Robust Modelling and Optimisation in Stochastic Processes Using Imprecise Probabilities, with an Application to Queueing Theory

Stavros Lopatatzidis

Promotoren: prof. dr. ir. G. De Cooman, dr. ir. J. De Bock, prof. dr. ir. S. De Vuyst
Proefschrift ingediend tot het behalen van de graad van
Doctor in de ingenieurswetenschappen: wiskundige ingenieurstechnieken

UNIVERSITEIT GENT

Vakgroep Elektronica en Informatiesystemen
Voorzitter: prof. dr. ir. R. Van de Walle
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2016 - 2017

ISBN 978-94-6355-036-9
NUR 916, 984
Wettelijk depot: D/2017/10.500/71

PROMOTOREN

prof. dr. ir. Gert de Cooman, UGent
dr. ir. Jasper De Bock, UGent
prof. dr. ir. Stijn De Vuyst, UGent

LEDEN VAN DE LEESCOMMISSIE

prof. dr. ir. Hendrik De Bie, UGent
prof. dr. Damjan Škulj, University of Ljubljana
prof. dr. ir. Matthias Troffaes, Durham University
prof. dr. ir. Joris Walraevens, UGent (secretaris)

OVERIGE LEDEN VAN DE EXAMENCOMMISSIE

prof. dr. ir. Alain Sarlette, UGent
prof. dr. ir. Jan Van Campenhout, UGent (voorzitter)

ADRES

Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur
Vakgroep Elektronica en Informatiesystemen
Onderzoeksgroep IDLab

Technologiepark-Zwijnaarde 914
9052 Zwijnaarde
België

FINANCIERING

Dit proefschrift bevat de resultaten van onderzoek gefinancierd door het Fonds voor Wetenschappelijk Onderzoek – Vlaanderen (FWO).

Contents

Preface **7**

Summary **10**

Samenvatting (Dutch summary) **13**

List of symbols **17**

1 Introduction **28**
1.1 General description of the ideas, motivation and main contribution . . . 28
1.2 Information about references . . . 32
1.3 Overview of the chapters . . . 32
1.4 Publications . . . 34

2 Modelling uncertainty **36**
2.1 Probability . . . 36
2.2 Expectation . . . 38
2.3 Sets of probability mass functions . . . 40
2.4 Lower and upper expectations . . . 44
2.5 Lower and upper expectations for a special class of sets of probability mass functions . . . 47
2.6 Natural Extension . . . 50
2.7 Unbounded gambles and conditional models . . . 56

3 Discrete-time stochastic processes and how to make them imprecise **57**
3.1 Notation . . . 58
3.2 Event trees . . . 59
3.3 Probability trees . . . 61
3.4 Conditional probability measures . . . 61
3.5 Expectations in probability trees . . . 72
3.6 General imprecise probability trees . . . 79

3.7 Imprecise probability trees derived from sets of conditional probability mass functions 80

4 A martingale-theoretic approach for imprecise discrete-time stochastic processes 86

4.1 Notation 87
4.2 Sub- and supermartingales 87
4.3 An alternative approach for global models 88
4.4 Connection with measure-theoretic approach 95

5 Imprecise discrete-time Markov chains 100

5.1 Markov chains 101
5.2 Homogeneous Markov chains 109
5.3 Towards imprecision 113
5.4 Imprecise Markov chains under epistemic irrelevance 115
5.5 Homogeneous imprecise Markov chains under epistemic irrelevance 126
5.6 Imprecise Markov chains under complete independence 131
5.7 Imprecise Markov chains under repetition independence . . . 142
5.8 Discussion 147

6 Imprecise discrete-time birth-death chains 151

6.1 Birth-death chains 152
6.2 Imprecise birth-death chains 152
6.3 First-passage and return times 155
6.4 Lower and upper expectations of upward first-passage times . 161
6.5 Lower and upper expectations of downward first-passage times 165
6.6 Lower and upper expectations of return times 167
6.7 Precise birth-death chains as a special case 168
6.8 Connection with measure-theoretic approach 177
6.A Useful properties 181

7 An application to queueing 189

7.1 Imprecision and queueing theory 190
7.2 The Geo/Geo/1/L queue and how to make it imprecise 191
7.3 Fixed-parameter repetition independence 197
7.4 Our example of an imprecise Geo/Geo/1/L queue 200
7.5 Expected queue length 202
7.6 Probability of different queue lengths 206
7.7 Turning on the server 210
7.8 Expected first-passage and return times 219

8 Conclusions 230

A Measure-theoretic probability **233**
A.1 Algebras, σ-algebras and generators 234
A.2 Probability measures 236
A.3 Measurable functions 237
A.4 Expectations in probability spaces 239

Bibliography **242**

Preface

Constantine P. Cavafy in his poem 'Ithaka' starts with the following lines:

"As you set out for Ithaka
hope the voyage is a long one,
full of adventure, full of discovery."

These lines perfectly describe my journey, with as destination the dissertation that you are about to read.

It all started a few years ago when I obtained my master's degree from Utrecht University and decided to aim for a doctoral degree. Among my applications to different universities, I was offered a doctoral position in Ghent University, which I accepted without hesitation and much thinking. This was a decision of life that I will never regret and forget, and moreover, I am proud of it. From the moment that I started doing research in Ghent University till now, it seems like a journey to me. A journey full of adventure, full of discovery, full of unique moments, full of beautiful memories. I stop here because this is a very big "full of...". Even if there were many stressful periods, in which the schedule was hectic and the pressure was intense, in the end the result was rewarding and I now look back on these days with nostalgia because they will probably not be repeated again. I learnt a lot of values during the years of my doctoral studies. I gained experience on how to do proper and solid research, I learnt and I was taught many fascinating aspects of mathematics, and what is more, I changed the way in which I perceive things. All the aforementioned were achieved through personal effort and, of course, through the help of other people, whom I would like to thank here.

First and foremost, I would like to thank my main supervisor Gert. Thanks to him, I learnt many things inside and outside research and mathematics. To me, he is not just a supervisor. He is a friend, a person that I admire and I am more than thankful to him for giving me the opportunity to work in his group. I will not forget the active discussions we had on various topics, related or not to research, the funny stories and the jokes he was telling, and the amazing time we had at his place and during our roadtrips when going to conferences. I am glad that I have met him.

I would also like to thank all the people, most of them friends by now, from my close working environment. First of all, I would like to thank my dear colleague and supervisor Jasper. I usually don't like to say big words about people, but I think he deserves the title of "The best researcher I have met in my life so far". Without him, I wouldn't be able to write this dissertation. The support and help he gave me all these years is invaluable, let alone the nice moments full of laughter that we shared. He is also the one that taught me to look into a problem from different angles till I find the most elegant solution. I owe a lot to this guy. Together with Gert and Jasper there is also one more person who completes what I call "The magic triplet" and that is Arthur. The most kind and lovely person I have probably met in my life. Always with positive energy and pure willingness to help you with anything. Especially during the stressful periods, he was the one who made it easier, not to mention the relaxed periods during which we had the fun of our lives. I believed that selflessness does not exist, but when I think of Arthur, I reconsider. Thank you Arthur. I would also like to thank Joris and Stijn, with whom I had a lot of interesting discussions about different research topics and how they can be tackled, and who helped me to publish my first result as first author which was the stepping stone for publishing more scientific results.

Many thanks as well to several other researchers, and especially to Quique, Erik, Matthias, Sebastien, Marcio, Alain, Alexander, Thomas and Meizhu, with whom I had nice moments in the office and at conferences. I would also like to thank the members of my examination committee, for accepting to read this dissertation and for providing useful comments. Finally, I would like to thank two more people that introduced me to the world of research and taught me many useful things. The first one is Nikos, with whom I had my first steps towards research. We managed to publish a result together back in 2010. The second one is Linda, the supervisor of my master's thesis in Utrecht University. She introduced me to the field of Bayesian networks and probabilistic systems and it is because of her that I found the doctoral position at Ghent University. She believed in me from the first moment and I am glad that we still work together from time to time.

Next, I would like to thank all the friends that supported me when I needed it the most, each in his/her own way. They always cared about me and wanted to know how things were going with my research. Some of them are friends of mine for more than 15 years and have offered me precious moments with a lot of fun that I hope I will never forget in my life. Starting with my buddies, some of them located in Athens and some others across Europe, many thanks to Nikos, Jack, Rafa, Alex, Anna, Michalis, Giorgos, Dimi, George, Emil, Vassiliki, Chara, Antonis, Themis, Mike, Panos, Labros, Yiannis, Costaras, Taseas, George junior, Yuli, Angeliki and Nikol. I would also like to thank the friends that I met in the Netherlands and Belgium during my studies, and especially Konstantinos, Agis, Vaggelis, Rais, Irini, Vally, Myrto, Nina, Filippos, Marpessa and Chris. Furthermore, I want to thank my buddies here in Ghent.

Many thanks to Liana and Tasia (a.k.a. Moons and Bells), Billy Maglo, Billy Vanilly, Costakis, Gregor, Elias, Mirella, Marilena, Tonia, Themis, Anouk, Jona, Eftichia and Frederik. Special thanks to my dear amigos and band mates Flojo and Simone—our jamming sessions were something to look forward to every week. One last, but not least, "Thank you" goes out to Melena for all the support and affection that she has given me, and in general for all the things she has done for me in her unique and special way during these years of my doctoral studies. Honestly guys, I thank you all from the depths of my heart.

I would also like to thank some relatives of mine that will always be there for me, ready to help in any situation. Just by asking me if I am fine and feeling proud of me, it means a lot to me. Therefore, many thanks to Aggelikoula, Panos, Katia, Chrisostomos, Elina, Akis, Elli, Manolis, Yiannis and Chrysoula. Regarding my family, I do not have the words to tell them how thankful I am, not only for their financial and psychological support, but also for their effort in trying to make me a better person. I feel lucky that I grew up in such a family, really. A huge "Thank you" to my mother Eleni, to my father Thanasis and to my brother Tasos.

Finally, I would like to thank music. I do not know how this world would be without music. So thank you to all the songs and kinds of music that accompanied me when I was working or when I wanted to relax, and created a chilling atmosphere that made me keep working even harder at times that I thought I would give up.

SUMMARY

Stochastic processes are systems that evolve through time, but whose specific time evolution is uncertain. This dissertation presents methods for computing robust inferences in such stochastic processes, and applies them to the field of queueing theory. In order to achieve this, we use imprecise probabilities.

We begin this dissertation by presenting some basic concepts from the theory of imprecise probabilities, which is a robust extension of probability theory. When we have a variable for which we are uncertain about its value, probability theory will assign a probability to each possible value by defining a single probability mass function on the set of possible values, whereas imprecise probability theory considers a set of probability mass functions. In the vast majority of cases, these sets of probability mass functions are assumed to be closed and convex and are then called credal sets. We show through a series of examples how these credal sets can be constructed and graphically represented. In addition, we also discuss cases where convexity might break down as this sometimes happens in the queueing models that we consider later in the text. Next, we show how to model uncertainty using lower expectation operators, which are non-linear operators that satisfy a number of axioms. Such a lower expectation operator can be associated with a set of probability mass functions and for a given function, its value can be interpreted as an exact lower bound on the set of possible expected values of that function. Alternatively, it can also be interpreted as a supremum buying price. A function then can be seen as a gamble that we are willing to buy for any price lower than its minimum expected value. We also discuss how natural extension allows us to extend the domain of a partially specified lower expectation operator.

Next in this dissertation, we introduce a theory of robust discrete-time stochastic processes. We deal with stochastic processes that have a finite state space—which is the set of possible states of the process at any time point—and we consider probability mass functions for the state of the process at any time point, conditional on the complete history of states. These conditional probability mass functions are what we call our local models, and we use them for deriving other more complex conditional or unconditional probabilities in our system and for computing expected values of functions that may even depend on the states in an infinite number of time points (also called

variables). In order to do that we borrow elements from the field of measure-theoretic probability, which is a sub-field of the wide area of measure theory. However, compared to the standard measure-theoretic approach which derives conditional probabilities from unconditional ones, our approach derives them directly from the local models using the definition of (full) conditional probability and coherent conditional probability, which in addition allows us to define conditional probabilities even when the event on which we condition has probability zero. The robust aspect of our theory lies in the fact that the local models are allowed to be partially specified, which gives rise to a set of stochastic processes, also called an imprecise stochastic process. We build our global models based on the local models, by considering two different approaches: (i) applying the framework of measure-theoretic probability to the set of stochastic processes and (ii) applying the martingale-theoretic framework. Since the approach based on measure-theoretic probability requires many assumptions, we mainly use the martingale-theoretic approach. At this point we have our first simple, but quite important, contribution: a generalisation of some basic properties that are typically satisfied in the precise-probabilistic versions of these systems. They are extensions of the axioms that are satisfied by lower expectation operators, and a law of iterated lower expectations. We generalise them by considering functions that may depend on the value of the states in an infinite number of time points (variables) and by dropping the assumption that the local sets of probability mass functions should be closed and convex. We also establish connections between the global models derived from these two approaches and for functions that depend on a finite number of variables, we find that both approaches coincide.

Moving on, we provide a detailed analysis of a well-known family of stochastic processes, which are discrete-time Markov chains with a finite state space. These models are particularly easy to use because the local models do not condition on the complete history of states but only on the latest state. We distinguish between time-homogeneous and time-inhomogeneous Markov chains, where the difference lies in whether the local models depend on time or not. As in the case of general imprecise stochastic processes, we then robustify these models by allowing our local models to be partially specified. In this way, we obtain a set of stochastic processes, called an imprecise Markov chain. In contrast with the standard assumption that the local models of an imprecise Markov chain are closed and convex sets of probability mass functions, we again consider general sets of probability mass functions. Furthermore, we impose different types of independence between the states of the process. For each of these concepts of independence, we show how to compute—exactly and tightly when possible—bounds on the expected values of functions of interest and we prove various properties. In particular, we find that the fewer assumptions we impose on the independence concept the wider are the bounds on the expected values and also the more efficient are the computations. We also show that our properties for imprecise stochastic processes remain valid here as well, and discuss and prove some additional ones.

Next, we focus on a specific type of Markov chains, called discrete-time birth-death chains with a finite state space. We first present some preliminaries about discrete-time birth-death chains, and then define our imprecise version of a birth-death chain as a special case of an imprecise Markov chain. For these imprecise birth-death chains, we then focus on the problem of computing expected first-passage and return times. That is, given the state of the process at any time, we are interested in the expected number of time steps we need in order to visit some other state of the process for the first time. A first contribution here is that we show, under some mild closedness and positivity assumptions on the local models, that any expected first-passage or return time is positive and finite. More importantly, we provide a recursive algorithm for computing exact bounds on such expected times. We also prove that the bounds on expected first-passage and return times are the same among different independence concepts and also among the different approaches used to obtain them, i.e. the martingale-theoretic and the measure-theoretic approaches.

In the final part of this dissertation, we apply our results to a queueing model. We first focus on the Geo/Geo/1/L queue, which stands for geometrically distributed interarrival and service times, with one server and maximum length L. We introduce an additional independence concept and examine the following queueing performance measures: expected queue length, expected first-passage times, expected return times, the probability of being in a state and the probability of "turning on the server", which means transitioning from state 0 to 1. For all these performance measures, we compute lower and upper bounds. Furthermore, we prove that the bounds on the expected queue length, the expected first-passage times, the expected return time to an empty queue, the expected return time to a full queue, the probability of having an empty queue and the probability of having a full queue, coincide no matter which type of independence we choose among the variables of the system. We also demonstrate that for the least strict independence concept, the expected value of the time average of a function on the state space can be more robust than the stationary marginal expected value.

SAMENVATTING

Dutch summary

Stochastische processen zijn systemen die in de tijd evolueren, en waarvan de specifieke tijdsevolutie onzeker is. Dit proefschrift brengt methoden aan voor het berekenen van robuuste gevolgtrekkingen in zulke stochastische processen, met een toepassing op het gebied van de wachtlijntheorie. Om de onzekerheid over de tijdsevolutie robuust te modelleren, gebruiken we de theorie van imprecieze waarschijnlijkheden, een uitbreiding van de waarschijnlijkheidsleer.

We beginnen dit proefschrift met enkele basisconcepten uit de theorie van imprecieze waarschijnlijkheden. Wanneer we onzeker zijn over de waarde van een veranderlijke, zal de waarschijnlijkheidsleer een waarschijnlijkheid voor elke waarde toewijzen door een enkele waarschijnlijkheidsmassafunctie op de verzameling van mogelijke waarden te definiëren, terwijl de theorie van imprecieze waarschijnlijkheden hiervoor een verzameling van (waarschijnlijkheids)massafuncties beschouwt. In de overgrote meerderheid van de gevallen wordt aangenomen dat zulke verzamelingen van massafuncties gesloten en convex zijn—ze worden dan credale verzamelingen genoemd. We tonen in een reeks voorbeelden hoe credale verzamelingen kunnen worden geconstrueerd en grafisch weergegeven. Daarnaast bespreken we ook gevallen waarin convexiteit wordt verlaten, zoals soms gebeurt in de wachtlijnmodellen die we verder in dit proefschrift bekijken. Vervolgens laten we zien hoe onzekerheid kan worden gemodelleerd met onderverwachtingswaarden, niet-lineaire operatoren die aan een aantal axioma's voldoen. Zo'n onderverwachtingswaarde-operator kan geassocieerd worden met een verzameling van massafuncties, en voor een gegeven functie van de onzekere veranderlijke kan haar onderverwachtingswaarde geïnterpreteerd worden als een exacte (of nauwe) ondergrens van de verzameling van haar mogelijke verwachtingswaarden. Alternatief kan ze ook worden gezien als een supremum aankoopprijs. Een functie van de onzekere veranderlijke wordt dan gezien als de onzekere numerieke uitkomst van een gok, die we bereid zijn te kopen voor elke prijs lager dan haar onderverwachtingswaarde. We bespreken ook hoe natuurlijke uitbreiding ons in staat stelt het domein van een gedeeltelijk gespecificeerde onderverwachtingswaarde-operator conservatief uit te breiden.

Vervolgens introduceren we een theorie van robuuste stochastische processen in discrete tijd. We beperken ons hierbij tot stochastische processen met een eindige toestandsruimte—de verzameling van mogelijke toestanden van het proces op elk willekeurig moment—en we beschouwen massafuncties voor de onzekere toestand van het proces op elk moment, conditioneel op de volledige geschiedenis van voorgaande toestanden. Deze conditionele waarschijnlijkheidsmassafuncties zijn wat we onze lokale modellen noemen. We gebruiken ze om andere, complexere conditionele of onconditionele waarschijnlijkheden voor het stochastisch proces af te leiden, en om (al dan niet conditionele) verwachtingswaarden te berekenen van functies die kunnen afhangen van de toestanden in een oneindig aantal tijdspunten—zulke functies zullen we veranderlijken noemen. Om dit te doen, ontlenen we elementen uit de maattheoretische waarschijnlijkheid, een deelveld van de maattheorie. Het verschil met de standaard maattheoretische aanpak, waar conditionele waarschijnlijkheden worden afgeleid uit onconditionele, is dat onze aanpak ze rechtstreeks uit de lokale modellen afleidt met behulp van zogeheten (volledige) conditionele waarschijnlijkheden en coherente conditionele waarschijnlijkheden, wat ons bovendien toestaat om conditionele waarschijnlijkheden te definiëren zelfs wanneer de gebeurtenis waarop we conditioneren waarschijnlijkheid nul heeft. Het robuuste aspect van onze aanpak komt hieruit voort dat we toelaten dat de lokale modellen gedeeltelijk gespecificeerd zijn, wat aanleiding geeft tot een verzameling van stochastische processen. Zo'n verzameling wordt ook wel een imprecies stochastisch proces genoemd. We ontwikkelen globale modellen op basis van de lokale modellen, door twee verschillende aanpakken uit te werken en met elkaar te vergelijken: (i) het toepassen van een maattheoretisch waarschijnlijkheidskader op de verzameling van stochastische processen; en (ii) een martingaaltheoretische aanpak. We zullen zien dat de maattheoretische aanpak ingewikkelder is en nogal veel aannames vereist, wat onze voorkeur voor de martingaaltheoretische aanpak verklaart. Dit leidt tot onze eerste vrij belangrijke bijdrage: een veralgemening van enkele gekende typische basiseigenschappen van stochastische processen naar een imprecies-probabilistische context: uitbreidingen van de axioma's waaraan onderverwachtingswaardeoperatoren voldoen, en een wet van herhaalde onderverwachtingswaarden. We veralgemenen bestaande resultaten ook door functies (veranderlijken) te beschouwen die afhangen van de waarden van de toestanden in een oneindig aantal tijdspunten, en door de veronderstelling te verlaten dat de lokale verzamelingen van waarschijnlijkheidsmassafuncties gesloten en convex moeten zijn. We vergelijken de globale modellen afgeleid binnen de maat- en de martingaaltheoretische aanpakken, en tonen aan dat op z'n minst voor functies die afhangen van een eindig aantal toestanden, de beide aanpakken samenvallen.

We gaan dan verder met een meer gedetailleerde analyse van een welbekende familie van stochastische processen: Markovketens in discrete tijd met een eindige toestandsruimte. Deze modellen zijn bijzonder makkelijk en efficiënt te gebruiken, omdat de lokale modellen niet conditioneel zijn op de volledige geschiedenis van toestanden, maar alleen op de laatste toestand. We onderscheiden tijdsinhomogene en tijdshomogene Markovketens, waarbij het verschil ligt in de al dan niet expliciete tijdsafhankelijkheid van de lokale modellen. Net zoals voor algemene imprecieze stochastische processen, maken we deze modellen robuuster door het mogelijk te maken onze lokale modellen slechts gedeeltelijk te specificeren. Op die manier krijgen we een verzameling van stochastische processen, ook wel imprecieze Markovketen genoemd. In tegenstelling tot de standaardveronderstelling dat de lokale modellen voor een imprecieze Markovketen gesloten en convexe verzamelingen van waarschijnlijkheidsmassafuncties zijn, bekijken wij in dit proefschrift opnieuw algemene verzamelingen van waarschijnlijkheidsmassafuncties. Bovendien bestuderen we gedetailleerd de gevolgen van het opleggen van verschillende soorten onafhankelijkheden tussen de toestanden van het proces: epistemische irrelevantie, complete onafhankelijkheid, en herhalingsonafhankelijkheid. Voor elk van die verschillende onafhankelijkheidsbegrippen bewijzen we verschillende eigenschappen, en bestuderen we hoe we—exacte en nauwe—grenzen op de verwachtingswaarden van relevante functies kunnen berekenen. In het bijzonder vinden we dat hoe minder veronderstellingen we over het onafhankelijkheidsconcept maken, hoe breder de grenzen op de verwachtingswaarden zijn, en hoe efficiënter de berekeningen. We geven aan dat de eerder bewezen algemene eigenschappen voor imprecieze stochastische processen ook hier geldig blijven, en we bespreken en bewijzen een aantal bijkomende eigenschappen.

Vervolgens richten we ons op een specifiek type Markovketens—de zogenoemde geboorte-en-doodketens in discrete tijd met een eindige toestandsruimte. Na een voorbereidende discussie definiëren we onze imprecieze versie van een geboorte-en-doodketen in discrete tijd als een bijzonder geval van een imprecieze Markovketen. Voor deze imprecieze geboorte-en-doodketens concentreren we ons op het probleem van het berekenen van verwachte eerste doorgangs- en terugkeertijden: gegeven de toestand van het proces op een bepaald moment, zijn we geïnteresseerd in het verwachte aantal tijdstappen nodig om voor het eerst een andere gegeven toestand van het proces te bezoeken. Onze eerste bijdrage hier is dat we onder enkele zwakke geslotenheids- en positiviteitsaannames op de lokale modellen laten zien dat elke verwachte eerste doorgangs- of terugkeertijd positief en eindig is. Een nog belangrijker bijdrage is een recursief algoritme voor het berekenen van exacte nauwe grenzen op zulke verwachte tijden. We bewijzen ook dat de grenzen op de verwachte eerste doorgangs- en de verwachte terugkeertijden dezelfde zijn onder de verschillende onafhankelijkheidsconcepten die we hierboven hebben genoemd, én binnen de maat- en martingaaltheoretische aanpakken.

In het laatste gedeelte van dit proefschrift passen we onze theoretische resultaten toe op een wachtlijnmodel. We richten ons op de Geo/Geo/1/L wachtlijn: een wachtlijn met geometrisch verdeelde aankomst en vertrek, met één bedieningsstation, en met maximale lengte L. We introduceren een extra onafhankelijkheidsconcept en onderzoeken een aantal wachtlijnprestatiematen: verwachte rijlengte, verwachte eerste doorgangtijden, verwachte terugkeertijden, de waarschijnlijkheid om een toestand te bereiken, en de waarschijnlijkheid om de server aan te zetten—wat een overgang van een lege naar een niet-lege toestand betekent. Voor al deze prestatiematen berekenen we onder- en bovengrenzen. Bovendien bewijzen we dat de grenzen op de verwachte rijlengte, de verwachte eerste doorgangtijden, de verwachte terugkeertijden naar een lege wachtlijn, de verwachte terugkeertijden tot een volle wachtlijn, de waarschijnlijkheid op een lege wachtlijn en de waarschijnlijkheid op een volle wachtlijn, samenvallen, ongeacht welk type onafhankelijkheid we tussen de toestanden van het systeem kiezen. We illustreren ook dat voor het minst strikte onafhankelijkheidsconcept de verwachtingswaarde van het tijdsgemiddelde van een functie van de toestand soms robuuster is dan de stationaire marginale verwachtingswaarde.

List of symbols

The list of symbols is ordered per category. Within each category, the symbols are ordered as they appear in the main text. The locations we provide correspond to the section where the symbols are located and/or to the page of their first use. Symbols that are only used locally are not included in the list.

Number sets

Symbol	Meaning	Location
$\mathbb{R}$	Set of real numbers	Page 36
$\mathbb{R}_{\geq 0}$	Set of non-negative real numbers	Page 44
$\mathbb{N}$	Set of natural numbers: $\{1,2,3,\dots\}$	Page 46
$\mathbb{N}_0$	Set of natural numbers with zero: $\mathbb{N}\cup\{0\}$	Page 46
$\overline{\mathbb{R}}$	Set of extended real numbers: $\mathbb{R}\cup\{-\infty,+\infty\}$	Page 235

Events, Sets and (Sets of) Functions

Symbol	Meaning	Location
X	Variable	Page 36
$\mathscr{X}$	State space: a non-empty finite set	Page 36 and Section 3.1$_{58}$
x	State value in $\mathscr{X}$	Sections 2.1$_{36}$ and 3.1$_{58}$
$2^{\mathscr{X}}$	Power set of $\mathscr{X}$	Section 2.1$_{36}$
$\mathscr{L}(\mathscr{X})$	Set of all real-valued functions on $\mathscr{X}$	Section 2.2$_{38}$
$\mathbb{I}_x$	Indicator of x, for $x\in\mathscr{X}$	Section 2.2$_{38}$
$\mathbb{I}_A$	Indicator of A, for $A\in 2^{\mathscr{X}}$	Section 2.2$_{38}$

f	Gamble: function in $\mathscr{L}(\mathscr{X})$	Section 2.2$_{38}$
$\mathscr{K}$	Subset of $\mathscr{L}(\mathscr{X})$	Section 2.6$_{50}$
X_n	Variable at time n	Section 3.1$_{58}$
$\{X_n\}_{n\in\mathbb{N}}$	Infinite sequence of variables	Section 3.1$_{58}$
Ω	Sample space $\mathscr{X}^{\mathbb{N}}$	Section 3.1$_{58}$
ω	Path: generic element of Ω	Section 3.1$_{58}$
x_n	State value at time n	Section 3.1$_{58}$
$X_{1:n}$	Sequence of variables from time 1 up to and including n	Section 3.1$_{58}$
$X_{m:n}$	Sequence of variables from time m up to and including n	Section 3.1$_{58}$
$x_{1:n}$	Situation: finite sequence of state values from time 1 up to and including n	Section 3.1$_{58}$
$x_{m+1:n}$	Finite sequence of state values from time $m+1$ up to and including n	Section 3.1$_{58}$
$\square$	Initial situation: situation at time 0	Section 3.1$_{58}$
$\mathscr{X}^*$	Set of all situations	Section 3.1$_{58}$
ω^n	First n state values of ω	Section 3.1$_{58}$
ω_n	State value of ω at time n	Section 3.1$_{58}$
$\Gamma(x_{1:n})$	All paths of which the first n state values are equal to $x_{1:n}$	Section 3.2$_{59}$
$\langle\mathscr{X}^*\rangle$	Algebra generated by the set of all situations	Section 3.2$_{59}$
$\sigma(\mathscr{X}^*)$	σ-Algebra generated by the set of all situations	Section 3.2$_{59}$
2^Ω	Set of all events in Ω: power set of Ω	Section 3.4.1$_{61}$
$2^\Omega_\varnothing$	Set of all events in Ω without the empty set: $2^\Omega\setminus\{\varnothing\}$	Section 3.4.1$_{61}$
A	Event: element of 2^Ω	Section 3.4.1$_{61}$
B	Conditioning event: element of $2^\Omega_\varnothing$	Section 3.4.1$_{61}$
$2^\Omega\times 2^\Omega_\varnothing$	Set of all conditional events	Section 3.4.1$_{61}$
$A\vert B$	Conditional event: element of $2^\Omega\times 2^\Omega_\varnothing$	Section 3.4.1$_{61}$
$\mathscr{C}_\sigma$	Domain of conditional events defined by Equation (3.2)$_{64}$	Section 3.4.2$_{64}$
$\mathscr{C}_{\mathscr{X}^*}$	Domain of conditional events defined by Equation (3.4)$_{65}$	Section 3.4.2$_{64}$

$\mathscr{C}$	Domain of conditional events defined by Equation (3.9)$_{67}$	Section 3.4.2$_{64}$
$\mathscr{C}^*$	Domain of conditional events defined by Equation (3.14)$_{70}$	Section 3.4.3$_{68}$
$\mathscr{C}^*_\sigma$	Domain of conditional events defined by Equation (3.15)$_{71}$	Section 3.4.3$_{68}$
$\mathscr{L}(\mathscr{X}^n)$	Set of all real-valued functions on $\mathscr{X}^n$	Section 3.5.1$_{72}$
h	Function in $\mathscr{L}(\mathscr{X}^n)$	Section 3.5.1$_{72}$
$h(X_{1:n})$	Real-valued n-measurable function	Section 3.5.1$_{72}$
s, t, u, v	Generic situations in $\mathscr{X}^*$	Section 4.1$_{87}$
$s\bullet$	$\Gamma(s)$: All paths that go through situation s	Section 4.1$_{87}$
$\mathscr{F}$	σ-Algebra on Ω	Section A.1$_{234}$
$\mathscr{F}_0$	Algebra on Ω	Section A.1$_{234}$
$\mathbb{B}_{\overline{\mathbb{R}}}$	Borel σ-algebra: the σ-algebra generated by all open subsets of $\mathbb{R}$	Section A.1$_{234}$
$(\Omega, \mathscr{F})$	Measurable space	Section A.2$_{236}$
$(\Omega, \mathscr{F}, P)$	Probability space	Section A.3$_{237}$
g	Extended real-valued function on Ω	Section A.3$_{237}$

Uncertainty models

Symbol	Meaning	Location
$p(\cdot)$	Probability mass function on $\mathscr{X}$	Section 2.1$_{36}$
$\Sigma_{\mathscr{X}}$	Set of all probability mass functions on $\mathscr{X}$	Section 2.1$_{36}$
P	Probability measure	Sections 2.1$_{36}$ and A.2$_{236}$
$P(A)$	Probability of the event A	Section 2.1$_{36}$ and Page 72
$E(\cdot)$	Expectation operator	Section 2.2$_{38}$
$E_p(\cdot)$	Expectation operator that corresponds to a probability mass function p	Section 2.2$_{38}$

$\mathscr{P}$	Non-empty subset of $\Sigma_{\mathscr{X}}$	Section 2.3$_{40}$
$\underline{p}(x)$	Lower probability mass of x	Section 2.3$_{40}$
$\overline{p}(x)$	Upper probability mass of x	Section 2.3$_{40}$
Φ	Credal set: closed and convex set of probability mass functions	Section 2.3$_{40}$
$\underline{E}$	Lower expectation operator	Section 2.4$_{44}$
$\overline{E}$	Upper expectation operator	Section 2.4$_{44}$
$\underline{E}(f)$	Lower expectation of f	Section 2.4$_{44}$
$\overline{E}(f)$	Upper expectation of f	Section 2.4$_{44}$
$\underline{E}_{\mathscr{P}}$	Lower expectation operator associated with the set of probability mass functions $\mathscr{P}$	Section 2.4$_{44}$
$\overline{E}_{\mathscr{P}}$	Upper expectation operator associated with the set of probability mass functions $\mathscr{P}$	Section 2.4$_{44}$
$\Phi_{\underline{E}}$	Closed and convex set of probability mass functions corresponding to $\underline{E}$	Section 2.4$_{44}$
$\underline{\mathscr{E}}(\cdot)$	Natural extension of a lower expectation assessment	Section 2.6$_{50}$
p	Probability tree: a function from $\mathscr{X} \times \mathscr{X}^*$ to $[0,1]$	Section 3.3$_{61}$
$p(\cdot \vert x_{1:n})$	Transition model associated with situation $x_{1:n}$: probability mass function on $\mathscr{X}$ conditional on $x_{1:n}$	Section 3.3$_{61}$
$p(X_1)$	Initial model: probability mass function on $\mathscr{X}$ conditional on $\square$	Section 3.3$_{61}$
$\mathbb{P}_{\mathscr{X}^*}$	Set of all probability trees	Section 3.3$_{61}$
$P(A \vert B)$	Probability of the event A conditional on the event B	Section 3.4.1$_{61}$
$\mathbb{P}_p$	Set of conditional probability measures on the domain $\mathscr{C}_\sigma$	Section 3.5$_{72}$
$E_P(g)$	Expectation of g with respect to P	Sections 3.5.1$_{72}$ and A.4$_{239}$
$E_P(g \vert B)$	Expectation of g with respect to P conditional on the event B	Section 3.5.2$_{74}$
$\mathscr{T}$	Imprecise probability tree: non-empty subset of $\mathbb{P}_{\mathscr{X}^*}$	Section 3.6$_{79}$
$\mathbb{P}_{\mathscr{T}}$	The set of conditional probability measures on $\mathscr{C}_\sigma$ corresponding to $\mathscr{T}$	Section 3.6$_{79}$

$\underline{E}_{\mathscr{T}}(g\|B)$	Lower expectation of g with respect to $\mathscr{T}$ conditional on the event B	Section 3.6$_{79}$
$\overline{E}_{\mathscr{T}}(g\|B)$	Upper expectation of g with respect to $\mathscr{T}$ conditional on the event B	Section 3.6$_{79}$
$\mathscr{P}_{x_{1:n}}$	Non-empty set of conditional probability mass functions associated with $x_{1:n}$	Section 3.7$_{80}$
$\mathscr{P}_{\Box}$	Non-empty set of conditional probability mass functions associated with $\Box$	Section 3.7$_{80}$
$\mathscr{T}_{\mathscr{P}}$	Imprecise probability tree constructed from the sets $\mathscr{P}_{x_{1:n}}$	Section 3.7$_{80}$
$\mathbb{P}_{\mathscr{P}}$	The set of conditional probability measures on $\mathscr{C}_{\sigma}$ corresponding to $\mathscr{T}_{\mathscr{P}}$	Section 3.7$_{80}$
$\underline{Q}(\cdot\|x_{1:n})$	Lower expectation operator associated with the set $\mathscr{P}_{x_{1:n}}$	Section 3.7$_{80}$
$\overline{Q}(\cdot\|x_{1:n})$	Upper expectation operator associated with the set $\mathscr{P}_{x_{1:n}}$	Section 3.7$_{80}$
$\underline{Q}_{\Box}(\cdot)$	Lower expectation operator associated with the set $\mathscr{P}_{\Box}$	Section 3.7$_{80}$
$\overline{Q}_{\Box}(\cdot)$	Upper expectation operator associated with the set $\mathscr{P}_{\Box}$	Section 3.7$_{80}$
$\underline{E}_{\mathscr{P}}(\cdot\|x_{1:m})$	Lower expectation with respect to $\mathbb{P}_{\mathscr{P}}$ conditional on $x_{1:m}$	Section 3.7$_{80}$
$\overline{E}_{\mathscr{P}}(\cdot\|x_{1:m})$	Upper expectation with respect to $\mathbb{P}_{\mathscr{P}}$ conditional on $x_{1:m}$	Section 3.7$_{80}$
$\mathscr{U}$	Process: real-valued function on $\mathscr{X}^*$	Section 4.2$_{87}$
$\Delta\mathscr{U}$	Process difference	Section 4.2$_{87}$
$\mathscr{M}$	Submartingale: process that satisfies Equation (4.1)$_{87}$	Section 4.2$_{87}$
$\mathscr{M}$	Supermartingale: process that satisfies Equation (4.2)$_{87}$	Section 4.2$_{87}$
$\underline{\mathbb{M}}$	Set of all uniformly bounded above submartingales	Section 4.2$_{87}$
$\overline{\mathbb{M}}$	Set of all uniformly bounded below supermartingales	Section 4.2$_{87}$
$\underline{E}_{\underline{Q}}(g\|s)$	Lower expectation of g conditional on s based on submartingales	Section 4.3.1$_{88}$
$\overline{E}_{\underline{Q}}(g\|s)$	Upper expectation of g conditional on s based on supermartingales	Section 4.3.1$_{88}$

(IMPRECISE) MARKOV CHAINS

Symbol	Meaning	Location
$q_n(\cdot\|x_n)$	Transition model of Markov chain associated with state value x_n at time n	Section 5.1.1$_{101}$
$q_\square(X_1)$	Initial model of Markov chain	Section 5.1.1$_{101}$
$\mathbb{P}_{\mathrm{M}}$	Set of all probability trees corresponding to Markov chains	Section 5.1.1$_{101}$
$f(X_n)$	Real-valued function that depends only on the state at time n	Section 5.1.2$_{105}$
T_n	Transition operator of Markov chain at time n	Section 5.1.2$_{105}$
$[f](X_{1:n})$	Time average: $\frac{1}{n}\sum_{i=1}^{n} f(X_i)$	Section 5.1.3$_{106}$
$\xi_m^n(\cdot,\cdot)$	Function defined by Equation (5.9)$_{106}$	Section 5.1.3$_{106}$
$q(\cdot\|x)$	Transition model of homogeneous Markov chain associated with state value x	Section 5.2.1$_{109}$
$\mathbb{P}_{\mathrm{HM}}$	Set of all probability trees corresponding to homogeneous Markov chains	Section 5.2.1$_{109}$
M	Transition matrix of homogeneous Markov chain	Section 5.2.1$_{109}$
T	Transition operator of homogeneous Markov chain	Section 5.2.2$_{110}$
E_∞	Limit expectation operator of ergodic homogeneous Markov chain	Section 5.2.2$_{110}$
$\mathscr{Q}_{n,x}$	Set of conditional probability mass functions associated with state value x at time n	Section 5.3$_{113}$
$\mathscr{Q}_\square$	Set of conditional probability mass functions associated with the initial situation	Section 5.3$_{113}$
$\underline{Q}_n(\cdot\|x)$	Lower expectation operator corresponding to $\mathscr{Q}_{n,x}$	Section 5.3$_{113}$
$\overline{Q}_n(\cdot\|x)$	Upper expectation operator corresponding to $\mathscr{Q}_{n,x}$	Section 5.3$_{113}$
$\underline{Q}_\square(\cdot)$	Lower expectation operator corresponding to $\mathscr{Q}_\square$	Section 5.3$_{113}$
$\overline{Q}_\square(\cdot)$	Upper expectation operator corresponding to $\mathscr{Q}_\square$	Section 5.3$_{113}$

$\mathscr{Q}_x$	Set of conditional probability mass functions associated with state value x	Section 5.3$_{113}$
$\underline{Q}(\cdot\|x)$	Lower expectation operator corresponding to $\mathscr{Q}_x$	Section 5.3$_{113}$
$\overline{Q}(\cdot\|x)$	Upper expectation operator corresponding to $\mathscr{Q}_x$	Section 5.3$_{113}$
$\mathscr{T}_{\mathscr{Q}}$	Imprecise probability tree constructed from the sets $\mathscr{Q}_{n,x}$ and $\mathscr{Q}_{\square}$	Section 5.4.1$_{115}$
$\mathbb{P}_{\mathscr{Q}}$	Set of conditional probability measures on $\mathscr{C}_\sigma$ corresponding to $\mathscr{T}_{\mathscr{Q}}$	Section 5.4.1$_{115}$
$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g\|B)$	Lower expectation of g with respect to $\mathbb{P}_{\mathscr{Q}}$ conditional on the event B in imprecise Markov chain under epistemic irrelevance	Section 5.4.1$_{115}$
$\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g\|B)$	Upper expectation of g with respect to $\mathbb{P}_{\mathscr{Q}}$ conditional on the event B in imprecise Markov chain under epistemic irrelevance	Section 5.4.1$_{115}$
$\underline{T}_n$	Lower transition operator of imprecise Markov chain	Section 5.4.2$_{116}$
$\overline{T}_n$	Upper transition operator of imprecise Markov chain	Section 5.4.2$_{116}$
$\underline{\xi}^n_m(\cdot,\cdot)$	Function defined by Equation (5.34)$_{118}$	Section 5.4.3$_{118}$
$\overline{\xi}^n_m(\cdot,\cdot)$	Function defined by Equation (5.35)$_{118}$	Section 5.4.3$_{118}$
$\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g\|s)$	Martingale-theoretic lower expectation of g conditional on situation s in imprecise Markov chain under epistemic irrelevance	Section 5.4.5$_{124}$
$\overline{E}^{\mathrm{ei}}_{\underline{Q}}(g\|s)$	Martingale-theoretic upper expectation of g conditional on situation s in imprecise Markov chain under epistemic irrelevance	Section 5.4.5$_{124}$
$\underline{E}^{\mathrm{ei}}_{\underline{Q}\|m+1}(g\|x)$	Martingale-theoretic lower expectation of a function g that does not depend on $X_{1:m}$ conditional on x in imprecise Markov chain under epistemic irrelevance	Section 5.4.5$_{124}$
$\overline{E}^{\mathrm{ei}}_{\underline{Q}\|m+1}(g\|x)$	Martingale-theoretic upper expectation of a function g that does not depend on $X_{1:m}$ conditional on x in imprecise Markov chain under epistemic irrelevance	Section 5.4.5$_{124}$

$\mathscr{T}_{\mathscr{Q}}^{\mathrm{M}}$	Set of probability trees corresponding to Markov chains constructed from the sets $\mathscr{Q}_{n,x}$ and $\mathscr{Q}_{\square}$	Section 5.6.1$_{131}$
$\mathbb{P}_{\mathscr{Q}}^{\mathrm{M}}$	Set of conditional probability measures on $\mathscr{C}_{\sigma}$ corresponding to $\mathscr{T}_{\mathscr{Q}}^{\mathrm{M}}$	Section 5.6.1$_{131}$
$\underline{E}_{\mathscr{Q}}^{\mathrm{ci}}(g\|B)$	Lower expectation of g with respect to $\mathbb{P}_{\mathscr{Q}}^{\mathrm{M}}$ conditional on the event B in imprecise Markov chain under complete independence	Section 5.6.1$_{131}$
$\overline{E}_{\mathscr{Q}}^{\mathrm{ci}}(g\|B)$	Upper expectation of g with respect to $\mathbb{P}_{\mathscr{Q}}^{\mathrm{M}}$ conditional on the event B in imprecise Markov chain under complete independence	Section 5.6.1$_{131}$
$\mathscr{T}_{\mathscr{Q}}^{\mathrm{HM}}$	Set of probability trees corresponding to homogeneous Markov chains constructed from the sets $\mathscr{Q}_{x}$ and $\mathscr{Q}_{\square}$	Section 5.7.1$_{142}$
$\mathbb{P}_{\mathscr{Q}}^{\mathrm{HM}}$	Set of conditional probability measures on $\mathscr{C}_{\sigma}$ corresponding to $\mathscr{T}_{\mathscr{Q}}^{\mathrm{HM}}$	Section 5.7.1$_{142}$
$\underline{E}_{\mathscr{Q}}^{\mathrm{ri}}(g\|B)$	Lower expectation of g with respect to $\mathbb{P}_{\mathscr{Q}}^{\mathrm{HM}}$ conditional on the event B in homogeneous imprecise Markov chain under repetition independence	Section 5.7.1$_{142}$
$\overline{E}_{\mathscr{Q}}^{\mathrm{ri}}(g\|B)$	Upper expectation of g with respect to $\mathbb{P}_{\mathscr{Q}}^{\mathrm{HM}}$ conditional on the event B in homogeneous imprecise Markov chain under repetition independence	Section 5.7.1$_{142}$

(IMPRECISE) BIRTH-DEATH CHAINS

Symbol	Meaning	Location
$\mathscr{X}_m$	Set $\{\ell, e, u\}$	Section 6.2$_{152}$
$\mathscr{X}_0$	Set $\{e, u\}$	Section 6.2$_{152}$
$\mathscr{X}_L$	Set $\{\ell, e\}$	Section 6.2$_{152}$
Φ_i	Set of probability mass functions on $\mathscr{X}_m$	Section 6.2$_{152}$
Φ_0	Set of probability mass functions on $\mathscr{X}_0$	Section 6.2$_{152}$
Φ_L	Set of probability mass functions on $\mathscr{X}_L$	Section 6.2$_{152}$

π_i	Probability mass functions in Φ_i	Section 6.2$_{152}$
π_0	Probability mass functions in Φ_0	Section 6.2$_{152}$
π_L	Probability mass functions in Φ_L	Section 6.2$_{152}$
(b_i, r_i, w_i)	Notation for $(\pi_i(\ell), \pi_i(e), \pi_i(u))$	Section 6.2$_{152}$
(r_0, w_0)	Notation for $(\pi_0(e), \pi_0(u))$	Section 6.2$_{152}$
(b_L, r_L)	Notation for $(\pi_L(\ell), \pi_L(e))$	Section 6.2$_{152}$
$\tau^n_{i\to j}$	First-passage time of j conditional on i at time n	Section 6.3$_{155}$
$\underline{\tau}_{i\to j}$	Martingale-theoretic lower expected first-passage time from i to j	Section 6.3$_{155}$
$\overline{\tau}_{i\to j}$	Martingale-theoretic upper expected first-passage time from i to j	Section 6.3$_{155}$
$\tau^M_{i\to j}$	Expected first-passage time from i to j in a birth-chain with transition matrix M according to the martingale-theoretic approach	Section 6.7.1$_{168}$
$E_P(\tau^n_{i\to j}\|X_n = i)$	Expected first-passage time from i to j in a birth-chain with probability measure P	Section 6.8$_{177}$
$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}\|X_n = i)$	Lower expected first-passage time from i to j with respect to $\mathbb{P}_{\mathscr{Q}}$ in an imprecise birth-death chain under epistemic irrelevance	Section 6.8$_{177}$
$\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}\|X_n = i)$	Upper expected first-passage time from i to j with respect to $\mathbb{P}_{\mathscr{Q}}$ in an imprecise birth-death chain under epistemic irrelevance	Section 6.8$_{177}$
$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to j}\|X_n = i)$	Lower expected first-passage time from i to j with respect to $\mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}$ in an imprecise birth-death chain under complete independence	Section 6.8$_{177}$
$\overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to j}\|X_n = i)$	Upper expected first-passage time from i to j with respect to $\mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}$ in an imprecise birth-death chain under complete independence	Section 6.8$_{177}$

$\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^{n}_{i\to j}\vert X_n=i)$	Lower expected first-passage time from i to j with respect to $\mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}}$ in an imprecise birth-death chain under repetition independence	Section 6.8$_{177}$
$\overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^{n}_{i\to j}\vert X_n=i)$	Upper expected first-passage time from i to j with respect to $\mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}}$ in an imprecise birth-death chain under repetition independence	Section 6.8$_{177}$

QUEUEING

Symbol	Meaning	Location
a	Probability of arrival in Geo/Geo/1/L queue	Section 7.2.1$_{191}$
d	Probability of departure in Geo/Geo/1/L queue	Section 7.2.1$_{191}$
$q_{\square}$	Initial model of Geo/Geo/1/L queue	Section 7.2.1$_{191}$
$q_{a,d}$	Probability tree of Geo/Geo/1/L queue with parameters a, d and $q_{\square}$	Section 7.2.1$_{191}$
$P(X_n=k)$	Probability of queue length k at time n	Section 7.2.1$_{191}$
$P(X=k)$	Probability of queue length k in the limit	Section 7.2.1$_{191}$
$[\underline{a},\overline{a}]$	Interval for arrival probability	Section 7.2.2$_{195}$
$[\underline{d},\overline{d}]$	Interval for departure probability	Section 7.2.2$_{195}$
$\mathscr{Q}_{\square}$	Initial model of imprecise Geo/Geo/1/L queue	Section 7.2.2$_{195}$
$\mathscr{T}^{\mathrm{O}}_{\mathscr{Q}}$	Set of probability trees of the form $q_{a,d}$, with $a\in[\underline{a},\overline{a}]$, $d\in[\underline{d},\overline{d}]$ and $q_{\square}\in\mathscr{Q}_{\square}$	Section 7.3$_{197}$
$\mathbb{P}^{\mathrm{O}}_{\mathscr{Q}}$	Set of conditional probability measures on $\mathscr{C}_{\sigma}$ corresponding to $\mathscr{T}^{\mathrm{O}}_{\mathscr{Q}}$	Section 7.3$_{197}$
$\underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(g\vert B)$	Lower expectation of g with respect to $\mathbb{P}^{\mathrm{O}}_{\mathscr{Q}}$ conditional on the event B in imprecise Geo/Geo/1/L queue under fixed-parameter repetition independence	Section 7.3$_{197}$
$\overline{E}^{\mathrm{fi}}_{\mathscr{Q}}(g\vert B)$	Upper expectation of g with respect to $\mathbb{P}^{\mathrm{O}}_{\mathscr{Q}}$ conditional on the event B in imprecise Geo/Geo/1/L queue under fixed-parameter repetition independence	Section 7.3$_{197}$

X_n	Queue length at time n	Section 7.5$_{202}$
$\frac{1}{n}\sum_{i=1}^{n} X_i$	Average queue length up to time n	Section 7.5.2$_{205}$
$\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1})$	Function for turning on the server at time $n+1$	Section 7.7$_{210}$
$\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1})$	Average of function for turning on the server up to time $n+1$	Section 7.7$_{210}$
$\underline{\psi}_k(\cdot)$	Function defined by Equation (7.29)$_{214}$	Section 7.7$_{210}$
$\overline{\psi}_k(\cdot)$	Function defined by Equation (7.30)$_{214}$	Section 7.7$_{210}$
$\underline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{i\to j} \vert X_n = i)$	Lower expected first-passage time from i to j with respect to $\mathbb{P}^{\text{O}}_{\mathscr{Q}}$ in an imprecise Geo/Geo/1/L queue under fixed-parameter repetition independence	Section 7.8.1$_{220}$
$\overline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{i\to j} \vert X_n = i)$	Upper expected first-passage time from i to j with respect to $\mathbb{P}^{\text{O}}_{\mathscr{Q}}$ in an imprecise Geo/Geo/1/L queue under fixed-parameter repetition independence	Section 7.8.1$_{220}$

VARIOUS OTHER OPERATORS

Symbol	Meaning	Location
$\text{ext}(\cdot)$	Operator that returns the extreme points of a set	Page 43
$\text{conv}(\cdot)$	Operator that returns the smallest convex set	Section 2.5$_{47}$
θ	Shift operator on $\mathbb{N}$ or on Ω or on functions g on Ω	Section 5.5.2$_{129}$
$\sigma(\cdot)$	Operator that generates a sigma algebra	Section A.1$_{234}$

1

INTRODUCTION

The theory of discrete-time stochastic processes can be extended by replacing the local probability mass functions of a stochastic process with sets of them. This yields a set of stochastic processes, which we call an imprecise stochastic process, and the purpose is to use this set in order to make robust inferences about systems whose time evolution is uncertain. This is a more reliable approach towards modelling uncertainty as it can handle cases where we only have partial knowledge about our uncertainty model, for example because the model has to be learnt from small amounts of data. It moreover provides a robust output instead of insisting on a single output that might be much less reliable.

In this dissertation, we study the theory of imprecise stochastic processes that is based on the theory of imprecise probabilities, focusing mainly on so-called imprecise Markov chains, and we develop efficient methods for computing tight bounds on various types of expectations. We start in this introduction by presenting the main ideas and concepts that will appear, and we also motivate our choice of using different mathematical frameworks for the derivation of the results further on in this dissertation. This introduction also provides some useful information about the internal and external references, a brief overview of each of the chapters, and a list of the main publications that led to this dissertation.

1.1 GENERAL DESCRIPTION OF THE IDEAS, MOTIVATION AND MAIN CONTRIBUTION

A stochastic process [31, 44] is a collection of variables that represents a system whose time evolution is uncertain. It is a popular model in various scientific fields, including biology, physics, queueing theory, telecommunications, engineering and economics. In general, the variables of a stochastic process are indexed by the real or the natural numbers, representing time. In the former case, we say that the stochastic process is modelled in continuous time, whereas in the latter case it is modelled in discrete time. In this is dissertation, we deal with discrete-time stochastic processes whose variables take values in a finite state space. The evolution of such a stochastic process depends on its parameters, which can be derived from data or from opinions of domain experts, or from a combination of both. These parameters are local probability mass functions that generate a (conditional) probability measure [45] on the possible time evolutions of the system. The goal is to use this probability measure to calculate probabilities of events, and expectations of functions, that are of interest to us.

This is easily illustrated by considering the following simple example, which is a stochastic process that is typically considered in the field of queueing theory. Suppose that we have a queue in a bank and that at predetermined times a customer can enter the queue with some arrival probability and a customer is serviced—in other words, leaves the queue—with some departure probability. Furthermore, let us assume that the arrival of a customer is independent of the service of any other customer and that the queue has a limited maximum capacity. We then have a discrete-time stochastic process, where at each time point we have a variable representing the number of customers that are in the queue. Suppose now that we observe the queue at a some time point and we see that is empty, then an interesting question that might be addressed would be the expected time till the queue becomes full for the first time. Other interesting questions are the expected (average) number of customers in the queue, the probability of the queue being full and the probability of the queue being empty.

In general, once we specify the parameters of a stochastic process, we are able to calculate various probabilities and expectations that give us insight in the behaviour of the system. For instance, in the example above, once we specify the probabilities of arrival and departure, we are able to answer the questions that we asked there. This makes stochastic processes useful and successful tools for making inferences in different fields. However, specifying the parameters of a stochastic process exactly is often unrealistic. In practice, we often only have partial knowledge about these parameters, for example because they are derived from small amounts of—possibly unreliable—data or because they are elicited from—possibly disagreeing—domain experts. For these reasons, there appears to be a need to allow for imprecise parameters, that is, to replace exactly specified parameters by sets/intervals of them. For instance, in the example mentioned before, this would correspond to a situation where the probability of an arriving customer and the probability of a departing customer are not specified precisely, but are only known to belong to certain intervals. Such an imprecise queueing system will be examined in Chapter 7$_{189}$

and we will also make inferences for them, by addressing questions like the ones we mentioned earlier.

An important aspect here is that imprecise parameters lead to imprecise inferences. Generally speaking, since we have sets of parameters, we can calculate the output of the process for each selection of parameters, and the aim is to find the tightest—if possible—bounds on this output. In this way, we are able to predict the range of outcomes that might arise, as a result of our imprecision about the parameters of the model. For instance, in the example mentioned before, instead of obtaining an expected number of customers, we would obtain an interval for this expectation. Furthermore, this interval-valued output would be robust with respect to our partial knowledge about the parameters of the model.

In order to achieve this type of robustness in the output of a stochastic process, we apply the framework of imprecise probabilities [5, 78] on stochastic processes. More specifically, instead of single parameters we give as input sets of them, and we provide bounds on the output by means of lower and upper expectations [78], which are basic tools for modelling uncertainty within the theory of imprecise probabilities. An important feature is that imprecise probabilities generalise probability theory, in the sense that if the input parameters were to be single values, then the two theories would coincide. The main advantage of imprecise probabilities is that they allow us to investigate how "sensitive", i.e. how robust, the output is when we vary the input parameters.

Among the many different types of stochastic processes that exist, we focus on so-called Markov chains [43,56,70], which are a special type of probability trees [40,64]. Over the years, these Markov chains have been applied to numerous scientific fields, in part because they have the advantage that they require only a limited number of input parameters, known as local models. Moreover, Markov chains offer an accurate representation of the behaviour of different types of systems, including queueing systems. When imprecision is incorporated into Markov chains and probability trees, we obtain what are called imprecise Markov chains [28, 37, 47, 68] and imprecise probability trees [5, 26]. The main difference between Markov chains and imprecise Markov chains is that the local models of an imprecise Markov chain are sets of probability mass functions. Another difference is that in imprecise Markov chains, we can choose among different types of independence, namely epistemic irrelevance [27], complete independence [15, 62] and repetition independence [14]. The choice of independence affects the interval-valued output, but also the computational methods used for obtaining this output. Although significant work has already been done on imprecise Markov chains, as well as on general imprecise stochastic processes, severe limitations are often imposed on the types of functions and local models that can be considered.

Our main contribution is the development of computational methods for imprecise Markov chains, for more general—than real-valued—types of functions, and more general—than closed and convex—types of local models.

Since we model uncertainty using lower and upper expectations, one of the most basic properties for their computation is the imprecise version of the law of iterated expectations. This property was already proved in Reference [27, Theorem 3.2] for real-valued functions that depend on a finite number of variables and for imprecise Markov chains under epistemic irrelevance whose local models are closed and convex sets of probability mass functions. In this dissertation, we prove multiple generalisations of it. First of all, we generalise it to imprecise stochastic processes—including imprecise Markov chains under epistemic irrelevance—whose local models are general sets of probability mass functions. We also use this generalisation to show how we can efficiently compute lower and upper expectations for a class of functions that we call time averages, and in addition prove various interesting properties for them. Secondly, we generalise the imprecise law of iterated expectations to functions that depend on an infinite number of variables, using a martingale-theoretic approach that is based on sub- and supermartingales. This approach has multiple advantages over the standard measure-theoretic approach, and we prove that it coincides with the latter when it comes to functions that depend on a finite number of variables.

We also define a special class of imprecise Markov chains, which we call imprecise birth-death chains. These processes have the structure of classical birth-death chains, but their local models are no longer single probability mass functions, but sets of them. For this class of processes, we focus on the computation of lower and upper expected first-passage and return times. In particular, we show that these lower and upper expectations satisfy a system of non-linear equations that can be efficiently solved in a simple and recursive way. Furthermore, we prove that it makes no difference whether we define lower and upper expected first-passage and return times using the martingale-theoretic or the measure-theoretic approach and we show that our results apply to all types of independence.

Finally, we apply our findings to an imprecise version of the Geo/Geo/1/L queueing model, which is obtained by adding imprecision to the parameters of the classical Geo/Geo/1/L queue and which is a special case of an imprecise birth-death chain. For this model, we also introduce a new type of independence, which we call fixed-parameter repetition independence, and we calculate bounds on the expectations of various performance measures that are commonly used in queueing theory. More specifically, we calculate the expected (average) queue length, the (average) probability of each queue length, the (average) probability of "turning on the server" and expected first-passage and return times, we show similarities and differences between the bounds on these expectations as we consider different types of independence concepts, and we prove a number of properties for them.

1.2 INFORMATION ABOUT REFERENCES

The references in this dissertation are both internal and external. The external references are bibliographic ones; they are enumerated near the end of this dissertation, and we refer to them using numbers in square brackets. For example, details about the theory of imprecise probabilities can be found in Walley's seminal book [78], as well as in more recent textbooks [5, 72]. The internal references are used for chapters, sections, subsections, appendices, equations, theorems, lemmas, corollaries, propositions and assumptions. We refer to them using a subscript that indicates the page on which they are presented, unless they are on the same page, in which case there is no subscript. For instance, Theorem 58_{156} is located on page 156. In the special case where the reference is on the previous page we use the symbol $\curvearrowleft$ and if it is on the next page we use the symbol $\curvearrowright$.

1.3 OVERVIEW OF THE CHAPTERS

This dissertation consists of eight chapters and one appendix. Apart form this introduction, Chapters 2_{36}–7_{189} present the theory and the results from which this dissertation was constructed. Chapter 8_{230} presents our conclusions and our ideas for further research, and Appendix A_{233} provides some basic information about measure-theoretic probability, which is needed for our results in Chapter 3_{57}. We now give a brief overview of the content of Chapters 2_{36}–7_{189}.

In Chapter 2_{36}, we begin by presenting some basic information about probability theory and we introduce the concept of a probability mass function. We then talk about sets of probability mass functions in order to establish the connection between classical probability theory and the theory of imprecise probabilities. The theory of imprecise probabilities often assumes closed and convex sets of probability mass functions, but we also consider cases where these assumptions are dropped. Afterwards, we present the concept of lower (and upper) expectations, which is our main tool for modelling uncertainty in this dissertation, and we prove some interesting properties for them when they are derived from a special class of closed and—possibly—non-convex sets of probability mass functions. We also show how we can construct lower and upper expectations from partial (probability) assessments, something that is called natural extension in the theory of imprecise probabilities, because our main goal in the rest of this dissertation is to build global models based on local assessments.

Chapter 3_{57} is about (imprecise) stochastic processes. We first introduce the concept of event trees, from which we jump to probability trees. Probability trees are defined through conditional probability mass functions and we construct our stochastic processes using these probability trees. We show how to derive unique conditional probability measures on algebras and σ-algebras of the sample space of a stochastic process based on the definition of a (coherent) conditional probability. Moreover, we show how we can compute expectations using the derived conditional probability measures, and we prove some properties for these expectations that will turn out to be useful for imprecise Markov chains. We end the chapter by defining imprecise stochastic processes, for which we prove a generalised version of the law of iterated expectations, for functions that depend on a finite number of state variables, and in the general case where the local models of the imprecise stochastic process are arbitrary sets of probability mass functions.

In Chapter 4_{86}, we define global lower and upper expectations based on the concept of sub- and supermartingales. This martingale-theoretic approach allows us to compute lower and upper expectations of functions that depend on an infinite number of state variables, which in addition need not be measurable. We show that these lower and upper expectations satisfy various useful properties such as, again, a generalised version of the law of iterated expectations. Furthermore, we establish the connection between the lower and upper expectations defined by the martingale-theoretic approach and the ones defined by the measure-theoretic approach, as discussed in Chapter 3_{57}, and we prove that they coincide for functions that depend on a finite number of state variables.

Next in this dissertation is our work on (imprecise) Markov chains, which can be found in Chapter 5_{100}. We start by giving a general description of Markov chains and we also present some properties that are specific to time-homogeneous Markov chains. Among them is a property of expected time averages in the limit, where time averages are a special class of functions that depend on a finite number of state variables. We then introduce imprecise Markov chains whose local models are general—so not necessarily closed and convex—sets of probability mass functions. Given that we can adopt various types of independence in imprecise Markov chains—epistemic irrelevance, complete independence and repetition independence—we discuss how this choice affects the resulting lower and upper expectations. In particular, we prove various properties that are satisfied by imprecise Markov chains under epistemic irrelevance and complete independence. Regarding time averages, we show that their lower and upper expectations coincide for epistemic irrelevance and complete independence and we also prove an interesting inequality for them that can be of practical use in a queueing context.

In Chapter 6_{151}, we introduce an imprecise version of birth-death chains. For these processes, we focus in particular on the computation of so-called lower and upper expected first-passage and return times. These first-passage and return times are expressed by a function that depends on an infinite number of state variables and can take infinite values as well, and for this reason we first define our lower and upper expected first-passage and return times through the martingale-theoretic approach. We derive a system of non-linear equations through which we show how to compute these expectations in a recursive way. In order to do this, we begin with lower and upper expected first-passage times and we show that they satisfy various properties. By using our results on lower and upper expected first-passage times, we manage to efficiently compute lower and upper expected return times. Moreover, we prove that any lower or upper expected first-passage or return time can be obtained for a specific (precise) birth-death chain that is compatible with the local models of the imprecise one. Additionally, we show that we obtain the same results for lower and upper expected first-passage and return times that are defined by the measure-theoretic approach, regardless of the chosen type of independence.

Finally, in Chapter 7189, we apply our results on imprecise Markov and birth-death chains to the imprecise Geo/Geo/1/L queueing model, which can be regarded as a Geo/Geo/1/L queue whose parameters belong to certain intervals. More specifically, where the Geo/Geo/1/L queueing model assumes a time-homogeneous probability of arrival and a time-homogeneous probability of departure, we here specify intervals for them and, depending on the independence concept, sometimes drop the assumption of time-homogeneity. In this context, we also introduce a fourth independence concept, which we call fixed-parameter repetition independence, according to which lower and upper expectations are derived from a single time-homogeneous pair of arrival and departure probabilities that are varied in their respective intervals. We then investigate the implications that each independence concept has on various expectation bounds. We first discuss about the lower and upper expected (average) queue length for which we find that it is not affected by the independence concept chosen and we prove that the lower expected (average) queue length is always obtained for the time-homogeneous largest departure probability and the time-homogeneous smallest arrival probability and vice versa for the upper expected (average) queue length. However, this is not always the case for the (average) probability of each individual queue length and for the (average) probability of "turning on the server", where we show that the type of independence affects the robustness in the output. In the cases of epistemic irrelevance and complete independence we even witness differences between the probability of a queue length in the limit and its respective average one. Finally, for lower and upper expected first-passage and return times, we prove that fixed-parameter independence leads to the same lower and upper expected first-passage times as the other three notions of independence but show that this is not necessarily true for lower and upper expected return times.

1.4 PUBLICATIONS

This dissertation is the product of research on imprecise stochastic processes and Markov chains that has resulted in a number of publications. Some of these have been published in international journals [25, 52, 53] and the rest of them in the proceedings of international conferences, either as papers [50] or as abstracts followed by a poster presentation [49, 51]. The main publications that constitute the core of this dissertation are the following:

- Gert de Cooman, Jasper De Bock and Stavros Lopatatzidis. Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems. Published in the International Journal of Approximate Reasoning [25].
- Stavros Lopatatzidis, Jasper De Bock and Gert de Cooman. Calculating bounds on expected return and first passage times in finite-state imprecise birth-death chains. Published in the proceedings of ISIPTA '15 [50].
- Stavros Lopatatzidis, Jasper De Bock and Gert de Cooman. Computing lower and upper expected first-passage and return times in imprecise birth–death chains. Published in the International Journal of Approximate Reasoning [52].
- Stavros Lopatatzidis, Jasper De Bock, Gert de Cooman, Stijn De Vuyst and Joris Walraevens. Robust queueing theory: an initial study using imprecise probabilities. Published in Queueing Systems [53].

Although this dissertation only deals with discrete-time stochastic processes, this does not mean that we did not look into continuous-time stochastic processes. In fact, we proposed a method for computing lower and upper expectations in imprecise continuous-time birth-death chains, which builds upon the method presented in Reference [69] and which is summarised in the following publication:

- Stavros Lopatatzidis, Jasper De Bock and Gert de Cooman. Computational methods for imprecise continuous-time birth-death processes: a preliminary study of flipping times. Abstract published in the proceedings of ISIPTA '15 [51].

The reason why the results in the aforementioned publication were left out is because we prefer to create a dissertation that focuses on a single topic, in this case imprecise discrete-time stochastic processes.

There is also a second publication that we did not include in this dissertation.

- Stavros Lopatatzidis, Jasper De Bock and Gert de Cooman. First steps towards Little's Law with imprecise probabilities. Abstract published in the proceedings of ISIPTA '13 [49].

It constitutes our initial attempt to merge imprecise probabilities with queueing theory, but has by now been superseded by the results in Chapter 7189.

2

MODELLING UNCERTAINTY

Consider a variable X taking values in a non-empty finite set $\mathscr{X}$. Suppose that we are uncertain about the value of the variable. A simple example is the outcome of a single coin toss. The variable is the side of the coin with possible values 'Heads' and 'Tails', and one plausible way to model our uncertainty is to specify a single *probability mass function* on these two values of the variable. Probability mass functions are considered perhaps as the most typical way to model uncertainty, though they are not the only ones.

In this chapter, we present the uncertainty models that we will use throughout this thesis, which are lower and upper expectations. We first introduce the notion of a probability mass function and its corresponding expectation operator, together with some basic properties. Then, we extend these results to sets of probability mass functions by using the framework of imprecise probabilities, where the concept of lower and upper expectations arises.

2.1 PROBABILITY

A probability mass function p on $\mathscr{X}$ is any (single) element of the set

$$\Sigma_{\mathscr{X}} := \Big\{ p \in \mathbb{R}^{\mathscr{X}} : \sum_{x \in \mathscr{X}} p(x) = 1 \text{ and } (\forall x \in \mathscr{X})\, p(x) \geq 0 \Big\}. \tag{2.1}$$

Any subset of $\mathscr{X}$ is called an *event* and the set of all events is the power set of $\mathscr{X}$, denoted by $2^{\mathscr{X}}$. In the example of the coin toss the events are $\{\text{Heads}\}$, $\{\text{Tails}\}$, the complete $\mathscr{X}$ and the empty set $\varnothing$. If an event is of the form $\{x\}$, with $x \in \mathscr{X}$, we call it an *atom*. For any event $A \in 2^{\mathscr{X}}$, the probability of A is denoted by $P(A)$ and is defined by

$$P(A) := \sum_{x \in A} p(x).$$

If $A = \mathscr{X}$, then we have the certain event with $P(\mathscr{X}) = 1$ and if $A = \varnothing$, then we have the impossible event with $P(\varnothing) = 0$. Moreover, for all $x \in \mathscr{X}$, we find that $P(\{x\}) = p(x)$. It is important to mention that P is a so-called *probability measure* and the difference with the probability mass function p is that P is defined on $2^{\mathscr{X}}$, whereas p is defined on $\mathscr{X}$. In other words, p is defined on elements and P on sets of elements.

Consider again the example of the coin toss and suppose now that we have a probability mass function, which assigns probability mass $1/3$ to 'Heads' and probability mass $2/3$ to 'Tails'. One could wonder what exactly these probabilities represent. There are two basic interpretations for probability. The first one is the *frequentist probability*[1], where the probability of an event is interpreted as its relative frequency in a long series of observations. The frequentist probability can be further characterised as *objective*. This means that the probabilities assigned to events come from measurements and/or recorded observations rather than subjective evaluations, something that often applies to such cases as the coin toss and the dice roll. The second interpretation is the *epistemic* one, where the probabilities of the events are interpreted in terms of knowledge and/or available evidence. Epistemic probabilities often express a degree of personal belief about events and therefore they are often further interpreted as subjective and personal. Both the frequentist and the epistemic interpretation can be divided into further interpretations as Walley proposes in Reference [78].

The difference between frequentist and epistemic probability is subtle. A frequentist probability can work perfectly under the epistemic interpretation. This follows from the principle of direct inference [78], which says that when the values of frequentist probabilities are known, they should be regarded as epistemic ones. Furthermore, the mathematics for the results presented throughout this dissertation do not depend on the chosen interpretation. However, the interpretations provide motivation for some of the methods and results that we will develop later on. Of great importance for our motivation is the behavioural interpretation, which is connected with the account of probability developed by de Finetti [29]. According to this account, probabilities reflect inclinations or dispositions to bet and we explain this concept in the next paragraph.

Suppose that we are offered a bet on an event. Suppose as well that we can buy or sell[2] this bet at possibly different prices. Buying a bet means that we pay a certain price and we receive the reward associated with its outcome. Similarly, selling a bet means that we receive a price and we pay back the reward associated with the outcome. According to de Finetti, the probability of an event $A \in 2^{\mathscr{X}}$ is the price at which we are willing to both buy and sell the

[1] Also known as physical probability or empirical probability or aleatory probability.

[2] 'Buy' and 'sell' are also found in the literature as 'bet in favour of' and 'bet against' respectively.

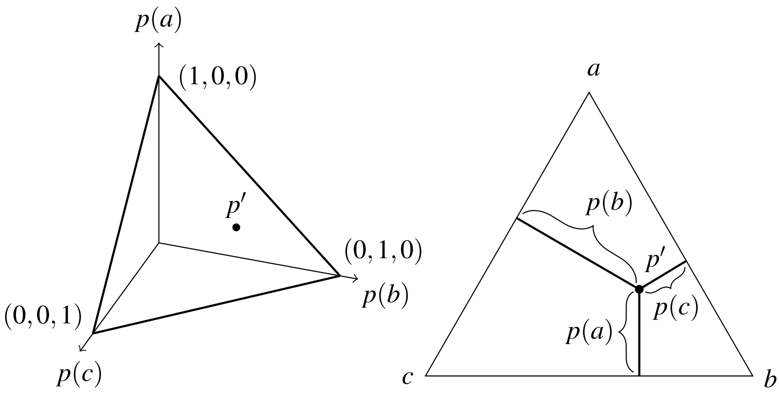

Figure 2.1: On the left, we depict the set $\Sigma_{\mathscr{X}}$ in the linear space $\mathbb{R}^{\mathscr{X}}$ for $\mathscr{X} = \{a,b,c\}$. On the right, we depict the same $\Sigma_{\mathscr{X}}$ in a two-dimensional space. For both representations, we also depict the probability mass function $p' = (3/10, 1/2, 1/5)$.

bet that has the reward one if the event happens or zero otherwise. The price that represents the probability of an event is also called its 'fair' price.

We close this section by discussing some ideas regarding representations of probability mass functions that will be useful in Section 2.3, where we consider a set of them. Suppose that we have $\mathscr{X} = \{a,b,c\}$. In this case $\mathscr{X}$ is ternary and the corresponding $\Sigma_{\mathscr{X}}$ can easily be depicted; see Figure 2.1. In the two-dimensional space, we represent $\Sigma_{\mathscr{X}}$ by means of an equilateral triangle of unit height. The elements p of $\Sigma_{\mathscr{X}}$ then correspond to points in this triangle. Every p is written as $(p(a), p(b), p(c))$, where the values $p(a), p(b)$ and $p(c)$ correspond to the probabilities of outcomes a, b and c respectively. For every such p, the value of $p(a)$ is equal to the perpendicular distance from p to the edge that opposes the corner that corresponds to a, and similarly for $p(b)$ and $p(c)$. This procedure can be similarly extended to higher dimensions (larger $\mathscr{X}$).

2.2 EXPECTATION

Any real-valued function on $\mathscr{X}$ is also called a *gamble* and the set of all gambles is denoted by $\mathscr{L}(\mathscr{X})$. We now introduce the *expectation operator* $E\colon \mathscr{L}(\mathscr{X}) \to \mathbb{R}$, which is a linear operator that satisfies the following properties:

P1. $\min f \leq E(f) \leq \max f$ for all $f \in \mathscr{L}(\mathscr{X})$; [bounds]

P2. $E(f+f') = E(f) + E(f')$ for all $f, f' \in \mathscr{L}(\mathscr{X})$; [finite-additivity]

P3. $E(\lambda f) = \lambda E(f)$ for all $f \in \mathscr{L}(\mathscr{X})$ and $\lambda \in \mathbb{R}$. [homogeneity]

For any real-valued function f in $\mathscr{L}(\mathscr{X})$, $E(f)$ is called the *expected value* (or simply *expectation*) of f. On a frequentist approach, the expected value of a function can be interpreted as the average of the function values observed in a long series of trials. However, the drawback of this interpretation is that it involves repetition. For instance, suppose that we are uncertain about the age of the King of Cambodia and that we are given a set of possible values, then the expected value in such a case is not consistent with the interpretation just mentioned.

In order to avoid such problems, we can also interpret the expected value in a behavioural way. In Section 2.136, we mentioned that we interpret the probability of an event as the fair price for betting on the event at which we gain one if the event happens or zero otherwise. A similar interpretation can be applied to more general gambles as well. We can interpret the expected value $E(f)$ of a real-valued function f as the fair price for the gamble f. In other words, for any gamble f in $\mathscr{L}(\mathscr{X})$ the price at which we are willing to either buy or sell f is $E(f)$.

The expectation of a function can be expressed in terms of probabilities. Consider any gamble $f \in \mathscr{L}(\mathscr{X})$ and any probability mass function p on $\mathscr{X}$. Then the corresponding expectation of f, denoted by $E_p(f)$, is defined by

$$E_p(f) := \sum_{x \in \mathscr{X}} p(x) f(x) \text{ for all } f \in \mathscr{L}(\mathscr{X}).$$

It is easy to see that E_p is the unique linear operator from $\mathscr{L}(\mathscr{X})$ to $\mathbb{R}$ that satisfies $E_p(\mathbb{I}_x) = p(x)$, $x \in \mathscr{X}$, where the indicator $\mathbb{I}_x \in \mathscr{L}(\mathscr{X})$ is defined by

$$\mathbb{I}_x(y) := \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases} \quad \text{for all } y \in \mathscr{X}.$$

This implies that E_p can be inferred from p. Conversely, if we have a linear operator E that satisfies (P1)↶–(P3), then there is a unique probability mass function p such that $E_p = E$.

Moreover, the behavioural interpretation of a probability $P(A)$ of an event $A \in 2^{\mathscr{X}}$ is a special case where our gamble f is equal to $\mathbb{I}_A$, where $\mathbb{I}_A$ is defined by

$$\mathbb{I}_A(y) := \begin{cases} 1 & \text{if } y \in A \\ 0 & \text{if } y \notin A \end{cases} \quad \text{for all } y \in \mathscr{X}.$$

Therefore, we understand that the expectation operator E can be regarded as an alternative, equivalent representation for p, but also as an equivalent way to model uncertainty.

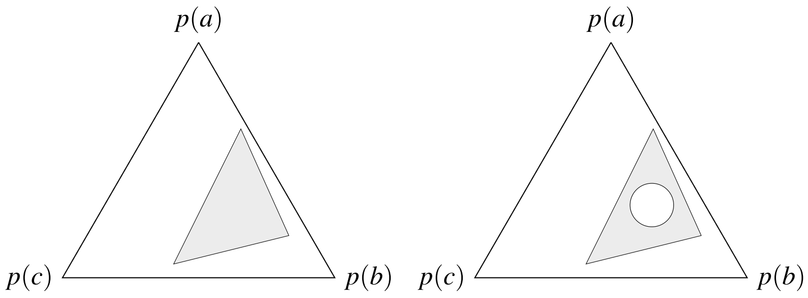

Figure 2.2: The grey areas depict two different sets of probability mass functions for $\mathscr{X} = \{a,b,c\}$.

2.3 SETS OF PROBABILITY MASS FUNCTIONS

In the previous section, we showed how to model uncertainty through an expectation operator E. The success of this approach crucially depends on the assumption that our uncertainty about the value of a variable can be described by a probability mass function p, and furthermore requires that p should be specified precisely. However, in practice, eliciting or assessing such a probability function can be difficult. This can happen especially when the probabilities are based on—possibly disagreeing—expert opinions or when they have to be learned from small amounts of—possibly poor quality—data. Whenever this is the case, the theory of imprecise probabilities [5, 72, 78] does not insist on the use of a single probability mass function p, but instead allows for the use of sets of probability mass functions. In Section 2.1$_{36}$, we saw that a probability mass function is a point in $\Sigma_{\mathscr{X}}$ and therefore, a set of probability mass functions can be any subset of $\Sigma_{\mathscr{X}}$; see for instance Figure 2.2.

