality holds because the limits of the right-hand side are real-valued. From this equality, we infer that

$$\lim_{n\to+\infty} E_P(f_n) = \lim_{n\to+\infty} E_P(g_n),$$

which is the equality we were after. □

We continue with a list of convenient properties of the expectation $E_P^{\mathrm{mc}}$ defined on the set $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F})$ of all $\mathcal{F}$-over and $\mathcal{F}$-under variables (see Taylor, 1985, Section 6-2).

**Lemma B.5.** *Consider a countably additive probability charge $P$ on a field $\mathcal{F}$ of events over some possibility space $\mathcal{X}$. Then for all $f$ and $g$ in $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F})$ and $\lambda$ in $\mathbb{R}_{\geq 0}$,*

(i) $E_P^{\mathrm{mc}}(f) = E_P(f)$ *whenever $f$ is $\mathcal{F}$-simple;*

(ii) $E_P^{\mathrm{mc}}(f) = -E_P^{\mathrm{mc}}(-f)$*;*

(iii) $E_P^{\mathrm{mc}}(\lambda f) = \lambda E_P^{\mathrm{mc}}(f)$*;*

(iv) $E_P^{\mathrm{mc}}(f+g) = E_P^{\mathrm{mc}}(f) + E_P^{\mathrm{mc}}(g)$ *if either $f, g \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})$ or $f, g \in \overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$;*

(v) $E_P^{\mathrm{mc}}(f) \leq E_P^{\mathrm{mc}}(g)$ *if* $f \leq g$ *and either* $f, g \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})$ *or* $f, g \in \overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$*;*

(vi) $E_P^{\mathrm{mc}}(f) \leq E_P^{\mathrm{mc}}(g)$ *whenever* $f \leq g$*,* $f \in \overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$ *and* $g \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})$*.*

The inner and outer Daniell extensions satisfy some convenient properties. The observant reader will notice that they are similar to the properties of coherent lower expectations, but this should not come as a surprise seeing that the inner Daniell extension is defined as a supremum and the outer Daniell extension as an infimum. Taylor (1985, Section 6-3) proves many of these properties in some shape or form, but we nevertheless provide a proof for the sake of completeness.

**Lemma B.6.** *Consider a countably additive probability charge* $P$ *on a field of events* $\mathcal{F}$ *over some possibility space* $\mathcal{X}$*. Then for any extended real variables* $f$ *and* $g$ *in* $\overline{\mathbb{V}}(\mathcal{X})$ *and any non-negative real number* $\lambda$*,*

(i) $E_P^{\mathrm{i}}(f) = E_P^{\mathrm{o}}(f) = E_P^{\mathrm{mc}}(f)$ *if* $f \in \overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F})$*;*

(ii) $E_P^{\mathrm{i}}(f) = E_P^{\mathrm{o}}(f) = E_P^{\mathrm{mc}}(f) = E_P(f)$ *if* $f \in \mathbb{S}(\mathcal{F})$*;*

(iii) $\inf f \leq E_P^{\mathrm{i}}(f) \leq E_P^{\mathrm{o}}(f) \leq \sup f$*;*

(iv) $E_P^{\mathrm{o}}(f) = -E_P^{\mathrm{i}}(-f)$*;*

(v) $E_P^{\mathrm{i}}(\lambda f) = \lambda E_P^{\mathrm{i}}(f)$ *and* $E_P^{\mathrm{o}}(\lambda f) = \lambda E_P^{\mathrm{o}}(f)$*;*

(vi) $E_P^{\mathrm{i}}(f) \leq E_P^{\mathrm{i}}(g)$ *and* $E_P^{\mathrm{o}}(f) \leq E_P^{\mathrm{o}}(g)$ *whenever* $f \leq g$*.*

*Proof.* For a proof of (i), we refer to (Taylor, 1985, Theorem 6-3 I). Property (ii) follows immediately from (i) and Lemma B.5 (i).

Note that in his definition of the outer Daniell integral $E_P^{\mathrm{o}}(f)$ of the extended real variable $f$, Taylor (1985) requires that there should be an $\mathcal{F}$-over variable $h$ such that $f \leq h$; similarly, for the inner Daniell integral $E_P^{\mathrm{i}}(f)$ he demands that there should be an $\mathcal{F}$-under variable $g$ such that $f \geq g$. In our setting, these two conditions are always satisfied because $\mathbb{S}(\mathcal{F})$ contains all constant gambles. To see why exactly, we recall from Lemma 2.39<sub>36</sub> that $\mu$ is an $\mathcal{F}$-simple variable for all $\mu$ in $\mathbb{R}$. Let $(\mu_n)_{n\in\mathbb{N}}$ be a non-decreasing sequence of real numbers such that $\lim_{n\to+\infty} \mu_n = +\infty$. Then $h := \text{p-w}\lim_{n\to+\infty} \mu_n$ is an $\mathcal{F}$-over variable, and $h \geq f$. Similarly, $g := \lim_{n\to+\infty} -\mu_n$ is an $\mathcal{F}$-under variable such that $g \leq f$.

Similar reasoning can be used to verify the outer inequalities in (iii). First, we deal with the case that $\inf f < +\infty$. Then we let $(\mu_n)_{n\in\mathbb{N}}$ be a non-increasing sequence of real numbers with $\lim_{n\to+\infty} \mu_n = \inf f$, so $h := \text{p-w}\lim_{n\to+\infty} \mu_n$ is a $\mathcal{F}$-under variable such that

$$h = \text{p-w}\lim_{n\to+\infty} \mu_n = \inf f \leq f.$$

Because $E_P(\mu_n) = \mu_n$ due to (ES1)<sub>37</sub>, we see that

$$E_P^{\mathrm{mc}}(h) = \lim_{n\to+\infty} E_P(\mu_n) = \lim_{n\to+\infty} \mu_n = \inf f$$

For this reason, $E_P^{\mathrm{i}}(f) \geq \inf f$. Second, we deal with the case that $\inf f = +\infty$. Then for all $a$ in $\mathbb{R}$, $a \leq f$ and therefore

$$E_P^{\mathrm{i}}(f) \geq E_P^{\mathrm{mc}}(a) = E_P(a) = a,$$

where the equalities follow from (ii)$_\curvearrowleft$ and (ES1)$_{37}$, respectively. Clearly, this implies that $E_P^{\mathrm{i}}(f) = +\infty = \inf f$, and therefore $\inf f \leq E_P^{\mathrm{i}}(f)$. An analoguous argument proves that $E_P^{\mathrm{o}}(f) \leq \sup f$. The inner inequality of (iii)$_\curvearrowleft$ follows almost immediately from Lemma B.5$_{454}$ (vi), because it follows from the latter that for any $g$ in $\overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$ and $h$ in $\overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F})$ such that $g \leq f \leq h$,

$$E_P^{\mathrm{mc}}(g) \leq E_P^{\mathrm{mc}}(h).$$

To verify (iv)$_\curvearrowleft$, we observe that

$$\begin{aligned}
E_P^{\mathrm{o}}(f) &= \inf\{E_P^{\mathrm{mc}}(h)\colon h \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F}), h \geq f\} = \inf\{-E_P^{\mathrm{mc}}(-h)\colon h \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F}), h \geq f\} \\
&= -\sup\{E_P^{\mathrm{mc}}(-h)\colon h \in \overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F}), h \geq f\} = -\sup\{E_P^{\mathrm{mc}}(h)\colon h \in \overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F}), h \leq -f\} \\
&= -E_P^{\mathrm{i}}(-f),
\end{aligned}$$

where the second equality follows from Lemma B.5$_{454}$ (ii) and the fourth inequality holds because $\overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F}) = -\overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$ – see Lemma B.4$_{453}$ (i).

Note that we only need to prove (v)$_\curvearrowleft$ and (vi)$_\curvearrowleft$ for the inner Daniel expectation $E_P^{\mathrm{i}}$, as the statement for the outer Daniel expectation $E_P^{\mathrm{o}}$ then follows from (iv)$_\curvearrowleft$.

Property (v)$_\curvearrowleft$ for $\lambda = 0$ follows immediately from (iii)$_\curvearrowleft$. For $\lambda > 0$, this property follows almost immediately from Eq. (5.8)$_{224}$ because by Lemma B.5$_{454}$ (iii),

$$E_P^{\mathrm{mc}}(h) = \lambda E_P^{\mathrm{mc}}\left(\frac{h}{\lambda}\right)$$

for any $\mathcal{F}$-under variable $h$ such that $h \leq \lambda f$.

Finally, (vi)$_\curvearrowleft$ follows immediately from Eq. (5.8)$_{224}$ because for any $\mathcal{F}$-under variable $h$ such that $h \leq f$, $h \leq g$ as well. □

We need these properties of the inner and outer extension to establish the following properties of the Daniell extension $E_P^{\mathrm{D}}$ on $\mathbb{D}_P^{\mathrm{D}}$.

**Theorem 5.9.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then*

DE1. $\mathbb{S}(\mathcal{F}) \subseteq \mathbb{D}_P^{\mathrm{D}}$ *and* $E_P^{\mathrm{D}}(f) = E_P(f)$ *for all $f$ in* $\mathbb{S}(\mathcal{F})$*;*

DE2. $\mathbb{S}(\mathcal{F}) \subseteq \overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F}) \subseteq \mathbb{D}_P^{\mathrm{D}}$ *and* $E_P^{\mathrm{D}}(f) = E_P^{\mathrm{mc}}(f)$ *for all $f$ in* $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F})$.

*Furthermore, for all D-integrable extended real variables $f$ and $g$ in $\mathbb{D}_P^{\mathrm{D}}$ and all real numbers $\mu$ in $\mathbb{R}$,*

DE3. $\inf f \leq E_P^{\mathrm{D}}(f) \leq \sup f$*;*

DE4. *$\mu f$ is D-integrable and* $E_P^{\mathrm{D}}(\mu f) = \mu E_P^{\mathrm{D}}(f)$*;*

DE5. *$f + g$ is D-integrable and $E_P^{\mathrm{D}}(f+g) = E_P^{\mathrm{D}}(f) + E_P^{\mathrm{D}}(g)$ whenever $f+g$ and $E_P^{\mathrm{D}}(f) + E_P^{\mathrm{D}}(g)$ are well-defined;*

DE6. $E_P^{\mathrm{D}}(f) \leq E_P^{\mathrm{D}}(g)$ *whenever* $f \leq g$.

*Finally, for all D-integrable real variables $f$ and $g$ in $\tilde{\mathbb{D}}_P^{\mathrm{D}}$,*

DE7. *$f \vee g$ and $f \wedge g$ also belong to $\tilde{\mathbb{D}}_P^{\mathrm{D}}$.*

*Proof.* Properties (DE1)$_{225}$ and (DE2)$_{225}$ follow immediately from (ii) and (i) of Lemma B.6$_{455}$, respectively. Similarly, (DE3)$_{225}$ follows immediately from Lemma B.6$_{455}$ (iii), and (DE6)$_{226}$ follows immediately from Lemma B.6$_{455}$ (vi).

Next, we verify (DE4)$_{225}$, and we will distinguish two cases based on the sign of $\mu$. In case $\mu \geq 0$, it follows from Lemma B.6$_{455}$ (v) that

$$E^{\mathrm{i}}_P(\mu f) = \mu E^{\mathrm{i}}_P(f) = \mu E^{\mathrm{D}}_P(f),$$

where for the last equality we have used that $f$ is D-integrable. Similarly $E^{\mathrm{o}}_P(\mu f) = \mu E^{\mathrm{o}}_P(f) = \mu E^{\mathrm{D}}_P(f)$. Because $E^{\mathrm{i}}_P(\mu f) = E^{\mathrm{o}}_P(\mu f) = \mu E^{\mathrm{D}}_P(f)$, $\mu f$ is D-integrable and $E^{\mathrm{D}}_P(\mu f) = \mu E^{\mathrm{D}}_P(f)$. In case $\mu < 0$, then it follows from properties (iv) and (v) of Lemma B.6$_{455}$ that

$$E^{\mathrm{i}}_P(\mu f) = -E^{\mathrm{o}}_P(-\mu f) = \mu E^{\mathrm{o}}_P(f) = \mu E^{\mathrm{D}}_P(f),$$

where for the last equality we have used that $f$ is D-integrable. Similarly, we find that $E^{\mathrm{o}}_P(\mu f) = \mu E^{\mathrm{D}}_P(f)$. Consequently, $\mu f$ is D-integrable and $E^{\mathrm{D}}_P(\mu f) = \mu E^{\mathrm{D}}_P(f)$.

Taylor (1985, Theorem 6–3 II) proves (DE5)$_{225}$ in case that $f$ and $g$ belong to $\tilde{\mathbb{D}}^{\mathrm{D}}_P$ – that is, in case their Daniell integrals are real-valued. For this reason, we only need to prove this property in case $E^{\mathrm{D}}_P(f)$ (and/or $E^{\mathrm{D}}_P(g)$) is not real-valued. We will only consider the case $E^{\mathrm{D}}_P(f) = +\infty$, the proof for $E^{\mathrm{D}}_P(f) = -\infty$ is analoguous.

Because $E^{\mathrm{D}}_P(f) + E^{\mathrm{D}}_P(g)$ is well-defined and $E^{\mathrm{D}}_P(f) = +\infty$, $E^{\mathrm{D}}_P(g)$ is either real-valued or equal to $+\infty$. We will only consider the case that $E^{\mathrm{D}}_P(g)$ is real-valued, the proof for the case that $E^{\mathrm{D}}_P(g) = +\infty$ is similar. To prove that $f+g$ is D-integrable, we take a closer look at the inner Daniell expectation $E^{\mathrm{i}}_P(f+g)$.

Fix any arbitrary natural number $N$. Because $E^{\mathrm{i}}_P(f) = E^{\mathrm{D}}_P(f) = +\infty$, it follows from Eqs. (5.6)$_{223}$ and (5.8)$_{224}$ that there is a non-increasing sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables that converges point-wise to some $\mathcal{F}$-under variable $h_f$ with $h_f \leq f$ and

$$\lim_{n\to+\infty} E_P(f_n) = E^{\mathrm{mc}}_P(h_f) \geq N + \frac{1}{N}. \tag{B.1}$$

Similarly, because $E^{\mathrm{i}}_P(g) = E^{\mathrm{D}}_P(g)$ is finite, there is a non-increasing sequence $(g_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables that converges point-wise to some $\mathcal{F}$-under variable $h_g$ with $h_g \leq g$ and

$$E^{\mathrm{D}}_P(g) - \frac{1}{N} \leq \lim_{n\to+\infty} E_P(g_n) = E^{\mathrm{mc}}_P(h_g) \leq E^{\mathrm{D}}_P(g). \tag{B.2}$$

Observe that $(f_n + g_n)_{n\in\mathbb{N}}$ is a non-increasing sequence of $\mathcal{F}$-simple variables that converges point-wise to the $\mathcal{F}$-under variable $h_f + h_g$. Therefore, it follows from Lemma B.5$_{454}$ (iv) that

$$E^{\mathrm{mc}}_P(h_f + h_g) = E^{\mathrm{mc}}_P(h_f) + E^{\mathrm{mc}}_P(h_g) \geq N + \frac{1}{N} + E^{\mathrm{D}}_P(g) - \frac{1}{N} = N + E^{\mathrm{D}}_P(g), \tag{B.3}$$

where the first inequality follows from Eqs. (B.1) and (B.2). Observe furthermore that because $h_f \leq f$ and $h_g \leq g$, $h_f + h_g \leq f + g$. It follows from this, Eq. (B.3) and Eq. (5.8)$_{224}$ that

$$E^{\mathrm{i}}_P(f+g) \geq E^{\mathrm{mc}}_P(h_f + h_g) \geq N + E^{\mathrm{D}}_P(g).$$

Because $N$ is an arbitrary natural number, we conclude that $E_P^{\mathrm{i}}(f+g) = +\infty$. By Lemma B.6455 (iii), this implies that $E_P^{\mathrm{o}}(f+g) = +\infty$ as well, so $f+g$ is D-integrable and $E_P^{\mathrm{D}}(f+g) = +\infty$.

Finally, Taylor (1985, Theorem 6-3 II (c)) proves (DE7)226. □

### B.1.3 Limit properties of the Daniell integral

Next, we examine the limit properties of the Daniell expectation. First up is the Monotone Convergence Theorem.

**Theorem 5.10.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Let $(f_n)_{n\in\mathbb{N}}$ be a non-decreasing sequence of D-integrable variables with $E_P^{\mathrm{D}}(f_1) > -\infty$. Then the point-wise limit of $(f_n)_{n\in\mathbb{N}}$ is D-integrable, and*

$$E_P^{\mathrm{D}}\left(\operatorname*{p-w\,lim}_{n\to+\infty} f_n\right) = \lim_{n\to+\infty} E_P^{\mathrm{D}}(f_n).$$

*The same holds in case $(f_n)_{n\in\mathbb{N}}$ is non-increasing and $E_P^{\mathrm{D}}(f_1) < +\infty$.*

*Proof.* Taylor (1985, Theorem 6-3 III) proves this for the case that $E_P^{\mathrm{D}}(f_n)$ is real-valued for every natural number $n$ and $\lim_{n\to+\infty} E_P^{\mathrm{D}}(f_n)$ is real-valued as well. In case $\lim_{n\to+\infty} E_P^{\mathrm{D}}(f_n)$ is not real-valued, the statement follows almost immediately from Lemma B.6455 (vi). □

Second, we have Lebesgue's Dominated Convergence Theorem.