We now introduce the concept of lower and upper probabilities. Given any non-empty set of probability mass functions $\mathscr{P} \subseteq \Sigma_{\mathscr{X}}$ the corresponding lower and upper probability mass of $x \in \mathscr{X}$ are defined by

$$\underline{p}(x) := \inf\left\{p(x) : p \in \mathscr{P}\right\} \text{ and } \overline{p}(x) := \sup\left\{p(x) : p \in \mathscr{P}\right\}, \tag{2.2}$$

respectively. The following example illustrates the concept of lower and upper probability masses.

Example 1. Let $\mathscr{X} := \{a,b,c\}$, consider the probability mass function

$$p^* = (p^*(a), p^*(b), p^*(c)) = (2/5, 2/5, 1/5)$$

and let $\varepsilon := 1/2$. We define the set of probability mass functions $\Phi^{\varepsilon}_{p^*}$ by

$$\Phi^{\varepsilon}_{p^*} := \left\{(1-\varepsilon)p^* + \varepsilon p : p \in \Sigma_{\mathscr{X}}\right\}, \tag{2.3}$$

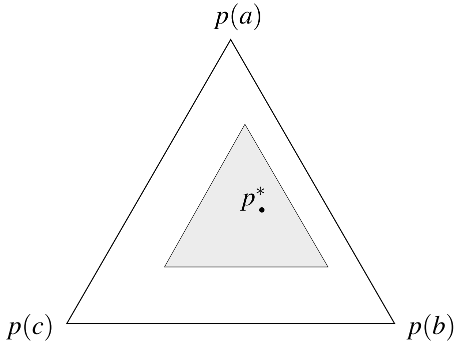

Figure 2.3: The grey area depicts the set $\Phi^{\varepsilon}_{p^*}$ of Example 1↶.

which corresponds to the the grey area in Figure 2.3. As can be seen from this figure, the set $\Phi^{\varepsilon}_{p^*}$ is the convex hull of three points, which have the following form:

$$(\overline{p}^*(a), \underline{p}^*(b), \underline{p}^*(c)),\ \ (\underline{p}^*(a), \overline{p}^*(b), \underline{p}^*(c)),\ \ (\underline{p}^*(a), \underline{p}^*(b), \overline{p}^*(c)).$$

The numerical values of the lower and upper probability masses in these expressions are given by

$$\underline{p}^*(a) = \underline{p}^*(b) = {}^1\!/_5,\ \ \overline{p}^*(a) = \overline{p}^*(b) = {}^7\!/_{10},\ \ \underline{p}^*(c) = {}^1\!/_{10},\ \ \overline{p}^*(c) = {}^3\!/_5.$$

They are easily obtained by combining Equations (2.2)↶ and (2.3)↶. Sets of probability mass functions defined by Equation (2.3) are called linear-vacuous mixtures [78, Section 2.9.2].[3] ◊

It is not necessary for a set of probability mass functions to be defined directly, as was the case in Example 1↶. It can also be specified indirectly, by means of partial constraints on probabilities. A particularly appealing way of doing so is to specify a probability interval $[\underline{p}(x), \overline{p}(x)]$ for every x in $\mathscr{X}$, and to let Φ be the largest subset of $\Sigma_{\mathscr{X}}$ that satisfies these bounds. We illustrate this in our next example.

Example 2. Let $\mathscr{X} := \{a, b, c\}$ and consider that

$$p(a) \in [{}^1\!/_5, {}^8\!/_{15}],\ p(b) \in [{}^1\!/_5, {}^8\!/_{15}]\ \textit{and}\ p(c) \in [{}^1\!/_{10}, {}^{13}\!/_{30}].$$

We denote by Φ_2 the largest set of probability mass functions $p \in \Sigma_{\mathscr{X}}$ that satisfies these constraints; see Figure 2.4↷. As can be seen from Figure 2.4↷, the set Φ_2 is a hexagon. Any vertex of the hexagon is a probability mass

[3] Also known as ε-contaminated models.

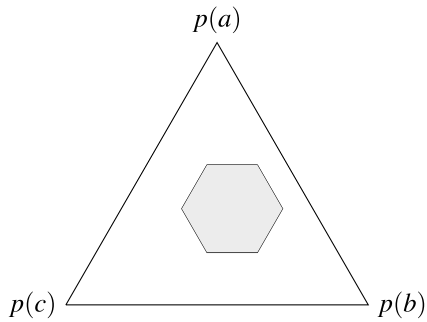

Figure 2.4: The grey area depicts the set Φ_2 of Example $2_{\curvearrowleft}$.

function with values an upper probability mass, a lower probability mass and the complement of their sum. For instance, the upper left vertex in Figure 2.4 is the probability mass function $(\overline{p}(a), \underline{p}(b), 1 - \overline{p}(a) - \underline{p}(b))$. These vertices correspond to the following points

$$\left(8/15, 1/5, 8/30\right), \left(8/15, 11/30, 1/10\right), \left(11/30, 8/15, 1/10\right),$$
$$\left(1/5, 8/15, 8/30\right), \left(1/5, 11/30, 13/30\right), \left(11/30, 1/5, 13/30\right)$$

and they can be identified in the above order starting from the upper left vertex of the polygon in Figure 2.4 and moving clockwise. ◊

However, as our next example should clarify, a set of probability mass functions is not always completely characterised by its lower and upper probability masses.

Example 3. Let Φ_3 be the circular disc of probability mass functions that is depicted in Figure $2.5_{\curvearrowright}$. In order to allow for an easy comparison, Figure $2.5_{\curvearrowright}$ also depicts the set Φ_2 of Example $2_{\curvearrowleft}$. These two sets clearly have the same lower and upper probabilities. ◊

The sets of probability mass functions presented in Examples 1_{40}–3, as well as the one on the left side of Figure 2.2_{40}, have something in common—they belong to the category of the so-called *credal sets*, denoted by Φ. The definition of a credal set goes as follows.

Definition 1. *Consider any set $\mathscr{X}$, then a set of probability mass functions $\Phi \subseteq \Sigma_{\mathscr{X}}$ is called a credal set if and only if it is non-empty, closed and convex.*

Regarding Definition 1, "closed" means that the boundary points of the credal set Φ are included in Φ and "convex" means that the line segment connecting any two points of Φ lies within Φ.

One important notion associated with credal sets is that of an *extreme point*, which is defined as follows.

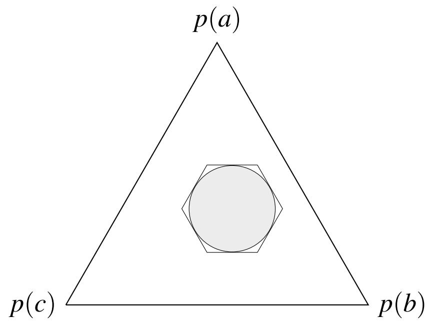

Figure 2.5: The grey zone depicts the set Φ_3 of Example 2_{41}. The set Φ_3 tangents to each side of the set Φ_2 that was described earlier in the same section.

Definition 2. *Consider a credal set Φ. Then, a probability mass function $p \in \Phi$ is called an extreme point, if for all $\theta \in [0,1]$ and all $p_1, p_2 \in \Phi$ such that $p_1 \neq p_2$, we have that*

$$p = \theta p_1 + (1-\theta)p_2 \Rightarrow \textit{ either } \theta = 1 \textit{ or } \theta = 0.$$

For any credal set Φ, we denote by $\mathrm{ext}(\Phi)$ the set of the extreme points of Φ.

The extreme points are the points in the boundary of the credal set that cannot be written as a linear combination of any two other points in the credal set. In Examples 1_{40} and 2_{41} the extreme points are the vertices of the polygons, whereas in Example $3_{\curvearrowleft}$ all the boundary points of the circular disc are extreme points. In the next section, we will see how the extreme points of a credal set can help us compute lower and upper expectations.

Often in the theory of imprecise probabilities it is taken for granted that sets of probability mass functions are closed and convex. However, this need not be the case; consider for instance the right-hand side of Figure 2.2_{40}. We close this section by providing an example, where the set of probability mass functions is specified from the product of probability intervals, as this might occur in a queueing model that we consider later on in Chapter 7_{189}.

Example 4. Let $\mathscr{X} := \{a,b,c\}$ and consider that

$$p(a) = p^\star(1-p^\circ),\ p(c) = p^\circ(1-p^\star) \textit{ and } p(b) = 1 - p(a) - p(c),$$

where $p^\star \in [1/5, 7/10]$ and $p^\circ \in [1/10, 7/10]$. We denote by Ψ_1 the largest set of probability mass functions $p \in \Sigma_{\mathscr{X}}$ that satisfies these constraints. The set Ψ_1 is closed but not convex; see Figure $2.6_{\curvearrowright}$.

Judging by Figure 2.6, we see that the set Ψ_1 has four "corners", which are obtained for the extreme values of $p^\star$ and p°. Starting from the corner of Ψ_1 that has the largest $p(a)$ value and moving clockwise, these corners correspond

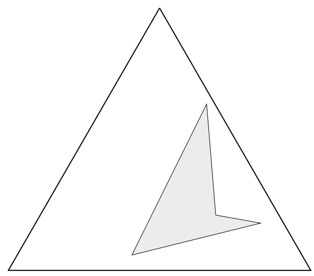

Figure 2.6: The grey zone depicts the set Ψ_1 of Example 4↶.

to the following probability mass functions:

$$p_1 = (63/100, 17/50, 3/100), p_2 = (21/100, 29/50, 21/100)$$
$$p_3 = (9/50, 37/50, 4/50), p_4 = (3/50, 19/50, 14/25)$$

The "corners" of such sets do not necessarily consist of lower or upper probabilities of $p(a), p(b)$ and $p(c)$, nor are they guaranteed to be extreme points. In this example, such a corner is the probability mass function p_2. ◊

2.4 LOWER AND UPPER EXPECTATIONS

A *lower expectation operator* (or simply *lower expectation*) $\underline{E}\colon \mathscr{L}(\mathscr{X}) \to \mathbb{R}$ is a non-linear operator that satisfies the following properties:[4]

C1. $\underline{E}(f) \geq \min f$ for all $f \in \mathscr{L}(\mathscr{X})$; [bounds]

C2. $\underline{E}(f+f') \geq \underline{E}(f)+\underline{E}(f')$ for all $f, f' \in \mathscr{L}(\mathscr{X})$; [superadditivity]

C3. $\underline{E}(\lambda f) = \lambda \underline{E}(f)$ for all $f \in \mathscr{L}(\mathscr{X})$ and real $\lambda \in \mathbb{R}_{\geq 0}$, [non-negative homogeneity]

where $\mathbb{R}_{\geq 0}$ is the set of all non-negative real numbers. For any $f \in \mathscr{L}(\mathscr{X})$, $\underline{E}(f)$ is the so-called *lower expected value*[5] (or simply *lower expectation*) of f. We call lower expectation assessment, or also, partial lower expectation specification, any partial function from $\mathscr{L}(\mathscr{X})$ to $\mathbb{R}$. It is possible to have lower expectation assessments that are not lower expectations or cannot be extended to lower expectations and we will see such cases in the next section.

From any given lower expectation $\underline{E}$ and for all $f \in \mathscr{L}(\mathscr{X})$, we can derive the *conjugate upper expectation* of $\underline{E}$ by $\overline{E}(f) = -\underline{E}(-f)$. Therefore, it

[4]These properties are also known as coherence axioms.
[5]Also called lower prevision, see References [5, 72, 78].

suffices to consider only one of them. We mainly focus on lower expectations. Lower and upper expectations satisfy also the following properties:

C4. $\underline{E}(f) \leq \underline{E}(f')$ and $\overline{E}(f) \leq \overline{E}(f')$ for all $f, f' \in \mathscr{L}(\mathscr{X})$ with $f \leq f'$;

C5. $\min f \leq \underline{E}(f) \leq \overline{E}(f) \leq \max f$ for all $f \in \mathscr{L}(\mathscr{X})$;

C6. $\overline{E}(f+f') \leq \overline{E}(f) + \overline{E}(f')$ for all $f, f' \in \mathscr{L}(\mathscr{X})$;

C7. $\overline{E}(\lambda f) = \lambda \overline{E}(f)$ for all $f \in \mathscr{L}(\mathscr{X})$ and real $\lambda \in \mathbb{R}_{\geq 0}$;

C8. $\underline{E}(f+\mu) = \underline{E}(f) + \mu$ and $\overline{E}(f+\mu) = \overline{E}(f) + \mu$ for all $f \in \mathscr{L}(\mathscr{X})$ and $\mu \in \mathbb{R}$.

Lower and upper expectations have different interpretations, which are proposed and extensively discussed by Walley in Reference [78]. Here we emphasise the behavioural interpretation as it will turn out to be our main motivation for developing the models in Chapter 4_{86}. In a minimal behavioural interpretation, the lower expected value $\underline{E}(f)$ represents a *supremum buying price* for the gamble f, in the sense that we are willing to buy the gamble for any price strictly lower than $\underline{E}(f)$. Similarly, $\overline{E}(f)$ is interpreted as an *infimum selling price* for the gamble f, in the sense that we are willing to sell the gamble for any price strictly higher than $\overline{E}(f)$.

Lower and upper expectations can also be associated with probabilities, in the sense that they can be derived from sets of probability mass functions $\mathscr{P} \subseteq \Sigma_{\mathscr{X}}$. Consider any $f \in \mathscr{L}(\mathscr{X})$ and any non-empty set $\mathscr{P} \subseteq \Sigma_{\mathscr{X}}$, then we define the lower and upper expectations of f as follows:

$$\underline{E}_{\mathscr{P}}(f) := \inf\{E_p(f) \colon p \in \mathscr{P}\} \text{ and } \overline{E}_{\mathscr{P}}(f) := \sup\{E_p(f) \colon p \in \mathscr{P}\}. \tag{2.4}$$

$\underline{E}_{\mathscr{P}}(f)$ satisfies properties $\mathrm{C1}_{\frown}$–$\mathrm{C3}_{\frown}$ and consequently, $\underline{E}_{\mathscr{P}}(f)$ and $\overline{E}_{\mathscr{P}}(f)$ satisfy properties C5–C8. In case $\mathscr{P} = \{p\}$, we have that $\underline{E}_{\mathscr{P}} = \overline{E}_{\mathscr{P}} = E_p$ and then the properties $\mathrm{C1}_{\frown}$–$\mathrm{C3}_{\frown}$ coincide with properties $\mathrm{P1}_{38}$–$\mathrm{P3}_{39}$ respectively. The expectations $\underline{E}_{\mathscr{P}}(f)$ and $\overline{E}_{\mathscr{P}}(f)$ are interpreted as bounds on the expected value $E_p(f)$, something that is known as the *sensitivity analysis* interpretation.

Conversely, if we are given a lower expectation $\underline{E}$ or an upper one $\overline{E}$ on $\mathscr{L}(\mathscr{X})$, we can produce a set of probability mass functions $\mathscr{P}$ such that Equation (2.4) holds for all $f \in \mathscr{L}(\mathscr{X})$. Note that $\mathscr{P}$ might not be unique. However, if $\mathscr{P}$ is a credal set, then it is unique [39]. Specifically, given a lower expectation $\underline{E}$ on $\mathscr{L}(\mathscr{X})$, then the corresponding credal set $\Phi_{\underline{E}}$ is defined by

$$\begin{aligned} \Phi_{\underline{E}} :=& \{p \in \Sigma_{\mathscr{X}} : E_p(f) \geq \underline{E}(f) \text{ for all } f \in \mathscr{L}(\mathscr{X})\} \\ =& \{p \in \Sigma_{\mathscr{X}} : E_p(f) \leq \overline{E}(f) \text{ for all } f \in \mathscr{L}(\mathscr{X})\} \\ =& \{p \in \Sigma_{\mathscr{X}} : \underline{E}(f) \leq E_p(f) \leq \overline{E}(f) \text{ for all } f \in \mathscr{L}(\mathscr{X})\}. \end{aligned} \tag{2.5}$$

In fact, the credal set $\Phi_{\underline{E}}$ is the largest set of probability mass functions such that the constraints imposed by the lower expectation $\underline{E}$ are satisfied. In this case, $\Phi_{\underline{E}}$ can be interpreted as our incomplete assessments. This also means that we can derive $\Phi_{\underline{E}}$ from the conjugate upper expectation of $\underline{E}$, i.e. $\overline{E}$.

We close this section by discussing how to derive lower and upper expectations from credal sets computationally. In Section 2.3$_{40}$, we saw some examples of credal sets that were either directly given or described through bounds on the probabilities of the elements of $\mathscr{X}$. Here, we also consider other ways in which a credal set can be described. One way to describe a credal set is through a finite number of linear inequality constraints, which are represented by $\mathbf{A}p \leq \mathbf{b}$, where $\mathbf{A}$ is a real-valued matrix $n \times |\mathscr{X}|$, p is a column vector of variables that assumes values in $\Sigma_{\mathscr{X}}$ and $\mathbf{b}$ is a real-valued column vector of size n, for some $n \in \mathbb{N}$, with $\mathbb{N}$ being the set of natural numbers excluding zero.[6] For each $i \in \{1,\dots,n\}$, the i-th row of $\mathbf{A}$, that is $\mathbf{A}_i$, and the i-th element of $\mathbf{b}$, that is $\mathbf{b}_i$, form an inequality $\mathbf{A}_i p \leq \mathbf{b}_i$, which is represented by a closed halfspace in $\mathbb{R}^{\mathscr{X}}$. The intersection of all these closed halfspaces yields a closed polytope [36, Theorem 3.1.1] and if the polytope is in addition non-empty, then it represents our credal set. There are also credal sets that can only be described by an infinite number of linear inequality constraints. For instance in order to describe the credal set in Figure 2.5$_{43}$, we need an infinite number of linear inequality constraints.

Suppose now that we have a finite or infinite number of linear inequality constraints $\mathbf{A}p \leq \mathbf{b}$ and that we want to compute the lower expectation of a function $f \in \mathscr{L}(\mathscr{X})$. According to Equation (2.4)$_{\curvearrowleft}$, the expectation $\underline{E}(f)$ is the infimum of $E_p(f)$ such that the constraints—describing the credal set—are satisfied. Since $E_p(f)$ is linear with respect to p, we have a linear program of the following form

$$\begin{aligned} &\text{minimise} && E_p(f) \\ &\text{subject to} && p \in \Sigma_{\mathscr{X}} \\ &\text{and} && \mathbf{A}p \leq \mathbf{b}. \end{aligned}$$

Moreover, in the aforementioned linear program if we maximise $E_p(f)$ instead of minimising it, then we obtain the upper expectation of f.

Another way to describe a credal set is through its extreme points (Definition 2$_{43}$), which are either given directly or are implied by a finite or infinite number of linear inequality constraints. Finding the extreme points for a given collection of linear inequality constraints is also known as the vertex enumeration problem. Consider any $f \in \mathscr{L}(\mathscr{X})$ and any credal set Φ with extreme points ext(Φ), then there is some $p \in \Phi$ such that $\underline{E}_{\Phi}(f) = E_p(f)$. Note that such a p might not be unique. Fortunately, we know from literature [78, The-

[6] The set of natural numbers including zero is denoted by $\mathbb{N}_0$.

orem 3.6.2] that

$$\underline{E}_{\Phi}(f) = \underline{E}_{\mathrm{ext}(\Phi)}(f) \text{ and } \overline{E}_{\Phi}(f) = \overline{E}_{\mathrm{ext}(\Phi)}(f). \tag{2.6}$$

In practice, however, it will rarely be feasible to solve a linear program or enumerate the extreme points when the number of linear constraints is infinite. For this reason, we prefer working with credal sets that are described by a finite number of linear constraints. In this case, linear programs can be solved through solvers or various procedures such as the simplex algorithm [19, 20] and Karmarkar's algorithm [42], which is an interior point method. As far as extreme points are concerned, we can enumerate the extreme points from a finite number of linear constraints using the Avis-Fukuda algorithm [6].

2.5 LOWER AND UPPER EXPECTATIONS FOR A SPECIAL CLASS OF SETS OF PROBABILITY MASS FUNCTIONS

In this section, we present some useful properties regarding the computation of lower and upper expectations, when the set of probability mass functions has a special structure that we define shortly after in Equation (2.9)$_{\curvearrowright}$.

First of all, we introduce the notion of a *convex hull* as we will often use it for the derivation of the results presented in this section. Consider any non-empty set $\mathscr{P} \subseteq \Sigma_{\mathscr{X}}$, then its convex hull, denoted by $\mathrm{conv}(\mathscr{P})$, is the smallest convex set including $\mathscr{P}$ and it is defined by

$$\mathrm{conv}(\mathscr{P}) := \left\{ \sum_{i=1}^{n} c_i p_i \colon n \in \mathbb{N},\, p_i \in \mathscr{P},\, c_i \in \mathbb{R}_{\geq 0} \text{ and } \sum_{i=1}^{n} c_i = 1 \right\}. \tag{2.7}$$

If the set $\mathscr{P}$ is finite, then it suffices to take convex combinations of at most $|\mathscr{P}|$ points in $\mathscr{P}$. An example of the convex hull of a non-convex set can be seen in Figure 2.7$_{\curvearrowright}$. The following property now tells us that for any $\mathscr{P} \subseteq \Sigma_{\mathscr{X}}$, it makes no difference whether lower and upper expectations are taken with respect to $\mathscr{P}$ or with respect to $\mathrm{conv}(\mathscr{P})$.

Theorem 1. *Consider any non-empty finite set $\mathscr{X}$ and any non-empty set of probability mass functions $\mathscr{P} \subseteq \Sigma_{\mathscr{X}}$. Then for all $f \in \mathscr{L}(\mathscr{X})$, it holds that $\underline{E}_{\mathscr{P}}(f) = \underline{E}_{\mathrm{conv}(\mathscr{P})}(f)$ and $\overline{E}_{\mathscr{P}}(f) = \overline{E}_{\mathrm{conv}(\mathscr{P})}(f)$.*

Proof. If $\mathscr{P}$ is convex, then the result follows trivially since the convex hull of a convex set is the set itself. Therefore, we only focus on the case where $\mathscr{P}$ is non-convex and we show that for all $p^* \in \mathrm{conv}(\mathscr{P})$ such that $p^* \notin \mathscr{P}$, there are $p', p'' \in \mathscr{P}$ such that $E_{p'}(f) \leq E_{p^*}(f) \leq E_{p''}(f)$, for all $f \in \mathscr{L}(\mathscr{X})$. This guarantees that $\underline{E}_{\mathscr{P}}(f) \leq E_{p^*}(f) \leq \overline{E}_{\mathscr{P}}(f)$ for all $f \in \mathscr{L}(\mathscr{X})$ and therefore also $\underline{E}_{\mathscr{P}}(f) \leq \underline{E}_{\mathrm{conv}(\mathscr{P})}(f)$ and $\overline{E}_{\mathrm{conv}(\mathscr{P})}(f) \leq \overline{E}_{\mathscr{P}}(f)$. The converse inequalities follow at once from $\mathscr{P} \subseteq \mathrm{conv}(\mathscr{P})$.

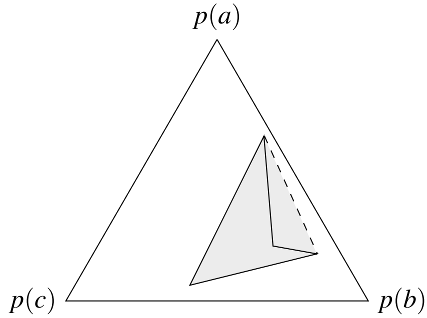

Figure 2.7: The grey zone depicts the convex hull of the set of probability mass functions Ψ_1 from Example 4_{43}.

Due to Equation $(2.7)_{\curvearrowleft}$, we find that

$$E_{p^*}(f) = \sum_{x\in\mathscr{X}} f(x)p^*(x) = \sum_{x\in\mathscr{X}} f(x)\sum_{i=1}^{n} c_i p_i(x), \tag{2.8}$$

where for all $i \in \{1,\ldots,n\}$, we have that $p_i \in \mathscr{P}$, $c_i \in \mathbb{R}_{\geq 0}$, and where $\sum_{i=1}^{n} c_i = 1$, for some $n \in \mathbb{N}$. It now follows from Equation (2.8) that

$$E_{p^*}(f) = \sum_{x\in\mathscr{X}} f(x)\sum_{i=1}^{n} c_i p_i(x) = \sum_{i=1}^{n} c_i \sum_{x\in\mathscr{X}} f(x)p_i(x) = \sum_{i=1}^{n} c_i E_{p_i}(f).$$

Now $\sum_{i=1}^{n} c_i E_{p_i}(f) \leq \max_{i=1}^{n} E_{p_i}(f)$. So there is some $p_j \in \mathscr{P}$ such that

$$\sum_{i=1}^{n} c_i E_{p_i}(f) \leq E_{p_j}(f).$$

Similarly, there is some $p_k \in \mathscr{P}$ such that

$$\sum_{i=1}^{n} c_i E_{p_i}(f) \geq \min_{i=1}^{n} E_{p_i}(f) = E_{p_k}(f). \qquad \square$$

For the rest of this section, we assume that $\mathscr{X} = \{a,b,c\}$. Consider as well any two probability intervals I_1 and I_2, that is

$$I_i := [\underline{p}_i, \overline{p}_i], \text{ where } 0 \leq \underline{p}_i \leq \overline{p}_i \leq 1 \text{ for all } i \in \{1,2\},$$

from which we derive the following set of probability mass functions

$$\Psi_{I_1,I_2} = \big\{p \in \Sigma_{\mathscr{X}} \colon p(a) = p_1(1-p_2),\ p(b) = p_2(1-p_1),$$
$$p(c) = 1 - p(a) - p(b), \text{ for some } p_1 \in I_1 \text{ and } p_2 \in I_2\big\}. \tag{2.9}$$

Note that Ψ_{I_1,I_2} is non-empty, closed but not necessarily convex; see Example 4_{43}. Before we present the next property of this section, we introduce one more set of probability mass functions. Consider any two probability intervals I_1 and I_2, then we define the set of "corner" points of Ψ_{I_1,I_2}, denoted by $\mathscr{V}_{I_1,I_2}$, as follows

$$\begin{aligned}\mathscr{V}_{I_1,I_2} := \big\{p \in \Sigma_{\mathscr{X}} : p(a) = p_1(1-p_2),\ p(b) = p_2(1-p_1),\\ p(c) = 1-p(a)-p(b), \text{ for some } p_1 \in \{\min(I_1),\max(I_1)\}\\ \text{and } p_2 \in \{\min(I_2),\max(I_2)\}\big\}.\end{aligned} \tag{2.10}$$

The following property now tells us that the convex hull of Ψ_{I_1,I_2} and the convex hull of $\mathscr{V}_{I_1,I_2}$ are the same.

Proposition 2. *Consider any two probability intervals I_1 and I_2. Consider as well the sets of probability mass functions Ψ_{I_1,I_2} and $\mathscr{V}_{I_1,I_2}$ defined by Equations* (2.9)$_\frown$ *and* (2.10)*, respectively. Then* $\text{conv}(\Psi_{I_1,I_2}) = \text{conv}(\mathscr{V}_{I_1,I_2})$.

Proof. We first show that $\text{conv}(\Psi_{I_1,I_2}) \subseteq \text{conv}(\mathscr{V}_{I_1,I_2})$. Let $I_1 = [\underline{p}_1, \overline{p}_1]$, $I_2 = [\underline{p}_2, \overline{p}_2]$. Due to Equation (2.9)$_\frown$, for any $p \in \Psi_{I_1,I_2}$, we have that $p(a) = p_1(1-p_2)$, $p(b) = p_2(1-p_1)$ and $p(c) = 1-p(a)-p(b)$ for some $p_1 \in [\underline{p}_1, \overline{p}_1]$ and $p_2 \in [\underline{p}_2, \overline{p}_2]$. Since p_1 can be written as a convex combination of $\underline{p}_1$ and $\overline{p}_1$ and p_2 as a convex combination of $\underline{p}_2$ and $\overline{p}_2$, we have that $p_1 = \lambda\underline{p}_1 + (1-\lambda)\overline{p}_1$ and that $p_2 = \lambda'\underline{p}_2 + (1-\lambda')\overline{p}_2$ with $\lambda, \lambda' \in [0,1]$, and we therefore find that

$$\begin{aligned}p(a) = \lambda\underline{p}_1 + (1-\lambda)\overline{p}_1 - \lambda\lambda'\underline{p}_1\underline{p}_2 - \lambda(1-\lambda')\underline{p}_1\overline{p}_2\\ -\lambda'(1-\lambda)\overline{p}_1\underline{p}_2 - (1-\lambda)(1-\lambda')\overline{p}_1\overline{p}_2.\end{aligned} \tag{2.11}$$

Now let $\lambda\lambda' = \lambda_1$, $\lambda(1-\lambda') = \lambda_2$, $\lambda'(1-\lambda) = \lambda_3$ and $(1-\lambda)(1-\lambda') = \lambda_4$. It is easy to see that $\sum_{i=1}^{4}\lambda_i = 1$ and also that $\lambda_1 + \lambda_2 = \lambda$ and $\lambda_3 + \lambda_4 = 1-\lambda$. Therefore, Equation (2.11) becomes

$$\begin{aligned}p(a) =& (\lambda_1+\lambda_2)\underline{p}_1 + (\lambda_3+\lambda_4)\overline{p}_1 - \lambda_1\underline{p}_1\underline{p}_2 - \lambda_2\underline{p}_1\overline{p}_2 - \lambda_3\overline{p}_1\underline{p}_2 - \lambda_4\overline{p}_1\overline{p}_2\\ =& \lambda_1\underline{p}_1(1-\underline{p}_2) + \lambda_2\underline{p}_1(1-\overline{p}_2) + \lambda_3\overline{p}_1(1-\underline{p}_2) + \lambda_4\overline{p}_1(1-\overline{p}_2).\end{aligned} \tag{2.12}$$

Similarly, we find that

$$p(b) = \lambda_1\underline{p}_2(1-\underline{p}_1) + \lambda_2\underline{p}_2(1-\overline{p}_1) + \lambda_3\overline{p}_2(1-\underline{p}_1) + \lambda_4\overline{p}_2(1-\overline{p}_1). \tag{2.13}$$

Hence, by combining Equations (2.12) and (2.13) with Equation $(2.7)_{47}$ for $\mathscr{P} = \mathscr{V}_{I_1,I_2}$, where $\mathscr{V}_{I_1,I_2}$ is defined by Equation (2.10), we infer that

$$p \in \Psi_{I_1,I_2} \Rightarrow p \in \text{conv}(\mathscr{V}_{I_1,I_2}),$$

which implies that $\Psi_{I_1,I_2} \subseteq \text{conv}(\mathscr{V}_{I_1,I_2})$ and therefore that

$$\text{conv}(\Psi_{I_1,I_2}) \subseteq \text{conv}(\text{conv}(\mathscr{V}_{I_1,I_2})) = \text{conv}(\mathscr{V}_{I_1,I_2}).$$

Conversely, since $\mathscr{V}_{I_1,I_2} \subseteq \Psi_{I_1,I_2}$, we find that $\text{conv}(\mathscr{V}_{I_1,I_2}) \subseteq \text{conv}(\Psi_{I_1,I_2})$, which completes the proof. □

Finally, we have the following property.

Proposition 3. *Consider any two probability intervals I_1 and I_2. Consider as well the sets of probability mass functions Ψ_{I_1,I_2} and $\mathscr{V}_{I_1,I_2}$ defined by Equations* (2.9)$_{48}$ *and* (2.10)$_{\frown}$*, respectively. Then for all $f \in \mathscr{L}(\mathscr{X})$, it holds that*

$$\underline{E}_{\Psi_{I_1,I_2}}(f) = \underline{E}_{\text{conv}(\Psi_{I_1,I_2})}(f) = \underline{E}_{\text{conv}(\mathscr{V}_{I_1,I_2})}(f) = \underline{E}_{\mathscr{V}_{I_1,I_2}}(f)$$

and similarly for the upper expectations.

Proof. The first and third equality follow immediately from Theorem 1_{47} and the second from Proposition $2_{\frown}$. □

From Proposition 3, we infer that when a set of probability mass functions is of the form Ψ_{I_1,I_2}, we can then compute lower and upper expectations using the finite set $\mathscr{V}_{I_1,I_2}$ instead, whose elements are probability mass functions derived from the bounds of the probability intervals I_1 and I_2.

2.6 NATURAL EXTENSION

We have seen that it is possible to specify a lower expectation from a given credal set and vice versa. In this section, we discuss specifying lower and upper expectations directly. Specifically, we examine how we can derive a lower expectation on $\mathscr{L}(\mathscr{X})$, when we are given a lower expectation assessment on a domain smaller than $\mathscr{L}(\mathscr{X})$.

Consider any $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$ and suppose that we have a lower expectation assessment $\underline{E}\colon \mathscr{K} \to \mathbb{R}$, then a plausible question is whether we can extend $\underline{E}$ to $\mathscr{L}(\mathscr{X})$ or not. To answer this, we first need to briefly explain the notion of coherence. In simple words, if a lower expectation assessment is coherent, it means that it is self-consistent. To understand this statement better, in the next paragraph we present two definitions of coherence.

First, we have the following definition [78, Section 2.5.4].

Definition 3. *Consider any $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$ and any function $\underline{E}\colon \mathscr{K} \to \mathbb{R}$. Then $\underline{E}$ is coherent if and only if*

$$\sup_{x\in\mathscr{X}} \left\{ \sum_{i=1}^{n} c_i[f_i(x) - \underline{E}(f_i)] - c_0[f_0(x) - \underline{E}(f_0)] \right\} \geq 0,$$

for all $n \in \mathbb{N}$, all $f_0, f_i \in \mathscr{K}$ and all $c_0, c_i \in \mathbb{R}_{\geq 0}$.

If $\underline{E}$ is coherent and $\mathscr{K} = \mathscr{L}(\mathscr{X})$, then $\underline{E}$ is a lower expectation; see Reference [78, Theorem 2.5.5].

We can also characterise coherence in terms of probability mass functions. Using $\underline{E}$, we first define the following set of probability mass functions:

$$\mathscr{P}_{\underline{E}} := \left\{ p \in \Sigma_{\mathscr{X}} : E_p(f) \geq \underline{E}(f) \text{ for all } f \in \mathscr{K} \right\}.$$

We then have the following property, which is known as the lower envelope theorem [78, Theorem 3.3.3 (b)].

Theorem 4. *Consider any $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$ and any function $\underline{E}\colon \mathscr{K} \to \mathbb{R}$. Then $\underline{E}$ is coherent if and only if $\underline{E}(f) = \min\{E_p(f)\colon p \in \mathscr{P}_{\underline{E}}\}$ for all $f \in \mathscr{K}$, that is, if and only if $\underline{E}$ is the lower envelope of $\mathscr{P}_{\underline{E}}$.*

In other words, $\underline{E}$ is coherent if and only if $\underline{E}_{\mathscr{P}_{\underline{E}}}$ coincides with $\underline{E}$ on $\mathscr{K}$.

Suppose now that we have a coherent $\underline{E}$ on a domain $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$, then we can extend it to a lower expectation on $\mathscr{L}(\mathscr{X})$, denoted by $\underline{\mathscr{E}}$ and called its *natural extension* [78, Theorem 3.4.1], which is defined as follows.

$$\underline{\mathscr{E}}(f) := \min \left\{ E_p(f)\colon p \in \mathscr{P}_{\underline{E}} \right\}, \tag{2.14}$$

for all $f \in \mathscr{L}(\mathscr{X})$. We understand that the definition of natural extension given by Equation (2.14) is directly connected with Theorem 4. We also present an alternative expression for natural extension [78, Definition 3.1.1] that is connected with Definition $3_{\curvearrowleft}$ and goes as follows.

$$\underline{\mathscr{E}}(f) := \sup \left\{ \alpha \in \mathbb{R} : f - \alpha \geq \sum_{i=1}^{n} c_i [f_i - \underline{E}(f_i)], \right.$$
$$\left. n \in \mathbb{N}, f_i \in \mathscr{K} \text{ and } c_i \in \mathbb{R}_{\geq 0} \right\}, \tag{2.15}$$

for all $f \in \mathscr{L}(\mathscr{X})$. Finally, we have the following result [78, Theorem 3.1.2 (d)], which is a trivial consequence of Theorem 4 and Equation (2.14).

Theorem 5. *$\underline{E}$ is coherent if and only if the natural extension of $\underline{E}$ coincides with $\underline{E}$ on $\mathscr{K}$, or in other words, if and only if it can be extended to a lower expectation on $\mathscr{L}(\mathscr{X})$.*

Similar results can be obtained for a coherent $\overline{E}$, but since $\overline{E}$ and $\underline{E}$ are connected through conjugacy, we only focus on $\underline{E}$ and $\underline{\mathscr{E}}$.

If the domain $\mathscr{K}$ is finite, then in Equation (2.15) we only need to consider sums up to $|\mathscr{K}|$. In that case, if we let $\mathscr{K} := \{f_1, \ldots, f_{|\mathscr{K}|}\}$, then for any

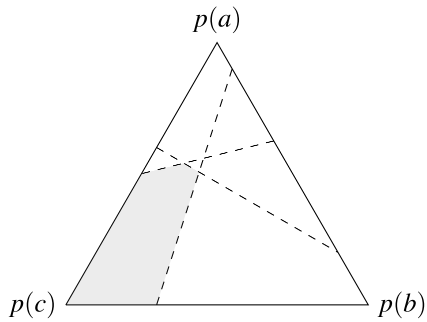

Figure 2.8: The grey area depicts the set of probability mass functions implied by the assessments of Example 5.

$f \in \mathscr{L}(\mathscr{X})$, we can obtain the natural extension $\underline{\mathscr{E}}(f)$ by solving the following linear program:

$$\begin{array}{ll} \text{maximise} & \alpha \\ \text{subject to} & \forall i \in \{1,\ldots,|\mathscr{K}|\},\ c_i \in \mathbb{R}_{\geq 0} \\ \text{and} & \forall x \in \mathscr{X} : f(x) - \alpha \geq \sum_{i=1}^{|\mathscr{K}|} c_i (f_i(x) - \underline{E}(f_i)). \end{array} \tag{2.16}$$

In essence, given a coherent $\underline{E}$ on a domain $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$, with natural extension we construct the smallest lower expectation on $\mathscr{L}(\mathscr{X})$ that extends $\underline{E}$.

We now provide an example in which we are given a coherent $\underline{E}$ on a finite domain $\mathscr{K} \subset \mathscr{L}(\mathscr{X})$ and we compute the lower expectation of a function in $\mathscr{L}(\mathscr{X}) \setminus \mathscr{K}$ using natural extension.

Example 5. Consider the set $\mathscr{X} = \{a,b,c\}$ and the following assessments[7]

$$f_1 := [1,2,3]^{\mathrm{T}} \textit{ with } \underline{E}(f_1) = 9/5,\ f_2 := [-4,4,2]^{\mathrm{T}} \textit{ with } \underline{E}(f_2) = -1 \textit{ and }$$
$$f_3 := [1,-6,3]^{\mathrm{T}} \textit{ with } \underline{E}(f_3) = 3/10,$$

where $\underline{E}$ is defined on the domain $\mathscr{K} := \{f_1, f_2, f_3\}$. A representation of these assessments is depicted in Figure 2.8.

We now compute the lower expectation of a function f in $\mathscr{L}(\mathscr{X})$ with values $f(a) = 2$, $f(b) = -4$ and $f(c) = -1$. According to the linear program given by (2.16), $\underline{\mathscr{E}}(f)$ is the supremum value of α satisfying the following

[7] Gambles are of the form $[f(a), f(b), f(c)]^{\mathrm{T}}$, where T stands for the transpose.

system of inequalities

$$\alpha \leq 2 + \frac{4}{5}c_1 + 3c_2 - \frac{7}{10}c_3$$
$$\alpha \leq -4 - \frac{1}{5}c_1 - 5c_2 + \frac{63}{10}c_3$$
$$\alpha \leq -1 - \frac{6}{5}c_1 - 3c_2 - \frac{27}{10}c_3,$$

where $c_1, c_2, c_3 \in \mathbb{R}_{\geq 0}$. By solving this linear program, we find that $\sup \alpha = -19/10$ for $c_1 = c_2 = 0$ and $c_3 = 1/3$.

This solution can be verified by computing the lower expectation of f using the credal set derived from $\underline{E}$, that is $\Phi_{\underline{E}}$, which is shown in Figure 2.8$_{\curvearrowleft}$. The credal set $\Phi_{\underline{E}}$ has five extreme points and due to Equation (2.6)$_{47}$, we know that at least one of them yields the lower expectation of f, which we expect to coincide with $\underline{\mathscr{E}}(f)$. Indeed, we find that $\underline{E}_{\Phi_{\underline{E}}}(f) = E_{p^*}(f) = -19/10$, where $p^* = (p^*(a), p^*(b), p^*(c)) = (0, \frac{3}{10}, \frac{7}{10})$. ◊

However, our lower expectation assessment, that is $\underline{E}$, on a domain $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$ might be incoherent. One reason why this may happen is because $\underline{E}$ does not *avoid sure loss* [78, Lemma 2.4.4 (a)], where avoiding sure loss is defined as follows.

$$\underline{E} \text{ avoids sure loss } \Leftrightarrow \sup_{x \in \mathscr{X}} \sum_{i=1}^{n} c_i[f_i(x) - \underline{E}(f_i)] \geq 0, \quad \text{for all } n \in \mathbb{N}, \text{ all } f_i \in \mathscr{K}, \text{ all } c_i \in \mathbb{R}_{\geq 0}. \tag{2.17}$$

If $\underline{E}$ does not avoid sure loss, we say that it *incurs sure loss*. In terms of probability mass functions, incurring sure loss means that the set of probability mass functions $\mathscr{P}_{\underline{E}}$ given by Equation (2.6)$_{51}$ is empty [78, Theorem 3.3.3 (a)]. The following example illustrates such a situation.

Example 6. Consider the set $\mathscr{X} = \{a, b, c\}$ and the following assessments

$$p(a) \geq 3/4,\ p(b) \geq 7/20 \text{ and } p(c) \geq 3/10. \tag{2.18}$$

Alternatively, we can interpret the aforementioned assessments by saying that we have a lower expectation assessment $\underline{E}$ on the domain $\mathscr{K} := \{f_1, f_2, f_3\}$, where

$$f_1 := \mathbb{I}_a \text{ with } \underline{E}(f_1) = 3/4,\ f_2 := \mathbb{I}_b \text{ with } \underline{E}(f_2) = 7/20 \text{ and}$$
$$f_3 := \mathbb{I}_c \text{ with } \underline{E}(f_3) = 3/10.$$

A representation of these assessments is depicted in Figure 2.9$_{\curvearrowright}$. Judging by Figure 2.9$_{\curvearrowright}$, we expect our assessments to incur sure loss since the intersection of the sets of probability mass functions that satisfy the constraints in

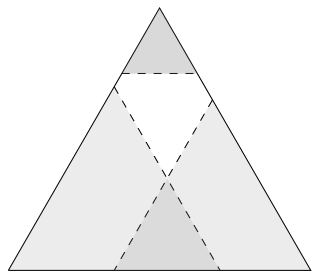

Figure 2.9: The dark grey regions depict the sets of probability mass functions implied by each of the three assessments of Example 6.

Equation (2.18) is the empty set. Indeed, we can verify that from the right-hand side of Equation (2.17). If we compute $\sup_{x\in\mathscr{X}} \sum_{i=1}^{3} c_i[f_i(x) - \underline{E}(f_i)]$ for $c_1 = c_2 = c_3 = 1$, the result is $-7/20$. Therefore, we infer that $\underline{E}$ incurs sure loss. $\Diamond$

Finally, it is possible that our lower expectation assessment is incoherent, but does avoid sure loss. In this case, natural extension corrects the lower expectation assessment upwards, in the sense that it constructs the smallest lower expectation implied by the lower expectation assessment. We illustrate this in the following example.

Example 7. Suppose that we have the set $\mathscr{X} = \{a,b,c\}$ and the following assessments

$$p(a) \in [1/4, 1/2],\ p(b) \in [1/10, 3/10] \text{ and } p(c) \in [3/10, 3/4].$$

Hence, we can say that we have the domain $\mathscr{K} := \{f_1, f_2, f_3, f_4, f_5, f_6\}$ and the following assessments:

$$f_1 := \mathbb{I}_a \text{ with } \underline{E}(f_1) = 1/4,\ f_2 := -\mathbb{I}_a \text{ with } \underline{E}(f_2) = -1/2,$$
$$f_3 := \mathbb{I}_b \text{ with } \underline{E}(f_3) = 1/10,\ f_4 := -\mathbb{I}_b \text{ with } \underline{E}(f_4) = -3/10,$$
$$f_5 := \mathbb{I}_c \text{ with } \underline{E}(f_5) = 3/10 \text{ and } f_6 := -\mathbb{I}_c \text{ with } \underline{E}(f_6) = -3/4.$$

A representation of these assessments is depicted in Figure 2.10.

Since the intersection of the constraints is non-empty, the assessments avoid sure loss. However, $\underline{E}$ is not coherent. Judging by Figure 2.10, we see that $\underline{E}(f_6)$ is inconsistent with the other assessments. We correct $\underline{E}(f_6)$ by computing $\underline{\mathscr{E}}(f_6)$ using the linear program given by (2.16)$_{52}$. Hence, we find that $\underline{\mathscr{E}}(f_6)$ is the supremum value of α satisfying the following system of

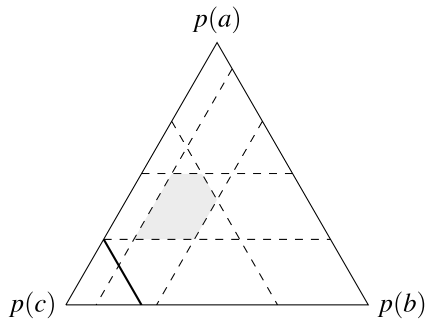

Figure 2.10: The grey area depicts the set of probability mass functions implied by the assessments of Example 7↶. The thick line indicates the assessment $\underline{E}(f_6) = -3/4$ which is inconsistent with the rest of the assessments.

inequalities

$$\begin{aligned}
\alpha &\leq -\frac{3}{4}c_1 + \frac{1}{2}c_2 + \frac{1}{10}c_3 - \frac{3}{10}c_4 + \frac{3}{10}c_5 - \frac{3}{4}c_6 \\
\alpha &\leq +\frac{1}{4}c_1 - \frac{1}{2}c_2 - \frac{9}{10}c_3 + \frac{7}{10}c_4 + \frac{3}{10}c_5 - \frac{3}{4}c_6 \\
\alpha &\leq -1 + \frac{1}{4}c_1 - \frac{1}{2}c_2 + \frac{1}{10}c_3 - \frac{3}{10}c_4 - \frac{7}{10}c_5 + \frac{1}{4}c_6,
\end{aligned}$$

where $c_1, c_2, c_3, c_4, c_5, c_6 \in \mathbb{R}_{\geq 0}$. By solving this linear program, we find that $\sup \alpha = -13/20$ for $c_1 = c_3 = 1$ and $c_2 = c_4 = c_5 = c_6 = 0$. Therefore, we have that $\underline{E}(f_6)$ should be $-13/20$ and not $-3/4$ as we initially assessed. ◊

At this point, we establish the connection between natural extension and the *incomplete* interpretation[8] of lower expectations. We have mentioned that the lower expectation of a gamble $f \in \mathscr{L}(\mathscr{X})$ can be interpreted as the supremum buying price of f, in the sense that we are willing to buy the gamble f for any price strictly lower than $\underline{E}(f)$. However, this interpretation does not say anything about prices $c \in \mathbb{R}$ such that $\underline{E}(f) \leq c$. According to the incomplete interpretation, we might be willing either to accept, or to reject, or even be completely undecided about such prices. Suppose that we are given a lower expectation assessment, i.e. some $\underline{E}$ on a domain $\mathscr{K} \subseteq \mathscr{L}(\mathscr{X})$, and that we reject any price $c > \underline{E}(f)$, for all $f \in \mathscr{K}$. If $\underline{E}$ is not coherent, then we might reject prices that we should also be willing to accept; see Example 7↶. Therefore, if for some $f \in \mathscr{K}$ we have that $\underline{E}(f) \leq \underline{\mathscr{E}}(f)$, we should be willing to accept any c such that $\underline{E}(f) \leq c < \underline{\mathscr{E}}(f)$. If we had decided to reject all prices that are greater that $\underline{E}(f)$, this would not be possible.

[8] Also known as non-exhaustive interpretation.

2.7 UNBOUNDED GAMBLES AND CONDITIONAL MODELS

All of the concepts and results discussed so far apply to real-valued functions on finite spaces, and when the lower expectations that we consider are unconditional. Although the basic ideas remain similar, technicalities arise when we deal with—possibly unbounded—real-valued functions on infinite spaces, with extended real-valued functions, or with conditional operators; see for example Reference [72, Part II] for a very general treatment. Fortunately, this will not be an issue here, because in the context of stochastic processes, these technicalities can be dealt with in a specific way. We will discuss this further in the next chapters.

3

Discrete-time stochastic processes and how to make them imprecise

In the previous chapter, we showed how to model uncertainty regarding a variable X taking values in a finite non-empty space $\mathscr{X}$. Suppose now that we have a sequence of variables $\{X_n\}_{n\in\mathbb{N}}$, where the value of each variable X_n takes values in $\mathscr{X}$. Any such sequence of variables is a stochastic process and our goal in this chapter is to model uncertainty in these kinds of processes.

A typical way to model uncertainty in stochastic processes, which is also the one we adopt here, is through the framework of *measure-theoretic probability*; see Appendix A233 for the basics of measure-theoretic probability. In order to build our probability measures, we first introduce the notion of a probability tree. Probability trees constitute a simple tool used in stochastic processes and we show that from any given probability tree, we can derive a so-called coherent conditional probability. Moreover, we show that our derived coherent conditional probability can be extended to a unique and σ-additive coherent conditional probability measure on the σ-algebra generated by the cylinder events, where the conditioning events are cylinder events. We then use this coherent conditional probability to compute expectations of measurable functions through Lebesgue integrals. Compared to the standard measure-theoretic approach, which derives conditional probabilities from unconditional ones, our approach derives them directly from a given probability tree. One of the advantages of this is that it allows us to define conditional probabilities even when the event on which we condition has zero probability.

Since it is difficult and often unrealistic to know precisely the parameters of a probability tree, we next extend our ideas to sets of probability trees. Any set of probability trees forms a so-called imprecise probability tree. We define global uncertainty models for general imprecise probability trees that are in fact lower and upper expectations of measurable extended real-valued functions. We then focus on imprecise probability trees that are constructed from sets of conditional probability mass functions. By allowing these sets of conditional probability mass functions to be general and not necessarily closed and convex, we also prove a more general law of iterated expectations for lower and upper expectations of real-valued functions that depend on a finite number of variables.

3.1 NOTATION

Before we discuss the main ideas behind stochastic processes, we first provide some useful notation. We consider infinite sequences of variables $\{X_n\}_{n\in\mathbb{N}}$, where each variable X_n is generally called a *state* or more specifically the *state at time n* and takes values in a non-empty finite set $\mathscr{X}$. This yields a stochastic process with *sample space* Ω defined by $\Omega := \mathscr{X}^{\mathbb{N}}$. A generic element of Ω is called a *path* and we denote it by ω. The set $\mathscr{X}$ is called the *state space* of the stochastic process and any element $x \in \mathscr{X}$ is called a *state value*. We also denote by x_n the *state value at time n*.

We now introduce notation for finite sequences of states and state values. For all $n \in \mathbb{N}$, the finite sequence of variables $X_1,\dots,X_n$ is denoted by $X_{1:n}$. For all $m,n \in \mathbb{N}$ with $m \leq n$, the finite sequence of states from time point m up to and including n is denoted by $X_{m:n}$. Any finite sequence $X_{1:n}$ takes values $x_{1:n} := (x_1,\dots,x_n)$ in $\mathscr{X}^n$ and any finite sequence of state values $x_{1:n} \in \mathscr{X}^n$ is called a *situation*. The set of all situations is denoted by $\mathscr{X}^*$ and defined by $\mathscr{X}^* := \{x_{1:n} \in \mathscr{X}^n \colon n \in \mathbb{N}_0\}$. For the special case of $n = 0$, we have the so-called *initial situation*, denoted by $\square$. Therefore, $\mathscr{X}^0 := \{\square\}$ and $x_{1:0}$ is the empty sequence. Moreover, for any path $\omega \in \Omega$, the initial sequence that consists of its first n elements is a situation in $\mathscr{X}^n$, denoted by ω^n, and its n-th element is a state value in $\mathscr{X}$, denoted by ω_n.

We also allow concatenation of situations with state values or sequences of state values. For all $n \in \mathbb{N}_0$, given any situation $x_{1:n} \in \mathscr{X}^n$ and any state value $x_{n+1} \in \mathscr{X}$, we denote their concatenation by $(x_{1:n}, x_{n+1})$, which is a situation in $\mathscr{X}^{n+1}$. Similarly, for all $m,n \in \mathbb{N}_0$ such that $m < n$, given any situation $x_{1:m} \in \mathscr{X}^m$ and any sequence of state values $x_{m+1:n} \in \mathscr{X}^{n-m}$, we denote their concatenation by $(x_{1:m}, x_{m+1:n})$, which is a situation in $\mathscr{X}^n$. Moreover, we allow concatenation of situations with states or sequences of states. For all $n \in \mathbb{N}_0$, given any situation $x_{1:n} \in \mathscr{X}^n$ and a state X_{n+1}, the concatenation of $x_{1:n}$ and X_{n+1} is denoted by $(x_{1:n}, X_{n+1})$. Similarly, for any $m < n$, we denote by $(x_{1:m}, X_{m+1:n})$ the concatenation of situation $x_{1:m}$ with the sequence of states $X_{m+1:n}$. The initial situation works as neutral element in the concatenation, in the sense that $(\square, x_1) = x_1$ and $(\square, x_{1:n}) = x_{1:n}$, for all $x_1 \in \mathscr{X}$ and all $x_{1:n} \in \mathscr{X}^*$.

3.2 EVENT TREES

In the previous chapter, we introduced the concept of an event as a subset of $\mathscr{X}$. From now on, when we refer to events, we mean subsets of Ω. The set of all possible events of Ω is the power set of Ω, denoted by 2^{Ω}. Our main goal is to specify the probabilities of various events in 2^{Ω}. When Ω is finite, we can specify the probabilities of all the elements in Ω and use these to specify the probability of any event in 2^{Ω}. However, in our present case this is not feasible because Ω is uncountably infinite. We can tackle this problem by specifying the probabilities of only some events from which we can then derive a probability measure on the algebra generated by these events and then extend this measure to the σ-algebra generated by this algebra; see Section A.1234 for more details on algebras and σ-algebras. In the rest of this section, we discuss a σ-algebra that is relevant for the stochastic processes that we consider and we leave the analysis about probabilities of events in this σ-algebra for the next section.

Given a sample space Ω, we focus on the algebra and the σ-algebra generated by the set of the so-called *cylinder events*. For all situations $x_{1:n} \in \mathscr{X}^*$, the corresponding cylinder event is the following event:

$$\Gamma(x_{1:n}) := \left\{ \omega \in \Omega \colon \omega^n = x_{1:n} \right\}.$$

Since there is a one-to-one correspondence between situations and cylinder events, we will generally use the notation $x_{1:n}$ for the event $\Gamma(x_{1:n})$, for all $x_{1:n} \in \mathscr{X}^*$. For example, when we define sets of events their elements will be written in terms of situations and we will even use the set of all situations $\mathscr{X}^*$ to refer to the set of all cylinder events. This also applies to probabilities, in the sense that the probability of the event $\Gamma(x_{1:n})$ will be denoted by $P(x_{1:n})$ instead of $P(\Gamma(x_{1:n}))$. Nevertheless, we will still use the notation $\Gamma(x_{1:n})$ when paths are involved, for instance 'for all ω in $\Gamma(x_{1:n})$' and 'the event $\Gamma(\omega^n)$ for some $\omega \in \Omega$'.

By considering complements, finite unions and intersections of events of the form $\Gamma(x_{1:n})$, we can construct the algebra generated by the set of all situations, which is denoted by $\langle \mathscr{X}^* \rangle$. The following result is a property of the events in $\langle \mathscr{X}^* \rangle$ that will be turn out to be useful later on when we build conditional probability measures; see Lemma 1470.

Lemma 6. *Consider any $A \in \langle \mathscr{X}^* \rangle$, then there is some $n \in \mathbb{N}_0$ and $C \subseteq \mathscr{X}^n$such that $A := \cup_{x_{1:n} \in C} \Gamma(x_{1:n})$.*

Proof. We provide a sketch of the proof. For any $m \in \mathbb{N}$ and $x_{1:m}$, the cylinder event $\Gamma(x_{1:m})$ and its complement depend on the first m states $X_{1:m}$. Therefore, all the cylinder events and their complements depend on a finite number of states. Moreover, all finite unions and intersections among the cylinder

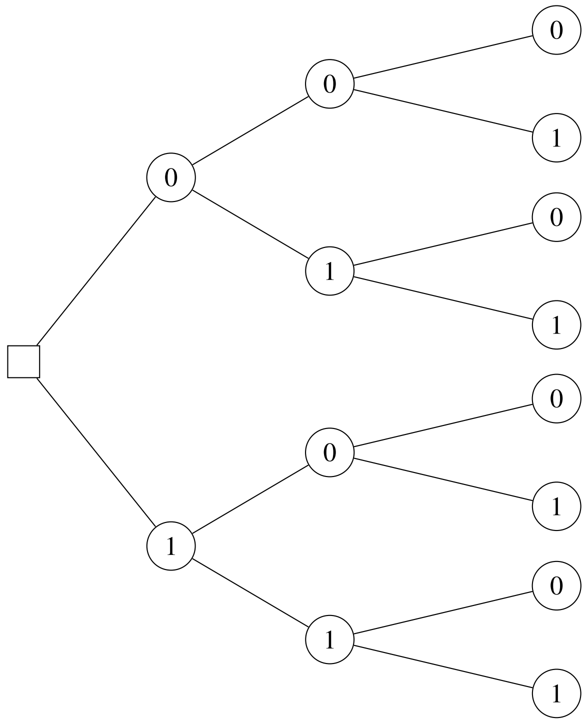

Figure 3.1: The initial part of the event tree of a stochastic process with state space $\mathscr{X} = \{0,1\}$.

events and their complements and the complements of these unions and intersections, will also depend on a finite number of states. Therefore, since $\langle \mathscr{X}^* \rangle$ is constructed from complements, finite unions and intersections of cylinder events, for any $A \in \langle \mathscr{X}^* \rangle$, there is $n \in \mathbb{N}_0$ such that A depends at most on the first n states $X_{1:n}$, which implies that there is some $C \subseteq \mathscr{X}^n$ such that $A := \cup_{x_{1:n} \in C} \Gamma(x_{1:n})$. □

Furthermore, by considering complements, countable unions and intersections we can generate the respective σ-algebra, which is denoted by $\sigma(\mathscr{X}^*)$. More details can be found in Section A.1$_{234}$ and specifically in Example 11$_{234}$, where we discuss the algebra and the σ-algebra generated by the cylinder sets of an infinite sequence of coin tosses.

Finally, we close this section by presenting a graphical representation of our sample space Ω. Since $\Omega = \mathscr{X}^{\mathbb{N}}$ and $\mathscr{X}$ is finite, we can depict our process by means of a tree and this tree is called an *event tree* [64]; see Figure 3.1.

3.3 PROBABILITY TREES

Consider the product set $\mathscr{X} \times \mathscr{X}^*$, whose elements are denoted by $(x_{n+1}|x_{1:n})$ for all $n \in \mathbb{N}_0$, all $x_{1:n} \in \mathscr{X}^n$ and all $x_{n+1} \in \mathscr{X}$. Consider as well a function $p\colon \mathscr{X} \times \mathscr{X}^* \to [0,1]$, such that for all $x_{1:n} \in \mathscr{X}^*$, $p(X_{n+1}|x_{1:n})$ is a probability mass function on $\mathscr{X}$, which we call a *local model*. For all $x_{1:n} \in \mathscr{X}^* \setminus \{\square\}$, the local model $p(X_{n+1}|x_{1:n})$ is called a *transition model* and for all $x_{n+1} \in \mathscr{X}$, the probability $p(x_{n+1}|x_{1:n})$ is called a *transition probability*. The local model $p(X_1|\square)$ is called an *initial model* and is simply denoted by $p(X_1)$. The set of all such functions p is denoted by $\mathbb{P}_{\mathscr{X}^*}$.

The local model $p(X_{n+1}|x_{1:n})$ is interpreted as a representation of some subject's beliefs about what will happen at time $n+1$ given that the process is in situation $x_{1:n}$. Basically, for all $n \in \mathbb{N}_0$, all $x_{1:n} \in \mathscr{X}^n$ and all $x_{n+1} \in \mathscr{X}$, the probability $p(x_{n+1}|x_{1:n})$ is the probability of the event $\Gamma(x_{1:n}, x_{n+1})$ conditional on the event $\Gamma(x_{1:n})$, where we use the notational convention introduced in Section 3.259.

Consider now the corresponding event tree of Ω, where a local model $p(X_{n+1}|x_{1:n})$ is attached to each situation $x_{1:n}$. This turns the event tree into a so-called *probability tree*; see References [64, Chapter 3] and [43, Section 1.9]. An example of a probability tree is depicted in Figure 3.2↷. A probability tree is characterised by the state space and the local models of a stochastic process. Since our sample space is of the form $\Omega = \mathscr{X}^{\mathbb{N}}$, our probability trees are characterised by $\mathscr{X}$ and the local models $p(X_{n+1}|x_{1:n})$, for all $n \in \mathbb{N}_0$ and all $x_{1:n} \in \mathscr{X}^*$. Therefore, a probability tree is fully described by a function p in $\mathbb{P}_{\mathscr{X}^*}$.

Probability trees are a simple tool to model uncertainty in stochastic processes. A well-established family of stochastic process that can be constructed in this way are the so-called *Markov chains*, which are also the ones that we will focus on in this dissertation. We leave the analysis of Markov chains for Chapter 5100 and here discuss general discrete-time stochastic process that are constructed from the local models of a given probability tree.

3.4 CONDITIONAL PROBABILITY MEASURES

We now discuss probabilities of events in $\sigma(\mathscr{X}^*)$ conditional on events in $\langle\mathscr{X}^*\rangle$. In order to determine such probabilities, we need a conditional probability measure. First, we present some preliminaries on conditional probability and then we show how we can define a conditional probability measure based on a given probability tree.

3.4.1 *Preliminaries on conditional probability*

As already mentioned, the set of all possible events of a sample space Ω is 2^Ω. We also let $2^\Omega_\varnothing := 2^\Omega \setminus \{\varnothing\}$. The product set $2^\Omega \times 2^\Omega_\varnothing$ is the complete set

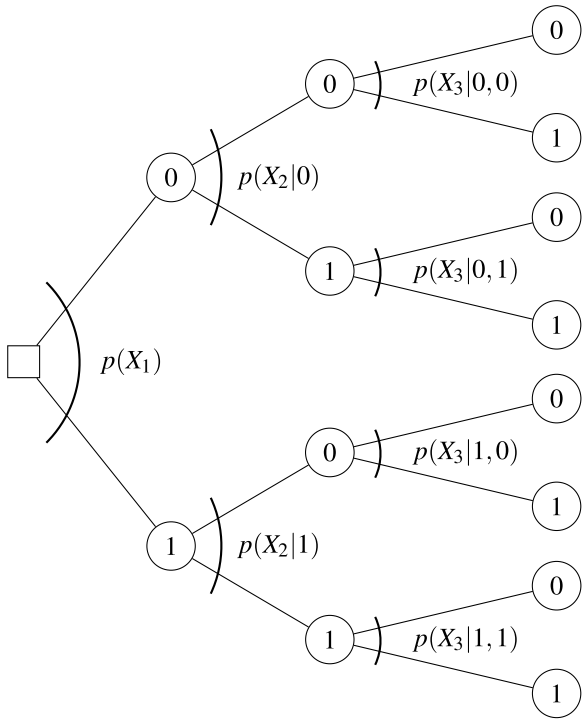

Figure 3.2: The initial part of the probability tree of a stochastic process with state space $\mathscr{X} = \{0,1\}$.

of all *conditional events*. Any element (A,B) of $2^\Omega \times 2^\Omega_\varnothing$, where $A \in 2^\Omega$ and $B \in 2^\Omega_\varnothing$, will be denoted by $(A|B)$ and is interpreted as *the event A conditional on the event B*. The definition of a conditional probability [16, 32] now goes as follows.

Definition 4 (Conditional probability)**.** *Consider any two algebras $\mathscr{A}, \mathscr{B}$ on Ω such that $\mathscr{B} \subseteq \mathscr{A}$ and the set of conditional events $\mathscr{C} := \mathscr{A} \times (\mathscr{B} \setminus \{\varnothing\})$. Then a conditional probability is a function $P\colon \mathscr{C} \to \mathbb{R}$ that, for all $A, C \in \mathscr{A}$ and all $B, D \in \mathscr{B} \setminus \{\varnothing\}$, satisfies the following properties:*

CP1. $P(A|B) \geq 0$*;*

CP2. $P(A|B) = 1$ *if* $B \subseteq A$*;*

CP3. $P(A \cup C|B) = P(A|B) + P(C|B)$ *if* $A \cap C = \varnothing$*;*

CP4. $P(A \cap D|B) = P(A|D \cap B)P(D|B)$ *if* $D \cap B \neq \varnothing$.

In case $\mathscr{B} = \mathscr{A}$, then P is called a full conditional probability *on $\mathscr{A}$. For all $A \in \mathscr{A}$, the unconditional probability $P(A|\Omega)$ is also denoted as $P(A)$.*

From properties CP1$_{\curvearrowleft}$–CP3$_{\curvearrowleft}$, we can easily derive additional properties. For all $A \in \mathscr{A}$ and $B \in \mathscr{B} \setminus \{\varnothing\}$, we have that

CP5. $0 \leq P(A|B) \leq 1$;

CP6. $P(A|B) = P(A \cap B|B)$;

CP7. $P(\varnothing|B) = 0$;

CP8. $P(\Omega|B) = 1$.

There are also other definitions for conditional probability. Perhaps the most common definition is the one introduced by Bayes,[1] who derives conditional probabilities from unconditional ones. According to Bayes' definition, we have that

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ for all } A \in 2^{\Omega} \text{ and all } B \in 2^{\Omega}_{\varnothing}, \text{ with } P(B) \neq 0.$$

Definition 4$_{\curvearrowleft}$ is based on de Finetti, who considers conditional probabilities to be primitive entities that are connected with unconditional probabilities in the following way:

$$P(A \cap B) = P(A|B)P(B), \text{ for all } A \in 2^{\Omega} \text{ and all } B \in 2^{\Omega}_{\varnothing}. \tag{3.1}$$

In fact, Equation (3.1) is a special case of property CP4$_{\curvearrowleft}$ [with $B = \Omega$ and D replaced by B]. Note also that Definition 4$_{\curvearrowleft}$ does not necessarily assume that the probabilities of the events that we condition on are positive. For example, in Equation (3.1), if $P(B) = 0$, then $P(A \cap B) = 0$, but $P(A|B)$ can assume any value in $[0,1]$.

We provide one more definition regarding conditional probability, which is that of a *coherent conditional probability* [8].[2]

Definition 5 (Coherent conditional probability)**.** *Consider any $\mathscr{C} \subseteq 2^{\Omega} \times 2^{\Omega}_{\varnothing}$ and a function $P : \mathscr{C} \to \mathbb{R}$. Then P is a coherent conditional probability if, for all $n \in \mathbb{N}$ and any choice of $A_i|B_i \in \mathscr{C}$ and $c_i \in \mathbb{R}$ for $i \in \{1, \ldots, n\}$, it holds that*

$$\max\left\{\sum_{i=1}^{n} c_i \mathbb{I}_{B_i}(\omega)(P(A_i|B_i) - \mathbb{I}_{A_i}(\omega)) \colon \omega \in B_0\right\} \geq 0,$$

where $B_0 := \cup_{i=1}^{n} B_i$.

[1] Also known as Bayes' rule.

[2] Regarding the definition of coherent conditional probability, some authors—see References [7, 8]—consider a supremum instead of a maximum. However, one can show that the supremum is actually a maximum since it is taken over a finite set of finite values. Moreover, the original definition also requires the infimum to be smaller than or equal to zero but this is implied by the supremum being greater than or equal to zero (it suffices to change the sign of c_i).

The interpretation of Definition $5_{\curvearrowleft}$ is based upon the concepts of supremum buying price and infimum selling price that we introduced in Section 2.4_{44}. In detail, any conditional probability $P(A|B)$ can be interpreted as the price at which are we are willing to either buy or sell the gamble that yields the reward one when the event A occurs and zero otherwise, provided that the event B happens. In case B does not happen, the gamble is called-off.

We see that the set of conditional events $\mathscr{C}$ in Definition $5_{\curvearrowleft}$ need not have any specific structure. However, when $\mathscr{C}$ has a structure as the one presented in Definition 4_{62}, we have the following property.

Theorem 7 ([57, Theorem 3])**.** *Consider any two algebras $\mathscr{A}, \mathscr{B}$ on Ω such that $\mathscr{B} \subseteq \mathscr{A}$ and the set of conditional events $\mathscr{C} := \mathscr{A} \times (\mathscr{B} \setminus \{\varnothing\})$. Then the function $P \colon \mathscr{C} \to \mathbb{R}$ is a conditional probability if and only if P is a coherent conditional probability on $\mathscr{C}$.*

Another useful property of coherent conditional probabilities is that they can be extended coherently to larger domains. The corresponding theorem goes as follows.

Theorem 8 ([57, Theorem 4])**.** *Consider any $\mathscr{C} \subseteq 2^{\Omega} \times 2^{\Omega}_{\varnothing}$ and any coherent conditional probability P on $\mathscr{C}$. Then for any $\mathscr{C}' \subseteq 2^{\Omega} \times 2^{\Omega}_{\varnothing}$ such that $\mathscr{C} \subseteq \mathscr{C}'$, P can be extended to a coherent conditional probability on $\mathscr{C}'$.*

3.4.2 *From probability trees to conditional probability measures*

Before we begin our analysis of the conditional probabilities that are of interest to us, we first introduce the concept of a σ-additive and coherent conditional probability. The definition of a σ-additive coherent conditional probability builds upon the definition of a σ-additive probability measure on an algebra given by Definition 14_{236}—see Appendix A_{233}—and goes as follows [8].

Definition 6 (σ-Additive coherent conditional probability)**.** *Consider any algebra $\mathscr{A}$ on Ω, any set $\mathscr{B}$ such that $\mathscr{B} \subseteq \mathscr{A} \setminus \{\varnothing\}$ and the set of conditional events $\mathscr{C} := \mathscr{A} \times \mathscr{B}$. Consider as well any coherent conditional probability P on $\mathscr{C}$. Then P is σ-additive if for each $B \in \mathscr{B} \setminus \{\varnothing\}$, $P(\cdot|B)$ is a σ-additive probability measure on $\mathscr{A}$.*

Because of Theorem 7, we understand that Definition 6 applies also for conditional probabilities by taking $\mathscr{B}$ to be an algebra such that $\mathscr{B} \subseteq \mathscr{A} \setminus \{\varnothing\}$. From now on, we will refer to σ-additive coherent conditional probabilities as *conditional probability measures*.

We focus on conditional probability measures on the following domain of events:

$$\mathscr{C}_{\sigma} := \Big\{ (A|B) \colon A \in \sigma(\mathscr{X}^*) \text{ and } B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\} \Big\}. \tag{3.2}$$

Starting from a single probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$, we will show that p corresponds to a set of conditional probability measures on $\mathscr{C}_\sigma$, denoted by $\mathbb{P}_p$, where each $P \in \mathbb{P}_p$ satisfies the following equation:

$$P(x_{1:n}, x_{n+1} | x_{1:n}) := p(x_{n+1} | x_{1:n}) \text{ for all } x_{1:n} \in \mathscr{X}^* \text{ and all } x_{n+1} \in \mathscr{X}. \quad (3.3)$$

Consider first the following domain of events:

$$\mathscr{C}_{\mathscr{X}^*} := \Big\{ (x_{1:n}, x_{n+1} | x_{1:n}) \colon x_{1:n} \in \mathscr{X}^* \text{ and } x_{n+1} \in \mathscr{X} \Big\}. \quad (3.4)$$

The following theorem says that from any probability tree, we can derive a unique coherent conditional probability on $\mathscr{C}_{\mathscr{X}^*}$. This property is known when conditional probability measures are derived from unconditional ones, but in our case is more general since we allow the probability of the conditioning event to be zero as well.

Theorem 9. *Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$. Then there is a unique coherent conditional probability P on $\mathscr{C}_{\mathscr{X}^*}$ that satisfies Equation* (3.3).

Proof. Fix any $m \in \mathbb{N}$. Consider any $n \in \mathbb{N}$ and for all $i \in \{1, \ldots, n\}$ consider any choice of $c_i \in \mathbb{R}$ and $A_i | B_i \in \mathscr{C}_{\mathscr{X}^*}$ such that for all $i \in \{1, \ldots, n\}$ there are $m_i \in \mathbb{N}$ and $x^i_{1:m_i} \in \mathscr{X}^{m_i}$ such that $A_i = x^i_{1:m_i}$ and $B_i = x^i_{1:m_i - 1}$, with $m_i \leq m$.

In order to prove that P is a coherent conditional probability, we need to show that

$$\max \left\{ \sum_{i=1}^{n} c_i \mathbb{I}_{B_i}(\omega) [P(A_i | B_i) - \mathbb{I}_{A_i}(\omega)] \colon \omega \in B_0 \right\} \geq 0, \quad (3.5)$$

with $B_0 := \cup_{i=1}^{n} B_i$. We will prove this using complete induction on m.

First, we prove that (3.5) holds for $m = 1$. In this case, it is easy to see that for all $i \in \{1, \ldots, n\}$, we have that $A_i = x^i$ for some $x^i \in \mathscr{X}$ and that $B_i = \Omega$. Let $N_x := \{i \in \{1, \ldots, n\} \colon A_i = x\}$, for all $x \in \mathscr{X}$, and $c_x := \sum_{i \in N_x} c_i$, and consider any $x^* \in \operatorname{argmin}_{x \in \mathscr{X}} c_x$. Then

$$\sum_{i=1}^{n} c_i P(A_i | B_i) = \sum_{i=1}^{n} c_i p(x^i) = \sum_{x \in \mathscr{X}} c_x p(x) \geq \sum_{x \in \mathscr{X}} c_{x^*} p(x) = c_{x^*} \sum_{x \in \mathscr{X}} p(x) = c_{x^*},$$

where we have used that $p(\cdot)$ is a probability mass function on $\mathscr{X}$. Choose now any $\omega' \in \Omega$ such that $\omega'_1 = x^*$, then

$$\begin{aligned} \sum_{i=1}^{n} c_i \mathbb{I}_{B_i}(\omega') [P(A_i | B_i) - \mathbb{I}_{A_i}(\omega')] &= \sum_{x \in \mathscr{X}} c_x [p(x) - \mathbb{I}_x(\omega')] \\ &\geq c_{x^*} - \sum_{x \in \mathscr{X}} c_x \mathbb{I}_x(\omega') = c_{x^*} - c_{x^*} = 0, \end{aligned}$$

where the first equality holds because $B_i = B_0 = \Omega$ and therefore $\mathbb{I}_{B_i}(\omega') = 1$ for all $i \in \{1,\ldots,n\}$, and the last equality holds because $\mathbb{I}_x(\omega') = 1$ for $x = x^*$ and $\mathbb{I}_x(\omega') = 0$ otherwise. Therefore, we see that (3.5)$_{\curvearrowleft}$ is satisfied.

Next, we assume that (3.5)$_{\curvearrowleft}$ holds for all $m \leq k$, and prove that it also holds for $m = k+1$. Let $S = \big\{i \in \{1,\ldots,n\} \colon m_i \leq k\big\}$. Then there are two possibilities. If $S \neq \varnothing$, then by the induction hypothesis, we know that

$$\max\left\{\sum_{i\in S} c_i \mathbb{I}_{B_i}(\omega)[P(A_i|B_i) - \mathbb{I}_{A_i}(\omega)] \colon \omega \in B_0^*\right\} \geq 0,$$

with $B_0^* \coloneqq \cup_{i\in S} B_i$. This implies that there is some $\omega^* \in B_0^* \subseteq B_0$ such that

$$\sum_{i\in S} c_i \mathbb{I}_{B_i}(\omega^*)[P(A_i|B_i) - \mathbb{I}_{A_i}(\omega^*)] \geq 0. \tag{3.6}$$

If $S = \varnothing$, then we let ω^* be any path in B_0.

Now let $S^* \coloneqq \{i \in \{1,\ldots,n\} \colon B_i = (\omega^*)^k\}$, $N_x^* \coloneqq \{i \in S^* \colon A_i = ((\omega^*)^k, x)\}$ and $c_x \coloneqq \sum_{i\in N_x^*} c_i$ for all $x \in \mathscr{X}$, and consider any $x^* \in \operatorname{argmin}_{x\in\mathscr{X}} c_x$. Then we find that

$$\sum_{i\in S^*} c_i P(A_i|B_i) = \sum_{x\in\mathscr{X}} c_x p(x|(\omega^*)^k) \geq \sum_{x\in\mathscr{X}} c_{x^*} p(x|(\omega^*)^k) = c_{x^*}, \tag{3.7}$$

where we have used that $p(\cdot|(\omega^*)^k)$ is a probability mass function on $\mathscr{X}$. If we now choose any $\omega^{**} \in B^*$ such that $(\omega^{**})^{k+1} = ((\omega^*)^k, x^*)$, then

$$\begin{aligned}\sum_{i\in S^*} c_i[P(A_i|B_i) - \mathbb{I}_{A_i}(\omega^{**})] &= \sum_{x\in\mathscr{X}} c_x[p(x|(\omega^*)^k) - \mathbb{I}_{((\omega^*)^k,x)}(\omega^{**})] \\ &\geq c_{x^*} - \sum_{x\in\mathscr{X}} c_x \mathbb{I}_{((\omega^*)^k,x)}(\omega^{**}) = c_{x^*} - c_{x^*} = 0,\end{aligned} \tag{3.8}$$

where the inequality comes from (3.7) and the last equality holds because $\mathbb{I}_{((\omega^*)^k,x)}(\omega^{**}) = 1$ when $x = x^*$ and $\mathbb{I}_{((\omega^*)^k,x)}(\omega^{**}) = 0$ otherwise.