**Theorem 5.11.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of D-integrable variables that converges point-wise. If there is a D-integrable variable $g$ with $E_P^{\mathrm{D}}(g) < +\infty$ such that $|f_n| \leq g$ for all $n$ in $\mathbb{N}$, then the point-wise limit of $(f_n)_{n\in\mathbb{N}}$ is D-integrable, and*

$$E_P^{\mathrm{D}}\left(\operatorname*{p-w\,lim}_{n\to+\infty} f_n\right) = \lim_{n\to+\infty} E_P^{\mathrm{D}}(f_n).$$

*Proof.* This is a special case of (Taylor, 1985, Theorem 6-3 IV (c)). □

## B.2 Daniell's extension and coherence

In this section, we explore the connection between Daniell's extension and the natural extension. First, we establish that the inner (and outer) Daniell extension is less conservative than the natural extension.

**Proposition 5.13.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then for any gamble $g$*

*in* $\mathbb{G}(\mathcal{X})$, $\underline{\mathcal{E}}_P(g) \leq E^{\mathrm{i}}_P(g)$ *and* $\overline{\mathcal{E}}_P(g) \geq E^{\mathrm{o}}_P(g)$. *Consequently,* $\mathbb{D}^{\mathrm{C}}_P$ *is included in* $\mathbb{D}^{\mathrm{D}}_P$, *and*

$$E^{\mathrm{C}}_P(g) = E^{\mathrm{D}}_P(g) \quad \textit{for all } g \in \mathbb{D}^{\mathrm{C}}_P.$$

*Proof.* We start with the first part of the statement. Observe that $\overline{\mathcal{E}}_P(g) = -\underline{\mathcal{E}}_P(-g)$ by definition, and that $E^{\mathrm{o}}_P(g) = -E^{\mathrm{i}}_P(-g)$ by Lemma B.6$_{455}$ (iv). Therefore, it clearly suffices for us to prove that $\underline{\mathcal{E}}_P(g) \leq E^{\mathrm{i}}_P(g)$.

To this end, we fix some gamble $g$ in $\mathbb{G}(\mathcal{X})$ and some positive real number $\epsilon$. Recall from Section 2.2.3$_{24}$ that $\underline{\mathcal{E}}_P$ is real valued. By Eq. (5.2)$_{217}$, there is an $\mathcal{F}$-simple variable $h_\epsilon$ such that $h_\epsilon \leq g$ and $\underline{\mathcal{E}}_P(g) \leq E_P(h_\epsilon) + \epsilon$. It furthermore follows from Lemma B.6$_{455}$ (ii) & (vi) that $E_P(h_\epsilon) = E^{\mathrm{i}}_P(h_\epsilon) \leq E^{\mathrm{i}}_P(g)$. Combining the two obtained inequalities, we see that $\underline{\mathcal{E}}_P(g) \leq E^{\mathrm{i}}_P(g) + \epsilon$. Because $\epsilon$ was an arbitrary positive real number, we infer from this inequality that $\underline{\mathcal{E}}_P(g) \leq E^{\mathrm{i}}_P(g)$, which is what we needed to prove.

To prove the second part of the statement, we recall that the gamble $g$ belongs to $\mathbb{D}^{\mathrm{C}}_P$ if and only $\underline{\mathcal{E}}_P(g) = \overline{\mathcal{E}}_P(g)$, and that $E^{\mathrm{i}}_P(g) \leq E^{\mathrm{o}}_P(g)$ due to (iii) in Lemma B.6$_{455}$. From this and the first part, it follows that $E^{\mathrm{i}}_P(g) = E^{\mathrm{o}}_P(g)$, so $g$ belongs to $\mathbb{D}^{\mathrm{D}}_P$, as required. □

Next, we establish that the restriction of the Daniell expectation $E^{\mathrm{D}}_P$ to gambles is a coherent expectation.

**Proposition 5.14.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then the restriction of the Daniell extension $E^{\mathrm{D}}_P$ to $\mathbb{G}(\mathcal{X}) \cap \mathbb{D}^{\mathrm{D}}_P$ is a coherent expectation.*

*Proof.* We intend to invoke Proposition 2.15$_{22}$, so we need to show that $\mathcal{G} := \mathbb{G}(\mathcal{X}) \cap \mathbb{D}^{\mathrm{D}}_P$ is a linear subspace of $\mathbb{G}(\mathscr{X})$ and that the restriction of $E^{\mathrm{D}}_P$ to $\mathcal{G}$ satisfies (E1)$_{22}$ and (E3)$_{22}$.

First, let us verify that $\mathcal{G}$ is a linear subspace of $\mathbb{G}(\mathcal{X})$. To this end, we recall from (DE4)$_{225}$ that for all $\lambda$ in $\mathbb{R}$ and $g$ in $\mathcal{G}$, $\lambda g$ again belongs to $\mathcal{G}$. Thus, what remains for is to show that for any two gambles $g$ and $h$ in $\mathcal{G}$, $g + h$ belongs to $\mathcal{G}$ as well. Observe that $g + h$ is well-defined and a gamble. Because $E^{\mathrm{D}}_P(g) + E^{\mathrm{D}}_P(h)$ is clearly well-defined due to (DE3)$_{225}$, it follows from (DE5)$_{225}$ that the gamble $g + h$ belongs to $\mathcal{G}$.

Next, we verify that $E^{\mathrm{D}}_P$ satisfies (E1)$_{22}$ and (E3)$_{22}$. The former follows immediately from (DE3)$_{225}$, and the latter from (DE5)$_{225}$. Thus, it follows from Proposition 2.15$_{22}$ that the restriction of $E^{\mathrm{D}}_P$ to $\mathcal{G} = \mathbb{G}(\mathcal{X}) \cap \mathbb{D}^{\mathrm{D}}_P$ is a coherent expectation. □

# Measure-theoretic probability theory C

In Sections 5.1.2$_{219}$ and 5.1.3$_{224}$, we extended the Dunford integral $E_P$ from $\mathcal{F}$-simple variables to more general variables by means of Daniell's extension, but this is not the most conventional way to go about this. It is much more common to follow the measure-theoretic framework for probability theory, as popularised by Kolmogorov (1933, 1950) – see (Feller, 1968; D. Williams, 1991; Billingsley, 1995; Fristedt et al., 1997; Shiryaev, 2016) for more recent introductions. Central to this framework are two important tools from measure theory: Lebesgue's extension of the Dunford integral $E_P$, and to a lesser extent also Carathéodory's extension of a probability charge to a probability measure.

In this appendix, we only introduce those parts of the ubiquitous measure-theoretic approach that we need in this dissertation, and more specifically in the proof of Theorem 5.19$_{230}$. Like most treatments of measure-theoretic probability, we start with probability charges and measures in Appendix C.1$_{\curvearrowright}$. Next, we take a look at expectations with respect to probability measures in Appendix C.2$_{466}$. Finally, we investigate the similarity of this framework with the Daniell extension in Appendix C.3$_{470}$.

Before we get going, we list some properties of probability charges. These properties are well known (see Shiryaev, 2016, Section 1.2), and are nothing more than specialisations of the properties of full conditional probabilities listed in Proposition 2.49$_{42}$.

**Lemma C.1.** *Consider a probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then for all events $A$ and $B$ in $\mathcal{F}$,*

PM4. $P(\varnothing) = 0$*;*

PM5. $P(A) \leq 1$*;*

PM6. $P(A^{\mathrm{c}}) = 1 - P(A)$;

PM7. $P(A) \leq P(B)$ *whenever* $A \subseteq B$*;*

PM8. $P(B) = P(A \cap B)$ *whenever* $P(A) = 1$.

Furthermore, we also establish that the following specific representation of an $\mathcal{F}$-simple variable always exists. At the same time, this result also establishes the equivalence of our definition of $\mathcal{F}$-simple variables to more strictly formulated definitions, for example that of Bhaskara Rao et al. (1983, Definition 4.12.2).

**Lemma C.2.** *Consider a field of events $\mathcal{F}$ on some possibility space $\mathcal{X}$. Then the real variable $f$ is $\mathcal{F}$-simple if and only if there are pairwise disjoint events $A_1$, …, $A_n$ in $\mathcal{F}$ with $\bigcup_{k=1}^n A_k = \mathcal{X}$ and real numbers $a_1, \dots, a_n$ such that*

$$f = \sum_{k=1}^n a_k \mathbb{I}_{A_k}.$$

*Proof.* The direct implication is immediate by Definition $2.38_{36}$, so we only need to prove the converse one. To this end, we let $f$ be an $\mathcal{F}$-simple variable. By Definition $2.38_{36}$, there are events $A_1, \dots, A_n$ in $\mathcal{F}$ and real numbers $a_1, \dots, a_n$ such that

$$f = \sum_{k=1}^n a_k \mathbb{I}_{A_k}. \tag{C.1}$$

We enumerate the constituents in

$$\left\{ \bigcap_{k=1}^n A'_k \colon A'_k \in \{A_k, A_k^{\mathrm{c}}\} \right\}$$

as $B_1, \dots, B_m$. Observe that $2 \leq m \leq 2^n$ by construction, that these constituents all belong to $\mathcal{F}$ and that they form a partition of $\mathcal{X}$ – meaning that $B_\ell \cap B_i = \emptyset$ for all $\ell$ and $i$ in $\{1, \dots, m\}$ with $\ell \neq i$ and that $\bigcup_{\ell=1}^m B_\ell = \mathcal{X}$. For this reason, $\sum_{\ell=1}^m \mathbb{I}_{B_\ell} = 1$, and therefore

$$f = \sum_{k=1}^n a_k \mathbb{I}_{A_k} \sum_{\ell=1}^m \mathbb{I}_{B_\ell} = \sum_{k=1}^n \sum_{\ell=1}^m a_k \mathbb{I}_{A_k} \mathbb{I}_{B_\ell} = \sum_{k=1}^n \sum_{\ell=1}^m a_k \mathbb{I}_{A_k \cap B_\ell} = \sum_{\ell=1}^m \sum_{k=1}^n a_k \mathbb{I}_{A_k \cap B_\ell}.$$

By construction, $A_k \cap B_\ell$ is equal to $B_\ell$ or $\emptyset$. For all $\ell$ in $\{1, \dots, m\}$, we let

$$\mathcal{K}_\ell := \{k \in \{1, \dots, n\} \colon A_k \cap B_\ell \neq \emptyset\} \quad \text{and} \quad b_\ell := \sum_{k \in \mathcal{K}_\ell} a_k,$$

where we adhere to the convention that the empty sum is equal to 0. This way,

$$f = \sum_{\ell=1}^m b_\ell \mathbb{I}_{B_\ell},$$

which is a representation of $f$ in the required form. □

## C.1 Countable additivity and probability measures

Our definition of the notion of countable additivity in Definition $5.4_{221}$ is somewhat atypical. Most authors prefer to focus on the probability charge $P$ and not on the corresponding expectation $E_P$. For this reason, they define the notion of countable additivity as the countable extension of the (finite) additivity axiom (P3)$_{34}$, thus explaining the name. More formally, they impose one of the following equivalent conditions.

**Lemma C.3.** *Consider a field $\mathcal{F}$ over some possibility space $\mathcal{X}$. A probability charge $P$ on the field $\mathcal{F}$ is* countably additive *if and only if one (and then all) of the following equivalent conditions holds.*

CA1. *for any event $A$ in $\mathcal{F}$ and any sequence $(A_n)_{n\in\mathbb{N}}$ of pairwise disjoint events in $\mathcal{F}$ such that $A = \bigcup_{n\in\mathbb{N}} A_n$, $P(A) = \sum_{n\in\mathbb{N}} P(A_n)$.*

CA2. *For any event $A$ in $\mathcal{F}$ and any non-decreasing sequence $(A_n)_{n\in\mathbb{N}}$ of events in $\mathcal{F}$ such that $A = \bigcup_{n\in\mathbb{N}} A_n$, $P(A) = \lim_{n\to+\infty} P(A_n)$.*

CA3. *For any event $A$ in $\mathcal{F}$ and any non-increasing sequence $(A_n)_{n\in\mathbb{N}}$ of events in $\mathcal{F}$ such that $A = \bigcap_{n\in\mathbb{N}} A_n$, $P(A) = \lim_{n\to+\infty} P(A_n)$.*

CA4. *For any non-increasing sequence $(A_n)_{n\in\mathbb{N}}$ of non-empty events in $\mathcal{F}$ such that $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$, $\lim_{n\to+\infty} P(A_n) = 0$.*

In the proof, we use the following intermediary technical lemma.

**Lemma C.4.** *Consider a field of events $\mathcal{F}$ on some possibility space $\mathcal{X}$. Then for any $\mathcal{F}$-simple variable $f$ and any real number $\alpha$, the level set $\{f > \alpha\}$ – as defined in Eq. (2.4)$_{14}$ – belongs to $\mathcal{F}$.*

*Proof.* Let $f = \sum_{k=1}^{n} a_k \mathbb{I}_{A_k}$ be a representation as in Lemma C.2$_{\curvearrowleft}$. Then clearly, the level set $\{f > \alpha\}$ is the union of those $A_k$'s such that $a_k > \alpha$. Because this is either the empty set or a finite union of events in the field $\mathcal{F}$, we infer from this that $\{f > \alpha\}$ belongs to $\mathcal{F}$ as well. □

*Proof of Lemma C.3.* For a proof that these conditions are equivalent, we refer to (Billingsley, 1995, Example 2.10) and (Shiryaev, 2016, Chapter 2, Section 1.2). Here we only verify that (CA1)–(CA4) are equivalent to the conditions of Definition 5.4$_{221}$. More specifically, it suffices to prove that Definition 5.4$_{221}$ (iii) is equivalent to (CA4).

First, we prove that Definition 5.4$_{221}$ (iii) implies (CA4). Let $(A_n)_{n\in\mathbb{N}}$ be any non-increasing sequence of events in $\mathcal{F}$ such that $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$. Then clearly, $(\mathbb{I}_{A_n})_{n\in\mathbb{N}}$ is a non-increasing sequence of $\mathcal{F}$-simple variables that converges point-wise to 0. Thus, it follows from Definition 5.4$_{221}$ (iii) that

$$\lim_{n\to+\infty} P(A_n) = \lim_{n\to+\infty} E_P(\mathbb{I}_{A_n}) = 0,$$

as required.

Next, we prove that (CA4) implies Definition 5.4$_{221}$ (iii). To that end, we let $(f_n)_{n\in\mathbb{N}}$ be any non-increasing sequence of $\mathcal{F}$-simple variables that converges point-wise to 0. Fix any positive real number $\epsilon$. It follows from Lemma C.4 that for any natural number $n$, the level set $A_n := \{f_n > \epsilon\}$ belongs to $\mathcal{F}$. Let $M := \max f_1$ and, for any natural number $n$, set

$$f'_n := \epsilon\mathbb{I}_{\mathcal{X}} + M\mathbb{I}_{\{f_n > \epsilon\}} = \epsilon\mathbb{I}_{\mathcal{X}} + M\mathbb{I}_{A_n}.$$

Observe that by construction and due to Lemma 2.39[36], $f'_n$ is an $\mathcal{F}$-simple function such that $f_n \leq f'_n$. Hence, it follows from (ES1)[37], (ES4)[37] and Eq. (2.19)[36] that

$$0 \leq E_P(f_n) \leq E_P(f'_n) = \epsilon + MP(A_n), \tag{C.2}$$

where for the first inequality we have used that $f_n \geq 0$ and for the last equality we have used (P1)[34].

Next, we observe that due to the monotonicity of $(f_n)_{n\in\mathbb{N}}$, the sequence $(A_n)_{n\in\mathbb{N}} = (\{f_n > \epsilon\})_{n\in\mathbb{N}}$ in $\mathcal{F}$ is non-increasing. Because $(f_n)_{n\in\mathbb{N}}$ furthermore converges point-wise to 0, it is clear that $\bigcap_{n\in\mathbb{N}} A_n = \bigcap_{n\in\mathbb{N}}\{f_n > \epsilon\} = \emptyset$. If all of the events in $(A_n)_{n\in\mathbb{N}}$ are non-empty, then it follows from (CA1)↶ that $\lim_{n\to+\infty} P(A_n) = 0$. If on the other hand $A_n = \emptyset$ for some $n$ in $\mathbb{N}$, then $P(A_k) = 0$ for all $k \geq n$ because $(A_n)_{n\in\mathbb{N}}$ is non-increasing and $P(\emptyset) = 0$ due to (PM4)[461]; consequently, $\lim_{n\to+\infty} P(A_n) = 0$. Because $\lim_{n\to+\infty} P(A_n) = 0$ in both cases, we conclude that

$$\lim_{n\to+\infty}\big(\epsilon + MP(A_n)\big) = \epsilon + M \lim_{n\to+\infty} P(A_n) = \epsilon.$$

From this and Eq. (C.2), it now follows that

$$0 \leq \liminf_{n\to+\infty} E_P(f_n) \leq \limsup_{n\to+\infty} E_P(f_n) \leq \epsilon,$$

because inequalities are preserved when taking limits. Since $\epsilon$ was an arbitrary positive real number, we infer from these inequalities that $\lim_{n\to+\infty} E_P(f_n) = 0$, as required. □

### C.1.1 σ-fields of events

While the structure of a field is nice, it does not necessarily consist of all the events that we are interested in. Our running example illustrates this nicely.

**Bruno's Example C.5.** In Bruno's Example 5.3[218], we defined the event $H_{\lim}$ that the machine always tosses heads as

$$H_{\lim} = \bigcap_{n\in\mathbb{N}} \{X_n = \mathtt{H}\} = \bigcap_{n\in\mathbb{N}} H_n, \tag{C.3}$$

where $\{X_n = \mathtt{H}\}$ and $H_n$ are as defined in Bruno's Example 2.33[32]. There, we also argued that while $H_{\lim}$ is a countable intersection of events that belong to the field $\mathcal{F}$, it does not belong to $\mathcal{F}$ itself. Observe that because $(H_n)_{n\in\mathbb{N}}$ is a non-increasing sequence of events, we can interpret $H_{\lim}$ as the 'monotone limit' of this sequence. ∲

This example shows that in some cases – especially when dealing with idealisations – it makes sense to consider 'limits' of events; in this particular example, a countable intersection of events. We usually get away with only requiring the following limit property.

**Definition C.6.** A σ*-field of events* $\mathcal{F} \subseteq \mathcal{P}(\mathcal{X})$ is a field of events over $\mathcal{X}$ such that

F4. $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$ for any sequence $(A_n)_{n\in\mathbb{N}}$ of events in $\mathcal{F}$.

Property (F4)$_{\curvearrowleft}$ is just the countable version of (F3)$_{32}$. Similarly to our discussion right after Definition 2.32$_{32}$, it follows from (F2)$_{32}$, (F4)$_{\curvearrowleft}$ and (the countable version of) De Morgan's laws that for any sequence $(A_n)_{n\in\mathbb{N}}$ of events in the σ-field of events $\mathcal{F}$, $\bigcap_{n\in\mathbb{N}} A_n$ also belongs to $\mathcal{F}$. In fact, there are two more equivalent 'limit' conditions (see, for example, Shiryaev, 2016, Section 2.1, Lemma 2).

**Lemma C.7.** *Consider a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then the following four statements are equivalent.*

(i) *$\mathcal{F}$ is a σ-field.*

(ii) *For any sequence $(A_n)_{n\in\mathbb{N}}$ of events in $\mathcal{F}$, $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{F}$.*

(iii) *For any non-decreasing sequence $(A_n)_{n\in\mathbb{N}}$ of events in $\mathcal{F}$, $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.*

(iv) *For any non-increasing sequence $(A_n)_{n\in\mathbb{N}}$ of events in $\mathcal{F}$, $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{F}$.*

Note that the trivial examples of fields of events that we gave right after Definition 2.32$_{32}$ are also σ-fields.

Just like we can enlarge any set of events $\mathcal{F}$ to a field, we can enlarge it to a σ-field of events; again, the set of all events $\mathcal{P}(\mathcal{X})$ is the trivial example. Recall from Lemma 2.34$_{33}$ that any set $\mathcal{F}$ of events generates the smallest field $\langle\mathcal{F}\rangle$ that includes it. A similar result holds for σ-fields; for a proof, see (Billingsley, 1995, Section 2, p. 21).

**Lemma C.8.** *Consider any family of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. If we let $\Sigma(\mathcal{F})$ be the collection of σ-fields over $\mathcal{X}$ that include $\mathcal{F}$, then*

$$\sigma(\mathcal{F}) := \bigcap_{\mathcal{F}'\in\Sigma(\mathcal{F})} \mathcal{F}'$$

*is the smallest σ-field that includes $\mathcal{F}$.*

Because $\sigma(\mathcal{F})$ is the smallest σ-field that includes $\mathcal{F}$, we call $\sigma(\mathcal{F})$ the *σ-field generated by $\mathcal{F}$*. Observe that the definition of $\sigma(\mathcal{F})$ is *not* constructive, in contrast to that of $\langle\mathcal{F}\rangle$. One inconvenient consequence of this is that, even in case every event in the family $\mathcal{F}$ has a nice 'closed form' expression, there usually is no such 'closed form' expression for a generic event $A$ in the generated σ-field $\sigma(\mathcal{F})$.

### C.1.2 Probability measures

Probability measures are, as suggested by the name, the work horses of the measure-theoretic approach to modelling uncertainty.

**Definition C.9.** Consider a σ-field $\mathcal{F}$ over some possibility space $\mathcal{X}$. A *probability measure $P$* on $\mathcal{F}$ is a probability charge on $\mathcal{F}$ that is countably additive.

While providing the definition of a probability measure is easy, actually defining one is – in general – pretty hard. One reason is that, as mentioned previously, there is no 'closed form' expression for a generic event in the σ-field. Fortunately, the following result allows us to construct a probability measure starting from a probability charge; see (Billingsley, 1995, Theorem 3.1) or (D. Williams, 1991, Theorem 1.7) for a proof and some more background.

**Theorem C.10** (Carathéodory's Extension Theorem)**.** *Consider a probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. If $P$ is countably additive, then there is a unique probability measure $P_\sigma$ on $\sigma(\mathcal{F})$ that extends $P$, in the sense that*

$$P_\sigma(A) = P(A) \quad \textit{for all } A \in \mathcal{F}.$$

**Bruno's Example C.11.** Recall from Bruno's Example 5.5$_{222}$ that Billingsley (1995, Theorem 2.3) shows that the probability charge $P$ that we have defined in Eq. (2.17)$_{35}$ is countably additive. Consequently, Theorem C.10 guarantees that there is a unique probability measure $P_\sigma$ on $\sigma(\mathcal{F})$ that extends $P$.

Recall from Bruno's Example C.5$_{464}$ that

$$H_{\lim} = \bigcap_{n\in\mathbb{N}} H_n.$$

Because $(H_n)_{n\in\mathbb{N}}$ is a non-increasing sequence of events in $\mathcal{F}$, it follows from Lemma C.7$_{\curvearrowleft}$ (iv) that $A_{\lim}$ belongs to the σ-field $\sigma(\mathcal{F})$ generated by the field $\mathcal{F}$. Because $P_\sigma$ is a probability measure on the σ-field $\sigma(\mathcal{F})$, we can invoke Lemma C.3$_{463}$ (CA3), to yield

$$P_\sigma(H_{\lim}) = \lim_{n\to+\infty} P_\sigma(H_n) = \lim_{n\to+\infty} P(H_n),$$

where the second equality holds because $P_\sigma$ is equal to $P$ on $\mathcal{F}$. Because $H_n = (X_{1:n} = (\mathtt{H},\dots,\mathtt{H}))$, it follows from the previous equality and Eq. (2.16)$_{35}$ that

$$P_\sigma(H_{\lim}) = \lim_{n\to+\infty} P(H_n) = \lim_{n\to+\infty} \prod_{k=1}^{n} p(\mathtt{H}) = \lim_{n\to+\infty} q^n = \begin{cases} 1 & \text{if } q = 1, \\ 0 & \text{otherwise.} \end{cases}$$

## C.2 Lebesgue extension

Classically, expectation in the measure-theoretic framework is defined through Lebesgue integration instead of Daniell integration, and these two approaches are seemingly different. More concretely, the Lebesgue extension is defined starting from a probability measure $P$ on a σ-field $\mathcal{F}$ of events, subsequently extended to 'measurable' non-negative extended real variables, and then finally to all 'measurable' extended real variables. However, as we will see in Appendix C.3$_{470}$, the Lebesgue and Daniell extensions are essentially identical. For a more detailed treatment of the standard approach, we refer to (D. Williams, 1991, Chapter 5), (Billingsley, 1995, Chapter 15), (Fristedt et al., 1997, Chapter 8) or (Shiryaev, 2016, Section 6)

### C.2.1 Measurable variables

First, we need to properly introduce measurability. We limit ourselves to measurability of extended real variables with respect to a σ-field; our reasons for this are pragmatic: it has a simple definition and it is the only type of measurability that we will need. For measurability of arbitrary – not necessarily extended real valued – variables, see (Halmos, 1974, Section 18), (D. Williams, 1991, Section 3) or (Billingsley, 1995, Section 13).

**Definition C.12.** Consider a σ-field of events $\mathcal{F}$ over a possibility space $\mathcal{X}$. An extended real variable $f$ is $\mathcal{F}$*-measurable* if it satisfies one (and then all) of the following four equivalent conditions.

(i) $\{f > \alpha\} = \{x \in \mathcal{X} \colon f(x) > \alpha\} \in \mathcal{F}$ for all $\alpha \in \mathbb{R}$.

(ii) $\{f \geq \alpha\} := \{x \in \mathcal{X} \colon f(x) \geq \alpha\} \in \mathcal{F}$ for all $\alpha \in \mathbb{R}$.

(iii) $\{f \leq \alpha\} := \{x \in \mathcal{X} \colon f(x) \leq \alpha\} \in \mathcal{F}$ for all $\alpha \in \mathbb{R}$.

(iv) $\{f < \alpha\} := \{x \in \mathcal{X} \colon f(x) < \alpha\} \in \mathcal{F}$ for all $\alpha \in \mathbb{R}$.

We denote the family of all $\mathcal{F}$-measurable (extended real) variables by $\mathbb{M}(\mathcal{F})$, and we use $\mathbb{M}(\mathcal{F})_{\geq 0}$ to denote the non-negative $\mathcal{F}$-measurable (extended real) variables. Important to note is that due to Lemma C.4<sub>463</sub>, any $\mathcal{F}$-simple variable is $\mathcal{F}$-measurable, so $\mathbb{S}(\mathcal{F}) \subseteq \mathbb{M}(\mathcal{F})$.

For an overview of the properties of $\mathcal{F}$-measurable variables, we refer to (Halmos, 1974, Section 20), (D. Williams, 1991, Sections 3.1 to 3.5) or (Shiryaev, 2016, Section 4.3 and 4.4). In our setting, we will make do with the following two quintessential properties.

**Lemma C.13.** *Consider a σ-field of events $\mathcal{F}$ over a possibility space $\mathcal{X}$, and a sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$*-measurable *variables. If $(f_n)_{n\in\mathbb{N}}$ converges point-wise, then this point-wise limit is $\mathcal{F}$-measurable.*

It is customary to split up any extended real variable $f$ in its non-negative part $f^+ := f \vee 0$ and its non-positive part $f^- := |f \wedge 0| = (-f) \vee 0$. Observe that both the non-negative part $f^+$ and the non-positive part $f^-$ are non-negative extended real variables, and that $f = f^+ - f^-$. Even more, the measurability of $f$ and its non-negative and non-positive parts are intertwined.

**Corollary C.14.** *Consider a σ-field of events $\mathcal{F}$ on a possibility space $\mathcal{X}$. Then the extended real variable $f$ is $\mathcal{F}$-measurable if and only if its non-negative part $f^+$ and its non-positive part $f^-$ are $\mathcal{F}$-measurable.*

*Proof.* Follows almost immediately from Definition C.12. □

### C.2.2 Lebesgue integration

From Corollary C.14↶, we know that any $\mathcal{F}$-measurable extended real variable can be decomposed into two $\mathcal{F}$-measurable non-negative extended real variables. It is precisely for this reason that we will focus on $\mathcal{F}$-measurable non-negative extended real variables first.

#### Expectation of non-negative measurable variables

Consider a probability measure $P$ on a σ-field $\mathcal{F}$ over some possibility space $\mathcal{X}$. We obtain the Lebesgue expectation $E_P^{\mathrm{L}}(f)$ of an $\mathcal{F}$-measurable non-negative extended real variable $f$ by approximating it from below with $\mathcal{F}$-simple variables. Formally – see for example (D. Williams, 1991, Section 5.2), (Fristedt et al., 1997, Section 8.1) or (Shiryaev, 2016, Section 6.2) – the extended real-valued operator $E_P^{\mathrm{L}}$ on $\mathbb{M}(\mathcal{F})_{\geq 0}$ is defined by

$$E_P^{\mathrm{L}}(f) := \sup\big\{E_P(g) \colon g \in \mathbb{S}(\mathcal{F}), g \leq f\big\} \quad \text{for all } f \in \mathbb{M}(\mathcal{F})_{\geq 0}. \tag{C.4}$$

Note that this is similar to how we determined $E_P^{\mathrm{C}}$ in Eqs. (5.2)[217] and (5.3)[217], but we now approximate *from below* only.

It is well-known – see (Denneberg, 1994, Chapter 5) or (König, 1997, Theorem 12.11) – that an alternative expression for the Lebesgue extension $E_P^{\mathrm{L}}(f)$ is the *Choquet integral*

$$E_P^{\mathrm{L}}(f) = \int_0^{+\infty} P(\{f > \alpha\})\,\mathrm{d}\alpha = \lim_{B \to +\infty} \int_0^{B} P(\{f > \alpha\})\,\mathrm{d}\alpha, \tag{C.5}$$

where the integrals are (improper) Riemann integrals.

#### Expectation of general measurable variables

Recall from Corollary C.14↶ that the extended real variable $f$ is $\mathcal{F}$-measurable if and only if its non-negative and non-positive parts $f^+$ and $f^-$ are $\mathcal{F}$-measurable non-negative extended real variables, and that $f = f^+ - f^-$. With this in mind, we call any $\mathcal{F}$-measurable variable $f$ *L-integrable* whenever $E_P^{\mathrm{L}}(f^+)$ and $E_P^{\mathrm{L}}(f^-)$ are not both infinite, and then define its *Lebesgue expectation* $E_P^{\mathrm{L}}(f)$ as

$$E_P^{\mathrm{L}}(f) := E_P^{\mathrm{L}}(f^+) - E_P^{\mathrm{L}}(f^-); \tag{C.6}$$

see for example (Fristedt et al., 1997, Section 8.1) or (Shiryaev, 2016, Section 6.2). Note that this is a proper definition because $E_P^{\mathrm{L}}(f^+)$ and $E_P^{\mathrm{L}}(f^-)$ are always non-negative and not both infinite by assumption, so the expression on the right-hand side is well-defined – that is, cannot lead to $+\infty - (+\infty)$. This way, the domain of the Lebesgue extension $E_P^{\mathrm{L}}$ is the class of L-integrable variables

$$\mathbb{D}_P^{\mathrm{L}} := \big\{f \in \mathbb{M}(\mathcal{F}) \colon \min\big\{E_P^{\mathrm{L}}(f^+), E_P^{\mathrm{L}}(f^-)\big\} < +\infty\big\}.$$

If both $E_P^{\mathrm{L}}(f^+)$ and $E_P^{\mathrm{L}}(f^-)$ are finite, then $E_P^{\mathrm{L}}(f)$ is finite as well; in this case, we say that the expectation is *absolutely convergent*. We denote the absolutely convergent part of the domain by

$$\begin{aligned}\tilde{\mathbb{D}}_P^{\mathrm{L}} &:= \{f \in \mathbb{M}(\mathcal{F})\colon \max\{E_P^{\mathrm{L}}(f^+), E_P^{\mathrm{L}}(f^-)\} < +\infty\}\\ &= \{f \in \mathbb{D}_P^{\mathrm{L}}\colon E_P^{\mathrm{L}}(|f|) < +\infty\},\end{aligned}$$

where the second equality follows from (LI4) further on.

The definition of the Lebesgue extension $E_P^{\mathrm{L}}$ in Eq. (C.6)$_{\curvearrowleft}$ is reminiscent of (DE5)$_{225}$, the additivity property of the Daniell extension $E_P^{\mathrm{D}}$. Even stronger, it turns out that the Lebesgue extension $E_P^{\mathrm{L}}$ has (almost) precisely the same properties as the Daniell extension $E_P^{\mathrm{D}}$. For example, L-integrable variables have the following properties that mirror those of Theorem 5.9$_{225}$; even though they can be found in virtually any treatise on measure-theoretic probability theory, including (Fristedt et al., 1997, Chapter 8) and (Shiryaev, 2016, Section 6), we mention them here for the sake of completeness.

**Theorem C.15.** *Consider a countably additive probability charge $P$ on a $\sigma$-field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then*

LI1. $\mathbb{S}(\mathcal{F}) \subseteq \mathbb{D}_P^{\mathrm{L}}$ *and* $E_P^{\mathrm{L}}(f) = E_P(f)$ *for all $f$ in* $\mathbb{S}(\mathcal{F})$.

*Furthermore, for all L-integrable extended real variables $f$ and $g$ in $\mathbb{D}_P^{\mathrm{L}}$ and all real numbers $\mu$ in $\mathbb{R}$,*

LI2. $\inf f \le E_P^{\mathrm{L}}(f) \le \sup f$;

LI3. $\mu f$ *is D-integrable and* $E_P^{\mathrm{L}}(\mu f) = \mu E_P^{\mathrm{L}}(f)$;

LI4. $f+g$ *is L-integrable and* $E_P^{\mathrm{L}}(f+g) = E_P^{\mathrm{L}}(f) + E_P^{\mathrm{L}}(g)$ *whenever* $f+g$ *and* $E_P^{\mathrm{L}}(f) + E_P^{\mathrm{L}}(g)$ *are well-defined;*

LI5. $E_P^{\mathrm{L}}(f) \le E_P^{\mathrm{L}}(g)$ *whenever* $f \le g$;

LI6. $|f|$ *is L-integrable and* $|E_P^{\mathrm{L}}(f)| \le E_P^{\mathrm{L}}(|f|)$.

Note that it follows immediately from the definition of $\mathbb{D}_P^{\mathrm{L}}$ and (LI2) that any $\mathcal{F}$-measurable variable $f$ that is bounded below or bounded above is L-integrable.