Let $S^{**} \coloneqq \{1,\ldots,n\} \setminus (S \cup S^*)$. Since $(\omega^{**})^{k+1} = ((\omega^*)^k, x^*)$, we know that $\mathbb{I}_{B_i}(\omega^{**}) = \mathbb{I}_{B_i}(\omega^*)$ and that $\mathbb{I}_{A_i}(\omega^{**}) = \mathbb{I}_{A_i}(\omega^*)$ for all $i \in S$. Also, we know that $\mathbb{I}_{B_i}(\omega^{**}) = 1$ for all $i \in S^*$ and that $\mathbb{I}_{B_i}(\omega^{**}) = 0$ for all $i \in S^{**}$. Therefore, by combining (3.6) with (3.8) and the fact that $\{1,\ldots,n\} = S \cup S^* \cup S^{**}$ we find that

$$\sum_{i=1}^{n} c_i \mathbb{I}_{B_i}(\omega^{**})[P(A_i|B_i) - \mathbb{I}_{A_i}(\omega^{**})] \geq 0.$$

It now only remains to show that ω^{**} belongs to B_0. Recall that $\omega^* \in B_0$. Since B_0 is an event that only depends on the first k states, and because $(\omega^*)^k = (\omega^{**})^k$, this implies that indeed $\omega^{**} \in B_0$.

Finally, the uniqueness of P follows from Equation (3.3)$_{\curvearrowleft}$ and the fact that there is a one-to-one correspondence between $\mathscr{C}_{\mathscr{X}^*}$ and $\mathscr{X} \times \mathscr{X}^*$. $\square$

From now on, we will refer to the coherent conditional probability of Theorem 9_{65} as *the* coherent conditional probability on $\mathscr{C}_{\mathscr{X}^*}$ derived from a probability tree p. In order to extend this unique coherent conditional probability to a conditional probability measure on $\mathscr{C}_\sigma$, we first show that it can be extended to a conditional probability measure on the following domain of conditional events:

$$\mathscr{C} := \Big\{(A|B) \colon A \in \langle \mathscr{X}^* \rangle \text{ and } B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\}\Big\}. \tag{3.9}$$

The corresponding lemma goes as follows.

Lemma 10. *Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$. Then there is a conditional probability measure P on $\mathscr{C}$ that satisfies Equation* $(3.3)_{65}$.

Proof. It follows from Theorem 9_{65} that there is a unique coherent conditional probability P^* on $\mathscr{C}_{\mathscr{X}^*}$ that satisfies Equation $(3.3)_{65}$. It then follows from Theorem 8_{64} that P^* can be extended to a coherent conditional probability P on $\mathscr{C}$. Since P^* is the restriction of P to $\mathscr{C}_{\mathscr{X}^*}$, we infer that P satisfies Equation $(3.3)_{65}$. Moreover, due to the structure of $\mathscr{C}$, it follows from Theorem 7_{64} that P is also a full conditional probability.

It now remains to prove that P is σ-additive. Due to property CP3_{62}, we have that $P(\cdot|B)$ is finitely additive on $\langle \mathscr{X}^* \rangle$ for each $B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\}$, and therefore, due to Theorem 95_{237}, we have that P is indeed σ-additive. □

Note that the conditional probability measure of Lemma 10 might not be unique. Hence, we understand that any probability tree corresponds to a set of conditional probability measures on $\mathscr{C}$ and it now remains to show that any conditional probability measure in this set can be extended to a conditional probability measure on $\mathscr{C}_\sigma$. In general, the extension of a conditional probability measure might not be σ-additive; see Reference [8]. Fortunately, in Reference [8] it is shown that when the domain of conditional events satisfies certain conditions, we can then extend a conditional probability measure and preserve σ-additivity. The corresponding theorem goes as follows.

Theorem 11 ([8, Theorem 2]). *Consider any algebra $\mathscr{A}$ on Ω, any set $\mathscr{B}$ such that $\mathscr{B} \subseteq \mathscr{A} \setminus \{\varnothing\}$ and the set of conditional events $\mathscr{C} := \mathscr{A} \times \mathscr{B}$. Consider as well any σ-additive and coherent conditional probability P' on $\mathscr{C}$, then there is a σ-additive and coherent extension P of P' to the domain $\mathscr{C}_\sigma := \sigma(\mathscr{A}) \times \mathscr{B}$.*

By taking $\mathscr{B} \cup \{\varnothing\}$ to be an algebra, we find that Theorem 11 applies also for conditional probabilities. Therefore, we have the following result.

Theorem 12. *Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$. Then there is a conditional probability measure P on $\mathscr{C}_\sigma$ that satisfies Equation* $(3.3)_{65}$.

Proof. It follows from Lemma 10 that there is a conditional probability measure P' on $\mathscr{C}$ that satisfies Equation $(3.3)_{65}$. Due to the structure of $\mathscr{C}$ and

Theorem $11_{\curvearrowleft}$, we know that P' can be extended to a conditional probability measure P on $\mathscr{C}_\sigma$. □

In summary, we started from a probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ and we constructed a unique coherent conditional probability on the domain $\mathscr{C}_{\mathscr{X}^*}$. Then we extended this coherent conditional probability to a conditional probability measure on $\mathscr{C}$ and we showed that the latter can be further extended to the domain $\mathscr{C}_\sigma$. Since the second of these three extensions might not be unique, we can only infer at this points that p corresponds to a set $\mathbb{P}_p$ of conditional probability measures on $\mathscr{C}_\sigma$.

3.4.3 *Properties of conditional probability measures*

Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ and its corresponding set of conditional probability measures $\mathbb{P}_p$ on $\mathscr{C}_\sigma$. We know from Equation $(3.2)_{64}$ that for all $(A|B) \in \mathscr{C}_\sigma$, we have that $B \in \langle \mathscr{X}^* \rangle \setminus \{\emptyset\}$, but from now on we will mainly be interested in conditional events $(A|B)$ such that $B \in \mathscr{X}^* \cup \mathscr{B}_{\mathscr{X}}$, where $\mathscr{B}_{\mathscr{X}}$ is defined as follows

$$\mathscr{B}_{\mathscr{X}} := \left\{ \bigcup_{x_{1:m} \in \mathscr{X}^m} \Gamma(x_{1:m}, x_{m+1}) \colon m \in \mathbb{N}_0 \text{ and } x_{m+1} \in \mathscr{X} \right\}. \tag{3.10}$$

For any event $B \in \mathscr{B}_{\mathscr{X}}$, there is some $m \in \mathbb{N}_0$ and some $x_{m+1} \in \mathscr{X}$ such that $B = \bigcup_{x_{1:m} \in \mathscr{X}^m} \Gamma(x_{1:m}, x_{m+1})$ and we will therefore denote this event B by $\mathscr{B}_{x_{m+1}}$. For any $(A|B) \in \mathscr{C}$ such that $B \in \mathscr{B}_{\mathscr{X}}$, we have that $B = \mathscr{B}_{x_{m+1}}$ for some $m \in \mathbb{N}_0$ and some $x_{m+1} \in \mathscr{X}$ and the probability $P(A|B)$ will then be denoted simply by $P(A|x_{m+1})$.

We now prove a property of probabilities of the form $P(A|x_{m+1})$. If we had used Bayes' rule to define these probabilities, then proving this result would be trivial. However, such a proof would also require the probabilities of the involved conditioning events to be strictly positive, which our approach does not impose. For that reason, the result below is not immediate, which is why we provide it with a proof.

Lemma 13. *Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ and any conditional probability measure P in $\mathbb{P}_p$. Consider as well any $n, m \in \mathbb{N}_0$, any $x_{m+1} \in \mathscr{X}$ and for all $i \in \{1, \dots, n\}$ consider any choice of $A_i \in \sigma(\mathscr{X}^*)$ and $c_i \in \mathbb{R}$. Then*

$$\begin{aligned} \min_{x_{1:m} \in \mathscr{X}^m} \sum_{i=1}^n c_i P(A_i|x_{1:m}, x_{m+1}) &\leq \sum_{i=1}^n c_i P(A_i|x_{m+1}) \\ &\leq \max_{x_{1:m} \in \mathscr{X}^m} \sum_{i=1}^n c_i P(A_i|x_{1:m}, x_{m+1}). \end{aligned}$$

Proof. We prove by contradiction that the left-hand side inequality holds; the proof for the right-hand side one is completely analogous. That is, we assume ex absurdo that $\sum_{i=1}^{n} c_i P(A_i|x_{m+1}) < \min_{x_{1:m}\in\mathscr{X}^m}\{\sum_{i=1}^{n} c_i P(A_i|x_{1:m},x_{m+1})\}$, and we show that P is not coherent. Specifically, we show that then

$$\max\left\{\sum_{i=1}^{n} c_i \mathbb{I}_{\mathscr{B}_{x_{m+1}}}(\omega)[P(A_i|x_{m+1}) - \mathbb{I}_{A_i}(\omega)] - \sum_{x_{1:m}\in\mathscr{X}^m}\sum_{i=1}^{n} c_i \mathbb{I}_{\Gamma(x_{1:m},x_{m+1})}(\omega)[P(A_i|x_{1:m},x_{m+1}) - \mathbb{I}_{A_i}(\omega)] \colon \omega\in B_0\right\} < 0, \tag{3.11}$$

where, going back to the notation of Definition 5_{63}, for all $x_{1:m}\in\mathscr{X}^m$ and all $i\in\{1,\dots,n\}$ we chose the coefficient of $P(A_i|x_{1:m},x_{m+1})$ to be $-c_i$ and due the definition of $\mathscr{B}_{x_{m+1}}$, we have that $B_0 = \mathscr{B}_{x_{m+1}}$. Indeed, consider any $\omega\in B_0 = \mathscr{B}_{x_{m+1}}$, then

$$\sum_{x_{1:m}\in\mathscr{X}^m} \mathbb{I}_{\Gamma(x_{1:m},x_{m+1})}(\omega) = \mathbb{I}_{\mathscr{B}_{x_{m+1}}}(\omega) = 1 \tag{3.12}$$

and also

$$\sum_{x_{1:m}\in\mathscr{X}^m} \mathbb{I}_{\Gamma(x_{1:m},x_{m+1})}(\omega)\sum_{i=1}^{n} c_i P(A_i|x_{1:m},x_{m+1}) \geq \min_{x_{1:m}\in\mathscr{X}^m}\sum_{i=1}^{n} c_i P(A_i|x_{1:m},x_{m+1}). \tag{3.13}$$

Hence

$$\begin{aligned}
&\sum_{i=1}^{n} c_i \mathbb{I}_{\mathscr{B}_{x_{m+1}}}(\omega)[P(A_i|x_{m+1}) - \mathbb{I}_{A_i}(\omega)] \\
&\qquad - \sum_{x_{1:m}\in\mathscr{X}^m}\sum_{i=1}^{n} c_i \mathbb{I}_{\Gamma(x_{1:m},x_{m+1})}(\omega)[P(A_i|x_{1:m},x_{m+1}) - \mathbb{I}_{A_i}(\omega)] \\
&\quad = \sum_{i=1}^{n} c_i P(A_i|x_{m+1}) - \sum_{i=1}^{n} c_i \mathbb{I}_{A_i}(\omega) + \sum_{i=1}^{n} c_i \mathbb{I}_{A_i}(\omega)\sum_{x_{1:m}\in\mathscr{X}^m}\mathbb{I}_{\Gamma(x_{1:m},x_{m+1})}(\omega) \\
&\qquad - \sum_{x_{1:m}\in\mathscr{X}^m}\mathbb{I}_{\Gamma(x_{1:m},x_{m+1})}(\omega)\sum_{i=1}^{n} c_i P(A_i|x_{1:m},x_{m+1}) \\
&\quad = \sum_{i=1}^{n} c_i P(A_i|x_{m+1}) - \sum_{x_{1:m}\in\mathscr{X}^m}\mathbb{I}_{\Gamma(x_{1:m},x_{m+1})}(\omega)\sum_{i=1}^{n} c_i P(A_i|x_{1:m},x_{m+1}) \\
&\quad \leq \sum_{i=1}^{n} c_i P(A_i|x_{m+1}) - \min_{x_{1:m}\in\mathscr{X}^m}\sum_{i=1}^{n} c_i P(A_i|x_{1:m},x_{m+1}) < 0,
\end{aligned}$$

where the two equalities follow from Equation (3.12), the first inequality from (3.13), and the strict inequality from the assumption. This implies that (3.11) holds, contradicting the coherence of P. □

Lemma 13_{68} will turn out to be useful in Chapter 5_{100}, where we discuss Markov chains. More specifically, we will see that when a stochastic process is a Markov process, then in some cases the inequalities in Lemma 13_{68} become equalities.

In the rest of this section, we focus on conditional events $(A|B) \in \mathscr{C}$ such that $B \in \mathscr{X}^*$. Recall that the coherent conditional probability on $\mathscr{C}_{\mathscr{X}^*}$ derived from a probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ can be extended to conditional probability measures on domains $\mathscr{C}$ and $\mathscr{C}_\sigma$, but that these extensions might not be unique. Fortunately, the extension of a coherent conditional probability on $\mathscr{C}_{\mathscr{X}^*}$ to the following domain:

$$\mathscr{C}^* := \left\{A|B \colon A \in \langle \mathscr{X}^* \rangle \text{ and } B \in \mathscr{X}^*\right\} \tag{3.14}$$

is unique. This is not surprising when conditional probability measures are derived from unconditional ones, for example using Bayes' rule. However, the property that we prove is more general since our conditional probability measures are also defined when the conditioning events have zero probability. The corresponding lemma goes as follows.

Lemma 14. *Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$. Then there is a unique conditional probability measure P on $\mathscr{C}^*$ that satisfies Equation* $(3.3)_{65}$. *Specifically, for any $x_{1:m} \in \mathscr{X}^*$, any $n \in \mathbb{N}_0$ and any $C \subseteq \mathscr{X}^n$, we have that*[3]

$$P(C|x_{1:m}) := \sum_{z_{1:n} \in C} P(z_{1:n}|x_{1:m}),$$

where

$$P(z_{1:n}|x_{1:m}) := \begin{cases} \prod_{i=m}^{n-1} p(z_{i+1}|z_{1:i}) & \text{if } m < n \text{ and } z_{1:m} = x_{1:m} \\ 1 & \text{if } m \geq n \text{ and } z_{1:n} = x_{1:n} \\ 0 & \text{otherwise.} \end{cases}$$

Proof. It follows from Theorem 12_{67} that there is at least one conditional probability measure on $\mathscr{C}_\sigma$ that satisfies Equation $(3.3)_{65}$. Recall that we let $\mathbb{P}_p$ be the set of all such conditional probability measures on $\mathscr{C}_\sigma$. We prove the statement by showing that the restriction to $\mathscr{C}^*$ is the same for all $P \in \mathbb{P}_p$.

Since P is a coherent conditional probability on $\mathscr{C}_\sigma$ and therefore on $\mathscr{C} \subseteq \mathscr{C}_\sigma$, it follows from Theorem 7_{64} that the restriction of P to $\mathscr{C}$ is a conditional probability, and therefore satisfies the properties CP1_{62}–CP8_{63}. We will use these to calculate the values of P on $\mathscr{C}^* \subseteq \mathscr{C}$.

We start by finding the probability of $z_{1:n}$ conditional on $x_{1:m}$, which is denoted by $P(z_{1:n}|x_{1:m})$, for all $z_{1:n} \in \mathscr{X}^*$. We distinguish among three cases.

[3] Due to Lemma 6_{59}, any event $A \in \langle \mathscr{X}^* \rangle$ can be expressed by $\cup_{z_{1:n} \in C} \Gamma(z_{1:n})$ for some $n \in \mathbb{N}_0$ and $C \subseteq \mathscr{X}^n$.

The first case is when $m \geq n$ and $z_{1:n} = x_{1:n}$, which implies that $\Gamma(x_{1:m}) \subseteq \Gamma(z_{1:n})$ and hence, due to property CP2$_{62}$ [with $A := \Gamma(z_{1:n})$ and $B := \Gamma(z_{1:m})$], we find that $P(z_{1:n}|x_{1:m}) = 1$.

The second case is when $m < n$ and $x_{1:m} = z_{1:m}$. Then

$$P(z_{1:n}|z_{1:m}) = P(z_{1:n}|z_{1:n-1})P(z_{1:n-1}|z_{1:m})$$

where the equality follows from property CP4$_{62}$ [with $A := \Gamma(z_{1:n})$, $B := \Gamma(z_{1:m})$ and $D := \Gamma(z_{1:n-1})$] combined with the fact that $\Gamma(z_{1:n}) = \Gamma(z_{1:n}) \cap \Gamma(z_{1:n-1})$ and that $\Gamma(z_{1:n-1}) = \Gamma(z_{1:n-1}) \cap \Gamma(z_{1:m})$. For all $k \in \{m+2, \ldots, n\}$, the probability $P(z_{1:k-1}|z_{1:m})$ is found similarly and for all $\ell \in \{1, \ldots, m+1\}$, we know that $P(z_{1:\ell-1}|z_{1:m}) = 1$, and finally $P(z_{1:n}|x_{1:m})$ is given by

$$P(z_{1:n}|x_{1:m}) = P(z_{1:n}|z_{1:m}) = \prod_{i=m}^{n-1} P(z_{1:i+1}|z_{1:i}) = \prod_{i=m}^{n-1} p(z_{i+1}|z_{1:i}),$$

where the first equality holds because $x_{1:m} = z_{1:m}$ and the last follows because P satisfies Equation (3.3)$_{65}$.

In all other cases, we have that $\Gamma(z_{1:n}) \cap \Gamma(x_{1:m}) = \varnothing$. It then follows from property CP7$_{63}$ [with $B := \Gamma(z_{1:m})$] that $P(z_{1:n}|x_{1:m}) = 0$.

In now remains to find the conditional probability of C conditional on $x_{1:m}$. Since C can be decomposed into disjoint cylinder events it follows from property CP3$_{62}$ that

$$P(C|x_{1:m}) = \sum_{z_{1:n} \in C} P(z_{1:n}|x_{1:m}),$$

which completes the proof. □

Finally, the coherent conditional probability on $\mathscr{C}_{\mathscr{X}^*}$ derived from a probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ can be extended uniquely to the following domain:

$$\mathscr{C}^*_\sigma := \left\{A|B\colon A \in \sigma(\mathscr{X}^*) \text{ and } B \in \mathscr{X}^*\right\}. \tag{3.15}$$

The corresponding theorem goes as follows.

Theorem 15. *Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$. Then there is a unique conditional probability measure P on $\mathscr{C}^*_\sigma$ that satisfies Equation* (3.3)$_{65}$.

Proof. It follows from Lemma 10$_{67}$ that there is a conditional probability measure on $\mathscr{C}$ that satisfies Equation (3.3)$_{65}$. Let $\mathbb{P}'_p$ be the set of all such conditional probability measures on $\mathscr{C}$. It then follows from Lemma 14$_\curvearrowleft$ that for all $P' \in \mathbb{P}'_p$, the restriction to $\mathscr{C}^*$ is unique. Let P^* be this unique restriction on $\mathscr{C}^*$. Since P^* is the restriction of conditional probability measures, it will also be σ-additive. Moreover, due to Theorem 11$_{67}$ and the structure of $\mathscr{C}^*$, we find that P^* can be extended to a conditional probability measure P on $\mathscr{C}^*_\sigma$.

It now remains to show that P is unique. For each $x_{1:n} \in \mathscr{X}^*$, $P^*(\cdot|x_{1:n})$ is a unique σ-additive probability measure since it is the restriction of P^* to the domain $\{(A|x_{1:n})\colon A \in \langle \mathscr{X}^* \rangle\}$. It then follows from Carathéodory's Theorem (Theorem 96_{237}) that $P^*(\cdot|x_{1:n})$ can be uniquely extended to a σ-additive probability measure on the domain $\{(A|x_{1:n})\colon A \in \sigma(\mathscr{X}^*)\}$. Since this is true for every $x_{1:n} \in \mathscr{X}^*$, we conclude that P is unique. □

We now close this section by establishing the connection between conditional and unconditional probability measures. In fact, for any probability tree in $p \in \mathbb{P}_{\mathscr{X}^*}$, any conditional probability measure $P \in \mathbb{P}_p$ on $\mathscr{C}_\sigma$ allows us to specify the probability of any event in $\sigma(\mathscr{X}^*)$. Indeed, for all $A \in \sigma(\mathscr{X}^*)$, the probability of A, denoted by $P(A)$, is the probability of A conditional on the set of all paths, i.e. $P(A|\square)$ since $\Gamma(\square) = \Omega$. Furthermore, since $\square \in \mathscr{X}^*$, we know from Theorem $15_{\curvearrowleft}$ that $P(A)$ is unique. Note that it is also possible to define an unconditional probability measure on $\sigma(\mathscr{X}^*)$ based on the local models of a probability tree—without defining a conditional probability measure—by using the Ionescu Tulcea Theorem; see Reference [41] for the original result, but also References [4, 67]. The disadvantage of that approach, though, is that conditional probabilities become ill-defined if the conditioning event has probability zero.

3.5 EXPECTATIONS IN PROBABILITY TREES

So far, we have seen that with a given a probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$, we can associate a set $\mathbb{P}_p$ of conditional probability measures on the domain $\mathscr{C}_\sigma$. In this section, we use such a conditional probability measure $P \in \mathbb{P}_p$ to compute expectations of measurable extended real-valued functions, i.e. measurable functions that take values in $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$. We first talk about unconditional expectations and then we move to conditional ones. We close the section by introducing the law of iterated expectations, which we will generalise later on, as this generalisation will turn out to be very useful for the computation of lower and upper expectations for imprecise stochastic processes.

3.5.1 *Unconditional Expectations*

Recall that for any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$, the conditional probability measures $P \in \mathbb{P}_p$ on $\mathscr{C}_\sigma$ that we considered above allow us to specify the probability of any event in $\sigma(\mathscr{X}^*)$ by conditioning on the initial situation, that is by considering the unique σ-additive probability measure $P(\cdot|\square)$, which we will also denote by P. This implies that we have a probability space $(\Omega, \sigma(\mathscr{X}^*), P)$ and we can use this probability space to compute expectations. For any extended real-valued function g on Ω such that g is measurable with respect to $\sigma(\mathscr{X}^*)$, we can compute the expectation of g with respect to P by computing the Lebesgue integral of g, given that the integral is defined; see Section A.4_{239} for more details and also Step 1_{241}–Step 4_{241} for an intuitive explanation why expectations of measurable functions are defined by Lebesgue integrals. This expectation of g is denoted by $E_P(g)$ and is given by

$$E_P(g) := \int_\Omega g(\omega)\mathrm{d}P(\omega).$$

We now show how to compute expectations of a special type of extended real-valued functions, the so-called *n-measurable functions*. Before we define n-measurable functions, we first introduce the set of all real-valued functions on $\mathscr{X}^n$, for some $n \in \mathbb{N}_0$,[4] which we denote by $\mathscr{L}(\mathscr{X}^n)$. An n-measurable function g is then a real-valued function on Ω such that there is some $h \in \mathscr{L}(\mathscr{X}^n)$ such that $g(\omega) = h(\omega^n)$ for all $\omega \in \Omega$. Therefore, any n-measurable function can be identified with an element of $\mathscr{L}(\mathscr{X}^n)$, and from now on, we will use $\mathscr{L}(\mathscr{X}^n)$ to denote the set of all n-measurable functions. In order to emphasise this, we will often denote an n-measurable function as $h(X_{1:n})$. Furthermore, n-measurable functions are also a subset of a class of functions, the so-called simple functions (Definition 18_{239}) and they satisfy various useful properties like the one presented below.

Lemma 16. *Consider any $n \in \mathbb{N}_0$ and any $h \in \mathscr{L}(\mathscr{X}^n)$, then h is measurable with respect to $\sigma(\mathscr{X}^*)$.*

Proof. Consider any $n \in \mathbb{N}_0$. Due to Lemma 99_{238}, we then have that for all $x_{1:n} \in \mathscr{X}^n$, the function $\mathbb{I}_{\Gamma(x_{1:n})}$ on Ω is measurable with respect to $\sigma(\mathscr{X}^*)$. Moreover, any $h \in \mathscr{L}(\mathscr{X}^n)$ can be written as

$$h = \sum_{x_{1:n} \in \mathscr{X}^n} h(x_{1:n})\mathbb{I}_{\Gamma(x_{1:n})}, \tag{3.16}$$

so h is simple, which, due to Lemma 100_{239}, implies that h is measurable with respect to $\sigma(\mathscr{X}^*)$. □

For all $n \in \mathbb{N}$ and all n-measurable functions $h \in \mathscr{L}(\mathscr{X}^n)$, the expectation of $h(X_{1:n})$ with respect to P, denoted by $E_P(h(X_{1:n}))$, is the Lebesgue integral of $h(X_{1:n})$, which in this specific case is given by

$$E_P\big(h(X_{1:n})\big) = \sum_{x_{1:n} \in \mathscr{X}^n} h(x_{1:n})P\big(x_{1:n}\big) = \sum_{x_{1:n} \in \mathscr{X}^n} h(x_{1:n}) \prod_{i=0}^{n-1} p(x_{i+1}|x_{1:i}),$$

where the first equality follows from Lemma 102_{240}, Lemma 105_{240} and Equation (3.16) and the second equality follows from the combination of Theorem 12_{67} with Lemma 14_{70}.

Finally, for any non-negative extended real-valued g on Ω and any non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n \in \mathbb{N}}$ such

[4] For $n = 0$, we have constant functions.

that $\lim_{n\to+\infty} h_n = g$, it follows from Theorem 104_{240} that

$$E_P(g) = \lim_{n\to+\infty} E_P(h_n).$$

Compared to general extended measurable real-valued functions, whose expectations are not always easily computable, there are examples of functions that are limits of non-decreasing sequences of non-negative n-measurable functions, whose expectations can be efficiently computed when the stochastic process is a time-homogeneous Markov chain. Such a function is the first passage time that is studied in Chapter 6_{151} and we will see that we can compute its expectation through closed-form expressions.

Observe that the expectations $E_P(h(X_{1:n}))$ and $E_P(g)$ do not depend on the specific choice of $P \in \mathbb{P}_p$ since in these cases P is restricted to $\mathscr{C}_\sigma^*$, and is therefore completely determined by p. This implies that any expectation with respect to a measure P on $\mathscr{C}_\sigma^*$ can be denoted by E_p instead of E_P. However, we prefer to stick to the notation E_P because, as we will see in Chapter 5_{100}, we also consider expectations with respect to conditional probability measures on $\mathscr{C}_\sigma$ and in this case E_p might not be able to determine E_P for all $P \in \mathbb{P}_p$.

3.5.2 *Conditional Expectations*

Regarding conditional expectations, things are very similar to the unconditional ones. We follow exactly the same reasoning, but this time the probability measure $P(\cdot|\square)$ is replaced by a probability measure $P(\cdot|B)$, for some $B \in \langle \mathscr{X}^* \rangle \setminus \{\emptyset\}$. For any $B \in \langle \mathscr{X}^* \rangle \setminus \{\emptyset\}$ and any extended real-valued function g on Ω such that g is measurable with respect to $\sigma(\mathscr{X}^*)$, the expectation of g with respect to P conditional on B is then denoted by $E_P(g|B)$ and is given by

$$E_P(g|B) := \int_\Omega g(\omega) \mathrm{d}P(\omega|B), \tag{3.17}$$

where the integral on the right-hand side is the Lebesgue integral. The reason why this indeed works is because $P(\cdot|B)$ is a σ-additive probability measure on $\sigma(\mathscr{X}^*)$; see Definition 6_{64}.

In the case of n-measurable functions, we have that for all $n, m \in \mathbb{N}_0$, all $x_{1:m} \in \mathscr{X}^m$ and all $h \in \mathscr{L}(\mathscr{X}^n)$, the expectation of h with respect to P conditional on $x_{1:m}$ is denoted by $E_P\big(h(X_{1:n})|x_{1:m}\big)$ and is given by

$$E_P\big(h(X_{1:n})|x_{1:m}\big) = \sum_{z_{1:n}\in\mathscr{X}^n} h(z_{1:n}) P\big(z_{1:n}|x_{1:m}\big).$$

In case $n > m$, then due to Theorem 12_{67} in combination with Lemma 14_{70}, we further find that

$$E_P\big(h(X_{1:n})|x_{1:m}\big) = \sum_{x_{m+1:n}\in\mathscr{X}^{n-m}} h(x_{1:n}) \prod_{i=m}^{n-1} p(x_{i+1}|x_{1:i}) \tag{3.18}$$

and if $n \leq m$, we have that $E_P(h(X_{1:n})|x_{1:m}) = h(x_{1:n})$.

Finally, for any $B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\}$, any non-negative extended real-valued g on Ω and any non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} h_n = g$, it holds that

$$E_P(g|B) = \lim_{n\to+\infty} E_P(h_n|B). \tag{3.19}$$

Indeed, since for any $B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\}$, the conditional probability measure $P(\cdot|B)$ works like a probability measure as in the unconditional case, Theorem 104_{240} implies the aforementioned equation.

3.5.3 *Law of iterated expectations*

An important property of stochastic processes is the so-called *law of iterated expectations*.[5] We present the property for a conditional probability measure $P \in \mathbb{P}_p$ on $\mathscr{C}_\sigma$ and for any n-measurable function. Note that our version of the law of iterated expectation presents an equality that holds everywhere, whereas under the standard measure-theoretic approach we would have an equality that holds "almost everywhere".

Theorem 17. *Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ and let $P \in \mathbb{P}_p$ be a coherent conditional probability measure on $\mathscr{C}_\sigma$. Consider as well any $n \in \mathbb{N}_0$ and any n-measurable function $h(X_{1:n})$. Then for all $m, m' \in \mathbb{N}_0$ such that $m \leq m' \leq n$, it holds that*

$$E_P\big(h(X_{1:n})|X_{1:m}\big) = E_P\Big(E_P\big(h(X_{1:n})|X_{1:m'}\big)\big|X_{1:m}\Big).$$

Proof. We prove the statement for the non-trivial case where $m < m' < n$. Due to Equation $(3.18)_\frown$, for all $x_{1:m} \in \mathscr{X}^m$, we have that

$$\begin{aligned}
E_P\big(h(X_{1:n})|x_{1:m}\big) &= \sum_{x_{m+1:n}\in\mathscr{X}^{n-m}} h(x_{1:n}) \prod_{i=m}^{n-1} p(x_{i+1}|x_{1:i}) \\
&= \sum_{x_{m+1:n}\in\mathscr{X}^{n-m}} h(x_{1:n}) \prod_{i=m'}^{n-1} p(x_{i+1}|x_{1:i}) \prod_{j=m}^{m'-1} p(x_{j+1}|x_{1:j}) \\
&= \sum_{x_{m+1:m'}\in\mathscr{X}^{m'-m}} E_P(h(X_{1:n})|x_{1:m}, x_{m+1:m'}) \prod_{i=m}^{m'-1} p(x_{i+1}|x_{1:i}) \\
&= E_P\Big(E_P\big(h(X_{1:n})|x_{1:m}, X_{m+1:m'}\big)|x_{1:m}\Big).
\end{aligned}$$

□

[5] Also known as law of total expectation.

The law of iterated expectations allows us to compute expectations recursively. We show how this is done for n-measurable functions. Suppose that we want to compute $E_P\big(h(X_{1:n})\big)$ for any $n \in \mathbb{N}$ and any function $h \in \mathscr{L}(\mathscr{X}^n)$. We first compute $E_P\big(h(X_{1:n})|x_{1:n-1}\big)$ for all $x_{1:n-1} \in \mathscr{X}^{n-1}$, then we plug the result in $E_P(\cdot|X_{1:n-2})$, we compute it for all $x_{1:n-2} \in \mathscr{X}^{n-2}$ and so on until we get to the initial situation. That is,

$$E_P\big(h(X_{1:n})\big) = E_P\Bigg(E_P\Big(\dots E_P\big(E_P(h(X_{1:n})|X_{1:n-1})\big|X_{1:n-2}\big)\dots\Big|X_1\Big)\Bigg). \tag{3.20}$$

Similarly, for the conditional expectation $E_P\big(h(X_{1:n})|x_{1:m}\big)$, we have that

$$E_P\big(h(X_{1:n})|x_{1:m}\big) = E_P\Big(\dots E_P\big(E_P(h(X_{1:n})|X_{1:n-1})\big|X_{1:n-2}\big)\dots\Big|x_{1:m}\Big), \tag{3.21}$$

for all $m \in \mathbb{N}_0$ such that $n \geq m$ and all $x_{1:m} \in \mathscr{X}^m$.

Computationally speaking, it may seem that the law of iterated expectations does not offer any advantage. However, we will later prove generalised versions of the law of iterated expectation that will turn out to be among the most useful properties for the derivation of the results in this dissertation.

3.5.4 *Additional properties*

We now prove some additional properties that are satisfied by the conditional expectations given by Equation (3.17)$_{74}$. The first property is about expectations of simple functions, and consequently n-measurable ones, and goes as follows.

Lemma 18. *Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ and any conditional probability measure $P \in \mathbb{P}_p$. Consider as well any $m \in \mathbb{N}_0$, any $x_{m+1} \in \mathscr{X}$ and any simple function h on Ω. Then it holds that*

$$\min_{x_{1:m} \in \mathscr{X}^m} E_P(h|x_{1:m}, x_{m+1}) \leq E_P(h|x_{m+1}) \leq \max_{x_{1:m} \in \mathscr{X}^m} E_P(h|x_{1:m}, x_{m+1}).$$

Proof. Since h is a simple function on Ω, it follows from Definition 18$_{239}$ that there is $n \in \mathbb{N}$, such that for all $i \in \{1, \dots, n\}$, there are $c_i \in \mathbb{R}$ and $A_i \in \sigma(\mathscr{X}^*)$ such that

$$h(\omega) = \sum_{i=1}^{n} c_i \mathbb{I}_{A_i}(\omega) \text{ for all } \omega \in \Omega. \tag{3.22}$$

Combining Lemma 102$_{240}$ with Lemma 105$_{240}$ and Equation (3.22), we find that

$$E_P(h|B) = \sum_{i=1}^{n} c_i P(A_i|B) \text{ for all } B \in \langle \mathscr{X}^* \rangle \setminus \{\emptyset\}$$

and the result now follows from Lemma 13$_{68}$. □

Our next result extends Lemma $18_{\curvearrowleft}$ to extended real-valued functions on Ω that are limits of non-decreasing sequences of non-negative n-measurable functions.

Lemma 19. *Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ and any conditional probability measure $P \in \mathbb{P}_p$. Consider as well any $m \in \mathbb{N}_0$, any $x_{m+1} \in \mathscr{X}$ and any extended real-valued g on Ω for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} h_n = g$. Then it holds that*

$$\min_{x_{1:m}\in\mathscr{X}^m} E_P(g|x_{1:m},x_{m+1}) \leq E_P(g|x_{m+1}) \leq \max_{x_{1:m}\in\mathscr{X}^m} E_P(g|x_{1:m},x_{m+1}).$$

Proof. Since the sequence $\{h_n\}_{n\in\mathbb{N}}$ consists of n-measurable functions, it follows from Lemma $18_{\curvearrowleft}$ that for all $n \in \mathbb{N}$,

$$\min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}) \leq E_P(h_n|x_{m+1}) \leq \max_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}).$$

Since the sequence $\{h_n\}_{n\in\mathbb{N}}$ is non-decreasing, we have that the sequences $\{\min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1})\}_{n\in\mathbb{N}}$, $\{\max_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1})\}_{n\in\mathbb{N}}$ and the sequence $\{E_P(h_n|x_{m+1})\}_{n\in\mathbb{N}}$ are non-decreasing as well, which further implies that $\lim_{n\to+\infty} E_P(h_n|x_{m+1})$, $\lim_{n\to+\infty}\min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1})$ and $\lim_{n\to+\infty}\max_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1})$ are real or equal to $+\infty$, and therefore, we find that

$$\begin{aligned}\lim_{n\to+\infty}\min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}) &\leq \lim_{n\to+\infty} E_P(h_n|x_{m+1})\\ &\leq \lim_{n\to+\infty}\max_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}).\end{aligned}$$

Finally, due to Lemma 20 and Equation $(3.19)_{75}$, we infer that

$$\min_{x_{1:m}\in\mathscr{X}^m} E_P(g|x_{1:m},x_{m+1}) \leq E_P(g|x_{m+1}) \leq \max_{x_{1:m}\in\mathscr{X}^m} E_P(g|x_{1:m},x_{m+1}). \qquad \square$$

Lemma 20. *Consider any probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ and any conditional probability measure $P \in \mathbb{P}_p$. Consider as well any $m \in \mathbb{N}_0$, any $x_{m+1} \in \mathscr{X}$ and any non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$. Then it holds that*

$$\begin{aligned}\min_{x_{1:m}\in\mathscr{X}^m}\lim_{n\to+\infty} E_P(h_n|x_{1:m},x_{m+1}) &= \lim_{n\to+\infty}\min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1});\\ \lim_{n\to+\infty}\max_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}) &= \max_{x_{1:m}\in\mathscr{X}^m}\lim_{n\to+\infty} E_P(h_n|x_{1:m},x_{m+1}).\end{aligned}$$

Proof. We first prove the second equality. Since the sequence $\{h_n\}_{n\in\mathbb{N}}$ is non-decreasing, we have that $h_k \leq h_{k+1}$, for all $k \in \mathbb{N}$, and it follows from Lemma 103_{240} that

$$E_P(h_k|x_{1:m},x_{m+1}) \leq E_P(h_{k+1}|x_{1:m},x_{m+1}) \text{ for all } x_{1:m} \in \mathscr{X}^m, \tag{3.23}$$

Then observe that

$$\lim_{n\to+\infty} \max_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}) = \sup_{n\in\mathbb{N}} \max_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1})$$
$$= \max_{x_{1:m}\in\mathscr{X}^m} \sup_{n\in\mathbb{N}} E_P(h_n|x_{1:m},x_{m+1}) = \max_{x_{1:m}\in\mathscr{X}^m} \lim_{n\to+\infty} E_P(h_n|x_{1:m},x_{m+1}),$$

where the first and the last equality holds because of Inequality (3.23)$_{\curvearrowleft}$, which implies that the limit is in fact a supremum, and the second equality holds because the maximum is a supremum and therefore, we can exchange the suprema.

We now prove that

$$\min_{x_{1:m}\in\mathscr{X}^m} \lim_{n\to+\infty} E_P(h_n|x_{1:m},x_{m+1}) \geq \lim_{n\to+\infty} \min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}).$$

Observe that

$$\lim_{n\to+\infty} \min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}) = \sup_{n\in\mathbb{N}} \min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1})$$
$$\leq \min_{x_{1:m}\in\mathscr{X}^m} \sup_{n\in\mathbb{N}} E_P(h_n|x_{1:m},x_{m+1}) = \min_{x_{1:m}\in\mathscr{X}^m} \lim_{n\to+\infty} E_P(h_n|x_{1:m},x_{m+1}),$$

where the first and the last equality holds because of Inequality (3.23)$_{\curvearrowleft}$, which implies that the limit is in fact a supremum, and the inequality holds because an infimum of suprema is greater than or equal to the supremum of infima.

It now remains to prove that

$$\min_{x_{1:m}\in\mathscr{X}^m} \lim_{n\to+\infty} E_P(h_n|x_{1:m},x_{m+1}) \leq \lim_{n\to+\infty} \min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}).$$

Since the sequence $\{h_n\}_{n\in\mathbb{N}}$ is non-decreasing, we have that for all $x_{1:m} \in \mathscr{X}^m$, $\{E_P(h_n|x_{1:m},x_{m+1})\}_{n\in\mathbb{N}}$ is also non-decreasing, which implies that the limit $\lim_{n\to+\infty} E_P(h_n|x_{1:m},x_{m+1})$ exists. For all $x_{1:m} \in \mathscr{X}^m$, let now $c_{x_{1:m}} := \lim_{n\to+\infty} E_P(h_n|x_{1:m},x_{m+1})$. We then have that either $c_{x_{1:m}} \in \mathbb{R}_{\geq 0}$ or $c_{x_{1:m}} = +\infty$. Let also $S := \{x_{1:m} \in \mathscr{X}^m \colon c_{x_{1:m}} \neq +\infty\}$. We then have that either $S = \varnothing$ or $S \neq \varnothing$.

Consider first the case $S = \varnothing$, which implies that $c_{x_{1:m}} = +\infty$ for all $x_{1:m} \in \mathscr{X}^m$. Fix any $\alpha \in \mathbb{R}_{\geq 0}$. Then for all $x_{1:m} \in \mathscr{X}^m$, there is $k_{x_{1:m}} \in \mathbb{N}$ such that for all $n \geq k_{x_{1:m}}$, it holds that $E_P(h_n|x_{1:m},x_{m+1}) \geq \alpha$. Let now $k := \max_{x_{1:m}\in\mathscr{X}^m} k_{x_{1:m}}$. Then for all $x_{1:m} \in \mathscr{X}^m$ and all $n \geq k$, we have that $E_P(h_n|x_{1:m},x_{m+1}) \geq \alpha$, which also implies that $\min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}) \geq \alpha$ and that

$$\lim_{n\to+\infty} \min_{x_{1:m}\in\mathscr{X}^m} E_P(h_n|x_{1:m},x_{m+1}) \geq \alpha, \tag{3.24}$$

where the limit exists, due to the fact that $\{E_P(h_n|x_{1:m},x_{m+1})\}_{n\in\mathbb{N}}$ is non-decreasing for all $x_{1:m} \in \mathscr{X}^m$. Since Equation (3.24) is true for all $\alpha \in \mathbb{R}_{\geq 0}$, we infer that $\lim_{n\to+\infty} \min_{x_{1:m}\in\mathscr{X}} E_P(h_n|x_{1:m},x_{m+1}) = +\infty$. Moreover, due to the definition of $c_{x_{1:m}}$ and the fact that $c_{x_{1:m}} = +\infty$ for all $x_{1:m} \in \mathscr{X}^m$, we also find that

$$\min_{x_{1:m} \in \mathscr{X}^m} \lim_{n \to +\infty} E_P(h_n | x_{1:m}, x_{m+1}) = +\infty$$

and therefore, we trivially have that

$$\min_{x_{1:m} \in \mathscr{X}^m} \lim_{n \to +\infty} E_P(h_n | x_{1:m}, x_{m+1}) \leq \lim_{n \to +\infty} \min_{x_{1:m} \in \mathscr{X}^m} E_P(h_n | x_{1:m}, x_{m+1}).$$

Consider now the case $S \neq \varnothing$. Fix any $\varepsilon > 0$ and let $c := \min_{x_{1:m} \in \mathscr{X}^m} c_{x_{1:m}}$. Since $S \neq \varnothing$, we infer that $c \in \mathbb{R}_{\geq 0}$. For all $x_{1:m} \in S$, there is $k_{x_{1:m}} \in \mathbb{N}$ such that for all $n \geq k_{x_{1:m}}$, it holds that $E_P(h_n | x_{1:m}, x_{m+1}) \geq c_{x_{1:m}} - \varepsilon \geq c - \varepsilon$. Similarly, for all $x_{1:m} \in \mathscr{X}^m \setminus S$, there is $k_{x_{1:m}} \in \mathbb{N}$ such that for all $n \geq k_{x_{1:m}}$, it holds that $E_P(h_n | x_{1:m}, x_{m+1}) \geq c \geq c - \varepsilon$. Let now $k := \max_{x_{1:m} \in \mathscr{X}^m} k_{x_{1:m}}$, then for all $n \geq k$ and all $x_{1:m} \in \mathscr{X}^m$, we have that $E_P(h_n | x_{1:m}, x_{m+1}) \geq c - \varepsilon$, which implies that $\min_{x_{1:m} \in \mathscr{X}^m} E_P(h_n | x_{1:m}, x_{m+1}) \geq c - \varepsilon$ and that

$$\lim_{n \to +\infty} \min_{x_{1:m} \in \mathscr{X}^m} E_P(h_n | x_{1:m}, x_{m+1}) \geq c - \varepsilon, \tag{3.25}$$

where the limit exists, due to the fact that $\{E_P(h_n | x_{1:m}, x_{m+1})\}_{n \in \mathbb{N}}$ is non-decreasing for all $x_{1:m} \in \mathscr{X}^m$. Since Equation (3.25) is true for all $\varepsilon > 0$, we then infer that

$$\begin{aligned}\lim_{n \to +\infty} \min_{x_{1:m} \in \mathscr{X}^m} E_P(h_n | x_{1:m}, x_{m+1}) \geq c &= \min_{x_{1:m} \in \mathscr{X}^m} c_{x_{1:m}} \\ &= \min_{x_{1:m} \in \mathscr{X}^m} \lim_{n \to +\infty} E_P(h_n | x_{1:m}, x_{m+1}). \quad \square\end{aligned}$$

The properties presented here will turn out be useful in Chapter 5100, where we analyse Markov chains. More specifically, the inequalities presented in Lemma 1977 will become equalities when the extended real-valued function g does not depend on the first $m-1$ states.

3.6 GENERAL IMPRECISE PROBABILITY TREES

Instead of a single probability tree, suppose now that we are given a set of them, denoted by $\mathscr{T}$. Clearly, $\mathscr{T} \subseteq \mathbb{P}_{\mathscr{X}^*}$. Since every $p \in \mathscr{T}$ has a corresponding set of conditional probability measures on $\mathscr{C}_\sigma$, the set $\mathscr{T}$ has a corresponding set of conditional probability measures as well, denoted by $\mathbb{P}_{\mathscr{T}}$ and defined by $\mathbb{P}_{\mathscr{T}} := \cup_{p \in \mathscr{T}} \mathbb{P}_p$. Any set of probability trees $\mathscr{T}$ is called an *imprecise probability tree* and the corresponding set of conditional probability measures $\mathbb{P}_{\mathscr{T}}$ is called an *imprecise stochastic process*.

Consider any $B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\}$ and any extended real-valued function g on Ω that is measurable with respect to $\sigma(\mathscr{X}^*)$, then for each $P \in \mathbb{P}_{\mathscr{T}}$ we have an expectation $E_P(g|B)$. Since $E_P(g|B)$ might not be defined for some $P \in \mathbb{P}_{\mathscr{T}}$, we only consider functions g such that $E_P(g|B)$ is defined for all $P \in \mathbb{P}_{\mathscr{T}}$. We now model our uncertainty through global—or alternatively joint—lower and upper expectations, denoted by $\underline{E}_{\mathscr{T}}(g|B)$ and $\overline{E}_{\mathscr{T}}(g|B)$ respectively. These expectations are simply the lower and upper envelopes of all expectations that correspond to some conditional probability measure in $\mathbb{P}_{\mathscr{T}}$, and they are defined by

$$\underline{E}_{\mathscr{T}}(g|B) := \inf\left\{E_P(g|B) \colon P \in \mathbb{P}_{\mathscr{T}}\right\} = \inf_{p \in \mathscr{T}} \inf_{P \in \mathbb{P}_{\mathscr{T}}} \left\{E_P(g|B)\right\} \tag{3.26}$$

$$\overline{E}_{\mathscr{T}}(g|B) := \sup\left\{E_P(g|B) \colon P \in \mathbb{P}_{\mathscr{T}}\right\} = \sup_{p \in \mathscr{T}} \sup_{P \in \mathbb{P}_{\mathscr{T}}} \left\{E_P(g|B)\right\}, \tag{3.27}$$

where $E_P(g|B)$ is given by Equation (3.17)$_{74}$. There are some types of measurable functions, whose lower and upper expectations are always well-defined for any set of probability trees. Among them are the n-measurable functions and the extended real-valued functions that are limits of non-decreasing sequences of non-negative n-measurable functions.

Regarding the computation of lower and upper expectations of measurable extended real-valued functions, optimising over all possible expectations might be computationally infeasible if the set $\mathbb{P}_{\mathscr{T}}$ is infinite. Luckily, there are types of imprecise stochastic processes for which we can efficiently compute lower and expectation s of measurable extended real-valued functions. One such type is introduced in the next section. Another type of imprecise stochastic processes that we deal with in Chapter 5$_{100}$ are the so-called imprecise Markov chains. For such imprecise Markov chains, we will see that we can select among different types of independence between the states of the process, and that the efficiency of the computation of expectations will depend on this selection.

3.7 IMPRECISE PROBABILITY TREES DERIVED FROM SETS OF CONDITIONAL PROBABILITY MASS FUNCTIONS

In this section, we discuss a specific way for obtaining an imprecise probability tree that is quite common in the practice of imprecise probabilities. Suppose that we have an event tree, where to each situation a set of conditional probability mass functions is attached. For all $x_{1:n} \in \mathscr{X}^*$, we denote by $\mathscr{P}_{x_{1:n}}$ the non-empty set of conditional probability mass functions associated with situation $x_{1:n}$—for the initial situation, this set is denoted by $\mathscr{P}_{\square}$. These sets of conditional probability mass functions form a collection of local models, which we denote by $\mathscr{P}$. By taking all possible combinations of the conditional probabilities of the local models $\mathscr{P}_{x_{1:n}}$, for all $x_{1:n} \in \mathscr{X}^*$, we obtain an imprecise probability tree, which is denoted by $\mathscr{T}_{\mathscr{P}}$ and defined as follows

$$\mathscr{T}_{\mathscr{P}} := \left\{p \in \mathbb{P}_{\mathscr{X}} \colon p(X_{n+1}|x_{1:n}) \in \mathscr{P}_{x_{1:n}} \text{ for all } n \in \mathbb{N}_0 \text{ and } x_{1:n} \in \mathscr{X}^*\right\}. \tag{3.28}$$

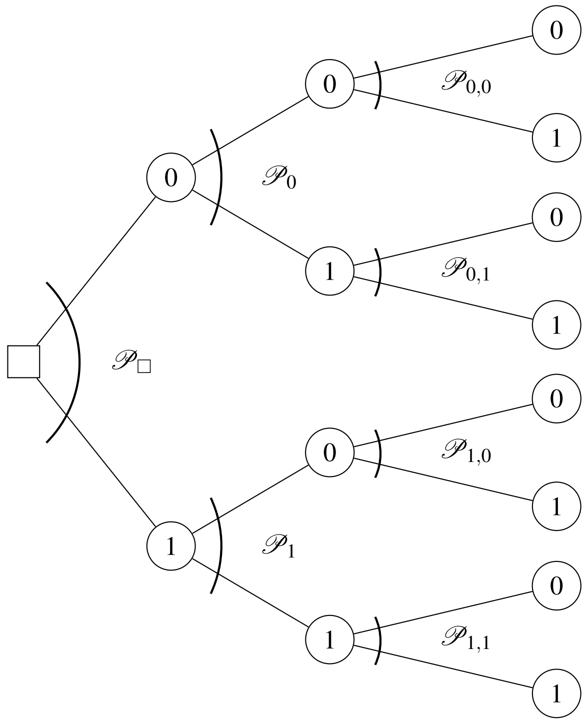

Figure 3.3: The initial part of the imprecise probability tree of an imprecise stochastic process with state space $\mathscr{X} = \{0,1\}$, initial model $\mathscr{P}_\square$ and transition models $\mathscr{P}_{x_{1:n}}$ for all $n \in \mathbb{N}$ and all $x_{1:n} \in \mathscr{X}^n$.

In other words, by choosing a local model from $\mathscr{P}_{x_{1:n}}$ in each situation $x_{1:n} \in \mathscr{X}^*$, we obtain a probability tree and $\mathscr{T}_{\mathscr{P}}$ is then the set of all such probability trees. In Figure 3.3 we have depicted the initial part of an imprecise probability tree $\mathscr{T}_{\mathscr{P}}$. Since $\mathscr{T}_{\mathscr{P}}$ is a set of probability trees, it has a corresponding set $\mathbb{P}_{\mathscr{T}_{\mathscr{P}}}$ of conditional probability measures on $\mathscr{C}_\sigma$, which we will also denote by $\mathbb{P}_{\mathscr{P}}$. The corresponding lower and upper expectations, as defined by Equations (3.26)$_\curvearrowleft$ and (3.27)$_\curvearrowleft$, will be denoted by $\underline{E}_{\mathscr{P}}$ and $\overline{E}_{\mathscr{P}}$.

Let us now discuss the computation of $\underline{E}_{\mathscr{P}}$ and $\overline{E}_{\mathscr{P}}$. First of all, for all $n \in \mathbb{N}_0$ and all $x_{1:n} \in \mathscr{X}^n$, we can associate with the set $\mathscr{P}_{x_{1:n}}$ a lower and an upper expectation operator, denoted by $\underline{Q}(\cdot|x_{1:n})$ and $\overline{Q}(\cdot|x_{1:n})$ respectively, which are defined by

$$\underline{Q}(f|x_{1:n}) := \inf\Big\{ \sum_{x\in\mathscr{X}} f(x)p(x) \colon p \in \mathscr{P}_{x_{1:n}} \Big\} \tag{3.29}$$

$$\overline{Q}(f|x_{1:n}) := \sup\Big\{ \sum_{x\in\mathscr{X}} f(x)p(x) \colon p \in \mathscr{P}_{x_{1:n}} \Big\}, \tag{3.30}$$

for all $f \in \mathscr{L}(\mathscr{X})$. If $n = 0$, we also let $\underline{Q}_{\square}(f) := \underline{Q}(f|\square)$ and $\overline{Q}_{\square}(f) := \overline{Q}(f|\square)$ for all $f \in \mathscr{L}(\mathscr{X})$.

Consider any n-measurable function $h(X_{1:n})$. If we let $\underline{Q}(h(X_{1:n})|X_{1:n-1})$ be the $(n-1)$-measurable function that is defined by

$$\underline{Q}(h(X_{1:n})|x_{1:n-1}) := \underline{Q}(h(x_{1:n-1},X_n)|x_{1:n-1}) \text{ for all } x_{1:n-1} \in \mathscr{X}^{n-1}, \tag{3.31}$$

and similarly for the upper one $\overline{Q}(h(X_{1:n})|X_{1:n-1})$, then we obtain the following result, which can be regarded as a generalised version of the law of iterated expectations.[6]

Theorem 21. *Consider any $n,m \in \mathbb{N}_0$ such that $n > m$ and any n-measurable function $h(X_{1:n})$. Then it holds that*

$$\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:m}) = \underline{Q}(\underline{Q}(\ldots\underline{Q}(h(X_{1:n})|X_{1:n-1})\ldots|X_{1:m+1})|X_{1:m});$$
$$\overline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:m}) = \overline{Q}(\overline{Q}(\ldots\overline{Q}(h(X_{1:n})|X_{1:n-1})\ldots|X_{1:m+1})|X_{1:m}).$$

Proof. We will only provide the proof for $\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:m})$; the proof for $\overline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:m})$ is completely analogous. Fix any $x_{1:m} \in \mathscr{X}^m$. We start by showing that $\underline{E}_{\mathscr{P}}(h(X_{1:n})|x_{1:m}) \geq \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m})$. Observe that

$$\begin{aligned}
\underline{E}_{\mathscr{P}}(h(X_{1:n})|x_{1:m}) &= \inf\{E_P(h(X_{1:n})|x_{1:m}) \colon P \in \mathbb{P}_{\mathscr{P}}\}\\
&= \inf\{E_P\big(E_P(h(X_{1:n})|X_{1:n-1}))|x_{1:m}\big) \colon P \in \mathbb{P}_{\mathscr{P}}\}\\
&\geq \inf\{E_P\big(E_{P'}(h(X_{1:n})|X_{1:n-1}))|x_{1:m}\big) \colon P,P' \in \mathbb{P}_{\mathscr{P}}\}\\
&\geq \inf\{E_P\big(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1}))|x_{1:m}\big) \colon P \in \mathbb{P}_{\mathscr{P}}\}\\
&= \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m}),
\end{aligned}$$

where the first and the last equality follow from Equation $(3.26)_{80}$, the second from Theorem 17_{75} and the second inequality holds because

$$E_P\big(E_{P'}(h(X_{1:n})|X_{1:n-1})|x_{1:m}\big) \geq E_P\big(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m}\big)$$

due to Lemma 103_{240} and the fact that, for all $x'_{1:n-1} \in \mathscr{X}^{n-1}$ and all $P' \in \mathbb{P}_{\mathscr{P}}$

$$\begin{aligned}
\underline{E}_{\mathscr{P}}(h(X_{1:n})|x'_{1:n-1}) &= \inf\{E_P(h(X_{1:n})|x'_{1:n-1}) \colon P \in \mathbb{P}_{\mathscr{P}}\}\\
&\leq E_{P'}(h(X_{1:n})|x'_{1:n-1}).
\end{aligned} \tag{3.32}$$

[6] A generalised law of iterated expectations for imprecise stochastic processes whose local models are closed and convex can be found in Reference [26, Theorem 7].

We now prove that $\underline{E}_{\mathscr{P}}(h(X_{1:n})|x_{1:m}) \leq \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m})$. Fix any $\varepsilon > 0$. For any $x'_{1:n-1} \in \mathscr{X}^{n-1}$, we have that

$$\underline{E}_{\mathscr{P}}(h(X_{1:n})|x'_{1:n-1}) = \inf\{E_P(h(X_{1:n})|x'_{1:n-1})\colon P \in \mathbb{P}_{\mathscr{P}}\},$$

and hence, we know that there is some $P^* \in \mathbb{P}_{\mathscr{P}}$ such that

$$E_{P^*}(h(X_{1:n})|x'_{1:n-1}) \leq \underline{E}_{\mathscr{P}}(h(X_{1:n})|x'_{1:n-1}) + \varepsilon/2,$$

where, because of Equation (3.18)$_{74}$, $E_{P^*}(h(X_{1:n})|x'_{1:n-1})$ depends only on the local model attached to $x'_{1:n-1}$. Therefore, because of Equation (3.28)$_{80}$ [the way the probability trees are constructed], there is some conditional probability measure $P' \in \mathbb{P}_{\mathscr{P}}$ such that

$$E_{P'}(h(X_{1:n})|X_{1:n-1}) \leq \underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1}) + \varepsilon/2, \tag{3.33}$$

where $E_{P'}(h(X_{1:n})|X_{1:n-1})$ depends only on the local models attached to situations in $\mathscr{X}^{n-1}$.

Moreover, there is some $P'' \in \mathbb{P}_{\mathscr{P}}$ such that

$$E_{P''}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m}) \leq \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m}) + \varepsilon/2, \tag{3.34}$$

where, because of Equation (3.18)$_{74}$, $E_{P''}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m})$ depends only on the local models attached to situations in $\cup_{i=1}^{n-2}\mathscr{X}^i$.

Due to Equation (3.28)$_{80}$ [the way the probability trees are constructed], we know that there is some $P \in \mathbb{P}_{\mathscr{P}}$ such that for all $x_{1:n-1} \in \mathscr{X}^{n-1}$, it holds that $P(X_{1:n}|x_{1:n-1}) = P'(X_{1:n}|x_{1:n-1})$, and for all $i \in \{1,\ldots,n-2\}$ and all $x_{1:i} \in \mathscr{X}^i$, that $P(X_{1:i+1}|x_{1:i}) = P''(X_{1:i+1}|x_{1:i})$. Therefore, we find that

$$E_P(h'(X_{1:n})|X_{1:n-1}) = E_{P'}(h'(X_{1:n})|X_{1:n-1}) \text{ for all } h'(X_{1:n}) \in \mathscr{L}(\mathscr{X}^n), \tag{3.35}$$

because both expectations depend only on the local models attached to situations in $\mathscr{X}^{n-1}$, and also that

$$E_P(h'(X_{1:n-1})|x_{1:m}) = E_{P''}(h'(X_{1:n-1})|x_{1:m}) \text{ for all } h'(X_{1:n-1}) \in \mathscr{L}(\mathscr{X}^{n-1}), \tag{3.36}$$

because both expectations depend only on the local models attached to situations in $\cup_{i=1}^{n-2}\mathscr{X}^i$. We now observe that

$$\begin{aligned} &E_P(h(X_{1:n})|x_{1:m}) = E_P(E_P(h(X_{1:n})|X_{1:n-1})|x_{1:m}) \\ &= E_{P''}(E_P(h(X_{1:n})|X_{1:n-1})|x_{1:m}) = E_{P''}(E_{P'}(h(X_{1:n})|X_{1:n-1})|x_{1:m}), \end{aligned} \tag{3.37}$$

where the first equality comes from Theorem 17$_{75}$, the second from Equation (3.36) and the last from Equation (3.35).

Due to Inequality (3.33) and Lemma 103$_{240}$, Inequality (3.34) now becomes

$$E_{P''}(E_{P'}(h(X_{1:n})|X_{1:n-1}) - \varepsilon/2|x_{1:m}) \leq \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m}) + \varepsilon/2,$$

and therefore, by combining Lemma 102_{240} with Lemma 106_{241}, we have that

$$E_{P''}(E_{P'}(h(X_{1:n})|X_{1:n-1})|x_{1:m}) \leq \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m}) + \varepsilon. \quad (3.38)$$

By combining Inequality (3.38) with Equation $(3.37)_{\frown}$, we infer that

$$E_P(h(X_{1:n})|x_{1:m}) \leq \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m}) + \varepsilon,$$

and therefore that

$$\underline{E}_{\mathscr{P}}(h(X_{1:n})|x_{1:m}) \leq \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m}) + \varepsilon, \quad (3.39)$$

which follows from Inequality $(3.32)_{82}$ [for $P' = P$ and $x'_{1:n-1} = x_{1:m}$]. Since Inequality (3.39) holds for any $\varepsilon > 0$, we have that

$$\underline{E}_{\mathscr{P}}(h(X_{1:n})|x_{1:m}) \leq \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|x_{1:m}).$$

Since $x_{1:m} \in \mathscr{X}^m$ was taken arbitrarily, we have that $\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:m}) = \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|X_{1:m})$ and continuing in this way, we find that

$$\begin{aligned}\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:m}) &= \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|X_{1:m}) \\ &= \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})|X_{1:n-2})|X_{1:m}) \\ &= \underline{E}_{\mathscr{P}}(\underline{E}_{\mathscr{P}}(\ldots \underline{E}_{\mathscr{P}}(h(X_{1:n})|X_{1:n-1})\ldots|X_{1:m+1})|X_{1:m}).\end{aligned}$$

The result now follows because

$$\begin{aligned}\underline{E}_{\mathscr{P}}(h(X_1)|\Box) &= \inf\{E_P(h(X_1))\colon P \in \mathbb{P}_{\mathscr{P}}\} \\ &= \inf\Big\{\sum_{x_1 \in \mathscr{X}} h(x_1)p(x_1)\colon p \in \mathscr{P}_{\Box}\Big\} = \underline{Q}_{\Box}(h) \quad (3.40)\end{aligned}$$

and because, for all $k \in \mathbb{N}$, all $h \in \mathscr{L}(\mathscr{X}^{k+1})$ and all $x_{1:k} \in \mathscr{X}^k$:

$$\begin{aligned}\underline{E}_{\mathscr{P}}(h(X_{1:k+1})|x_{1:k}) &= \inf\{E_P(h(X_{1:k+1})|x_{1:k})\colon P \in \mathbb{P}_{\mathscr{P}}\} \\ &= \inf\Big\{\sum_{x_{k+1} \in \mathscr{X}} h(x_{1:k}, x_{k+1})p(x_{k+1}|x_{1:k})\colon p(x_{k+1}|x_{1:k}) \in \mathscr{P}_{x_{1:k}}\Big\} \\ &= \underline{Q}(h(X_{1:k+1})|x_{1:k}), \quad (3.41)\end{aligned}$$

where the first equalities in both equations come from Equation $(3.26)_{80}$, the second from Equation $(3.18)_{74}$ and the last from Equation $(3.29)_{82}$. $\Box$

Theorem 21_{82} tells us that for any $n, m \in \mathbb{N}_0$ such that $n > m$, we can compute the lower or upper expectation of any n-measurable function conditional on any situation $x_{1:m} \in \mathscr{X}^m$ by solving local optimisation problems instead of optimising over all probability trees in $\mathscr{T}_{\mathscr{P}}$. Note that the sets $\mathscr{P}_{x_{1:n}}$, from which we derive the lower and upper expectations $\underline{Q}(\cdot|x_{1:n})$ and $\overline{Q}(\cdot|x_{1:n})$, are kept general so far. Therefore, Theorem 21_{82} covers also the cases where $\mathscr{P}_{x_{1:n}}$ are credal sets or derived from two probability intervals I_1 and I_2 and have the form of Ψ_{I_1,I_2} given by Equation $(2.9)_{48}$.

For the purposes of this dissertation, we also need efficient computational methods for lower and upper expectations of more general extended measurable real-valued functions on Ω. Results that are similar to Theorem 17_{75} apply for such functions as well. However, for the computation of the global models of more general extended measurable real-valued function on Ω, we adopt an alternative approach that is based on submartingales rather than probability trees, and which is described in Chapter $4_{\curvearrowright}$. The reason why we do that is because this approach is in some respects more powerful than the one based on imprecise probability trees, does not require any measurability conditions for the function under study, and can be shown to coincide with the approach based on imprecise probability trees for at least some types of functions.

4

A MARTINGALE-THEORETIC APPROACH FOR IMPRECISE DISCRETE-TIME STOCHASTIC PROCESSES

In the previous chapter, we introduced imprecise stochastic processes whose local models were sets of conditional probability mass functions. Here we focus on imprecise stochastic processes whose local models are lower and upper expectations. Our approach for deriving global models is based on the concept of sub- and supermartingales. We derive our global models directly from the local lower and upper expectations of the imprecise stochastic process, which implies that our approach is not necessarily connected with (all kinds of ways of constructing) sets of probability trees. The derived global models allow us to define global lower and upper expectations of any—not necessarily measurable—extended real-valued function on Ω.

We then prove a number of properties for these global models. Among them are generalised versions of C1$_{44}$–C8$_{45}$ and the law of iterated expectations. Moreover, we investigate how the martingale-theoretic approach is connected with the measure-theoretic one introduced in the previous chapter. For n-measurable functions, we find that the two approaches coincide (for specific ways of constructing the set of probability trees). For extended real-valued functions on Ω that are limits of non-decreasing sequences of non-negative n-measurable functions, the martingale-theoretic approach is "at most as precise as" the measure-theoretic one, in the sense that the global lower expectations obtained by the martingale-theoretic approach are shown to be smaller than or equal to the ones obtained by the measure-theoretic approach and vice versa for the global upper expectations.

4.1 NOTATION

The notation introduced in Section 3.158 still applies and we now add some more elements. Instead of the notation $x_{1:n}$, we also use the generic notations s, t, u and v for situations in $\mathscr{X}^*$. Moreover, instead of $\Gamma(s)$, we will also use the notation $(s\bullet)$ for all the paths that go through situation s. Regarding concatenation, for all $n \in \mathbb{N}_0$, the concatenation of any situation $x_{1:n} \in \mathscr{X}^n$ with the state X_{n+1} at time $n+1$ will also be denoted by $(x_{1:n}, \cdot)$, where '$\cdot$' represents the generic value of the next state X_{n+1}. We also concatenate situations with infinite sequences of states and this is denoted by $(x_{1:n}, X_{n+1:\infty})$ or $(s, X_{n+1:\infty})$, for all $n \in \mathbb{N}_0$ and all $s, x_{1:n} \in \mathscr{X}^n$. Recall that the initial situation works as neutral element in the concatenation.

4.2 SUB- AND SUPERMARTINGALES

We first provide some basic information concerning sub- and supermartingales, starting with some preliminaries about processes. A *real process* $\mathscr{U}$ is a real-valued map defined on $\mathscr{X}^*$, which associates a real number $\mathscr{U}(x_{1:n}) \in \mathbb{R}$ with any situation $x_{1:n} \in \mathscr{X}^*$. A *gamble process* is a map from $\mathscr{X}^*$ to $\mathscr{L}(\mathscr{X})$, which associates with any situation $x_{1:n} \in \mathscr{X}^*$ a gamble in $\mathscr{L}(\mathscr{X})$—in fact, a gamble on X_{n+1}. With any real process $\mathscr{U}$, we can always associate a corresponding gamble process $\Delta\mathscr{U}$, called the *process difference*. For every situation $x_{1:n} \in \mathscr{X}^*$, the corresponding gamble $\Delta\mathscr{U}(x_{1:n}) \in \mathscr{L}(\mathscr{X})$ is defined by

$$\Delta\mathscr{U}(x_{1:n})(x_{n+1}) := \mathscr{U}(x_{1:n}, x_{n+1}) - \mathscr{U}(x_{1:n}) \text{ for all } x_{n+1} \in \mathscr{X}.$$

As mentioned in the preamble of this chapter, the local models of the imprecise stochastic processes that we consider are represented by lower and upper expectations, that is $\underline{Q}(\cdot|x_{1:n})$ and $\overline{Q}(\cdot|x_{1:n})$ respectively for all $x_{1:n} \in \mathscr{X}^*$, where $\underline{Q}(\cdot|x_{1:n})$ satisfies the properties introduced in the beginning of Section 2.444 and $\overline{Q}(\cdot|x_{1:n})$ is its conjugate upper expectation. These lower and upper expectations could be derived from local sets of probability mass functions $\mathscr{P}_{x_{1:n}}$ using Equations (3.29)82 and (3.30)82, but they can also be given directly.

A *submartingale* $\mathscr{M}$ is then a real process such that

$$\underline{Q}(\Delta\mathscr{M}(x_{1:n})|x_{1:n}) \geq 0 \text{ for all } n \in \mathbb{N}_0 \text{ and all } x_{1:n} \in \mathscr{X}^n \tag{4.1}$$

or, in other words, a process that is expected to increase. A *supermartingale* is a real process $\mathscr{M}$ such that $-\mathscr{M}$ is a submartingale, or equivalently, because of conjugacy, such that

$$\overline{Q}(\Delta\mathscr{M}(x_{1:n})|x_{1:n}) \leq 0 \text{ for all } n \in \mathbb{N}_0 \text{ and all } x_{1:n} \in \mathscr{X}^n \tag{4.2}$$

or, in other words, a process that is expected to decrease. A submartingale is *uniformly bounded above* if there is some $c \in \mathbb{R}$, such that $\mathscr{M}(x_{1:n}) \leq c$ for all $x_{1:n} \in \mathscr{X}^*$. A supermartingale $\mathscr{M}$ is *uniformly bounded below* if $-\mathscr{M}$ is uniformly bounded above, or equivalently, if there is some $c \in \mathbb{R}$, such that $\mathscr{M}(x_{1:n}) \geq c$ for all $x_{1:n} \in \mathscr{X}^*$. The set of all uniformly bounded above submartingales is denoted by $\underline{\mathbb{M}}$, and the set of all uniformly bounded below supermartingales is denote by $\overline{\mathbb{M}}$; clearly, we have that $\overline{\mathbb{M}} = -\underline{\mathbb{M}}$.

4.3 AN ALTERNATIVE APPROACH FOR GLOBAL MODELS

We now present an approach for defining global models that is based on sub- and supermartingales as well as some basic properties that are satisfied by these global models. This approach is inspired by the game-theoretic probability framework proposed by Shafer and Vovk; see Reference [25, 26, 66]. The properties of this approach are (at least for now) better understood and it can easily be applied to any—not necessarily measurable—extended-real valued function. Another advantage of this approach is that it works directly with the lower and upper expectations that represent our local models.

4.3.1 *Defining global models*

The global models that we are about to construct will provide lower and upper expectations of extended real-valued functions on Ω. The link between the lower and upper expectations of extended real-valued functions and the local models of our imprecise stochastic processes is established by means of the sub- and supermartingales of the previous subsection.

First, for any real process $\mathscr{U}$ (and therefore, in particular, for any sub- or supermartingale), we consider the extended real-valued functions $\liminf \mathscr{U}$ and $\limsup \mathscr{U}$ on Ω, defined for all $\omega \in \Omega$ by

$$\liminf \mathscr{U}(\omega) := \liminf_{n\to\infty} \mathscr{U}(\omega^n) \text{ and } \limsup \mathscr{U}(\omega) := \limsup_{n\to\infty} \mathscr{U}(\omega^n).$$

Next, we use these functions to define conditional global lower and upper expectations. For any $s \in \mathscr{X}^*$ and any extended real-valued function g on Ω, the conditional global lower expectation of g is denoted by $\underline{E}_{\underline{Q}}(g|s)$ and defined as

$$\underline{E}_{\underline{Q}}(g|s) := \sup\big\{\mathscr{M}(s) \colon \mathscr{M} \in \underline{\mathbb{M}} \text{ and } \limsup \mathscr{M}(\omega) \leq g(\omega) \\ \text{for all } \omega \in \Gamma(s)\big\}. \quad (4.3)$$

Similarly, the conjugate conditional global upper expectation of g is $\overline{E}_{\overline{Q}}(g|s)$ and defined as

$$\overline{E}_{\underline{Q}}(g|s) := \inf\big\{\mathscr{M}(s) \colon \mathscr{M} \in \overline{\mathbb{M}} \text{ and } \liminf \mathscr{M}(\omega) \geq g(\omega) \\ \text{for all } \omega \in \Gamma(s)\big\}. \quad (4.4)$$

For $s = \Box$ in the definitions above, we obtain the unconditional global lower and upper expectations of g, which we denote by $\underline{E}_Q(g)$ and $\overline{E}_Q(g)$. Notice that, because of the conjugacy between $\underline{Q}$ and $\overline{Q}$, we can write $\underline{E}_{\overline{Q}}(g|s)$ instead of $\underline{E}_Q(g|s)$ and $\overline{E}_{\overline{Q}}(g|s)$ instead of $\overline{E}_Q(g|s)$.

We now provide an interpretation of the aforementioned definitions. Basically, a submartingale $\mathscr{M}$ can be interpreted as a capital process, in the sense that it represents the evolution of a subject's monetary capital, of which the local changes $\Delta\mathscr{M}(x_{1:n})$ are expected (on 'average') to either increase this capital or keep it steady because their local lower expectation $\underline{Q}(\Delta\mathscr{M}(x_{1:n})|x_{1:n})$ is non-negative, for all $x_{1:n} \in \mathscr{X}^*$. The assumption is then that, since all these local increases are expected to be at most non-negative, the value of $\limsup \mathscr{M}(x_{1:n}, X_{n+1:\infty})$ should be expected to be at least the starting capital $\mathscr{M}(x_{1:n})$. Therefore, $\mathscr{M}(x_{1:n})$ can be seen as a supremum buying price for the—possibly extended—real-valued gamble $\limsup \mathscr{M}(x_{1:n}, X_{n+1:\infty})$. This implies that for a given extended real-valued function g on Ω, if for all $\omega \in \Gamma(x_{1:n})$, $\limsup \mathscr{M}(\omega) \leq g(\omega)$, then $\mathscr{M}(x_{1:n})$ is a buying price for g, which, due to the behavioural interpretation of lower expectations as supremum buying prices, leads us to Equation (4.3)$_{\curvearrowleft}$. Equation (4.4)$_{\curvearrowleft}$ follows from a similar argument or from conjugacy. The reason why we only consider submartingales that are bounded above is because otherwise it is possible to end up with $\overline{E}_Q(g|s) = -\infty$ and $\underline{E}_Q(g|s) = +\infty$; see Reference [25, Example 1]. Also, from an interpretational point of view, that a supermartingale is bounded below means that the subject is not allowed to borrow unlimited amounts of money. Detailed technical, interpretational and philosophical discussions about the aforementioned definitions and other closely related so-called game-theoretic definitions of lower and upper expectations can be found in References [26,66].

4.3.2 *Properties of the global models*

In this section, we present some useful properties that are satisfied by the global models defined by Equations (4.3)$_{\curvearrowleft}$ and (4.4)$_{\curvearrowleft}$. We start with a technical lemma that is used multiple times in the proofs of this section.

Lemma 22. *Consider any submartingale $\mathscr{M}$ and any situation $s \in \mathscr{X}^*$, then:*

$$\mathscr{M}(s) \leq \sup_{\omega\in\Gamma(s)} \liminf \mathscr{M}(\omega) \leq \sup_{\omega\in\Gamma(s)} \limsup \mathscr{M}(\omega).$$

Proof. Consider any real α, and assume that $\mathscr{M}(s) > \alpha$. Assume that $s = x_{1:n}$ with $n \in \mathbb{N}_0$. Since $\mathscr{M}$ is a submartingale, we know that $\underline{Q}(\mathscr{M}(x_{1:n}, \cdot) - \mathscr{M}(x_{1:n})|x_{1:n}) \geq 0$, and therefore, it follows from properties C5$_{45}$ and C8$_{45}$ and the assumption, that

$$\max \mathscr{M}(x_{1:n}, \cdot) \geq \underline{Q}(\mathscr{M}(x_{1:n}, \cdot)|x_{1:n}) \geq \mathscr{M}(x_{1:n}) > \alpha,$$

implying that there is some $x_{n+1} \in \mathscr{X}$ such that $\mathscr{M}(x_{1:n+1}) > \alpha$. Repeating the same argument over and over again, this leads to the conclusion that there is some $\omega \in \Gamma(x_{1:n})$ such that $\mathscr{M}(\omega^{n+k}) > \alpha$ for all $k \in \mathbb{N}_0$, whence $\liminf \mathscr{M}(\omega) \geq \alpha$, and therefore also $\sup_{\omega \in \Gamma(x_{1:n})} \liminf \mathscr{M}(\omega) \geq \alpha$. The rest of the proof is now immediate. □

The next three results show the behaviour of the defined global models on n-measurable functions.[1]

Proposition 23. *Consider any $m, n \in \mathbb{N}_0$ such that $n \geq m$, any $x_{1:m} \in \mathscr{X}^m$ and any n-measurable extended real-valued function $h(X_{1:n})$. Then*

$$\underline{E}_{\underline{Q}}\big(h(X_{1:n})|x_{1:m}\big) = \sup\big\{\mathscr{M}(x_{1:m}) \colon \mathscr{M} \in \underline{\mathbb{M}} \text{ and } (\forall x_{m+1:n} \in \mathscr{X}^{n-m})\mathscr{M}(x_{1:n}) \leq h(x_{1:n})\big\};$$

$$\overline{E}_{\underline{Q}}\big(h(X_{1:n})|x_{1:m}\big) = \inf\big\{\mathscr{M}(x_{1:m}) \colon \mathscr{M} \in \underline{\mathbb{M}} \text{ and } (\forall x_{m+1:n} \in \mathscr{X}^{n-m})\mathscr{M}(x_{1:n}) \geq h(x_{1:n})\big\}.$$

Proof. We sketch the idea of the proof of the equality for the lower expectations; the proof for the upper expectations is completely similar. For simplicity of notation, let

$$R := \sup\big\{\mathscr{M}(x_{1:m}) \colon \mathscr{M} \in \underline{\mathbb{M}} \text{ and } (\forall x_{m+1:n} \in \mathscr{X}^{n-m})\mathscr{M}(x_{1:n}) \leq h(x_{1:n})\big\}.$$

First, consider any submartingale $\mathscr{M}$ such that $\mathscr{M}(x_{1:n}) \leq h(x_{1:n})$ for all $x_{m+1:n} \in \mathscr{X}^{n-m}$. Consider the submartingale $\mathscr{M}'$ derived from $\mathscr{M}$ by keeping it constant as soon as any situation in $\mathscr{X}^n$ is reached , then clearly $\mathscr{M}'$ is bounded above, $\limsup \mathscr{M}'(\omega) \leq h(\omega)$ for all $\omega \in \Gamma(x_{1:m})$, and $\mathscr{M}(x_{1:m}) = \mathscr{M}'(x_{1:m})$. Hence it follows from the definition of the global model in Equation $(4.3)_{88}$ that $\mathscr{M}(x_{1:m}) \leq \underline{E}_{\underline{Q}}(h(X_{1:n})|x_{1:m})$, whence $R \leq \underline{E}_{\underline{Q}}(h(X_{1:n})|x_{1:m})$.