**Lemma C.16.** *Consider a countably additive probability charge $P$ on a $\sigma$-field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then any $\mathcal{F}$-measurable variable $f$ that is bounded below or bounded above belongs to $\mathbb{D}_P^{\mathrm{L}}$.*

*Proof.* Recall from (just before) Corollary C.14$_{467}$ that $f^+$ and $f^-$ are non-negative $\mathcal{F}$-measurable variables such that $f = f^+ - f^-$. Because $f$ is bounded below or above by assumption, either $f^+$ or $f^-$ is bounded (or both). Hence, it follows from (LI2) that $E_P^{\mathrm{L}}(f^+) < +\infty$ or $E_P^{\mathrm{L}}(f^-) < +\infty$ (or both), and this implies that $f$ is L-integrable. □

The Lebesgue extension $E_P^{\mathrm{L}}$ also satisfies two familiar limit properties. First, we have the Monotone Convergence Theorem, so the counterpart of

Theorem $5.10_{226}$ – see (Fristedt et al., 1997, Theorem 11) or (Shiryaev, 2016, Section 6) for a proof.

**Theorem C.17.** *Consider a probability measure $P$ on a σ-field $\mathcal{F}$ of events over some possibility space $\mathcal{X}$. Let $(f_n)_{n\in\mathbb{N}}$ be a non-decreasing sequence of $\mathcal{F}$-measurable variables in $\mathbb{D}_P^{\mathrm{L}}$ such that $E_P^{\mathrm{L}}(f_1) > -\infty$. Then the point-wise limit of $(f_n)_{n\in\mathbb{N}}$ is an $\mathcal{F}$-measurable variable in $\mathbb{D}_P^{\mathrm{L}}$, and*

$$E_P^{\mathrm{L}}\Big(\operatorname{p-w}\lim_{n\to+\infty} f_n\Big) = \lim_{n\to+\infty} E_P^{\mathrm{L}}(f_n).$$

*A similar result holds for any non-increasing sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-measurable variables in $\mathbb{D}_P^{\mathrm{L}}$.*

Second, we have Lebesgue's Dominated Convergence Theorem, so the counterpart of Theorem $5.11_{226}$ (see Shiryaev, 2016, Section 6).

**Theorem C.18.** *Consider a probability measure $P$ on a σ-field $\mathcal{F}$ of events over some possibility space $\mathcal{X}$. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of $\mathcal{F}$-measurable variables in $\mathbb{D}_P^{\mathrm{L}}$ and $g$ an $\mathcal{F}$-measurable variable in $\mathbb{D}_P^{\mathrm{L}}$ with $E_P^{\mathrm{L}}(g) < +\infty$ such that $|f_n| \leq g$ for all $n$ in $\mathbb{N}$. If $(f_n)_{n\in\mathbb{N}}$ converges point-wise, then this point-wise limit is an $\mathcal{F}$-measurable variable in $\mathbb{D}_P^{\mathrm{L}}$, and*

$$E_P^{\mathrm{L}}\Big(\operatorname{p-w}\lim_{n\to+\infty} f_n\Big) = \lim_{n\to+\infty} E_P^{\mathrm{L}}(f_n).$$

## C.3 Carathéodory & Lebesgue versus Daniell

We end our treatment of measure-theoretical expectation with a comparison of the two approaches to extending the expectation operator. Recall from Appendix $\mathrm{C.1.2}_{465}$ that it is customary to construct a probability measure $P_\sigma$ from a countably additive probability charge $P$ on a field through Carathéodory's Extension Theorem. This extension $P_\sigma$ then leads to the Lebesgue extension $E_{P_\sigma}^{\mathrm{L}}$. More concretely, the classical approach can be summarised in three steps:

(i) Construct a countably additive probability charge $P$ on a field $\mathcal{F}$.

(ii) Use Carathéodory's Theorem to extend $P$ to the unique probability measure $P_\sigma$ on $\sigma(\mathcal{F})$ that coincides with $P$ on $\mathcal{F}$.

(iii) Define the Lebesgue extension $E_{P_\sigma}^{\mathrm{L}}$ as follows:

 a) start from the expectation $E_{P_\sigma}$ defined on $\mathbb{S}(\sigma(\mathcal{F}))$;

 b) extend $E_{P_\sigma}$ to $E_{P_\sigma}^{\mathrm{L}}$ on $\mathbb{M}_{\geq 0}(\sigma(\mathcal{F}))$;

 c) and finally extend $E_{P_\sigma}^{\mathrm{L}}$ to all L-integrable variables.

However, if we immediately go for the Daniell extension, then we do not have to invoke Caratheodory's Theorem. More precisely, we can use the following less standard approach:

1. Construct a countably additive probability charge $P$ on a field $\mathcal{F}$.
2. Define the Daniell extension $E_P^{\mathrm{D}}$ as follows:
   a) start from the expectation $E_P$ on the $\mathcal{F}$-simple variables;
   b) extend $E_P$ to $E_P^{\mathrm{mc}}$ on $\overline{\mathbb{V}}_{\mathrm{u}}^{\mathrm{o}}(\mathcal{F})$;
   c) and finally extend $E_P^{\mathrm{D}}$ to all D-integrable variables.

These two approaches turn out to be essentially equivalent.

**Theorem C.19.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Let $P_\sigma$ be the unique probability measure on $\sigma(\mathcal{F})$ that coincides with $P$ on $\mathcal{F}$. Then $\mathbb{D}_{P_\sigma}^{\mathrm{L}} \subseteq \mathbb{D}_P^{\mathrm{D}}$ and*

$$E_{P_\sigma}^{\mathrm{L}}(f) = E_P^{\mathrm{D}}(f) \quad \textit{for all } f \in \mathbb{D}_{P_\sigma}^{\mathrm{L}}.$$

Because our proof for this result is rather long, we start with proving the following intermediary results. First and foremost, we will need the well-known fact that any $\mathcal{F}$-measurable non-negative extended real variable is the point-wise limit of a non-decreasing sequence of $\mathcal{F}$-simple (non-negative) variables (see Billingsley, 1995, Theorem 13.4 and 13.5).

**Lemma C.20.** *Consider a σ-field of events $\mathcal{F}$ over a possibility space $\mathcal{X}$ and a non-negative extended real variable $f$. Then $f$ is $\mathcal{F}$-measurable if and only if there is a non-decreasing sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables that converges point-wise to $f$.*

*Proof.* For a proof of the direct implication, we refer to (Billingsley, 1995, Theorem 13.5). To prove the converse implication, we observe that $f_n$ is $\mathcal{F}$-measurable due to Lemma C.4463. The statement now follows from this and the well-known result – see for example (Billingsley, 1995, Theorem 13.4 (i)) – that the point-wise limit $f$ of a non-decreasing sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-measurable (real) variables is $\mathcal{F}$-measurable as well. □

Next, we make the crucial observation that the restriction of the Daniell extension $E_P^{\mathrm{D}}$ to indicators gives a probability measure; our statement and proof are inspired by those of Taylor (1985, Theorems 6-5 II and III)

**Proposition C.21.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then the set of events*

$$\mathcal{A}_P := \left\{A \in \mathcal{P}(\mathcal{X}) : \mathbb{I}_A \in \mathbb{D}_P^{\mathrm{D}}\right\}$$

*includes $\mathcal{F}$ and is a σ-field. Furthermore,*

$$\pi_P : \mathcal{A}_P \to \mathbb{R} : A \mapsto \pi_P(A) := E_P^{\mathrm{D}}(\mathbb{I}_A)$$

*is a probability measure on $\mathcal{A}_P$ that coincides with $P$ on $\mathcal{F}$.*

*Proof.* Recall that for any event $A$ in $\mathcal{F}$, $\mathbb{I}_A$ is an $\mathcal{F}$-simple variable. By Lemma B.6$_{455}$ (ii), $\mathbb{I}_A$ is D-integrable and

$$\pi_P(A) = E_P^{\mathrm{D}}(\mathbb{I}_A) = E_P(A) = P(A).$$

Consequently, $\mathcal{A}_P$ clearly includes $\mathcal{F}$ and $\pi_P$ coincides with $P$ on $\mathcal{F}$.

To complete the proof, we verify that $\mathcal{A}_P$ is a σ-field – that is, that $\mathcal{A}_P$ satisfies (F1)$_{32}$–(F4)$_{464}$ – at the same time verifying that $\pi_P$ is a probability measure – that is, that $\pi_P$ satisfies (P1)$_{34}$–(P3)$_{34}$ and (CA4)$_{463}$.

First, we observe that $\mathbb{I}_{\mathcal{X}}$ is an $\mathcal{F}$-simple variable, so it follows from (DE1)$_{225}$ that $\mathcal{X}$ belongs to $\mathcal{A}_P$ – that is, $\mathbb{I}_{\mathcal{X}}$ is D-integrable – and

$$\pi_P(\mathcal{X}) = E_P^{\mathrm{D}}(\mathbb{I}_{\mathcal{X}}) = E_P(\mathbb{I}_{\mathcal{X}}) = P(\mathcal{X}) = 1,$$

where the final equality follows from (P1)$_{34}$. This settles (F1)$_{32}$ and (P1)$_{34}$.

Next, we take any $A$ in $\mathcal{A}_P$. Observe that $\mathbb{I}_{A^{\mathrm{c}}} = 1 - \mathbb{I}_A$. Recall that $1 = \mathbb{I}_{\mathcal{X}}$ is D-integrable. Because $\mathbb{I}_A$ is D-integrable, it follows from (DE4)$_{225}$ that $-\mathbb{I}_A$ is D-integrable and from (DE3)$_{225}$ that $-1 \leq E_P^{\mathrm{D}}(-\mathbb{I}_A) \leq 0$. Therefore, it follows from (DE5)$_{225}$ that $\mathbb{I}_{A^{\mathrm{c}}} = 1 - \mathbb{I}_A$ is D-integrable, so $A^{\mathrm{c}}$ belongs to $\mathcal{A}_P$. This settles (F2)$_{32}$.

Property (P2)$_{34}$ follows immediately from (DE3)$_{225}$, because

$$\pi_P(A) = E_P^{\mathrm{D}}(\mathbb{I}_A) \geq 0 \quad \text{for all } A \in \mathcal{A}_P.$$

To verify (F3)$_{32}$ and (P3)$_{34}$, we fix any two events $A$ and $B$ in $\mathcal{A}_P$. Observe that if $A \cap B = \emptyset$, then it follows immediately from (DE3)$_{225}$ and (DE5)$_{225}$ that $A \cup B$ belongs to $\mathcal{A}_P$ because $\mathbb{I}_{A\cup B} = \mathbb{I}_A + \mathbb{I}_B$ is D-integrable, and

$$\pi_P(A\cup B) = E_P^{\mathrm{D}}(\mathbb{I}_{A\cup B}) = E_P^{\mathrm{D}}(\mathbb{I}_A + \mathbb{I}_B) = E_P^{\mathrm{D}}(\mathbb{I}_A) + E_P^{\mathrm{D}}(\mathbb{I}_B) = \pi_P(A) + \pi_P(B).$$

This settles (P3)$_{34}$, and (F3)$_{32}$ in case $A \cap B = \emptyset$. In case $A \cap B \neq \emptyset$, (F3)$_{32}$ follows immediately from (DE3)$_{225}$ and (DE7)$_{226}$ because $\mathbb{I}_{A\cup B} = \mathbb{I}_A \vee \mathbb{I}_B$.

Thus far, we have verified that $\mathcal{A}_P$ is a field of events and that $\pi_P$ is a probability charge on $\mathcal{A}_P$. To complete our proof, we now show that $\mathcal{A}_P$ is a σ-field and $P$ a probability measure.

To verify (F4)$_{464}$, we let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events in $\mathcal{A}_P$. For any $n$ in $\mathbb{N}$, we let $B_n := \bigcup_{k=1}^n A_k$. Observe that due to the previous, $B_n$ belongs to $\mathcal{A}_P$ and $\mathbb{I}_{B_n}$ is D-integrable for all $n$ in $\mathbb{N}$. Because furthermore $(\mathbb{I}_{B_n})_{n\in\mathbb{N}}$ is non-decreasing due to (DE6)$_{226}$ and $0 \leq E_P^{\mathrm{D}}(\mathbb{I}_{B_1})$ due to (DE3)$_{225}$, it follows from Theorem 5.10$_{226}$ that $\text{p-w}\lim_{n\to+\infty}\mathbb{I}_{B_n}$ is D-integrable. Because $\text{p-w}\lim_{n\to+\infty}\mathbb{I}_{B_n} = \mathbb{I}_{\bigcup_{n\in\mathbb{N}} A_n}$, we have shown that $\bigcup_{n\in\mathbb{N}} A_n$ belongs to $\mathcal{A}_P$, as required.

Finally, we verify (CA4)$_{463}$. To this end, we let $(A_n)_{n\in\mathbb{N}}$ be a non-increasing sequence of non-empty events in $\mathcal{A}_P$ such that $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$. Then it is clear that $(\mathbb{I}_{A_n})_{n\in\mathbb{N}}$ is a non-increasing sequence of variables in $\mathbb{D}_P^{\mathrm{D}}$ with $\text{p-w}\lim_{n\to+\infty}\mathbb{I}_{A_n} = \mathbb{I}_\emptyset$. Observe that $E_P^{\mathrm{D}}(\mathbb{I}_{A_1}) \leq 1 < +\infty$ due to (DE3)$_{225}$, that $\mathbb{I}_\emptyset$ belongs to $\mathbb{D}_P^{\mathrm{D}}$ due to (F1)$_{32}$ and (F2)$_{32}$, and that $E_P^{\mathrm{D}}(\mathbb{I}_\emptyset) = 0$ due to (DE3)$_{225}$. Therefore, it follows from Theorem 5.10$_{226}$ that

$$\lim_{n\to+\infty}\pi_P(A_n) = \lim_{n\to+\infty}E_P^{\mathrm{D}}\big(\mathbb{I}_{A_n}\big) = E_P^{\mathrm{D}}(\mathbb{I}_\emptyset) = 0$$

as required. □

Using the foregoing result, we more or less immediately obtain the following two corollaries.

**Corollary C.22.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Let $P_\sigma$ be the unique probability measure on $\sigma(\mathcal{F})$ that extends $P$. Then any event $A$ in $\sigma(\mathcal{F})$ is D-measurable, and*

$$P_\sigma(A) = E_P^{\mathrm{D}}(\mathbb{I}_A).$$

*Proof.* Let $\pi_P$ be the probability measure as defined in Proposition C.21$_{471}$. Because its domain $\mathcal{A}_P$ is a σ-field that includes $\mathcal{F}$, it also includes the smallest σ-field $\sigma(\mathcal{F})$ that is generated by $\mathcal{F}$. It now follows from Proposition C.21$_{471}$ that the restriction of $\pi_P$ to $\sigma(\mathcal{F})$ is a probability measure that coincides with $P$ on $\mathcal{F}$. Because $P_\sigma$ is the unique probability measure on $\sigma(\mathcal{F})$ that coincides with $P$ on $\mathcal{F}$ by Theorem C.10$_{466}$, $P_\sigma$ has to coincide with $\pi_P$ on $\sigma(\mathcal{F})$. The statement now follows immediately. □

**Corollary C.23.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then any $\sigma(\mathcal{F})$-simple variable $f$ is D-integrable, and*

$$E_{P_\sigma}(f) = E_P^{\mathrm{D}}(f),$$

*where $P_\sigma$ is the unique probability measure on $\sigma(\mathcal{F})$ that extends $P$.*

*Proof.* Follows almost immediately from Eq. (2.19)$_{36}$, Corollary C.22, (DE3)$_{225}$, (DE4)$_{225}$ and (DE5)$_{225}$. □

With the help of the foregoing corollary, we can establish that any non-negative $\sigma(\mathcal{F})$-measurable variable is both L-integrable and D-integrable.

**Lemma C.24.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then any non-negative $\sigma(\mathcal{F})$-measurable variable $f$ belongs to $\mathbb{D}_{P_\sigma}^{\mathrm{L}}$ and $\mathbb{D}_P^{\mathrm{D}}$, and*

$$E_{P_\sigma}^{\mathrm{L}}(f) = E_P^{\mathrm{D}}(f).$$

*Proof.* By Lemma C.20$_{471}$, there is a non-decreasing sequence $(f_n)_{n\in\mathbb{N}}$ of $\sigma(\mathcal{F})$-simple variables that converges point-wise to $f$. Observe that $\min f_1 > -\infty$ because $f_1$ is $\sigma(\mathcal{F})$-simple.

First, we establish that $f$ belongs to $\mathbb{D}_P^{\mathrm{D}}$. To this end, we recall from Corollary C.23 that, for all $n$ in $\mathbb{N}$, $f_n$ is D-integrable and $E_{P_\sigma}(f_n) = E_P^{\mathrm{D}}(f_n)$. Furthermore, we observe that it follows from (DE3)$_{225}$ that $E_P^{\mathrm{D}}(f_1) > -\infty$ because $\min f_1 > -\infty$. Therefore, it follows from Theorem 5.10$_{226}$ that $f$ belongs to $\mathbb{D}_P^{\mathrm{D}}$ and

$$E_P^{\mathrm{D}}(f) = \lim_{n\to+\infty} E_P^{\mathrm{D}}(f_n) = \lim_{n\to+\infty} E_{P_\sigma}(f_n).$$

Second, we establish that $f$ belongs to $\mathbb{D}_{P_\sigma}^{\mathrm{L}}$. To this end, we fix some $n$ in $\mathbb{N}$. By Lemma C.4$_{463}$, $f_n$ is $\sigma(\mathcal{F})$-measurable because $f_n$ is $\sigma(\mathcal{F})$-simple. It furthermore follows from (LI1)$_{469}$ that $f_n$ is L-integrable with $E^{\mathrm{L}}_{P_\sigma}(f_n) = E_{P_\sigma}(f_n)$ because $f_n$ is $\sigma(\mathcal{F})$-simple, and from (LI2)$_{469}$ that $E^{\mathrm{L}}_{P_\sigma}(f_1) > -\infty$ because $\min f_1 > -\infty$. Therefore, it follows from Theorem C.17$_{470}$ that $f$ belongs to $\mathbb{D}^{\mathrm{L}}_{P_\sigma}$ and

$$E^{\mathrm{L}}_{P_\sigma}(f) = \lim_{n\to+\infty} E^{\mathrm{L}}_{P_\sigma}(f_n) = \lim_{n\to+\infty} E_{P_\sigma}(f_n).$$

This settles that $f$ belongs to $\mathbb{D}^{\mathrm{L}}_{P_\sigma}$ and $\mathbb{D}^{\mathrm{D}}_{P}$. Because $E^{\mathrm{L}}_{P_\sigma}(f)$ and $E^{\mathrm{D}}_{P}(f)$ are both equal to $\lim_{n\to+\infty} E_{P_\sigma}(f_n)$ due to the previous, we conclude that $E^{\mathrm{L}}_{P_\sigma}(f) = E^{\mathrm{D}}_{P}(f)$. □

The result that we are after follows more or less immediately from Eq. (C.6)$_{468}$ and Lemma C.24$_{\curvearrowleft}$.

*Proof of Theorem C.19$_{471}$.* Take any extended real variable $f$ in $\mathbb{D}^{\mathrm{L}}_{P_\sigma}$. Then $f$ is $\sigma(\mathcal{F})$-measurable by assumption, and both $f^+$ and $f^-$ are non-negative and $\sigma(\mathcal{F})$-measurable due to Corollary C.14$_{467}$. Thus, it follows from Lemma C.24$_{\curvearrowleft}$ that

$$E^{\mathrm{L}}_{P_\sigma}(f^+) = E^{\mathrm{D}}_{P}(f^+) \quad \text{and} \quad E^{\mathrm{L}}_{P_\sigma}(f^-) = E^{\mathrm{D}}_{P}(f^-).$$

Furthermore, because at least one of $E^{\mathrm{L}}_{P_\sigma}(f^+)$ and $E^{\mathrm{L}}_{P_\sigma}(f^-)$ is real valued, it follows from Eq. (C.6)$_{468}$, (DE4)$_{225}$ and (DE5)$_{225}$ that

$$E^{\mathrm{L}}_{P_\sigma}(f) = E^{\mathrm{L}}_{P_\sigma}(f^+) - E^{\mathrm{L}}_{P_\sigma}(f^-) = E^{\mathrm{D}}_{P}(f^+) - E^{\mathrm{D}}_{P}(f^-) = E^{\mathrm{D}}_{P}(f^+) + E^{\mathrm{D}}_{P}(-f^-) = E^{\mathrm{D}}_{P}(f).$$

This proves that $\mathbb{D}^{\mathrm{L}}_{P_\sigma} \subseteq \mathbb{D}^{\mathrm{D}}_{P}$, as well as the equality of the statement. □

It follows immediately from Theorem C.19$_{471}$ and Lemma C.16$_{469}$ that any $\sigma(\mathcal{F})$-measurable variable that is bounded below or bounded above belongs to $\mathbb{D}^{\mathrm{D}}_{P}$.

**Corollary C.25.** *Consider a countably additive probability charge $P$ on a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then any $\sigma(\mathcal{F})$-measurable variable $f$ that is bounded below or bounded above belongs to $\mathbb{D}^{\mathrm{D}}_{P}$.*

*Proof.* Follows immediately from Theorem C.19$_{471}$ and Lemma C.16$_{469}$. □

A more important consequence of Theorem C.19$_{471}$ is the following utterly important theorem; in our proof, we also rely on Lemma C.13$_{467}$ and Corollary C.25.

**Theorem 5.12.** *Consider a field of events $\mathcal{F}$ over some possibility space $\mathcal{X}$. Then for any countably additive probability charge $P$ on $\mathcal{F}$,*

$$\mathbb{S}(\mathcal{F}) \subseteq \overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathcal{F}) \subseteq \overline{\mathbb{V}}_{\lim}(\mathcal{F}) \subseteq \mathbb{D}^{\mathrm{D}}_{P}.$$

*Proof.* Recall from Lemma B.5<sub>454</sub> (i) that $\mathbb{S}(\mathcal{F}) \subseteq \overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathcal{F})$. Because every variable $f$ in $\overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathcal{F})$ is the monotone limit of a sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables, it also belongs to $\overline{\mathbb{V}}_{\mathrm{lim}}(\mathcal{F})$; hence, $\overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathcal{F}) \subseteq \overline{\mathbb{V}}_{\mathrm{lim}}(\mathcal{F})$. Thus, it remains for us to show the third inclusion in the statement.

Fix some $f$ in $\overline{\mathbb{V}}_{\mathrm{lim}}(\mathcal{F})$. Then by definition, there is a sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{F}$-simple variables that converges point-wise to $f$, and that is either uniformly bounded below or uniformly bounded above. For all $n$ in $\mathbb{N}$, $f_n$ is trivially $\sigma(\mathcal{F})$-simple because it is $\mathcal{F}$-simple, and therefore also trivially $\sigma(\mathcal{F})$-measurable by (LI1)$_{469}$. Hence, it follows from Lemma C.13$_{467}$ that $f$ is $\sigma(\mathcal{F})$-measurable. Note that $f$ is either bounded above or bounded below because $(f_n)_{n\in\mathbb{N}}$ is either uniformly bounded above or uniformly bounded below, so it follows from Corollary C.25$_{\curvearrowleft}$ that $f$ belongs to $\mathbb{D}^{\mathrm{D}}_{P}$. Since $f$ was an arbitrary variable in $\overline{\mathbb{V}}_{\mathrm{lim}}(\mathcal{F})$, this proves that $\overline{\mathbb{V}}_{\mathrm{lim}}(\mathcal{F}) \subseteq \mathbb{D}^{\mathrm{D}}_{P}$, as required. □

# *List of symbols*

## Sets and sequences

### Number sets

| | |
|---|---|
| $\mathbb{N}$ | the set of natural numbers · Section $1.5_8$ |
| $\mathbb{Z}_{\geq 0}$ | the set of non-negative integers · Section $1.5_8$ |
| $\mathbb{Z}$ | the set of integers · Section $1.5_8$ |
| $\mathbb{R}, \mathbb{R}_{\geq 0}, \mathbb{R}_{>0}$ | the set of (non-negative/positive) reals · Section $1.5_8$ |
| $\overline{\mathbb{R}}$ | the set of extended reals · Section $1.5_8$ |

### Sequences of time points

| | |
|---|---|
| $s, t, r$ | time points in $\mathbb{R}_{\geq 0}$ |
| $u, v, w$ | ordered sequences of time points · p. 60 |
| $()$ | the empty sequence of time points · p. 60 |
| $\mathscr{U}, \mathscr{U}_{\neq ()}$ | the set of (non-empty) sequences of time points · Eqs. $(3.5)_{60}$ and $(3.6)_{60}$ |
| $\min u$ | first time point in the sequence of time points $u$ · p. 61 |
| $\max u$ | last time point in the sequence of time points $u$ · p. 61 |
| $u \preccurlyeq v$ | $v$ only contains time points that are also in $u$ or that succeed the last time point of $u$ · Eq. $(3.7)_{61}$ |
| $\mathscr{U}_{\succcurlyeq u}$ | the set of sequences of time points $v$ such that $v \succcurlyeq u$ · p. 61 |
| $u \prec v$ | the last time point of $u$ preceeds the first time point of $v$ · Eq. $(3.8)_{61}$ |
| $\mathscr{U}_{\succ u}$ | the set of non-empty sequences of time points $v$ such that $v \succ u$ · p. 61 |
| $u \cup v$ | ordered union of the time points in $u$ and $v$ · p. 61 |
| $u \cap v$ | ordered intersection of the time points in $u$ and $v$ · p. 61 |
| $u \setminus v$ | sequence of time points in $u$ that are not in $v$ · p. 61 |

**Grids of time points**

| | |
|---|---|
| $\mathscr{U}_{[s,r]}$ | the set of grids over $[s,r]$ – that is, sequences of time points $v=(t_0,\dots,t_n)$ with $t_0=s$ and $t_n=r$ · Eq. (5.19)$_{233}$ |
| $\Delta(v)$ | width of the largest subinterval of the grid $v$ · Eq. (5.20)$_{233}$ |

## Modelling uncertainty

| | |
|---|---|
| $X$ | uncertain outcome of an experiment · Section 2.1.1$_{12}$ |
| $\mathcal{X}$ | possibility space · Section 2.1.1$_{12}$ |
| $x,y,z$ | outcomes in $\mathcal{X}$ · Section 2.1.1$_{12}$ |
| $\mathcal{P}(\mathcal{X})$ | the set of events – that is, the set of all subsets of $\mathcal{X}$ · Section 2.1.2$_{13}$ |
| $\mathcal{P}(\mathcal{X})_{\supset\emptyset}$ | the set of non-empty events · Section 2.1.2$_{13}$ |

**Variables**

| | |
|---|---|
| $\overline{\mathbb{V}}(\mathcal{X})$ | the set of extended real variables on $\mathcal{X}$ · Section 2.1.3$_{13}$ |
| $\mathbb{V}(\mathcal{X})$ | the set of real variables on $\mathcal{X}$ · Section 2.1.3$_{13}$ |
| $\mathbb{G}(\mathcal{X})$ | the set of gambles – that is, bounded real variables – on $\mathcal{X}$ · Section 2.1.3$_{13}$ |
| $f\wedge g$ | point-wise minimum of $f$ and $g$ · Eq. (2.1)$_{14}$ |
| $f\vee g$ | point-wise maximum of $f$ and $g$ · Eq. (2.2)$_{14}$ |
| $f=g$ | $f(x)=g(x)$ for all $x$ in $\mathcal{X}$ · Section 2.1.3$_{13}$ |
| $f\leq g$ | $f(x)\leq g(x)$ for all $x$ in $\mathcal{X}$ · Section 2.1.3$_{13}$ |
| $f\lessdot g$ | $f(x)<g(x)$ for all $x$ in $\mathcal{X}$ · Section 2.1.3$_{13}$ |
| $f<g$ | $f\leq g$ and $f\neq g$ · Section 2.1.3$_{13}$ |
| $\mathbb{I}_A$ | indicator of the event $A$ in $\mathcal{P}(\mathcal{X})$ · Eq. (2.3)$_{14}$ |
| $\{f\gtrdot\alpha\}$ | level set · Eq. (2.4)$_{14}$ |
| $\text{p-w}\lim_{n\to+\infty} f_n$ | point-wise limit of the sequence $(f_n)_{n\in\mathbb{N}}$ in $\overline{\mathbb{V}}(\mathcal{X})$ · Eq. (5.4)$_{221}$ |
| $(f_n)_{n\in\mathbb{N}}\nearrow f$ | $(f_n)_{n\in\mathbb{N}}$ is non-decreasing and converges point-wise to $f$ · Section 5.1.2$_{221}$ |
| $(f_n)_{n\in\mathbb{N}}\searrow f$ | $(f_n)_{n\in\mathbb{N}}$ is non-increasing and converges point-wise to $f$ · Section 5.1.2$_{221}$ |

**Probability charges**

| | |
|---|---|
| $\mathcal{F}$ | field of events · Definition 2.32$_{32}$ |
| $P$ | probability charge on a field $\mathcal{F}$ · Definition 2.36$_{34}$ |

| | |
|---|---|
| $\mathbb{S}(\mathcal{F})$ | the set of $\mathcal{F}$-simple variables · Definition 2.38 36 |
| $E_P$ | expectation on $\mathbb{S}(\mathcal{F})$ corresponding to the probability charge $P$ on $\mathcal{F}$ · Eq. (2.19) 36 |

**Coherent conditional probabilities**

| | |
|---|---|
| $P$ | coherent conditional probability on $\mathcal{D} \subseteq \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})_{\supset\emptyset}$ · Definition 2.51 44 |
| $\mathbb{CS}(\mathcal{D})$ | set of couples of simple variables and conditioning events corresponding to the structure of fields $\mathcal{D}$ · Eq. (2.27) 47 |
| $E_P$ | (conditional) expectation on $\mathbb{CS}(\mathcal{D})$ corresponding to the coherent conditional probability $P$ on the structure of fields $\mathcal{D}$ · Eq. (2.26) 47 |

**Daniell integration**

| | |
|---|---|
| $\overline{\mathbb{V}}^{\mathrm{o}}(\mathcal{F}), \overline{\mathbb{V}}_{\mathrm{u}}(\mathcal{F})$ | the set of $\mathcal{F}$-over/under variables · p. 222 |
| $\overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathcal{F})$ | the set of $\mathcal{F}$-over and under variables · Eq. (5.5) 222 |
| $E_P^{\mathrm{mc}}$ | extension of $E_P$ from $\mathbb{S}(\mathcal{F})$ to $\overline{\mathbb{V}}^{\mathrm{o}}_{\mathrm{u}}(\mathcal{F})$ · Eq. (5.6) 223 |
| $E_P^{\mathrm{i}}$ | inner Daniell extension of $E_P$ from $\mathbb{S}(\mathcal{F})$ to $\overline{\mathbb{V}}(\mathcal{X})$ · Eq. (5.8) 224 |
| $E_P^{\mathrm{o}}$ | outer Daniell extension of $E_P$ from $\mathbb{S}(\mathcal{F})$ to $\overline{\mathbb{V}}(\mathcal{X})$ · Eq. (5.9) 224 |
| $\mathbb{D}_P^{\mathrm{D}}$ | the set of D-integrable variables for $P$ · Eq. (5.12) 225 |
| $E_P^{\mathrm{D}}$ | Daniell extension of $E_P$ from $\mathbb{S}(\mathcal{F})$ to $\mathbb{D}_P^{\mathrm{D}}$ · p. 225 |
| $\overline{\mathbb{V}}_{\mathrm{lim}}(\mathcal{F})$ | the set of extended real variables that are the point-wise limit of a uniformly bounded below or above sequence of $\mathcal{F}$-simple variables · Eq. (5.14) 227 |

## Jump processes in general

| | |
|---|---|
| $\mathscr{X}$ | state space · Section 3.1.1 55 |
| $x, y, z$ | states in $\mathscr{X}$ · Section 3.1.1 55 |
| $x_u$ | tuple of states indexed by the time points in $u$ · p. 61 |
| $\mathscr{X}_u$ | the set of tuples of states indexed by the time points in $u$ · p. 61 |
| $x_{()}$ | empty tuple of states · p. 61 |
| $\omega$ | path – that is, a map from $\mathbb{R}_{\geq 0}$ to $\mathscr{X}$ · Section 3.1.1 55 |
| $\omega\|_u$ | restriction of the path $\omega$ to the time points in $u$ · Eq. (3.9) 61 |
| $\tilde{\Omega}_{\mathscr{X}}, \tilde{\Omega}$ | the set of paths with state space $\mathscr{X}$ · Section 3.1.1 55 |
| $\Omega_{\mathscr{X}}, \Omega$ | the set of càdlàg paths with state space $\mathscr{X}$ · Definition 3.4 57 |

**Finitary events**

| | |
|---|---|
| $X_t$ | projector variable · Eq. (3.3) 59 |

| | |
|---|---|
| $X_u$ | tuple of projector variables indexed by the time points in $u$ · Eq. (3.10)62 |
| $\{X_v \in B\}$ | cylinder/finitary event · Eq. (3.12)62 |
| $\mathscr{H}$ | the set of conditioning events of the form $\{X_u = x_u\}$ · Eq. (3.14)62 |
| $\mathscr{F}_u$ | the field of finitary events of the form $\{X_v \in B\}$ with $v \succcurlyeq u$ · Eq. (3.16)63 |

**Idealised events and variables**

| | |
|---|---|
| $\mathscr{J}_{[s,r]}$ | jump times over $[s,r]$ · Eq. (5.16)231 |
| $\eta_{[s,r]}$ | number of jumps over $[s,r]$ · Eq. (5.18)232 |
| $\eta_v$ | number of jumps over the grid $v$ · Eq. (5.21)233 |
| $H^{S,G}_{\mathscr{R}}$ | until event – that is, the event of hitting the set $G$ of goal states while passing through the safe states in $S$ over the time points in $\mathscr{R}$ · Eq. (6.3)281 |
| $h^{S,G}_{\mathscr{R}}$ | indicator of the until event $H^{S,G}_{\mathscr{R}}$ · Section 6.2281 |
| $H^{G}_{\mathscr{R}}$ | event of hitting the set $G$ of goal states over the time points in $\mathscr{R}$ · Section 6.2.3287 |
| $\tau^G_{[s,+\infty[}$ | hitting time of the set $G$ of goal states after $s$ · Eq. (6.12)289 |
| $\tau^G_{[s,r]}$ | truncated hitting time corresponding to $\tau^G_{[s,+\infty[}$ · Eq. (6.13)290 |
| $\tau^G_v$ | approximating hitting time corresponding to $\tau^G_{[\min v,\max v]}$ · Eq. (6.14)290 |
| $\int_s^r f_t(X_t)\,\mathrm{d}t$ | Riemann integral of $f_t(X_t)$ over $[s,r]$ · Eq. (6.26)299 |
| $\langle f_\bullet\rangle_v$ | approximating Riemann sum for the piece-wise continuous family $(f_t)_{t\in[s,r]}$ corresponding to the grid $v$ over $[s,r]$ · Eq. (6.27)300 |
| $[\![f]\!]_{[s,r]}$ | temporal average of $f(X_t)$ over $[s,r]$ · Section 6.4.4305 |
| $[\![f]\!]_v$ | approximating Riemann sum over the grid $v$ for the temporal average of $f(X_t)$ over $[\min v,\max v]$ · Section 6.4.4305 |