For the converse inequality, consider any bounded above submartingale $\mathscr{M}$ for which it holds that $\limsup \mathscr{M}(\omega) \leq h(\omega)$ for all $\omega \in \Gamma(x_{1:m})$. Fix any $x_{m+1:n} \in \mathscr{X}^{n-m}$, then it follows from the n-measurability of $h(X_{1:n})$ that $\limsup \mathscr{M}(\omega) \leq h(x_{1:n})$ for all $\omega \in \Gamma(x_{1:n})$, whence

$$\mathscr{M}(x_{1:n}) \leq \sup_{\omega \in \Gamma(x_{1:n})} \limsup \mathscr{M}(\omega) \leq h(x_{1:n}),$$

where the first inequality follows from Lemma $22_{\frown}$ [for $s := x_{1:n}$]. This implies that $\mathscr{M}(x_{1:m}) \leq R$, and therefore also $\underline{E}_{\underline{Q}}(h(X_{1:n})|x_{1:m}) \leq R$. □

[1] In Chapter 3_{57} we introduced the convention that n-measurable functions are real-valued. As we will also work with n-measurable functions that are extended real-valued, whenever this is the case, we will explicitly mention that the n-measurable function is extended real-valued—see Proposition 23 and Corollary $24_{\frown}$.

Corollary 24. *Consider any $m, n \in \mathbb{N}_0$ such that $n \geq m$, any $x_{1:m} \in \mathscr{X}^m$ and any n-measurable extended real-valued function $h(X_{1:n})$. Then*

$$\underline{E}_{\underline{Q}}\big(h(X_{1:n})|x_{1:m}\big) = \sup\big\{\underline{E}_{\underline{Q}}\big(h'(X_{1:n})|x_{1:m}\big) \colon h' \in \mathscr{L}(\mathscr{X}^n) \text{ and } (\forall x_{m+1:n} \in \mathscr{X}^{n-m}) h'(x_{1:n}) \leq h(x_{1:n})\big\};$$
$$\overline{E}_{\underline{Q}}\big(h(X_{1:n})|x_{1:m}\big) = \inf\big\{\overline{E}_{\underline{Q}}\big(h'(X_{1:n})|x_{1:m}\big) \colon h' \in \mathscr{L}(\mathscr{X}^n) \text{ and } (\forall x_{m+1:n} \in \mathscr{X}^{n-m}) h'(x_{1:n}) \geq h(x_{1:n})\big\}.$$

Proof. We give the proof for the lower expectations; the proof for the upper expectations is completely similar. For simplicity of notation, let

$$R := \sup\big\{\underline{E}_{\underline{Q}}\big(h'(X_{1:n})|x_{1:m}\big) \colon h' \in \mathscr{L}(\mathscr{X}^n) \text{ and } (\forall x_{m+1:n} \in \mathscr{X}^{n-m}) h'(x_{1:n}) \leq h(x_{1:n})\big\}.$$

It follows from Proposition 23$_{\curvearrowleft}$ that $\underline{E}_{\underline{Q}}(h'(X_{1:n})|x_{1:m}) \leq \underline{E}_{\underline{Q}}(h(X_{1:n})|x_{1:m})$ for all $h' \in \mathscr{L}(\mathscr{X}^n)$ such that $h'(x_{1:m}, x_{m+1:n}) \leq h(x_{1:m}, x_{m+1:n})$ for all $x_{m+1:n} \in \mathscr{X}^{n-m}$, and therefore also $R \leq \underline{E}_{\underline{Q}}(h(X_{1:n})|x_{1:m})$.

Conversely, consider any submartingale $\mathscr{M}$ such that $\mathscr{M}(x_{1:m}, x_{m+1:n}) \leq h(x_{1:m}, x_{m+1:n})$ for all $x_{m+1:n} \in \mathscr{X}^{n-m}$. If we define the n-measurable function $h'(X_{1:n})$ on Ω by letting $h'(x_{1:n}) := \mathscr{M}(x_{1:n})$ for all $x_{1:n} \in \mathscr{X}^n$, then it follows from Proposition 23$_{\curvearrowleft}$ that $\mathscr{M}(x_{1:m}) \leq \underline{E}_{\underline{Q}}(h'(X_{1:n})|x_{1:m})$, and since by assumption $h'(x_{1:m}, x_{m+1:n}) \leq h(x_{1:m}, x_{m+1:n})$ for all $x_{m+1:n} \in \mathscr{X}^{n-m}$, also that $\underline{E}_{\underline{Q}}(h'(X_{1:n})|x_{1:m}) \leq R$. Hence $\mathscr{M}(x_{1:m}) \leq R$, and therefore, by Proposition 23$_{\curvearrowleft}$, $\underline{E}_{\underline{Q}}(h(X_{1:n})|x_{1:m}) \leq R$. □

Corollary 25. *Consider any $n \in \mathbb{N}_0$, any situation $x_{1:n} \in \mathscr{X}^n$, and any $(n+1)$-measurable function $h(X_{1:n+1})$. Then*

$$\underline{E}_{\underline{Q}}(h(X_{1:n+1})|x_{1:n}) = \underline{Q}(h(x_{1:n}, \cdot)|x_{1:n});$$
$$\overline{E}_{\underline{Q}}(h(X_{1:n+1})|x_{1:n}) = \overline{Q}(h(x_{1:n}, \cdot)|x_{1:n}).$$

Proof. We give the proof for the lower expectation; the proof for the upper expectation is completely similar.

First, consider any $\mathscr{M} \in \underline{\mathbb{M}}$ such that $\mathscr{M}(x_{1:n}, \cdot) \leq h(x_{1:n} \cdot)$, then it follows from properties C4$_{45}$ and C8$_{45}$ and the submartingale character of $\mathscr{M}$ that

$$\underline{Q}(h(x_{1:n}, \cdot)|x_{1:n}) \geq \underline{Q}(\mathscr{M}(x_{1:n}, \cdot)|x_{1:n}) \geq \mathscr{M}(x_{1:n}),$$

so Proposition 23$_{\curvearrowleft}$ guarantees that $\underline{E}_{\underline{Q}}(h(X_{1:n+1})|x_{1:n}) \leq \underline{Q}(h(x_{1:n}, \cdot)|x_{1:n})$.

To show that the inequality is actually an equality, consider any submartingale $\mathscr{M}$ such that $\mathscr{M}(x_{1:n}) = \underline{Q}(h(x_{1:n}, \cdot)|x_{1:n})$ and $\mathscr{M}(x_{1:n}, \cdot) = h(x_{1:n}, \cdot)$. □

We end this section by proving three interesting and very useful results about the global models. The first summarises and extends properties first proved by Shafer and Vovk (see for instance References [66, Chapter 8.3], [65, Section 2] and [5, Section 6.3]), in showing that these global models satisfy properties that extend the properties C1$_{44}$–C8$_{45}$ for lower and upper expectations from gambles to extended real-valued functions. We provide, for the sake of completeness, proofs that are very close to the ones given by Shafer and Vovk [65, Section 2].

Proposition 26. *Consider any $s \in \mathscr{X}^*$, any $\mu \in \mathbb{R}$, any $\lambda \in \mathbb{R}_{\geq 0}$ and any extended real-valued functions g, g' on Ω. Then*

G1. $\underline{E}_Q(g|s) \geq \inf\{g(\omega)\colon \omega \in \Gamma(s)\}$;

G2. $\underline{E}_Q(g+g'|s) \geq \underline{E}_Q(g|s) + \underline{E}_Q(g'|s)$;

G3. $\underline{E}_Q(\lambda g|s) = \lambda \underline{E}_Q(g|s)$;

G4. *if $g' \leq g$ on $\Gamma(s)$, then*

$$\underline{E}_Q(g'|s) \leq \underline{E}_Q(g|s) \text{ and } \overline{E}_Q(g'|s) \leq \overline{E}_Q(g|s);$$

as a consequence, if $g' = g$ on $\Gamma(s)$, then

$$\underline{E}_Q(g'|s) = \underline{E}_Q(g|s) \text{ and } \overline{E}_Q(g'|s) = \overline{E}_Q(g|s);$$

G5. $\inf\{g(\omega)\colon \omega \in \Gamma(s)\} \leq \underline{E}_Q(g|s) \leq \overline{E}_Q(g|s) \leq \sup\{g(\omega)\colon \omega \in \Gamma(s)\}$;

G6. $\underline{E}_Q(g+\mu|s) = \mu + \underline{E}_Q(g|s)$ *and* $\overline{E}_Q(g+\mu|s) = \mu + \overline{E}_Q(g|s)$.

In these expressions, we use the convention that $\infty + \infty = \infty$, $-\infty + (-\infty) = -\infty$, $-\infty + \infty = \infty + (-\infty) = -\infty$, $a + \infty = \infty + a = \infty$, $a + (-\infty) = -\infty + a = -\infty$ for all real a, and $0 \cdot \pm\infty = \pm\infty \cdot 0 = 0$.[2]

Proof. G1. If $\inf\{g(\omega)\colon \omega \in \Gamma(s)\} = -\infty$, then the inequality is trivially satisfied. Consider therefore any real $L \leq \inf\{g(\omega)\colon \omega \in \Gamma(s)\}$, and the submartingale $\mathscr{M}$ that assumes the constant value L everywhere. Then surely $\mathscr{M}$ is bounded above, $\limsup \mathscr{M}(s\bullet) = L \leq g(s\bullet)$ and $\mathscr{M}(s) = L$, so Equation (4.3)$_{88}$ guarantees that indeed $L \leq \underline{E}_Q(g|s)$.

G2. When $\underline{E}_Q(g|s)$ or $\underline{E}_Q(g'|s)$ is equal to $-\infty$, then so is their sum, and the inequality holds trivially. Assume therefore that both $\underline{E}_Q(g|s) > -\infty$ and $\underline{E}_Q(g'|s) > -\infty$. This implies that there are bounded above submartingales $\mathscr{M}_1$

[2] This is the extended addition that is convenient for working with lower expectations; for the dual upper expectations, we need to introduce a dual operator, defined by $a +^* b := -[(-a)+(-b)]$ for all extended real a and b.

and $\mathscr{M}_2$ such that $\limsup \mathscr{M}_1(s\bullet) \leq g(s\bullet)$ and $\limsup \mathscr{M}_2(s\bullet) \leq g'(s\bullet)$. Consider any such submartingales $\mathscr{M}_1$ and $\mathscr{M}_2$, then it follows from property $\mathrm{C2}_{44}$ of the local models that $\mathscr{M} := \mathscr{M}_1 + \mathscr{M}_2$ is a bounded above submartingale as well. Since

$$\limsup \mathscr{M}(s\bullet) \leq \limsup \mathscr{M}_1(s\bullet) + \limsup \mathscr{M}_2(s\bullet) \leq g(s\bullet) + g'(s\bullet),$$ [3]

where the second inequality holds due to the convention introduced and the fact that $\limsup \mathscr{M}_1(s\bullet) \leq g(s\bullet)$ and $\limsup \mathscr{M}_2(s\bullet) \leq g'(s\bullet)$, we infer from Equation $(4.3)_{88}$ that indeed $\underline{E}_Q(g+g'|s) \geq \underline{E}_Q(g|s) + \underline{E}_Q(g'|s)$.

$\mathrm{G3}_{\curvearrowleft}$. For $\lambda > 0$, it suffices to observe that if $\mathscr{M}$ is a bounded above submartingale such that $\limsup \mathscr{M}(s\bullet) \leq g(s\bullet)$, then the process $\lambda\mathscr{M}$ is also a bounded above submartingale such that $\limsup[\lambda\mathscr{M}(s\bullet)] \leq \lambda g(s\bullet)$, and *vice versa*. For $\lambda = 0$, we infer on the one hand from property $\mathrm{G1}_{\curvearrowleft}$ and Lemma 22_{89} that $\underline{E}_Q(\lambda g|s) = \underline{E}_Q(0|s) = 0$, and on the other hand we also know that $0 \cdot \underline{E}_Q(g|s) = 0$.

$\mathrm{G4}_{\curvearrowleft}$. Due to conjugacy, it suffices to prove the first inequality. It is trivially satisfied if $\underline{E}_Q(g'|s) = -\infty$. Assume therefore that $\underline{E}_Q(g'|s) > -\infty$, meaning that there is some bounded above submartingale $\mathscr{M}$ such that $\limsup \mathscr{M}(s\bullet) \leq g'(s\bullet)$. Consider any such submartingale $\mathscr{M}$, then we have that

$$\limsup \mathscr{M}(s\bullet) \leq g'(s\bullet) \leq g(s\bullet),$$

which implies that $\mathscr{M}(s) \leq \underline{E}(g|s)$, and therefore that $\underline{E}_Q(g'|s) \leq \underline{E}_Q(g|s)$.

$\mathrm{G5}_{\curvearrowleft}$. Suppose *ex absurdo* that $\underline{E}_Q(g|s) > \overline{E}_Q(g|s) = -\underline{E}_Q(-g|s)$. This implies that also $\underline{E}_Q(g|s) + \underline{E}_Q(-g|s) > 0$, but then property $\mathrm{G2}_{\curvearrowleft}$ tells us that also $\underline{E}_Q(g+(-g)|s) > 0$. Now the extended real-valued function $g+(-g)$ assumes only the values 0 and $-\infty$, and therefore $g+(-g) \leq 0$, so we infer from property $\mathrm{G4}_{\curvearrowleft}$ that $\underline{E}_Q(g+(-g)|s) \leq \underline{E}_Q(0|s) = 0$, where the last equality follows from property $\mathrm{G3}_{\curvearrowleft}$. This is a contradiction. The remaining inequalities are now trivial.

$\mathrm{G6}_{\curvearrowleft}$. Due to conjugacy, it suffices to prove the first equality. If $\mathscr{M}$ is a bounded above submartingale such that $\limsup \mathscr{M}(s\bullet) \leq g(s\bullet) + \mu$, then $\mathscr{M} - \mu$ is a bounded above submartingale such that $\limsup[\mathscr{M}(g\bullet) - \mu] \leq g(s\bullet)$, and *vice versa*. □

Our second result follows immediately from the definition of the global model given by Equation $(4.3)_{88}$ and therefore it is stated without proof.

[3]The first inequality holds for bounded above submartingales, but may fail for more general ones. Indeed, assume that on some path ω, $\mathscr{M}_1(\omega^n) = 2n$ and $\mathscr{M}_2(\omega^n) = -n$. Then $\limsup \mathscr{M}_1(\omega) = +\infty$, $\limsup \mathscr{M}_2(\omega) = -\infty$, and $\limsup[\mathscr{M}_1(\omega) + \mathscr{M}_2(\omega)] = +\infty$, so the inequality is violated.

Proposition 27. *Consider any $n \in \mathbb{N}_0$ and any $x_{1:n} \in \mathscr{X}^n$. Then for any extended real-valued function g on Ω, we have that*

$$\underline{E}_Q(g|x_{1:n}) = \underline{E}_Q(g(x_{1:n}, X_{n+1:\infty})|x_{1:n}).$$

If we regard the conditional lower expectation $\underline{E}_Q(g|x_{1:m})$ as a function of $x_{1:m}$, and we interpret this function $\underline{E}_Q(g|X_{1:m})$ as a (possibly extended) real-valued function on Ω, then we obtain our third result, which can be regarded as a generalisation of the law of iterated expectations. Our formulation generalises a result by Shafer and Vovk [66, Proposition 8.7], whose proof can only be guaranteed to work for bounded real-valued functions; we provide a proof that is better suited for dealing with extended real-valued ones.

Theorem 28 (Law of iterated expectations)**.** *Consider any $n, m \in \mathbb{N}_0$ such that $m \leq n$. Then for any extended real-valued function g on Ω, it holds that*

$$\underline{E}_Q(g|X_{1:m}) = \underline{E}_Q(\underline{E}_Q(g|X_{1:n})|X_{1:m}).$$

Proof. Fix any $z_{1:m} \in \mathscr{X}^m$. Due to Proposition 27$_\curvearrowleft$, we prove that

$$\underline{E}_Q(g|z_{1:m}) = \underline{E}_Q(\underline{E}_Q(g|z_{1:m}, X_{m+1:n})|z_{1:m}).$$

First, consider any bounded above submartingale $\mathscr{M} \in \underline{\mathbb{M}}$ for which it holds that $\limsup \mathscr{M}(z_{1:m}\bullet) \leq g(z_{1:m}\bullet)$. Then also, for any $x_{m+1:n} \in \mathscr{X}^{n-m}$, $\limsup \mathscr{M}(z_{1:m}, x_{m+1:n}\bullet) \leq g(z_{1:m}, x_{m+1:n}\bullet)$, which, due to Equation $(4.3)_{88}$ [for $s := (z_{1:m}, x_{m+1:n})$], implies that $\mathscr{M}(z_{1:m}, x_{m+1:n}) \leq \underline{E}_Q(g|z_{1:m}, x_{m+1:n})$. Furthermore, this implies that $\mathscr{M}(z_{1:m}, X_{m+1:n}) \leq \underline{E}_Q(g|z_{1:m}, X_{m+1:n})$, and therefore we infer from Proposition 26_{92} [property $\overline{\text{G}}4_{92}$ for $s := z_{1:m}$], that also $\underline{E}_Q(\mathscr{M}(z_{1:m}, X_{m+1:n})|z_{1:m}) \leq \underline{E}_Q(\underline{E}_Q(g|z_{1:m}, X_{m+1:n})|z_{1:m})$. Since it follows trivially from Proposition 23_{90} that $\mathscr{M}(z_{1:m}) \leq \underline{E}_Q(\mathscr{M}(z_{1:m}, X_{m+1:n})|z_{1:m})$, this allows us to infer that $\mathscr{M}(z_{1:m}) \leq \underline{E}_Q(\underline{E}_Q(g|z_{1:m}, X_{m+1:n})|z_{1:m})$. If we now use Equation $(4.3)_{88}$ [for $s := z_{1:m}$], we find that

$$\underline{E}_Q(g|z_{1:m}) \leq \underline{E}_Q(\underline{E}_Q(g|z_{1:m}, X_{m+1:n})|z_{1:m}).$$

For the converse inequality, consider any function $h \in \mathscr{L}(\mathscr{X}^n)$ such that $h(z_{1:m}, x_{m+1:n}) \leq \underline{E}_Q(g|z_{1:m}, x_{m+1:n})$ for all $x_{m+1:n} \in \mathscr{X}^{n-m}$. Fix any $\varepsilon > 0$. It then follows from Equation $(4.3)_{88}$ [for $s := (z_{1:m}, x_{m+1:n})$] that, for any $x_{m+1:n} \in \mathscr{X}^{n-m}$, there is some bounded above submartingale $\mathscr{M}_{x_{m+1:n}}$ such that

$$\mathscr{M}_{x_{m+1:n}}(z_{1:m}, x_{m+1:n}) \geq h(z_{1:m}, x_{m+1:n}) - \frac{\varepsilon}{2}$$
$$\text{and } \limsup \mathscr{M}_{x_{m+1:n}}(z_{1:m}, x_{m+1:n}\bullet) \leq g(z_{1:m}, x_{m+1:n}\bullet).$$

Now consider any $h' \in \mathscr{L}(\mathscr{X}^n)$ such that

$$h'(z_{1:m}, x_{m+1:n}) = \mathscr{M}_{x_{m+1:n}}(z_{1:m}, x_{m+1:n}) \geq h(z_{1:m}, x_{m+1:n}) - \frac{\varepsilon}{2} \quad \text{for all } x_{m+1:n} \in \mathscr{X}^{n-m},$$

then it follows from Proposition 23_{90} that there is a submartingale $\mathscr{M}'$ such that $\mathscr{M}'(z_{1:m}) > \underline{E}_Q(h'(X_{1:n})|z_{1:m}) - \frac{\varepsilon}{2}$ and $\mathscr{M}'(z_{1:m}, x_{m+1:n}) \leq h'(z_{1:m}, x_{m+1:n})$ for all $x_{m+1:n} \in \mathscr{X}^{n-m}$. Now consider a submartingale $\mathscr{M}$ that assumes the constant value $\mathscr{M}'(z_{1:m})$ in all situations $t \in \mathscr{X}^*$ such that either $t = z_{1:m}$ or $\Gamma(t) \not\subseteq \Gamma(z_{1:m})$ and that moreover

$$\Delta\mathscr{M}(z_{1:m}, x_{m+1:k}) = \begin{cases} \Delta\mathscr{M}'(z_{1:m}, x_{m+1:k}) & \text{if } k < n \\ \Delta\mathscr{M}_{x_{m+1:n}}(z_{1:m}, x_{m+1:k}) & \text{if } k \geq n \end{cases} \quad \text{for all } k \geq m \text{ and } x_{m+1:k} \in \mathscr{X}^{k-m}.$$

It then follows that $\mathscr{M}(z_{1:m}, x_{m+1:k}) \leq \mathscr{M}_{x_{m+1:n}}(z_{1:m}, x_{m+1:k})$ for all $k \geq n$ and $x_{m+1:k} \in \mathscr{X}^{k-m}$ and, therefore, we find that $\mathscr{M}$ is bounded above and that $\limsup \mathscr{M}(z_{1:m}, x_{m+1:n} \bullet) \leq g(z_{1:m}, x_{m+1:n} \bullet)$ for all $x_{m+1:n} \in \mathscr{X}^{n-m}$, which implies that $\limsup \mathscr{M}(z_{1:m} \bullet) \leq g(z_{1:m}, \bullet)$, and also that $\underline{E}_Q(g|z_{1:m}) \geq \mathscr{M}(z_{1:m})$, by applying Equation $(4.3)_{88}$ [for $s = z_{1:m}$]. Since $\mathscr{M}(z_{1:m}) = \mathscr{M}'(z_{1:m}) > \underline{E}_Q(h'(X_{1:n})|z_{1:m}) - \frac{\varepsilon}{2}$, we find that $\underline{E}_Q(g|z_{1:m}) > \underline{E}_Q(h'(X_{1:n})|z_{1:m}) - \frac{\varepsilon}{2}$. Furthermore, because $h'(z_{1:m}, X_{m+1:n}) \geq h(z_{1:m}, X_{m+1:n}) - \frac{\varepsilon}{2}$, it follows from Proposition 27_{93} and property $G4_{92}$ that

$$\begin{aligned}\underline{E}_Q(h'(X_{1:n})|z_{1:m}) =& \underline{E}_Q(h'(z_{1:m}, X_{m+1:n})|z_{1:m}) \\ \geq& \underline{E}_Q\left(h(z_{1:m}, X_{m+1:n}) - \frac{\varepsilon}{2} \middle| z_{1:m}\right) = \underline{E}_Q\left(h(X_{1:n}) - \frac{\varepsilon}{2} \middle| z_{1:m}\right),\end{aligned}$$

which, due to $G6_{92}$, implies that $\underline{E}_Q(h'(X_{1:n})|z_{1:m}) \geq \underline{E}_Q(h(X_{1:n})|z_{1:m}) - \frac{\varepsilon}{2}$. Hence, we find that $\underline{E}_Q(g|z_{1:m}) > \underline{E}_Q(h(X_{1:n})|z_{1:m}) - \varepsilon$. Since this holds for any $\varepsilon > 0$, we find that $\underline{E}_Q(g|z_{1:m}) \geq \underline{E}_Q(h(X_{1:n})|z_{1:m})$ and since this holds for any $h \in \mathscr{L}(\mathscr{X}^n)$ such that $h(z_{1:m}, x_{m+1:n}) \leq \underline{E}_Q(g|z_{1:m}, x_{m+1:n})$ for all $x_{m+1:n} \in \mathscr{X}^{n-m}$, it follows from Corollary 24_{91} that

$$\underline{E}_Q(g|z_{1:m}) \geq \underline{E}_Q(\underline{E}_Q(g|z_{1:m}, X_{m+1:n})|z_{1:m}). \qquad \square$$

The conditional upper expectations of Equation $(4.4)_{88}$ satisfy suitably adapted versions of Proposition 27_{93} and Theorem $28_{\curvearrowleft}$; they follow immediately from conjugacy and the respective version for lower expectations.

4.4 CONNECTION WITH MEASURE-THEORETIC APPROACH

Consider any imprecise stochastic process whose local models are sets of conditional probability mass functions $\mathscr{P}_{x_{1:n}}$ for all $x_{1:n} \in \mathscr{X}^*$, then we derive our global lower and upper expectations $\underline{E}_{\mathscr{P}}$ and $\overline{E}_{\mathscr{P}}$ using Equations (3.26)$_{80}$ and (3.27)$_{80}$—for $\mathbb{P}_{\mathscr{T}} = \mathbb{P}_{\mathscr{P}}$. Since from each $\mathscr{P}_{x_{1:n}}$ we can derive local models $\underline{Q}(\cdot|x_{1:n})$ and $\overline{Q}(\cdot|x_{1:n})$ using Equations (3.29)$_{82}$ and (3.30)$_{82}$, we can also derive global lower and upper expectations $\underline{E}_{\underline{Q}}$ and $\overline{E}_{\underline{Q}}$ using Equations (4.3)$_{88}$ and (4.4)$_{88}$. Therefore, we have two types of global models, namely the measure-theoretic and the martingale-theoretic ones, and our goal in this section is to investigate the connection between them. That is, we ask ourselves whether the global lower and upper expectations $\underline{E}_{\underline{Q}}$ and $\overline{E}_{\underline{Q}}$ coincide with $\underline{E}_{\mathscr{P}}$ and $\overline{E}_{\mathscr{P}}$ respectively for various types of measurable extended real-valued functions on Ω.

Starting with n-measurable functions, we find that the global models derived from the martingale-theoretic approach and the respective ones derived from the measure-theoretic approach coincide if the event on which we condition belongs to $\mathscr{X}^*$.[4]

Theorem 29. *Consider any $m,n \in \mathbb{N}_0$ such that $m \leq n$, any $x_{1:m} \in \mathscr{X}^m$ and any n-measurable function $h(X_{1:n})$. It then holds that*

$$\underline{E}_{\underline{Q}}(h(X_{1:n})|x_{1:m}) = \underline{E}_{\mathscr{P}}(h(X_{1:n})|x_{1:m});$$
$$\overline{E}_{\underline{Q}}(h(X_{1:n})|x_{1:m}) = \overline{E}_{\mathscr{P}}(h(X_{1:n})|x_{1:m}).$$

Proof. We provide the proof for the lower expectations; the proof for the upper expectations is completely similar. It follows from Theorem 28$_{94}$ and Corollary 25$_{91}$ that

$$\underline{E}_{\underline{Q}}(h(X_{1:n})|x_{1:m}) = \underline{Q}(\underline{Q}(\dots\underline{Q}(h(X_{1:n})|X_{1:n-1})\dots|X_{1:m+1})|x_{1:m}), \qquad (4.5)$$

which, by Theorem 21$_{82}$, implies that indeed

$$\underline{E}_{\underline{Q}}(h(X_{1:n})|x_{1:m}) = \underline{E}_{\mathscr{P}}(h(X_{1:n})|x_{1:m}). \qquad \square$$

In the remainder of this section, we show how the two approaches are connected for extended real-valued functions on Ω that are limits of non-decreasing sequences of n-measurable functions. This is done in several steps. First, we recall a property proved by [24, Theorem 3] that is satisfied by the global models of the martingale-theoretic approach. This property states that the global lower expectation of the extended real-valued function on Ω coincides with the limit of the global lower expectations of the n-measurable functions of the non-decreasing sequence.

[4] A similar result for imprecise Markov chains is proved in Reference [26, Section 8]. Moreover, the two approaches also coincide with Williams' natural extension. More information on Williams' natural extension can be found in References [74, 83, 84].

Theorem 30 ([24, Theorem 3])**.** *Consider any situation $s \in \mathscr{X}^*$ and any extended real-valued function g on Ω for which there is a non-decreasing sequence of n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that* $\lim_{n\to+\infty} h_n = g$. *Then*

$$\underline{E}_{\underline{Q}}(g|s) = \lim_{n\to+\infty} \underline{E}_{\underline{Q}}(h_n|s).$$

In our next step, we investigate the connection between the two approaches for stochastic processes, i.e when the local models are (precise, linear) expectations, denoted by $Q(\cdot|x_{1:n})$ for all $x_{1:n} \in \mathscr{X}^*$. In this case, the collection of all local models $Q(\cdot|x_{1:n})$ will be denoted by Q, and for any $s \in \mathscr{X}^*$, the global lower and upper expectation of any extended real-valued function g on Ω derived from the martingale-theoretic approach will be denoted by $\underline{E}_Q(g|s)$ and $\overline{E}_Q(g|s)$ respectively. Moreover, since each $Q(X_{n+1}|x_{1:n})$ can be uniquely represented by a probability mass function $p(X_{n+1}|x_{1:n})$, we have a probability tree p that—because of Theorem 1571—corresponds to a unique conditional probability measure P on $\mathscr{C}_\sigma^*$ and therefore, we infer that there is a one-to-one correspondence between Q and P. The following property tells us that $\underline{E}_Q(g|s)$ coincides with $E_P(g|s)$ when g is a limit of a non-decreasing sequence of non-negative n-measurable functions.

Theorem 31. *Consider any situation $s \in \mathscr{X}^*$ and any extended real-valued function g on Ω for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that* $\lim_{n\to+\infty} h_n = g$. *Then it holds that*

$$\underline{E}_Q(g|s) = E_P(g|s).$$

Proof. It follows from Theorem 30↶ that

$$\underline{E}_Q(g|s) = \lim_{n\to+\infty} \underline{E}_Q(h_n|s). \tag{4.6}$$

Due to Equation (3.19)75, we have that

$$E_P(g|s) = \lim_{n\to+\infty} E_P(h_n|s). \tag{4.7}$$

Moreover, it follows from Theorem 29↶ and the unicity of p and its corresponding P that for all $n \in \mathbb{N}$

$$\underline{E}_Q(h_n|s) = E_P(h_n|s),$$

which implies that

$$\lim_{n\to+\infty} \underline{E}_Q(h_n|s) = \lim_{n\to+\infty} E_P(h_n|s) \tag{4.8}$$

The result now follows from Equations (4.6) , (4.7) and (4.8). □

Before we prove our main result of this section, we present one more property that is satisfied by the global models of the martingale-theoretic approach, which goes as follows.

Lemma 32. *Consider any imprecise stochastic process whose local models are lower expectations* $\underline{Q}(\cdot|t)$*, and any stochastic process whose local models are expectations* $Q(\cdot|t)$ *such that* $Q(f|t) \geq \underline{Q}(f|t)$ *for all* $t \in \mathscr{X}^*$ *and all* $f \in \mathscr{L}(\mathscr{X})$*. For any* $s \in \mathscr{X}^*$ *and any extended real-valued function* g *on* Ω*, it then holds that*

$$\underline{E}_{\underline{Q}}(g|s) \leq \underline{E}_Q(g|s) \leq \overline{E}_Q(g|s) \leq \overline{E}_{\underline{Q}}(g|s).$$

Proof. First of all, it follows from property G5$_{92}$ that $\underline{E}_Q(g|s) \leq \overline{E}_Q(g|s)$. Regarding the rest of the inequalities, we only prove that $\underline{E}_{\underline{Q}}(g|s) \leq \underline{E}_Q(g|s)$; the proof for the upper case is completely similar.

Consider any bounded above submartingale $\mathscr{M}$ with respect to $\underline{Q}(\cdot|\cdot)$ such that $\limsup \mathscr{M}(\omega) \leq g(\omega)$ for all $\omega \in \Gamma(s)$. For all $t \in \mathscr{X}^*$, it then holds that

$$Q(\Delta\mathscr{M}(t)|t) \geq \underline{Q}(\Delta\mathscr{M}(t)|t) \geq 0,$$

where the first inequality holds because $Q(f|t) \geq \underline{Q}(f|t)$ for all $f \in \mathscr{L}(\mathscr{X})$ and the second follows from Equation (4.1)$_{87}$. Therefore, we infer that $\mathscr{M}$ is also a submartingale with respect to $Q(\cdot|\cdot)$ and consequently, we have that $\mathscr{M}(s) \leq \underline{E}_Q(g|s)$. Hence $\underline{E}_{\underline{Q}}(g|s) \leq \underline{E}_Q(g|s)$. □

Finally, we have the following property which tells us that the global lower and upper expectations of the martingale-theoretic approach are "at most as precise as"—at least as conservative as—those derived from the measure-theoretic approach. Note that for all $t \in \mathscr{X}^*$, the local models $\underline{Q}(\cdot|t)$ and $\overline{Q}(\cdot|t)$ are now given by Equations (3.29)$_{82}$ and (3.30)$_{82}$.

Theorem 33. *Consider any situation* $s \in \mathscr{X}^*$ *and any extended real-valued function* g *on* Ω *for which there is a non-decreasing sequence of non-negative n-measurable functions* $\{h_n\}_{n\in\mathbb{N}}$ *such that* $\lim_{n\to+\infty} h_n = g$*. Then it holds that*

$$\underline{E}_{\underline{Q}}(g|s) \leq \underline{E}_{\mathscr{P}}(g|s) \leq \overline{E}_{\mathscr{P}}(g|s) \leq \overline{E}_{\underline{Q}}(g|s).$$

Proof. We only prove that $\underline{E}_{\underline{Q}}(g|s) \leq \underline{E}_{\mathscr{P}}(g|s)$ and that $\overline{E}_{\mathscr{P}}(g|s) \leq \overline{E}_{\underline{Q}}(g|s)$; the proof for $\underline{E}_{\mathscr{P}}(g|s) \leq \overline{E}_{\mathscr{P}}(g|s)$ follows from Equations (3.26)$_{80}$ and (3.27)$_{80}$.

Let $\mathscr{G}$ be the set of all collections Q such that $Q(f|t) \geq \underline{Q}(f|t)$, for all $t \in \mathscr{X}^*$ and all $f \in \mathscr{L}(\mathscr{X})$, then it follows from Lemma 32 that for all $Q \in \mathscr{G}$, $\underline{E}_{\underline{Q}}(g|s) \leq \underline{E}_Q(g|s)$, which implies that

$$\underline{E}_{\underline{Q}}(g|s) \leq \inf_{Q\in\mathscr{G}} \underline{E}_Q(g|s). \tag{4.9}$$

It now follows from Theorem 31$_{\curvearrowleft}$ that each $P \in \mathbb{P}_{\mathscr{P}}$ has a corresponding collection $Q \in \mathscr{G}$ such that $\underline{E}_Q(g|s) = E_P(g|s)$. Hence

$$\inf_{Q\in\mathscr{G}} \underline{E}_Q(g|s) \leq \inf_{P\in\mathbb{P}_{\mathscr{P}}} E_P(g|s). \tag{4.10}$$

Combining now Equation (4.10)$_{\curvearrowleft}$ with Equation (3.26)$_{80}$, we have that

$$\inf_{Q \in \mathscr{G}} \underline{E}_Q(g|s) \leq \underline{E}_{\mathscr{P}}(g|s),$$

which, due to Inequality (4.9)$_{\curvearrowleft}$, implies that $\underline{E}_{\underline{Q}}(g|s) \leq \underline{E}_{\mathscr{P}}(g|s)$.

We now prove that $\overline{E}_{\underline{Q}}(g|s) \geq \overline{E}_{\mathscr{P}}(g|s)$. It follows from Lemma 32$_{\curvearrowleft}$ that for all $Q \in \mathscr{G}$, $\overline{E}_{\underline{Q}}(g|s) \geq \overline{E}_Q(g|s)$, which implies that

$$\sup_{Q \in \mathscr{G}} \overline{E}_Q(g|s) \leq \overline{E}_{\underline{Q}}(g|s). \tag{4.11}$$

It also follows from property G5$_{92}$ that $\underline{E}_Q(g|s) \leq \overline{E}_Q(g|s)$, for all $Q \in \mathscr{G}$, which implies that

$$\sup_{Q \in \mathscr{G}} \underline{E}_Q(g|s) \leq \sup_{Q \in \mathscr{G}} \overline{E}_Q(g|s). \tag{4.12}$$

Moreover, due to Theorem 31$_{97}$, each $P \in \mathbb{P}_{\mathscr{P}}$ has a corresponding collection $Q \in \mathscr{G}$ such that $\underline{E}_Q(g|s) = E_P(g|s)$ and therefore, we find that

$$\sup_{P \in \mathbb{P}_{\mathscr{P}}} E_P(g|s) \leq \sup_{Q \in \mathscr{G}} \underline{E}_Q(g|s),$$

and by combining the aforementioned inequality with Inequality (4.12), we infer that

$$\sup_{P \in \mathbb{P}_{\mathscr{P}}} E_P(g|s) \leq \sup_{Q \in \mathscr{G}} \overline{E}_Q(g|s) \tag{4.13}$$

Since $\overline{E}_{\mathscr{P}}(g|s) = \sup_{P \in \mathbb{P}_{\mathscr{P}}} E_P(g|s)$, by combining Inequality (4.13) with Inequality (4.11), we finally find that $\overline{E}_{\mathscr{P}}(g|s) \leq \overline{E}_{\underline{Q}}(g|s)$. □

In Chapter 6$_{151}$, we will see that for the function of first-passage times, which is a measurable extended real-valued function on Ω, the inequalities presented in Theorem 33$_{\curvearrowleft}$ are in fact equalities. Since we do not have an example where these inequalities are strict, we conjecture that the martingale-theoretic approach may coincide with the measure-theoretic one on the domain of measurable functions on Ω.

5

IMPRECISE DISCRETE-TIME MARKOV CHAINS

We now focus on a family of discrete-time stochastic processes, the so-called Markov chains, which are named after Andrey Andreyevich Markov who first introduced them [55]. These are stochastic processes that satisfy the Markov property, which is a "memorylessness" property, in the sense that our beliefs about the future state of the process depend only on the present state of the process and not on its past states. Due to the Markov property, Markov chains can be described by a reduced number of local models, which makes them applicable to various scientific fields and also reduces the complexity of inference problems.

We provide a brief analysis of classical discrete-time Markov chains. Similarly to the general stochastic processes that were introduced in Chapter 357, we build our Markov chains from probability trees. We also introduce two additional types of functions. The first are the functions $f(X_n)$ that depend on a single state at some time point $n \in \mathbb{N}$; we show how we can compute their expectations conditional on a state value. The second type consists of what we call time averages, which are a special family of n-measurable functions, defined as the average of $f(X_i)$ as i varies from 1 to n, and we show an alternative way for the efficient computation of their expectations. Moreover, if the Markov chain is time-homogeneous, i.e. when the local models do not depend on time, in which case we simply call it homogeneous, we present additional properties that are satisfied by the expectations of the two aforementioned types of functions in the limit as n becomes infinitely large.

After this, we discuss imprecise Markov chains, which are imprecise stochastic processes that satisfy an imprecise version of the Markov property. The existing results on imprecise Markov chains assume that the local models are credal sets—closed and convex sets of probability mass functions. We extend these results by considering general sets of conditional probability mass functions and develop new results of our own. Furthermore, we also consider different independence concepts between the variables of the global models. Choosing a different independence concept leads to differences in the global lower and upper expectations of different types of functions, but also in the efficiency of the computational methods used to calculate these expectations. More specifically, the more stringent an independence concept is, the more 'precise' are the expectations. Finally, we prove that the global lower and upper expectations of a function that depends on a single state X_n are more 'imprecise' than the respective expectations of its time average, when both expectations are taken to the limit, i.e. for $n \to +\infty$.

5.1 MARKOV CHAINS

We begin by providing useful information for modelling uncertainty in (precise) Markov chains. We first discuss the Markov property and then discuss the computation of expectations of various types of functions.

5.1.1 *The Markov property*

Consider any stochastic process characterised by a probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ and suppose that for all conditional probability measures $P \in \mathbb{P}_p$, the following equality holds:

$$P(X_{n+1}|x_{1:n}) = P(X_{n+1}|x_n), \text{ for all } n \in \mathbb{N} \text{ and all } x_{1:n} \in \mathscr{X}^n. \tag{5.1}$$

Equation (5.1) is known as the *Markov property*[1] and any discrete-time stochastic process that satisfies it is called a *Markov chain*—see Reference [43, Definition 2.1.1].[2] Equation (5.1) tells us that for all $n \in \mathbb{N}$, the state X_{n+1} is independent of the states $X_{1:n-1}$ conditional on the value of X_n. In other words, if we knew the values of the states $X_{1:n-1}$, this would not change our beliefs about X_{n+1} given the value of X_n.

We now present a sufficient and necessary condition for a probability tree to satisfy the *Markov property*.[3]

[1] Also known as Markov condition.

[2] In order to avoid confusion with the terminology used by the authors in Reference [43], we call a Markov chain what they call a Markov process and we call a homogeneous Markov chain—see Section 5.2_{109}—what they call a Markov chain.

[3] This result may seem standard, but since our measure-theoretic framework is based on the definitions of conditional probability and coherent conditional probability, it is not straightforward that the condition presented in Theorem 34 is sufficient for the Markov property to hold, in particular when the probability of x_n is zero.

Theorem 34. *A probability tree $p \in \mathbb{P}_{\mathscr{X}^*}$ satisfies the Markov property if and only if for all $n \in \mathbb{N}$ and all $x_n \in \mathscr{X}$ there is some probability mass function $q_n(X_{n+1}|x_n)$ such that*

$$p(X_{n+1}|x_{1:n-1},x_n) = q_n(X_{n+1}|x_n) \textit{ for all } x_{1:n-1} \in \mathscr{X}^{n-1}. \tag{5.2}$$

Proof. Recall that any $P \in \mathbb{P}_p$ is a coherent conditional probability on the domain $\mathscr{C}_\sigma$, and therefore it follows from the combination of Equation $(3.2)_{64}$ with Theorem 7_{64} that P is also a conditional probability, and it further follows from property CP6_{63} that for all $x_{1:k-1} \in \mathscr{X}^{k-1}$ and all $x_k, x_{k+1} \in \mathscr{X}$

$$P(x_{k+1}|x_{1:k-1},x_k) = P(x_{1:k-1},x_k,x_{k+1}|x_{1:k-1},x_k),$$

which, due to Equation $(3.3)_{65}$, implies that

$$P(x_{k+1}|x_{1:k-1},x_k) = p(x_{k+1}|x_{1:k-1},x_k). \tag{5.3}$$

For the "only if" part, it now follows from Equation $(5.1)_{\curvearrowleft}$ and (5.3) that $P(X_{n+1}|x_n) = P(X_{n+1}|x_{1:n}) = p(X_{n+1}|x_{1:n})$ for all $n \in \mathbb{N}$ and all $x_{1:n} \in \mathscr{X}^n$, which implies that $p(X_{n+1}|x_{1:n})$ does not depend on the state values $x_{1:n-1}$ and hence, Equation (5.2) is satisfied.

For the "if" part, suppose that Equation (5.2) holds. Consider any $P \in \mathbb{P}_p$, any $k \in \mathbb{N}$ and any $x_k, x_{k+1} \in \mathscr{X}$. Then it follows from Lemma 13_{68} [with $m = k-1$, $n = 1$, $c_1 = 1$ and $A_1 = \cup_{x_{1:k-1} \in \mathscr{X}^{k-1}} (x_{1:k-1}, x_k, x_{k+1})$] that

$$\min_{x_{1:k-1} \in \mathscr{X}^{k-1}} P(x_{k+1}|x_{1:k-1},x_k) \leq P(x_{k+1}|x_k) \leq \max_{x_{1:k-1} \in \mathscr{X}^{k-1}} P(x_{k+1}|x_{1:k-1},x_k). \tag{5.4}$$

Furthermore, it follows from Equations (5.2) and (5.3) that

$$P(x_{k+1}|x_{1:k-1},x_k) = p(x_{k+1}|x_{1:k-1},x_k) = q_k(x_{k+1}|x_k) \text{ for all } x_{1:k-1} \in \mathscr{X}^{k-1}. \tag{5.5}$$

If we combine Equation (5.5) with the inequalities in Equation (5.4), we find that $P(x_{k+1}|x_k) = q_k(x_{k+1}|x_k)$, which, in combination with Equation (5.5), implies that

$$P(x_{k+1}|x_k) = q_k(x_{k+1}|x_k) = P(x_{k+1}|x_{1:k-1},x_k) \text{ for all } x_{1:k-1} \in \mathscr{X}^{k-1}.$$

Hence, Equation $(5.1)_{\curvearrowleft}$ is indeed satisfied. □

Theorem $34_{\curvearrowleft}$ tells us that the probability tree p of a Markov chain satisfies Equation (5.2) and vice versa. The initial model $p(X_1|\square)$ in this case will be denoted by $q_\square(X_1)$. The set of all probability trees that satisfy Equation (5.2) will be denoted by $\mathbb{P}_\text{M}$ and clearly, $\mathbb{P}_\text{M} \subset \mathbb{P}_{\mathscr{X}^*}$.

It follows from Equation (5.2) that a Markov chain is a process such that at each time point $n \in \mathbb{N}$, the transition probability $p(x_{n+1}|x_{1:n})$ from any situation $x_{1:n} \in \mathscr{X}^n$ to a situation $x_{1:n+1} \in \mathscr{X}^{n+1}$ depends only on the last state value

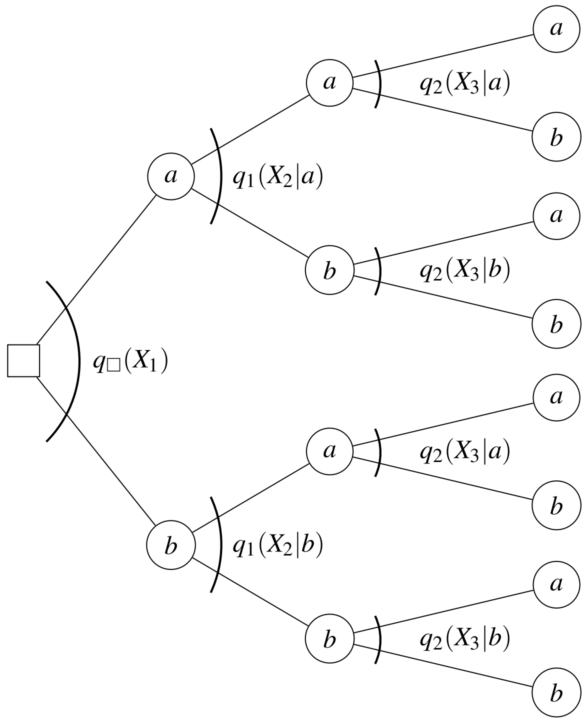

Figure 5.1: The initial part of a Markov chain with state space $\mathscr{X} = \{a,b\}$.

x_n and on the time point n. An example of a Markov chain is depicted in Figure 5.1.

Since Markov chains are special stochastic processes, they satisfy all the properties presented in Chapter 3$_{57}$. Consider now any extended real-valued function g on Ω that does not depend on the first m states $X_{1:m}$ for some $m \in \mathbb{N}$, i.e. $g(\omega) = g(\omega')$ for all $\omega, \omega' \in \Omega$ such that $\omega_k = \omega'_k$ for all $k > m$, and for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} h_n = g$. Then the expectation of any such function g satisfies the following Markov property.

Theorem 35. *Consider any Markov probability tree $p \in \mathbb{P}_{\mathrm{M}}$ and any conditional probability measure $P \in \mathbb{P}_p$. Consider as well any $m \in \mathbb{N}_0$, any $x_{1:m} \in \mathscr{X}^m$, any $x_{m+1} \in \mathscr{X}$ and any measurable extended (non-negative)*[4] *real-valued function g on Ω that does not depend on the first m states $X_{1:m}$ for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$*

[4]The result can be extended to functions that are bounded below.

such that $\lim_{n\to+\infty} h_n = g$. *Then*

$$E_P(g|x_{1:m}, x_{m+1}) = E_P(g|x_{m+1}).$$

Proof. Fix any $x'_{1:m} \in \mathscr{X}^m$ and let $\{h'_n\}_{n\in\mathbb{N}}$ be the sequence of n-measurable functions such that for all $i \in \{1,\dots,m\}$, h'_i is any i-measurable function, and for all $j \in \mathbb{N}$ such that $j > m$, we have that $h'_j(z_{1:j}) = h_j(x'_{1:m}, z_{m+1:j})$ for all $z_{1:j} \in \mathscr{X}^j$, where clearly h'_j does not depend on the first m states. Let now $g' := \lim_{n\to+\infty} h'_n$, then also g' does not depend on the first m states and moreover, for all $z_{1:m} \in \mathscr{X}^m$, we have that

$$g'(z_{1:m}\bullet) = \lim_{n\to+\infty} h'_n(z_{1:m}, X_{m+1:n}) = \lim_{n\to+\infty} h_n(x'_{1:m}, X_{m+1:n}) = g(x'_{1:m}\bullet),$$

and since g does not depend on the first m states, we infer that $g' = g$. Hence, for any measurable extended real-valued function g on Ω that does not depend on the states $X_{1:m}$ and for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} h_n = g$, we can always construct a non-decreasing sequence of non-negative n-measurable functions $\{h'_n\}_{n\in\mathbb{N}}$ such that for all $n' > m$, $h'_{n'}$ does not depend on the states $X_{1:m}$, and that $\lim_{n\to+\infty} h'_n = g$. Therefore, w.l.o.g, we can assume that for all $k > m$, h_k does not depend on the first m states $X_{1:m}$.

For all $z_{1:m} \in \mathscr{X}^m$ and all $k > m+1$, it follows from Equation (3.18)$_{74}$ that

$$E_P(h_k(X_{1:k})|z_{1:m}, x_{m+1}) \\ = \sum_{x_{m+2:k}\in\mathscr{X}^{k-m-1}} h_k(z_{1:m}, x_{m+1}, x_{m+2:k}) \prod_{i=m+1}^{k-1} p(x_{i+1}|z_{1:m}, x_{m+1:i}),$$

which due to Equation (5.2)$_{102}$ becomes

$$E_P(h_k(X_{1:k})|z_{1:m}, x_{m+1}) = \\ \sum_{x_{m+2:k}\in\mathscr{X}^{k-m-1}} h_k(z_{1:m}, x_{m+1}, x_{m+2:k}) \prod_{i=m+1}^{k-1} q_i(x_{i+1}|x_i).$$

Since h_k does not depend on the first m states $X_{1:m}$, we infer that

$$E_P(h_k(X_{1:k})|z_{1:m}, x_{m+1}) = E_P(h_k(X_{1:k})|x_{1:m}, x_{m+1}).$$

Since this is true for every $k > m+1$, we find that

$$\lim_{k\to+\infty} E_P(h_k(X_{1:k})|x_{1:m}, x_{m+1}) = \lim_{k\to+\infty} E_P(h_k(X_{1:k})|z_{1:m}, x_{m+1}),$$

and due to Equation (3.19)$_{75}$, we find that

$$E_P(g|x_{1:m}, x_{m+1}) = E_P(g|z_{1:m}, x_{m+1}).$$

Since this is true for every $z_{1:m} \in \mathscr{X}^m$, it follows from Lemma 19$_{77}$ that $E_P(g|x_{1:m}, x_{m+1}) = E_P(g|x_{m+1})$. □

5.1.2 *Expectations of functions that depend on a single state*

Apart from general n-measurable functions and extended real-valued functions that are limits of non-decreasing sequences of non-negative n-measurable ones, from now on we will also consider an additional type of functions whose expectations are often studied in Markov chains. These are n-measurable functions that only depend on the state at time n, for some $n \in \mathbb{N}$. That is, any n-measurable function $h(X_{1:n})$ such that for all $x_{1:n} \in \mathscr{X}^n$, its value is given by $f(x_n)$, for some $f \in \mathscr{L}(\mathscr{X})$. From now on, any such function will be denoted generically by $f(X_n)$.

Clearly, such a function $f(X_n)$ does not depend on the first $n-1$ states and therefore, it follows from Theorem 35_{103} that for $n > m$, $E_P(f(X_n)|x_m) = E_P(f(X_n)|x_{1:m})$ for all $x_{1:m} \in \mathscr{X}^m$. Hence, we can use Equation $(3.21)_{76}$ to compute it. However, there is also an alternative, more efficient approach for computing $E_P(f(X_n)|x_m)$ that is based on so-called *transition operators*.

Transition operators are linear operators denoted by T_n and given by

$$T_n\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})\colon f \mapsto T_n f \text{ for all } n \in \mathbb{N},$$

where, for all $f \in \mathscr{L}(\mathscr{X})$, $T_n f$ is a function in $\mathscr{L}(\mathscr{X})$ defined by

$$T_n f(x_n) := \sum_{x_{n+1} \in \mathscr{X}} f(x_{n+1}) q_n(x_{n+1}|x_n) \text{ for all } x_n \in \mathscr{X}, \tag{5.6}$$

where $q_n(X_{n+1}|x_n)$ is the local model of the Markov probability tree coming from Equation $(5.2)_{102}$. For $n = 0$, the initial model $q_\square$ is associated with an expectation operator denoted by $E_\square$. The following lemma now shows that $E_P(f(X_n)|x_m)$ can be expressed in terms of transition operators.[5]

Lemma 36. *Consider any $p \in \mathbb{P}_\mathrm{M}$ and any $P \in \mathbb{P}_p$. Consider as well any $n, m \in \mathbb{N}$ such that $n > m$, any $x_m \in \mathscr{X}$ and any function $f \in \mathscr{L}(\mathscr{X})$. Then*

$$E_P\big(f(X_n)|x_m\big) = T_m T_{m+1} \dots T_{n-1} f(x_m).$$

For the unconditional case, we have that

$$E_P\big(f(X_n)\big) = E_\square\big(T_1 \dots T_{n-1} f\big).$$

Proof. The lemma follows directly from the combination of Equations $(3.20)_{76}$ and $(3.21)_{76}$ with Theorem 35_{103} and Equations $(3.18)_{74}$, $(5.2)_{102}$ and (5.6). □

[5] A very similar result for homogeneous Markov chains can be found in Reference [81, Theorem 9.1.1].

5.1.3 *Expectations of time averages*

We introduce one more type of functions whose expectations are of interest to us. These are the so-called *time averages*. Consider any $n \in \mathbb{N}$ and any function $f \in \mathscr{L}(\mathscr{X})$, then the time average of f up to and including time point n is an n-measurable function, denoted by $[f](X_{1:n})$ and given by

$$[f](X_{1:n}) := \frac{1}{n}\sum_{i=1}^{n} f(X_i). \tag{5.7}$$

Consider any $p \in \mathbb{P}_{\mathrm{M}}$ and any $P \in \mathbb{P}_p$. Then for any $n, m \in \mathbb{N}_0$, any $x_{1:m} \in \mathscr{X}^m$ and any $f \in \mathscr{L}(\mathscr{X})$, it follows from Equation (5.7) and Lemma 102_{240} that

$$E_P\big([f](X_{1:n})|x_{1:m}\big) = \frac{1}{n}\left[\sum_{i=1}^{n} E_P\big(f(X_i)|x_{1:m}\big)\right]. \tag{5.8}$$

If $i > m$, then $E_P(f(X_i)|x_{1:m})$ is given by Lemma $36_{\frown}$. If $i \leq m$, it follows from the sentence after Equation $(3.21)_{76}$ that $E_P(f(X_i)|x_{1:m}) = f(x_i)$.

However, there is also an alternative approach for the computation of expectations of time averages that is based on transition operators, which as we will see in Section $5.4.3_{118}$, can be generalised to imprecise Markov chains. In this approach, we use a special function that is derived from transition operators. In particular, for any $k, m \in \mathbb{N}$ such that $k > m$ and any $f', f'' \in \mathscr{L}(\mathscr{X})$, the value of this function is denoted by $\xi_m^k(f', f'')$ and is a real-valued function on $\mathscr{X}$ that is defined by

$$\xi_m^k(f', f'') := \begin{cases} T_m f'' & \text{if } k = m+1 \\ \xi_m^{k-1}(f', f' + T_{k-1} f'') & \text{if } k \geq m+2. \end{cases} \tag{5.9}$$

Moreover, we let

$$\xi_m^k(f') := \xi_m^k(f', f'). \tag{5.10}$$

Observe also that for any $\lambda \in \mathbb{R}$, any $n \in \mathbb{N}$, any $x_n \in \mathscr{X}$ and any $f \in \mathscr{L}(\mathscr{X})$, it follows from Equation $(5.6)_{\frown}$ that $T_n(\lambda f)(x_n) = \lambda T_n f(x_n)$ and therefore, it follows that

$$\xi_m^k(\lambda f', \lambda f'') = \lambda \xi_m^k(f', f'') \text{ for all } f', f'' \in \mathscr{L}(\mathscr{X}) \text{ and all } m, k \in \mathbb{N} \text{ such that } k > m. \tag{5.11}$$

Finally, we can compute expectations of time averages according to the following lemma.

Lemma 37. *Consider any $p \in \mathbb{P}_{\mathrm{M}}$ and any $P \in \mathbb{P}_p$. Consider as well any $n, m \in \mathbb{N}$ such that $n > m$, any $x_{1:m} \in \mathscr{X}^m$ and any function $f \in \mathscr{L}(\mathscr{X})$. Then*

$$E_P\big([f](X_{1:n})|x_{1:m}\big) = \frac{1}{n}\left[\sum_{i=1}^{m} f(x_i) + \xi_m^n(f)(x_m)\right].$$

For the unconditional case, with $n > 1$*, we have that*

$$E_P\big([f](X_{1:n})\big) = \frac{1}{n} E_{\square}\big(f + \xi_1^n(f)\big).$$

Proof. We will only provide the proof for the conditional expectations; the proof for the unconditional ones is completely analogous.

We first prove by induction that for all $k \in \mathbb{N}$ such that $k > m$ and all $f', f'' \in \mathscr{L}(\mathscr{X})$ the following holds:

$$E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)\right] \middle| x_{1:m}\right) = \frac{1}{k}\left[\sum_{i=1}^{m} f'(x_i) + \xi_m^k(f', f'')(x_m)\right]. \quad (5.12)$$

Indeed, for $k = m+1$, we observe that

$$\begin{aligned}
&E_P\left(\frac{1}{m+1}\left[\sum_{i=1}^{m} f'(X_i) + f''(X_{m+1})\right] \middle| x_{1:m}\right) \\
&= \frac{1}{m+1} E_P\left(\sum_{i=1}^{m} f'(X_i) + f''(X_{m+1}) \middle| x_{1:m}\right) \\
&= \frac{1}{m+1} \sum_{x_{m+1} \in \mathscr{X}} \left(\sum_{i=1}^{m} f'(x_i) + f''(x_{m+1})\right) p(x_{m+1}|x_{1:m}) \\
&= \frac{1}{m+1}\left(\sum_{i=1}^{m} f'(x_i) + \sum_{x_{m+1} \in \mathscr{X}} f''(x_{m+1}) p(x_{m+1}|x_{1:m})\right) \\
&= \frac{1}{m+1}\left(\sum_{i=1}^{m} f'(x_i) + \sum_{x_{m+1} \in \mathscr{X}} f''(x_{m+1}) q_m(x_{m+1}|x_m)\right) \\
&= \frac{1}{m+1}\left(\sum_{i=1}^{m} f'(x_i) + T_m f''(x_m)\right) \\
&= \frac{1}{m+1}\left(\sum_{i=1}^{m} f'(x_i) + \xi_m^{m+1}(f', f'')(x_m)\right),
\end{aligned}$$

where the first equality follows from Lemma 102_{240}, the second from Equation $(3.18)_{74}$, the third holds because we have that $\sum_{x_{m+1} \in \mathscr{X}} p(x_{m+1}|x_{1:m}) = 1$ since $p(X_{m+1}|x_{1:m})$ is a probability mass function on $\mathscr{X}$, the fourth follows from Equation $(5.2)_{102}$, the fifth from Equation $(5.6)_{105}$ and the last from Equation $(5.9)_{\curvearrowleft}$.

Consider now any $k > m+1$ and assume that Equation (5.12) is true for $k-1$. It follows from Theorem 17_{75} that

$$\begin{aligned}
&E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)\right] \middle| x_{1:m}\right) \\
&= E_P\left(E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)\right] \middle| X_{1:k-1}\right) \middle| x_{1:m}\right) \quad (5.13)
\end{aligned}$$

and we observe that

$$
\begin{aligned}
&E_P\left(E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+f''(X_k)\right]\middle|X_{1:k-1}\right)\middle|x_{1:m}\right)\\
&\quad= E_P\left(\frac{1}{k}\sum_{i=1}^{k-1} E_P\big(f'(X_i)\big|X_{1:k-1}\big)+\frac{1}{k}E_P\big(f''(X_k)\big|X_{1:k-1}\big)\middle|x_{1:m}\right)\\
&\quad= E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+E_P(f''(X_k)|X_{1:k-1})\right]\middle|x_{1:m}\right)\\
&\quad= E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+E_P(f''(X_k)|X_{k-1})\right]\middle|x_{1:m}\right),
\end{aligned}
\tag{5.14}
$$

where the first equality follows from Lemma 102_{240}, the second follows from the argument after Equation $(5.8)_{106}$ and the last follows Theorem 35_{103}.

We now find that

$$
\begin{aligned}
&E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+f''(X_k)\right]\middle|x_{1:m}\right)\\
&= E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+E_P(f''(X_k)|X_{k-1})\right]\middle|x_{1:m}\right)\\
&= E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+T_{k-1}f''(X_{k-1})\right]\middle|x_{1:m}\right)\\
&= E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-2} f'(X_i)+f'(X_{k-1})+T_{k-1}f''(X_{k-1})\right]\middle|x_{1:m}\right)\\
&= E_P\left(\frac{k-1}{k(k-1)}\left[\sum_{i=1}^{k-2} f'(X_i)+f'(X_{k-1})+T_{k-1}f''(X_{k-1})\right]\middle|x_{1:m}\right)\\
&= E_P\left(\frac{1}{k-1}\left[\sum_{i=1}^{k-2} \frac{k-1}{k}f'(X_i)+\frac{k-1}{k}\big(f'(X_{k-1})+T_{k-1}f''(X_{k-1})\big)\right]\middle|x_{1:m}\right),
\end{aligned}
$$

where the first equality follows from Equations $(5.13)_{\curvearrowleft}$ and (5.14), the second follows from Lemma 36_{105}. Since $\frac{k-1}{k}(f'(X_{k-1})+T_{k-1}f''(X_{k-1}))$ can be regarded as a function of X_{k-1}, it follows from the aforementioned equation and the induction hypothesis that

$$\begin{aligned}
&E_P\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+f''(X_k)\right]\middle|x_{1:m}\right)\\
&\quad=\frac{1}{k-1}\left[\sum_{i=1}^{m}\frac{k-1}{k}f'(x_i)+\xi_m^{k-1}\left(\frac{k-1}{k}f',\frac{k-1}{k}(f'+T_{k-1}f'')\right)(x_m)\right]\\
&\quad=\frac{1}{k}\left[\sum_{i=1}^{m}f'(x_i)+\xi_m^{k-1}\left(f',f'+T_{k-1}f''\right)(x_m)\right]\\
&\quad=\frac{1}{k}\left[\sum_{i=1}^{m}f'(x_i)+\xi_m^{k}(f',f'')(x_m)\right],
\end{aligned}$$

where the second equality follows from Equation $(5.11)_{106}$ and the last equality follows from Equation $(5.9)_{106}$. Finally, let $k=n$ and $f'=f''=f$, then the result follows from Equation $(5.10)_{106}$. □

5.2 HOMOGENEOUS MARKOV CHAINS

We now discuss about a subclass of Markov chains, the so-called homogeneous Markov chains and we present additional properties that are satisfied by the expectations of functions that depend on a single state and of time averages when they are taken to the limit.

5.2.1 *Preliminaries*

Consider any Markov chain and suppose now that its local models depend only on the last state value and not on the time point. That is,

$$p(X_{n+1}|x_{1:n})=q(X_{n+1}|x_n)\text{ for all }x_{1:n}\in\mathscr{X}^*, \tag{5.15}$$

where $q(X_{n+1}|x_n)$ is a probability mass function on $\mathscr{X}$ for every $x_n\in\mathscr{X}$. The initial model $p(X_1|\square)$ is again denoted by $q_\square(X_1)$. Any Markov chain that satisfies Equation (5.15) is called a *homogeneous Markov chain*.[6] The set of all probability trees that satisfy Equation (5.15) will be denoted by $\mathbb{P}_{\text{HM}}$ and we observe that $\mathbb{P}_{\text{HM}}\subset\mathbb{P}_{\text{M}}\subset\mathbb{P}_{\mathscr{X}^*}$. An example of a homogeneous Markov chain is depicted in Figure $5.2_{\curvearrowright}$.

Since the transition models of a homogeneous Markov chain depend only on the state values in $\mathscr{X}$ and not on the time point n, they can be summarised by a single so-called *transition matrix*. For any homogeneous Markov chain with ordered state space $\mathscr{X}$, this transition matrix is a $|\mathscr{X}|\times|\mathscr{X}|$ matrix, where for any $i,j\in\mathscr{X}$ the element of the i-th row and j-th column represents the transition probability $q(j|i)$—also denoted by $q_{i,j}$. The form of

[6] Also called a stationary Markov chain.

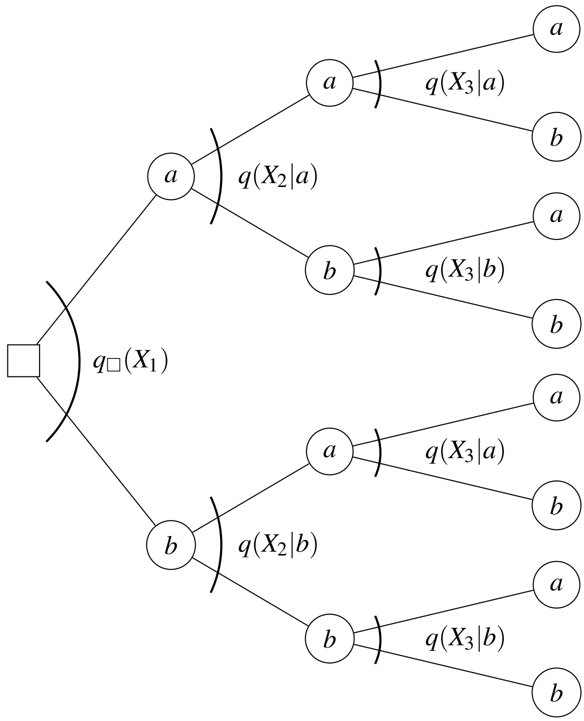

Figure 5.2: The initial part of a homogeneous Markov chain with state space $\mathscr{X} = \{a,b\}$.

the transition matrix M of a homogeneous Markov chain with state space $\mathscr{X} = \{0,\ldots,L\}$, where $L \in \mathbb{N}$, can be seen below.

$$M = \begin{pmatrix} q_{0,0} & q_{0,1} & \cdots & \cdots & q_{0,L} \\ q_{1,0} & q_{1,1} & \cdots & \cdots & q_{1,L} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ q_{L-1,0} & \cdots & \cdots & q_{L-1,L-1} & q_{L-1,L} \\ q_{L,0} & \cdots & \cdots & q_{L,L-1} & q_{L,L} \end{pmatrix}$$

Clearly, each row i of the stochastic matrix is the transition model $q(\cdot|i)$.

5.2.2 *Properties of functions that depend on a single state*

For functions that depend on a single state, we observe that all the properties presented in Section 5.1.2$_{105}$ will still hold. One difference is that since our local models do not depend on time, we now have a single transition operator

T instead of the many T_n, which is a linear operator given by

$$T\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})\colon f \mapsto Tf,$$

where, for all $f \in \mathscr{L}(\mathscr{X})$, Tf is a function in $\mathscr{L}(\mathscr{X})$ defined by

$$Tf(x) := \sum_{y\in\mathscr{X}} f(y)q(y|x) \text{ for all } x \in \mathscr{X}. \tag{5.16}$$

Obviously, the transition operator T is an alternative way to represent the transition matrix M, since $Tf = Mf$ for all $f \in \mathscr{L}(\mathscr{X})$.[7] For $n = 0$, we still have a probability mass function $p_\Box$ that is represented by an expectation operator $E_\Box$. Therefore, the (conditional) expectation of any function $f(X_n)$ is given by Lemma 36_{105}, where now T_n is replaced with T for all $n \in \mathbb{N}$.

We now introduce our definition of an *ergodic Markov chain*, which goes as follows.[8]

Definition 7. *A homogeneous Markov chain with a transition operator T is called* ergodic *if for all $f \in \mathscr{L}(\mathscr{X})$,* $\lim_{n\to+\infty} T^n f$ *exists and is a constant function.*

The reason why ergodicity is important is because it implies the existence of a unique limit expectation operator E_∞. In particular, for any ergodic homogeneous Markov chain with a transition operator T and any $f \in \mathscr{L}(\mathscr{X})$, if we let $E_\infty(f)$ be the constant value of $\lim_{n\to+\infty} T^n f$, then

$$\begin{aligned}\lim_{n\to+\infty} E_P(f(X_n)) &= \lim_{n\to+\infty} E_\Box(T^{n-1}f) = \lim_{n\to+\infty} \sum_{x\in\mathscr{X}} q_\Box(x)T^{n-1}f(x)\\ &= \sum_{x\in\mathscr{X}} q_\Box(x) \lim_{n\to+\infty} T^n f(x) = \sum_{x\in\mathscr{X}} q_\Box(x)E_\infty(f) = E_\infty(f),\end{aligned} \tag{5.17}$$

where the first equality comes from Lemma 36_{105} and the fact that the Markov chain is homogeneous. Moreover, the limit expectation operator E_∞ does not depend on the initial expectation operator $E_\Box$, and it is furthermore the only such operator that it is T-*invariant*, in the sense that $E_\infty = E_\infty \circ T$; see also References [43, Theorem 4.1.6] and [54, Chapter 7].

One obvious question that comes to mind is under which conditions a homogeneous Markov chain is ergodic. A sufficient condition for ergodicity is regularity [28, Definition 4.1].[9]

[7] This indicates that the linear operator T works on functions and is the dual of the linear operator M that is usually taken to work on probabilities.

[8] There are numerous definitions for an ergodic Markov chain that do not necessarily deal with the existence of a unique limiting distribution. More information about these definitions can be found in Reference [43, Chapter V].

[9] Definition $8_\frown$ is a special case of the definition of a regularly absorbing homogeneous imprecise Markov chain [28]—we just need to consider the case where $\overline{T} = \underline{T} = T$ and where the top class is the complete set $\mathscr{X}$. Furthermore, Definition $8_\frown$ can also be found in—amongst others—References [43, Theorem 4.1.2] and [54, Chapter 7], but in these cases it is connected with definitions for ergodicity that are different from Definition 7.

Definition 8. *Consider a homogeneous Markov chain with transition matrix M. If there is some $m \in \mathbb{N}$ such that $M^m(x,y) > 0$ for all $x,y \in \mathscr{X}$, then we call this Markov chain* regular.

There are two ways to compute the limit expectation operator E_∞ of an ergodic Markov chain and consequently the limit expectation of any $f \in \mathscr{L}(\mathscr{X})$. The first is to compute E_∞ directly by solving the system of equations implied by $E_\infty = E_\infty \circ T$. Let $\mathscr{X} = \{0,\dots,L\}$, where $L \in \mathbb{N}$, and let also $[\pi_1,\dots,\pi_L]$ be the probability mass function on $\mathscr{X}$ that corresponds to the limit operator E_∞. We can then compute $[\pi_1,\dots,\pi_L]$ by solving the following system of equations:

$$\pi_j = \sum_{i \in \mathscr{X}} \pi_i q_{i,j} \text{ for all } j \in \mathscr{X},$$

where $q_{i,j}$ are the entries of the transition matrix M, for all $i,j \in \mathscr{X}$. In other words, π^{T} is a left eigenvector of M with eigenvalue 1. The second approach is to compute the limit $\lim_{n\to+\infty} T^n$ explicitly, which also yields the expectation operator E_∞. As we will see in Section $5.5.1_{126}$, this second approach can be generalised to imprecise Markov chains.

5.2.3 *Properties of time averages*

Regarding time averages, the expectation of any time average $[f](X_{1:n})$ is given by Equation $(5.8)_{106}$ or Lemma 37_{106}, where now T_n is replaced with T for all $n \in \mathbb{N}$. Furthermore, for ergodic homogeneous Markov chains, we have the following useful property which says that the limiting expectation of any function that depends on a single state and the respective one of its time average coincide.

Theorem 38. *Consider an ergodic homogeneous Markov chain characterised by a probability tree p. Consider as well any $P \in \mathbb{P}_p$ and any $f \in \mathscr{L}(\mathscr{X})$. Then*

$$\lim_{n\to+\infty} E_P\big([f](X_{1:n})\big) = \lim_{n\to+\infty} E_P\big(f(X_n)\big) = E_\infty(f).$$

Proof. Due to Equation $(5.17)_{\curvearrowleft}$, we have that $\lim_{n\to+\infty} E_P\big(f(X_n)\big) = E_\infty(f)$, which implies that for all $\varepsilon > 0$ there is some $k_\varepsilon \in \mathbb{N}$ such that for all $k' \geq k_\varepsilon$, it holds that

$$E_\infty(f) - \varepsilon \leq E_P(f(X_{k'})) \leq E_\infty(f) + \varepsilon. \tag{5.18}$$

Observe now that

$$\liminf_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} E_P\big(f(X_i)\big)$$
$$\geq \liminf_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{k_\varepsilon - 1} E_P\big(f(X_i)\big) + \liminf_{n\to+\infty} \frac{1}{n} \sum_{i=k_\varepsilon}^{n} E_P\big(f(X_i)\big). \tag{5.19}$$

Since $\sum_{i=1}^{k_\varepsilon-1} E_P(f(X_i))$ is a constant and $n \to +\infty$, we infer that

$$\liminf_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{k_\varepsilon-1} E_P\big(f(X_i)\big) = 0. \tag{5.20}$$

Moreover, for all $n \geq k_\varepsilon$, we have that

$$\frac{1}{n} \sum_{i=k_\varepsilon}^{n} E_P\big(f(X_i)\big) \geq \frac{n-k_\varepsilon+1}{n}\big(E_\infty(f)-\varepsilon\big).$$

Hence, it follows that

$$\liminf_{n\to+\infty} \frac{1}{n} \sum_{i=k_\varepsilon}^{n} E_P\big(f(X_i)\big) \geq \liminf_{n\to+\infty} \frac{n-k_\varepsilon+1}{n}\big(E_\infty(f)-\varepsilon\big) = E_\infty(f)-\varepsilon, \tag{5.21}$$

where the equality holds because $E_\infty(f)-\varepsilon$ is constant and $\liminf_{n\to+\infty} \frac{n-k_\varepsilon}{n} = 1$. By combining Equations (5.20) and (5.21) with Inequality (5.19)$_{\curvearrowleft}$, we find that

$$\liminf_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} E_P\big(f(X_i)\big) \geq E_\infty(f)-\varepsilon. \tag{5.22}$$

Similarly,

$$\limsup_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} E_P\big(f(X_i)\big) \leq E_\infty(f)+\varepsilon. \tag{5.23}$$

It now follows from Inequalities (5.22) and (5.23) that

$$E_\infty(f)-\varepsilon \leq \liminf_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} E_P\big(f(X_i)\big) \leq \limsup_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} E_P\big(f(X_i)\big) \leq E_\infty(f)+\varepsilon.$$

Since this holds for any $\varepsilon > 0$, we infer that

$$\begin{aligned}\liminf_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} E_P\big(f(X_i)\big) &= \limsup_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} E_P\big(f(X_i)\big) \\ &= \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} E_P\big(f(X_i)\big) = E_\infty(f). \end{aligned} \tag{5.24}$$

Finally, the result follows from Equations (5.8)$_{106}$ and (5.24). □

5.3 TOWARDS IMPRECISION

We have seen that the Markov property is a condition that can be expressed in terms of the probability measure or the local models of a stochastic process, or in terms of expectations. When generalising this condition to the case of imprecise probability trees, it is conventional to express it in terms of lower (or upper) expectations. In particular, for an imprecise probability tree $\mathscr{T}$, we require that for all $n \in \mathbb{N}$, all $x_{1:n} \in \mathscr{X}^n$ and all $f \in \mathscr{L}(\mathscr{X})$:

$$\underline{E}_{\mathscr{T}}\big(f(X_{n+1})|x_{1:n}\big) = \underline{E}_{\mathscr{T}}\big(f(X_{n+1})|x_n\big) \text{ and}$$
$$\overline{E}_{\mathscr{T}}\big(f(X_{n+1})|x_{1:n}\big) = \overline{E}_{\mathscr{T}}\big(f(X_{n+1})|x_n\big), \quad (5.25)$$

where the expectations $\underline{E}_{\mathscr{T}}(f(X_{n+1})|x_{1:n})$ and $\overline{E}_{\mathscr{T}}(f(X_{n+1})|x_{1:n})$ are given by Equations $(3.26)_{80}$ and $(3.27)_{80}$ respectively. Equation (5.25) is known as the *imprecise Markov property* and any imprecise stochastic process that satisfies it is called an *imprecise Markov chain*.

Imprecise Markov chains are part of the theory of imprecise probabilities and they have been studied in—amongst others—References [28,37,47,68,75–77]. What is interesting about imprecise Markov chains is that we can adopt different independence concepts among the states of the chain. Independence concepts are often useful as they allow us to reduce the complexity of inference problems. In Sections $5.4_{\frown}$—5.7_{142}, we start from the most general independence concept and we move to more stringent ones by imposing additional constraints as we move along. In Section 5.8_{147}, we also discuss similarities and differences that arise in the global lower and upper expectations under the different independence concepts.

For each of the independence concepts that we will consider, and each of the corresponding types of imprecise Markov chains, our starting point will be a collection of local models. More specifically, for every $n \in \mathbb{N}$ and every $x \in \mathscr{X}$, we consider a set $\mathscr{Q}_{n,x}$ of conditional probability mass functions $q_n(\cdot|x)$ on $\mathscr{X}$. Furthermore, we also consider an initial model $\mathscr{Q}_\square$, which is again a set of probability mass functions on $\mathscr{X}$. Each set $\mathscr{Q}_{n,x}$ has a corresponding lower and upper expectation, which are denoted by $\underline{Q}_n(\cdot|x)$ and $\overline{Q}_n(\cdot|x)$ and defined by

$$\underline{Q}_n(f|x) := \inf\left\{\sum_{y\in\mathscr{X}} f(y)p(y) \colon p \in \mathscr{Q}_{n,x}\right\} \quad (5.26)$$

$$\overline{Q}_n(f|x) := \sup\left\{\sum_{y\in\mathscr{X}} f(y)p(y) \colon p \in \mathscr{Q}_{n,x}\right\} \quad (5.27)$$

for all $f \in \mathscr{L}(\mathscr{X})$. The lower and upper expectations that correspond to the initial model $\mathscr{Q}_\square$ are denoted by $\underline{Q}_\square(\cdot)$ and $\overline{Q}_\square(\cdot)$ respectively.