**Norms and operators on $\mathbb{G}(\mathscr{X})$**

| | |
|---|---|
| $\langle\bullet,\bullet\rangle$ | inner product on $\mathbb{G}(\mathscr{X})$ · p. 75 |
| $\lVert\bullet\rVert_\infty, \lVert\bullet\rVert$ | supremum norm on $\mathbb{G}(\mathscr{X})$ · Eq. (3.26)75 |
| $\lVert\bullet\rVert_{\mathrm{v}}$ | variation semi-norm on $\mathbb{G}(\mathscr{X})$ · Eq. (4.5)179 |
| $\lVert\bullet\rVert_{\mathrm{c}}$ | centred semi-norm on $\mathbb{G}(\mathscr{X})$ · Eq. (4.6)179 |
| $\lVert\bullet\rVert_{\mathrm{op}}$ | operator norm induced by the supremum norm $\lVert\bullet\rVert$ · Eq. (3.27)77 |
| $M(x,y)$ | $(x,y)$-component of the linear operator $M$ · p. 76 |
| $I$ | identity operator · p. 76 |

| | |
|---|---|
| $(e^{tG})_{t\in\mathbb{R}_{\geq 0}}$ | semi-group generated by the linear operator $G$ · Eq. (3.29)$_{79}$ |

## Precise jump processes

| | |
|---|---|
| $\mathscr{D}_{\mathscr{X}}, \mathscr{D}$ | domain of a jump process with state space $\mathscr{X}$ · Eq. (3.17)$_{65}$ |
| $P$ | jump process · Definition 3.12$_{65}$ |
| $\mathbb{JS}$ | domain of the (conditional) expectation $E_P$ corresponding to the jump process $P$ · Eq. (3.24)$_{68}$ |
| $\mathbb{JD}$ | domain of the extended (conditional) expectation $E_P^{\mathrm{D}}$ corresponding to the countably additive jump process $P$ · Eq. (5.15)$_{229}$ |
| $P_{p_0,Q}$ | homogeneous Markovian jump process characterised by $p_0$ and $Q$ · Section 3.2.4$_{87}$ |
| $E_{p_0,Q}$ | (conditional) expectation corresponding to $P_{p_0,Q}$ · Section 3.2.4$_{87}$ |

## Operators

| | |
|---|---|
| $T$ | transition operator · Definition 3.25$_{80}$ |
| $Q$ | rate operator · Definition 3.27$_{81}$ |
| $\mathfrak{Q}_{\mathscr{X}}, \mathfrak{Q}$ | the set of rate operators on $\mathbb{G}(\mathscr{X})$ · Definition 3.27$_{81}$ |
| $T_{t,r}, T_{t,r}^{\{X_u=x_u\}}$ | (history-dependent) transition operator corresponding to the jump process $P$ · Eqs. (3.33)$_{84}$ and (3.35)$_{84}$ |
| $Q_{t,r}^{\{X_u=x_u\}}$ | history-dependent 'rate' operator corresponding to the jump process $P$ · Eq. (3.52)$_{94}$ |
| $\partial_+ T_{t,t}^{\{X_u=x_u\}}$ | the set of right-sided accumulation points of $Q_{t,\bullet}^{\{X_u=x_u\}}$ · Eq. (3.55)$_{95}$ |
| $\partial_- T_{t,t}^{\{X_u=x_u\}}$ | the set of left-sided accumulation points of $Q_{\bullet,t}^{\{X_u=x_u\}}$ · Eq. (3.56)$_{95}$ |

## Imprecise jump processes

| | |
|---|---|
| $\mathscr{P}$ | imprecise jump process · Definition 3.39$_{88}$ |
| $\underline{P}_{\mathscr{P}}, \overline{P}_{\mathscr{P}}$ | lower/upper (conditional) probability corresponding to $\mathscr{P}$ · Eqs. (3.42)$_{88}$ and (3.43)$_{88}$ |
| $\underline{E}_{\mathscr{P}}, \overline{E}_{\mathscr{P}}$ | lower/upper (conditional) expectation corresponding to $\mathscr{P}$ · Eqs. (3.44)$_{89}$ and (3.45)$_{89}$ |
| $\mathscr{M}$ | set of initial probability mass functions · p. 90 |
| $\mathscr{Q}$ | set of rate operators · p. 90 |

| | |
|---|---|
| $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ | the set of homogeneous Markovian jump processes that are consistent with $\mathscr{M}$ and $\mathscr{Q}$ · Eq. (3.46)$_{90}$ |
| $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ | the set of Markovian jump processes that are consistent with $\mathscr{M}$ and $\mathscr{Q}$ · Eq. (3.46)$_{90}$ |
| $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ | the set of jump processes that are consistent with $\mathscr{M}$ and $\mathscr{Q}$ · Eq. (3.46)$_{90}$ |

**Operators**

| | |
|---|---|
| $d_{\mathfrak{Q}}(Q,\mathscr{Q})$ | distance between the rate operator $Q$ and the set of rate operators $\mathscr{Q}$ · Eq. (3.54)$_{94}$ |
| $\lVert\mathscr{Q}\rVert_{\mathrm{op}}$ | supremum of $\lVert Q\rVert_{\mathrm{op}}$ for $Q$ in $\mathscr{Q}$ · Eq. (3.63)$_{100}$ |
| $\underline{T},\overline{T}$ | lower and its conjugate upper transition operator · Definition 3.60$_{107}$ |
| $\underline{Q},\overline{Q}$ | lower and its conjugate upper rate operator · Definition 3.63$_{109}$ |
| $\underline{Q}_{\mathscr{Q}}$ | lower envelope of the set of rate operators $\mathscr{Q}$ · Eq. (3.71)$_{109}$ |
| $\mathscr{Q}_{\underline{Q}}$ | the set of rate operators that dominate the lower rate operator $\underline{Q}$ · Eq. (3.72)$_{110}$ |
| $(e^{t\underline{Q}})_{t\in\mathbb{R}_{\geq 0}}$ | (non-linear) semi-group generated by the lower rate operator $\underline{Q}$ · Proposition 3.74$_{114}$ |
| $\underline{E}_{\mathscr{M}}$ | lower envelope of $\{E_p\colon p\in\mathscr{M}\}$ · Eq. (3.76)$_{117}$ |

**Ergodicity**

| | |
|---|---|
| $\rho(\underline{T})$ | coefficient of ergodicity of the lower transition operator $\underline{T}$ · Eq. (4.19)$_{190}$ |
| $\bullet\hookrightarrow\bullet$ | upper reachability relation corresponding to the lower rate operator $\underline{Q}$ · Definition 4.31$_{192}$ |
| $\bullet\longrightarrow\bullet$ | lower reachability relation corresponding to the lower rate operator $\underline{Q}$ · Definition 4.31$_{192}$ |
| $\mathscr{X}_{\underline{Q}}$ | top class of the ergodic lower rate operator $\underline{Q}$ · Proposition 4.33$_{193}$ |
| $E_{\mathrm{lim}}$ | limit expectation of the ergodic rate operator $Q$ · Section 6.4.4$_{307}$ |
| $p_{\mathrm{lim}}$ | unique probability mass function such that $E_{p_{\mathrm{lim}}}=E_{\mathrm{lim}}$ · Section 7.3.2$_{376}$ |

**Lumping**

| | |
|---|---|
| $\hat{\mathscr{X}}$ | lumped state space · Section 7.1.1$_{342}$ |
| $\Lambda$ | lumping map from $\mathscr{X}$ to $\hat{\mathscr{X}}$ · Section 7.1.1$_{342}$ |

| | |
|---|---|
| $\Lambda^{-1}$ | inverse of the lumping map $\Lambda$ · Eq. (7.2)$_{342}$ |
| $\hat{\Omega}, \Omega_{\hat{\mathscr{X}}}$ | the set of lumped paths · Eq. (7.4)$_{343}$ |
| $\hat{X}_t$ | projector variable for $\hat{\Omega}$ · p. 344 |
| $\{\hat{X}_v \in B\}$ | cylinder/finitary event for $\hat{\Omega}$ · Eq. (7.5)$_{344}$ |
| $\hat{\mathscr{H}}$ | the set of conditioning events for $\hat{\Omega}$ · Eq. (7.6)$_{344}$ |
| $\hat{\mathscr{F}}_u$ | the field of finitary events of the form $\{\hat{X}_v \in B\}$ with $v \succcurlyeq u$ · Eq. (7.6)$_{344}$ |
| $\hat{\mathscr{D}}, \mathscr{D}_{\hat{\mathscr{X}}}$ | domain of a lumped jump process · Eq. (7.8)$_{344}$ |
| $\mathscr{D}^{\star}$ | structure of fields that includes $\mathscr{D}$ · Eq. (7.14)$_{346}$ |
| $P^{\star}$ | extension of the jump process $P$ to $\mathscr{D}^{\star}$ · p. 347 |
| $\Lambda_{\Omega}^{-1}$ | inverse on $\mathcal{P}(\hat{\Omega})$ of the lumping map $\Lambda$ · Eq. (7.9)$_{344}$ |
| $\hat{P}$ | lumped jump process · Eq. (7.15)$_{347}$ |
| $\hat{f}^{\uparrow\Omega}$ | cylindrical extension of the extended real variable $\hat{f}$ on $\hat{\Omega}$ to an extended real variable on $\Omega$ · Eq. (7.23)$_{350}$ |
| $\hat{p}_p$ | lumped probability mass function on $\hat{\mathscr{X}}$ induced by the probability mass function $p$ on $\mathscr{X}$ · Eq. (7.25)$_{354}$ |
| $\hat{\mathscr{M}}_{\mathscr{M}}$ | set of lumped probability mass functions induced by those in the set $\mathscr{M}$ · Eq. (7.26)$_{354}$ |
| $\underline{\hat{Q}}_{\mathscr{Q}}$ | lumped lower rate operator corresponding to the set $\mathscr{Q}$ of rate operators · Eq. (7.27)$_{355}$ |
| $\hat{\mathscr{Q}}_{\mathscr{Q}}$ | set of lumped rate operators that dominate $\underline{\hat{Q}}_{\mathscr{Q}}$ · Eq. (7.33)$_{358}$ |
| $f^{\downarrow\min}, f^{\downarrow\max}$ | transformation of the gamble $f$ on $\mathscr{X}_v$ to one on $\hat{\mathscr{X}}_v$ · Eqs. (7.38)$_{364}$ and (7.39)$_{364}$ |

# *Bibliography*


William J. Anderson (1991). *Continuous-Time Markov Chains. An Applications-Oriented Approach.* Springer. DOI: 10.1007/978-1-4612-3038-0 (Cited on pp. 55, 197).

Thomas Augustin, Frank P.A. Coolen, Gert de Cooman & Matthias C.M. Troffaes, eds. (2014). *Introduction to Imprecise Probabilities.* Wiley (Cited on p. 16).

Siamak Azodolmolky, Mirosław Klinkowski, Eva Marin, Davide Careglio, Josep Solé Pareta & Ioannis Tomkos (2009). A survey on physical layer impairments aware routing and wavelength assignment algorithms in optical networks. In: *Computer Networks* **53**.7, pp. 926–944. DOI: 10.1016/j.comnet.2008.11.014 (Cited on p. 405).

Christel Baier, Boudewijn Haverkort, Holger Hermanns & Joost-Pieter Katoen (2003). Model-checking algorithms for continuous-time Markov chains. In: *IEEE Transactions on Software Engineering* **29**.6, pp. 524–541. DOI: 10.1109/TSE.2003.1205180 (Cited on p. 281).

Christel Baier & Joost-Pieter Katoen (2008). *Principles of Model Checking.* MIT Press (Cited on p. 281).

Frank Ball & Geoffrey F. Yeo (1993). Lumpability and marginalisability for continuous-time Markov chains. In: *Journal of Applied Probability* **30**.3, pp. 518–528. DOI: 10.2307/3214762 (Cited on pp. 338, 342).

Patrizia Berti, Eugenio Regazzini & Pietro Rigo (1991). Coherent statistical inference and Bayes theorem. In: *The Annals of Statistics* **19**.1, pp. 366–381. DOI: 10.1214/aos/1176347988 (Cited on p. 16).

Patrizia Berti & Pietro Rigo (2002). On coherent conditional probabilities and disintegrations. In: *Annals of Mathematics and Artificial Intelligence* **35**, pp. 71–82. DOI: 10.1023/A:1014514730066 (Cited on pp. 16, 40, 229).

Dimitris Bertsimas & John N. Tsitsiklis (1997). *Introduction to Linear Optimization.* Athena Scientific (Cited on p. 377).

François Besnard & Lina Bertling (2010). An approach for condition-based maintenance optimization applied to wind turbine blades. In: *IEEE Transactions on Sustainable Energy* **1**.2, pp. 77–83. DOI: 10.1109/TSTE.2010.2049452 (Cited on p. 3).

Jeff Bezanson, Stefan Karpinski, Viral B. Shah & Alan Edelman (2012). Julia: A fast dynamic language for technical computing. arXiv: 1209.5145 [cs:PL] (Cited on p. 415).

K. P. S. Bhaskara Rao & M. Bhaskara Rao (1983). *Theory of Charges. A Study of Finitely Additive Measures*. Academic Press (Cited on pp. 33, 219, 462).

Patrick Billingsley (1995). *Probability and Measure*. 3rd ed. Wiley (Cited on pp. 11, 33, 35, 40, 55, 62, 220, 222, 245, 249–251, 253, 461, 463, 465–467, 471).

Patrick Billingsley (1999). *Convergence of Probability Measures*. 2nd ed. Wiley (Cited on p. 57).

Alexandr A. Borovkov (2013). *Probability Theory*. Springer. DOI: 10.1007/978-1-4471-5201-9 (Cited on pp. 55, 245, 251).

Bernard Bru & Marc Yor (2002). Comments on the life and mathematical legacy of Wolfgang Doeblin. In: *Finance and Stochastics* **6**.1, pp. 3–47. DOI: 10.1007/s780-002-8399-0 (Cited on p. 3).

Peter Buchholz (2005). An improved method for bounding stationary measures of finite Markov processes. In: *Performance Evaluation* **62**.1, pp. 349–365. DOI: 10.1016/j.peva.2005.07.002 (Cited on pp. 338, 343, 373, 374, 382–384).

Dmitri Burago, Yuri Burago & Sergei Ivanov (2011). *A Course in Metric Geometry*. American Mathematical Society (Cited on pp. 94, 128).

Cletus J. Burke & Murray Rosenblatt (1958). A Markovian function of a Markov chain. In: *The Annals of Mathematical Statistics* **29**.4, pp. 1112–1122. DOI: 10.1214/aoms/1177706444 (Cited on pp. 4, 337, 338, 342, 372, 373).

Kai Lai Chung (1960). *Markov Chains with Stationary Transition Probabilities*. Springer (Cited on pp. 3, 55, 74).

Jaume Comellas & Gabriel Junyent (2015). On the worthiness of flexible grid in elastic optical networks. In: *41st European Conference on Optical Communication (ECOC 2015)*. IEEE. DOI: 10.1109/ECOC.2015.7341797 (Cited on p. 405).

Aura Conci & Carlos S. Kubrusly (2017). Distance between sets – A survey. In: *Advances in Mathematical Sciences and Applications* **26**. arXiv: 1808.02574 [math.FA] (Cited on pp. 94, 128).

Michael G. Crandall & Amnon Pazy (1969). Semi-groups of nonlinear contractions and dissipative sets. In: *Journal of Functional Analysis* **3**.3, pp. 376–418. DOI: 10.1016/0022-1236(69)90032-9 (Cited on p. 79).

Lucio Crisma, Patrizia Gigante & Pietro Millossovich (1997). A notion of coherent prevision for arbitrary random quantities. In: *Journal of the Italian Statistical Society* **6**.3, pp. 233–243. DOI: 10.1007/bf03178914 (Cited on p. 219).

Percy J. Daniell (1918). A general form of integral. In: *Annals of Mathematics* **19**.4, pp. 279–294. DOI: 10.2307/1967495. JSTOR: 1967495 (Cited on pp. xv, xix, 6, 216, 220–222).

Jasper De Bock (2015). Credal Networks under Epistemic Irrelevance: Theory and Algorithms. PhD thesis. Ghent University (Cited on pp. 5, 166).

Jasper De Bock (2017a). Credal networks under epistemic irrelevance. In: *International Journal of Approximate Reasoning* **85**, pp. 107–138. DOI: 10.1016/j.ijar.2017.03.012 (Cited on p. 166).

Jasper De Bock (2017b). The limit behaviour of imprecise continuous-time Markov chains. In: *Journal of Nonlinear Science* **27**.1, pp. 159–196. DOI: 10.1007/s00332-016-9328-3 (Cited on pp. 54, 77, 108, 109, 111–113, 115, 137, 191–194).