We will also consider the case where $\mathscr{Q}_{n,x}$ does not depend on n, and will then be denoted by $\mathscr{Q}_x$. Each $\mathscr{Q}_x$ has a corresponding lower and upper expectation, which are now denoted by $\underline{Q}(\cdot|x)$ and $\overline{Q}(\cdot|x)$ and defined by

$$\underline{Q}(f|x) := \inf\left\{\sum_{y\in\mathscr{X}} f(y)p(y) \colon p \in \mathscr{Q}_x\right\} \quad (5.28)$$

$$\overline{Q}(f|x) := \sup\left\{\sum_{y\in\mathscr{X}} f(y)p(y) \colon p \in \mathscr{Q}_x\right\} \quad (5.29)$$

for all $f \in \mathscr{L}(\mathscr{X})$. For the initial situation, we again have the lower and upper expectations $\underline{Q}_{\square}(\cdot)$ and $\overline{Q}_{\square}(\cdot)$.

5.4 IMPRECISE MARKOV CHAINS UNDER EPISTEMIC IRRELEVANCE

We start by introducing the independence concept of epistemic irrelevance in imprecise Markov chains, and we discuss how to compute lower and upper expectations of various functions for this type of imprecise Markov chain.

5.4.1 *Epistemic irrelevance*

In Section 3.780, we introduced a specific way for obtaining an imprecise stochastic process, by using sets of conditional probability mass functions. We started by associating with each situation $x_{1:n} \in \mathscr{X}^*$ a non-empty set of conditional probability mass functions $\mathscr{P}_{x_{1:n}}$ on $\mathscr{X}$ and then derived an imprecise probability tree $\mathscr{T}_{\mathscr{P}}$ using Equation (3.28)80.

We here consider the special case where the sets $\mathscr{P}_{x_{1:n}}$ are derived from the local models in Section 5.3113. In particular, we assume that

$$\mathscr{P}_{x_{1:n}} = \mathscr{Q}_{n,x_n} \text{ for all } n \in \mathbb{N} \text{ and all } x_{1:n} \in \mathscr{X}^n \tag{5.30}$$

and that $\mathscr{P}_{\square} = \mathscr{Q}_{\square}$. In this case, we use $\mathscr{T}_{\mathscr{Q}}$ as an alternative notation for $\mathscr{T}_{\mathscr{P}}$.

Since the imprecise probability tree $\mathscr{T}_{\mathscr{Q}}$ is a set of probability trees, it has a corresponding set of conditional probability measures on $\mathscr{C}_{\sigma}$, which we denote by $\mathbb{P}_{\mathscr{Q}}$. Furthermore, since each set $\mathscr{P}_{x_{1:n}}$ can be associated with a lower and an upper expectation operator, denoted by $\underline{Q}(\cdot|x_{1:n})$ and $\overline{Q}(\cdot|x_{1:n})$ respectively, due to Equation (5.30), we now have that $\underline{Q}(\cdot|\square) = \underline{Q}_{\square}(\cdot)$ and $\overline{Q}(\cdot|\square) = \overline{Q}_{\square}(\cdot)$, and that

$$\underline{Q}(\cdot|x_{1:n}) = \underline{Q}_n(\cdot|x_n) \text{ and } \overline{Q}(\cdot|x_{1:n}) = \overline{Q}_n(\cdot|x_n) \text{ for all } n \in \mathbb{N} \text{ and all } x_{1:n} \in \mathscr{X}^n, \tag{5.31}$$

where $\underline{Q}_n(\cdot|x_n)$ and $\overline{Q}_n(\cdot|x_n)$ are given by Equations (5.26)↶ and (5.27)↶.

The imprecise stochastic process that we have just derived by using the local models $\mathscr{Q}_{n,x}$ and $\mathscr{Q}_{\square}$ is called an imprecise Markov chain under *epistemic irrelevance* [10, 23, 27, 28, 78]. The reason why it is called like that is because the variables $X_{1:n-1}$ are epistemically irrelevant to X_{n+1} given X_n, which means that if we already know X_n, then additionally observing $X_{1:n-1}$ does not affect our beliefs about X_{n+1}. Epistemic irrelevance is therefore clearly a notion of independence. It is called 'epistemic' in order to emphasise that it is a statement about beliefs; in this case, beliefs about local probability mass functions, in the form of sets of candidates for them. It is called 'irrelevance'—instead of 'independence'—to emphasise that it is asymmetric. An example of an imprecise Markov chain under epistemic irrelevance is shown in Figure 5.3↷. Since

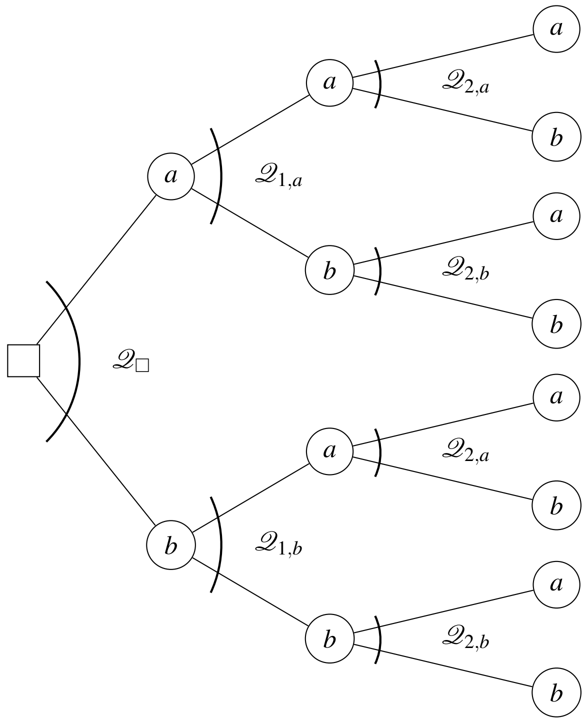

Figure 5.3: The initial part of an imprecise Markov chain under epistemic irrelevance with state space $\mathscr{X} = \{a, b\}$.

an imprecise Markov chain under epistemic irrelevance is a special case of a general imprecise stochastic process, all properties presented in Sections 3.6$_{79}$ and 3.7$_{80}$ will still hold. Global lower and upper expectations of measurable extended real-valued functions g on Ω conditional on $B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\}$ in imprecise Markov chains under epistemic irrelevance are defined by Equations (3.26)$_{80}$ and (3.27)$_{80}$—for $\mathbb{P}_{\mathscr{T}} = \mathbb{P}_{\mathscr{Q}}$—and will be denoted by $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|B)$ and $\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|B)$ respectively. In the coming sections, we show that these global lower and upper expectations of different types of functions satisfy various versions of the imprecise Markov property and focus on the computational aspects of these global expectations.

5.4.2 *Global lower and upper expectations of functions that depend on a single state*

We first show that the global lower and upper expectations of functions that depend on a single state satisfy an *imprecise Markov property* when the independence concept is epistemic irrelevance. This imprecise Markov property is a direct consequence of Theorem 44_{123} and therefore, it is stated without proof.

Theorem 39. *Consider an imprecise Markov chain under epistemic irrelevance. Consider as well any $m,n \in \mathbb{N}$ such that $n > m$, any $x_{1:m-1} \in \mathscr{X}^{m-1}$, any $x_m \in \mathscr{X}$ and any function $f \in \mathscr{L}(\mathscr{X})$. Then*

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_m) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_{1:m-1},x_m);$$
$$\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_m) = \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_{1:m-1},x_m).$$

In the case of Markov chains, we showed that we can compute expectations of functions that depend on single state using transition operators. We now generalise the idea of transition operators to imprecise Markov chains under epistemic irrelevance. We introduce lower and upper transition operators [28], which will be used for computing global lower and upper expectations of different types of functions. For any $n \in \mathbb{N}$, this time we have the non-linear operators $\underline{T}_n$ and $\overline{T}_n$ given by

$$\underline{T}_n \colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) \colon f \mapsto \underline{T}_n f;$$
$$\overline{T}_n \colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X}) \colon f \mapsto \overline{T}_n f,$$

where for all $f \in \mathscr{L}(\mathscr{X})$, $\underline{T}_n f$ and $\overline{T}_n f$ are functions in $\mathscr{L}(\mathscr{X})$ defined by

$$\underline{T}_n f(x) := \underline{Q}_n(f|x) = \inf\left\{ \sum_{y\in\mathscr{X}} f(y)p(y) \colon p \in \mathscr{Q}_{n,x} \right\} \tag{5.32}$$

$$\overline{T}_n f(x) := \overline{Q}_n(f|x) = \sup\left\{ \sum_{y\in\mathscr{X}} f(y)p(y) \colon p \in \mathscr{Q}_{n,x} \right\} \tag{5.33}$$

for all $x \in \mathscr{X}$.

Our next result is a generalisation of Lemma 36_{105} and states that the global lower and upper expectation of any function that depends on a single state can be expressed in terms of lower and upper transition operators.[10]

Lemma 40. *Consider an imprecise Markov chain under epistemic irrelevance. Consider as well any $m,n \in \mathbb{N}$ such that $n > m$, any $x_m \in \mathscr{X}$ and any function $f \in \mathscr{L}(\mathscr{X})$. Then*

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_m) = \underline{T}_m \underline{T}_{m+1} \cdots \underline{T}_{n-1} f(x_m);$$
$$\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_m) = \overline{T}_m \overline{T}_{m+1} \cdots \overline{T}_{n-1} f(x_m).$$

[10] A similar version of Lemma 40 for imprecise Markov chains whose local models are credal sets and where the functions under study are general n-measurable functions can be found in Reference [28].

For the unconditional case, we have that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)) = \underline{Q}_{\square}(\underline{T}_1 \dots \underline{T}_{n-1} f);$$
$$\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)) = \overline{Q}_{\square}(\overline{T}_1 \dots \overline{T}_{n-1} f).$$

Proof. The statement of the lemma follows directly from the combination of Theorem 39↶, Equations (5.32)↶ and (5.33)↶ with Theorem 21_{82}. □

The practical importance of this result is that it provides us with an efficient method for computing lower and upper expectations of functions that depend on a single state. The reason for this efficiency, is because the number of local optimisation problems that we need to solve is linear with respect to the number of time points that are considered. Indeed, at each time point, evaluating the lower or upper transition operator only requires us to compute $|\mathscr{X}|$ (lower or upper) expected values. The resulting function is then plugged into the lower or upper transition operator associated with the previous time point and we repeat this procedure till we reach the situation on which we conditioned.

5.4.3 *Global lower and upper expectations of time averages*

Regarding global lower and upper expectations of time averages, we can compute them using functions that are based on lower and upper transition operators. For any $k,m \in \mathbb{N}$ such that $k > m$ and any $f', f'' \in \mathscr{L}(\mathscr{X})$, we now introduce $\underline{\xi}^k_m(f', f'')$ and $\overline{\xi}^k_m(f', f'')$, which are real-valued functions on $\mathscr{X}$ that are defined by

$$\underline{\xi}^k_m(f', f'') := \begin{cases} \underline{T}_m f'' & \text{if } k = m+1 \\ \underline{\xi}^{k-1}_m(f', f' + \underline{T}_{k-1} f'') & \text{if } k \geq m+2. \end{cases} \tag{5.34}$$

and

$$\overline{\xi}^k_m(f', f'') := \begin{cases} \overline{T}_m f'' & \text{if } k = m+1 \\ \overline{\xi}^{k-1}_m(f', f' + \overline{T}_{k-1} f'') & \text{if } k \geq m+2. \end{cases} \tag{5.35}$$

We also let

$$\underline{\xi}^k_m(f') := \underline{\xi}^k_m(f', f') \text{ and } \overline{\xi}^k_m(f') := \overline{\xi}^k_m(f', f'). \tag{5.36}$$

Observe that for any $\lambda \in \mathbb{R}_{\geq 0}$, any $n \in \mathbb{N}$, any $x_n \in \mathscr{X}$ and any $f \in \mathscr{L}(\mathscr{X})$, since $\underline{T}_n(\lambda f)$ is an infimum of expectations it follows from Lemma 102_{240} that $\underline{T}_n(\lambda f)(x_n) = \lambda \underline{T}_n f(x_n)$ and therefore, it follows that

$$\underline{\xi}^k_m(\lambda f', \lambda f'') = \lambda \underline{\xi}^k_m(f', f'') \text{ and } \overline{\xi}^k_m(\lambda f', \lambda f'') = \lambda \overline{\xi}^k_m(f', f'') \tag{5.37}$$

for all $f', f'' \in \mathscr{L}(\mathscr{X})$ and all $m,k \in \mathbb{N}$ such that $k > m$.

We are now ready to present the following lemma which is a generalisation of Lemma 37_{106}.[11]

Lemma 41. *Consider an imprecise Markov chain under epistemic irrelevance. Consider as well any $m,n \in \mathbb{N}$ such that $n > m$, any $x_{1:m} \in \mathscr{X}^m$ and any function $f \in \mathscr{L}(\mathscr{X})$. Then*

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) = \frac{1}{n}\left[\sum_{i=1}^{m} f(x_i) + \underline{\xi}^{n}_{m}(f)(x_m)\right];$$
$$\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) = \frac{1}{n}\left[\sum_{i=1}^{m} f(x_i) + \overline{\xi}^{n}_{m}(f)(x_m)\right].$$

For the unconditional case, with $n > 1$, we have that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})) = \frac{1}{n}\underline{Q}_{\square}\big(f + \underline{\xi}^{n}_{1}(f)\big);$$
$$\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})) = \frac{1}{n}\overline{Q}_{\square}\big(f + \overline{\xi}^{n}_{1}(f)\big).$$

Proof. We will only provide the proof for the global lower expectations conditional on $x_{1:m}$; the proofs for the global upper ones and for the unconditional case are completely analogous.

We first prove by induction that for all $k \in \mathbb{N}$ such that $k > m$ and all $f', f'' \in \mathscr{L}(\mathscr{X})$ the following holds:

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)\right]\middle| x_{1:m}\right) = \frac{1}{k}\left[\sum_{i=1}^{m} f'(x_i) + \underline{\xi}^{k}_{m}(f',f'')(x_m)\right]. \tag{5.38}$$

For $k = m+1$, we observe that

$$\begin{aligned}
&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{m+1}\left[\sum_{i=1}^{m} f'(X_i) + f''(X_{m+1})\right]\middle| x_{1:m}\right)\\
&= \frac{1}{m+1}\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\left[\sum_{i=1}^{m} f'(X_i) + f''(X_{m+1})\right]\middle| x_{1:m}\right)\\
&= \frac{1}{m+1}\underline{Q}\left(\left[\sum_{i=1}^{m} f'(x_i) + f''(X_{m+1})\right]\middle| x_{1:m}\right)\\
&= \frac{1}{m+1}\left[\sum_{i=1}^{m} f'(x_i) + \underline{Q}(f''(X_{m+1})|x_{1:m})\right],
\end{aligned} \tag{5.39}$$

where the first equality follows from the definition of $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}$ combined with Lemma 102_{240}, the second from Theorem 21_{82} and Equation $(3.31)_{82}$, and

[11] For the more general case of credal networks under epistemic irrelevance, an extended version of this result can be found in Reference [21], under the additional assumption that the local models are closed and convex.

the last equality holds because, since $\sum_{i=1}^{m} f'(x_i)$ is a constant and $\underline{Q}(\cdot|x_{1:m})$ is an infimum of expectations, it follows from Lemmas 102_{240} and 106_{241} that $\sum_{i=1}^{m} f'(x_i)$ can be taken out of the infimum.

Furthermore, we find that

$$\begin{aligned}
&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{m+1}\left[\sum_{i=1}^{m} f'(X_i)+f''(X_{m+1})\right]\middle|x_{1:m}\right)\\
&=\frac{1}{m+1}\left[\sum_{i=1}^{m} f'(x_i)+\underline{Q}(f''(X_{m+1})|x_{1:m})\right]\\
&=\frac{1}{m+1}\left[\sum_{i=1}^{m} f'(x_i)+\underline{Q}_m(f''(X_{m+1})|x_m)\right]\\
&=\frac{1}{m+1}\left[\sum_{i=1}^{m} f'(x_i)+\underline{T}_m f''(x_m)\right]=\frac{1}{m+1}\left[\sum_{i=1}^{m} f'(x_i)+\underline{\xi}_m^{m+1}(f',f'')(x_m)\right],
\end{aligned}$$

where the first equality follows from Equation $(5.39)_{\curvearrowleft}$, the second equality from Equation $(5.31)_{115}$, the third equality from Equation $(5.32)_{117}$ and the last equality from Equation $(5.34)_{118}$.

Now consider any $k>m+1$ and assume that Equation $(5.38)_{\curvearrowleft}$ is true for $k-1$. It follows from Theorem 21_{82} that

$$\begin{aligned}
&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+f''(X_k)\right]\middle|x_{1:m}\right)\\
&=\underline{Q}\left(\dots\underline{Q}\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+f''(X_k)\right]\middle|X_{1:k-1}\right)\dots\middle|x_{1:m}\right)\\
&=\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\underline{Q}\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+f''(X_k)\right]\middle|X_{1:k-1}\right)\middle|x_{1:m}\right). \qquad (5.40)
\end{aligned}$$

Consider also any $z_{1:k-1}\in\mathscr{X}^{k-1}$ and observe that

$$\begin{aligned}
&\underline{Q}\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+f''(X_k)\right]\middle|z_{1:k-1}\right)=\underline{Q}\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(z_i)+f''(X_k)\right]\middle|z_{1:k-1}\right)\\
&=\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(z_i)+\underline{Q}(f''(X_k)|z_{1:k-1})\right]=\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(z_i)+\underline{Q}_{k-1}(f''(X_k)|z_{k-1})\right]
\end{aligned} \qquad (5.41)$$

where the first equality follows from Equation $(3.31)_{82}$, the last equality follows from Equation $(5.31)_{115}$ and the second holds because, since $\sum_{i=1}^{k-1} f'(z_i)$ is a constant and $\underline{Q}(\cdot|z_{1:k-1})$ is an infimum of expectations, it follows from Lemmas 106_{241} and 102_{240} that $\sum_{i=1}^{m} f'(x_i)$ and $\frac{1}{k}$ can be taken out of the infimum. Furthermore, observe that

$$
\begin{aligned}
&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+f''(X_k)\right]\middle|x_{1:m}\right)\\
&=\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+\underline{Q}_{k-1}(f''(X_k)|X_{k-1})\right]\middle|x_{1:m}\right)\\
&=\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+\underline{T}_{k-1}f''(X_{k-1})\right]\middle|x_{1:m}\right)\\
&=\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{k}\left[\sum_{i=1}^{k-2} f'(X_i)+f'(X_{k-1})+\underline{T}_{k-1}f''(X_{k-1})\right]\middle|x_{1:m}\right)\\
&=\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{k-1}{k(k-1)}\left[\sum_{i=1}^{k-2} f'(X_i)+f'(X_{k-1})+\underline{T}_{k-1}f''(X_{k-1})\right]\middle|x_{1:m}\right)\\
&=\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{k-1}\left[\sum_{i=1}^{k-2} \frac{k-1}{k}f'(X_i)+\frac{k-1}{k}\Big(f'(X_{k-1})+\underline{T}_{k-1}f''(X_{k-1})\Big)\right]\middle|x_{1:m}\right),
\end{aligned}
\tag{5.42}
$$

where the first equality follows from the combination of Equation (5.40)$_{\curvearrowleft}$ with Equation (5.41)$_{\curvearrowleft}$ and the second equality follows from Equation (5.32)$_{117}$. If we regard now $\frac{k-1}{k}(f'(X_{k-1})+\underline{T}_{k-1}f''(X_{k-1}))$ as a function on $\mathscr{X}$, it follows from Equation (5.42) and the inductions hypothesis that

$$
\begin{aligned}
&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{k}\left[\sum_{i=1}^{k-1} f'(X_i)+f''(X_k)\right]\middle|x_{1:m}\right)\\
&\qquad=\frac{1}{k-1}\left[\sum_{i=1}^{m}\frac{k-1}{k}f'(x_i)+\underline{\xi}_m^{k-1}\left(\frac{k-1}{k}f',\frac{k-1}{k}f'+\underline{T}_{k-1}f''\right)(x_m)\right]\\
&\qquad=\frac{1}{k}\left[\sum_{i=1}^{m} f'(x_i)+\underline{\xi}_m^{k-1}(f',f'+\underline{T}_{k-1}f'')(x_m)\right]\\
&\qquad=\frac{1}{k}\left[\sum_{i=1}^{m} f'(x_i)+\underline{\xi}_m^{k}(f',f'')(x_m)\right],
\end{aligned}
$$

where the second equality follows from Equation (5.37)$_{118}$ and the last equality follows from Equation (5.34)$_{118}$. Finally, let $k=n$ and $f'=f''=f$, then the result follows from Equation (5.36)$_{118}$. □

Similarly to what happened for functions that depend on a single state, we find that lower and upper transition operators facilitate the computation of lower and upper expectations of time averages. As we can see from Equations (5.34)$_{118}$ and (5.35)$_{118}$ and Lemma 41$_{119}$, the only difference is that we additionally plug the function f into every lower or upper transition operator. Here too, this method allows for efficient computations whose complexity is linear in the number of time points.

5.4.4 *Global lower and upper expectations of measurable extended real-valued functions*

In this section, we discuss global lower and upper expectations of extended (non-negative) real-valued functions g on Ω that do not depend on the first m states $X_{1:m}$ for some $m \in \mathbb{N}$, and for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} h_n = g$. These global lower and upper expectations satisfy an imprecise Markov property, which is stated in Theorem $44_{\curvearrowright}$.

Before we present Theorem $44_{\curvearrowright}$, we first show two properties that are satisfied by the aforementioned global lower and upper expectations which will be used in the proof of Theorem $44_{\curvearrowright}$.

Lemma 42. *Consider an imprecise Markov chain under epistemic irrelevance. Consider as well any $P \in \mathbb{P}_{\mathscr{Q}}$, any $m \in \mathbb{N}$, any $x_{1:m} \in \mathscr{X}^m$, any $x_{m+1} \in \mathscr{X}$ and any extended real-valued function g on Ω that does not depend on the first m states $X_{1:m}$ for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} h_n = g$. Then there is some $P' \in \mathbb{P}_{\mathscr{Q}}$ such that*

$$E_P(g|x_{1:m},x_{m+1}) = E_{P'}(g|z_{1:m},x_{m+1}) \text{ for all } z_{1:m} \in \mathscr{X}^m.$$

Proof. Let $p \in \mathscr{T}_{\mathscr{Q}}$ be the probability tree that corresponds to P. It follows from Equations $(5.30)_{115}$ and $(3.28)_{80}$ that there is some $p' \in \mathscr{T}_{\mathscr{Q}}$ such that for all $i \in \mathbb{N}\setminus\{1,\dots,m+1\}$ and all $z'_{1:i} \in \mathscr{X}^i$,

$$p'(X_{i+1}|z'_{1:i}) = p(X_{i+1}|x_{1:m},x_{m+1},z'_{m+2:i}).$$

Let $P' \in \mathbb{P}_{\mathscr{Q}}$ be any conditional probability measure whose corresponding probability tree is p'. Consider now any $z_{1:m} \in \mathscr{X}^m$, then it follows from the argumentation in the beginning of proof of Theorem 35_{103} that, w.l.o.g, we can assume for all $k > m$, that h_k does not depend on the first m states $X_{1:m}$ and therefore for all $k > m+1$, due to Equation $(3.18)_{74}$ and the properties of p' and p, that

$$\begin{aligned}
E_P(h_k(X_{1:k})|x_{1:m+1}) &= \sum_{x_{m+2:k}\in\mathscr{X}^{k-m-1}} h(x_{1:m},x_{m+1},x_{m+2:k}) \prod_{j=m+1}^{k-1} p(x_{j+1}|x_{1:j}) \\
&= \sum_{x_{m+2:k}\in\mathscr{X}^{k-m-1}} h(x_{1:m},x_{m+1},x_{m+2:k}) \prod_{j=m+1}^{k-1} p'(x_{j+1}|z_{1:m},x_{m+1:j}) \\
&= \sum_{x_{m+2:k}\in\mathscr{X}^{k-m-1}} h(z_{1:m},x_{m+1},x_{m+2:k}) \prod_{j=m+1}^{k-1} p'(x_{j+1}|z_{1:m},x_{m+1:j}) \\
&= E_{P'}(h_k(X_{1:k})|z_{1:m},x_{m+1}).
\end{aligned}$$

Since $E_P(h_k(X_{1:k})|x_{1:m}, x_{m+1}) = E_{P'}(h_k(X_{1:k})|z_{1:m}, x_{m+1})$ for all $k > m+1$, we further infer that

$$\lim_{n\to+\infty} E_P(h_n(X_{1:n})|x_{1:m}, x_{m+1}) = \lim_{n\to+\infty} E_{P'}(h_n(X_{1:n})|z_{1:m}, x_{m+1}),$$

which, due to Equation (3.19)$_{75}$, implies that

$$E_P(g|x_{1:m}, x_{m+1}) = E_{P'}(g|z_{1:m}, x_{m+1}). \qquad \square$$

Lemma 43. *Consider an imprecise Markov chain under epistemic irrelevance. Consider as well any $m \in \mathbb{N}$, any $x_{1:m} \in \mathscr{X}^m$, any $x_{m+1} \in \mathscr{X}$ and any extended real-valued function g on Ω that does not depend on the first m states $X_{1:m}$ for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that* $\lim_{n\to+\infty} h_n = g$. *Then*

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m}, x_{m+1}) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|z_{1:m}, x_{m+1}) \textit{ for all } z_{1:m} \in \mathscr{X}^m.$$

Proof. Fix any $\varepsilon > 0$. Then there is some $P \in \mathbb{P}_{\mathscr{Q}}$ such that

$$E_P(g|x_{1:m}, x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m}, x_{m+1}) + \varepsilon. \tag{5.43}$$

Consider any $z_{1:m} \in \mathscr{X}^m$, then it follows from Lemma 42$_{\curvearrowleft}$ that there is some $P' \in \mathbb{P}_{\mathscr{Q}}$ such that

$$E_{P'}(g|z_{1:m}, x_{m+1}) = E_P(g|x_{1:m}, x_{m+1}). \tag{5.44}$$

Since $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|z_{1:m}, x_{m+1}) \leq E_{P'}(g|z_{1:m}, x_{m+1})$, it follows from Equation (5.44) and Inequality (5.43) that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|z_{1:m}, x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m}, x_{m+1}) + \varepsilon$ and since this holds for any $\varepsilon > 0$, we infer that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|z_{1:m}, x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m}, x_{m+1})$. The result now follows from symmetry. $\square$

Theorem 44. *Consider an imprecise Markov chain under epistemic irrelevance. Consider as well any $m \in \mathbb{N}$, any $x_{1:m} \in \mathscr{X}^m$, any $x_{m+1} \in \mathscr{X}$ and any extended real-valued function g on Ω that does not depend on the first m states $X_{1:m}$ for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that* $\lim_{n\to+\infty} h_n = g$. *Then*

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m}, x_{m+1}) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{m+1}) \textit{ and } \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m}, x_{m+1}) = \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{m+1}).$$

Proof. We will only provide the proof for the global lower expectations; the proof for the global upper ones is completely analogous.

We first prove that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m}, x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{m+1})$. Fix any $\varepsilon > 0$. Then there is some $P \in \mathbb{P}_{\mathscr{Q}}$ such that

$$E_P(g|x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{m+1}) + \varepsilon. \tag{5.45}$$

It follows from Lemma 19_{77} that

$$\min_{z_{1:m}\in\mathscr{X}^m} E_P(g|z_{1:m},x_{m+1}) \leq E_P(g|x_{m+1}),$$

and since $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|z_{1:m},x_{m+1}) \leq E_P(g|z_{1:m},x_{m+1})$ for all $z_{1:m}\in\mathscr{X}^m$, we find that

$$\min_{z_{1:m}\in\mathscr{X}^m} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|z_{1:m},x_{m+1}) \leq E_P(g|x_{m+1}). \tag{5.46}$$

It now follows from Lemma $43_{\curvearrowleft}$ that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m},x_{m+1}) = \min_{z_{1:m}\in\mathscr{X}^m} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|z_{1:m},x_{m+1}). \tag{5.47}$$

By combining Equation (5.47) with Inequalities (5.46) and $(5.45)_{\curvearrowleft}$, we infer that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m},x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{m+1})+\varepsilon$, and since this holds for any $\varepsilon>0$, we infer that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m},x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{m+1})$.

It now remains to prove that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m},x_{m+1})$. Fix any $\varepsilon>0$. Then there is some $P^*\in\mathbb{P}_{\mathscr{Q}}$ such that

$$E_{P^*}(g|x_{1:m},x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m},x_{m+1})+\varepsilon. \tag{5.48}$$

It now follows from Lemma 42_{122} that there is some $P'\in\mathbb{P}_{\mathscr{Q}}$ such that

$$E_{P'}(g|z_{1:m},x_{m+1}) = E_{P^*}(g|x_{1:m},x_{m+1}) \text{ for all } z_{1:m}\in\mathscr{X}^m. \tag{5.49}$$

Also, it follows from Equation (5.49) and Lemma 19_{77} that $E_{P'}(g|z_{1:m},x_{m+1}) = E_{P'}(g|x_{m+1})$ for all $z_{1:m}\in\mathscr{X}^m$, and therefore, due to Equation (5.49), we find that

$$E_{P'}(g|x_{m+1}) = E_{P^*}(g|x_{1:m},x_{m+1}). \tag{5.50}$$

Since $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{m+1}) \leq E_{P'}(g|x_{m+1})$, it follows from Equation (5.50) and Inequality (5.48) that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m},x_{m+1})+\varepsilon$, and since this holds for any $\varepsilon>0$, we infer that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{m+1}) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|x_{1:m},x_{m+1})$, which completes the proof. □

5.4.5 *A martingale-theoretic approach for global lower and upper expectations*

So far, we have constructed global lower and upper expectations in imprecise Markov chains under epistemic irrelevance using the measure-theoretic approach. We now investigate how we can construct them using the martingale-theoretic approach. From the analysis that took place in Chapter 4_{86}, we know that the martingale-theoretic approach works with local models that are lower and upper expectation operators rather than sets of conditional probability mass functions. Therefore, instead of working with the sets of conditional probability mass functions $\mathscr{Q}_{n,x}$, we here consider their corresponding lower and upper expectation operators $\underline{Q}_n(\cdot|x)$ and $\overline{Q}_n(\cdot|x)$, with $n \in \mathbb{N}$ and $x \in \mathscr{X}$, and use Equation $(5.31)_{115}$ to define our local models. For any situation $s \in \mathscr{X}^*$ and any extended real-valued function g on Ω, the global lower and upper expectations of g conditional on s are then defined by Equations $(4.3)_{88}$ and $(4.4)_{88}$ and will be denoted by $\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s)$ and $\overline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s)$ respectively.

Since martingale-theoretic imprecise Markov chains under epistemic irrelevance can be regarded as a special case of general martingale-theoretic imprecise stochastic processes, we infer that all the properties presented in Sections $4.3.2_{89}$ and 4.4_{95} will still hold. Hence, it follows from Theorem 29_{96} that the global lower and upper expectations of functions that depend on a finite number of states, for instance functions that depend on a single state and time averages, defined by the martingale-theoretic approach will coincide with the respective ones defined by the measure-theoretic approach. However, we were only able to prove inequalities, not equalities, for global lower and upper expectations of extended real-valued functions that are limits of non-decreasing sequences of non-negative n-measurable functions; see Theorem 33_{98}.

Note that the martingale-theoretic approach allows conditioning only on single situations. Therefore, it might seem as if the global models defined by this approach do not satisfy the imprecise Markov property, because the latter needs conditioning on multiple situations. Fortunately, these global models satisfy a property that is similar to the imprecise Markov one. Specifically, for any $m \in \mathbb{N}_0$, any situation $s \in \mathscr{X}^m$, any $x \in \mathscr{X}$ and any extended real-valued function g on Ω that does not depend on the first m states, we can show that the global lower and upper expectations $\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x)$ and $\overline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x)$ do not depend on s, in the sense that for any situation $t \in \mathscr{X}^m$, it holds that $\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|t,x)$ and $\overline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) = \overline{E}^{\mathrm{ei}}_{\underline{Q}}(g|t,x)$.

Proposition 45. *Consider an imprecise Markov chain under epistemic irrelevance. Consider as well any $m \in \mathbb{N}_0$, any $s,t \in \mathscr{X}^m$, any $x \in \mathscr{X}$ and any extended real-valued function g on Ω that does not depend on the first m states, then*

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|t,x) \text{ and } \overline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) = \overline{E}^{\mathrm{ei}}_{\underline{Q}}(g|t,x).$$

Proof. We will only provide the proof for the global lower expectations; the proof for the global upper ones is completely analogous.

Consider any $\mathscr{M} \in \underline{\mathbb{M}}$ such that $\limsup \mathscr{M}(s,x\bullet) \leq g(s,x\bullet)$, and let $\mathscr{M}'$ be the real process defined by letting $\mathscr{M}'(t,x,u) := \mathscr{M}(s,x,u)$ for all $u \in \mathscr{X}^*$, and letting $\mathscr{M}'(v) = \mathscr{M}(s,x)$ for all $v \in \mathscr{X}^*$ such that $\Gamma(v) \not\subset \Gamma(t,x)$. $\mathscr{M}'$ is clearly a bounded above submartingale because $\mathscr{M}$ is, and moreover $\mathscr{M}'(t,x) = \mathscr{M}(s,x)$ and $\limsup \mathscr{M}'(t,x\bullet) = \limsup \mathscr{M}(s,x\bullet) \leq g(s,x\bullet) = g(t,x\bullet)$, whence, by Equation $(4.3)_{88}$

$$\begin{aligned}\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) &= \sup\{\mathscr{M}(s,x)\colon \mathscr{M} \in \underline{\mathbb{M}} \text{ and } \limsup \mathscr{M}(s,x\bullet) \leq g(s,x\bullet)\}\\ &\leq \sup\{\mathscr{M}'(t,x)\colon \mathscr{M}' \in \underline{\mathbb{M}} \text{ and } \limsup \mathscr{M}'(t,x\bullet) \leq g(t,x\bullet)\} = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|t,x)\end{aligned}$$

and since this holds for any $s,t \in \mathscr{X}^m$, we infer that $\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|t,x)$. □

Proposition 45 allows us to introduce a new notation for global lower and upper expectations of extended real-valued functions g on Ω that do not depend on the first m states $X_{1:m}$, with $m \in \mathbb{N}_0$:

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}|m+1}(g|x) := \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) \text{ and } \overline{E}^{\mathrm{ei}}_{\underline{Q}|m+1}(g|x) := \overline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) \text{ for all } s \in \mathscr{X}^m. \tag{5.51}$$

Obviously, $\underline{E}^{\mathrm{ei}}_{\underline{Q}|m+1}(g|\cdot)$ and $\overline{E}^{\mathrm{ei}}_{\underline{Q}|m+1}(g|\cdot)$ are extended real-valued functions on $\mathscr{X}$.

5.5 HOMOGENEOUS IMPRECISE MARKOV CHAINS UNDER EPISTEMIC IRRELEVANCE

Consider now an imprecise Markov chain under epistemic irrelevance where Equation $(5.30)_{115}$ is replaced by

$$\mathscr{P}_{x_{1:n}} = \mathscr{Q}_{x_n} \text{ for all } n \in \mathbb{N} \text{ and all } x_{1:n} \in \mathscr{X}^n \tag{5.52}$$

and where, consequently, Equation $(5.31)_{115}$ is replaced by

$$\underline{Q}(\cdot|x_{1:n}) = \underline{Q}(\cdot|x_n) \text{ and } \overline{Q}(\cdot|x_{1:n}) = \overline{Q}(\cdot|x_n) \text{ for all } n \in \mathbb{N} \text{ and all } x_{1:n} \in \mathscr{X}^n, \tag{5.53}$$

with $\underline{Q}(\cdot|x_n)$ and $\overline{Q}(\cdot|x_n)$ given by Equations $(5.28)_{114}$ and $(5.29)_{114}$.

This particular type of imprecise stochastic process is called a homogeneous imprecise Markov chain under epistemic irrelevance. An example of such an imprecise Markov chain is shown in Figure 5.4. Clearly, a homogeneous imprecise Markov chain under epistemic irrelevance is simply an imprecise Markov chain under epistemic irrelevance whose local models do not depend on time. Therefore, all the properties presented in Sections $5.4.2_{116}$—$5.4.5_{124}$ will still hold. Moreover, due to the homogeneity of the imprecise Markov chain, we will see in Sections 5.5.1 and $5.5.2_{129}$ that additional properties are satisfied by some types of global lower and upper expectations.

5.5.1 *Properties of global lower and upper expectations of functions that depend on a single state*

Similarly to Section $5.4.2_{116}$, we can compute global lower and upper expectations of functions that depend on a single state in homogeneous imprecise Markov chains under epistemic irrelevance by using lower and upper transition operators. This time the lower and upper transition operators do not depend on time. Specifically, we have the transition operators $\underline{T}$ and $\overline{T}$, which

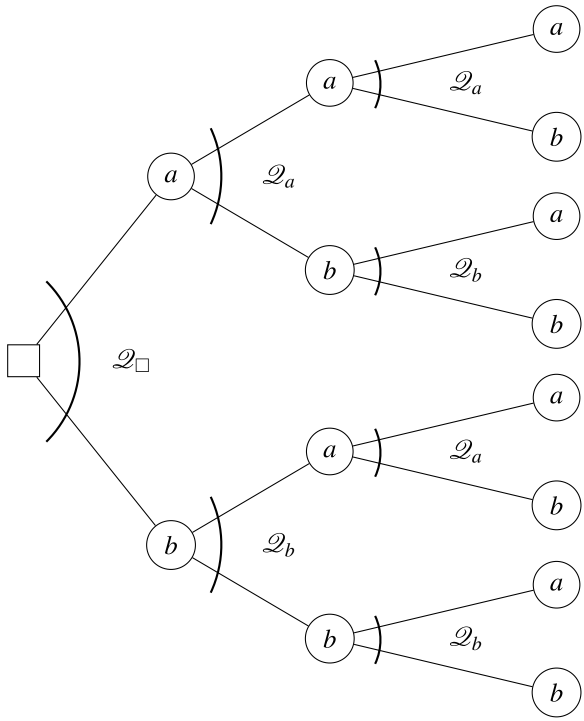

Figure 5.4: The initial part of a homogeneous imprecise Markov chain under epistemic irrelevance with state space $\mathscr{X} = \{a,b\}$.

are non-linear operators that are given by

$$\underline{T}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})\colon f \mapsto \underline{T}f;$$
$$\overline{T}\colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})\colon f \mapsto \overline{T}f,$$

where for all $f \in \mathscr{L}(\mathscr{X})$, $\underline{T}f$ and $\overline{T}f$ are functions in $\mathscr{L}(\mathscr{X})$ defined by

$$\underline{T}f(x) := \underline{Q}(f|x) = \inf\left\{\sum_{y\in\mathscr{X}} f(y)p(y)\colon p \in \mathcal{Q}_x\right\} \tag{5.54}$$

$$\overline{T}f(x) := \overline{Q}(f|x) = \sup\left\{\sum_{y\in\mathscr{X}} f(y)p(y)\colon p \in \mathcal{Q}_x\right\}, \tag{5.55}$$

for all $x \in \mathscr{X}$. Therefore, due to Lemma 40_{117}, for all $m,n \in \mathbb{N}$ such that $n > m$, all $x_{1:m} \in \mathscr{X}^m$ and all $f \in \mathscr{L}(\mathscr{X})$, the global lower and upper expectation of

f at time n conditional on $x_{1:m}$ are computed as follows:[12]

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_{1:m}) = \underline{T}^{n-m}f(x_m) \text{ and } \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_{1:m}) = \overline{T}^{n-m}f(x_m), \quad (5.56)$$

where we denote by $\underline{T}^{n-m}$ the $(n-m)$-fold composition of $\underline{T}$ itself, that is

$$\underline{T}^{n-m} := \overbrace{\underline{T} \times \cdots \times \underline{T}}^{n-m},$$

and similarly for $\overline{T}^{n-m}$. Analogously for the unconditional case, we find that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)) = \underline{Q}_{\square}(\underline{T}^{n-1}f) \text{ and } \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)) = \overline{Q}_{\square}(\overline{T}^{n-1}f). \quad (5.57)$$

As in the case of homogeneous Markov chains, also in homogeneous imprecise Markov chains under epistemic irrelevance the global lower and upper expectations of functions that depend on a single state satisfy various properties when taken to the limit. We first introduce the concept of ergodicity for homogeneous imprecise Markov chains under epistemic irrelevance [38].

Definition 9. *A homogeneous imprecise Markov chain under epistemic irrelevance with a lower transition operator $\underline{T}$ is called* ergodic *if for all $f \in \mathscr{L}(\mathscr{X})$, $\lim_{n\to+\infty} \underline{T}^n f$ exists and is a constant function.*

We now show that ergodicity implies the existence of a unique limit expectation operator $\underline{E}_{\infty}$. For any ergodic homogeneous imprecise Markov chain under epistemic irrelevance with a lower transition operator $\underline{T}$ and any $f \in \mathscr{L}(\mathscr{X})$, if we let $\underline{E}_{\infty}(f)$ be the constant value of $\lim_{n\to+\infty} \underline{T}^n f$, it follows from Equation (5.57) that $\lim_{n\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)) = \lim_{n\to+\infty} \underline{Q}_{\square}(\underline{T}^{n-1}f)$. Since for all $k \in \mathbb{N}$, it holds that $\min \underline{T}^{k-1}f \leq \underline{Q}_{\square}(\underline{T}^{k-1}f) \leq \max \underline{T}^{k-1}f$, and since the limit of the left- and right-hand side exist and are both equal to $\underline{E}_{\infty}(f)$ we find that

$$\lim_{n\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)) = \underline{E}_{\infty}(f).$$

Furthermore, the limit expectation operator $\underline{E}_{\infty}$ does not depend on the initial lower expectation operator $\underline{Q}_{\square}$ and it is the only such operator that is $\underline{T}$-*invariant*, in the sense that $\underline{E}_{\infty} = \underline{E}_{\infty} \circ \underline{T}$; see also Reference [28, Theorem 5.1].

A sufficient condition for a homogeneous imprecise Markov chain under epistemic irrelevance to be ergodic is to be *regularly absorbing* [28, Definition 4.1].[13] The definition of a regularly absorbing homogeneous imprecise Markov chain under epistemic irrelevance—presented in Reference [28]—is mainly expressed in terms of upper transition operators. Since we prefer working with lower transition operators and due to the conjugacy between lower and

[12] A similar result for homogeneous imprecise Markov chains under epistemic irrelevance, whose local models are credal sets, can be found in Reference [28].

[13] An extensive work on necessary and sufficient conditions can be found in Reference [38].

upper transition operators, i.e. $\overline{T}(f) = -\underline{T}(-f)$ for all $f \in \mathscr{L}(\mathscr{X})$, we will make use of the following equivalent definition.

Definition 10. *A homogeneous Markov chain under epistemic irrelevance is called* regularly absorbing *if the following holds:*

$$\mathscr{Y} := \left\{ x \in \mathscr{X} : (\exists m \in \mathbb{N})(\forall k \geq m)(\forall y \in \mathscr{X}) - \underline{T}^k(-\mathbb{I}_x)(y) > 0 \right\} \neq \emptyset,$$

and if moreover for all $y \in \mathscr{X} \setminus \mathscr{Y}$ there is some $m \in \mathbb{N}$ such that $\underline{T}^m \mathbb{I}_{\mathscr{Y}}(y) > 0$.

Note that if a homogeneous imprecise Markov chain under epistemic irrelevance is regularly absorbing, then it is also ergodic, but not vice versa.

Regarding the computation of $\underline{E}_\infty$, unfortunately, we cannot compute it easily or directly by solving the system $\underline{E}_\infty = \underline{E}_\infty \circ \underline{T}$ because this is a system of non-linear equations and it is not guaranteed that it can be solved. Luckily, we can compute the global lower expectation of any function $f \in \mathscr{L}(\mathscr{X})$ in the limit by computing the limit $\lim_{n \to +\infty} \underline{T}^n f$ explicitly.

5.5.2 *Properties of the global lower and upper expectations defined by the martingale-theoretic approach*

In the beginning of Section 5.5_{126}, we mention that a homogeneous imprecise Markov chain under epistemic irrelevance is a special case of an imprecise Markov chain under epistemic irrelevance. This implies that we can also derive our global lower and upper expectations by using the martingale-theoretic approach, where now the local models, i.e. the lower and upper expectation operators, satisfy Equation $(5.53)_{126}$. All the properties presented in Section $5.4.5_{124}$ will still hold and we also present additional properties that hold due to the homogeneity of the local models.

The first property is a Markov property for the global models. It states that all global conditional lower and upper expectations are completely determined by the global conditional expectations $\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\cdot|x)$ and $\overline{E}^{\mathrm{ei}}_{\underline{Q}}(\cdot|x)$, $x \in \mathscr{X}$.

Lemma 46. *Consider an imprecise Markov chain under epistemic irrelevance. Consider any extended real-valued function g on Ω, any situation $s \in \mathscr{X}^*$ and any $x \in \mathscr{X}$, then*

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g(s\bullet)|x) \text{ and } \overline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) = \overline{E}^{\mathrm{ei}}_{\underline{Q}}(g(s\bullet)|x).$$

A perhaps more familiar way of writing this is

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g(X_{1:\infty})|s,x) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g(s,X_{1:\infty})|x) \text{ and } \overline{E}^{\mathrm{ei}}_{\underline{Q}}(g(X_{1:\infty})|s,x) = \overline{E}^{\mathrm{ei}}_{\underline{Q}}(g(s,X_{1:\infty})|x).$$

Proof. We will only provide the proof for the global lower expectations; the proof for the global upper ones is completely analogous.

Consider, for ease of notation, the extended real-valued function g' on Ω such that $g' := g(s\bullet)$. Consider any bounded above submartingale $\mathscr{M}$ such that $\limsup \mathscr{M}(s,x\bullet) \leq g(s,x\bullet)$, and let $\mathscr{M}'$ be the real process that is defined by $\mathscr{M}'(u) := \mathscr{M}(s,u)$ for all $u \in \mathscr{X}^*$. $\mathscr{M}'$ is clearly a bounded above submartingale because $\mathscr{M}$ is, and also $\mathscr{M}'(x) = \mathscr{M}(s,x)$ and $\limsup \mathscr{M}'(x\bullet) = \limsup \mathscr{M}(s,x\bullet) \leq g(s,x\bullet) = g'(x\bullet)$, whence, by Equation (4.3)$_{88}$

$$\begin{aligned}\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) &= \sup\{\mathscr{M}(s,x)\colon \mathscr{M} \in \underline{\mathbb{M}} \text{ and } \limsup \mathscr{M}(s,x\bullet) \leq g(s,x\bullet)\}\\ &\leq \sup\{\mathscr{M}'(x)\colon \mathscr{M}' \in \underline{\mathbb{M}} \text{ and } \limsup \mathscr{M}'(x\bullet) \leq g'(x\bullet)\} = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g'|x).\end{aligned}$$

Conversely, consider now any bounded above submartingale $\mathscr{M}$ such that $\limsup \mathscr{M}(x\bullet) \leq g'(x\bullet)$, and let $\mathscr{M}'$ be the real process defined by letting $\mathscr{M}'(s,x,u) := \mathscr{M}(x,u)$ for all $u \in \mathscr{X}^*$, and letting $\mathscr{M}'(t) := \mathscr{M}(x)$ in all situations $t \in \mathscr{X}^*$ such that $\Gamma(t) \not\subseteq \Gamma(s,x)$. Then $\mathscr{M}'$ is clearly a bounded above submartingale because $\mathscr{M}$ is, and moreover we have that $\mathscr{M}'(s,x) = \mathscr{M}(x)$ and $\limsup \mathscr{M}'(s,x\bullet) = \limsup \mathscr{M}(x\bullet) \leq g'(x\bullet) = g(s,x\bullet)$, whence, again by Equation (4.3)$_{88}$

$$\begin{aligned}\underline{E}^{\mathrm{ei}}_{\underline{Q}}(g'|x) &= \sup\{\mathscr{M}(x)\colon \mathscr{M} \in \underline{\mathbb{M}} \text{ and } \limsup \mathscr{M}(x\bullet) \leq g'(x\bullet)\}\\ &\leq \sup\{\mathscr{M}'(s,x)\colon \mathscr{M}' \in \underline{\mathbb{M}} \text{ and } \limsup \mathscr{M}'(s,x\bullet) \leq g(s,x\bullet)\}\\ &= \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x).\end{aligned}$$

□

We now introduce the *shift operator* θ on $\mathbb{N}$ by letting $\theta(n) := n+1$ for all $n \in \mathbb{N}$. This induces a shift operator on Ω: $\theta\omega$ is the path with $(\theta\omega)_n := \omega_{\theta(n)} = \omega_{n+1}$ for all $n \in \mathbb{N}$. And this also induces a shift operation on functions g on Ω: θg is the function defined by $(\theta g)(\omega) := g(\theta\omega)$ for all $\omega \in \Omega$.

Proposition 47. *Let $m \in \mathbb{N}_0$. If a function g on Ω does not depend on the first m states, then θg does not depend on the first $m+1$ states.*

Proof. Assume g does not depend on the first m states, so there is some function g' such that $g(s\bullet) = g'$ for all $s \in \mathscr{X}^m$. Then for all $x \in \mathscr{X}$, $s \in \mathscr{X}^m$ and all $\omega \in \Omega$:

$$(\theta g)(x,s,\omega) = g(\theta(x,s,\omega)) = g(s,\omega) = g'(\omega),$$

which concludes the proof. □

Therefore, for any function g on *paths* that does not depend on the first m states, it holds that $g(\omega) = g(s,\theta^m\omega)$ for all $s \in \mathscr{X}^m$ and $\omega \in \Omega$, and we can also write it as $g = g(s\theta^m\bullet)$.

Our last result is a shift invariance property that is satisfied by the global lower and upper expectations.

Proposition 48. *Consider an imprecise Markov chain under epistemic irrelevance. Let $m \in \mathbb{N}$ and consider any extended real-valued function g on Ω that does not depend on the first $m-1$ states. Then for all $k \in \mathbb{N}_0$:*

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}|m}(g|\cdot) = \underline{E}^{\mathrm{ei}}_{\underline{Q}|m+k}(\theta^k g|\cdot) \text{ and } \overline{E}^{\mathrm{ei}}_{\underline{Q}|m}(g|\cdot) = \overline{E}^{\mathrm{ei}}_{\underline{Q}|m+k}(\theta^k g|\cdot).$$

Proof. We will only provide the proof for the global lower expectations; the proof for the global upper ones is completely analogous.

It clearly suffices to prove the statement for $k = 1$. Consider any $s \in \mathscr{X}^{m-1}$ and any $x, y \in \mathscr{X}$, then it follows from Equation $(5.51)_{126}$ and Lemma 46_{129} that

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}|m}(g|x) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g|s,x) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(g(s\,\bullet)|x)$$

and also that

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}|m+1}(\theta g|x) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(\theta g|y,s,x) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(\theta g(y,s\,\bullet)|x).$$

Now observe that $\theta g(y,s\,\bullet) = g(\theta(y,s\,\bullet)) = g(s\,\bullet)$.[14] □

Proposition 48 will turn out to be useful in Chapter 6_{151}, where we compute global lower and upper expected first-passage and return times in a special type of homogeneous imprecise Markov chains, the so-called imprecise time-homogeneous birth-death chains.

5.6 IMPRECISE MARKOV CHAINS UNDER COMPLETE INDEPENDENCE

In this section, we present a more stringent independence concept for imprecise Markov chains than epistemic irrelevance, called complete independence. As far as the computation of expectations is concerned, complete independence coincides with epistemic irrelevance when it comes to global lower and upper expectations of functions that depend on a single state and time averages. However, as far as we can tell, this might not be the case for more general functions.

5.6.1 *Complete independence*

Consider an imprecise Markov chain under epistemic irrelevance and let $\mathscr{T}_{\mathscr{Q}}$ be the the corresponding imprecise probability tree that we defined in Section $5.4.1_{115}$. Now let

$$\mathscr{T}^{\mathrm{M}}_{\mathscr{Q}} := \{p \in \mathbb{P}_{\mathrm{M}} \colon p \in \mathscr{T}_{\mathscr{Q}}\} = \mathscr{T}_{\mathscr{Q}} \cap \mathbb{P}_{\mathrm{M}}. \tag{5.58}$$

[14] Note that Proposition 48 might not hold for imprecise Markov chains under epistemic irrelevance that are not homogeneous. Since the local models at times n and $n-1$ conditional on x might differ, $\underline{E}^{\mathrm{ei}}_{\underline{Q}|n}(g|x)$ and $\underline{E}^{\mathrm{ei}}_{\underline{Q}|n+1}(\theta g|x)$ might differ as well.

This imprecise probability tree $\mathscr{T}^{\mathrm{M}}_{\mathscr{Q}}$ has a corresponding set of conditional probability measures on $\mathscr{C}_\sigma$, which we denote by $\mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}$. Clearly, $\mathscr{T}^{\mathrm{M}}_{\mathscr{Q}} \subseteq \mathscr{T}_{\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{M}}_{\mathscr{Q}} \subseteq \mathbb{P}_{\mathscr{Q}}$, where $\mathbb{P}_{\mathscr{Q}}$ is the corresponding set of conditional probability measures on $\mathscr{C}_\sigma$ of $\mathscr{T}_{\mathscr{Q}}$.

The imprecise stochastic process associated with the set $\mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}$ is called an imprecise Markov chain under *complete independence* [15, 62]. We infer from Equation (5.58)$_{\curvearrowleft}$ that the difference between epistemic irrelevance and complete independence is that under the former concept we consider probability trees that are derived from all possible combinations of the conditional probabilities of the local models, whereas under the latter we consider only those that correspond to Markov chains.

For any measurable extended real-valued function g on Ω and any $B \in \langle \mathscr{X}^* \rangle \setminus \{\emptyset\}$, the global lower and upper expectation of g conditional on B will be denoted by $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|B)$ and $\overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|B)$ respectively and they are defined as follows:

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|B) := \inf\left\{E_P(g|B) \colon P \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}\right\}; \tag{5.59}$$

$$\overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|B) := \sup\left\{E_P(g|B) \colon P \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}\right\}. \tag{5.60}$$

If the function g under study is a measurable extended real-valued one on Ω that does not depend on the first m states $X_{1:m}$, for some $m \in \mathbb{N}$, and is a limit of a non-decreasing sequence of non-negative n-measurable functions, then its global lower and upper expectation satisfy the following imprecise Markov property.

Theorem 49. *Consider an imprecise Markov chain under complete independence. Consider as well any $m \in \mathbb{N}_0$, any $x_{1:m} \in \mathscr{X}^m$, any $x_{m+1} \in \mathscr{X}$ and any measurable extended real-valued function g on Ω that does not depend on the first m states $X_{1:m}$ for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} h_n = g$. Then*

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|x_{1:m}, x_{m+1}) = \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|x_{m+1}) \text{ and } \overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|x_{1:m}, x_{m+1}) = \overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|x_{m+1}).$$

Proof. We will only provide the proof for the global lower expectations; the proof for the global upper ones is completely analogous. Observe that

$$\begin{aligned}\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|x_{1:m}, x_{m+1}) &= \inf\left\{E_P(g|x_{1:m+1}) \colon P \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}\right\} \\ &= \inf\left\{E_P(g|x_{m+1}) \colon P \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}\right\} = \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|x_{m+1}),\end{aligned}$$

where the first and the last equality follow from Equation (5.59) and the second equality follows from Theorem 35_{103} since every $P \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}$ is a conditional probability measure that corresponds to a Markov chain. □

We close this section with a technical lemma that will be used in the proofs of Theorems 51_{137} and 52_{138} further on.

Lemma 50. *Consider an initial model $\mathscr{Q}_\square$, and for each $n \in \mathbb{N}$ and each $x \in \mathscr{X}$, a set of conditional probability mass functions $\mathscr{Q}_{n,x}$ as introduced in Section 5.3_{113}. Consider as well any $m \in \mathbb{N}_0$, any $n \in \mathbb{N}$, any $x_{1:m} \in \mathscr{X}^m$ and any $f', f'' \in \mathscr{L}(\mathscr{X})$. Then*

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n) \middle| x_{1:m}\right) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n) \middle| x_{1:m}\right);$$
$$\overline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n) \middle| x_{1:m}\right) = \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n) \middle| x_{1:m}\right).$$

Proof. We will only provide the proof for the global lower expectations; the proof for the global upper ones is completely analogous.

If $m = 0$ and $n = 1$, we have that $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(f''(X_1)) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f''(X_1)) = \underline{Q}_\square(f'')$. If $m \neq 0$ and $n \leq m$, it follows from the argument following Equation $(3.18)_{74}$ that $E_P(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n)|x_{1:m}) = \sum_{i=1}^{n-1} f'(x_i) + f''(x_n)$ for all $P \in \mathbb{P}_{\mathscr{Q}}$. Therefore, and because $\mathbb{P}^{\mathrm{M}}_{\mathscr{Q}} \subseteq \mathbb{P}_{\mathscr{Q}}$, it follows from the definitions of $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}$ and $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}$ that

$$\begin{aligned}
&\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n) \middle| x_{1:m}\right) \\
&= \inf\left\{E_P\left(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n) \middle| x_{1:m}\right) : P \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}\right\} \\
&= \sum_{i=1}^{n-1} f'(x_i) + f''(x_n) = \inf\left\{E_P\left(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n) \middle| x_{1:m}\right) : P \in \mathbb{P}_{\mathscr{Q}}\right\} \\
&\qquad = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n) \middle| x_{1:m}\right).
\end{aligned}$$

We now prove the rest of the cases using induction. That is, for all $n \in \mathbb{N}$ such that $n > \max\{m, 1\}$ and all $f', f'' \in \mathscr{L}(\mathscr{X})$, we will prove that

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n) \middle| x_{1:m}\right) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{n-1} f'(X_i) + f''(X_n) \middle| x_{1:m}\right). \tag{5.61}$$

Consider any $k > \max\{m, 1\}$ and assume that Equation (5.61) holds for $n = k - 1$. We will prove that it holds for $n = k$ as well.

It follows from Theorem 21_{82} [repeatedly] that

$$\begin{aligned}
&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)\,\middle|\, x_{1:m}\right) \\
&\qquad = \underline{Q}\left(\dots \underline{Q}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)\,\middle|\, X_{1:k-1}\right)\dots \,\middle|\, x_{1:m}\right) \\
&\qquad = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\underline{Q}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)\,\middle|\, X_{1:k-1}\right)\,\middle|\, x_{1:m}\right).
\end{aligned} \tag{5.62}$$

Observe that for any $z_{1:k-1} \in \mathscr{X}^{k-1}$

$$\begin{aligned}
&\underline{Q}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)\,\middle|\, z_{1:k-1}\right) = \underline{Q}\left(\sum_{i=1}^{k-1} f'(z_i) + f''(X_k)\,\middle|\, z_{1:k-1}\right) \\
&\quad = \sum_{i=1}^{k-1} f'(z_i) + \underline{Q}(f''(X_k)|z_{1:k-1}) = \sum_{i=1}^{k-1} f'(z_i) + \underline{Q}_{k-1}(f''(X_k)|z_{k-1}) \\
&\quad = \sum_{i=1}^{k-1} f'(z_i) + \underline{T}_{k-1} f''(z_{k-1}),
\end{aligned} \tag{5.63}$$

where the first equality follows from Equation $(3.31)_{82}$; the second holds because, since $\sum_{i=1}^{k-1} f'(z_i)$ is a constant and $\underline{Q}(\cdot|z_{1:k-1})$ is an infimum of expectations, it follows from Lemmas 102_{240} and 106_{241} that $\sum_{i=1}^{k-1} f'(z_i)$ can be taken out; the third equality follows from Equation $(5.31)_{115}$; and the last equality follows from Equation $(5.32)_{117}$. Since Equation (5.63) holds for any $z_{1:k-1} \in \mathscr{X}^{k-1}$, it follows from Equation (5.62) that

$$\begin{aligned}
\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)\,\middle|\, x_{1:m}\right) &= \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + \underline{T}_{k-1} f''(X_{k-1})\,\middle|\, x_{1:m}\right) \\
&= \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-2} f'(X_i) + f'(X_{k-1}) + \underline{T}_{k-1} f''(X_{k-1})\,\middle|\, x_{1:m}\right).
\end{aligned}$$

Due to the induction hypothesis and the fact that $f'(X_{k-1}) + \underline{T}_{k-1} f''(X_{k-1})$ can be regarded as a function on $\mathscr{X}$, the aforementioned equation implies that

$$\begin{aligned}
&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)\,\middle|\, x_{1:m}\right) \\
&\qquad = \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-2} f'(X_i) + f'(X_{k-1}) + \underline{T} f''(X_{k-1})\,\middle|\, x_{1:m}\right) \\
&\qquad = \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + \underline{T} f''(X_{k-1})\,\middle|\, x_{1:m}\right).
\end{aligned} \tag{5.64}$$

We now prove Equation $(5.61)_{133}$ by showing first that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)|x_{1:m}) \geq \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)|x_{1:m})$ and then that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)|x_{1:m}) \leq \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k)|x_{1:m})$.

Fix any $\varepsilon > 0$. Then there is some $P_\varepsilon \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}$ such that

$$E_{P_\varepsilon}\left(\sum_{i=1}^{k-1} f'(X_i) + \underline{T} f''(X_{k-1}) \middle| x_{1:m}\right) \leq \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + \underline{T} f''(X_{k-1}) \middle| x_{1:m}\right) + \frac{\varepsilon}{2}, \tag{5.65}$$

and for any $x \in \mathscr{X}$, there is also some $p_x \in \mathscr{Q}_{k-1,x}$ such that

$$\sum_{y \in \mathscr{X}} f''(y) p_x(y) \leq \inf\left\{\sum_{y \in \mathscr{X}} f''(y) q(y) \colon q \in \mathscr{Q}_{k-1,x}\right\} + \frac{\varepsilon}{2} = \underline{T}_{k-1} f''(x) + \frac{\varepsilon}{2}, \tag{5.66}$$

where the equality follows from Equation $(5.32)_{117}$.

Let $p \in \mathscr{T}^{\mathrm{M}}_{\mathscr{Q}}$ be the probability tree that corresponds to P_ε. It follows from Equation $(5.58)_{131}$ that there is some $p^* \in \mathscr{T}^{\mathrm{M}}_{\mathscr{Q}}$ such that for all $i \in \{1, \ldots, k-2\}$ and all $z_{1:i} \in \mathscr{X}^i$, it holds that $p^*(X_{i+1}|z_{1:i}) = p(X_{i+1}|z_{1:i})$, and for all $z_{1:k-2} \in \mathscr{X}^{k-2}$ and all $x \in \mathscr{X}$, it holds that $p^*(X_k|z_{1:k-2}, x) = p_x(X_k)$. Consider now any $P^* \in \mathbb{P}_{p^*}$, then $P^* \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}$ and also

$$\begin{aligned} &E_{P^*}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right) \\ &= E_{P^*}\left(E_{P^*}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| X_{1:k-1}\right) \middle| x_{1:m}\right) \\ &= E_{P^*}\left(\sum_{i=1}^{k-1} f'(X_i) + E_{P^*}(f''(X_k)|X_{1:k-1}) \middle| x_{1:m}\right), \end{aligned} \tag{5.67}$$

where the first equality follows from Theorem 17_{75}, and the second equality from Lemmas 102_{240} and 106_{241} and the argument after Equation $(3.18)_{74}$. For all $z_{1:k-1} \in \mathscr{X}^{k-1}$, it follows from Equation $(3.18)_{74}$, the structure of p^* and Inequality (5.66) that

$$\begin{aligned} E_{P^*}(f''(X_k)|z_{1:k-1}) &= \sum_{y \in \mathscr{X}} f''(y) p^*(y|z_{1:k-1}) \\ &= \sum_{y \in \mathscr{X}} f''(y) p_{z_{k-1}}(y) \leq \underline{T}_{k-1} f''(z_{k-1}) + \frac{\varepsilon}{2}. \end{aligned}$$

Hence, by combining Equation (5.67) with Lemma 103_{240}, we further find that

$$\begin{aligned}
&E_{P^*}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right)\\
&\quad\leq E_{P^*}\left(\sum_{i=1}^{k-1} f'(X_i) + \underline{T}_{k-1}f''(X_{k-1}) + \frac{\varepsilon}{2} \middle| x_{1:m}\right)\\
&\quad= E_{P_\varepsilon}\left(\sum_{i=1}^{k-1} f'(X_i) + \underline{T}_{k-1}f''(X_{k-1}) + \frac{\varepsilon}{2} \middle| x_{1:m}\right)\\
&\quad= E_{P_\varepsilon}\left(\sum_{i=1}^{k-1} f'(X_i) + \underline{T}_{k-1}f''(X_{k-1}) \middle| x_{1:m}\right) + \frac{\varepsilon}{2}\\
&\quad\leq \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + \underline{T}_{k-1}f''(X_{k-1}) \middle| x_{1:m}\right) + \varepsilon, \qquad (5.68)
\end{aligned}$$

where the first equality follows from Equation $(3.18)_{74}$ and the structure of p^*, the last equality holds because of Lemmas 102_{240} and 106_{241} and the last inequality follows from Inequality $(5.65)_{\frown}$. Since Inequality (5.68) holds for all $\varepsilon > 0$, we infer that

$$E_{P^*}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right) \leq \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + \underline{T}_{k-1}f''(X_{k-1}) \middle| x_{1:m}\right),$$

and therefore, due to Equation $(5.64)_{134}$, also that

$$E_{P^*}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right)$$

and since $P^* \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}$, also that

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right).$$

Finally, observe that

$$\begin{aligned}
&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right)\\
&\quad= \inf\left\{E_P\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right) : P \in \mathbb{P}_{\mathscr{Q}}\right\}\\
&\quad\leq \inf\left\{E_P\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right) : P \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}\right\}\\
&\quad= \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} f'(X_i) + f''(X_k) \middle| x_{1:m}\right),
\end{aligned}$$

where the first and the last equality follow from the definition of global lower expectations under epistemic irrelevance and complete independence, respectively, and the inequality holds because $\mathbb{P}^{\mathrm{M}}_{\mathscr{Q}} \subseteq \mathbb{P}_{\mathscr{Q}}$. $\square$

5.6.2 *Global lower and upper expectations of functions that depend on a single state*

For global lower and upper expectations of functions that depend on a single state, it makes no difference whether we adopt epistemic irrelevance or complete independence. The corresponding theorem goes as follows.[15]

Theorem 51. *Consider an initial model $\mathscr{Q}_\square$ and for each $n \in \mathbb{N}$ and each $x \in \mathscr{X}$, a set of conditional probability mass functions $\mathscr{Q}_{n,x}$ as introduced in Section 5.3_{113}. Consider as well any $n, m \in \mathbb{N}$ such that $n > m$, any $x_m \in \mathscr{X}$ and any function $f \in \mathscr{L}(\mathscr{X})$. Then*

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(f(X_n)|x_m) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_m) \text{ and } \overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(f(X_n)|x_m) = \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_m).$$

For the unconditional case, for any $n \in \mathbb{N}$ and any $f \in \mathscr{L}(\mathscr{X})$, we have that

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(f(X_n)) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)) \text{ and } \overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(f(X_n)) = \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)).$$

Proof. We only provide the proof the global lower expectations; the proof for the global upper ones is completely analogous.

It follows from Lemma 50_{133} [for $f'' = f$, $f' = 0$ and $m = 0$] that

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(f(X_n)) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)).$$

Similarly, for any $x_{1:m-1} \in \mathscr{X}^{m-1}$, Lemma 50_{133} [again for $f'' = f$ and $f' = 0$] implies that

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(f(X_n)|x_{1:m}) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_{1:m})$$

and therefore, it follows from Theorems 49_{132} and 39_{117} that $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(f(X_n)|x_m) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_m)$. $\square$

5.6.3 *Global lower and upper expectations of time averages*

Similarly to the case of global lower and upper expectations of functions that depend on a single state, the global lower and upper expectations of time averages in imprecise Markov chains under complete independence coincide with the respective ones under epistemic irrelevance. The corresponding theorem goes as follows.

[15] A similar theorem for homogeneous imprecise Markov chains whose local models are credal sets was proved in Reference [5, Theorem 11.4].

Theorem 52. *Consider an initial model $\mathscr{Q}_\square$ and for each $n \in \mathbb{N}$ and each $x \in \mathscr{X}$, a set of conditional probability mass functions $\mathscr{Q}_{n,x}$ as introduced in Section 5.3*$_{113}$*. Consider as well any $m \in \mathbb{N}_0$, any $n \in \mathbb{N}$, any $x_{1:m} \in \mathscr{X}^m$ and any function $f \in \mathscr{L}(\mathscr{X})$. Then*

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m});$$
$$\overline{E}^{\mathrm{ci}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) = \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}).$$

Proof. Observe that

$$\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) = \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\middle| x_{1:m}\right)$$
$$= \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\middle| x_{1:m}\right) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}),$$

where the first and the last equality follow from Equation (5.7)$_{106}$ and the second equality follows from Lemma 50$_{133}$ [for $f'' = f' = \frac{1}{n}f$]. □

5.6.4 *Global lower and upper expectations of more general functions*

For general n-measurable functions or extended real-valued functions that are limits of non-decreasing sequences of non-negative n-measurable functions, the global lower and upper expectations for imprecise Markov chains under complete independence may not coincide with the respective ones under epistemic irrelevance. This is clarified through the following example.

Example 8. Consider the set $\mathscr{X} = \{a, b\}$, the interval $I = [1/4, 3/4]$ and the following sets of probability mass functions on $\mathscr{X}$:

$$\mathscr{Q}_\square := \{(1/2, 1/2)\},\ \mathscr{Q}_{2,b} := \{(0,1)\} \text{ and}$$
$$\mathscr{Q}_{2,a} = \mathscr{Q}_{1,a} = \mathscr{Q}_{1,b} := \{(q, 1-q) \colon q \in I\}. \quad (5.69)$$

Consider also any imprecise Markov chain under epistemic irrelevance and any imprecise Markov chain under complete independence, whose state space is $\mathscr{X}$ and local models at times $n = 0, 1, 2$ are given by Equation (5.69). Consider as well the function h in $\mathscr{L}(\mathscr{X}^3)$ with values $h(a,\cdot,a) = h(b,\cdot,b) = 0$ and $h(a,\cdot,b) = h(b,\cdot,a) = 1$.

We will calculate the lower expectations $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(h(X_{1:3}))$ and $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(h(X_{1:3}))$. Starting with $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(h(X_{1:3}))$, it follows from Theorem 21$_{82}$ that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(h(X_{1:3})) = \underline{Q}_\square(\underline{Q}(\underline{Q}(h(X_{1:3})|X_{1:2})|X_1)) = \underline{Q}_\square(h''(X_1)), \quad (5.70)$$

with $h''(X_1) := \underline{Q}(h'(X_{1:2})|X_1)$ and $h'(X_{1:2}) := \underline{Q}(h(X_{1:3})|X_{1:2})$.

For $h'(a,a)$, it follows from Equation $(3.31)_{82}$ that

$$h'(a,a) = \underline{Q}(h(X_{1:3})|a,a) = \underline{Q}(h(a,a,X_3)|a,a)$$

and, due to Equation $(5.31)_{115}$, that $\underline{Q}(h(a,a,X_3)|a,a) = \underline{Q}_2(h(a,a,X_3)|a)$. Therefore, it follows from Equation $(5.26)_{114}$ that

$$\begin{aligned} h'(a,a) = \underline{Q}_2(h(a,a,X_3)|a) &= \inf\left\{ \sum_{y\in\mathscr{X}} h(a,a,y)p(y) \colon p \in \mathscr{Q}_{2,a} \right\} \\ &= \inf\left\{ h(a,a,a)p(a) + h(a,a,b)p(b) \colon p \in \mathscr{Q}_{2,a} \right\} \\ &= \inf\left\{ h(a,a,a)q + h(a,a,b)(1-q) \colon q \in I \right\} \\ &= \inf\left\{ 1-q \colon q \in I \right\} = 1 - \frac{3}{4} = \frac{1}{4}, \end{aligned}$$

where the fourth equality follows from Equality $(5.69)_{\curvearrowleft}$ and the sixth holds because I is a closed interval consisting of strictly positive values and therefore the infimum $\inf\{1-q \colon q \in I\}$ is obtained for the largest value in I.

Since, due to Equality $(5.69)_{\curvearrowleft}$, $\mathscr{Q}_{2,b}$ consists of a single probability mass function, it is easy to see that $h'(a,b) = \underline{Q}_2(h(a,b,X_3)|b) = 1$.

Similarly, we find that $h'(a,b) = \underline{Q}_2(h(b,a,X_3)|a) = 1/4$ and $h'(b,b) = \underline{Q}_2(h(b,b,X_3)|b) = 0$.

We then calculate $h''(X_1)$. For $X_1 = a$, it follows from Equations $(3.31)_{82}$ and $(5.31)_{115}$ that

$$h''(a) = \underline{Q}(h'(X_{1:2})|a) = \underline{Q}(h'(a,X_2)|a) = \underline{Q}_1(h'(a,X_2)|a).$$

Since we know that $h'(a,a) = 1/4$ and $h'(a,b) = 1$, it follows from Equation $(5.26)_{114}$ that

$$\begin{aligned} h''(a) = \underline{Q}_1(h'(a,X_2)|a) &= \inf\left\{ \sum_{y\in\mathscr{X}} h'(a,y)p(y) \colon p \in \mathscr{Q}_{1,a} \right\} \\ &= \inf\left\{ h'(a,a)p(a) + h'(a,b)p(b) \colon p \in \mathscr{Q}_{1,a} \right\} \\ &= \inf\left\{ h'(a,a)q + h'(a,b)(1-q) \colon q \in I \right\} \\ &= \inf\left\{ \frac{1}{4}q + 1 - q \colon q \in I \right\} = \inf\left\{ 1 - \frac{3}{4}q \colon q \in I \right\} = 1 - \frac{3}{4}\cdot\frac{3}{4} = \frac{7}{16}, \end{aligned}$$

where the fourth equality follows from Equality $(5.69)_{\curvearrowleft}$ and the seventh holds because I is a closed interval consisting of strictly positive values and therefore the infimum $\inf\{1-\frac{3}{4}q \colon q \in I\}$ is obtained for the largest value in I. In a completely similar way, we find that $h''(b) = 1/16$. Finally, we have that

$$\begin{aligned} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(h(X_{1:3})) &= \underline{Q}_{\square}(h''(X_1)) = \inf\left\{ \sum_{y\in\mathscr{X}} h''(y)p(y) \colon p \in \mathscr{Q}_{\square} \right\} \\ &= h''(a)\frac{1}{2} + h''(b)\frac{1}{2} = \frac{7}{16}\cdot\frac{1}{2} + \frac{1}{16}\cdot\frac{1}{2} = \frac{1}{2}\left(\frac{7}{16}+\frac{1}{16}\right) = \frac{1}{4} \end{aligned} \tag{5.71}$$

where the first equality follows from Equation (5.70)$_{138}$, the second equality follows from the definition of $\underline{Q}_{\square}(\cdot)$ and the third equality follows from Equality (5.69)$_{138}$.

We now calculate $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(h(X_{1:3}))$, for which we have that

$$
\begin{aligned}
\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(h(X_{1:3})) &= \inf\left\{E_P(h(X_{1:3}))\colon P \in \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}\right\}\\
&= \inf\left\{\sum_{x_{1:3}\in\mathscr{X}^3} h(x_{1:3})\prod_{i=0}^{2} p(x_{i+1}|x_{1:i})\colon p \in \mathscr{T}^{\mathrm{M}}_{\mathscr{Q}}\right\}\\
&= \inf\Big\{h(a,a,a)p(a|a,a)p(a|a)p(a|\square) + h(a,a,b)p(b|a,a)p(a|a)p(a|\square)\\
&\qquad + h(a,b,a)p(a|a,b)p(b|a)p(a|\square) + h(a,b,b)p(b|a,b)p(b|a)p(a|\square)\\
&\qquad + h(b,a,a)p(a|b,a)p(a|b)p(b|\square) + h(b,a,b)p(b|b,a)p(a|b)p(b|\square)\\
&\qquad + h(b,b,a)p(a|b,b)p(b|b)p(b|\square) + h(b,b,b)p(b|b,b)p(b|b)p(b|\square)\\
&\qquad \colon p \in \mathscr{T}^{\mathrm{M}}_{\mathscr{Q}}\Big\}\\
&= \inf\Big\{q_2(b|a)q_1(a|a)q_{\square}(a) + q_2(b|b)q_1(b|a)q_{\square}(a)\\
&\qquad + q_2(a|a)q_1(a|b)q_{\square}(b) + q_2(a|b)q_1(b|b)q_{\square}(b)\colon q_2(\cdot|a) \in \mathscr{Q}_{2,a},\\
&\qquad\quad q_2(\cdot|b) \in \mathscr{Q}_{2,b},\ q_1(\cdot|a) \in \mathscr{Q}_{1,a},\ q_1(\cdot|b) \in \mathscr{Q}_{1,b} \text{ and } q_{\square} \in \mathscr{Q}_{\square}\Big\}\\
&= \inf\left\{(1-q'')q'\frac{1}{2} + (1-q')\frac{1}{2} + 2(1-q')\frac{1}{2} + q''q^*\frac{1}{2}\colon q^*, q', q'' \in I\right\}\\
&= \inf\left\{q''q^*\frac{1}{2} - q''q'\frac{1}{2} + \frac{1}{2}\colon q^*, q', q'' \in I\right\}\\
&= \inf\left\{q''\frac{1}{2}(q^* - q') + \frac{1}{2}\colon q^*, q', q'' \in I\right\} = \frac{3}{4}\cdot\frac{1}{2}\left(\frac{1}{4}-\frac{3}{4}\right) + \frac{1}{2}\\
&= -\frac{3}{4}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2} = -\frac{3}{16} + \frac{1}{2} = \frac{5}{16}
\end{aligned}
\tag{5.72}
$$

where the first equality follows from Equation (5.59)$_{132}$, the second follows from Equation (3.18)$_{74}$, the fourth equality follows from Equation (5.58)$_{131}$, the fifth follows from Equation (5.69)$_{138}$ and the eighth equality holds because the infimum is obtained for the smallest possible value of $q''(q*-q')$ and therefore, it is easy to see that this value is obtained for $q'' = q' = 3/4$ and $q^* = 1/4$.

Finally, due to Equations (5.71)$_{\curvearrowleft}$ and (5.72), we infer that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(h(X_{1:3})) < \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(h(X_{1:3}))$. ◊

Although there are also cases in which epistemic irrelevance and complete independence coincide, such as the lower and upper expected first-passage times of Chapter 6$_{151}$, this example illustrates that this is not true in general. Therefore, we will require new computational methods for computing global lower and upper expectations that correspond to imprecise Markov chains under complete independence.

However, unfortunately, there are no known efficient methods for computing global lower and upper expectations of general n-measurable functions, let alone of more general measurable functions, in imprecise Markov chains under complete independence. For this reason, we now propose a brute-force approach for approximating global lower and upper expectations of general n-measurable functions for small $n \in \mathbb{N}$.[16] We describe the approach for conditional expectations; the unconditional ones follow similarly.