Jasper De Bock & Gert de Cooman (2015). Extreme lower previsions. In: *Journal of Mathematical Analysis and Applications* **421**.2, pp. 1042–1080. DOI: 10.1016/j.jmaa.2014.07.067 (Cited on p. 445).

Jasper De Bock, Alexander Erreygers & Thomas Krak (2021). Sum-product laws and efficient algorithms for imprecise Markov chains. In: *Proceedings of the 37th Conference on Uncertainty in Artificial Intelligence (UAI 2021)* (Cited on pp. 8, 166, 277).

Gert de Cooman, Jasper De Bock & Stavros Lopatatzidis (2016). Imprecise stochastic processes in discrete time: Global models, imprecise Markov chains, and ergodic theorems. In: *International Journal of Approximate Reasoning* **76**, pp. 18–46. DOI: 10.1016/j.ijar.2016.04.009 (Cited on p. 308).

Gert de Cooman & Filip Hermans (2008). Imprecise probability trees: Bridging two theories of imprecise probability. In: *Artificial Intelligence* **172**.11, pp. 1400–1427. DOI: 10.1016/j.artint.2008.03.001 (Cited on pp. 54, 106, 107, 170).

Gert de Cooman, Filip Hermans, Alessandro Antonucci & Marco Zaffalon (2010). Epistemic irrelevance in credal nets: The case of imprecise Markov trees. In: *International Journal of Approximate Reasoning* **51**.9, pp. 1029–1052. DOI: 10.1016/j.ijar.2010.08.011 (Cited on p. 166).

Gert de Cooman, Filip Hermans & Erik Quaeghebeur (2009). Imprecise Markov chains and their limit behavior. In: *Probability in the Engineering and Informational Sciences* **23**.4, pp. 597–635. DOI: 10.1017/S0269964809990039 (Cited on pp. 54, 106–108, 170, 190).

Gert de Cooman & Erik Quaeghebeur (2012). Exchangeability and sets of desirable gambles. In: *International Journal of Approximate Reasoning* **53**.3, pp. 363–395. DOI: 10.1016/j.ijar.2010.12.002 (Cited on p. 388).

Bruno de Finetti (1970). *Teoria Delle Probabilità*. Giuliu Einaudi (Cited on pp. xiii, xvii, 16, 488).

Bruno de Finetti (2017). *Theory of Probability. A critical introductory treatment*. Trans. by Antonio Machí & Adrian Smith. Vol. 1. Reprint of (de Finetti, 1974, 1975). Wiley (Cited on pp. 12, 16, 17, 19–22, 26).

Bruno de Finetti (1974, 1975). *Theory of Probability*. Trans. by Antonio Machí & Adrian Smith. 2 vols. English translation of (de Finetti, 1970). Wiley (Cited on pp. 16, 488).

Robert Denk, Michael Kupper & Max Nendel (2018). Kolmogorov-type and general extension results for nonlinear expectations. In: *Banach Journal of Mathematical Analysis* **12**.3, pp. 515–540. DOI: 10.1215/17358787-2017-0024 (Cited on p. 440).

Dieter Denneberg (1994). *Non-Additive Measure and Integral*. Springer. DOI: 10.1007/978-94-017-2434-0 (Cited on p. 468).

Salem Derisavi, Holger Hermanns & William H. Sanders (2003). Optimal state-space lumping in Markov chains. In: *Information Processing Letters* **87**.6, pp. 309–315. DOI: 10.1016/S0020-0190(03)00343-0 (Cited on p. 338).

John D. Diener & William H. Sanders (1995). Empirical comparison of uniformization methods for continuous-time Markov chains. In: *Computations with Markov chains. Proceedings of the 2nd International Workshop on the Numerical Solution of Markov Chains*. Springer, pp. 547–570. DOI: 10.1007/978-1-4615-2241-6_29 (Cited on pp. 197, 364).

Wolfgang Doeblin (1938). Sur l'équation matricielle $A(t+s) = A(t)A(s)$ et ses applications aux probabilités en chaîne. In: *Bulletin des Sciences Mathématiques* **62**, pp. 21–32 (Cited on pp. 3, 77).

Joseph L. Doob (1953). *Stochastic Processes*. Wiley (Cited on pp. 3, 11, 55, 74, 245, 251).

Lester E. Dubins (1975). Finitely additive conditional probabilities, conglomerability and disintegrations. In: *Annals of Probability* **3**.1, pp. 89–99. DOI: 10.1214/aop/1176996451 (Cited on pp. 40, 41).

Nelson Dunford & Jacob T. Schwartz (1958). *Linear Operators. Part I: General Theory*. Interscience Publishers (Cited on p. 78).

Gustav Elfving (1937). Zur Theorie der Markoffschen Ketten. In: *Acta Societatis Scientiarum Fennicae* **2** (8). Translated to English as (Elfving, 1954) (Cited on pp. 77, 489).

Gustav Elfving (1954). *On the theory of Markoff chains*. Tech. rep. English translation of (Elfving, 1937). North Carolina State University, Department of Statistics, Institute of Statistics (Cited on p. 489).

Klaus-Jochen Engel & Rainer Nagel (2000). *One-Parameter Semigroups for Linear Evolution Equations*. Vol. 194. Graduate Texts in Mathematics. Springer. DOI: 10.1007/b97696 (Cited on p. 78).

Alexander Erreygers & Jasper De Bock (2017a). Imprecise continuous-time Markov chains: Efficient computational methods with guaranteed error bounds. In: *Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2017)*. PMLR, pp. 145–156 (Cited on pp. 7, 157, 194, 198, 489).

Alexander Erreygers & Jasper De Bock (2017b). Imprecise continuous-time Markov chains: Efficient computational methods with guaranteed error bounds. Extended preprint of (Erreygers & De Bock, 2017a). arXiv: 1702.07150 [math.PR] (Cited on pp. 115, 157, 181, 194).

Alexander Erreygers & Jasper De Bock (2018a). Computing inferences for large-scale continuous-time Markov chains by combining lumping with imprecision. In: *Uncertainty Modelling in Data Science (Proceedings of SMPS 2018)*. Springer, pp. 78–86 (Cited on pp. 7, 489).

Alexander Erreygers & Jasper De Bock (2018b). Computing inferences for large-scale continuous-time Markov chains by combining lumping with imprecision. Extended preprint of (Erreygers & De Bock, 2018a). arXiv: 1804.01020 [math.PR] (Cited on p. 345).

Alexander Erreygers & Jasper De Bock (2019a). Bounding inferences for large-scale continuous-time Markov chains: A new approach based on lumping and imprecise Markov chains. In: *International Journal of Approximate Reasoning* **115**, pp. 96–133. DOI: 10.1016/j.ijar.2019.09.003 (Cited on pp. 7, 338).

Alexander Erreygers & Jasper De Bock (2019b). First steps towards an imprecise Poisson process. In: *Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications (ISIPTA 2019)*. Vol. 103. PMLR, pp. 175–184 (Cited on pp. 7, 56, 441).

Alexander Erreygers & Jasper De Bock (2021). Extending the domain of imprecise jump processes from simple variables to measurable ones. In: *Proceedings of the Twelfth International Symposium on Imprecise Probabilities: Theories and Applications (ISIPTA 2021)*. Vol. 147. PMLR, pp. 140–149 (Cited on pp. 7, 230, 277, 440).

Alexander Erreygers, Jasper De Bock, Gert de Cooman & Arthur Van Camp (2019). Optimal control of a linear system subject to partially specified input noise. In: *International Journal of Robust and Nonlinear Control*. DOI: 10.1002/rnc.4583 (Cited on p. 7).

Alexander Erreygers & Enrique Miranda (2020). A study of the set of probability measures compatible with comparative judgements. In: *Information Processing and Management of Uncertainty in Knowledge-Based Systems (Proceedings of IPMU 2020)*. Springer, pp. 165–179. DOI: 10.1007/978-3-030-50143-3_13 (Cited on p. 8).

Alexander Erreygers & Enrique Miranda (2021). A graphical study of comparative probabilities. In: *Journal of Mathematical Psychology* **104**, p. 102582. DOI: 10.1016/j.jmp.2021.102582 (Cited on p. 8).

Alexander Erreygers, Cristina Rottondi, Giacomo Verticale & Jasper De Bock (2018a). Imprecise Markov models for scalable and robust performance evaluation of flexi-grid spectrum allocation policies. Extended preprint of (Erreygers, Rottondi, Verticale & De Bock, 2018b). arXiv: 1801.05700 [cs.NI] (Cited on pp. 417, 426).

Alexander Erreygers, Cristina Rottondi, Giacomo Verticale & Jasper De Bock (2018b). Imprecise Markov models for scalable and robust performance evaluation of flexi-grid spectrum allocation policies. In: *IEEE Transactions on Communications* **66**.11, pp. 5401–5414. DOI: 10.1109/TCOMM.2018.2846235 (Cited on pp. 7, 403, 490).

William Feller (1968). *An Introduction to Probability Theory and Its Applications*. 3rd ed. Vol. 1. Wiley (Cited on pp. 3, 11, 55, 339, 461).

Mark D Feuer & Sheryl L Woodward (2012). Advanced ROADM networks. In: *National Fiber Optic Engineers Conference*. Optical Society of America. DOI: 10.1364/NFOEC.2012.NW3F.3 (Cited on p. 405).

Giuliana Franceschinis & Richard R. Muntz (1994). Bounds for quasi-lumpable Markov chains. In: *Performance Evaluation* **20**.1, pp. 223–243. DOI: 10.1016/0166-5316(94)90015-9 (Cited on pp. 338, 340, 341, 343, 353, 364, 365, 373, 374, 378, 379, 382, 384).

Maurice Fréchet (1938). Récherches théoriques modernes sur les probabilités. Second Livre: Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles. In: *Traité du calcul des probabilités et de ses applications*. Ed. by Émile Borel. Tome I, Fasc. III (Cited on pp. 3, 77, 83).

Bert E. Fristedt & Lawrence F. Gray (1997). *A Modern Approach to Probability Theory*. Birkhäuser Basel. DOI: 10.1007/978-1-4899-2837-5 (Cited on pp. 40, 245, 461, 466, 468–470).

Arnab Ganguly, Tatjana Petrov & Heinz Koeppl (2014). Markov chain aggregation and its applications to combinatorial reaction networks. In: *Journal of Mathematical Biology* **69**.3, pp. 767–797. DOI: 10.1007/s00285-013-0738-7 (Cited on pp. 3, 343, 373).

Ori Gerstel, Masahiko Jinno, Andrew Lord & SJ Ben Yoo (2012). Elastic optical networking: a new dawn for the optical layer? In: *IEEE Communications Magazine* **50**.2, s12–s20. DOI: 10.1109/MCOM.2012.6146481 (Cited on pp. 404, 405).

Iosif I. Gikhman & Anatoly V. Skorokhod (1969). *Introduction to the Theory of Random Processes*. Trans. by Inc. Scripta Technica. W.B. Saunders Company (Cited on p. 55).

Daniel T. Gillespie (1977). Exact stochastic simulation of coupled chemical reactions. In: *The Journal of Physical Chemistry* **81**.25, pp. 2340–2361. DOI: 10.1021/j100540a008 (Cited on p. 419).

Jack Hachigian (1963). Collapsed Markov chains and the Chapman-Kolmogorov equation. In: *The Annals of Mathematical Statistics* **34**.1, pp. 233–237 (Cited on p. 338).

Ernst Hairer, Syvert P. Nørsett & Gerhard Wanner (2008). *Solving Ordinary Differential Equations I. Nonstiff Problems*. 3rd ed. Springer. DOI: 10.1007/978-3-540-78862-1 (Cited on pp. 174, 175).

Alan Hájek (2019). Interpretations of probability. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2019. Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/ (Cited on p. 15).

Paul R. Halmos (1974). *Measure Theory*. Springer (Cited on p. 467).

Darald J. Hartfiel (1985). On the solutions to $x'(t) = A(t)x(t)$ over all $A(t)$, where $P \leq A(t) \leq Q$. In: *Journal of Mathematical Analysis and Applications* **108**.1, pp. 230–240. DOI: 10.1016/0022-247X(85)90020-4 (Cited on p. 4).

Darald J. Hartfiel (1994). Lumping in Markov set-chains. In: *Stochastic Processes and their Applications* **50**.2, pp. 275–279. DOI: 10.1016/0304-4149(94)90124-4 (Cited on p. 338).

Darald J. Hartfiel (1998). *Markov Set-Chains*. Springer. DOI: 10.1007/bfb0094586 (Cited on p. 110).

Filip Hermans & Gert de Cooman (2012). Characterisation of ergodic upper transition operators. In: *International Journal of Approximate Reasoning* **53**.4, pp. 573–583. DOI: 10.1016/j.ijar.2011.12.008 (Cited on pp. 54, 108, 189–191, 206, 207).

Filip Hermans & Damjan Škulj (2014). Stochastic processes. In: *Introduction to Imprecise Probabilities*. Ed. by Thomas Augustin, Frank P.A. Coolen, Gert de Cooman & Matthias C.M. Troffaes. Wiley. Chap. 11 (Cited on p. 110).

Einar Hille & Ralph S. Phillips (1957). *Functional Analysis and Semi-Groups*. Vol. 31. Colloquium Publications. American Mathematical Society (Cited on p. 78).

Marius Iosifescu (1980). *Finite Markov Processes and their Applications*. Wiley (Cited on pp. 3, 55, 189, 305, 307).

Arieh Iserles (2009). *A First Course in the Numerical Analysis of Differential Equations*. 2nd ed. Cambridge University Press. DOI: 10.1017/cbo9780511995569 (Cited on pp. 174, 175).

ITU-T (2020). *Spectral grids for WDM applications: DWDM frequency grid*. Version 3. HDL: 11.1002/1000/11482 (Cited on p. 405).

Arne Jensen (1953). Markoff chains as an aid in the study of Markoff processes. In: *Scandinavian Actuarial Journal* **1953**.sup1, pp. 87–91. DOI: 10.1080/03461238.1953.10419459 (Cited on p. 364).

Masahiko Jinno, Hidehiko Takara, Bartlomiej Kozicki, Yukio Tsukishima, Yoshiaki Sone & Shinji Matsuoka (2009). Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. In: *IEEE Communications Magazine* **47**.11, pp. 66–73. DOI: 10.1109/MCOM.2009.5307468 (Cited on p. 405).

Joost-Pieter Katoen, Daniel Klink, Martin Leucker & Verena Wolf (2012). Three-valued abstraction for probabilistic systems. In: *The Journal of Logic and Algebraic Programming* **81**.4, pp. 356–389. DOI: 10.1016/j.jlap.2012.03.007 (Cited on pp. 281, 338, 363, 364).

Richard Kay (1986). A Markov model for analysing cancer markers and disease states in survival studies. In: *Biometrics* **42**.4, pp. 855–865. DOI: 10.2307/2530699. JSTOR: 2530699 (Cited on p. 3).

John G. Kemeny & J. Laurie Snell (1960). *Finite Markov Chains*. Springer-Verlag (Cited on p. 190).

Maurice G. Kendall & Buckland R. William (1971). *A Dictionary of Statistical Terms*. 3rd ed. Oliver and Boyd (Cited on p. 55).

Joobum Kim, Xue Wang, Shuyi WangYan, Miguel Razo, Tacca & Andrea Fumagalli (2015). Blocking fairness in two-service EONs with uneven arrival rates. In: *2015 17th International Conference on Transparent Optical Networks (ICTON 2015)*. IEEE. DOI: 10.1109/ICTON.2015.7193430 (Cited on p. 406).

Joobum Kim, Shuyi Yan, Andrea Fumagalli, Eiji Oki & Naoaki Yamanaka (2015). An analytical model of spectrum fragmentation in a two-service elastic optical link. In: *2015 IEEE Global Communications Conference (GLOBECOM 2015)*. IEEE. DOI: 10.1109/GLOCOM.2015.7417243 (Cited on pp. xv, xx, 3, 343, 403, 405–407, 421–424, 427).

Joobum Kim, Shuyi Yan, Andrea Fumagalli, Eiji Oki & Naoaki Yamanaka (2016). Two-service analytical model for partially-shared elastic optical link spectrum. In: *2016 IEEE 17th International Conference on High Performance Switching and Routing (HPSR 2016)*. IEEE, pp. 75–80. DOI: 10.1109/HPSR.2016.7525642 (Cited on pp. 405–407).

Andrei N. Kolmogorov (1931). Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. In: *Mathematische Annalen* **104**. Translated to English as (Kolmogorov, 1992). DOI: 10.1007/BF01457949 (Cited on pp. 3, 77, 493).

Andrei N. Kolmogorov (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Translated to English as (Kolmogorov, 1950). Springer (Cited on pp. xiii, xvii, 461, 493).

Andrei N. Kolmogorov (1950). *Foundations of the Theory of Probability*. English translation of (Kolmogorov, 1933). Chelsea (Cited on pp. 461, 493).

Andrei N. Kolmogorov (1992). On analytical methods in probability theory. In: *Selected Works of A. N. Kolmogorov. Volume II: Probability Theory and Mathematical Statistics*. Ed. and trans. by Albert N. Shiryaev. English translation of (Kolmogorov, 1931). Springer. Chap. 9. DOI: 10.1007/978-94-011-2260-3_9 (Cited on p. 493).