Consider any $n, m \in \mathbb{N}$ such that $n > m$ and any situation $x_{1:m} \in \mathscr{X}^m$. Consider as well any n-measurable function $h(X_{1:n})$ and suppose that we want to calculate $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(h(X_{1:n})|x_{1:m})$ and $\overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(h(X_{1:n})|x_{1:m})$. We first select $k \in \mathbb{N}$ conditional probability mass functions from each local model $\mathscr{Q}_{i,x}$, for all $i \in \{m, \dots, n-1\}$ and all $x \in \mathscr{X}$—for $i = m$, we only need to consider k conditional probability mass functions from $\mathscr{Q}_{m,x_m}$. We then construct the imprecise probability tree by considering all possible combinations of the selected conditional probability mass functions. For each probability tree in the constructed imprecise probability, we compute the global conditional expectation of $h(X_{1:n})$, among which the lowest value is the approximation for $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(h(X_{1:n})|x_{1:m})$ and the highest value for $\overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(h(X_{1:n})|x_{1:m})$. In the unconditional case, we follow the same procedure and additionally, we select k probability mass functions from the initial model $\mathscr{Q}_{\square}$.

This approach is quite inefficient in general. And, not only does it not guarantee that we will achieve the global infimum or supremum, but it is also computationally expensive and we have no idea about the error. For any $n, m \in \mathbb{N}_0$ such that $n > m$, we need to calculate $k^{(n-m-1)|\mathscr{X}|+1}$ global expectations.

5.6.5 *Homogeneous imprecise Markov chains under complete independence*

For homogeneous imprecise Markov chains under complete independence, everything mentioned in Sections 5.6.1$_{131}$–5.6.4$_{138}$ holds in the exact same way and the only difference now is that the local models do not depend on time and hence, they satisfy Equation (5.52)$_{126}$. Therefore, we can compute global lower and upper expectations of functions that depend on a single state and time averages using lower and upper transition operators $\underline{T}$ and $\overline{T}$. In the case of general n-measurable functions, for small $n \in \mathbb{N}$, we can again use the approach described in Section 5.6.4$_{138}$.

[16] When n is large, it might not be feasible to compute expectations of general n-measurable functions even under epistemic irrelevance. By applying Theorem 21$_{82}$ to any n-measurable function $h(X_{1:n})$, we need to solve $\sum_{i=i}^{n-1} |\mathscr{X}|^i + 1$ optimisation problems in order to find $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(h(X_{1:n}))$.

5.7 IMPRECISE MARKOV CHAINS UNDER REPETITION INDEPENDENCE

We now present an even more stringent independence concept than epistemic irrelevance and complete independence, which can be used in homogeneous imprecise Markov chains, and which is called repetition independence. But, as we will see, under repetition independence, we may not be able to efficiently compute global lower and upper expectations of various types of functions, even for those that depend only on a single state.

5.7.1 *Repetition independence*

Consider a homogeneous imprecise Markov chain under epistemic irrelevance and let $\mathscr{T}_{\mathscr{Q}}$ be the corresponding imprecise probability tree as defined in Section $5.4.1_{115}$. We now impose an additional constraint that is stronger that the one imposed on (homogeneous) imprecise Markov chains under complete independence. Instead of the imprecise probability $\mathscr{T}_{\mathscr{Q}}$ or the imprecise probability $\mathscr{T}_{\mathscr{Q}}^{\mathrm{M}}$ given by Equation $(5.58)_{131}$, we consider the set

$$\mathscr{T}_{\mathscr{Q}}^{\mathrm{HM}} := \left\{p \in \mathbb{P}_{\mathrm{HM}} \colon p \in \mathscr{T}_{\mathscr{Q}}\right\} = \mathscr{T}_{\mathscr{Q}} \cap \mathbb{P}_{\mathrm{HM}}. \tag{5.73}$$

This imprecise probability tree $\mathscr{T}_{\mathscr{Q}}^{\mathrm{HM}}$ has a corresponding set of conditional probability measures on $\mathscr{C}_{\sigma}$, which we denote by $\mathbb{P}_{\mathscr{Q}}^{\mathrm{HM}}$. It is easy to see that $\mathscr{T}_{\mathscr{Q}}^{\mathrm{HM}} \subseteq \mathscr{T}_{\mathscr{Q}}^{\mathrm{M}} \subseteq \mathscr{T}_{\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{Q}}^{\mathrm{HM}} \subseteq \mathbb{P}_{\mathscr{Q}}^{\mathrm{M}} \subseteq \mathbb{P}_{\mathscr{Q}}$.

The set $\mathbb{P}_{\mathscr{Q}}^{\mathrm{M}}$ corresponds to an imprecise stochastic process that is called a homogeneous imprecise Markov chain under *repetition independence* [14]. Judging by Equation (5.73), we see that under repetition independence, from all combinations of the probability mass functions of the local models we consider only the probability trees that correspond to homogeneous imprecise Markov chains, whereas in the case of complete independence we consider those that correspond to—not necessarily homogeneous—Markov chains and in the case of epistemic irrelevance we consider all possible combinations.

For any measurable extended real-valued function g on Ω and any $B \in \langle\mathscr{X}^*\rangle \setminus \{\varnothing\}$, the global lower and upper expectation of g conditional on B will be denoted by $\underline{E}_{\mathscr{Q}}^{\mathrm{ri}}(g|B)$ and $\overline{E}_{\mathscr{Q}}^{\mathrm{ri}}(g|B)$, respectively, and are defined as follows:

$$\underline{E}_{\mathscr{Q}}^{\mathrm{ri}}(g|B) := \inf\left\{E_P(g|B) \colon P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{HM}}\right\}; \tag{5.74}$$

$$\overline{E}_{\mathscr{Q}}^{\mathrm{ri}}(g|B) := \sup\left\{E_P(g|B) \colon P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{HM}}\right\}. \tag{5.75}$$

Global lower and upper expectations of measurable extended real-valued functions on Ω that do not depend on the first m states $X_{1:m}$, for some $m \in \mathbb{N}$, and that are limits of some non-decreasing sequence of non-negative n-measurable functions satisfy the following imprecise Markov property.

Theorem 53. *Consider a homogeneous imprecise Markov chain under repetition independence. Consider as well any* $m \in \mathbb{N}_0$, *any* $x_{1:m} \in \mathscr{X}^m$, *any* $x_{m+1} \in \mathscr{X}$ *and any measurable extended real-valued function* g *on* Ω *that does not depend on the first* m *states* $X_{1:m}$ *and for which there is a non-decreasing sequence of non-negative* n*-measurable functions* $\{h_n\}_{n\in\mathbb{N}}$ *such that* $\lim_{n\to+\infty} h_n = g$. *Then*

$$\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(g|x_{1:m},x_{m+1}) = \underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(g|x_{m+1}) \text{ and } \overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(g|x_{1:m},x_{m+1}) = \overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(g|x_{m+1}).$$

Proof. We will only provide the proof for the global lower expectations; the proof for the global upper ones is completely analogous.

Observe that

$$\begin{aligned}\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(g|x_{1:m},x_{m+1}) &= \inf\left\{E_P(g|x_{1:m},x_{m+1}) \colon P \in \mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}}\right\} \\ &= \inf\left\{E_P(g|x_{m+1}) \colon P \in \mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}}\right\} = \underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(g|x_{m+1}),\end{aligned}$$

where the first and the last equality follow from Equation (5.74)$_{\curvearrowleft}$ and the second equality follows from Theorem 35$_{103}$ since every $P \in \mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}}$ is a conditional probability measure that corresponds to a homogeneous Markov chain. □

5.7.2 *Global lower and upper expectations*

In general, there are no known efficient methods for computing global lower and upper expectations in homogeneous Markov chains under repetition independence. Furthermore, the results may differ from those obtained under epistemic irrelevance and complete independence. In fact, even for the computation of global lower and upper expectations of functions that depend on a single state, the following example illustrates that we can no longer use lower and upper transition operators.

Example 9. Consider the set $\mathscr{X} = \{a,b\}$, the interval $I = [1/4, 3/4]$ and the following sets of probability mass functions on $\mathscr{X}$:

$$\mathscr{Q}_\square := \{(1/2,1/2)\},\ \mathscr{Q}_a := \{(q,1-q) \colon q \in I\} \text{ and } \mathscr{Q}_b := \{(1,0)\}. \quad (5.76)$$

Consider also any homogeneous imprecise Markov chain under epistemic irrelevance and any homogeneous imprecise Markov under repetition independence, whose state space is $\mathscr{X}$ and local models at times $n = 0,1,2$ are given by Equation (5.76). Consider as well the function f in $\mathscr{L}(\mathscr{X})$ defined by $f(a) := 0$ and $f(b) := 1$.

We will calculate the lower expectations $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_3))$ and $\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(f(X_3))$. Starting with $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_3))$—which is equal to $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(f(X_3))$ due to Theorem 51$_{137}$—it follows from Equation (5.57)$_{128}$ that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_3)) = \underline{Q}_\square(\underline{T}^2 f).$$

For $\underline{T}f(a)$, it follows from Equation $(5.54)_{127}$ that

$$\begin{aligned}\underline{T}f(a) &= \inf\left\{\sum_{y\in\mathscr{X}} f(y)p(y)\colon p\in\mathscr{Q}_a\right\} = \inf\big\{f(a)p(a)+f(b)p(b)\colon p\in\mathscr{Q}_a\big\}\\ &= \inf\big\{f(a)q+f(b)(1-q)\colon q\in I\big\} = \inf\big\{1-q\colon q\in I\big\}\\ &= \inf\big\{1-q\colon q\in I\big\} = 1-\frac{3}{4} = \frac{1}{4},\end{aligned}$$

where the third equality follows from Equation $(5.76)_{\curvearrowleft}$ and the sixth holds because I is a closed interval consisting of strictly positive values and therefore the infimum $\inf\{1-q\colon q\in I\}$ is obtained for the largest value in I.

Since, due to Equation $(5.76)_{\curvearrowleft}$, $\mathscr{Q}_b$ consists of a single probability mass function, it is easy to see that $\underline{T}f(b)=0$.

Moving on, we calculate $\underline{T}^2f(a)$ and we find that

$$\begin{aligned}\underline{T}^2f(a) &= \inf\left\{\sum_{y\in\mathscr{X}} \underline{T}f(y)p(y)\colon p\in\mathscr{Q}_a\right\}\\ &= \inf\big\{\underline{T}f(a)p(a)+\underline{T}f(b)p(b)\colon p\in\mathscr{Q}_a\big\}\\ &= \inf\big\{\underline{T}f(a)q+\underline{T}f(b)(1-q)\colon q\in I\big\} = \inf\left\{\frac{1}{4}q\colon q\in I\right\} = \frac{1}{4}\cdot\frac{1}{4} = \frac{1}{16}\end{aligned}$$

where the first equality follows from Equation $(5.54)_{127}$ and the fact that $\underline{T}f$ is a function on $\mathscr{X}$, the third equality follows from Equation $(5.76)_{\curvearrowleft}$ and the fifth holds because I is a closed interval consisting of strictly positive values and therefore the infimum $\inf\{\frac{1}{4}q\colon q\in I\}$ is obtained for the smallest value in I. Similarly, we find that $\underline{T}^2f(b)=1/4$.

Finally, it follows from Equation $(5.57)_{128}$ that

$$\begin{aligned}\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_3)) &= \underline{Q}_{\square}(\underline{T}^2f) = \inf\left\{\sum_{y\in\mathscr{X}} \underline{T}^2f(y)p(y)\colon p\in\mathscr{Q}_{\square}\right\}\\ &= \inf\big\{\underline{T}^2f(a)p(a)+\underline{T}^2f(b)p(b)\colon p\in\mathscr{Q}_{\square}\big\} = \frac{1}{16}\cdot\frac{1}{2}+\frac{1}{4}\cdot\frac{1}{2}\\ &= \frac{1}{2}(\frac{1}{16}+\frac{1}{4}) = \frac{1}{2}\cdot\frac{5}{16} = \frac{5}{32}.\end{aligned} \tag{5.77}$$

We now calculate $\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(f(X_3))$, for which we have that

$$\begin{aligned}\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(f(X_3)) &= \inf\left\{E_P(f(X_3))\colon P\in\mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}}\right\}\\ &= \inf\left\{\sum_{x_{1:3}\in\mathscr{X}^3} f(x_3)\prod_{i=0}^{2} p(x_{i+1}|x_{1:i})\colon p\in\mathscr{T}^{\mathrm{HM}}_{\mathscr{Q}}\right\}\end{aligned}$$

$$
\begin{aligned}
&= \inf\Big\{ f(a)p(a|a,a)p(a|a)p(a|\Box) + f(b)p(b|a,a)p(a|a)p(a|\Box) \\
&\quad + f(a)p(a|a,b)p(b|a)p(a|\Box) + f(b)p(b|a,b)p(b|a)p(a|\Box) \\
&\quad + f(a)p(a|b,a)p(a|b)p(b|\Box) + f(b)p(b|b,a)p(a|b)p(b|\Box) \\
&\quad + f(a)p(a|b,b)p(b|b)p(b|\Box) + f(b)p(b|b,b)p(b|b)p(b|\Box) \\
&\quad : p \in \mathscr{T}_{\mathscr{Q}}^{\mathrm{HM}} \Big\} \\
&= \inf\Big\{ q(b|a)q(a|a)q_\Box(a) + q(b|b)q(b|a)q_\Box(a) \\
&\quad + q(b|a)q(a|b)q_\Box(b) + q(b|b)q(b|b)q_\Box(b) \\
&\quad : q(\cdot|a) \in \mathscr{Q}_a,\ q(\cdot|b) \in \mathscr{Q}_b \text{ and } q_\Box \in \mathscr{Q}_\Box \Big\} \\
&= \inf\Big\{ \big(1-q(a|a)\big)q(a|a)\frac{1}{2} + \big(1-q(a|a)\big)\frac{1}{2} : q(\cdot|a) \in \mathscr{Q}_a \Big\} \\
&= \inf\Big\{ \big(1-q'\big)q'\frac{1}{2} + \big(1-q'\big)\frac{1}{2} : q' \in I \Big\} = \inf\{-q'^2\frac{1}{2} + \frac{1}{2} : q' \in I\} \\
&= -\Big(\frac{3}{4}\Big)^2 \cdot \frac{1}{2} + \frac{1}{2} = -\frac{9}{16} \cdot \frac{1}{2} + \frac{1}{2} = -\frac{9}{32} + \frac{1}{2} = \frac{7}{32},
\end{aligned}
\tag{5.78}
$$

where the first equality follows from Equation (5.59)$_{132}$, the second follows from Equation (3.18)$_{74}$, the fourth equality follows from Equation (5.73)$_{142}$ and the definition of f, the fifth and sixth equality follow from Equation (5.76)$_{143}$ and the eighth equality holds because I is a closed interval consisting of strictly positive values and therefore the infimum $\inf\{-q'^2\frac{1}{2} + \frac{1}{2} : q' \in I\}$ is obtained for the largest value in I.

Finally, due to Equations (5.77)$_{\curvearrowleft}$ and (5.78), we infer that $\underline{E}_{\mathscr{Q}}^{\mathrm{ei}}(f(X_3)) < \underline{E}_{\mathscr{Q}}^{\mathrm{ri}}(f(X_3))$. ◊

Nevertheless, there are some specific cases where epistemic irrelevance, complete independence and repetition independence lead to the same result. The following proposition, which will turn out to be useful in Chapter 7$_{189}$ when we compare these three independence concepts with a fourth one, provides a first example.

Proposition 54. *Consider an initial model $\mathscr{Q}_\Box$ and for each $x \in \mathscr{X}$, a set of conditional probability mass functions $\mathscr{Q}_x$, as introduced in Section 5.3$_{113}$. Consider as well any function $f \in \mathscr{L}(\mathscr{X})$ and any $k \in \{1,2\}$. Then*

$$
\underline{E}_{\mathscr{Q}}^{\mathrm{ei}}(f(X_k)) = \underline{E}_{\mathscr{Q}}^{\mathrm{ci}}(f(X_k)) = \underline{E}_{\mathscr{Q}}^{\mathrm{ri}}(f(X_k));
$$
$$
\overline{E}_{\mathscr{Q}}^{\mathrm{ei}}(f(X_k)) = \overline{E}_{\mathscr{Q}}^{\mathrm{ci}}(f(X_k)) = \overline{E}_{\mathscr{Q}}^{\mathrm{ri}}(f(X_k)).
$$

Proof. We will only provide the proof for the global lower expectations; the proof for the global upper ones is completely analogous.

The fact that $\underline{E}_{\mathscr{Q}}^{\mathrm{ei}}(f(X_k)) = \underline{E}_{\mathscr{Q}}^{\mathrm{ci}}(f(X_k))$ follows directly from Theorem 51$_{137}$. Therefore, we will prove that $\underline{E}_{\mathscr{Q}}^{\mathrm{ei}}(f(X_k)) = \underline{E}_{\mathscr{Q}}^{\mathrm{ri}}(f(X_k))$.

For $k = 1$, we find that

$$
\begin{aligned}
\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_1)) &= \inf\Big\{E_P(f(X_1))\colon P \in \mathbb{P}_{\mathscr{Q}}\Big\} = \inf\Big\{\sum_{x\in\mathscr{X}} p(x)f(x)\colon p \in \mathscr{T}_{\mathscr{Q}}\Big\} \\
&= \inf\Big\{\sum_{x\in\mathscr{X}} q_{\square}(x)f(x)\colon q_{\square} \in \mathscr{Q}_{\square}\Big\} = \inf\Big\{\sum_{x\in\mathscr{X}} p(x)f(x)\colon p \in \mathscr{T}^{\mathrm{HM}}_{\mathscr{Q}}\Big\} \\
&= \inf\Big\{E_P(f(X_1))\colon P \in \mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}}\Big\} = \underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(f(X_1)),
\end{aligned}
$$

and similarly for $k = 2$, we have that

$$
\begin{aligned}
\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_2)) &= \inf\Big\{E_P(f(X_2))\colon P \in \mathbb{P}_{\mathscr{Q}}\Big\} \\
&= \inf\Big\{\sum_{x_{1:2}\in\mathscr{X}^2} f(x_2)p(x_2|x_1)p(x_1|\square)\colon p \in \mathscr{T}_{\mathscr{Q}}\Big\} \\
&= \inf\Big\{\sum_{x_{1:2}\in\mathscr{X}^2} f(x_2)q(x_2|x_1)q_{\square}(x_1)\colon \\
&\qquad\qquad (\forall x_1 \in \mathscr{X})\, q(\cdot|x_1) \in \mathscr{Q}_{x_1}, q_{\square} \in \mathscr{Q}_{\square}\Big\} \\
&= \inf\Big\{\sum_{x_{1:2}\in\mathscr{X}^2} f(x_2)p(x_2|x_1)p(x_1|\square)\colon p \in \mathscr{T}^{\mathrm{HM}}_{\mathscr{Q}}\Big\} \\
&= \inf\Big\{E_P(f(X_2))\colon P \in \mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}}\Big\} = \underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(f(X_2)),
\end{aligned}
$$

where the first equality follows from the definition of $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}$, the second and fifth equality follow from Equation $(3.18)_{74}$, the third equality follows from Equation $(3.28)_{80}$ [the way probability trees are constructed under epistemic irrelevance, where $\mathscr{P}_{\square} = \mathscr{Q}_{\square}$ and $\mathscr{P}_{x_{1:n}} = \mathscr{Q}_{x_n}$ for all $x_{1:n} \in \mathscr{X}^* \setminus \{\square\}$], the fourth equality follows from Equations $(3.28)_{80}$ and $(5.73)_{142}$ [the way probability trees are constructed under repetition independence] and the last equality follows from Equation $(5.74)_{142}$. □

Another special case where epistemic irrelevance, complete independence and repetition independence coincide are the lower and upper expected first-passage and return times in imprecise birth-death chains that we will study in Chapter 6.

However, in general, and as was illustrated by Example 9, repetition independence does not coincide with epistemic irrelevance or complete independence. In those cases, new computational methods are required. However, unfortunately, exact efficient methods are not available. Nevertheless, it is possible to approximate global lower and upper expectations of functions that depend on a single state and n-measurable ones, for small $n \in \mathbb{N}$. We consider an approach that is similar to the one for global lower and upper expectations in imprecise Markov chains under complete independence that was introduced in Section $5.6.4_{138}$. Here too, this approach does not guarantee that a global infimum or supremum is achieved, nor do we know anything about the error, and it can generally be quite inefficient.

Consider any $n, m \in \mathbb{N}$ such that $n > m$ and any situation $x_{1:m} \in \mathscr{X}^m$. Consider as well any n-measurable function $h(X_{1:n})$ and suppose that we want to calculate $\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(h(X_{1:n})|x_{1:m})$ and $\overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(h(X_{1:n})|x_{1:m})$. We first select $k \in \mathbb{N}$ conditional probability mass functions from each local model $\mathscr{Q}_x$ for all $x \in \mathscr{X}$. We then construct the imprecise probability tree by considering all possible combinations of the selected conditional probability mass functions—there are $k^{|\mathscr{X}|}$ in total, except for the case $n = m+1$ where there are k. For each probability tree in the constructed imprecise probability tree, we compute the global conditional expectation of $h(X_{1:n})$, among which the lowest value is the approximation for $\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(h(X_{1:n})|x_{1:m})$ and the highest value is the approximation for $\overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(h(X_{1:n})|x_{1:m})$. The same method can be used also for the approximation of $\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(f(X_n)|x_{1:m})$ and $\overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(f(X_n)|x_{1:m})$ for any $f \in \mathscr{L}(\mathscr{X})$. In the unconditional case, we can follow a similar procedure if we additionally select k probability mass functions from the initial model $\mathscr{Q}_\square$—there are now $k^{|\mathscr{X}|+1}$ combinations.

5.8 DISCUSSION

We have seen that we can adopt various types of independence when working with imprecise Markov chains. In this section, we briefly discuss the effects that the different independence concepts have on the global lower and upper expectations.

Our first remark is that the more constraints we impose on the global models, the more precise our models become, in the sense that the distance between lower and upper expectation decreases. The reason why this happens is because the more stringent the independence concept is, the less conditional probability measures we consider for deriving our global models. For imprecise Markov chains, the global models under epistemic irrelevance use the set $\mathbb{P}_{\mathscr{Q}}$, whereas the respective ones under complete independence use the set $\mathbb{P}^{\mathrm{M}}_{\mathscr{Q}}$, for which it holds that $\mathbb{P}^{\mathrm{M}}_{\mathscr{Q}} \subseteq \mathbb{P}_{\mathscr{Q}}$. Similarly, for the homogeneous case, where we also consider global models under repetition independence that use the set $\mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}}$, we have that $\mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}} \subseteq \mathbb{P}^{\mathrm{M}}_{\mathscr{Q}} \subseteq \mathbb{P}_{\mathscr{Q}}$. These observations are formalised in the following lemmas; the proofs are obvious and are therefore omitted.

Lemma 55. *Consider an initial model $\mathscr{Q}_\square$ and for each $n \in \mathbb{N}$ and each $x \in \mathscr{X}$, a set of conditional probability mass functions $\mathscr{Q}_{n,x}$ as introduced in Section 5.3*$_{113}$*. Consider also any $B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\}$ and any measurable extended real-valued function g on Ω for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} h_n = g$. Then*

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|B) \leq \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|B) \leq \overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|B) \leq \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|B).$$

Lemma 56. *Consider an initial model $\mathscr{Q}_\square$ and for each $x \in \mathscr{X}$, a set of conditional probability mass functions $\mathscr{Q}_x$, as introduced in Section 5.3*$_{113}$*. Consider as well any $B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\}$ and any measurable extended real-valued function g on Ω for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} h_n = g$. It then holds that*

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|B) \leq \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|B) \leq \underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(g|B) \leq \overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(g|B) \leq \overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(g|B) \leq \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(g|B).$$

Finally, we present an interesting property of lower and upper expectations of functions that depend on a single state and their respective time averages, when these are taken over a time window whose width becomes infinitely large.

Lemma 57. *Consider a homogeneous imprecise Markov chain under epistemic irrelevance. Consider as well any $m \in \mathbb{N}_0$, any $x_{1:m} \in \mathscr{X}^m$ and any function $f \in \mathscr{L}(\mathscr{X})$. Then*

$$\begin{aligned}\liminf_{n\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_{1:m}) &\leq \liminf_{n\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) \\ &\leq \limsup_{n\to+\infty} \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) \leq \limsup_{n\to+\infty} \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_{1:m}).\end{aligned}$$

Proof. The middle inequality follows directly from the definitions of lim inf and lim sup. For the rest of the inequalities, we will only provide the proof for the global lower expectations; the proof for the global upper ones is completely analogous.

For any $n \in \mathbb{N}$, it follows from the definition of $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\cdot|x_{1:m})$ and Equation $(5.7)_{106}$ that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) = \inf\left\{E_P\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\middle| x_{1:m}\right) : P \in \mathbb{P}_{\mathscr{Q}}\right\},$$

and by combining this with Lemma 102_{240}, we find that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) = \inf\left\{\frac{1}{n}\sum_{i=1}^{n} E_P(f(X_i)|x_{1:m}) \colon P \in \mathbb{P}_{\mathscr{Q}}\right\}.$$

Since coefficients can be taken out of the infimum, we further infer that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) = \frac{1}{n}\inf\left\{\sum_{i=1}^{n} E_P(f(X_i)|x_{1:m}) \colon P \in \mathbb{P}_{\mathscr{Q}}\right\}. \tag{5.79}$$

Combining Equation (5.79) with the fact that an infimum of a sum is greater than or equal to the sum of infima, we have that

$$\begin{aligned}\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) &\geq \frac{1}{n}\sum_{i=1}^{n} \inf\left\{E_P(f(X_i)|x_{1:m}) \colon P \in \mathbb{P}_{\mathscr{Q}}\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m}),\end{aligned}$$

whence

$$\liminf_{n\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}([f](X_{1:n})|x_{1:m}) \geq \liminf_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m}). \tag{5.80}$$

Since f is real-valued and for all $n \in \mathbb{N}$, $\min f \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_{1:m}) \leq \max f$, we infer that for all $n \in \mathbb{N}$, $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_{1:m})$ is real-valued and therefore, that $\liminf_{n\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|x_{1:m})$ is also real-valued. Fix now any $k \in \mathbb{N}$. Then for all $n > k$, we have that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m}) &= \frac{1}{n}\sum_{i=1}^{k} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m}) + \frac{1}{n}\sum_{i=k+1}^{n} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m}) \\
&\geq \frac{1}{n}\sum_{i=1}^{k} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m}) + \frac{n-k}{n} \inf_{\ell\geq k+1} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_\ell)|x_{1:m})
\end{aligned}$$

and since this hold for all $n > k$, we further find that

$$\begin{aligned}
&\liminf_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m}) \\
&\quad\geq \liminf_{n\to+\infty} \left(\frac{1}{n}\sum_{i=1}^{k} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m}) + \frac{n-k}{n} \inf_{\ell\geq k+1} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_\ell)|x_{1:m})\right) \\
&\quad\geq \liminf_{n\to+\infty} \left(\frac{1}{n}\sum_{i=1}^{k} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m})\right) + \liminf_{n\to+\infty} \left(\frac{n-k}{n} \inf_{\ell\geq k+1} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_\ell)|x_{1:m})\right),
\end{aligned} \tag{5.81}$$

where the second inequality follows from the fact that an infimum of a sum is greater or equal to the sum of infima. Since $\sum_{i=1}^{k} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m})$ is a constant, we have that

$$\liminf_{n\to+\infty} \left(\frac{1}{n}\sum_{i=1}^{k} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m})\right) = 0. \tag{5.82}$$

Since $\lim_{n\to+\infty} \frac{n-k}{n} = 1$ and $\inf_{\ell\geq k+1} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_\ell)|x_{1:m})$ does not depend on n, we infer that

$$\liminf_{n\to+\infty} \left(\frac{n-k}{n} \inf_{\ell\geq k+1} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_\ell)|x_{1:m})\right) = \inf_{\ell\geq k+1} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_\ell)|x_{1:m}). \tag{5.83}$$

Combining Equations (5.82) and (5.83) with Inequality (5.81), we find that

$$\liminf_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m}) \geq \inf_{\ell\geq k+1} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_\ell)|x_{1:m})$$

and since this holds for any $k \in \mathbb{N}$, we have that

$$\begin{aligned}
\liminf_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_i)|x_{1:m}) &\geq \lim_{k\to+\infty} \inf_{\ell\geq k+1} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_\ell)|x_{1:m}) \\
&= \liminf_{k\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_k)|x_{1:m}),
\end{aligned} \tag{5.84}$$

where the equality follows from the definition of the limit inferior. Finally, the result follows by combining Equation (5.84)↶ with Equation (5.80)↶. □

Although Lemma 57_{148} is formulated and proved for homogeneous imprecise Markov chains epistemic irrelevance, it should be clear that the proof does not depend on the independence concept chosen and therefore, Lemma 57_{148} holds also for complete and repetition independence. For complete independence, since the global lower and upper expectations of functions that depend on a single state and time averages coincide with the respective ones under epistemic irrelevance, they coincide when they are taken to the limit as well. Finally, as we will see in Section 7.6_{206}, under epistemic irrelevance and complete independence the inequalities in Lemma 57_{148} are typically strict, whereas for repetition independence, they tend to be equalities.

6

IMPRECISE DISCRETE-TIME BIRTH-DEATH CHAINS

In this chapter, we discuss a special class of homogeneous imprecise Markov chains, which we call imprecise birth-death chains. We first give a brief description of a precise birth-death chain and then we introduce our version of an imprecise birth-death chain by allowing our local models to be sets of probability mass functions and adopting—at least in the beginning—the independence concept of epistemic irrelevance.

We focus on—upward and downward—first-passage and return times,[1] which are expressed by an extended real-valued function on the set of all paths Ω and we provide a method for computing their global lower and upper expected values. For precise birth-death chains, expected first-passage times have been studied in Reference [71]. For imprecise birth-death chains, we here first define global lower and upper expected first-passage and return times according to the martingale-theoretic approach. Under mild assumptions, we then prove that any such global lower or upper expected first-passage and return time is real-valued. Next, we show that global lower and upper expected first-passage and return times satisfy a system of non-linear equations, and we develop a simple recursive method for solving it.

Along the way, we also prove some useful properties that are satisfied by our global lower and upper expected first-passage and return times. In particular, when the local models of an imprecise birth-death chain are closed sets of probability mass functions, any lower or upper expected first-passage or return time can always be obtained by considering a specific precise birth-death chain.

[1] Also known as recurrence times.

All our methods can also be used in the measure-theoretic approach, because, as we will see, any global lower or upper expected first-passage or return time defined according to the measure-theoretic approach coincides with the corresponding one defined according to the martingale-theoretic approach. Similarly, for the measure-theoretic approach, we also prove that the choice of the independence concept does not affect the global lower and upper expected first-passage and return times.

6.1 BIRTH-DEATH CHAINS

Birth-death chains [81, Section 9.4] are a special type of homogeneous Markov chains, where transitions from a given state are possible only to that same state or to adjacent ones. They constitute a class of models that are characterised by a specific structure and allow our computational methods to be simplified in certain cases. They are used in various scientific fields, including evolutionary biology [2, Chapter 3] and queueing theory [34].

Since birth-death chains are homogeneous Markov chains, they are completely determined by an initial model and by their transition models, the latter of which can be summarised by a transition matrix—see Equation $(5.2.1)_{110}$—that is in addition tridiagonal due to the fact that transitions from a given state are possible only to that same state or to adjacent ones. Hence, the transition matrix of a birth-death chain with a finite ordered state space $\mathscr{X} = \{0,\dots,L\}$, where $L \in \mathbb{N}$, has the following form:

$$M = \begin{pmatrix} r_0 & w_0 & 0 & \cdots & \cdots & 0 \\ b_1 & r_1 & w_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & b_{L-1} & r_{L-1} & w_{L-1} \\ 0 & \cdots & \cdots & 0 & b_L & r_L \end{pmatrix}, \tag{6.1}$$

where the elements of each row sum to 1. For all $i \in \mathscr{X} \setminus \{0,L\}$, we will assume that the probabilities w_i, b_i and r_i are strictly positive, and similarly for r_0, w_0, b_L, r_L. Such a birth-death chain not only has a a tree-like representation—see Figure 5.2_{110}—it also has a chain-like representation, as depicted in Figure $6.1_{\curvearrowright}$.

6.2 IMPRECISE BIRTH-DEATH CHAINS

The concept of a birth-death chain can be made imprecise by letting its local models be sets of conditional probability mass functions. This yields a special class of homogeneous imprecise Markov chains. In the rest of this section, we explain how they are constructed and described.

For every $i \in \mathscr{X} \setminus \{0,L\}$, consider any set of conditional probability mass functions $\mathscr{R}_i$ on $\mathscr{X}_m := \{\ell, e, u\}$, where m stands for middle and ℓ, e and u

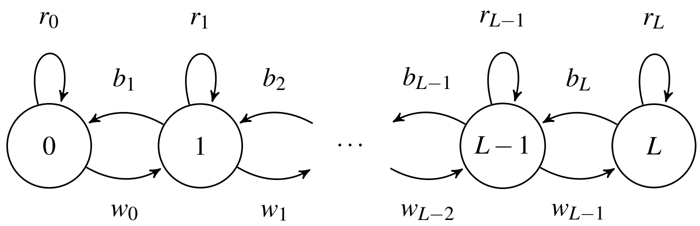

Figure 6.1: A birth-death chain with state space $\mathscr{X} = \{0, \ldots, L\}$

stand for lower, equal and upper, respectively, and where for any $\pi_i \in \mathscr{R}_i$, we make use of the notational convention that $(b_i, r_i, w_i) = (\pi_i(\ell), \pi_i(e), \pi_i(u))$.[2] Consider as well any set of conditional probability mass functions $\mathscr{R}_0$ on $\mathscr{X}_0 := \{e, u\}$ and any set of conditional probability mass functions $\mathscr{R}_L$ on $\mathscr{X}_L := \{\ell, e\}$. For any $\pi_0 \in \mathscr{R}_0$ and $\pi_L \in \mathscr{R}_L$, we also make use of the notational conventions that $(r_0, w_0) = (\pi_0(e), \pi_0(u))$ and $(b_L, r_L) = (\pi_L(\ell), \pi_L(e))$. Finally, we also adopt notational conventions for the lower and upper probability masses that correspond to the sets $\mathscr{R}_i$ for all $i \in \mathscr{X}$. Similarly to Equation (2.2)$_{40}$, we define the following:

$$(\forall i \in \mathscr{X})\ \underline{r}_i := \inf\{r_i \colon \pi_i \in \mathscr{R}_i\} \text{ and } \overline{r}_i := \sup\{r_i \colon \pi_i \in \mathscr{R}_i\}; \tag{6.2}$$

$$(\forall i \in \mathscr{X} \setminus \{0\})\ \underline{b}_i := \inf\{b_i \colon \pi_i \in \mathscr{R}_i\} \text{ and } \overline{b}_i := \sup\{b_i \colon \pi_i \in \mathscr{R}_i\}; \tag{6.3}$$

$$(\forall i \in \mathscr{X} \setminus \{L\})\ \underline{w}_i := \inf\{w_i \colon \pi_i \in \mathscr{R}_i\} \text{ and } \overline{w}_i := \sup\{w_i \colon \pi_i \in \mathscr{R}_i\}. \tag{6.4}$$

For reasons of mathematical convenience, we will restrict ourselves to sets $\mathscr{R}_i$ that satisfy the following assumption.

Assumption 6.1. *For every $i \in \mathscr{X}$, $\mathscr{R}_i$ is closed and consists of strictly positive probability mass functions.*

It follows from Assumption 6.1 that the infima and the suprema in Equations (6.2)—(6.4) are actually minima and maxima. Moreover, Assumption 6.1 implies—amongst other useful consequences such as Theorem 58$_{156}$ further on—that $0 < \underline{w}_i \leq \overline{w}_i < 1$ for all $i \in \mathscr{X} \setminus \{L\}$ and that $0 < \underline{b}_i \leq \overline{b}_i < 1$ for all $i \in \mathscr{X} \setminus \{0\}$.

We now use the sets of conditional probability mass functions $\mathscr{R}_i$ to define corresponding sets of conditional probability mass functions $\mathscr{Q}_i$ on $\mathscr{X}$. For all $i \in \mathscr{X} \setminus \{0, L\}$, a conditional probability mass function $q(\cdot|i) \in \Sigma_{\mathscr{X}}$ belongs to

[2] In Propositions 67$_{166}$, 64$_{164}$ and 59$_{162}$, we consider sets $\mathscr{R}$ of conditional probability mass functions on $\mathscr{X}_m$ that are not necessarily associated with a state value $i \in \mathscr{X} \setminus \{0, L\}$. There too, similarly, for each $\pi \in \mathscr{R}$, we make use of the notational convention that $(b, r, w) = (\pi(\ell), \pi(e), \pi(u))$.

$\mathscr{Q}_i$ if and only if there is some $\pi_i \in \mathscr{R}_i$ such that

$$q(j|i) = \begin{cases} b_i & \text{if } j = i-1 \\ r_i & \text{if } j = i \\ w_i & \text{if } j = i+1 \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } j \in \mathscr{X}. \tag{6.5}$$

Similarly, $q(\cdot|0)$ belongs to $\mathscr{Q}_0$ if and only if there is some $\pi_0 \in \mathscr{R}_0$ such that

$$q(j|0) = \begin{cases} r_0 & \text{if } j = 0 \\ w_0 & \text{if } j = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } j \in \mathscr{X} \tag{6.6}$$

and $q(j|L)$ belongs to $\mathscr{Q}_L$ if and only if there is some $\pi_L \in \mathscr{R}_L$ such that

$$q(j|L) = \begin{cases} b_L & \text{if } j = L-1 \\ r_L & \text{if } j = L \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } j \in \mathscr{X}. \tag{6.7}$$

For all $i \in \mathscr{X}$, the set of conditional probability mass functions $\mathscr{Q}_i$ has a corresponding lower and an upper expectation operator, which are denoted by $\underline{Q}(\cdot|i)$ and $\overline{Q}(\cdot|i)$ and, for all $f \in \mathscr{L}(\mathscr{X})$, are defined as

$$\underline{Q}(f|i) := \min\left\{ \sum_{x\in\mathscr{X}} f(x)p_i(x) \colon p_i \in \mathscr{R}_i \right\} \tag{6.8}$$

$$\overline{Q}(f|i) := \max\left\{ \sum_{x\in\mathscr{X}} f(x)p_i(x) \colon p_i \in \mathscr{R}_i \right\}. \tag{6.9}$$

The sets of conditional probability mass functions that we have just introduced, and their corresponding lower and upper expectation operators, can now be used to construct a special type of imprecise Markov chain, which we call an imprecise birth-death chain.[3] For the initial situation we have an arbitrary set of probability mass functions $\mathscr{Q}_\square$ on $\mathscr{X}$ with corresponding lower and upper expectation operators $\underline{Q}_\square$ and $\overline{Q}_\square$. Moreover, for all $f \in \mathscr{L}(\mathscr{X})$ and all $i \in \mathscr{X}$, the lower and upper transition operators $\underline{T}$ and $\overline{T}$ are now defined by $\underline{T}f(i) := \underline{Q}(f|i)$ and $\overline{T}f(i) := \overline{Q}(f|i)$, where $\underline{Q}(f|i)$ and $\overline{Q}(f|i)$ are given by Equations (6.8) and (6.9) respectively.

Since an imprecise birth-death chain is a special case of a homogeneous imprecise Markov chain, we can again choose among different types of independence for building our global models. In Sections 6.3$_{\curvearrowright}$—6.7$_{168}$, our imprecise birth-death chains adopt epistemic irrelevance and we use the martingale-theoretic approach to study lower and upper expected first-passage and return

[3] Similar models have already been considered in Reference [17].

times. In Section 6.8_{177}, we will also consider lower and upper expected first-passage and return times based on the measure-theoretic approach under different independence concepts and we will prove that they coincide with the corresponding ones based on the martingale-theoretic approach regardless of the independence concept chosen.

6.3 FIRST-PASSAGE AND RETURN TIMES

We now introduce the variable that represents first-passage (or return) times and we define its lower and upper expectation in an imprecise birth-death chain under epistemic irrelevance, using the martingale-theoretic approach.

Consider a time $n \in \mathbb{N}$, two—possibly identical—states i and j in $\mathscr{X}$ and any $x_{1:n-1} \in \mathscr{X}^{n-1}$. Suppose that the imprecise birth-death chain starts out at time n in the situation $(x_{1:n-1}, i)$, then we ask ourselves how long it will take to reach the state value j, or if $i = j$, for the imprecise birth-death chain to return to the state value i. To study this, we introduce the extended real-valued function $\tau^n_{i\to j}$ given by:[4]

$$\tau^n_{i\to j}(\omega) := \begin{cases} 0 & \text{if } \omega_n \neq i \\ \inf\{m \in \mathbb{N} \colon \omega_{n+m} = j\} & \text{if } \omega_n = i. \end{cases} \tag{6.10}$$

We call this number of time-steps the *first-passage time* of j conditional on $X_{1:n} = (x_{1:n-1}, i)$, and when $i = j$, we call it the *return time* of i. The so-called upward and downward first-passage times correspond to the cases $i < j$ and $i > j$, respectively. The goal is to compute the lower and upper expectations $\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^n_{i\to j}|x_{1:n-1}, i)$ and $\overline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^n_{i\to j}|x_{1:n-1}, i)$ of these first-passage and return times.

In order to do that, we first observe that $\theta\tau^n_{i\to j} = \tau^{n+1}_{i\to j}$. Consider now the lower expectation $\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^n_{i\to j}|x_{1:n-1}, i)$ for any $x_{1:n-1} \in \mathscr{X}^{n-1}$. Then, since $\tau^n_{i\to j}$ clearly does not depend on the first $n-1$ states $X_{1:n-1}$, we infer from Proposition 45_{125} and Equation $(5.51)_{126}$ that

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^n_{i\to j}|x_{1:n-1}, i) = \underline{E}^{\mathrm{ei}}_{\underline{Q}|n}(\tau^n_{i\to j}|i). \tag{6.11}$$

Moreover, we infer from Proposition 48_{131} that

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}|n+1}(\tau^{n+1}_{i\to j}|i) = \underline{E}^{\mathrm{ei}}_{\underline{Q}|n+1}(\theta\tau^n_{i\to j}|i) = \underline{E}^{\mathrm{ei}}_{\underline{Q}|n}(\tau^n_{i\to j}|i),$$

so we conclude that $\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^n_{i\to j}|x_{1:n-1}, i)$ neither depends on the initial segment $x_{1:n-1}$, nor on its length $n-1$. A similar conclusion holds for $\overline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^n_{i\to j}|x_{1:n-1}, i)$. In order to reflect these findings in our notation, we will from now on denote

[4] The reason why this function is extended real-valued is because $\tau^n_{i\to j}(\omega) = +\infty$ if $\omega_{n+m} \neq j$ for all $m \in \mathbb{N}$.

the lower and upper expected first-passage time from i to j by $\underline{\tau}_{i\to j}$ and $\overline{\tau}_{i\to j}$, respectively; they are defined by

$$\underline{\tau}_{i\to j} := \underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^{1}_{i\to j}|i) = \underline{E}^{\mathrm{ei}}_{\underline{Q}|n}(\tau^{n}_{i\to j}|i) \tag{6.12}$$

and

$$\overline{\tau}_{i\to j} := \overline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^{1}_{i\to j}|i) = \overline{E}^{\mathrm{ei}}_{\underline{Q}|n}(\tau^{n}_{i\to j}|i). \tag{6.13}$$

The following theorem establishes a first convenient property of $\underline{\tau}_{i\to j}$ and $\overline{\tau}_{i\to j}$, which follows from Assumption 6.1_{153}.

Theorem 58. *For all $i,j \in \mathscr{X}$, the lower and upper expected first-passage times $\underline{\tau}_{i\to j}$ and $\overline{\tau}_{i\to j}$ are real-valued and strictly positive.*

Proof. Since it follows from Equation $(6.10)_{\curvearrowleft}$ that $\inf\{\tau^{1}_{i\to j}(\omega)\colon \omega\in\Gamma(i)\} \geq 1$ and from Equations (6.12) and (6.13) that

$$\underline{\tau}_{i\to j} = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^{1}_{i\to j}|i) \text{ and } \overline{\tau}_{i\to j} = \overline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^{1}_{i\to j}|i), \tag{6.14}$$

property $\mathrm{G5}_{92}$ implies that $1 \leq \underline{\tau}_{i\to j} \leq \overline{\tau}_{i\to j}$, and therefore, the only thing that we still need to prove is that $\overline{\tau}_{i\to j} < +\infty$. We will do this by showing that there is some bounded above supermartingale $\mathscr{M}_0$ in $\overline{\mathbb{M}}$ such that

$$\liminf \mathscr{M}_0(\omega) \geq \tau^{1}_{i\to j}(\omega) \text{ for all } \omega\in\Gamma(i). \tag{6.15}$$

Indeed, since it follows from Equations (6.14) and $(4.4)_{88}$ that

$$\overline{\tau}_{i\to j} = \overline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^{1}_{i\to j}|i) := \inf\{\mathscr{M}(i)\colon \mathscr{M}\in\overline{\mathbb{M}} \text{ and } \liminf\mathscr{M}(\omega) \geq \tau^{1}_{i\to j}(\omega) \text{ for all } \omega\in\Gamma(i)\},$$

Equation (6.15) will then imply that $\overline{\tau}_{i\to j} \leq \mathscr{M}_0(i) < +\infty$.

Consider the values $\varepsilon^{u}_{0} := 1/\underline{w}_0$ and $\varepsilon^{d}_{L} := 1/\underline{b}_L$. Using ε^{u}_{0} and ε^{d}_{L}, we now define recursively, for all $x\in\mathscr{X}\setminus\{0,L\}$:

$$\varepsilon^{u}_{x} := \frac{1}{\underline{w}_x} + \frac{\overline{b}_x}{\underline{w}_x}\varepsilon^{u}_{x-1} \text{ and } \varepsilon^{d}_{x} := \frac{1}{\underline{b}_x} + \frac{\overline{w}_x}{\underline{b}_x}\varepsilon^{d}_{x+1}. \tag{6.16}$$

Due to Assumption 6.1_{153}, we have that ε^{u}_{0} and ε^{d}_{L}, as well as ε^{u}_{x} and ε^{d}_{x}, for all $x\in\mathscr{X}\setminus\{0,L\}$, are strictly positive and real-valued. Now let $\Delta_j\in\mathscr{L}(\mathscr{X})$ be defined by

$$\Delta_j(i') := \begin{cases} 0 & \text{if } i' = j \\ \sum_{\ell=i'}^{j-1}\varepsilon^{u}_{\ell} & \text{if } i' < j \\ \sum_{\ell=j+1}^{i'}\varepsilon^{d}_{\ell} & \text{if } i' > j \end{cases} \quad \text{for all } i'\in\mathscr{X}, \tag{6.17}$$

and consider the real process $\mathscr{M}_0$, defined for all $m \in \mathbb{N}_0$ and $x_{1:m} \in \mathscr{X}^m$ by

$$\mathscr{M}_0(x_{1:m}) := \begin{cases} 1+\overline{Q}(\Delta_j|i) & \text{if } m \in \{0,1\} \text{ or } x_1 \neq i; \\ m-1+\Delta_j(x_m) & \text{if } x_1 = i,\, m \geq 2 \text{ and} \\ & \quad (\forall k \in \{2,\dots,m-1\})\, x_k \neq j; \\ \mathscr{M}_0(x_{1:m-1}) & \text{if } x_1 = i,\, m \geq 2 \text{ and} \\ & \quad (\exists k \in \{2,\dots,m-1\})\, x_k = j. \end{cases} \tag{6.18}$$

In the remainder of this proof, we show that Inequality (6.15)$_{\curvearrowleft}$ holds, by proving that $\mathscr{M}_0 \in \overline{\mathbb{M}}$ and that $\liminf \mathscr{M}_0(\omega) \geq \tau^1_{i\to j}(\omega)$ for all $\omega \in \Gamma(i)$.

We start by proving that $\liminf \mathscr{M}_0(\omega) \geq \tau^1_{i\to j}(\omega)$ for all $\omega \in \Gamma(i)$. We consider two cases: $\tau^1_{i\to j}(\omega) < +\infty$ and $\tau^1_{i\to j}(\omega) = +\infty$. If $\tau^1_{i\to j}(\omega) < +\infty$, then with $m := \tau^1_{i\to j}(\omega)+1$, Equation (6.10)$_{155}$ implies that

$$\omega_m = j \text{ and } (\forall k \in \{2,\dots,m-1\})\omega_k \neq j,$$

and therefore, because of Equations (6.18) and (6.17)$_{\curvearrowleft}$, for all $n \geq m$, it follows that

$$\mathscr{M}_0(\omega^n) = \mathscr{M}_0(\omega^m) = m-1+\Delta_j(j) = \tau^1_{i\to j}(\omega),$$

which implies that $\liminf_{n\to\infty} \mathscr{M}_0(\omega^n) = \tau^1_{i\to j}(\omega)$. If $\tau^1_{i\to j}(\omega) = +\infty$, Equation (6.10)$_{155}$ implies that $\omega_k \neq j$ for all $k \geq 2$, and therefore, it follows from Equation (6.18) that

$$\liminf_{n\to\infty} \mathscr{M}_0(\omega^n) = \liminf_{n\to\infty}(n-1+\Delta_j(\omega_n)) \geq \liminf_{n\to\infty}(n-1) = +\infty = \tau^1_{i\to j}(\omega),$$

where the inequality holds because, due to Equation (6.17)$_{\curvearrowleft}$, we have that

$$\Delta_j(\omega_n) \geq 0.$$

We now prove that $\mathscr{M}_0$ belongs to $\overline{\mathbb{M}}$. From Equation (6.17)$_{\curvearrowleft}$, we infer that $\Delta_j \geq 0$ and therefore, it follows from property C5$_{45}$ that $\overline{Q}(\Delta_j|i) \geq 0$. Hence, due to Equation (6.18), it follows that $\mathscr{M}_0$ is bounded below by 0. Therefore, in order to prove that $\mathscr{M}_0 \in \overline{\mathbb{M}}$, it remains to prove that $\mathscr{M}_0$ is a supermartingale, or equivalently, that $\overline{Q}_\square(\Delta\mathscr{M}_0(\square)) \leq 0$ and $\overline{Q}(\Delta\mathscr{M}_0(x_{1:m})|x_m) \leq 0$ for all $m \in \mathbb{N}$ and $x_{1:m} \in \mathscr{X}^m$.

The first inequality is easily proved: since Equation (6.18) implies that $\Delta\mathscr{M}_0(\square) = 0$, it follows from property C5$_{45}$ that $\overline{Q}_\square(\Delta\mathscr{M}_0(\square)) = 0$. So, consider any $m \in \mathbb{N}$ and any $x_{1:m} \in \mathscr{X}^m$, then we need to prove that

$$\overline{Q}(\Delta\mathscr{M}_0(x_{1:m})|x_m) \leq 0.$$

We distinguish amongst three types of situations $x_{1:m}$.

If $x_1 \neq i$ or $x_k = j$ for at least one k in $\{2,\dots,m\}$, then as before, Equation (6.18) implies that $\Delta\mathscr{M}_0(x_{1:m}) = 0$, and therefore, it follows from property C5$_{45}$ that $\overline{Q}(\Delta\mathscr{M}_0(x_{1:m})|x_m) = 0$.

If $m = 1$ and $x_1 = i$, then

$$\begin{aligned}\overline{Q}(\Delta\mathscr{M}_0(i)|i) =& \overline{Q}(1+\Delta_j - [1+\overline{Q}(\Delta_j|i)]|i)\\ =& \overline{Q}(\Delta_j - \overline{Q}(\Delta_j|i)|i) = \overline{Q}(\Delta_j|i) - \overline{Q}(\Delta_j|i) = 0,\end{aligned}$$

where the first equality follows from Equation $(6.18)_{\frown}$ and the third equality from property $C8_{45}$.

The remaining situations $x_{1:m}$ are those for which $m \geq 2$, $x_1 = i$ and $x_k \neq j$ for all k in $\{2, \ldots, m\}$. Before tackling this type of situation, we first present some useful equations. For all $x \in \mathscr{X}$, it follows from Equation (6.18) that

$$\Delta\mathscr{M}_0(x_{1:m})(x) = (m+1) - 1 + \Delta_j(x) - (m-1+\Delta_j(x_m)) = 1 + \Delta_j(x) - \Delta_j(x_m). \tag{6.19}$$

Combining Equation (6.19) with Equation (6.17) results in

$$\Delta\mathscr{M}_0(x_{1:m})(x_m) = 1. \tag{6.20}$$

Also, if $x_m \neq L$, then since $x_m \neq j$, we find that

$$\Delta\mathscr{M}_0(x_{1:m})(x_m+1) = 1 + \Delta_j(x_m+1) - \Delta_j(x_m) = \begin{cases} 1 - \varepsilon^u_{x_m} & \text{if } x_m < j; \\ 1 + \varepsilon^d_{x_m+1} & \text{if } x_m > j; \end{cases} \tag{6.21}$$

similarly, if $x_m \neq 0$, we have that

$$\Delta\mathscr{M}_0(x_{1:m})(x_m-1) = 1 + \Delta_j(x_m-1) - \Delta_j(x_m) = \begin{cases} 1 + \varepsilon^u_{x_m-1} & \text{if } x_m < j; \\ 1 - \varepsilon^d_{x_m} & \text{if } x_m > j. \end{cases} \tag{6.22}$$

We now consider three cases: $x_m = 0$, $x_m = L$ and $x_m \notin \{0, L\}$.

If $x_m = 0$, it follows from Equations $(6.6)_{154}$ and $(6.9)_{154}$ that

$$\begin{aligned}\overline{Q}(\Delta\mathscr{M}_0(x_{1:m})|x_m) &= \max_{\pi_0 \in \mathscr{R}_0} \{(1-w_0)\Delta\mathscr{M}_0(x_{1:m})(0) + w_0\Delta\mathscr{M}_0(x_{1:m})(1)\}\\ &= \max_{\pi_0 \in \mathscr{R}_0} \{(1-w_0) + w_0(1-\varepsilon^u_0)\} = \max_{\pi_0 \in \mathscr{R}_0} \{-w_0\varepsilon^u_0\} + 1\\ &= -\underline{w}_0\varepsilon^u_0 + 1 = 0,\end{aligned}$$

where the second equality follows from Equations (6.20) and (6.21) and the fourth holds because ε^u_0 is strictly positive.

If $x_m = L$, it follows from Equations $(6.7)_{154}$ and $(6.9)_{154}$ that

$$\begin{aligned}\overline{Q}(\Delta\mathscr{M}_0(x_{1:m})|x_m) &= \max_{\pi_L \in \mathscr{R}_L} \{b_L\Delta\mathscr{M}_0(x_{1:m})(L-1) + (1-b_L)\Delta\mathscr{M}_0(x_{1:m})(L)\}\\ &= \max_{\pi_L \in \mathscr{R}_L} \{b_L(1-\varepsilon^d_L) + (1-b_L))\} = \max_{\pi_L \in \mathscr{R}_L} \{-b_L\varepsilon^d_L\} + 1\\ &= -\underline{b}_L\varepsilon^d_L + 1 = 0,\end{aligned}$$

where the second equality follows from Equations (6.20)$_{\curvearrowleft}$ and (6.22)$_{\curvearrowleft}$ and the fourth holds because ε_L^d is strictly positive.

If $x_m \notin \{0, L\}$, it follows from Equations (6.5)$_{154}$ and (6.9)$_{154}$ that

$$\begin{aligned}
\overline{Q}(\Delta\mathscr{M}_0(x_{1:m})|x_m) &= \max_{\pi_{x_m}\in\mathscr{R}_{x_m}} \Big\{ b_{x_m}\Delta\mathscr{M}_0(x_{1:m})(x_m-1) \\
&\qquad\qquad + (1-b_{x_m}-w_{x_m})\Delta\mathscr{M}_0(x_{1:m})(x_m) \\
&\qquad\qquad + w_{x_m}\Delta\mathscr{M}_0(x_{1:m})(x_m+1) \Big\} \\
&= \begin{cases} \displaystyle\max_{\pi_{x_m}\in\mathscr{R}_{x_m}} \{ b_{x_m}\varepsilon^u_{x_m-1} - w_{x_m}\varepsilon^u_{x_m} \} + 1 & \text{if } x_m < j; \\ \displaystyle\max_{\pi_{x_m}\in\mathscr{R}_{x_m}} \{ -b_{x_m}\varepsilon^d_{x_m} + w_{x_m}\varepsilon^d_{x_m+1} \} + 1 & \text{if } x_m > j, \end{cases}
\end{aligned}$$

where the last equality holds because of Equations (6.20)$_{\curvearrowleft}$–(6.22)$_{\curvearrowleft}$. Hence, if $x_m < j$, we find that

$$\begin{aligned}
\overline{Q}(\Delta\mathscr{M}_0(x_{1:m})|x_m) &= \max_{\pi_{x_m}\in\mathscr{R}_{x_m}} \{ b_{x_m}\varepsilon^u_{x_m-1} - w_{x_m}\varepsilon^u_{x_m} \} + 1 \\
&\leq \max_{\pi_{x_m}\in\mathscr{R}_{x_m}} \{ b_{x_m}\varepsilon^u_{x_m-1} \} + \max_{\pi_{x_m}\in\mathscr{R}_{x_m}} \{ -w_{x_m}\varepsilon^u_{x_m} \} + 1 \\
&= \overline{b}_{x_m}\varepsilon^u_{x_m-1} - \underline{w}_{x_m}\varepsilon^u_{x_m} + 1 = 0,
\end{aligned}$$

where the second equality holds because $\varepsilon^u_{x_m-1}$ and $\varepsilon^u_{x_m}$ are strictly positive and the third equality follows from (6.16)$_{156}$. Similarly, if $x_m > j$, we find that

$$\begin{aligned}
\overline{Q}(\Delta\mathscr{M}_0(x_{1:m})|x_m) &= \max_{\pi_{x_m}\in\mathscr{R}_{x_m}} \{ -b_{x_m}\varepsilon^d_{x_m} + w_{x_m}\varepsilon^d_{x_m+1} \} + 1 \\
&\leq \max_{\pi_{x_m}\in\mathscr{R}_{x_m}} \{ -b_{x_m}\varepsilon^d_{x_m} \} + \max_{\pi_{x_m}\in\mathscr{R}_{x_m}} \{ w_{x_m}\varepsilon^d_{x_m+1} \} + 1 \\
&= -\underline{b}_{x_m}\varepsilon^d_{x_m} + \overline{w}_{x_m}\varepsilon^d_{x_m+1} + 1 = 0.
\end{aligned}$$ □

In the rest of this section, we will derive a system of non-linear equations for lower and upper expected first-passage times. The starting point for this derivation is the fact that

$$\underline{\tau}_{i\to j} = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^1_{i\to j}|i) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^1_{i\to j}|i, X_2)|i), \tag{6.23}$$

which is a direct consequence of Theorem 28$_{94}$ and Proposition 27$_{93}$. Next, in order to express $\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^1_{i\to j}|i, X_2)$ in terms of lower expected first-passage times, we start by observing that

$$\begin{aligned}
\tau^1_{i\to j}(i, X_{2:\infty}) &= \begin{cases} 1 & \text{if } X_2 = j \\ 1 + \tau^2_{z\to j}(i, z, X_{3:\infty}) & \text{if } X_2 = z \neq j \end{cases} \\
&= 1 + \sum_{z\in\mathscr{X}\setminus\{j\}} \mathbb{I}_z(X_2)\tau^2_{z\to j}(i, z, X_{3:\infty}),
\end{aligned} \tag{6.24}$$

where we adopt the convention that $0 \cdot +\infty = 0$. If we now consider any $z \in \mathscr{X} \setminus \{j\}$, then we infer from Equation $(6.24)_{\curvearrowleft}$, Proposition 27_{93} [repeatedly] and property $\mathrm{G6}_{92}$ that

$$\begin{aligned}\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^1_{i\to j}|i,z) =&\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^1_{i\to j}(i,z,X_{3:\infty})|i,z) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(1+\tau^2_{i\to j}(i,z,X_{3:\infty})|i,z)\\ =&1+\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^2_{i\to j}(i,z,X_{3:\infty})|i,z) = 1+\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^2_{i\to j}|i,z) = 1+\underline{\tau}_{z\to j}.\end{aligned}$$

Similarly, we infer from Equation $(6.24)_{\curvearrowleft}$, Proposition 27_{93} and property $\mathrm{G5}_{92}$ that

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^1_{i\to j}|i,j) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^1_{i\to j}(i,j,X_{3:\infty})|i,j) = \underline{E}^{\mathrm{ei}}_{\underline{Q}}(1|i,j) = 1.$$

Hence, from the two aforementioned equations it follows that

$$\underline{E}^{\mathrm{ei}}_{\underline{Q}}(\tau^1_{i\to j}|i,X_2) = 1 + \sum_{z\in\mathscr{X}\setminus\{j\}} \mathbb{I}_z(X_2)\underline{\tau}_{z\to j},$$

and by combining this with Equation $(6.23)_{\curvearrowleft}$, we find that

$$\begin{aligned}\underline{\tau}_{i\to j} =&\underline{E}^{\mathrm{ei}}_{\underline{Q}}\left(1+\sum_{z\in\mathscr{X}\setminus\{j\}} \mathbb{I}_z(X_2)\underline{\tau}_{z\to j}\middle|i\right) = 1+\underline{E}^{\mathrm{ei}}_{\underline{Q}}\left(\sum_{z\in\mathscr{X}\setminus\{j\}} \mathbb{I}_z(X_2)\underline{\tau}_{z\to j}\middle|i\right)\\ =&1+\underline{Q}\left(\sum_{z\in\mathscr{X}\setminus\{j\}} \mathbb{I}_z\underline{\tau}_{z\to j}\middle|i\right),\end{aligned}$$

where the second equality follows from property $\mathrm{G6}_{92}$ and the last equality from Corollary 25_{91} combined with the fact that $\underline{\tau}_{z\to j}$ is real-valued due to Theorem 58_{156}.

Because of Equations $(6.5)_{154}$–$(6.7)_{154}$, we now finally obtain the following system of non-linear equations: for all $j \in \mathscr{X}$, we have that

$$\underline{\tau}_{0\to j} = 1 + \min_{\pi_0\in\mathscr{R}_0}\left\{r_0\mathbb{I}_{\neg j}(0)\underline{\tau}_{0\to j} + w_0\mathbb{I}_{\neg j}(1)\underline{\tau}_{1\to j}\right\} \tag{6.25}$$

and

$$\underline{\tau}_{L\to j} = 1 + \min_{\pi_L\in\mathscr{R}_L}\left\{b_L\mathbb{I}_{\neg j}(L-1)\underline{\tau}_{L-1\to j} + r_L\mathbb{I}_{\neg j}(L)\underline{\tau}_{L\to j}\right\} \tag{6.26}$$

and, for all $i \in \mathscr{X}/\{0,L\}$, we have that

$$\underline{\tau}_{i\to j} = 1 + \min_{\pi_i\in\mathscr{R}_i}\left\{b_i\mathbb{I}_{\neg j}(i-1)\underline{\tau}_{i-1\to j} + r_i\mathbb{I}_{\neg j}(i)\underline{\tau}_{i\to j} + w_i\mathbb{I}_{\neg j}(i+1)\underline{\tau}_{i+1\to j}\right\}, \tag{6.27}$$

where we let $\mathbb{I}_{\neg j} := 1 - \mathbb{I}_j$. Using a completely analogous derivation, we also find that

$$1+\overline{Q}\left(\sum_{z\in\mathscr{X}\setminus\{j\}} \mathbb{I}_z\overline{\tau}_{z\to j}\middle|i\right), \tag{6.28}$$

which gives rise to a similar system of non-linear equations.

In order to compute $\underline{\tau}_{i\to j}$ and $\overline{\tau}_{i\to j}$, we need to solve these systems of equations. However, since the equations in these systems are non-linear, it is not always feasible to solve them directly. Fortunately, as we will show in the following three sections, it is possible to transform them into simple recursive expressions, which can then be used to compute $\underline{\tau}_{i\to j}$ and $\overline{\tau}_{i\to j}$ for all $i,j\in\mathscr{X}$.

6.4 LOWER AND UPPER EXPECTATIONS OF UPWARD FIRST-PASSAGE TIMES

We start by computing lower expectations of upward first-passage times, that is, for all $i,j\in\mathscr{X}$ such that $i<j$, we will compute $\underline{\tau}_{i\to j}$. We initially focus on computing $\underline{\tau}_{i\to i+1}$, for $i\in\mathscr{X}\setminus\{L\}$, and then show that any lower expected upward first-passage time can be obtained as a sum of such terms. Similar results are obtained for upper expected upward first-passage times.

Finding $\underline{\tau}_{0\to 1}$ is easy. It follows from Equation (6.25), with $j=1$, that

$$\begin{aligned}\underline{\tau}_{0\to 1} = 1+\min_{\pi_0\in\mathscr{R}_0} r_0\underline{\tau}_{0\to 1} &= 1+\min_{\pi_0\in\mathscr{R}_0}(1-w_0)\underline{\tau}_{0\to 1}\\ &= 1+\underline{\tau}_{0\to 1}-\max_{\pi_0\in\mathscr{R}_0} w_0\underline{\tau}_{0\to 1} = 1+\underline{\tau}_{0\to 1}-\overline{w}_0\underline{\tau}_{0\to 1},\end{aligned}$$

where the second equality holds because π_0 is a probability mass function on a binary set and the last equality holds because we know from Theorem 58_{156} that $\underline{\tau}_{0\to 1}$ is real-valued and therefore finite. Hence, since we know from Theorem 58_{156} that $\underline{\tau}_{0\to 1}$ is strictly positive and real-valued, it follows that

$$\underline{\tau}_{0\to 1} = \frac{1}{\overline{w}_0}. \tag{6.29}$$

Finding $\underline{\tau}_{0\to j}$, for $j\in\{2,\dots,L\}$, is more involved. We start by establishing a connection with $\underline{\tau}_{1\to j}$. By applying Equation $(6.25)_{\curvearrowleft}$, we find that

$$\begin{aligned}\underline{\tau}_{0\to j} = 1+\min_{\pi_0\in\mathscr{R}_0}\{r_0\underline{\tau}_{0\to j}+w_0\underline{\tau}_{1\to j}\} &= 1+\min_{\pi_0\in\mathscr{R}_0}\{(1-w_0)\underline{\tau}_{0\to j}+w_0\underline{\tau}_{1\to j}\}\\ &= 1+\underline{\tau}_{0\to j}+\min_{\pi_0\in\mathscr{R}_0} w_0(\underline{\tau}_{1\to j}-\underline{\tau}_{0\to j}),\end{aligned}$$

which implies, due to Theorem 58_{156}, that

$$\min_{\pi_0\in\mathscr{R}_0} w_0(\underline{\tau}_{1\to j}-\underline{\tau}_{0\to j}) = -1. \tag{6.30}$$

Since the minimum in Equation (6.30) is negative and w_0 is a probability and therefore non-negative, it must be that $\underline{\tau}_{1\to j}-\underline{\tau}_{0\to j}<0$. Therefore, Equation (6.30) is minimised for $w_0=\overline{w}_0$ and we find that

$$\underline{\tau}_{0\to j} = \frac{1}{\overline{w}_0}+\underline{\tau}_{1\to j}. \tag{6.31}$$

By combining Equations (6.29)$_{\curvearrowleft}$ and (6.31)$_{\curvearrowleft}$, we see that

$$\underline{\tau}_{0\to j} = \underline{\tau}_{0\to 1} + \underline{\tau}_{1\to j} \text{ for all } j \in \{2,\dots,L\}. \tag{6.32}$$

Since we already know $\underline{\tau}_{0\to 1}$—see Equation (6.29)$_{\curvearrowleft}$—we are now left to find $\underline{\tau}_{1\to j}$.

We first consider the case $j=2$. There, it follows from Equation (6.27)$_{160}$, with $i=1$ and $j=2$, that

$$\begin{aligned}\underline{\tau}_{1\to 2} =& 1 + \min_{\pi_1 \in \mathscr{R}_1} \{b_1 \underline{\tau}_{0\to 2} + r_1 \underline{\tau}_{1\to 2}\} \\ =& 1 + \min_{\pi_1 \in \mathscr{R}_1} \{b_1 \underline{\tau}_{0\to 2} + (1-b_1-w_1)\underline{\tau}_{1\to 2}\} \\ =& 1 + \underline{\tau}_{1\to 2} + \min_{\pi_1 \in \mathscr{R}_1} \{b_1(\underline{\tau}_{0\to 2} - \underline{\tau}_{1\to 2}) - w_1 \underline{\tau}_{1\to 2}\},\end{aligned}$$

which implies, due to Theorem 58$_{156}$, that

$$\min_{\pi_1 \in \mathscr{R}_1} \{b_1(\underline{\tau}_{0\to 2} - \underline{\tau}_{1\to 2}) - w_1 \underline{\tau}_{1\to 2}\} = -1.$$

By applying Equation (6.32) for $j=2$ we then find that

$$\min_{\pi_1 \in \mathscr{R}_1} \{b_1 \underline{\tau}_{0\to 1} - w_1 \underline{\tau}_{1\to 2}\} = -1. \tag{6.33}$$

Therefore, and because we already know the value of $\underline{\tau}_{0\to 1}$, it follows from Assumption 6.1$_{153}$ and the following lemma that $\underline{\tau}_{1\to 2}$ is the unique solution to Equation (6.33).

Proposition 59. *Consider a closed set $\mathscr{R}$ on $\mathscr{X}_m$ that consists of strictly positive probability mass functions and let c be a real constant. Then*

$$\min_{\pi \in \mathscr{R}} \{bc - w\mu\}$$

is a strictly decreasing function of μ.

Proof. Consider any $\mu_1, \mu_2 \in \mathbb{R}$, such that $\mu_2 > \mu_1$. Then,

$$\begin{aligned}\min_{\pi \in \mathscr{R}} \{bc - w\mu_1\} &= \min_{\pi \in \mathscr{R}} \{bc - w\mu_2 + w(\mu_2 - \mu_1)\} \\ &\geq \min_{\pi \in \mathscr{R}} \{bc - w\mu_2\} + \min_{\pi \in \mathscr{R}} \{w(\mu_2 - \mu_1)\} > \min_{\pi \in \mathscr{R}} \{bc - w\mu_2\}\end{aligned}$$

where the last inequality holds because, since $\mu_2 - \mu_1 > 0$,

$$\min_{\pi \in \mathscr{R}} \{w(\mu_2 - \mu_1)\} = (\mu_2 - \mu_1) \min_{\pi \in \mathscr{R}} w,$$

where $\min_{\pi \in \mathscr{R}} w > 0$ because the closed set $\mathscr{R}$ consists of strictly positive probability mass functions. □

This shows that the unique solution $\underline{\tau}_{1\to 2}$ is furthermore easy to compute. Indeed, it follows from Proposition $59_{\curvearrowleft}$ that a simple bisection method suffices.

Next, we consider the case $j \in \{3,\ldots,L\}$. By applying Equation $(6.27)_{160}$, for such a j and with $i = 1$, we find that

$$\begin{aligned}
\underline{\tau}_{1\to j} &= 1 + \min_{\pi_1 \in \mathscr{R}_1} \{b_1 \underline{\tau}_{0\to j} + r_1 \underline{\tau}_{1\to j} + w_1 \underline{\tau}_{2\to j}\} \\
&= 1 + \min_{\pi_1 \in \mathscr{R}_1} \{b_1 \underline{\tau}_{0\to j} + (1 - b_1 - w_1)\underline{\tau}_{1\to j} + w_1 \underline{\tau}_{2\to j}\} \\
&= 1 + \underline{\tau}_{1\to j} + \min_{\pi_1 \in \mathscr{R}_1} \{b_1(\underline{\tau}_{0\to j} - \underline{\tau}_{1\to j}) + w_1(\underline{\tau}_{2\to j} - \underline{\tau}_{1\to j})\},
\end{aligned}$$

which implies, due to Theorem 58_{156}, that

$$\min_{\pi_1 \in \mathscr{R}_1} \{b_1(\underline{\tau}_{0\to j} - \underline{\tau}_{1\to j}) + w_1(\underline{\tau}_{2\to j} - \underline{\tau}_{1\to j})\} = -1.$$

In combination with Equation $(6.32)_{\curvearrowleft}$, this results in

$$\min_{\pi_1 \in \mathscr{R}_1} \{b_1 \underline{\tau}_{0\to 1} + w_1(\underline{\tau}_{2\to j} - \underline{\tau}_{1\to j})\} = -1. \tag{6.34}$$

Since we know from Assumption 6.1_{153} and Proposition $59_{\curvearrowleft}$ that the equation

$$\min_{\pi_1 \in \mathscr{R}_1} \{b_1 \underline{\tau}_{0\to 1} + w_1 \mu\} = -1$$

has a unique solution μ, it follows directly from Equations $(6.33)_{\curvearrowleft}$ and (6.34) that

$$\underline{\tau}_{1\to j} = \underline{\tau}_{1\to 2} + \underline{\tau}_{2\to j} \text{ for all } j \in \{3,\ldots,L\}. \tag{6.35}$$

At this point, we already know how to compute $\underline{\tau}_{0\to 1}$ and $\underline{\tau}_{1\to 2}$ and we have also established the following additivity property:

$$\underline{\tau}_{i\to j} = \underline{\tau}_{i\to i+1} + \underline{\tau}_{i+1\to j}$$

for all $i \in \{0,1\}$ and $j \in \{i+2,\ldots,L\}$. By continuing in this way, we obtain the following two results, which are direct consequences of Lemma 81_{181} in Appendix $6.A_{181}$.

Proposition 60. *For any $i \in \mathscr{X} \setminus \{0,L\}$, we have that*

$$\min_{\pi_i \in \mathscr{R}_i} \{b_i \underline{\tau}_{i-1\to i} - w_i \underline{\tau}_{i\to i+1}\} = -1. \tag{6.36}$$

Proposition 61. *For all $i,j \in \mathscr{X}$ such that $i+1 < j$, we have that*

$$\underline{\tau}_{i\to j} = \underline{\tau}_{i\to i+1} + \underline{\tau}_{i+1\to j}.$$

For all $i \in \mathscr{X} \setminus \{L\}$, the value of $\underline{\tau}_{i\to i+1}$ can therefore be computed recursively. For $i = 0$, we simply apply Equation $(6.29)_{161}$. For any other $i \in \mathscr{X} \setminus \{0, L\}$, it follows from Assumption 6.1_{153} and Propositions 59_{162} and $60_{\curvearrowleft}$ that $\underline{\tau}_{i\to i+1}$ is the unique solution to Equation (6.36), which can be obtained by means of a bisection method. In this equation, the value of $\underline{\tau}_{i-1\to i}$ has already been computed earlier on in this recursive procedure.

The following additivity result is a consequence of Proposition $61_{\curvearrowleft}$.

Corollary 62. *For all $i, j \in \mathscr{X}$ such that $i < j$, we have that*

$$\underline{\tau}_{i\to j} = \sum_{k=i}^{j-1} \underline{\tau}_{k\to k+1}.$$

Proof. For $j = i+1$, this result is trivial. For $j = i+2$, it follows from Proposition $61_{\curvearrowleft}$ that

$$\underline{\tau}_{i\to i+2} = \underline{\tau}_{i\to i+1} + \underline{\tau}_{i+1\to i+2}.$$

Similarly, for $j > i+2$, by applying Proposition $61_{\curvearrowleft}$ multiple times, we find that

$$\begin{aligned}\underline{\tau}_{i\to j} &= \underline{\tau}_{i\to i+1} + \underline{\tau}_{i+1\to j} = \underline{\tau}_{i\to i+1} + \underline{\tau}_{i+1\to i+2} + \underline{\tau}_{i+2\to j} \\ &= \underline{\tau}_{i\to i+1} + \underline{\tau}_{i+1\to i+2} + \ldots + \underline{\tau}_{j-1\to j} = \sum_{k=i}^{j-1} \underline{\tau}_{k\to k+1}.\end{aligned}$$ □

It implies that the recursive techniques that we developed in this section can be used to compute any lower expected upward first-passage time.

Similar results can be proved for upper expectations of upward first-passage times. We only provide the final expressions; the derivation is completely analogous. In this case, the starting point is that

$$\overline{\tau}_{0\to 1} = \frac{1}{\underline{p}_0}. \tag{6.37}$$

For all $i \in \mathscr{X} \setminus \{0, L\}$, the value of $\overline{\tau}_{i\to i+1}$ can then be computed recursively, due to Assumption 6.1_{153} and the next two results. The first result is a direct consequence of Lemma 82_{182} in Appendix $6.\mathrm{A}_{181}$.

Proposition 63. *For all $i \in \mathscr{X} \setminus \{0, L\}$, we have that*

$$\max_{\pi_i \in \mathscr{R}_i} \{b_i \overline{\tau}_{i-1\to i} - w_i \overline{\tau}_{i\to i+1}\} = -1. \tag{6.38}$$

Proposition 64. *Consider a closed set $\mathscr{R}$ on $\mathscr{X}_m$ that consists of strictly positive probability mass functions and let c be a real constant. Then*

$$\max_{\pi \in \mathscr{R}} \{bc - w\mu\}$$

is a strictly decreasing function of μ.

Proof of Proposition 64. Consider any $\mu_1, \mu_2 \in \mathbb{R}$, such that $\mu_2 > \mu_1$. Then,

$$\begin{aligned}\max_{\pi\in\mathscr{R}}\{bc - w\mu_2\} &= \max_{\pi\in\mathscr{R}}\{bc - w\mu_1 + w(\mu_1 - \mu_2)\} \\ &\leq \max_{\pi\in\mathscr{R}}\{bc - w\mu_1\} + \max_{\pi\in\mathscr{R}}\{w(\mu_1 - \mu_2)\} < \max_{\pi\in\mathscr{R}}\{bc - w\mu_1\}\end{aligned}$$

where the last inequality holds because

$$\max_{\pi\in\mathscr{R}}\{w(\mu_1 - \mu_2)\} = (\mu_1 - \mu_2)\max_{\pi\in\mathscr{R}} w$$

where $\max_{\pi\in\mathscr{R}} w > 0$ because the closed set $\mathscr{R}$ consists of strictly positive probability mass functions. □

Due to our next result, this recursive technique allows us to compute arbitrary upper expected upward first-passage times.

Proposition 65. *For all $i, j \in \mathscr{X}$ such that $i < j$, we have that*

$$\overline{\tau}_{i\to j} = \sum_{k=i}^{j-1} \overline{\tau}_{k\to k+1}.$$

Proof. For $j = i+1$, this result is trivial. For $j = i+2$, it follows from Lemma 82_{182} that

$$\overline{\tau}_{i\to i+2} = \overline{\tau}_{i\to i+1} + \overline{\tau}_{i+1\to i+2}$$

Similarly, for $j > i+2$, by applying Lemma 82_{182} multiple times, we find that

$$\begin{aligned}\overline{\tau}_{i\to j} &= \overline{\tau}_{i\to i+1} + \overline{\tau}_{i+1\to j} = \overline{\tau}_{i\to i+1} + \overline{\tau}_{i+1\to i+2} + \overline{\tau}_{i+2\to j} \\ &= \overline{\tau}_{i\to i+1} + \overline{\tau}_{i+1\to i+2} + \ldots + \overline{\tau}_{j-1\to j} = \sum_{k=i}^{j-1} \overline{\tau}_{k\to k+1}.\end{aligned}$$

□

6.5 LOWER AND UPPER EXPECTATIONS OF DOWNWARD FIRST-PASSAGE TIMES

Lower and upper expectations of downward first-passage times can be computed in more or less the same way. The main difference is that the recursive expressions now start from the other side, that is, from $i = L$.[5] We find that

$$\underline{\tau}_{L\to L-1} = \frac{1}{\overline{b}_L} \text{ and } \overline{\tau}_{L\to L-1} = \frac{1}{\underline{b}_L}. \tag{6.39}$$

[5] Our presentation of—and proofs for—the results in this section are adapted versions of the ones in Section 6.4_{161}. An alternative method would be to observe that a downward first-passage time from i to j is the same as an upward first-passage time from $L-i$ to $L-j$ in a new imprecise birth-death chain, obtained by reversing the order of the states, and by switching the role of w and b accordingly.

For all $i \in \mathscr{X} \setminus \{0,L\}$, due to Assumption 6.1[153], the values of $\underline{\tau}_{i\to i-1}$ and $\overline{\tau}_{i\to i-1}$ can now be computed recursively, using the next two results. The first result is a direct consequence of Lemmas 83[184] and 84[186] in Appendix 6.A[181].

Proposition 66. *For all $i \in \mathscr{X} \setminus \{0,L\}$, we have that*

$$\min_{\pi_i \in \mathscr{R}_i} \{-b_i \underline{\tau}_{i\to i-1} + w_i \underline{\tau}_{i+1\to i}\} = -1 \text{ and } \max_{\pi_i \in \mathscr{R}_i} \{-b_i \overline{\tau}_{i\to i-1} + w_i \overline{\tau}_{i+1\to i}\} = -1.$$

Proposition 67. *Consider a closed set $\mathscr{R}$ on $\mathscr{X}_m$ that consists of strictly positive probability mass functions and let c be a real constant. Then*

$$\min_{\pi \in \mathscr{R}} \{-b\mu + wc\} \text{ and } \max_{\pi \in \mathscr{R}} \{-b\mu + wc\}$$

are strictly decreasing functions of μ.

Proof. Consider any $\mu_1, \mu_2 \in \mathbb{R}$, such that $\mu_2 > \mu_1$. Then

$$\begin{aligned}\min_{\pi \in \mathscr{R}} \{-b\mu_1 + wc\} &= \min_{\pi \in \mathscr{R}} \{b(\mu_2 - \mu_1) - b\mu_2 + wc\} \\ &\geq \min_{\pi \in \mathscr{R}} \{b(\mu_2 - \mu_1)\} + \min_{\pi \in \mathscr{R}} \{-b\mu_2 + wc\} > \min_{\pi \in \mathscr{R}} \{-b\mu_2 + wc\},\end{aligned}$$

where the last inequality holds because

$$\min_{\pi \in \mathscr{R}} \{b(\mu_2 - \mu_1)\} = (\mu_2 - \mu_1) \min_{\pi \in \mathscr{R}} b$$

where $\min_{\pi \in \mathscr{R}} b > 0$ because the closed set $\mathscr{R}$ consists of strictly positive probability mass functions.

Similarly,

$$\begin{aligned}\max_{\pi \in \mathscr{R}} \{-b\mu_2 + wc\} &= \max_{\pi \in \mathscr{R}} \{b(\mu_1 - \mu_2) - b\mu_1 + wc\} \\ &\leq \max_{\pi \in \mathscr{R}} \{b(\mu_1 - \mu_2)\} + \max_{\pi \in \mathscr{R}} \{-b\mu_1 + wc\} < \max_{\pi \in \mathscr{R}} \{-b\mu_1 + wc\},\end{aligned}$$

where the last inequality holds because

$$\max_{\pi \in \mathscr{R}} \{b(\mu_1 - \mu_2)\} = (\mu_1 - \mu_2) \max_{\pi \in \mathscr{R}} b,$$

where $\max_{\pi \in \mathscr{R}} b > 0$ because the closed set $\mathscr{R}$ consists of strictly positive probability mass functions. □

Once we have computed $\underline{\tau}_{i\to i-1}$ and $\overline{\tau}_{i\to i-1}$ for all $i \in \mathscr{X} \setminus \{L\}$, the following result enables us to easily obtain all other lower and upper expected downward first-passage times.

Proposition 68. *For all $i, j \in \mathscr{X}$ such that $i > j$, we have that*

$$\underline{\tau}_{i\to j} = \sum_{k=j}^{i-1} \underline{\tau}_{k+1\to k} \text{ and } \overline{\tau}_{i\to j} = \sum_{k=j}^{i-1} \overline{\tau}_{k+1\to k}.$$

Proof. We first prove the lower case. For $j = i-1$, this result is trivial. For $j = i-2$, it follows from Lemma 83_{184} that

$$\underline{\tau}_{i\to i-2} = \underline{\tau}_{i\to i-1} + \underline{\tau}_{i-1\to i-2}$$

Similarly, for $j < i-2$, by applying Lemma 83_{184} multiple times, we find that

$$\begin{aligned}\underline{\tau}_{i\to j} &= \underline{\tau}_{i\to i-1} + \underline{\tau}_{i-1\to j} = \underline{\tau}_{i\to i-1} + \underline{\tau}_{i-1\to i-2} + \underline{\tau}_{i-2\to j} \\ &= \underline{\tau}_{i\to i-1} + \underline{\tau}_{i-1\to i-2} + \ldots + \underline{\tau}_{j+1\to j} = \sum_{k=j}^{i-1} \underline{\tau}_{k+1\to k}.\end{aligned}$$

Next, we prove the upper case. For $j = i-1$, this result is trivial. For $j = i-2$, it follows from Lemma 84_{186} that

$$\overline{\tau}_{i\to i-2} = \overline{\tau}_{i\to i-1} + \overline{\tau}_{i-1\to i-2}$$

Similarly, for $j < i-2$, by applying Lemma 84_{186} multiple times, we find that

$$\begin{aligned}\overline{\tau}_{i\to j} &= \overline{\tau}_{i\to i-1} + \overline{\tau}_{i-1\to j} = \overline{\tau}_{i\to i-1} + \overline{\tau}_{i-1\to i-2} + \overline{\tau}_{i-2\to j} \\ &= \overline{\tau}_{i\to i-1} + \overline{\tau}_{i-1\to i-2} + \ldots + \overline{\tau}_{j+1\to j} = \sum_{k=j}^{i-1} \overline{\tau}_{k+1\to k}.\end{aligned}$$ □

6.6 LOWER AND UPPER EXPECTATIONS OF RETURN TIMES

Given the results in the previous two sections, lower and upper expected return times can now be computed very easily. By applying Equations $(6.25)_{160}$–$(6.27)_{160}$, with j equal to 0, L and i, respectively, we find that

$$\underline{\tau}_{0\to 0} = 1 + \min_{\pi_0 \in \mathscr{R}_0} w_0 \underline{\tau}_{1\to 0} = 1 + \underline{w}_0 \underline{\tau}_{1\to 0} \tag{6.40}$$

and

$$\underline{\tau}_{L\to L} = 1 + \min_{\pi_L \in \mathscr{R}_L} b_L \underline{\tau}_{L-1\to L} = 1 + \underline{b}_L \underline{\tau}_{L-1\to L} \tag{6.41}$$

and, for all $i \in \mathscr{X} \setminus \{0, L\}$, that

$$\underline{\tau}_{i\to i} = 1 + \min_{\pi_i \in \mathscr{R}_i} \{b_i \underline{\tau}_{i-1\to i} + w_i \underline{\tau}_{i+1\to i}\}. \tag{6.42}$$

In these expressions, the lower expected first-passage times $\underline{\tau}_{1\to 0}$, $\underline{\tau}_{L-1\to L}$, $\underline{\tau}_{i-1\to i}$ and $\underline{\tau}_{i+1\to i}$ can be computed using the recursive techniques that we developed in the previous two sections. Similarly, for the upper expectation case, we find that

$$\overline{\tau}_{0\to 0} = 1 + \max_{\pi_0 \in \mathscr{R}_0} w_0 \overline{\tau}_{1\to 0} = 1 + \overline{w}_0 \overline{\tau}_{1\to 0} \tag{6.43}$$

and

$$\overline{\tau}_{L\to L} = 1 + \max_{\pi_L \in \mathscr{R}_L} b_L \overline{\tau}_{L-1\to L} = 1 + \overline{b}_L \overline{\tau}_{L-1\to L} \tag{6.44}$$

and, for all $i \in \mathscr{X} \setminus \{0, L\}$, that

$$\overline{\tau}_{i\to i} = 1 + \max_{\pi_i \in \mathscr{R}_i} \{b_i \overline{\tau}_{i-1\to i} + w_i \overline{\tau}_{i+1\to i}\}. \tag{6.45}$$

Again, the upper expected first-passage times $\overline{\tau}_{1\to 0}$, $\overline{\tau}_{L-1\to L}$, $\overline{\tau}_{i-1\to i}$ and $\overline{\tau}_{i+1\to i}$ that appear in these expressions can be computed with the recursive techniques that were introduced above.

6.7 PRECISE BIRTH-DEATH CHAINS AS A SPECIAL CASE

Since birth-death chains are a special case of imprecise birth-death chains, our results for imprecise birth-death chains can also be applied to birth-death chains. We now study this special case in some detail, and establish a connection with the more general one. For now, we will still be focusing on the martingale-theoretic approach. The measure-theoretic approach will be considered in Section 6.8_{177}.

6.7.1 *Expected first-passage and return times in precise birth-death chains*

Clearly, a birth-death chain can be regarded as a special type of imprecise birth-death chain. It corresponds to the case where all sets of probability mass functions are singletons, that is, $\mathscr{Q}_\square = \{q_\square\}$ and, for all $i \in \mathscr{X}$, $\mathscr{R}_i = \{\pi_i\}$. We refer to this special type of imprecise birth-death chain as a precise birth-death chain. For these precise birth-death chains, as the following result implies, lower and upper expected first-passage and return times coincide.

Proposition 69. *Consider an imprecise birth-death chain such that, for all $i \in \mathscr{X}$, $\mathscr{R}_i = \{\pi_i\}$. Then*

$$\underline{\tau}_{i\to j} = \overline{\tau}_{i\to j} \textit{ for all } i, j \in \mathscr{X}.$$

Proof. This result is an immediate consequence of the fact that, in this case, our recursive equations for computing $\underline{\tau}_{i\to j}$ become identical to the equations that we use to compute $\overline{\tau}_{i\to j}$. For example, for upward first-passage times, Equation $(6.29)_{161}$ is now identical to Equation $(6.37)_{164}$ because $\underline{w}_0 = w_0 =$

$\overline{w}_0$, and Equation (6.36)$_{163}$ is now identical to Equation (6.38)$_{164}$ because, since $\mathscr{R}_i = \{\pi_i\}$, the minimum and maximum disappear. Similar observations can also be made for all the other recursive equations in Section 6.4$_{161}$–Section 6.6$_{167}$. □

Notice that this result does not require that the initial set of probability mass functions $\mathscr{Q}_\square$ should be a singleton. This is not surprising: since none of the methods in Section 6.4$_{161}$–Section 6.6$_{167}$ require the use of $\mathscr{Q}_\square$, it follows that $\mathscr{Q}_\square$ has no effect on first-passage or return times. Therefore, for our present purposes, all the relevant parameters of a birth-death chain can be represented by a single transition matrix M, the form of which is given by Equation (6.1)$_{152}$. Due to Proposition 69$_\frown$, with any such matrix M, we can associate a unique expected first-passage time from $i \in \mathscr{X}$ to $j \in \mathscr{X}$, defined by

$$\tau^M_{i\to j} := \underline{\tau}_{i\to j} = \overline{\tau}_{i\to j} \text{ for all } i,j \in \mathscr{X}. \tag{6.46}$$

If $i = j$, then $\tau^M_{i\to i}$ is called an expected return time. As it turns out, we can derive closed-form expressions for these expected first-passage and return times. The following lemma presents such an expression for a specific type of expected upward first-passage times.

Proposition 70. *Consider a precise birth-death chain of which the transition matrix M is given by Equation* (6.1)$_{152}$. *Then for all $i \in \mathscr{X} \setminus \{L\}$, we have that*

$$\tau^M_{i\to i+1} = \sum_{k=0}^{i} \frac{\prod_{\ell=k+1}^{i} b_\ell}{\prod_{m=k}^{i} w_m}. \tag{6.47}$$

Proof. We provide a proof by induction. Since $\overline{w}_0 = w_0$, it follows from Equations (6.46) and (6.29)$_{161}$ that $\tau^M_{0\to 1} = \underline{\tau}_{0\to 1} = 1/\overline{w}_0 = 1/w_0$, proving Equation (6.47) for $i = 0$.

Consider now any $i \in \mathscr{X} \setminus \{0, L\}$ and let us assume, as our induction hypothesis, that the result is true for $i-1$. Since $\mathscr{R}_i = \{\pi_i\}$, it follows from Proposition 60$_{163}$ and Equation (6.46) that $b_i \tau^M_{i-1\to i} - w_i \tau^M_{i\to i+1} = -1$, and therefore, Assumption 6.1$_{153}$ implies that

$$\begin{aligned}\tau^M_{i\to i+1} &= \frac{1}{w_i} + \frac{b_i}{w_i}\tau^M_{i-1\to i} = \frac{1}{w_i} + \frac{b_i}{w_i}\sum_{k=0}^{i-1} \frac{\prod_{\ell=k+1}^{i-1} b_\ell}{\prod_{m=k}^{i-1} w_m}\\ &= \frac{1}{w_i} + \sum_{k=0}^{i-1} \frac{\prod_{\ell=k+1}^{i} b_\ell}{\prod_{m=k}^{i} w_m} = \sum_{k=0}^{i} \frac{\prod_{\ell=k+1}^{i} b_\ell}{\prod_{m=k}^{i} w_m},\end{aligned}$$

where the second equality follows from the induction hypothesis. □

Based on this result, it is now easy to obtain expressions for all the other expected upward first-passage times, because it follows from Corollary 62$_{164}$ and

Equation (6.46)$_{\curvearrowleft}$ that

$$\tau^M_{i\to j} = \sum_{k=i}^{j-1} \tau^M_{k\to k+1} \text{ for all } i,j \in \mathscr{X} \text{ such that } i<j. \tag{6.48}$$

Similarly, for expected downward first-passage times, it follows from Proposition 68$_{167}$ and Equation (6.46)$_{\curvearrowleft}$ that

$$\tau^M_{i\to j} = \sum_{k=j}^{i-1} \tau^M_{k+1\to k} \text{ for all } i,j \in \mathscr{X} \text{ such that } i>j, \tag{6.49}$$

where the individual terms in the summation are given by Proposition 71.

Proposition 71. *Consider a precise birth-death chain of which the transition matrix M is given by Equation* (6.1)$_{152}$*. Then for all $i \in \mathscr{X} \setminus \{0\}$, we have that*

$$\tau^M_{i\to i-1} = \sum_{k=i}^{L} \frac{\prod_{\ell=i}^{k-1} w_\ell}{\prod_{m=i}^{k} b_m}.$$

Proof. This proof is completely analogous to the proof of Proposition 70$_{\curvearrowleft}$. Again, we provide a proof by induction. Since $\overline{b}_L = b_L$, it follows from Equations (6.46)$_{\curvearrowleft}$ and (6.39)$_{165}$ that $\tau^M_{L\to L-1} = \underline{\tau}_{L\to L-1} = 1/\overline{b}_L = 1/b_L$, proving the result for $i=L$.

Consider now any $i \in \mathscr{X} \setminus \{0,L\}$ and let us assume, as our induction hypothesis, that the result is true for $i+1$. Since $\mathscr{R}_i = \{\pi_i\}$, it follows from Proposition 66$_{166}$ and Equation (6.46)$_{\curvearrowleft}$ that $-b_i\tau^M_{i\to i-1} + w_i\tau^M_{i+1\to i} = -1$, and therefore, Assumption 6.1$_{153}$ implies that

$$\begin{aligned}\tau^M_{i\to i-1} &= \frac{1}{b_i} + \frac{w_i}{b_i}\tau^M_{i+1\to i} = \frac{1}{b_i} + \frac{w_i}{b_i}\sum_{k=i+1}^{L}\frac{\prod_{\ell=i+1}^{k-1} w_\ell}{\prod_{m=i+1}^{k} b_m}\\ &= \frac{1}{b_i} + \sum_{k=i+1}^{L}\frac{\prod_{\ell=i}^{k-1} w_\ell}{\prod_{m=i}^{k} b_m} = \sum_{k=i}^{L}\frac{\prod_{\ell=i}^{k-1} w_\ell}{\prod_{m=i}^{k} b_m},\end{aligned}$$

where the second equality follows from the induction hypothesis. □

Closed-form expressions for expected return times can now be derived from Equations (6.40)$_{167}$–(6.45)$_{168}$, which, for precise birth-death chains, reduce to the following simple expressions:

$$\tau^M_{0\to 0} = 1 + w_0\tau^M_{1\to 0} \text{ and } \tau^M_{L\to L} = 1 + b_L\tau^M_{L-1\to L} \tag{6.50}$$

and

$$\tau^M_{i\to i} = 1 + b_i\tau^M_{i-1\to i} + w_i\tau^M_{i+1\to i} \text{ for all } i \in \mathscr{X}\setminus\{0,L\}. \tag{6.51}$$

6.7.2 *Connecting imprecise birth-death chains with precise ones.*

Birth-death chains are not just a special case of imprecise birth-death chains, they are also closely related to them in a more general way. In particular, for first-passage and return times, the lower and upper expectations of an imprecise birth-death chain are achieved by a (precise) birth-death chain.

Since we already know from Section 6.7.1$_{168}$ that the initial model $q_\square \in \mathscr{Q}_\square$ of birth-death chains does not influence their expected first-passage and return times, we will conveniently represent them by means of their transition matrix M. Depending on the type of bound that we are considering, a different type of transition matrix M will be needed to achieve the bound. We will specify the essential features of these different types by means of selection methods. For any given imprecise birth-death chain, such a selection method describes a specific way of choosing a transition matrix M.

For lower expected upward first-passage times, we use the following selection method.

Selection Method LU$_k$
Let M be any transition matrix of the form in Equation (6.1)$_{152}$ such that

1. if $k \neq 0$, then $w_0 = \overline{w}_0$;
2. for all $\ell \in \{1,\ldots,k-1\}$

$$(b_\ell, r_\ell, w_\ell) \in \operatorname*{argmin}_{\pi_\ell \in \mathscr{R}_\ell} \{b_\ell \underline{\tau}_{\ell-1\to\ell} - w_\ell \underline{\tau}_{\ell\to\ell+1}\}.$$

Indeed, as our next result establishes, for any given imprecise birth-death chain, its lower expected upward first-passage time can be obtained by a birth-death chain whose transition matrix M is selected according to the method above.

Theorem 72. *Consider an imprecise birth-death chain, some $k \in \mathscr{X}$, and a birth death chain whose transition matrix M is obtained from the imprecise birth-death chain by means of Selection Method LU$_k$. Then for all $i,j \in \mathscr{X}$ such that $i < j \leq k$, $\underline{\tau}_{i\to j} = \tau^M_{i\to j}$.*

Proof. Due to Corollary 62$_{164}$ and Equation (6.48)$_\curvearrowleft$, it clearly suffices to prove that

$$\underline{\tau}_{i\to i+1} = \tau^M_{i\to i+1} \text{ for all } i < k. \tag{6.52}$$

We provide a proof by induction. Since we know from Selection Method LU$_k$ 1 that $\overline{w}_0 = w_0$, it follows from Equation (6.29)$_{161}$ and Proposition 70$_{169}$ that $\underline{\tau}_{0\to 1} = 1/\overline{w}_0 = 1/w_0 = \tau^M_{0\to 1}$, which proves Equation (6.52) for $i = 0$.

Consider now any $i \in \{1,\ldots,k-1\}$ and let us assume, as our induction hypothesis, that Equation (6.52) is true for $i-1$, that is, $\underline{\tau}_{i-1\to i} = \tau^M_{i-1\to i}$. Then on the one hand, if we apply Proposition 60$_{163}$ to the imprecise birth-death chain, it follows from Selection Method LU$_k$ 2 that $b_i \underline{\tau}_{i-1\to i} - w_i \underline{\tau}_{i\to i+1} = -1$.

On the other hand, if we apply Proposition 60_{163} to the birth-death chain with transition matrix M, we find that $b_i\tau^M_{i-1\to i} - w_i\tau^M_{i\to i+1} = -1$. By combining these two statements with the induction hypothesis, it follows that $w_i\underline{\tau}_{i\to i+1} = w_i\tau^M_{i\to i+1}$, which, because of Assumption 6.1_{153}, implies that $\underline{\tau}_{i\to i+1} = \tau^M_{i\to i+1}$, as required. □

This result is at its most powerful if we choose $k = L$. In that case, it follows that all the lower expected upward first-passage times $\underline{\tau}_{i\to j}$, with $i, j \in \mathscr{X}$ such that $i < j$, can be obtained by the same birth-death chain.

Similar results also hold for upper expected upward first-passage times.

Selection Method UU$_k$
Let M be any transition matrix of the form in Equation $(6.1)_{152}$ such that

1. if $k \neq 0$, then $w_0 = \underline{w}_0$;
2. for all $\ell \in \{1,\dots,k-1\}$,

$$(b_\ell, r_\ell, w_\ell) \in \operatorname*{argmax}_{\pi_\ell \in \mathscr{R}_\ell}\{b_\ell\overline{\tau}_{\ell-1\to\ell} - w_\ell\overline{\tau}_{\ell\to\ell+1}\}.$$

Theorem 73. *Consider an imprecise birth-death chain, some $k \in \mathscr{X}$, and a birth death chain whose transition matrix M is obtained from the imprecise birth-death chain by means of Selection Method UU$_k$. Then for all $i, j \in \mathscr{X}$ such that $i < j \leq k$, $\overline{\tau}_{i\to j} = \tau^M_{i\to j}$.*

Proof. Due to Proposition 65_{165} and Equation $(6.48)_{170}$, it suffices to prove that

$$\overline{\tau}_{i\to i+1} = \tau^M_{i\to i+1} \text{ for all } i < k. \tag{6.53}$$

We provide a proof by induction. Since we know from Selection Method UU$_k$ 1 that $\underline{w}_0 = w_0$, it follows from Equation $(6.37)_{164}$ and Proposition 70_{169} that $\overline{\tau}_{0\to 1} = 1/\underline{w}_0 = 1/w_0 = \tau^M_{0\to 1}$, which proves Equation (6.53) for $i = 0$.

Consider now any $i \in \{1,\dots,k-1\}$ and let us assume, as our induction hypothesis, that Equation (6.53) is true for $i-1$, that is, $\overline{\tau}_{i-1\to i} = \tau^M_{i-1\to i}$. Then on the one hand, if we apply Proposition 63_{164} to the imprecise birth-death chain, it follows from Selection Method UU$_k$ 2↶ that $b_i\overline{\tau}_{i-1\to i} - w_i\overline{\tau}_{i\to i+1} = -1$. On the other hand, if we apply Proposition 63_{164} to the birth-death chain with transition matrix M, we find that $b_i\tau^M_{i-1\to i} - w_i\tau^M_{i\to i+1} = -1$. By combining these two statements with the induction hypothesis, it follows that $w_i\overline{\tau}_{i\to i+1} = w_i\tau^M_{i\to i+1}$, which, because of Assumption 6.1_{153}, implies that $\overline{\tau}_{i\to i+1} = \tau^M_{i\to i+1}$, as required. □

As before, this result is most powerful if we choose $k = L$, because it then implies that every upper expected upward first-passage time can be obtained by the same birth-death chain.

At first sight, it seems as though Theorems 72_{171} and $73_{\curvearrowleft}$ could provide us with a simple method for computing lower and upper expected upward first-passage times, thereby providing an alternative for the recursive equations in Section 6.4_{161}. All we have to do is (a) construct a transition matrix M according to an appropriate selection method and then (b) use this matrix M to apply the closed-form expressions in Section $6.7.1_{168}$. However, this method is not practical. The issue here is step (a). For example, executing Selection Method LU$_k$ 2_{171} is not just a matter of choosing $w_\ell = \overline{w}_\ell$ and $b_\ell = \underline{b}_\ell$, because, depending on the shape of $\mathscr{R}_\ell$, it may not be possible to attain these extrema simultaneously. Therefore, finding the optimal tuples (b_ℓ, r_ℓ, w_ℓ) requires us to know the value of $\underline{\tau}_{\ell\to\ell+1}$ for all $\ell \in \{0,\ldots,k-1\}$. However, in practice, we don't know these values yet. In fact, the whole point of computing lower expected upward first-passage times is to obtain these values. Therefore, Theorems 72_{171} and $73_{\curvearrowleft}$ should not be regarded as the basis of a computational method. Instead, their main importance is the theoretical insight that the lower and upper expected upward first-passage times that correspond to an imprecise birth death chain are achieved by (precise) birth-death chains.

Completely analogous conclusions can be drawn for lower and upper downward first-passage times, using the following selection methods and results.

Selection Method LD$_k$
Let M be any stochastic matrix of the form in Equation $(6.1)_{152}$ such that

1. if $k \neq L$, then $b_L = \overline{b}_L$;
2. for all $\ell \in \{k+1,\ldots,L-1\}$,

$$(b_\ell, r_\ell, w_\ell) \in \operatorname*{argmin}_{\pi_\ell \in \mathscr{R}_\ell} \{-b_\ell \underline{\tau}_{\ell\to\ell-1} + w_\ell \underline{\tau}_{\ell+1\to\ell}\}.$$

Theorem 74. *Consider an imprecise birth-death chain, some $k \in \mathscr{X}$, and a birth death chain whose transition matrix M is obtained from the imprecise birth-death chain by means of Selection Method LD$_k$. Then for all $i, j \in \mathscr{X}$ such that $k \leq j < i$, $\underline{\tau}_{i\to j} = \tau^M_{i\to j}$.*

Proof of Theorem 74. Due to Proposition 68_{167} and Equation $(6.49)_{170}$, it suffices to prove that

$$\underline{\tau}_{i\to i-1} = \tau^M_{i\to i-1} \text{ for all } i > k. \tag{6.54}$$

We provide a proof by induction. Since we know from Selection Method LD$_k$ 1 that $\overline{b}_L = b_L$, it follows from Equation $(6.39)_{165}$ and Proposition 71_{170} that $\underline{\tau}_{L\to L-1} = 1/\overline{b}_L = 1/b_L = \tau^M_{L\to L-1}$, which proves Equation (6.54) for $i = L$.

Consider now any $i \in \{k+1,\ldots,L-1\}$ and let us assume, as our induction hypothesis, that Equation (6.54) is true for $i+1$, that is, $\underline{\tau}_{i+1\to i} = \tau^M_{i+1\to i}$. Then on the one hand, if we apply Proposition 66_{166} to the imprecise birth-death chain, it follows from Selection Method LD$_k$ 2 that $-b_i \underline{\tau}_{i\to i-1} +$

$w_i \underline{\tau}_{i+1\to i} = -1$. On the other hand, if we apply Proposition 66_{166} to the birth-death chain with transition matrix M, we find that $-b_i \tau^M_{i\to i-1} + w_i \tau^M_{i+1\to i} = -1$. By combining these two statements with the induction hypothesis, it follows that $b_i \underline{\tau}_{i\to i-1} = b_i \tau^M_{i\to i-1}$, which, because of Assumption 6.1_{153}, implies that $\underline{\tau}_{i\to i-1} = \tau^M_{i\to i-1}$, as required. □

Selection Method UD$_k$
Let M be any transition matrix of the form in Equation $(6.1)_{152}$ such that

1. if $k \neq L$, then $b_L = \underline{b}_L$;
2. for all $\ell \in \{k+1, \ldots, L-1\}$,

$$(b_\ell, r_\ell, w_\ell) \in \operatorname*{argmax}_{\pi_\ell \in \mathscr{R}_\ell} \{-b_\ell \overline{\tau}_{\ell\to\ell-1} + w_\ell \overline{\tau}_{\ell+1\to\ell}\}.$$

Theorem 75. *Consider an imprecise birth-death chain, some $k \in \mathscr{X}$, and a birth death chain whose transition matrix M is obtained from the imprecise birth-death chain by means of Selection Method UD$_k$. Then for all $i, j \in \mathscr{X}$ such that $k \leq j < i$, $\overline{\tau}_{i\to j} = \tau^M_{i\to j}$.*

Proof of Theorem 75. Due to Proposition 68_{167} and Equation $(6.49)_{170}$, it suffices to prove that

$$\overline{\tau}_{i\to i-1} = \tau^M_{i\to i-1} \text{ for all } i > k. \tag{6.55}$$

We provide a proof by induction. Since we know from Selection Method UD$_k$ 1 that $\underline{b}_L = b_L$, it follows from Equation $(6.39)_{165}$ and Proposition 71_{170} that $\overline{\tau}_{L\to L-1} = 1/\underline{b}_L = 1/b_L = \tau^M_{L\to L-1}$, which proves Equation (6.55) for $i = L$.

Consider now any $i \in \{k+1, \ldots, L-1\}$ and let us assume, as our induction hypothesis, that Equation (6.55) is true for $i+1$, that is, $\overline{\tau}_{i+1\to i} = \tau^M_{i+1\to i}$. Then on the one hand, if we apply Proposition 66_{166} to the imprecise birth-death chain, it follows from Selection Method UD$_k$ 2 that $-b_i \overline{\tau}_{i\to i-1} + p_i \overline{\tau}_{i+1\to i} = -1$. On the other hand, if we apply Proposition 66_{166} to the birth-death chain with transition matrix M, we find that $-b_i \tau^M_{i\to i-1} + p_i \tau^M_{i+1\to i} = -1$. By combining these two statements with the induction hypothesis, it follows that $b_i \overline{\tau}_{i\to i-1} = b_i \tau^M_{i\to i-1}$, which, because of Assumption $(6.1)_{153}$, implies that $\overline{\tau}_{i\to i-1} = \tau^M_{i\to i-1}$, as required. □

These results are at their most powerful if we choose $k = 0$. In that case, all the lower expected downward first-passage times $\underline{\tau}_{i\to j}$, with $i, j \in \mathscr{X}$ such that $j < i$, can be obtained by the same birth-death chain, and similarly for the upper expected downward first-passage times.

This is not the case for lower and upper expected return times: there may not be a single birth-death chain for which all lower expected return times are obtained, nor is there guaranteed to be a single birth-death chain for which all the upper expected return times are obtained. Nevertheless, as we show in

Theorems 76 and 77↷ below, it is always possible to select one birth-death chain for each specific lower expected return time and one for each specific upper expected return time, using the following selection methods.

Selection Method LR$_k$
Let M be any transition matrix of the form in Equation (6.1)$_{152}$ such that

1. if $k \neq 0$, then $w_0 = \overline{w}_0$, and if $k = 0$, then $w_0 = \underline{w}_0$;
2. for all $\ell \in \{1, \ldots, k-1\}$,
$$(b_\ell, r_\ell, w_\ell) \in \operatorname*{argmin}_{\pi_\ell \in \mathscr{R}_\ell} \{b_\ell \underline{\tau}_{\ell-1\to\ell} - w_\ell \underline{\tau}_{\ell\to\ell+1}\}.$$
3. if $k \neq 0$ and $k \neq L$, then
$$(b_k, r_k, w_k) \in \operatorname*{argmin}_{\pi_k \in \mathscr{R}_k} \{b_k \underline{\tau}_{k-1\to k} + w_k \underline{\tau}_{k+1\to k}\};$$
4. for all $\ell \in \{k+1, \ldots, L-1\}$,
$$(b_\ell, r_\ell, w_\ell) \in \operatorname*{argmin}_{\pi_\ell \in \mathscr{R}_\ell} \{-b_\ell \underline{\tau}_{\ell\to\ell-1} + w_\ell \underline{\tau}_{\ell+1\to\ell}\}.$$
5. if $k \neq L$, then $b_L = \overline{b}_L$, and if $k = L$, then $b_L = \underline{b}_L$.

Theorem 76. *Consider an imprecise birth-death chain, some $k \in \mathscr{X}$, and a birth death chain whose transition matrix M is obtained from the imprecise birth-death chain by means of Selection Method LR$_k$. Then $\underline{\tau}_{k\to k} = \tau^M_{k\to k}$.*

Proof. First observe that, since Selection Method LR$_k$ implies Selection Method LU$_k$ and LD$_k$, we can use Theorems 72$_{171}$ and 74$_{173}$ to find that

$$\left(k \neq 0 \Rightarrow \underline{\tau}_{k-1\to k} = \tau^M_{k-1\to k}\right) \text{ and } \left(k \neq L \Rightarrow \underline{\tau}_{k+1\to k} = \tau^M_{k+1\to k}\right) \tag{6.56}$$

We now prove the theorem for $k = 0$. Since we know from Selection Method LR$_k$ 1 that $\underline{w}_0 = w_0$, it follows from Equations (6.40)$_{167}$ and (6.56) that
$$\underline{\tau}_{0\to 0} = 1 + \underline{w}_0 \underline{\tau}_{1\to 0} = 1 + w_0 \underline{\tau}_{1\to 0} = 1 + w_0 \tau^M_{1\to 0}$$
and therefore, due to Equation (6.50)$_{170}$, we infer that $\underline{\tau}_{0\to 0} = \tau^M_{0\to 0}$.

The case $k = L$ is proved similarly. Since we know from Selection Method LR$_k$ 5 that $\underline{b}_L = b_L$, it follows from Equations (6.41)$_{167}$ and (6.56) that

$$\underline{\tau}_{L\to L} = 1 + \underline{b}_L \underline{\tau}_{L-1\to L} = 1 + b_L \underline{\tau}_{L-1\to L} = 1 + b_L \tau^M_{L-1\to L}$$

and therefore, due to Equation (6.50)$_{170}$, we find that $\underline{\tau}_{L\to L} = \tau^M_{L\to L}$.

It remains now to prove the theorem for the case $k \in \{1, \ldots, L-1\}$. Since we know from Selection Method LR$_k$ 3$_{\curvearrowleft}$ that

$$(b_k, r_k, w_k) \in \operatorname*{argmin}_{\pi_k \in \mathscr{R}_k} \{b_k \underline{\tau}_{k-1\to k} + w_k \underline{\tau}_{k+1\to k}\},$$

it follows from Equations (6.42)$_{167}$ and (6.56)$_{\curvearrowleft}$ that

$$\underline{\tau}_{k\to k} = 1 + q_k \underline{\tau}_{k-1\to k} + w_k \underline{\tau}_{k+1\to k} = 1 + b_k \tau^M_{k-1\to k} + w_k \tau^M_{k+1\to k}$$

and therefore, Equation (6.51)$_{170}$ implies that $\underline{\tau}_{k\to k} = \tau^M_{k\to k}$, as required. □

Selection Method UR$_k$
Let M be any transition matrix of the form in Equation (6.1)$_{152}$ such that

1. if $k \neq 0$, then $w_0 = \underline{w}_0$, and if $k = 0$, then $w_0 = \overline{w}_0$;
2. for all $\ell \in \{1, \ldots, k-1\}$,
$$(b_\ell, r_\ell, w_\ell) \in \operatorname*{argmax}_{\pi_\ell \in \mathscr{R}_\ell} \{b_\ell \overline{\tau}_{\ell-1\to\ell} - w_\ell \overline{\tau}_{\ell\to\ell+1}\}.$$
3. if $k \neq 0$ and $k \neq L$, then
$$(b_k, r_k, w_k) \in \operatorname*{argmax}_{\pi_k \in \mathscr{R}_k} \{b_k \overline{\tau}_{k-1\to k} + w_k \overline{\tau}_{k+1\to k}\};$$
4. for all $\ell \in \{k+1, \ldots, L-1\}$,
$$(b_\ell, r_\ell, w_\ell) \in \operatorname*{argmax}_{\pi_\ell \in \mathscr{R}_\ell} \{-b_\ell \overline{\tau}_{\ell\to\ell-1} + w_\ell \overline{\tau}_{\ell+1\to\ell}\}.$$
5. if $k \neq L$, then $b_L = \underline{b}_L$, and if $k = L$, then $b_L = \overline{b}_L$.

Theorem 77. *Consider an imprecise birth-death chain, some $k \in \mathscr{X}$, and a birth death chain whose transition matrix M is obtained from the imprecise birth-death chain by means of Selection Method UR$_k$. Then $\overline{\tau}_{k\to k} = \tau^M_{k\to k}$.*

Proof. First observe that, since Selection Method UR$_k$ implies Selection Method UU$_k$ and UD$_k$, we can use Theorems 73$_{172}$ and 75$_{174}$ to find that

$$\left(k \neq 0 \Rightarrow \overline{\tau}_{k-1\to k} = \tau^M_{k-1\to k}\right) \text{ and } \left(k \neq L \Rightarrow \overline{\tau}_{k+1\to k} = \tau^M_{k+1\to k}\right) \quad (6.57)$$

We now prove the theorem for $k = 0$. Since we know from Selection Method UR$_k$ 1 that $\overline{w}_0 = w_0$, it follows from Equations (6.43)$_{168}$ and (6.57) that

$$\overline{\tau}_{0\to 0} = 1 + \overline{w}_0 \overline{\tau}_{1\to 0} = 1 + w_0 \overline{\tau}_{1\to 0} = 1 + w_0 \tau^M_{1\to 0}$$

and therefore, due to Equation $(6.50)_{170}$, we infer that $\overline{\tau}_{0\to0} = \tau^M_{0\to0}$.

The case $k = L$ is proved similarly. Since we know from Selection Method UR_k $5_{\curvearrowleft}$ that $\overline{q}_L = q_L$, it follows from Equations $(6.44)_{168}$ and $(6.57)_{\curvearrowleft}$ that

$$\overline{\tau}_{L\to L} = 1 + \overline{b}_L \overline{\tau}_{L-1\to L} = 1 + b_L \overline{\tau}_{L-1\to L} = 1 + b_L \tau^M_{L-1\to L}$$

and therefore, due to Equation $(6.50)_{170}$, we find that $\overline{\tau}_{L\to L} = \tau^M_{L\to L}$.

It remains now to prove the theorem for the case $k \in \{1,\dots,L-1\}$. Since we know from Selection Method LR_k $3_{\curvearrowleft}$ that

$$(b_k, r_k, w_k) \in \operatorname*{argmax}_{\pi_k \in \mathscr{R}_k} \{b_k \overline{\tau}_{k-1\to k} + w_k \overline{\tau}_{k+1\to k}\},$$

it follows from Equations $(6.45)_{168}$ and $(6.57)_{\curvearrowleft}$ that

$$\overline{\tau}_{k\to k} = 1 + b_k \overline{\tau}_{k-1\to k} + w_k \overline{\tau}_{k+1\to k} = 1 + b_k \tau^M_{k-1\to k} + w_k \tau^M_{k+1\to k}$$

and therefore, Equation $(6.51)_{170}$ implies that $\overline{\tau}_{k\to k} = \tau^M_{k\to k}$, as required. □

6.8 CONNECTION WITH MEASURE-THEORETIC APPROACH

The lower and upper expected first-passage and return times that we have discussed, so far, are defined by the martingale-theoretic approach. We now study the case where lower and upper expected first-passage and return times are defined by the measure-theoretic approach, and establish the connection with the corresponding ones defined by the martingale-theoretic approach.

Since an imprecise birth-death chain is a special case of a homogeneous imprecise Markov chain, it is associated with an imprecise probability tree $\mathscr{T}_{\mathscr{Q}}$, $\mathscr{T}^{\text{M}}_{\mathscr{Q}}$ or $\mathscr{T}^{\text{HM}}_{\mathscr{Q}}$, depending on the chosen independence concept. In this particular case, $\mathscr{Q}$ is now the collection of local models of the imprecise birth-death chain that was introduced in Section 6.2_{152}, i.e. $\mathscr{Q}_{\square}$ and, for all $i \in \mathscr{X}$, $\mathscr{Q}_i$. Of course, any (precise) birth-death chain for which the initial model $q_{\square}$ belongs to $\mathscr{Q}_{\square}$ and each of the transition models belongs to $\mathscr{Q}_i$ is an element of the imprecise probability tree $\mathscr{T}^{\text{HM}}_{\mathscr{Q}}$, and therefore also of $\mathscr{T}^{\text{M}}_{\mathscr{Q}}$ and $\mathscr{T}_{\mathscr{Q}}$. However, $\mathscr{T}^{\text{M}}_{\mathscr{Q}}$ and $\mathscr{T}_{\mathscr{Q}}$ also contain other probability trees, which do not necessarily correspond to a (homogeneous) birth-death chain.

As we will show here, for the purposes of computing lower and upper expected first-passage or return times, the probability trees in $\mathscr{T}_{\mathscr{Q}}$ and $\mathscr{T}^{\text{M}}_{\mathscr{Q}}$ that do not correspond to a (homogeneous) birth-death chain are not essential, because for first-passage and return times, their lower and upper expectations are achieved by the birth-death chains in $\mathscr{T}^{\text{HM}}_{\mathscr{Q}}$.

Since the measure-theoretic approach as we have described it applies only to measurable functions, we first need to make sure that first-passage times are measurable. As the following result shows, any first-passage and return time is measurable, and is moreover the limit of a non-decreasing sequence of non-negative n-measurable functions.

Lemma 78. *Consider any $m \in \mathbb{N}$ and any $i, j \in \mathscr{X}$. Then $\tau^m_{i\to j}$ is measurable and moreover, there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\tau^m_{i\to j} = \lim_{n\to+\infty} h_n$.*

Proof. It suffices to prove that there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\tau^m_{i\to j} = \lim_{n\to+\infty} h_n$ and it will then follow from Lemma 16_{73} and Theorem 98_{238} that $\tau^m_{i\to j}$ is also measurable.

For all $\ell \in \{1,\dots,m\}$, let h_ℓ be defined by $h_\ell(\omega) := 0$ for all $\omega \in \Omega$. For all $n \in \mathbb{N}\setminus\{1,\dots,m\}$ and all $\omega \in \Omega$, let h_n be defined by

$$h_n(\omega) := \begin{cases} 0 & \text{if } \omega_m \neq i; \\ n-m & \text{if } \omega_m = i \text{ and } (\forall k \in \{m+1,\dots,n\})\omega_k \neq j; \\ k^*(\omega,n) & \text{if } \omega_m = i \text{ and } (\exists k \in \{m+1,\dots,n\})\omega_k = j, \end{cases} \tag{6.58}$$

where $k^*(\omega,n) := \min\{k \in \{m+1,\dots,n\} \colon \omega_k = j\} - m$. Clearly, h_n is n-measurable and non-negative for all $n \in \mathbb{N}$, and the sequence $\{h_n\}_{n\in\mathbb{N}}$ is non-decreasing.

For all $\omega \in \Omega$ with $\omega_m \neq i$, it follows from Equations $(6.10)_{155}$ and (6.58) that $\tau^m_{i\to j}(\omega) = h_n(\omega) = 0$ for all $n \in \mathbb{N}$, which implies that

$$\tau^m_{i\to j}(\omega) = \lim_{n\to+\infty} h_n(\omega) = 0. \tag{6.59}$$

For all $\omega \in \Omega$ with $\omega_m = i$, we have that either j is reached at some time $k \in \mathbb{N}\setminus\{1,\dots,m\}$ or j is never reached. In the first case, it follows from Equations $(6.10)_{155}$ and (6.58) that $\tau^m_{i\to j}(\omega) = h_n(\omega) = k-m$ for all $n \in \mathbb{N}\setminus\{1,\dots,k-1\}$, which implies that

$$\tau^m_{i\to j}(\omega) = \lim_{n\to+\infty} h_n(\omega) = k-m. \tag{6.60}$$

In case j is never reached, it follows from Equation $(6.10)_{155}$ that $\tau^m_{i\to j}(\omega) = +\infty$ and from Equation (6.58) that $h_n(\omega) = n-m$ for all $n \in \mathbb{N}\setminus\{1,\dots,m\}$, which implies that $\lim_{n\to+\infty} h_n(\omega) = +\infty$ and therefore that

$$\tau^m_{i\to j}(\omega) = \lim_{n\to+\infty} h_n(\omega) = +\infty. \tag{6.61}$$

It now follows from Equations (6.59), (6.60) and (6.61) that indeed

$$\tau^m_{i\to j}(\omega) = \lim_{n\to+\infty} h_n(\omega) \text{ for all } \omega \in \Omega. \qquad \square$$

Since first-passage and return times are measurable, lower and upper expected first-passage and return times in imprecise birth-death chains can now be defined straightforwardly by the measure-theoretic approach and we can adopt different independence concepts. For all $n \in \mathbb{N}$ and all $i, j \in \mathscr{X}$, the lower and upper expectations under epistemic irrelevance $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)$ and $\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)$ are defined by Equations $(3.26)_{80}$ and $(3.27)_{80}$, for $\mathbb{P}_{\mathscr{T}} = \mathbb{P}_{\mathscr{Q}}$ and with

$$(X_n = i) = \bigcup_{x_{1:n-1}\in\mathscr{X}^{n-1}} \Gamma(x_{1:n-1}, i).$$

Similarly, lower and upper expected first-passage and return times under complete independence are denoted by $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)$ and $\overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)$ and defined by Equations $(5.59)_{132}$ and $(5.60)_{132}$, and the respective ones under repetition independence are denoted by $\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)$ and $\overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)$ and defined by Equations $(5.74)_{142}$ and $(5.75)_{142}$.

Next, we investigate the connection between the measure-theoretic approach and the martingale-theoretic approach in birth-death chains. In the beginning of Section $6.7.1_{168}$, we mentioned that a birth-death chain can be regarded as a imprecise birth-death chain whose local models, that is $\mathscr{Q}_\square$ and $\mathscr{Q}_i$ for all $i \in \mathscr{X}$, are singletons. Besides the fact that in this case the local models $\mathscr{Q}_i$ can be represented by a single transition matrix M, it also follows that $\mathscr{T}_{\mathscr{Q}}$ now consists of a single unique probability tree $p \in \mathscr{T}^{\mathrm{HM}}_{\mathscr{Q}}$. Since $\mathscr{T}^{\mathrm{HM}}_{\mathscr{Q}} \subseteq \mathscr{T}^{\mathrm{M}}_{\mathscr{Q}} \subseteq \mathscr{T}_{\mathscr{Q}}$, this implies that $\mathscr{T}^{\mathrm{HM}}_{\mathscr{Q}} = \mathscr{T}^{\mathrm{M}}_{\mathscr{Q}} = \mathscr{T}_{\mathscr{Q}} = \{p\}$. As the following theorem shows, for any $P \in \mathbb{P}_p$, the expected first-passage or return time $E_P(\tau^n_{i\to j}|X_n = i)$ coincides with the respective one defined by the martingale-theoretic approach, that is, $\tau^M_{i\to j}$.

Theorem 79. *Consider a birth-death chain with unique probability tree p, of which the transition matrix M is given by Equation* $(6.1)_{152}$. *Consider as well any $n \in \mathbb{N}$ and any $P \in \mathbb{P}_p$. For all $i, j \in \mathscr{X}$, it then holds that*

$$\tau^M_{i\to j} = E_P(\tau^n_{i\to j}|X_n = i).$$

Proof. Consider any $x_{1:n-1} \in \mathscr{X}^{n-1}$ and observe that

$$\begin{aligned}\tau^M_{i\to j} &= \underline{\tau}_{i\to j} = \underline{E}^{\mathrm{ei}}_{Q}(\tau^n_{i\to j}|x_{1:n-1}, i) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|x_{1:n-1}, i) = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)\\ &\leq E_P(\tau^n_{i\to j}|X_n = i) \leq \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|x_{1:n-1}, i)\\ &\leq \overline{E}^{\mathrm{ei}}_{Q}(\tau^n_{i\to j}|x_{1:n-1}, i) = \overline{\tau}_{i\to j} = \tau^M_{i\to j},\end{aligned}$$

where the first and the last equality follow from Equation $(6.46)_{169}$, the second equality follows from Equations $(6.12)_{156}$ and $(6.11)_{155}$, the third and the fourth equality follow from Theorem 44_{123} combined with Lemma $78_{\curvearrowleft}$, the fifth equality follows from Equation $(6.13)_{156}$ and the upper expectation version of Equation $(6.11)_{155}$, and finally the first and last inequality follow from Theorem 33_{98} combined with Lemma $78_{\curvearrowleft}$. Since we have that $\tau^M_{i\to j} \leq E_P(\tau^n_{i\to j}|X_n = i) \leq \tau^M_{i\to j}$, we also infer that $\tau^M_{i\to j} = E_P(\tau^n_{i\to j}|X_n = i)$. □

We are now ready to present the main result of this section, which is a generalisation of Theorem 79. In particular, this theorem says that lower and upper expected first-passage and return times defined by the martingale-theoretic approach coincide with the corresponding ones defined by the measure-theoretic approach, under any of the independence concepts we have been considering.

Theorem 80. *Consider an imprecise birth-death chain, any $n \in \mathbb{N}$ and any $i, j \in \mathscr{X}$. Then*

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \underline{\tau}_{i\to j};$$
$$\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \overline{\tau}_{i\to j}.$$

Proof. We will only provide the proof for the lower expectations; the proof for the upper ones is completely analogous.

Let M be the transition matrix obtained from the imprecise birth-death chain by means of Selection Method LU$_k$ or Selection Method LD$_k$ or Selection Method LR$_k$ depending on whether $i < j$ or $i > j$ or $i = j$. Then it follows from Theorems 72_{171}, 74_{173} and 76_{175} that

$$\underline{\tau}_{i\to j} = \tau^M_{i\to j}. \tag{6.62}$$

Since M is constructed from probability mass functions in the sets $\mathscr{R}_x$, and consequently in local models $\mathscr{Q}_x$, for $x \in \mathscr{X}$, it follows from Equation $(5.73)_{142}$ that there is some $p \in \mathscr{T}^{\mathrm{HM}}_{\mathscr{Q}}$ such that $p(y|s,x) = M(x,y)$ for all $x, y \in \mathscr{X}$ and all $s \in \mathscr{X}^*$, where $M(x,y)$ is the element of M at row x and column y. Consider any such probability tree p, then there is a birth-death chain with unique probability tree p, of which the transition matrix is M. It then follows from Theorem $79_{\curvearrowleft}$ that for any $n \in \mathbb{N}$ and any $P \in \mathbb{P}_p$, it holds that $\tau^M_{i\to j} = E_P(\tau^n_{i\to j}|X_n = i)$, and therefore, due to Equation (6.62), that

$$\underline{\tau}_{i\to j} = E_P(\tau^n_{i\to j}|X_n = i). \tag{6.63}$$

Since $p \in \mathscr{T}^{\mathrm{HM}}_{\mathscr{Q}}$ and $P \in \mathbb{P}_p$, we infer that $P \in \mathbb{P}^{\mathrm{HM}}_{\mathscr{Q}}$. Hence, it follows from Equation $(5.74)_{142}$ that $E_P(\tau^n_{i\to j}|X_n = i) \geq \underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)$, and due to Equation (6.63), we find that

$$\underline{\tau}_{i\to j} \geq \underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i). \tag{6.64}$$

Consider now any $x_{1:n-1} \in \mathscr{X}^{n-1}$. Then

$$\begin{aligned}\underline{\tau}_{i\to j} &= \underline{E}^{\mathrm{ei}}_{Q}(\tau^n_{i\to j}|x_{1:n-1}, i) \leq \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|x_{1:n-1}, i)\\ &= \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) \leq \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) \leq \underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i),\end{aligned} \tag{6.65}$$

where the first equality follows from Equations $(6.12)_{156}$ and $(6.11)_{155}$, the second equality follows from Theorem 44_{123} combined with Lemma 78_{178}, the first inequality follows from Theorem 33_{98} combined with Lemma 78_{178}

and the last two inequalities follow from Lemma 56$_{147}$, again combined with Lemma 78$_{178}$.

By combining Equation (6.64)$_{\curvearrowleft}$ with Equation (6.65)$_{\curvearrowleft}$, it follows that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n=i)=\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n=i)=\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n=i)=\underline{\tau}_{i\to j}. \qquad \square$$

In Chapter 7$_{189}$, we will introduce an additional independence concept and we will see that Theorem 80$_{\curvearrowleft}$ applies to this concept also for lower and upper expected first-passage times but not always for lower and upper expected return times.

6.A USEFUL PROPERTIES

In this appendix, we have gathered some useful properties that allow us to prove more elegantly some of the results stated in Sections 6.4$_{161}$ and 6.5$_{165}$.

Lemma 81. *For all $k\in\mathscr{X}\setminus\{0,L\}$, we have that*

$$\min_{\pi_k\in\mathscr{R}_k}\{b_k\underline{\tau}_{k-1\to k}-w_k\underline{\tau}_{k\to k+1}\}=-1,$$

and that, for all ℓ in $\mathscr{X}$ such that $\ell>k$:

$$\underline{\tau}_{k-1\to\ell}=\underline{\tau}_{k-1\to k}+\underline{\tau}_{k\to\ell}.$$

Proof. For $k=1$, we have proved in the main text that the lemma holds; see Equations (6.32)$_{162}$ and (6.33)$_{162}$. We will now generalise it using induction. Assuming that the lemma is true for $k-1$, with $k\in\mathscr{X}\setminus\{0,1,L\}$, we prove that it is also true for k.

Consider any $\ell>k$. By taking Equation (6.27)$_{160}$, for $i=k-1$ and $j=\ell$, we find that

$$\begin{aligned}\underline{\tau}_{k-1\to\ell}&=1+\min_{\pi_{k-1}\in\mathscr{R}_{k-1}}\{b_{k-1}\underline{\tau}_{k-2\to\ell}+r_{k-1}\underline{\tau}_{k-1\to\ell}+w_{k-1}\underline{\tau}_{k\to\ell}\}\\&=1+\min_{\pi_{k-1}\in\mathscr{R}_{k-1}}\{b_{k-1}\underline{\tau}_{k-2\to\ell}+(1-b_{k-1}-w_{k-1})\underline{\tau}_{k-1\to\ell}+w_{k-1}\underline{\tau}_{k\to\ell}\}\\&=1+\underline{\tau}_{k-1\to\ell}+\min_{\pi_{k-1}\in\mathscr{R}_{k-1}}\{b_{k-1}(\underline{\tau}_{k-2\to\ell}-\underline{\tau}_{k-1\to\ell})-w_{k-1}(\underline{\tau}_{k-1\to\ell}-\underline{\tau}_{k\to\ell})\},\end{aligned}$$

which, due to Theorem 58$_{156}$, implies that

$$\min_{\pi_{k-1}\in\mathscr{R}_{k-1}}\{b_{k-1}(\underline{\tau}_{k-2\to\ell}-\underline{\tau}_{k-1\to\ell})-w_{k-1}(\underline{\tau}_{k-1\to\ell}-\underline{\tau}_{k\to\ell})\}=-1.$$

In combination with the induction hypothesis, which implies that $\underline{\tau}_{k-2\to\ell}=\underline{\tau}_{k-2\to k-1}+\underline{\tau}_{k-1\to\ell}$, the equation above results in

$$\min_{\pi_{k-1}\in\mathscr{R}_{k-1}}\{b_{k-1}\underline{\tau}_{k-2\to k-1}-w_{k-1}(\underline{\tau}_{k-1\to\ell}-\underline{\tau}_{k\to\ell})\}=-1. \tag{6.66}$$

Due to Proposition 59_{162} and the induction hypothesis, which implies that

$$\min_{\pi_{k-1}\in\mathscr{R}_{k-1}} \{b_{k-1}\underline{\tau}_{k-2\to k-1} - w_{k-1}\underline{\tau}_{k-1\to k}\} = -1.$$

we infer from Equation $(6.66)_{\curvearrowleft}$ that $\underline{\tau}_{k-1\to\ell} - \underline{\tau}_{k\to\ell} = \underline{\tau}_{k-1\to k}$, and therefore that

$$\underline{\tau}_{k-1\to\ell} = \underline{\tau}_{k-1\to k} + \underline{\tau}_{k\to\ell} \tag{6.67}$$

By taking now Equation $(6.27)_{160}$, for $i=k$ and $j=k+1$, we find that

$$\begin{aligned}\underline{\tau}_{k\to k+1} &= 1 + \min_{\pi_k\in\mathscr{R}_k} \{b_k\underline{\tau}_{k-1\to k+1} + r_k\underline{\tau}_{k\to k+1}\} \\ &= 1 + \min_{\pi_k\in\mathscr{R}_k} \{b_k\underline{\tau}_{k-1\to k+1} + (1-b_k-w_k)\underline{\tau}_{k\to k+1}\} \\ &= 1 + \underline{\tau}_{k\to k+1} + \min_{\pi_k\in\mathscr{R}_k} \{b_k(\underline{\tau}_{k-1\to k+1} - \underline{\tau}_{k\to k+1}) - w_k\underline{\tau}_{k\to k+1}\},\end{aligned}$$

which, due to Theorem 58_{156}, implies that

$$\min_{\pi_k\in\mathscr{R}_k} \{b_k(\underline{\tau}_{k-1\to k+1} - \underline{\tau}_{k\to k+1}) - w_k\underline{\tau}_{k\to k+1}\} = -1.$$

By combining this with Equation (6.67), for $\ell = k+1$, we find that

$$\min_{\pi_k\in\mathscr{R}_k} \{b_k\underline{\tau}_{k-1\to k} - w_k\underline{\tau}_{k\to k+1}\} = -1. \qquad \square$$

Lemma 82. *For all $k \in \mathscr{X}\setminus\{0,L\}$, we have that*

$$\max_{\pi_k\in\mathscr{R}_k} \{b_k\overline{\tau}_{k-1\to k} - w_k\overline{\tau}_{k\to k+1}\} = -1,$$

and that, for all ℓ in $\mathscr{X}$ such that $\ell > k$:

$$\overline{\tau}_{k-1\to\ell} = \overline{\tau}_{k-1\to k} + \overline{\tau}_{k\to\ell}.$$

Proof. For any $k \in \mathscr{X}\setminus\{0,L\}$ and $\ell \in \mathscr{X}$ such that $\ell > k$, applying Equation $(6.28)_{160}$ for $i=k$ and $j=\ell$ yields

$$\begin{aligned}\overline{\tau}_{k\to\ell} &= 1 + \overline{Q}\left(\sum_{z\in\mathscr{X}\setminus\{\ell\}} \mathbb{I}_z\overline{\tau}_{z\to\ell}\middle| k\right) \\ &= 1 + \max_{\pi_k\in\mathscr{R}_k} \{b_k\mathbb{I}_{\neg\ell}(k-1)\overline{\tau}_{k-1\to\ell} + r_k\mathbb{I}_{\neg\ell}(k)\overline{\tau}_{k\to\ell} + w_k\mathbb{I}_{\neg\ell}(k+1)\overline{\tau}_{k+1\to\ell}\}.\end{aligned} \tag{6.68}$$

We now first prove the case $k=1$. By applying Equation $(6.28)_{160}$ for $i=0$, we find that

$$\begin{aligned}\overline{\tau}_{0\to j} &= 1 + \overline{Q}\left(\sum_{z\in\mathscr{X}\setminus\{\ell\}} \mathbb{I}_z\overline{\tau}_{z\to\ell}\middle| 0\right) \\ &= 1 + \max_{\pi_0\in\mathscr{R}_0} \left\{r_0\mathbb{I}_{\neg j}(0)\overline{\tau}_{0\to j} + w_0\mathbb{I}_{\neg j}(1)\overline{\tau}_{1\to j}\right\}.\end{aligned} \tag{6.69}$$

Consider any $\ell \in \mathscr{X}$ such that $\ell > 1$. By applying Equation $(6.69)_{\curvearrowleft}$ for $j = \ell$, we then find that

$$\begin{aligned}\overline{\tau}_{0\to\ell} &= 1 + \max_{\pi_0 \in \mathscr{R}_0} \left\{ r_0 \overline{\tau}_{0\to\ell} + w_0 \overline{\tau}_{1\to\ell} \right\} = 1 + \max_{\pi_0 \in \mathscr{R}_0} \left\{ (1 - w_0)\overline{\tau}_{0\to\ell} + w_0 \overline{\tau}_{1\to\ell} \right\} \\ &= 1 + \overline{\tau}_{0\to\ell} + \max_{\pi_0 \in \mathscr{R}_0} \left\{ -w_0(\overline{\tau}_{0\to\ell} - \overline{\tau}_{1\to\ell}) \right\},\end{aligned}$$

which, due to Theorem 58_{156}, implies that

$$\max_{\pi_0 \in \mathscr{R}_0} \left\{ -w_0(\overline{\tau}_{0\to\ell} - \overline{\tau}_{1\to\ell}) \right\} = -1 \Rightarrow \overline{\tau}_{0\to\ell} = \frac{1}{\underline{w}_0} + \overline{\tau}_{1\to\ell}.$$

By combining this with Equation $(6.37)_{164}$, we find that

$$\overline{\tau}_{0\to\ell} = \overline{\tau}_{0\to 1} + \overline{\tau}_{1\to\ell}. \tag{6.70}$$

By taking Equation $(6.68)_{\curvearrowleft}$, for $k = 1$ and $\ell = 2$, we find that

$$\begin{aligned}\overline{\tau}_{1\to 2} &= 1 + \max_{\pi_1 \in \mathscr{R}_1} \{ b_1 \overline{\tau}_{0\to 2} + r_1 \overline{\tau}_{1\to 2} \} \\ &= 1 + \max_{\pi_1 \in \mathscr{R}_1} \{ b_1 \overline{\tau}_{0\to 2} + (1 - b_1 - w_1)\overline{\tau}_{1\to 2} \} \\ &= 1 + \overline{\tau}_{1\to 2} + \max_{\pi_1 \in \mathscr{R}_1} \{ b_1(\overline{\tau}_{0\to 2} - \overline{\tau}_{1\to 2}) - w_1 \overline{\tau}_{1\to 2} \}\end{aligned}$$

which, due to Theorem 58_{156}, implies that

$$\max_{\pi_1 \in \mathscr{R}_1} \{ b_1(\overline{\tau}_{0\to 2} - \overline{\tau}_{1\to 2}) - w_1 \overline{\tau}_{1\to 2} \} = -1.$$

By combining this with Equation (6.70), for $\ell = 2$, we find that

$$\max_{\pi_1 \in \mathscr{R}_1} \{ b_1 \overline{\tau}_{0\to 1} - w_1 \overline{\tau}_{1\to 2} \} = -1.$$

We will now generalise our proof using induction. Assuming that the lemma is true for $k-1$, with $k \in \mathscr{X} \setminus \{0, 1, L\}$, we prove that it is also true for k. Consider any $\ell > k$. It follows from Equation $(6.68)_{\curvearrowleft}$ that

$$\begin{aligned}\overline{\tau}_{k-1\to\ell} &= 1 + \max_{\pi_{k-1} \in \mathscr{R}_{k-1}} \{ b_{k-1} \overline{\tau}_{k-2\to\ell} + r_{k-1} \overline{\tau}_{k-1\to\ell} + w_{k-1} \overline{\tau}_{k\to\ell} \} \\ &= 1 + \max_{\pi_{k-1} \in \mathscr{R}_{k-1}} \{ b_{k-1} \overline{\tau}_{k-2\to\ell} + (1 - b_{k-1} - w_{k-1})\overline{\tau}_{k-1\to j} + w_{k-1} \overline{\tau}_{k\to\ell} \} \\ &= 1 + \overline{\tau}_{k-1\to\ell} + \max_{\pi_{k-1} \in \mathscr{R}_{k-1}} \{ b_{k-1}(\overline{\tau}_{k-2\to\ell} - \overline{\tau}_{k-1\to\ell}) - w_{k-1}(\overline{\tau}_{k-1\to\ell} - \overline{\tau}_{k\to\ell}) \},\end{aligned}$$

which, due to Theorem 58_{156}, implies that

$$\max_{\pi_{k-1} \in \mathscr{R}_{k-1}} \{ b_{k-1}(\overline{\tau}_{k-2\to\ell} - \overline{\tau}_{k-1\to\ell}) - w_{k-1}(\overline{\tau}_{k-1\to\ell} - \overline{\tau}_{k\to\ell}) \} = -1.$$

In combination with the induction hypothesis, which implies that $\overline{\tau}_{k-2\to\ell} = \overline{\tau}_{k-2\to k-1} + \overline{\tau}_{k-1\to\ell}$, the equation above results in

$$\max_{\pi_{k-1}\in\mathscr{R}_{k-1}} \{b_{k-1}\overline{\tau}_{k-2\to k-1} - w_{k-1}(\overline{\tau}_{k-1\to\ell} - \overline{\tau}_{k\to\ell})\} = -1. \tag{6.71}$$

Due to Proposition 64$_{164}$ and the induction hypothesis, which implies that

$$\max_{\pi_{k-1}\in\mathscr{R}_{k-1}} \{b_{k-1}\overline{\tau}_{k-2\to k-1} - w_{k-1}\overline{\tau}_{k-1\to k}\} = -1,$$

we infer from Equation (6.71) that $\overline{\tau}_{k-1\to\ell} - \overline{\tau}_{k\to\ell} = \overline{\tau}_{k-1\to k}$, and therefore that

$$\overline{\tau}_{k-1\to\ell} = \overline{\tau}_{k-1\to k} + \overline{\tau}_{k\to\ell} \tag{6.72}$$

By taking now Equation (6.68)$_{182}$, for $\ell = k+1$, we find that

$$\begin{aligned}
\overline{\tau}_{k\to k+1} &= 1 + \max_{\pi_k\in\mathscr{R}_k} \{b_k\overline{\tau}_{k-1\to k+1} + r_k\overline{\tau}_{k\to k+1}\} \\
&= 1 + \max_{\pi_k\in\mathscr{R}_k} \{b_k\overline{\tau}_{k-1\to k+1} + (1-b_k-w_k)\overline{\tau}_{k\to k+1}\} \\
&= 1 + \overline{\tau}_{k\to k+1} + \max_{\pi_k\in\mathscr{R}_k} \{b_k(\overline{\tau}_{k-1\to k+1} - \overline{\tau}_{k\to k+1}) - w_k\overline{\tau}_{k\to k+1}\},
\end{aligned}$$

which, due to Theorem 58$_{156}$, implies that

$$\max_{\pi_k\in\mathscr{R}_k} \{b_k(\overline{\tau}_{k-1\to k+1} - \overline{\tau}_{k\to k+1}) - w_k\overline{\tau}_{k\to k+1}\} = -1.$$

By combining this with Equation (6.72), for $\ell = k+1$, we find that

$$\max_{\pi_k\in\mathscr{R}_k} \{b_k\overline{\tau}_{k-1\to k} - w_k\overline{\tau}_{k\to k+1}\} = -1. \qquad \square$$

Lemma 83. *For all $k \in \mathscr{X}\setminus\{0,L\}$, we have that*

$$\min_{\pi_k\in\mathscr{R}_k} \{-b_k\underline{\tau}_{k\to k-1} + w_k\underline{\tau}_{k+1\to k}\} = -1,$$

and that, for all ℓ in $\mathscr{X}$ such that $k > \ell$:

$$\underline{\tau}_{k+1\to\ell} = \underline{\tau}_{k+1\to k} + \underline{\tau}_{k\to\ell}.$$

Proof. We first prove the case $k = L-1$. Consider any $\ell \in \mathscr{X}$ such that $\ell < L-1$. By taking Equation (6.26)$_{160}$, for $j = \ell$, we find that

$$\begin{aligned}
\underline{\tau}_{L\to\ell} &= 1 + \min_{\pi_L\in\mathscr{R}_L} \left\{b_L\underline{\tau}_{L-1\to\ell} + r_L\underline{\tau}_{L\to\ell}\right\} \\
&= 1 + \min_{\pi_L\in\mathscr{R}_L} \left\{b_L\underline{\tau}_{L-1\to\ell} + (1-b_L)\underline{\tau}_{L\to\ell}\right\} \\
&= 1 + \underline{\tau}_{L\to\ell} + \min_{\pi_L\in\mathscr{R}_L} \left\{-b_L(\underline{\tau}_{L\to\ell} - \underline{\tau}_{L-1\to\ell})\right\},
\end{aligned}$$

which, due to Theorem 58$_{156}$, implies that

$$\min_{\pi_L \in \mathscr{R}_L} \{ -b_L(\underline{\tau}_{L\to\ell} - \underline{\tau}_{L-1\to\ell})\} = -1 \Rightarrow \underline{\tau}_{L\to\ell} = \frac{1}{\overline{b}_L} + \underline{\tau}_{L-1\to\ell}.$$

By combining this with Equation (6.39)$_{165}$, we find that

$$\underline{\tau}_{L\to\ell} = \underline{\tau}_{L\to L-1} + \underline{\tau}_{L-1\to\ell}. \tag{6.73}$$

By applying Equation (6.27)$_{160}$ for $i = L-1$ and $j = L-2$, we find that

$$\begin{aligned}
\underline{\tau}_{L-1\to L-2} &= 1 + \min_{\pi_{L-1} \in \mathscr{R}_{L-1}} \{r_{L-1}\underline{\tau}_{L-1\to L-2} + w_{L-1}\underline{\tau}_{L\to L-2}\} \\
&= 1 + \min_{\pi_{L-1} \in \mathscr{R}_{L-1}} \{(1 - b_{L-1} - p_{L-1})\underline{\tau}_{L-1\to L-2} + w_{L-1}\underline{\tau}_{L\to L-2}\} \\
&= 1 + \underline{\tau}_{L-1\to L-2} + \min_{\pi_{L-1} \in \mathscr{R}_{L-1}} \{-b_{L-1}\underline{\tau}_{L-1\to L-2} \\
&\qquad\qquad + w_{L-1}(\underline{\tau}_{L\to L-2} - \underline{\tau}_{L-1\to L-2})\}
\end{aligned}$$

which, due to Theorem 58$_{156}$, implies that

$$\min_{\pi_{L-1} \in \mathscr{R}_{L-1}} \{-b_{L-1}\underline{\tau}_{L-1\to L-2} + w_{L-1}(\underline{\tau}_{L\to L-2} - \underline{\tau}_{L-1\to L-2})\} = -1.$$

Combining the equation above with Equation (6.73), for $\ell = L-2$, we find that

$$\min_{\pi_{L-1} \in \mathscr{R}_{L-1}} \{-b_{L-1}\underline{\tau}_{L-1\to L-2} + w_{L-1}\underline{\tau}_{L\to L-1}\} = -1,$$

We will now generalise the proof using induction. Assuming that the lemma is true for $k+1$, with $k \in \mathscr{X} \setminus \{0, L-1, L\}$, we prove that it is also true for k.

Consider any $\ell \in \mathscr{X}$ such that $\ell < k$. By applying Equation (6.27)$_{160}$ for $i = k+1$ and $j = \ell$, we find that

$$\begin{aligned}
\underline{\tau}_{k+1\to\ell} &= 1 + \min_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{b_{k+1}\underline{\tau}_{k\to\ell} + r_{k+1}\underline{\tau}_{k+1\to\ell} + w_{k+1}\underline{\tau}_{k+2\to\ell}\} \\
&= 1 + \min_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{b_{k+1}\underline{\tau}_{k\to\ell} + (1 - b_{k+1} - w_{k+1})\underline{\tau}_{k+1\to\ell} + w_{k+1}\underline{\tau}_{k+2\to\ell}\} \\
&= 1 + \underline{\tau}_{k+1\to\ell} + \min_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{-b_{k+1}(\underline{\tau}_{k+1\to\ell} - \underline{\tau}_{k\to\ell}) \\
&\qquad\qquad + w_{k+1}(\underline{\tau}_{k+2\to\ell} - \underline{\tau}_{k+1\to\ell})\},
\end{aligned}$$

which, due to Theorem 58$_{156}$, implies that

$$\min_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{-b_{k+1}(\underline{\tau}_{k+1\to\ell} - \underline{\tau}_{k\to\ell}) + w_{k+1}(\underline{\tau}_{k+2\to\ell} - \underline{\tau}_{k+1\to\ell})\} = -1.$$

In combination with the induction hypothesis, which implies that $\underline{\tau}_{k+2\to\ell} = \underline{\tau}_{k+2\to k+1} + \underline{\tau}_{k+1\to\ell}$, the equation above results in

$$\min_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{-b_{k+1}(\underline{\tau}_{k+1\to\ell} - \underline{\tau}_{k\to\ell}) + w_{k+1}\underline{\tau}_{k+2\to k+1}\} = -1. \tag{6.74}$$

Due to Proposition 67_{166} and the induction hypothesis, which implies that

$$\min_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{-b_{k+1}\underline{\tau}_{k+1\to k} + w_{k+1}\underline{\tau}_{k+2\to k+1}\} = -1,$$

we infer from Equation $(6.74)_{\curvearrowleft}$ that $\underline{\tau}_{k+1\to\ell} - \underline{\tau}_{k\to\ell} = \underline{\tau}_{k+1\to k}$, and therefore that

$$\underline{\tau}_{k+1\to\ell} = \underline{\tau}_{k+1\to k} + \underline{\tau}_{k\to\ell} \tag{6.75}$$

By taking now Equation $(6.27)_{160}$, for $i = k$ and $j = k-1$, we find that

$$\begin{aligned}
\underline{\tau}_{k\to k-1} &= 1 + \min_{\pi_k \in \mathscr{R}_k} \{r_k\underline{\tau}_{k\to k-1} + w_k\underline{\tau}_{k+1\to k-1}\} \\
&= 1 + \min_{\pi_k \in \mathscr{R}_k} \{(1-b_k-w_k)\underline{\tau}_{k\to k-1} + w_k\underline{\tau}_{k+1\to k-1}\} \\
&= 1 + \underline{\tau}_{k\to k-1} + \min_{\pi_k \in \mathscr{R}_k} \{-b_k\underline{\tau}_{k\to k-1} + w_k(\underline{\tau}_{k+1\to k-1} - \underline{\tau}_{k\to k-1})\},
\end{aligned}$$

which, due to Theorem 58_{156}, implies that

$$\min_{\pi_k \in \mathscr{R}_k} \{-b_k\underline{\tau}_{k\to k-1} + w_k(\underline{\tau}_{k+1\to k-1} - \underline{\tau}_{k\to k-1})\} = -1.$$

By combining this with Equation (6.75), for $\ell = k-1$, we find that

$$\min_{\pi_k \in \mathscr{R}_k} \{-b_k\underline{\tau}_{k-1\to k} + w_k\underline{\tau}_{k+1\to k}\} = -1. \qquad \square$$

Lemma 84. *For all $k \in \mathscr{X} \setminus \{0, L\}$, we have that*

$$\max_{\pi_k \in \mathscr{R}_k} \{-b_k\overline{\tau}_{k\to k-1} + w_k\overline{\tau}_{k+1\to k}\} = -1,$$

and that, for all ℓ in $\mathscr{X}$ such that $k > \ell$:

$$\overline{\tau}_{k+1\to\ell} = \overline{\tau}_{k+1\to k} + \overline{\tau}_{k\to\ell}.$$

Proof. We first prove the case $k = L-1$. By taking Equation $(6.28)_{160}$, for $i = L$, we find that

$$\begin{aligned}
\overline{\tau}_{L\to j} &= 1 + \overline{Q}\left(\sum_{z\in\mathscr{X}\setminus\{\ell\}} \mathbb{I}_z\overline{\tau}_{z\to\ell} \middle| L\right) \\
&= 1 + \max_{\pi_L \in \mathscr{R}_L} \left\{b_L\mathbb{I}_{\neg j}(L-1)\overline{\tau}_{L-1\to j} + r_L\mathbb{I}_{\neg j}(L)\overline{\tau}_{L\to j}\right\}.
\end{aligned} \tag{6.76}$$

Consider any $\ell \in \mathscr{X}$ such that $\ell < L-1$. By applying Equation (6.76) for $j = \ell$, we find that

$$\begin{aligned}
\overline{\tau}_{L\to\ell} &= 1 + \max_{\pi_L \in \mathscr{R}_L} \left\{b_L\overline{\tau}_{L-1\to\ell} + r_L\overline{\tau}_{L\to\ell}\right\} \\
&= 1 + \max_{\pi_L \in \mathscr{R}_L} \left\{b_L\overline{\tau}_{L-1\to\ell} + (1-b_L)\overline{\tau}_{L\to\ell}\right\} \\
&= 1 + \overline{\tau}_{L\to\ell} + \max_{\pi_L \in \mathscr{R}_L} \left\{-b_L(\overline{\tau}_{L\to\ell} - \overline{\tau}_{L-1\to\ell})\right\},
\end{aligned}$$

which, due to Theorem 58_{156}, implies that

$$\max_{\pi_L \in \mathscr{R}_L} \{ -b_L(\overline{\tau}_{L\to\ell} - \overline{\tau}_{L-1\to\ell})\} = -1 \Rightarrow \overline{\tau}_{L\to\ell} = \frac{1}{\underline{b}_L} + \overline{\tau}_{L-1\to\ell}.$$

By combining this with Equation $(6.39)_{165}$, we find that

$$\overline{\tau}_{L\to\ell} = \overline{\tau}_{L\to L-1} + \overline{\tau}_{L-1\to\ell}. \tag{6.77}$$

By taking Equation $(6.68)_{182}$, for $k = L-1$ and $\ell = L-2$, we find that

$$\begin{aligned}
&\overline{\tau}_{L-1\to L-2} = 1 + \max_{\pi_{L-1} \in \mathscr{R}_{L-1}} \{r_{L-1}\overline{\tau}_{L-1\to L-2} + w_{L-1}\overline{\tau}_{L\to L-2}\} \\
&= 1 + \max_{\pi_{L-1} \in \mathscr{R}_{L-1}} \{(1 - b_{L-1} - w_{L-1})\overline{\tau}_{L-1\to L-2} + w_{L-1}\overline{\tau}_{L\to L-2}\} \\
&= 1 + \overline{\tau}_{L-1\to L-2} + \max_{\pi_{L-1} \in \mathscr{R}_{L-1}} \{-b_{L-1}\overline{\tau}_{L-1\to L-2} \\
&\qquad\qquad + w_{L-1}(\overline{\tau}_{L\to L-2} - \overline{\tau}_{L-1\to L-2})\}
\end{aligned}$$

which, due to Theorem 58_{156}, implies that

$$\max_{\pi_{L-1} \in \mathscr{R}_{L-1}} \{-b_{L-1}\overline{\tau}_{L-1\to L-2} + w_{L-1}(\overline{\tau}_{L\to L-2} - \overline{\tau}_{L-1\to L-2})\} = -1.$$

Combining the equation above with Equation (6.77), for $\ell = L-2$, we find that

$$\max_{\pi_{L-1} \in \mathscr{R}_{L-1}} \{-b_{L-1}\overline{\tau}_{L-1\to L-2} + p_{L-1}\overline{\tau}_{L\to L-1}\} = -1.$$

We will now generalise the proof using induction. Assuming that the lemma is true for $k+1$, with $k \in \mathscr{X} \setminus \{0, L-1, L\}$, we prove that it is also true for k.