Heinz König (1997). *Measure and Integration. An Advanced Course in Basic Procedures and Applications*. Springer. DOI: 10.1007/978-3-540-89502-2 (Cited on p. 468).

Thomas Krak (2021). Continuous-Time Imprecise-Markov Chains: Theory and Algortihms. PhD thesis. Ghent University (Cited on p. 5).

Thomas Krak, Jasper De Bock & Arno Siebes (2017). Imprecise continuous-time Markov chains. In: *International Journal of Approximate Reasoning* **88**, pp. 452–528. DOI: 10.1016/j.ijar.2017.06.012 (Cited on pp. xiii, xiv, xvii, xviii, 4, 5, 43, 44, 54, 55, 58, 60, 62, 65, 66, 74, 75, 84–88, 90–93, 98, 101, 104–106, 108, 110, 111, 117, 118, 120, 124, 125, 132, 134, 138, 139, 141, 143–145, 147, 149, 153, 157, 158, 163, 170, 174–178, 182, 192, 244).

Thomas Krak, Alexander Erreygers & Jasper De Bock (2018). An imprecise probabilistic estimator for the transition rate matrix of a continuous-time Markov chain. In: *Uncertainty Modelling in Data Science (Proceedings of SMPS 2018)*, pp. 124–132. DOI: 10.1007/978-3-319-97547-4_17. arXiv: 1804.01330 [math.PR] (Cited on pp. 7, 90).

Thomas Krak, Natan T'Joens & Jasper De Bock (2019). Hitting times and probabilities for imprecise Markov chains. In: *Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications*. Vol. 103. PMLR, pp. 265–275. arXiv: 1905.08781 [math.PR] (Cited on pp. 289, 292).

Vidyadhar G. Kulkarni (2011). *Introduction to Modeling and Analysis of Stochastic Systems*. Springer. DOI: 10.1007/978-1-4419-1772-0 (Cited on p. 305).

Jean-François Le Gall (2016). *Brownian Motion, Martingales, and Stochastic Calculus*. Springer. DOI: 10.1007/978-3-319-31089-3 (Cited on pp. 55, 57, 259).

Isaac Levi (1983). *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. MIT press (Cited on p. 27).

Paul Lévy (1955). W. Doeblin (V. Doblin) (1915-1940). In: *Revue d'histoire des sciences et de leurs applications* **8**.2, pp. 107–115. DOI: 10.2307/45242851. JSTOR: 45242851 (Cited on p. 3).

Stavros Lopatatzidis (2017). Robust Modelling and Optimisation in Stochastic Processes Using Imprecise Probabilities, With an Application to Queueing Theory. PhD thesis. Ghent University (Cited on pp. 5, 33, 34, 44, 51, 229).

Andrei A. Markov (1906). Extension of the law of large numbers to dependent quantities. Russian. In: *Bulletin of the Physio-Mathematical Society of the University of Kazan*. 2nd ser. **15**, pp. 135–156 (Cited on p. 2).

Enrique Miranda & Marco Zaffalon (2017). Full conglomerability. In: *Journal of Statistical Theory and Practice* **11**.4, pp. 634–669. DOI: 10.1080/15598608.2017.1295890 (Cited on p. 239).

Isao Miyadera (1992). *Nonlinear semigroups*. American Mathematical Society (Cited on p. 79).

Cleve Moler & Charles Van Loan (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. In: *SIAM Review* **45**.1, pp. 3-4–9. DOI: 10.1137/S00361445024180 (Cited on p. 160).

Mitio Nagumo (1936). Einige analytische Untersuchungen in linearen, metrischen Ringen. In: *Japanese journal of mathematics: Transactions and abstracts* **13**, pp. 61–80. DOI: 10.4099/jjm1924.13.0_61 (Cited on p. 78).

D. S. Nathan (1935). One-parameter groups of transformations in abstract vector spaces. In: *Duke Mathematical Journal* **1**.4, pp. 518–526. DOI: 10.1215/S0012-7094-35-00139-9 (Cited on p. 78).

Max Nendel (2020). Markov chains under nonlinear expectation. In: *Mathematical Finance* **31**.1, pp. 474–507. DOI: 10.1111/mafi.12289 (Cited on p. 439).

Max Nendel (2021). On nonlinear expectations and Markov chains under model uncertainty. In: *International Journal of Approximate Reasoning* **130**, pp. 226–245. DOI: 10.1016/j.ijar.2020.12.013 (Cited on p. 439).

James R. Norris (1997). *Markov Chains*. Cambridge University Press. DOI: 10.1017/CBO9780511810633 (Cited on pp. 3, 55, 79, 87, 287, 289, 291, 292, 307, 308, 374, 408, 417, 419).

Krzysztof Pawlikowski (1990). Steady-state simulation of queueing processes: Survey of problems and solutions. In: *ACM Computing Surveys* **22**.2, pp. 123–170. DOI: 10.1145/78919.78921 (Cited on p. 419).

Amnon Pazy (1983). Semigroups of operators in Banach spaces. In: *Equadiff 82*. Springer, pp. 508–524 (Cited on p. 78).

Shige Peng (2005). Nonlinear expectations and nonlinear Markov chains. In: *Chinese Annals of Mathematics* **26**.02, pp. 159–184. DOI: 10.1142/S0252959905000154 (Cited on p. 440).

Erik Quaeghebeur (2009). Learning from Samples Using Coherent Lower Previsions. PhD thesis. Ghent University (Cited on p. 5).

Erik Quaeghebeur (2014). Desirability. In: *Introduction to Imprecise Probabilities*. Ed. by Thomas Augustin, Frank P.A. Coolen, Gert de Cooman & Matthias C.M. Troffaes. Wiley. Chap. 2 (Cited on pp. 17, 19).

Eugenio Regazzini (1985). Finitely additive conditional probabilities. In: *Rendiconti del Seminario Matematico e Fisico di Milano* **55**.1, pp. 69–89. DOI: 10.1007/BF02924866 (Cited on pp. 16, 40, 41, 44, 45, 47, 48).

R. Tyrrell Rockafellar & Roger J.B. Wets (1998). *Variational Analysis*. Springer. DOI: 10.1007/978-3-642-02431-3 (Cited on pp. 94, 95).

Cristina Rottondi, Alexander Erreygers, Giacomo Verticale & Jasper De Bock (2017). Modelling spectrum assignment in a two-service flexi-grid optical link with imprecise continuous-time Markov chains. In: *Proceedings of the 13th International Conference on the Design of Reliable Communication Networks (DRCN 2017)*. VDE Verlag, pp. 39–46 (Cited on pp. 6, 403).

Halsey L. Royden (1968). *Real Analysis*. 2nd ed. Macmillan (Cited on p. 220).

Gerardo Rubino & Bruno Sericola (1993). A finite characterization of weak lumpable Markov processes. Part II: The continuous time case. In: *Stochastic Processes and their Applications* **45**.1, pp. 115–125. DOI: 10.1016/0304-4149(93)90063-A (Cited on p. 338).

Walter Rudin (1976). *Principles of Mathematical Analysis*. 3rd ed. McGraw-Hill (Cited on pp. 297, 299).

Youcef Saad & Martin H. Schultz (1986). GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. In: *SIAM Journal on Scientific and Statistical Computing* **7**.3, pp. 856–869. DOI: 10.1137/0907058 (Cited on p. 418).

Eric Schechter (1997). *Handbook of Analysis and Its Foundations*. Academic Press (Cited on pp. 61, 75–77, 94, 126, 179, 296, 297, 299, 301, 446).

Eugene Seneta (1979). Coefficients of ergodicity: Structure and applications. In: *Advances in Applied Probability* **11**.3, pp. 576–590 (Cited on p. 197).

Eugene Seneta (1981). *Non-negative Matrices and Markov Chains*. Springer (Cited on p. 374).

Eugene Seneta (1996). Markov and the birth of chain dependence theory. In: *International Statistical Review / Revue Internationale de Statistique* **64**.3, pp. 255–263. DOI: 10.2307/1403785. JSTOR: 1403785 (Cited on p. 2).

Eugene Seneta (2006). Markov and the the creation of Markov chains. In: *Markov Anniversary Meeting 2006*. Boson Books, pp. 1–20 (Cited on p. 2).

Glenn Shafer & Vladimir Vovk (2001). *Probability and Finance: It's Only a Game!* Wiley. DOI: 10.1002/0471249696 (Cited on p. 440).

Glenn Shafer & Vladimir Vovk (2019). *Game-Theoretic Foundations for Probability and Finiance*. Wiley. DOI: 10.1002/9781118548035 (Cited on p. 440).

Zhi-shu Shen, Hiroshi Hasegawa & Ken-ichi Sato (2014). A novel flexible grid/semi-flexible grid optical path network design algorithm that reserves exclusive frequency slots for high bitrate signals. In: *2014 International Conference on Optical Network Design and Modeling (ONDM 2014)*. IEEE, pp. 287–292 (Cited on p. 405).

Zhi-shu Shen, Hiroshi Hasegawa, Ken-ichi Sato, Takafumi Tanaka & Akira Hirano (2014). A novel elastic optical path network that utilizes bitrate-specific anchored frequency slot arrangement. In: *Optics Express* **22**.3, pp. 3169–3179. DOI: 10.1364/OE.22.003169 (Cited on p. 405).

Albert N. Shiryaev (2016). *Probability-1. Mathematical Foundations of Probability Theory*. Springer. Chap. 2 (Cited on pp. 36, 40, 220, 461, 463, 465–470).

Damjan Škulj (2009). Discrete time Markov chains with interval probabilities. In: *International Journal of Approximate Reasoning* **50**.8, pp. 1314–1329. DOI: 10.1016/j.ijar.2009.06.007 (Cited on p. 110).

Damjan Škulj (2015). Efficient computation of the bounds of continuous time imprecise Markov chains. In: *Applied Mathematics and Computation* **250**, pp. 165–180. DOI: 10.1016/j.amc.2014.10.092 (Cited on pp. 4, 54, 109, 115, 174, 175).

Damjan Škulj & Robert Hable (2013). Coefficients of ergodicity for Markov chains with uncertain parameters. In: *Metrika* **76**.1, pp. 107–133. DOI: 10.1007/s00184-011-0378-0 (Cited on pp. 54, 189–191, 197, 443, 444).

Gerard L. G. Sleijpen & Diederik R. Fokkema (1993). BiCGstab(l) for linear equations involving unsymmetric matrices with complex spectrum. In: *Electronic Transactions on Numerical Analysis* **1**, pp. 11–32 (Cited on p. 418).

Peter Sonneveld & Martin B. van Gijzen (2009). IDR(s): A family of simple and fast algorithms for solving large nonsymmetric systems of linear equations. In: *SIAM Journal on Scientific Computing* **31**.2, pp. 1035–1062. DOI: 10.1137/070685804 (Cited on p. 418).

William J. Stewart (2009). *Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling*. Princeton University Press (Cited on pp. 55, 364, 374, 376, 379, 382, 418).

Natan T'Joens & Jasper De Bock (2021). Average behaviour in discrete-time imprecise Markov chains: A study of weak ergodicity. In: *International Journal of Approximate Reasoning* **132**, pp. 181–205. DOI: 10.1016/j.ijar.2021.03.001. arXiv: 2102.04793 [math.PR] (Cited on p. 308).

Natan T'Joens, Thomas Krak, Jasper De Bock & Gert de Cooman (2019). A recursive algorithm for computing inferences in imprecise Markov chains. In: *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. Springer, pp. 455–465. DOI: 10.1007/978-3-030-29765-7_38. arXiv: 1905.12968 [math.PR] (Cited on p. 166).

Angus E. Taylor (1985). *General Theory of Functions and Integration*. Dover (Cited on pp. 8, 220, 223, 225, 226, 451, 453–455, 457, 458, 471).

Petar Todorovic (1992). *An Introduction to Stochastic Processes and Their Applications*. Springer. DOI: 10.1007/978-1-4613-9742-7 (Cited on pp. 245, 251).

Ioannis Tomkos, Eleni Palkopoulou & Marianna Angelou (2012). A survey of recent developments on flexible/elastic optical networking. In: *2012 14th International Conference on Transparent Optical Networks (ICTON 2012)*. IEEE. DOI: 10.1109/ICTON.2012.6254409 (Cited on p. 405).

Antonio Tornambè (1995). *Discrete-Event System Theory: An Introduction*. World Scientific (Cited on pp. 55, 189, 191, 376, 379).

Matthias C.M. Troffaes (2013). Model checking for infinite horizon Markov chains. Talk at the Sixth Workshop on Principles and Methods of Statistical Inference with Interval Probability (WPMSIIP 2013). Lugano, Switzerland (Cited on p. 217).

Matthias C.M. Troffaes & Simon Blake (2013). A robust data driven approach to quantifying common-cause failure in power networks. In: *Proceedings* *of the Eighth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2013)*, pp. 311–317 (Cited on p. 313).

Matthias C.M. Troffaes & Gert de Cooman (2014). *Lower Previsions*. Wiley Series in Probability and Statistics. Wiley (Cited on pp. 5, 8, 16, 17, 21–25, 27, 28, 30, 31, 35–38, 44, 47, 48, 50, 51, 217, 219, 241, 242, 271).

Matthias C.M. Troffaes, Jacob Gledhill, Damjan Škulj & Simon Blake (2015). Using imprecise continuous time Markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair. In: *Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA 2015)*, pp. 287–294 (Cited on pp. 3, 313, 314, 319–321, 373, 417).

Matthias C.M. Troffaes, Thomas Krak & Henna Bains (2019). Two-state imprecise Markov chains for statistical modelling of two-state non-Markovian processes. In: *Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications*. Vol. 103. PMLR, pp. 394–403 (Cited on p. 116).

Antti Valmari & Giuliana Franceschinis (2010). Simple $O(m \log n)$ Time Markov chain lumping. In: *Tools and Algorithms for the Construction and Analysis of Systems*. Springer Berlin Heidelberg, pp. 38–52 (Cited on p. 338).

Arthur Van Camp (2017). Choice Functions as a Tool to Model Uncertainty. PhD thesis. Ghent University (Cited on p. 5).

Charles F. Van Loan (1975). *A study of the matrix exponential*. Tech. rep. Numerical Analysis Report No. 10. Reissued as MIMS EPrint 2006.397. University of Manchester (Cited on p. 79).

Peter Walley (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall (Cited on pp. xiii, xvii, 15, 16, 20–22, 24–31, 38).

Xue Wang, Joobum Kim, Shuyi Yan, Miguel Razo, Marco Tacca & Andrea Fumagalli (2014). Blocking probability and fairness in two-rate Elastic Optical Networks. In: *2014 16th International Conference on Transparent Optical Networks (ICTON 2014)*. IEEE. DOI: 10.1109/ICTON.2014.6876707 (Cited on p. 406).

Peter Whittle (2000). *Probability Via Expectation*. 4th ed. Springer. DOI: 10.1007/978-1-4612-0509-8 (Cited on p. 80).

David Williams (1991). *Probability with Martingales*. Cambridge University Press. DOI: 10.1017/cbo9780511813658 (Cited on pp. 32, 36, 220, 452, 461, 466–468).

Peter M. Williams (1975). *Notes on conditional previsions*. Tech. rep. Reprinted as (P. M. Williams, 2007). School of Mathematical and Physical Sciences, University of Sussex (Cited on pp. xiii, xvii, 12, 16, 44, 47, 499).

Peter M. Williams (2007). Notes on conditional previsions. In: *International Journal of Approximate Reasoning* **44**.3. Reprint of (P. M. Williams, 1975), pp. 366–383. DOI: 10.1016/j.ijar.2006.07.019 (Cited on pp. 12, 16, 44, 47, 498).

Ronald W. Wolff (1982). Poisson arrivals see time averages. In: *Operations Research* **30**.2, pp. 223–231. DOI: 10.1287/opre.30.2.223. JSTOR: 170165 (Cited on pp. 412, 417).

Shuyi Yan, Miguel Razo, Marco Tacca & Andrea Fumagalli (2014). A blocking probability estimator for the multi-application and multi-resource constraint problem. In: *2014 International Conference on Computing, Networking and Communications (ICNC 2014)*, pp. 921–926. DOI: 10.1109/ICCNC.2014.6785460 (Cited on p. 427).

Yawei Yin, Mingyang Zhang, Zuqing Zhu & S. J. B. Yoo (2013). Fragmentation-aware routing, modulation and spectrum assignment algorithms in elastic optical networks. In: *Optical Fiber Communication Conference*. Optical Society of America. DOI: 10.1364/OFC.2013.OW3A.5 (Cited on p. 406).

Kôsaku Yosida (1936). On the group embedded in the metrical complete ring. In: *Japanese journal of mathematics: Transactions and abstracts* **13**, pp. 7–26. DOI: 10.4099/jjm1924.13.0_7 (Cited on p. 78).

Yiming Yu, Jie Zhang, Yongli Zhao, Xuping Cao, Xuefeng Lin & Wanyi Gu (2013). The first single-link exact model for performance analysis of flexible grid WDM networks. In: *National Fiber Optic Engineers Conference*. Optical Society of America. DOI: 10.1364/NFOEC.2013.JW2A.68 (Cited on p. 406).

Guoying Zhang, Marc De Leenheer, Annalisa Morea & Biswanath Mukherjee (2013). A survey on OFDM-based elastic core optical networking. In: *IEEE Communications Surveys & Tutorials* **15**.1, pp. 65–87. DOI: 10.1109/SURV.2012.010912.00123 (Cited on p. 405).