Consider any $\ell < k$. By taking Equation $(6.28)_{160}$, for $i = k+1$ and $j = \ell$, we find that

$$\begin{aligned}
&\overline{\tau}_{k+1\to\ell} = 1 + \max_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{b_{k+1}\overline{\tau}_{k\to\ell} + r_{k+1}\overline{\tau}_{k+1\to\ell} + w_{k+1}\overline{\tau}_{k+2\to\ell}\} \\
&= 1 + \max_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{b_{k+1}\overline{\tau}_{k\to j} + (1 - b_{k+1} - w_{k+1})\overline{\tau}_{k+1\to j} + w_{k+1}\overline{\tau}_{k+2\to\ell}\} \\
&= 1 + \overline{\tau}_{k+1\to\ell} + \max_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{-b_{k+1}(\overline{\tau}_{k+1\to\ell} - \overline{\tau}_{k\to\ell}) \\
&\qquad\qquad + w_{k+1}(\overline{\tau}_{k+2\to\ell} - \overline{\tau}_{k+1\to\ell})\},
\end{aligned}$$

which, due to Theorem 58_{156}, implies that

$$\max_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{-b_{k+1}(\overline{\tau}_{k+1\to\ell} - \overline{\tau}_{k\to\ell}) + w_{k+1}(\overline{\tau}_{k+2\to\ell} - \overline{\tau}_{k+1\to\ell})\} = -1.$$

In combination with the induction hypothesis, which implies that $\overline{\tau}_{k+2\to\ell} = \overline{\tau}_{k+2\to k+1} + \overline{\tau}_{k+1\to\ell}$, the equation above results in

$$\max_{\pi_{k+1} \in \mathscr{R}_{k+1}} \{-b_{k+1}(\overline{\tau}_{k+1\to\ell} - \overline{\tau}_{k\to\ell}) + w_{k+1}\overline{\tau}_{k+2\to k+1}\} = -1. \tag{6.78}$$

Due to Proposition 67_{166} and the induction hypothesis, which implies that

$$\max_{\pi_{k+1}\in\mathscr{R}_{k+1}} \left\{-b_{k+1}\overline{\tau}_{k+1\to k} + w_{k+1}\overline{\tau}_{k+2\to k+1}\right\} = -1,$$

we infer from Equation (6.78)↶ that $\overline{\tau}_{k+1\to\ell} - \overline{\tau}_{k\to\ell} = \overline{\tau}_{k+1\to k}$, and therefore that

$$\overline{\tau}_{k+1\to\ell} = \overline{\tau}_{k+1\to k} + \overline{\tau}_{k\to\ell}. \tag{6.79}$$

By taking Equation $(6.28)_{160}$, for $i = k$ and $j = k-1$, we find that

$$\begin{aligned}
\overline{\tau}_{k\to k-1} &= 1 + \max_{\pi_k\in\mathscr{R}_k} \left\{r_k\overline{\tau}_{k\to k-1} + w_k\overline{\tau}_{k+1\to k-1}\right\} \\
&= 1 + \max_{\pi_k\in\mathscr{R}_k} \left\{(1-b_k-w_k)\overline{\tau}_{k\to k-1} + w_k\overline{\tau}_{k+1\to k-1}\right\} \\
&= 1 + \overline{\tau}_{k\to k-1} + \max_{\pi_k\in\mathscr{R}_k} \left\{-b_k\overline{\tau}_{k\to k-1} + w_k\left(\overline{\tau}_{k+1\to k-1} - \overline{\tau}_{k\to k-1}\right)\right\},
\end{aligned}$$

which, due to Theorem 58_{156}, implies that

$$\max_{\pi_k\in\mathscr{R}_k} \left\{-b_k\overline{\tau}_{k\to k-1} + w_k\left(\overline{\tau}_{k+1\to k-1} - \overline{\tau}_{k\to k-1}\right)\right\} = -1.$$

By combining this with Equation (6.79), for $\ell = k-1$, we find that

$$\max_{\pi_k\in\mathscr{R}_k} \left\{-b_k\overline{\tau}_{k-1\to k} + w_k\overline{\tau}_{k+1\to k}\right\} = -1. \qquad \square$$

7

AN APPLICATION TO QUEUEING

In this chapter our aim is to demonstrate the use of imprecise probabilities for robust queueing analysis. A queueing model is a stochastic process, for which the transition models are derived from an arrival and a departure process. We focus on the Geo/Geo/1/L queue, where the arrival and the departure processes obey geometric distributions, in the sense that at each time point an arrival occurs with probability $a \in [0,1]$ and a departure occurs with probability $d \in [0,1]$. A Geo/Geo/1/L queue is a special case of a birth-death chain and therefore allows for the efficient computation of expectations of various functions.

Afterwards, we introduce an imprecise version of the Geo/Geo/1/L queue. More specifically, instead of a single pair of arrival and departure probabilities, we assume that these probabilities belong to intervals and can take any value within the respective interval at each time point. This yields a special type of imprecise birth-death chain and we also introduce an additional independence concept that seems suitable for such an imprecise chain. We call this independence concept fixed-parameter repetition independence because it borrows elements from—but it is more stringent than—repetition independence.

We also present some properties and numerical results regarding various performance measures that are used in queueing theory. We construct an imprecise Geo/Geo/1/L queue and we calculate the lower and upper expected (average) queue length, (average) probability of each queue length, (average) probability of 'turning on the server', and first-passage and return times under the different independence concepts. For some of these performance measures, we prove that it makes no difference which independence concept is adopted. In the rest of the cases, we discuss the differences that arise.

7.1 IMPRECISION AND QUEUEING THEORY

The Achilles heel of queueing models for the use of decision support and prediction in practical applications is that, usually, it is difficult to specify the parameters of these models in a way that is both exact and reliable. In the most general form, a queueing model is treated as a stochastic model $Y = s(X)$, where the probabilistic information about X is typically expressed in terms of parameters. The problem, then, is to obtain a probabilistic description of Y, either analytically, algorithmically, by estimation or otherwise. The confidence we can put in our inferences about Y depends on how skilful the model $s(\cdot)$ is and how confident we are about our probabilistic description of X.

The goal of queueing theory in the last hundred years has been to analyse increasingly generalised, complicated and intricate models $s(\cdot)$. Although successful on its own, the drawback of this approach is that it often inspires overconfidence in its results, because the task of gaining confidence in the—parameters of the—probability model for X is not treated with the same attention. Once numbers are produced, their dependence on the parameters of the queueing model is often too easily forgotten. For this reason, we will analyse queueing models from an imprecise point of view in the same way we have done so far for stochastic processes. For example, when the probability of an arrival or the probability of a departure in a queue needs to be specified, we will allow for set-valued assessments. This is a more prudent and arguably more honest approach which, to some extent, allows accounting for the unreliability of the estimation of the input parameters in the queueing model.

The framework of imprecise probabilities is not the only attempt towards a more robust description of uncertainty in queueing theory. Amongst others, we have the Dempster–Shafer theory [63], interval probabilities [79] and fuzzy set theory [86]. To some extent, one can also characterise the influence of parameter uncertainty on performance measures strictly from within the framework of precise probabilities, for example by studying perturbations of the Markov chain underlying the queueing model [3, 12]. For small enough ε, the limiting distribution of a perturbated Markov chain with transition matrix $M_\varepsilon = M + \varepsilon M'$ is a power series in ε with coefficients that can be calculated from M' and both the stationary distribution and the deviation matrix of the original chain M [61]. Such methods provide a sensitivity analysis [11, 59, 80, 85]. Although very useful in many cases, they cover only small perturbations in one direction—that of M'—and are generally not able to relate sets of input parameters to performance bounds when general independence concepts are adopted like, for instance, the one of epistemic irrelevance.

7.2 THE GEO/GEO/1/L QUEUE AND HOW TO MAKE IT IMPRECISE

For the sake of demonstration and due to the fact that the merging between imprecise probabilities and queueing theory is quite unexplored, the queueing model that we consider is admittedly kept as simple as possible. That is, we consider a Geo/Geo/1/L queue which we then make imprecise by allowing the local models to be sets of probability mass functions instead of just single ones. This simplicity serves all the better to exhibit the implications of the local models when these are no longer single probability mass functions but sets of them. We also present an additional independence concept that can be applied to the imprecise version of the Geo/Geo/1/L .

7.2.1 *The Geo/Geo/1/L queue*

One standard way to describe a queueing model is by using Kendall's notation. That is, a queueing model is denoted by $A/D/S/L$, where A stands for the arrival process, D for the departure process, S for the number of servers and L for the maximum length of the queue. Here we focus on the Geo/Geo/1/L queue, which is a simple but quite common example in queueing theory. Following Kendall's notation, our system is a single-server queue of maximum length L, where $L \in \mathbb{N}$. Arrivals and departures occur according to geometrical distributions, i.e. if we are given a probability a for the arrival to occur it means that the probability of the first arrival at time point k is given by $(1-a)^{k-1}a$ and similarly for the departure. Arrivals and departures are assumed to be (stochastically) independent. Another assumption is that the content of the queue is observed between consecutive time points and we assume that at each time point, a departure occurs prior to an arrival, a convention called either Departures-First (DF) or Early Arrival System (EAS) [13].[1] We choose this priority in order to avoid zero time servicing when the queue is empty. Furthermore, we assume that an item stays in the queue until served and that the service discipline is work-conserving.

We now provide a more detailed description of this Geo/Geo/1/L queue; see also Reference [1, Section 6.2]. Assume that, at any time point and given any possible queue length, we have a probability of arrival and a probability of departure, which are denoted by a and d respectively. At each time point $n \in \mathbb{N}$, the possible queue lengths are represented by the state space $\mathscr{X} = \{0, \ldots, L\}$. The probability of any length in $\mathscr{X}$ at time point $n+1$ conditional on the previous length $x_n \in \mathscr{X}$ does not depend on time, but only on the values x_n, a and d. This implies that our transition models are conditional probability

[1] In queueing theory, it is typical to say that the content of the queue is observed during consecutive time slots and that arrivals and departures occur at slot boundaries. In our case, a slot is the time between two discrete time points.

mass functions $q(\cdot|x_n)$ which—according to the assumptions described in the previous paragraph—are defined as follows:

If $x_n = 0$, then

$$q(x_{n+1}|0) := \begin{cases} 1-a & \text{if } x_{n+1} = 0 \\ a & \text{if } x_{n+1} = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } x_{n+1} \in \mathscr{X}. \tag{7.1}$$

If $x_n \in \mathscr{X} \setminus \{0, L\}$, then for all $x_{n+1} \in \mathscr{X}$:

$$q(x_{n+1}|x_n) := \begin{cases} d(1-a) & \text{if } x_{n+1} = x_n - 1 \\ da + (1-d)(1-a) & \text{if } x_{n+1} = x_n \\ (1-d)a & \text{if } x_{n+1} = x_n + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{7.2}$$

Finally, if $x_n = L$, then

$$q(x_{n+1}|L) := \begin{cases} d(1-a) & \text{if } x_{n+1} = L-1 \\ 1-d(1-a) & \text{if } x_{n+1} = L \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } x_{n+1} \in \mathscr{X}. \tag{7.3}$$

A graphical representation of these transition models is shown in Figure 7.1$_{\curvearrowright}$. They can also be described by the following transition matrix:

$$M = \begin{pmatrix} 1-a & a & 0 & \cdots \\ d(1-a) & da+(1-d)(1-a) & (1-d)a & \cdots \\ \vdots & \ddots & \ddots & \ddots \\ \cdots & d(1-a) & da+(1-d)(1-a) & (1-d)a \\ 0 & \cdots & d(1-a) & 1-d(1-a) \end{pmatrix} \tag{7.4}$$

For the initial situation, we assume a probability mass function $q_{\square}$ on $\mathscr{X}$. It clearly follows from Equations (6.1)$_{152}$ and (7.4) that a Geo/Geo/1/L queue is a birth-death chain, and consequently it can also have a chain-like representation; see Figure 7.2$_{194}$. The probability tree that is associated with the local models that we have just introduced will be denoted by $q_{a,d}$.

Additionally to the assumptions mentioned in the beginning of this section, we assume that $0 < a < 1$ and $0 < d < 1$. The reason for this assumption is that it guarantees ergodicity. In particular, consider a Geo/Geo/1/L queue of which the transition matrix is given by Equation (7.4), where now the three diagonals consist of strictly positive elements, then it follows from Definition 8$_{112}$ that the queue is regular and therefore, ergodic. An ergodic Geo/Geo/1/L queue offers analytical formulas for computing the probability $\lim_{n\to+\infty} P(X_n = k)$,

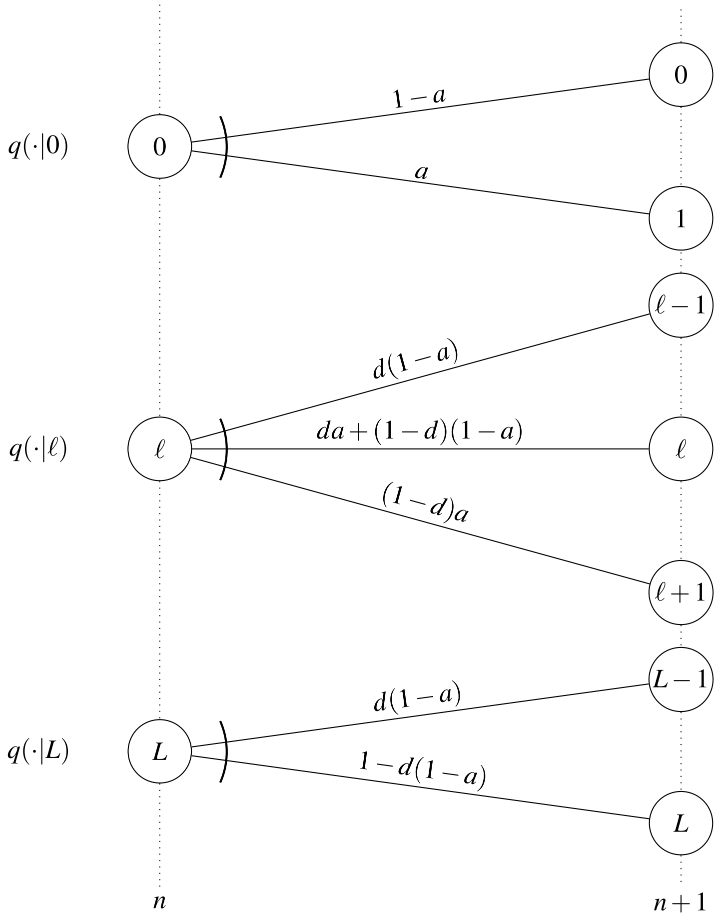

Figure 7.1: At any time point n and for each possible queue length $x_n \in \{0,\dots,L\}$, we have a conditional probability mass function $q(\cdot|x_n)$, where a and d are the probabilities of arrival and departure respectively.

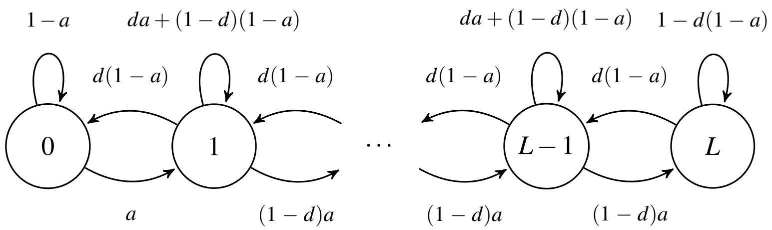

Figure 7.2: All the possible transitions from one state to another when the probability of arrival is a and the one for departure is d.

which we simply denote by $P(X=k)$ for all $k \in \mathscr{X}$.[2] In order to find these limit probabilities of the queue lengths, we use the so-called *balance equations* [35]:

$$aP(X=0) = d(1-a)P(X=1)$$

and

$$(1-d)aP(X=k-1) = d(1-a)P(X=k) \quad \text{for} \quad 2 \leq k \leq L.$$

By combining these equations with the fact that $\sum_{k\in\mathscr{X}} P(X=k) = 1$ and solving the resulting system of equations, we find—if $0 < a < 1$ and $0 < d < 1$—that

$$P(X=0) = \frac{d-a}{d - \dfrac{(1-d)^L a^{L+1}}{d^L(1-a)^L}}. \tag{7.5}$$

and

$$P(X=k) = \frac{(1-d)^{k-1}a^k}{d^k(1-a)^k} P(X=0) \quad \text{for} \quad 2 \leq k \leq L. \tag{7.6}$$

Alternatively, we can compute—not only in the limit—the probabilities of the possible queue lengths using the law of iterated expectations by regarding the probability of a queue length as the expectation of the indicator of this length.[3] Consider any $n \in \mathbb{N}$ and any $k \in \mathscr{X}$, then we know that $(X_n = k) := \cup_{x_{1:n-1}\in\mathscr{X}^{n-1}} \Gamma(x_{1:n-1}, k)$. Let $A_{n,k} := (X_n = k)$, it then follows from Lemma 105_{240} that $E_P(\mathbb{I}_{A_{n,k}}) = P(X_n = k)$. Furthermore, it is easy to see that $\mathbb{I}_{A_{n,k}}(X_{1:\infty}) = \mathbb{I}_k(X_n)$, where the indicator $\mathbb{I}_k$ is a function on $\mathscr{X}$. Now let T be the transition operator defined by Equation $(5.16)_{111}$ that is associated with the transition models of the Geo/Geo/1/L queue, it then follows from Lemma 36_{105} with respect to the homogeneous transition operator T that

$$P(X_n = k) = E_{\square}(T^{n-1}\mathbb{I}_k),$$

[2] In queueing theory, X represents the length of the queue when the latter is in a steady state.

[3] This approach is also know as computing the transient solution. More details about the transient solution for homogeneous Markov chains can be found in Reference [1, Section 3.8].

where $E_\square$ is the expectation operator that corresponds to the initial model $q_\square$. Moreover, we have that

$$P(X=k) = \lim_{n\to+\infty} E_\square(T^{n-1}\mathbb{I}_k) = \lim_{n\to+\infty} E_P\big(\mathbb{I}_k(X_n)\big),$$

and since the queue is ergodic, it follows from Equation $(5.17)_{111}$ that

$$P(X=k) = E_\infty(\mathbb{I}_k), \tag{7.7}$$

where E_∞ is the unique limit expectation operator of the ergodic Geo/Geo/1/L queue.

In the classic framework of precise probabilities, the Geo/Geo/1/L queue poses no problems and all performance measures can be obtained easily. One reason why this happens is because the Geo/Geo/1/L queue is a (homogeneous) birth-death chain and many performance measures are computed directly by just using the probability of arrival and the probability of departure. For instance, under the assumption that both the arrival and the departure probability are strictly positive, we can compute expectations in the limit of general functions in $\mathscr{L}(\mathscr{X})$ and time averages by using a closed-form expression. More specifically, for any $f \in \mathscr{L}(\mathscr{X})$, it follows from Equation $(5.17)_{111}$ that $\lim_{n\to+\infty} E_P(f(X_n)) = E_\infty(f)$, and due to Equation (7.7), we find that

$$E_\infty(f) = \sum_{x\in\mathscr{X}} f(x)P(X=x),$$

where $P(X=x)$ is given by Equations $(7.5)_{\curvearrowleft}$ and $(7.6)_{\curvearrowleft}$ for all $x \in \mathscr{X}$. Moreover, it follows from Theorem 38_{112} that $\sum_{x\in\mathscr{X}} f(x)P(X=x)$ is equal to the limit expectation of the time average of f. When there is imprecision, however, this may no longer be the case. The reason for this is that we will allow the probabilities of arrival and departure to vary over time and history of state values, which implies that we consider also non-homogeneous queues.

7.2.2 *An imprecise version of the Geo/Geo/1/L queue*

We now generalise the concept of a Geo/Geo/1/L queue by making it imprecise. More specifically, suppose that instead of a single pair a and d, we consider intervals of arrival and departure probabilities, denoted by $[\underline{a},\overline{a}]$ and $[\underline{d},\overline{d}]$, respectively, such that $0 < \underline{a} \leq \overline{a} < 1$ and $0 < \underline{d} \leq \overline{d} < 1$. Using $[\underline{a},\overline{a}]$ and $[\underline{d},\overline{d}]$, we first define sets of probability mass functions on the sets $\mathscr{X}_m$, $\mathscr{X}_0$ and $\mathscr{X}_L$—introduced in the beginning of Section 6.2_{152}—as follows:

$$\mathscr{R}_0 := \Big\{\pi_0 \in \Sigma_{\mathscr{X}_0} \colon \pi_0(e) = 1-a,\ \pi_0(u) = a \text{ and } a \in [\underline{a},\overline{a}]\Big\}; \tag{7.8}$$

$$\mathscr{R}_L := \Big\{\pi_L \in \Sigma_{\mathscr{X}_L} \colon \pi_L(\ell) = d(1-a),\ \pi_L(e) = 1-d(1-a), \\ a \in [\underline{a},\overline{a}] \text{ and } d \in [\underline{d},\overline{d}]\Big\}. \tag{7.9}$$

For all $i \in \mathscr{X} \setminus \{0,L\}$, we have that

$$\mathscr{R}_i := \Big\{\pi_i \in \Sigma_{\mathscr{X}_m} \colon \pi_i(\ell) = d(1-a),\ \pi_i(e) = da + (1-d)(1-a),\\ \pi_i(u) = (1-d)a,\ a \in [\underline{a},\overline{a}] \text{ and } d \in [\underline{d},\overline{d}]\Big\}. \quad (7.10)$$

For each $x \in \mathscr{X}$, we now consider a transition model $\mathscr{Q}_x$ on $\mathscr{X}$ that is defined by $\mathscr{R}_x$ according to Equations $(6.5)_{154}$–$(6.7)_{154}$, and consider also an initial model $\mathscr{Q}_\square$ on $\mathscr{X}$. These local models form a specific imprecise birth-death chain, which we call an *imprecise Geo/Geo/1/L queue*. Furthermore, since we assume that $0 < \underline{a} \leq \overline{a} < 1$ and $0 < \underline{d} \leq \overline{d} < 1$, this imprecise queue satisfies Assumption 6.1_{153}. Also, due to our notational convention that $(b_i, r_i, w_i) = (\pi_i(\ell), \pi_i(e), \pi_i(u))$ for all $\pi_i \in \mathscr{R}_i$ and all $i \in \mathscr{X} \setminus \{0,L\}$, $(r_0, w_0) = (\pi_0(e), \pi_0(u))$ for all $\pi_0 \in \mathscr{R}_0$ and $(b_L, r_L) = (\pi_L(\ell), \pi_L(e))$ for all $\pi_L \in \mathscr{R}_L$, we infer that $(r_0, w_0) = (1-a, a)$, $(b_L, r_L) = (d(1-a), 1-d(1-a))$ and $(b_i, r_i, w_i) = (d(1-a), da + (1-d)(1-a), a(1-d))$, where $a \in [\underline{a},\overline{a}]$ and $d \in [\underline{d},\overline{d}]$.

Since an imprecise Geo/Geo/1/L queue is an imprecise birth-death chain, and consequently a homogeneous imprecise Markov chain, we can choose between different independence concepts and approaches for building our global models. In particular, we can choose among epistemic irrelevance, complete independence and repetition independence, as introduced in Chapter 5_{100}. Apart from these three independence concepts, we can also adopt an additional independence concept that is suitable for imprecise Geo/Geo/1/L queues, and which will be introduced in Section $7.3_{\curvearrowright}$. Regarding the approach used for defining our global models, we can distinguish between the measure-theoretic and the martingale-theoretic approach.

For the bounds on the performance measures that are presented in Sections 7.5_{202}—7.8_{219}, we take into consideration all possible independence concepts. In summary, these bounds are lower and upper expectations of functions that depend on one or two states, time averages and first-passage times. When it comes to global lower and upper expectations of functions that depend on a single state and time averages, it follows from Theorems 51_{137} and 52_{138} that epistemic irrelevance coincides with complete independence, but not necessarily with the rest of the independence concepts; see Example 9_{143} and also Sections 7.6_{206} and 7.7_{210}. As far as lower and upper expected first-passage and return times are concerned, it makes no difference whether we adopt epistemic irrelevance, complete or repetition independence. In Section $7.8.2_{225}$, however, we will see that this is not the case for the additional independence concept. For epistemic irrelevance, regarding the approach used to define our global models, we consider both the measure-theoretic approach and the martingale-theoretic approach. Not surprisingly, the chosen approach does not affect our results, and we can therefore adopt computational methods that are based on either of them. Regarding functions that depend on one or two states and time averages, it follows from Theorem 29_{96} that the measure-theoretic approach coincides with the martingale-theoretic one. Furthermore, due to Theorem 80_{180}, this is also the case for lower and upper expected first-passage and return times.

We end this section by presenting a useful simplification of the expression for $\underline{T}f(x)$ that is given by Equation $(5.54)_{127}$, for the specific case of our imprecise Geo/Geo/1/L queue of which the transition models $\mathscr{Q}_x$ are now derived from the sets $\mathscr{R}_x$ that are given by Equations $(7.8)_{195}$–$(7.10)_{\frown}$. First of all, we find that

$$\begin{aligned}\underline{T}f(0) &= \min\Big\{(1-a)f(0)+af(1)\colon a\in[\underline{a},\overline{a}]\Big\}\\ &= \min\Big\{(1-a)f(0)+af(1)\colon a\in\{\underline{a},\overline{a}\}\Big\}.\end{aligned} \tag{7.11}$$

Secondly, we have that

$$\begin{aligned}\underline{T}f(L) &= \min\Big\{d(1-a)f(L-1)+[1-d(1-a)]f(L)\colon a\in[\underline{a},\overline{a}],\ d\in[\underline{d},\overline{d}]\Big\}\\ &= \min\Big\{d(1-a)f(L-1)+[1-d(1-a)]f(L)\colon a\in\{\underline{a},\overline{a}\},\ d\in\{\underline{d},\overline{d}\}\Big\}.\end{aligned} \tag{7.12}$$

Moreover, for all $x\in\mathscr{X}\setminus\{0,L\}$, we find that

$$\begin{aligned}\underline{T}f(x) = \min\Big\{&d(1-a)f(x-1)+[(1-d)(1-a)+da]f(x)\\ &+(1-d)af(x+1)\colon a\in[\underline{a},\overline{a}],\ d\in[\underline{d},\overline{d}]\Big\}\\ = \min\Big\{&d(1-a)f(x-1)+[(1-d)(1-a)+da]f(x)\\ &+(1-d)af(x+1)\colon a\in\{\underline{a},\overline{a}\},\ d\in\{\underline{d},\overline{d}\}\Big\},\end{aligned} \tag{7.13}$$

where the second equality comes from Proposition 3_{50} (for $I_1=[\underline{d},\overline{d}]$, $I_2=[\underline{a},\overline{a}]$ and $\Psi_{I_1,I_2}=\mathscr{Q}_x$). Similar simplifications apply to $\overline{T}f$ as well.

7.3 FIXED-PARAMETER REPETITION INDEPENDENCE

In this section, we introduce an independence concept that borrows elements from—but it is more stringent than—repetition independence. Consider the sets $\mathscr{Q}_x$ that are given by Equations $(7.8)_{195}$—$(7.10)_{\frown}$ and, for the initial situation $\square$, consider any set of probability mass functions $\mathscr{Q}_\square$ on $\mathscr{X}$. Each of the independence concepts that we have so far seen then leads to a different set of probability trees: $\mathscr{T}_{\mathscr{Q}}$, $\mathscr{T}_{\mathscr{Q}}^{\mathrm{M}}$ and $\mathscr{T}_{\mathscr{Q}}^{\mathrm{HM}}$. Furthermore, for every $a\in[\underline{a},\overline{a}]$, $d\in[\underline{d},\overline{d}]$ and $q_\square\in\mathscr{Q}_\square$, the corresponding probability tree $q_{a,d}$, as introduced in Section 7.2.1$_{191}$, clearly belongs to each of these sets. However, there might be other probability trees in these sets that are not of the form $q_{a,d}$. For example, consider any probability tree of which the transition model associated with the queue length 0 is given by Equation (7.1)$_{192}$ for $a = \underline{a}$ and for which the rest of the transition models are given by Equations (7.2)$_{192}$ and (7.3)$_{192}$ for $a = \overline{a}$ and $d = \underline{d}$. Such a probability tree belongs to $\mathscr{T}_{\mathscr{Q}}$, $\mathscr{T}_{\mathscr{Q}}^{\mathrm{M}}$ and $\mathscr{T}_{\mathscr{Q}}^{\mathrm{HM}}$, but is not of the form $q_{a,d}$. This observation leads us to introduce a fourth type of imprecise probability tree which is denoted by $\mathscr{T}_{\mathscr{Q}}^{\mathrm{O}}$ and defined by

$$\mathscr{T}_{\mathscr{Q}}^{\mathrm{O}} := \left\{ q_{a,d} \colon a \in [\underline{a}, \overline{a}],\, d \in [\underline{d}, \overline{d}] \text{ and } q_{\square} \in \mathscr{Q}_{\square} \right\}, \tag{7.14}$$

where, as in Section 7.2.1$_{191}$, $q_{a,d}$ is a probability tree which is obtained for a single homogeneous arrival probability a, a single homogeneous departure probability d and a single probability mass function $q_{\square}$ on $\mathscr{X}$.[4] The imprecise probability tree $\mathscr{T}_{\mathscr{Q}}^{\mathrm{O}}$ has a corresponding set of conditional probability measures on $\mathscr{C}_{\sigma}$, which we denote by $\mathbb{P}_{\mathscr{Q}}^{\mathrm{O}}$. The set $\mathbb{P}_{\mathscr{Q}}^{\mathrm{O}}$ corresponds to an imprecise Geo/Geo/1/L queue and we call the independence concept that is satisfied by the states of the process *fixed-parameter repetition independence*. Clearly, we have that $\mathscr{T}_{\mathscr{Q}}^{\mathrm{O}} \subseteq \mathscr{T}_{\mathscr{Q}}^{\mathrm{HM}} \subseteq \mathscr{T}_{\mathscr{Q}}^{\mathrm{M}} \subseteq \mathscr{T}_{\mathscr{Q}}$ and, consequently, also that $\mathbb{P}_{\mathscr{Q}}^{\mathrm{O}} \subseteq \mathbb{P}_{\mathscr{Q}}^{\mathrm{HM}} \subseteq \mathbb{P}_{\mathscr{Q}}^{\mathrm{M}} \subseteq \mathbb{P}_{\mathscr{Q}}$.

Regarding lower and upper expectations under fixed-parameter independence, for any measurable extended real-valued function g on Ω and any $B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\}$, the global lower and upper expectation of g conditional on B, denoted by $\underline{E}_{\mathscr{Q}}^{\mathrm{fi}}(g|B)$ and $\overline{E}_{\mathscr{Q}}^{\mathrm{fi}}(g|B)$ respectively, are defined by

$$\underline{E}_{\mathscr{Q}}^{\mathrm{fi}}(g|B) := \inf\left\{ E_P(g|B) \colon P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{O}} \right\}; \tag{7.15}$$

$$\overline{E}_{\mathscr{Q}}^{\mathrm{fi}}(g|B) := \sup\left\{ E_P(g|B) \colon P \in \mathbb{P}_{\mathscr{Q}}^{\mathrm{O}} \right\}. \tag{7.16}$$

Combining the fact that $\mathbb{P}_{\mathscr{Q}}^{\mathrm{O}} \subseteq \mathbb{P}_{\mathscr{Q}}^{\mathrm{HM}} \subseteq \mathbb{P}_{\mathscr{Q}}^{\mathrm{M}} \subseteq \mathbb{P}_{\mathscr{Q}}$ with the definitions of the global lower and upper expectation under the different independence concepts, we have the following property, which is trivial and therefore stated without proof.

Lemma 85. *Consider an initial model $\mathscr{Q}_{\square}$ and for each $x \in \mathscr{X}$, a set of conditional probability mass functions $\mathscr{Q}_x$ as defined in Section 7.2.2$_{195}$. Consider as well any $B \in \langle \mathscr{X}^* \rangle \setminus \{\varnothing\}$ and any measurable extended real-valued function g on Ω for which there is a non-decreasing sequence of non-negative n-measurable functions $\{h_n\}_{n \in \mathbb{N}}$ such that $\lim_{n \to +\infty} h_n = g$. It then holds that*

$$\underline{E}_{\mathscr{Q}}^{\mathrm{ei}}(g|B) \leq \underline{E}_{\mathscr{Q}}^{\mathrm{ci}}(g|B) \leq \underline{E}_{\mathscr{Q}}^{\mathrm{ri}}(g|B) \leq \underline{E}_{\mathscr{Q}}^{\mathrm{fi}}(g|B) \leq \overline{E}_{\mathscr{Q}}^{\mathrm{fi}}(g|B) \leq \overline{E}_{\mathscr{Q}}^{\mathrm{ri}}(g|B) \leq \overline{E}_{\mathscr{Q}}^{\mathrm{ci}}(g|B) \leq \overline{E}_{\mathscr{Q}}^{\mathrm{ei}}(g|B).$$

[4] For readers that are familiar with the concept of "separately specified rows" [18], it may be of interest to know that, in contrast with imprecise Markov chains under epistemic irrelevance, complete independence and repetition independence, the local models of an imprecise Markov chain under fixed-parameter repetition independence may not satisfy this property.

Computationally speaking, we cannot always compute global lower and upper expectations under fixed-parameter repetition independence exactly. Furthermore, the results may differ from those obtained under epistemic irrelevance, complete independence and repetition independence. The following example illustrates this.

Example 10. Consider the set $\mathscr{X} = \{0,1,2\}$ and the intervals $I_a = [1/5, 2/5]$ and $I_d = [2/5, 3/5]$. Consider also any imprecise Geo/Geo/1/L queue under epistemic irrelevance and any imprecise Geo/Geo/1/L queue under fixed-parameter repetition independence such that, for both of these models, the state space is $\mathscr{X}$, the initial model is $\mathscr{Q}_\square := (1/3, 1/3, 1/3)$ and the local models at time $n=1$ are derived from Equations $(7.8)_{195}$–$(7.10)_{196}$ with $[\underline{a}, \overline{a}] := I_a$ and $[\underline{d}, \overline{d}] := I_d$.

We will calculate the lower expectations $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_1(X_2))$—which coincides with $\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\mathbb{I}_1(X_2))$ and $\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\mathbb{I}_1(X_2))$ due to Proposition 54_{145}—and $\underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\mathbb{I}_1(X_2))$, and will show that they differ. Starting with $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_1(X_2))$ it follows from Equation $(5.57)_{128}$ that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_1(X_2)) = \underline{Q}_\square(\underline{T}\mathbb{I}_1). \tag{7.17}$$

For $\underline{T}\mathbb{I}_1(0)$, it follows from Equation $(7.11)_{197}$ that

$$\begin{aligned}\underline{T}\mathbb{I}_1(0) &= \min\Big\{(1-a)\mathbb{I}_1(0) + a\mathbb{I}_1(1)\colon a \in I_a\Big\}\\ &= \min\Big\{(1-a)\mathbb{I}_1(0) + a\mathbb{I}_1(1)\colon a \in \big\{\frac{1}{5}, \frac{2}{5}\big\}\Big\}\\ &= \min\Big\{a\colon a \in \big\{\frac{1}{5}, \frac{2}{5}\big\}\Big\} = \frac{1}{5}.\end{aligned}$$

Similarly, by using Equations $(7.12)_{197}$ and $(7.13)_{197}$, we find that $\underline{T}\mathbb{I}_1(2) = 6/25$ and $\underline{T}\mathbb{I}_1(1) = 11/25$.

Finally, it follows from Equation (7.17) that

$$\begin{aligned}\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_1(X_2)) &= \underline{Q}_\square(\underline{T}\mathbb{I}_1) = \inf\Big\{\sum_{y\in\mathscr{X}} \underline{T}\mathbb{I}_1(y)p(y)\colon p \in \mathscr{Q}_\square\Big\}\\ &= \inf\big\{\underline{T}\mathbb{I}_1(0)p(0) + \underline{T}\mathbb{I}_1(1)p(1) + \underline{T}\mathbb{I}_1(2)p(2)\colon p \in \mathscr{Q}_\square\big\}\\ &= \frac{1}{3}\cdot\frac{1}{5} + \frac{1}{3}\cdot\frac{11}{25} + \frac{1}{3}\cdot\frac{6}{25} = \frac{1}{3}\Big(\frac{1}{5} + \frac{11}{25} + \frac{6}{25}\Big) = \frac{1}{3}\cdot\frac{22}{25} = \frac{22}{75}.\end{aligned} \tag{7.18}$$

We now calculate $\underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\mathbb{I}_1(X_2))$, for which we have that

$$\begin{aligned}\underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\mathbb{I}_1(X_2)) &= \inf\Big\{E_P(\mathbb{I}_1(X_2))\colon P \in \mathbb{P}^{\mathrm{O}}_{\mathscr{Q}}\Big\}\\ &= \inf\Big\{\sum_{x_{1:2}\in\mathscr{X}^2} \mathbb{I}_1(x_2)\prod_{i=0}^{1} p(x_{i+1}|x_{1:i})\colon p \in \mathscr{T}^{\mathrm{O}}_{\mathscr{Q}}\Big\}\end{aligned}$$

$$
\begin{aligned}
&= \inf\Big\{\mathbb{I}_1(0)q(0|0)q_\square(0) + \mathbb{I}_1(1)q(1|0)q_\square(0) + \mathbb{I}_1(0)q(0|1)q_\square(1) \\
&\qquad + \mathbb{I}_1(1)q(1|1)q_\square(1) + \mathbb{I}_1(2)q(2|1)q_\square(1) + \mathbb{I}_1(1)q(1|2)q_\square(2) \\
&\qquad + \mathbb{I}_1(2)q(2|2)q_\square(2) \colon q_{a,d} \in \mathscr{T}_{\mathscr{Q}}^{\mathrm{O}}\Big\} \\
&= \inf\Big\{a\frac{1}{3} + [da + (1-d)(1-a)]\frac{1}{3} + d(1-a)\frac{1}{3} \colon a \in I_a,\ d \in I_d\Big\} \\
&= \inf\Big\{\frac{1}{3} + da\frac{1}{3} \colon a \in I_a,\ d \in I_d\Big\} = \inf\Big\{\frac{1}{3}(1+da) \colon a \in I_a,\ d \in I_d\Big\} \\
&= \frac{1}{3}(1 + \frac{2}{5}\cdot\frac{1}{5}) = \frac{1}{3}\cdot\frac{27}{25} = \frac{27}{75},
\end{aligned} \tag{7.19}
$$

where the first equality follows from Equation $(7.15)_{198}$, the second equality follows from Equation $(3.18)_{74}$, the third follows from Equation $(7.14)_{198}$, the fourth equality follows from Equations $(7.1)_{192}$–$(7.3)_{192}$ and the seventh equality holds because I_a and I_d are closed intervals consisting of strictly positive values and therefore the infimum $\inf\{\frac{1}{3}(1+da) \colon a \in I_a,\ d \in I_d\}$ is obtained for the smallest value in I_d and the smallest value in I_a.

Finally, due to Equations $(7.18)_{\curvearrowleft}$ and (7.19), we infer that $\underline{E}_{\mathscr{Q}}^{\mathrm{ei}}(\mathbb{I}_1(X_2) < \underline{E}_{\mathscr{Q}}^{\mathrm{fi}}(\mathbb{I}_1(X_2))$. ◊

For that reason, we here propose a method for approximating global lower and upper expectations under fixed-parameter repetition independence which is similar to the one for repetition independence described in Section $5.7.2_{143}$. For instance, consider any $n, m \in \mathbb{N}$ such that $n > m$, any situation $x_{1:m} \in \mathscr{X}^m$, any function $f \in \mathscr{L}(\mathscr{X})$ and suppose that we want to compute $\underline{E}_{\mathscr{Q}}^{\mathrm{fi}}(f(X_n)|x_{1:m})$. This time, we select k probabilities from $[\underline{a}, \overline{a}]$ and k probabilities from $[\underline{d}, \overline{d}]$, and we compute the global expectations for all possible combinations of the selected probabilities. The smallest value among these global expectations is then our approximation for the lower global expectation. In the unconditional case, we additionally select k probability mass functions from $\mathscr{Q}_\square$. In a similar way we can approximate (conditional) global upper expectations of functions $f \in \mathscr{L}(\mathscr{X})$ and also global lower and upper expectations of their respective time average. With this method we need to compute k^2 global expectations in the conditional case and k^3 global expectations in the unconditional case.

7.4 OUR EXAMPLE OF AN IMPRECISE GEO/GEO/1/L QUEUE

From the analysis in Section $7.2.2_{195}$, we infer that any imprecise Geo/Geo/1/L queue can be constructed by specifying a maximum queue length L, an interval of arrival probabilities, an interval of departure probabilities and an initial model. For the bounds on the performance measures that are calculated in Sections 7.5_{202}–7.8_{219}, we have considered an imprecise Geo/Geo/1/L queue with $L = 7$, where the probability interval for an arrival is $[0.5, 0.6]$ and the one for a departure $[0.7,0.8]$. For the initial model $\mathscr{Q}_\square$, we used a *vacuous* model—the set of all probability mass functions on $\mathscr{X}$, namely $\Sigma_\mathscr{X}$.

For repetition independence and fixed-parameter repetition independence, as mentioned before, exact computations are in general infeasible. Therefore, in those cases, we approximate global lower and upper expectations by selecting a number of probabilities and then computing the global expectations for all possible combinations of the selected probabilities—see Sections 5.7.2$_{143}$ and 7.3$_{197}$. In particular, we select eleven different values for the arrival and departure probabilities, and eight different probability mass functions for the different initial models: arrival probabilities take values in the set $\{0.5, 0.51, \ldots, 0.6\}$, departure probabilities in the set $\{0.7, 0.71, \ldots, 0.8\}$, and initial models are selected from the extreme points of $\Sigma_\mathscr{X}$.

Note that, in contrast with the approximation methods described in Sections 5.7.2$_{143}$ and 7.3$_{197}$, the number of selected probability mass functions for the initial model is not the same as the number of selected arrival and departure probabilities. The reason why we selected only the extreme points of $\Sigma_\mathscr{X}$ for the initial model is because our approximations for unconditional global lower and upper expectations under (fixed-parameter) repetition independence will not be better for any other selection of initial models in $\Sigma_\mathscr{X}$. We only explain in more detail why this statement is true for unconditional global lower expectations under fixed-parameter repetition independence of functions that depend on a single state. The cases of unconditional global upper expectations of functions that depend on a single state under fixed-parameter independence, unconditional global lower and upper expectations of time averages under fixed-parameter repetition independence, and also all the respective ones under repetition independence, are completely analogous.

Indeed, consider any $n \in \mathbb{N}$, any $f \in \mathscr{L}(\mathscr{X})$, any $a \in [\underline{a}, \overline{a}]$, any $d \in [\underline{d}, \overline{d}]$ and any $q_\square \in \Sigma_\mathscr{X}$. Let $q_{a,d} \in \mathscr{T}^{\mathrm{O}}_{\mathscr{Q}}$ be the probability tree that is derived from a, d and $q_\square$. Let also T be the transition operator that is associated with the transition models of $q_{a,d}$, and $E_\square$ be the expectation operator that corresponds to $q_\square$. Consider now any $P \in \mathbb{P}_{q_{a,d}}$, then it follows from Lemma 36$_{105}$—for the homogeneous transition operator T—that $E_P(f(X_n)) = E_\square(T^{n-1}f)$. Since $T^{n-1}f$ is a real-valued function on $\mathscr{X}$, it then follows from property P1$_{38}$ that

$$\min T^{n-1}f \leq E_\square(T^{n-1}f) \leq \max T^{n-1}f.$$

Let $y^* \in \operatorname{argmin}_{x \in \mathscr{X}} T^{n-1}f(x)$ and let also $q^*_\square$ be the probability mass function in $\Sigma_\mathscr{X}$ that assigns the probability mass 1 to y^* and 0 to any $x \in \mathscr{X} \setminus \{y^*\}$. Clearly, $q^*_\square$ is one of the extreme points of $\Sigma_\mathscr{X}$. Let now $E^*_\square$ be the expectation operator that corresponds to $q^*_\square$, then we have that $E^*_\square(T^{n-1}f) = \min T^{n-1}f$ and therefore that $E^*_\square(T^{n-1}f) \leq E_\square(T^{n-1}f)$.

7.5 EXPECTED QUEUE LENGTH

The first performance measure that we deal with is the expected queue length. At any time point $n \in \mathbb{N}$, the queue length is given by X_n, which implies that it depends on a single state and hence, it follows from Theorem 51_{137} that epistemic irrelevance coincides with complete independence. Furthermore, it follows from Theorem 29_{96} that the measure-theoretic approach coincides with the martingale-theoretic one. We adopt the measure-theoretic approach and we compute the global lower and upper expected queue length at time n, for increasing values of n.

Under epistemic irrelevance and complete independence, we calculate the global lower and upper expected queue length using Equation $(5.57)_{128}$. For repetition independence and fixed-parameter repetition independence, we calculate the global lower and upper expected queue length according the procedures described in Sections $5.7.2_{143}$ and 7.3_{197} respectively using the selected arrival and departure probabilities that were presented in Section 7.4_{200}.

We start by calculating unconditional global lower and upper expected queue length at time n, and also upper expected queue length at time n conditional on the event that the queue is empty at time 1, for increasing values of n, under both epistemic irrelevance and fixed-parameter repetition independence; see Figure $7.3_{\curvearrowright}$. In the previous paragraph, we mentioned that we calculate global lower and upper expectations under epistemic irrelevance and complete independence using Equation $(5.57)_{128}$, which applies to unconditional global lower and upper expectations. However, our experiments consider conditional ones as well. This is not a problem because we can regard the global upper expected queue length at time n conditional on the event that the queue is empty at time 1 as an unconditional global upper expectation where the initial model is the probability mass function in $\Sigma_{\mathscr{X}}$ that assigns the probability mass 1 to the empty queue—state value 0—and probability mass 0 to any $x \in \mathscr{X} \setminus \{0\}$. Let $q^*_{\square}$ be the aforementioned probability mass function in $\Sigma_{\mathscr{X}}$, then for all $f \in \mathscr{L}(\mathscr{X})$, it follows from Equation $(5.56)_{128}$ that $\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n)|X_1 = 0) = \overline{T}^{n-1} f(0)$, and we further observe that

$$\sum_{y \in \mathscr{X}} q^*_{\square}(y)\overline{T}^{n-1} f(y) = \overline{T}^{n-1} f(0),$$

which, because of Equation $(5.57)_{128}$, is equal to $\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(X_n)$ with respect to the initial model $\mathscr{Q}_{\square} = \{q^*_{\square}\}$.

Under epistemic irrelevance, we observe convergence for both the global lower and upper expectation, regardless of whether we use the vacuous initial model or start from an empty queue, where the latter, as explained earlier, can be regarded as a special case of an unconditional expectation. This is due to the fact that the probability intervals of arrival and departure consist of strictly positive probabilities, which can be shown to imply that there is some $n' \in \mathbb{N}$ such that for all $\ell \geq n'$, $-\underline{T}^{\ell}(-\mathbb{I}_x) > 0$ for all $x \in \mathscr{X}$, and it then follows

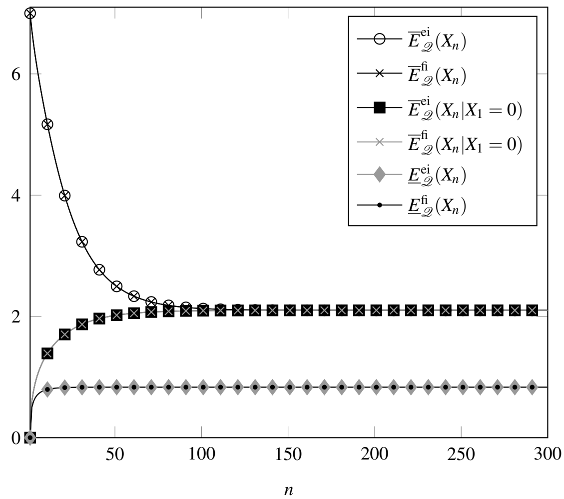

Figure 7.3: Global lower and upper expected queue length.

from Definition 10_{129} that the imprecise Geo/Geo/1/L queue is regularly absorbing. Therefore, the imprecise queue is also ergodic and it follows from Definition 9_{128} and the analysis in Section $5.5.1_{126}$ that the global lower and upper expected queue length converge to a value that is independent of the initial model or initial state. Furthermore, for any single homogeneous arrival and departure probability the corresponding Geo/Geo/1/L queue is also regular (Definition 8_{112}), and consequently ergodic, which, due to Equation $(5.17)_{111}$, implies that the global lower and upper expected queue length under fixed-parameter independence also converges to a value that is independent of the initial model or initial state.

7.5.1 *Monotonicity*

In Figure 7.3, we also observe that the results under epistemic irrelevance coincide with those under fixed-parameter repetition independence, which due to Lemma 85_{198}, implies that both concepts also coincide with complete independence and repetition independence. Therefore, we will not refer to complete independence and repetition independence for the rest of this section. Under fixed-parameter repetition independence, we obtain, reasonably, the global lower expected queue length for the smallest probability of arrival (0.5) and the largest probability of departure (0.8). This happens due to the "monotonicity" of the function used. In particular, under epistemic irrelevance, although we do not require the use of a single homogeneous probability of arrival and departure, one can show that for these functions, the optimal choice for the arrival probabilities is to always consider the minimum value, and the optimal choice for the departure probability is to always consider the maximum value.

This is due to the "monotonicity" of the argument function, as made clear in the following theorem—similar (suitably adapted) results hold for the global upper expectation, and for functions $f \in \mathscr{L}(\mathscr{X})$ that are non-increasing on $\mathscr{X}$ rather than non-decreasing.

Theorem 86. *Consider any $n \in \mathbb{N}$ and any $f \in \mathscr{L}(\mathscr{X})$ such that*

$$f(k) \leq f(k+1) \text{ for all } k \in \{0,\dots,L-1\}. \tag{7.20}$$

Then, in an imprecise Geo/Geo/1/L queue under epistemic irrelevance with parameters in intervals $[\underline{a},\overline{a}]$ and $[\underline{d},\overline{d}]$, the lower expected value $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n))$ is obtained for homogeneous parameters $\underline{a}$ and $\overline{d}$, and the upper expected value $\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(f(X_n))$ for $\overline{a}$ and $\underline{d}$.

Proof. We provide the proof for global lower expectations; the proof for the global upper ones is completely analogous.

Due to Equation $(5.57)_{128}$, the result follows—by induction—if for any function $f \in \mathscr{L}(\mathscr{X})$ that satisfies Equation (7.20), we can show (a) that $\underline{T}f$ is obtained for $a = \underline{a}$ and $d = \overline{d}$ and (b) that $\underline{T}f$ also satisfies Equation (7.20).

For all $x \in \{1,\dots,L\}$, let $m_x := f(x) - f(x-1) \geq 0$, where the inequality follows from Equation (7.20). We first prove (a). For $x = 0$, Equation $(7.11)_{197}$ implies that

$$\begin{aligned}\underline{T}f(0) &= \min\big\{(1-a)f(0) + af(1)\colon a \in \{\underline{a},\overline{a}\}\big\}\\ &= \min\big\{(f(0) + am_1\colon a \in \{\underline{a},\overline{a}\}\big\} = f(0) + \underline{a}m_1,\end{aligned} \tag{7.21}$$

where the last step holds because $m_1 \geq 0$. Similarly, for $x \in \{1,\dots,L-1\}$, Equation $(7.13)_{197}$ implies that

$$\begin{aligned}\underline{T}f(x) = \min\big\{&d(1-a)f(x-1) + [(1-d)(1-a) + da]f(x)\\ &+ (1-d)af(x+1)\colon a \in \{\underline{a},\overline{a}\}, d \in \{\underline{d},\overline{d}\}\big\}\\ = \min\big\{&f(x) - d(1-a)m_x + (1-d)am_{x+1}\colon a \in \{\underline{a},\overline{a}\}, d \in \{\underline{d},\overline{d}\}\big\}\\ = f(x) &- \overline{d}(1-\underline{a})m_x + (1-\overline{d})\underline{a}m_{x+1},\end{aligned} \tag{7.22}$$

where the last step holds because $m_x \geq 0$ and $m_{x+1} \geq 0$. Finally, for $x = L$, Equation $(7.12)_{197}$ implies that

$$\begin{aligned}\underline{T}f(L) &= \min\big\{d(1-a)f(L-1) + [1-d(1-a)]f(L)\colon a \in \{\underline{a},\overline{a}\}, d \in \{\underline{d},\overline{d}\}\big\}\\ &= \min\big\{f(L) - d(1-a)m_L\colon a \in \{\underline{a},\overline{a}\}, d \in \{\underline{d},\overline{d}\}\big\}\\ &= f(L) - \overline{d}(1-\underline{a})m_L,\end{aligned} \tag{7.23}$$

where the last step holds because $m_L \geq 0$. This concludes the proof of (a).

We now prove (b): $\underline{T}f(x+1)-\underline{T}f(x)\geq 0$ for all $x\in\{0,\dots,L-1\}$. For $x=0$, this holds because it follows from Equations (7.21)$_\curvearrowleft$ and (7.22)$_\curvearrowleft$ that

$$\begin{aligned}\underline{T}f(1)-\underline{T}f(0)&=\big(f(1)-\overline{d}(1-\underline{a})m_1+(1-\overline{d})\underline{a}m_2\big)-(f(0)+\underline{a}m_1)\\&=m_1-\overline{d}(1-\underline{a})m_1+(1-\overline{d})\underline{a}m_2-\underline{a}m_1\\&\geq m_1-\overline{d}(1-\underline{a})m_1-\underline{a}m_1=(1-\underline{a})(1-\overline{d})m_1\geq 0\end{aligned}$$

For $x\in\{1,\dots,L-2\}$, this holds because it follows from Equation (7.22)$_\curvearrowleft$ that

$$\begin{aligned}\underline{T}f(x+1)-\underline{T}f(x)&=\big(f(x+1)-\overline{d}(1-\underline{a})m_{x+1}+(1-\overline{d})\underline{a}m_{x+2}\big)\\&\qquad-\big(f(x)-\overline{d}(1-\underline{a})m_x+(1-\overline{d})\underline{a}m_{x+1}\big)\\&=m_{x+1}-\overline{d}(1-\underline{a})m_{x+1}+(1-\overline{d})\underline{a}m_{x+2}+\overline{d}(1-\underline{a})m_x-(1-\overline{d})\underline{a}m_{x+1}\\&\geq m_{x+1}-\overline{d}(1-\underline{a})m_{x+1}-(1-\overline{d})\underline{a}m_{x+1}=(1-\underline{a})(1-\overline{d})m_{x+1}\geq 0.\end{aligned}$$

For $x=L-1$, this holds because it follows from Equations (7.22)$_\curvearrowleft$ and (7.23)$_\curvearrowleft$ that

$$\begin{aligned}\underline{T}f(L)-\underline{T}f(L-1)&=\big(f(L)-\overline{d}(1-\underline{a})m_L\big)\\&\qquad-\big(f(L-1)-\overline{d}(1-\underline{a})m_{L-1}+(1-\overline{d})\underline{a}m_L\big)\\&=m_L-\overline{d}(1-\underline{a})m_L+\overline{d}(1-\underline{a})m_{L-1}-(1-\overline{d})\underline{a}m_L\\&\geq m_L-\overline{d}(1-\underline{a})m_L-(1-\overline{d})\underline{a}m_L\\&=\big((1-\overline{d})(1-\underline{a})+\overline{d}\underline{a}\big)\,m_L\geq 0.\end{aligned}$$

□

7.5.2 *Expected average queue length*

The average queue length at any time point $n\in\mathbb{N}$ is the average of the queue length over the time points 1 through n, that is $\frac{1}{n}\sum_{i=1}^{n}X_i$. For the global lower and upper expected average queue length, things are similar to the case of the global lower and upper expected queue length. Since the average queue length is a time average, it follows from Theorem 52$_{138}$ that epistemic irrelevance coincides with complete independence and furthermore, it follows from Theorem 29$_{96}$ that the measure-theoretic approach coincides with the martingale-theoretic one. Again adopting the measure-theoretic approach, we calculate the global lower and upper expected average queue length under epistemic irrelevance (or complete independence) according to Lemma 41$_{119}$ using the homogeneous lower and upper transition operators $\underline{T}$ and $\overline{T}$, and the corresponding ones under repetition independence and fixed-parameter repetition independence according the procedures described in Sections 5.7.2$_{143}$ and 7.3$_{197}$ respectively.

We first calculate the global lower and upper expected average queue length at time n, for increasing values of n, under both epistemic irrelevance and fixed-

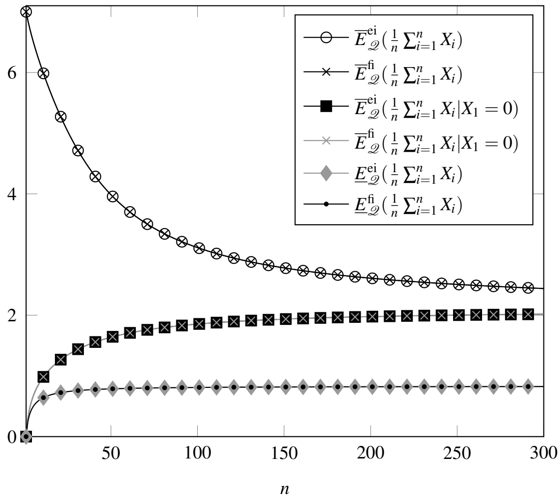

Figure 7.4: Global lower and upper expected average queue length.

parameter repetition independence; see Figure 7.4. We find again that epistemic irrelevance coincides with fixed-parameter independence and therefore, due to Lemma 85$_{198}$, both concepts also coincide with complete independence and repetition independence. For fixed-parameter repetition independence, although the convergence is rather slow, Figure 7.5$_{\curvearrowright}$ seems to suggest that the lower expected queue length and lower expected average queue length converge to the same value. Furthermore, since we have just seen that neither of these two objects depends on the chosen independence concept, the same is true for repetition independence, complete independence and epistemic irrelevance. Hence, in this case, we find that the inequality in Lemma 57$_{148}$ is actually an equality.

7.6 PROBABILITY OF DIFFERENT QUEUE LENGTHS

For a precise Geo/Geo/1/L queue P, we know from Section 7.2.1$_{191}$ that for any $n \in \mathbb{N}$ and any $k \in \mathscr{X}$, the probability $P(X_n = k)$ of queue length k at time n is equal to $E_P(\mathbb{I}_k(X_n))$. In fact, this equality clearly holds for any conditional probability measure P. Since we now consider local models as defined in Section 7.2.2$_{195}$, which are associated with various sets of conditional probability measures, we can use this equality to derive similar equalities for global lower and upper probabilities of queue lengths. We will only do this explicitly for epistemic irrelevance. The derivations for the other independence concepts are analogous.

Consider an imprecise Geo/Geo/1/L queue under epistemic irrelevance, then we know that it corresponds to a set of conditional probability measures

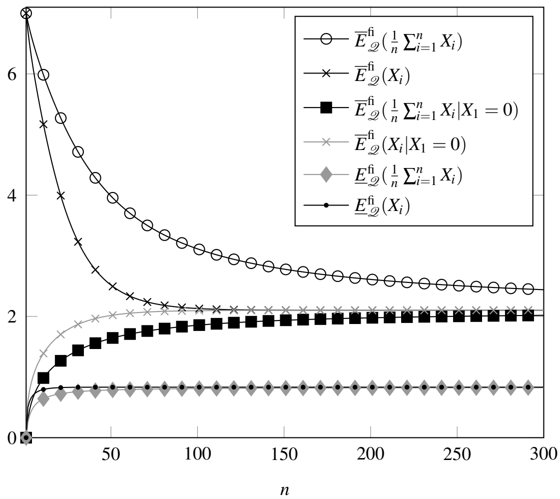

Figure 7.5: Global lower and upper expected (average) queue length.

$\mathbb{P}_{\mathscr{Q}}$. The global lower probability of queue length k at time n is therefore defined by $\inf\{P(X_n=k)\colon P\in\mathbb{P}_{\mathscr{Q}}\}$, and we observe that

$$\inf\left\{P(X_n=k)\colon P\in\mathbb{P}_{\mathscr{Q}}\right\}=\inf\left\{E_P(\mathbb{I}_k(X_n))\colon P\in\mathbb{P}_{\mathscr{Q}}\right\}=\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_k(X_n)).$$

For the global upper probability of queue length k at time n, we find that

$$\sup\left\{P(X_n=k)\colon P\in\mathbb{P}_{\mathscr{Q}}\right\}=\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_k(X_n)).$$

We also discuss global lower and upper average probabilities of the different queue lengths. In the precise-probabilistic framework, the average probability of queue length k up to time n is given by $\frac{1}{n}\sum_{i=1}^{n}P(X_i=k)$ and due to Lemma 102_{240} and the fact that $P(X_i=k)=E_P(\mathbb{I}_k(X_i))$, we find that

$$\frac{1}{n}\sum_{i=1}^{n}P(X_i=k)=E_P\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_k(X_i)\right). \tag{7.24}$$

Consider now an imprecise Geo/Geo/1/L queue under epistemic irrelevance, then the global lower average probability of queue length k up to time n is defined by $\inf\{\frac{1}{n}\sum_{i=1}^{n}P(X_i=k)\colon P\in\mathbb{P}_{\mathscr{Q}}\}$ and we observe that

$$\begin{aligned}\inf\left\{\frac{1}{n}\sum_{i=1}^{n}P(X_i=k)\colon P\in\mathbb{P}_{\mathscr{Q}}\right\}&=\inf\left\{E_P(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_k(X_i))\colon P\in\mathbb{P}_{\mathscr{Q}}\right\}\\&=\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_k(X_i)\right),\end{aligned}$$

where the first equality follows from Equation $(7.24)_{\curvearrowleft}$. For the upper case, we have that

$$\sup\left\{\frac{1}{n}\sum_{i=1}^{n} P(X_i = k) \colon P \in \mathbb{P}_{\mathscr{Q}}\right\} = \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_k(X_i)\right).$$

Similar arguments hold for the other independence concepts.

Since the different queue lengths at time n are functions that depend on a single state and their respective average ones are time averages, it follows from Theorem 51_{137} that epistemic irrelevance coincides with complete independence, and it also follows from Theorem 29_{96} that for epistemic irrelevance, the measure-theoretic approach coincides with the martingale-theoretic approach. Hence, we will not refer to complete independence and the martingale-theoretic approach for the remainder of this section. By adopting the measure-theoretic approach, we calculate the lower and upper (average) probability for all $k \in \mathscr{X}$ using the same methods that were used for the global lower and upper (average) expected queue length.

We begin by showing, in Tables $7.1_{\curvearrowright}$– 7.3_{211}, the global lower and upper (average) probability of every possible queue length at time n under epistemic irrelevance and fixed-parameter repetition independence as n approaches infinity. For the lower and upper probabilities under fixed-parameter repetition independence we used Equations $(7.5)_{194}$ and $(7.6)_{194}$. Moreover, for the results under fixed-parameter repetition independence, we provide—between parentheses—the probabilities of arrival and departure (a,d), for which the global lower or upper expectation was obtained.

For $k \in \{0,7\}$, as in the case of the expected queue length, the results under epistemic irrelevance coincide with the respective ones under fixed-parameter repetition independence and therefore, due to Lemma 85_{198}, also under complete independence and repetition independence. Again, this is due to the "monotonicity" of the function used. However, this is not the case for the other queue lengths, i.e. $k \in \{1,2\dots,6\}$, and in the rest of this section, we discuss these as well as other differences.

The probability mass functions used for the calculation of lower and upper probabilities of the different states depend on the chosen independence concept. For example, as we can see in Table 7.2_{210}, under fixed-parameter repetition independence, the global upper probability of having queue length 1 in the limit is obtained for $a = 0.55$ and $d = 0.8$. On the other hand, as we know from Equation $(5.57)_{128}$ in combination with Equations $(7.11)_{197}$—$(7.13)_{197}$, the optimisation problem that needs to be solved under epistemic irrelevance only considers extreme values of a, that is $\underline{a}$ or $\overline{a}$. This implies that, in this case, the probability tree used depends on the chosen type of independence. It is therefore not surprising that the values of $\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_1(X_n))$ and $\overline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\mathbb{I}_1(X_n))$, for $n \to +\infty$, are different. As is to be expected from Lemma 85_{198}, we have that $\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_1(X_n)) \geq \overline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\mathbb{I}_1(X_n))$, and in this case, we further infer that

k	0	7
$\lim_{n\to+\infty} \underline{E}^{\text{ei}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.1486	0.0001
$\lim_{n\to+\infty} \underline{E}^{\text{ei}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_k(X_i))$	0.1486	0.0001
$\lim_{n\to+\infty} \overline{E}^{\text{ei}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.3750	0.0225
$\lim_{n\to+\infty} \overline{E}^{\text{ei}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_k(X_i))$	0.3750	0.0225
$\lim_{n\to+\infty} \underline{E}^{\text{fi}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.1486 $(0.6, 0.7)$	0.0001 $(0.5, 0.8)$
$\lim_{n\to+\infty} \underline{E}^{\text{fi}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_k(X_i))$	0.1486 $(0.6, 0.7)$	0.0001 $(0.5, 0.8)$
$\lim_{n\to+\infty} \overline{E}^{\text{fi}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.3750 $(0.5, 0.8)$	0.0225 $(0.6, 0.7)$
$\lim_{n\to+\infty} \overline{E}^{\text{fi}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_k(X_i))$	0.3750 $(0.5, 0.8)$	0.0225 $(0.6, 0.7)$

Table 7.1: Global lower and upper (average) probabilities of queue lengths 0 and 7 under epistemic irrelevance and fixed-parameter repetition independence for $n \to +\infty$.

$\overline{E}^{\text{ei}}_{\mathscr{Q}}(\mathbb{I}_1(X_n)) > \overline{E}^{\text{fi}}_{\mathscr{Q}}(\mathbb{I}_1(X_n))$. Similar conclusions can be drawn for $\underline{E}^{\text{ei}}_{\mathscr{Q}}(\mathbb{I}_1(X_n))$, for which it holds that $\underline{E}^{\text{ei}}_{\mathscr{Q}}(\mathbb{I}_1) > \underline{E}^{\text{fi}}_{\mathscr{Q}}(\mathbb{I}_1(X_n))$ as shown in Table 7.2$_{\curvearrowright}$. The global lower and upper probabilities under repetition independence lie between the respective ones under epistemic irrelevance and fixed-parameter repetition independence.

Judging from Tables 7.1– 7.3$_{211}$, we see that the global lower and upper average probabilities under fixed-parameter repetition independence are included between the global lower and upper probabilities, when both are taken to the limit. Under fixed-parameter repetition independence, this is to be expected. More specifically, the lower and upper probabilities will coincide with the respective ones of the averages, due to Theorem 38$_{112}$, since for each selection of arrival and departure probability the resulting Geo/Geo/1/L queue is regular, and therefore ergodic.

Under epistemic irrelevance, Lemma 57$_{148}$ only guarantees an inequality. A lower (or upper) expectation and a corresponding average one might be obtained for different probability trees. In other words, under epistemic irrelevance, the 'worst-case' scenario in the limit is never better than the 'average worst-case' scenario. Our results in Tables 7.1– 7.3$_{211}$—where the limits inferior and superior in the lemma are actually limits—confirm this result. Also, as we can see, in some cases, strict inequalities can be observed. We stress that both scenarios are practically relevant. The probability of being in state k at

k	1	2	3
$\lim_{n\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.2909	0.1081	0.0276
$\lim_{n\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_k(X_i))$	0.3093	0.1100	0.0289
$\lim_{n\to+\infty} \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.5344	0.2684	0.1683
$\lim_{n\to+\infty} \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_k(X_i))$	0.5173	0.2577	0.1492
$\lim_{n\to+\infty} \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.3185	0.1172	0.0293
	$(0.6, 0.7)$	$(0.5, 0.8)$	$(0.5, 0.8)$
$\lim_{n\to+\infty} \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_k(X_i))$	0.3185	0.1172	0.0293
	$(0.6, 0.7)$	$(0.5, 0.8)$	$(0.5, 0.8)$
$\lim_{n\to+\infty} \overline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.4775	0.2065	0.1316
	$(0.55, 0.8)$	$(0.6, 0.72)$	$(0.6, 0.7)$
$\lim_{n\to+\infty} \overline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_k(X_i))$	0.4775	0.2065	0.1316
	$(0.55, 0.8)$	$(0.6, 0.72)$	$(0.6, 0.7)$

Table 7.2: Global lower and upper (average) probabilities of queue lengths 1, 2 and 3 under epistemic irrelevance and fixed-parameter repetition independence for $n \to +\infty$.

time n is important for a single customer who would arrive at time instant n, while the average probability of being in state k is important from the system operator's point of view.

The global lower and upper average probabilities under repetition independence are not given, but will of course lie within the respective average ones under epistemic irrelevance and fixed-parameter repetition independence.

Finally, in order to get an idea of how lower and upper (average) probabilities of queue lengths evolve over time, Figures $7.6_{\curvearrowright}$ and 7.7_{212} depict the global lower and upper probabilities under epistemic irrelevance and fixed-parameter independence and the respective average ones, for queue length 1. In Figure 7.8_{212}, we provide a direct comparison of the global lower and upper probabilities under epistemic irrelevance with the respective average ones, again for queue length 1.

7.7 TURNING ON THE SERVER

In this section, we present our results for the (average) probability of "turning on the server". Consider any $n \in \mathbb{N}$, then turning on the server at time $n+1$ means that the queue length at time $n+1$ is 1 and the queue length at time n is 0. This is expressed by the event $S_{n+1} := \cup_{x_{1:n-1}\in\mathscr{X}^{n-1}} \Gamma(x_{1:n-1}, 0, 1)$, and it follows from Lemma 105_{240} that $E_P(\mathbb{I}_{S_{n+1}}) = P(S_{n+1})$. Moreover, since $\mathbb{I}_{S_{n+1}}$

k	4	5	6
$\lim_{n\to+\infty} \underline{E}^{\text{ei}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.0069	0.0017	0.00044
$\lim_{n\to+\infty} \underline{E}^{\text{ei}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_k(X_i))$	0.0073	0.0018	0.00046
$\lim_{n\to+\infty} \overline{E}^{\text{ei}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.1053	0.0648	0.0388
$\lim_{n\to+\infty} \overline{E}^{\text{ei}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_k(X_i))$	0.0902	0.0559	0.0352
$\lim_{n\to+\infty} \underline{E}^{\text{fi}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.0073 $(0.5,0.8)$	0.0018 $(0.5,0.8)$	0.00046 $(0.5,0.8)$
$\lim_{n\to+\infty} \underline{E}^{\text{fi}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_k(X_i))$	0.0073 $(0.5,0.8)$	0.0018 $(0.5,0.8)$	0.00046 $(0.5,0.8)$
$\lim_{n\to+\infty} \overline{E}^{\text{fi}}_{\mathscr{Q}}(\mathbb{I}_k(X_n))$	0.0846 $(0.6,0.7)$	0.0544 $(0.6,0.7)$	0.0350 $(0.6,0.7)$
$\lim_{n\to+\infty} \overline{E}^{\text{fi}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_k(X_i))$	0.0846 $(0.6,0.7)$	0.0544 $(0.6,0.7)$	0.0350 $(0.6,0.7)$

Table 7.3: Global lower and upper (average) probabilities of queue lengths 4, 5 and 6 under epistemic irrelevance and fixed-parameter repetition independence for $n \to +\infty$.

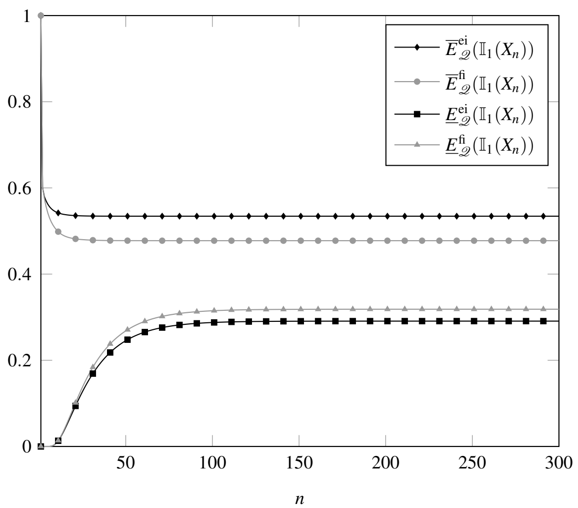

Figure 7.6: Global lower and upper probability of queue length 1 under epistemic irrelevance and fixed-parameter repetition independence.

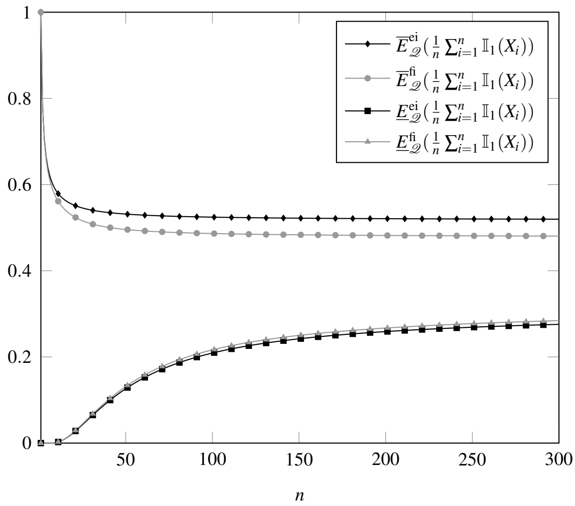

Figure 7.7: Global lower and upper average probability of queue length 1 under epistemic irrelevance and fixed-parameter repetition independence.

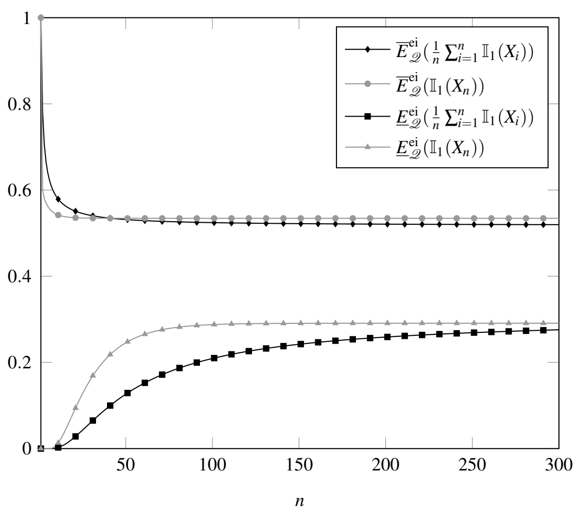

Figure 7.8: Global lower and upper (average) probability of queue length 1 under epistemic irrelevance.

is $(n+1)$-measurable and does not depend on the first $n-1$ states $X_{1:n-1}$, $\mathbb{I}_{S_{n+1}}$ can be replaced by a function that depends only on the states $X_{n:n+1}$, which in this case is $\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1})$. Hence, the probability of turning on the server at time $n+1$ is given by $E_P(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1}))$.

The average probability of turning on the server up to time $n+1$ is defined by $\frac{1}{n}\sum_{i=1}^{n} P(S_{i+1})$. Also in this case, we can express such a probability in terms of indicators in the following way:

$$\frac{1}{n}\sum_{i=1}^{n} P(S_{i+1}) = \frac{1}{n}\sum_{i=1}^{n} E_P(\mathbb{I}_{S_{i+1}}) = E_P\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{S_{i+1}}\right) = E_P\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1})\right), \tag{7.25}$$

where the second equality follows from Lemma 102$_{240}$.

Since we work with local models as defined in Section 7.2.2$_{195}$, and consequently with sets of conditional probability measures, we need to define the global lower and upper (average) probability of turning on the server at time $n+1$ for any $n \in \mathbb{N}$. We only show how these probability bounds are defined under epistemic irrelevance; for the other independence concepts these are defined in a similar way. Consider an imprecise Geo/Geo/1/L queue under epistemic irrelevance. This implies that we have a set of conditional probability measures $\mathbb{P}_{\mathscr{Q}}$ and the global lower probability is then the infimum over all $P(S_{n+1})$ for $P \in \mathbb{P}_{\mathscr{Q}}$. Therefore, we find that it is given by

$$\begin{aligned}\inf\left\{P(S_{n+1})\colon P \in \mathbb{P}_{\mathscr{Q}}\right\} &= \inf\left\{E_P(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1}))\colon P \in \mathbb{P}_{\mathscr{Q}}\right\}\\ &= \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1})).\end{aligned}$$

Similarly, the global upper probability of turning on the server at time $n+1$ is given by

$$\sup\left\{P(S_{n+1})\colon P \in \mathbb{P}_{\mathscr{Q}}\right\} = \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1})).$$

Next, we discuss global lower and upper average probabilities of turning on the server up to time $n+1$. Consider again an imprecise Geo/Geo/1/L queue under epistemic irrelevance. Then the global lower average probability of turning on the server up to time $n+1$ is given by

$$\begin{aligned}\inf\left\{\frac{1}{n}\sum_{i=1}^{n} P(\mathbb{I}_{S_{i+1}})\colon P \in \mathbb{P}_{\mathscr{Q}}\right\} &= \inf\left\{E_P\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1})\right)\colon P \in \mathbb{P}_{\mathscr{Q}}\right\}\\ &= \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1})\right),\end{aligned}$$

where the first equality follows from Equation (7.25). The upper average probability of turning on the server up to time $n+1$ is

$$\sup\left\{\frac{1}{n}\sum_{i=1}^{n} P(\mathbb{I}_{S_{i+1}})\colon P \in \mathbb{P}_{\mathscr{Q}}\right\} = \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1})\right).$$

Similar expressions apply for the rest of the independence concepts as well.

We will now show how we compute the global lower and upper (average) probability of turning on the server when the independence concept is epistemic irrelevance. Consider any $n \in \mathbb{N}$ and any $x_{1:n} \in \mathscr{X}^n$, and observe that

$$\begin{aligned}
&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1})|x_{1:n}) = \underline{Q}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1})|x_{1:n}) \\
&\quad = \underline{Q}(\mathbb{I}_0(x_n)\mathbb{I}_1(X_{n+1})|x_{1:n}) = \mathbb{I}_0(x_n)\underline{Q}(\mathbb{I}_1(X_{n+1})|x_{1:n}) = \mathbb{I}_0(x_n)\underline{Q}(\mathbb{I}_1(X_{n+1})|x_n) \\
&\quad = \mathbb{I}_0(x_n)\underline{T}\mathbb{I}_1(x_n),
\end{aligned} \tag{7.26}$$

where the first equality follows from Theorem 21_{82}, the second follows from Equation $(3.31)_{82}$, the third holds because, since $\mathbb{I}_0(x_n)$ is a positive coefficient and $\underline{Q}(\cdot|x_{1:n})$ is an infimum of expectations, it follows from Lemma 102_{240} that $\mathbb{I}_0(x_n)$ can be taken out, the fourth equality follows from Equation $(5.53)_{126}$ and the last follows from Equation $(5.54)_{127}$.

Finally, observe that

$$\begin{aligned}
\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1})) &= \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1})|X_{1:n})) \\
&= \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\underline{T}\mathbb{I}_1(X_n)) = \underline{Q}_{\square}\big(\underline{T}^{n-1}(\mathbb{I}_0\underline{T}\mathbb{I}_1)\big),
\end{aligned} \tag{7.27}$$

where the first equality follows from Theorem 21_{82}, the second follows from Equation (7.26) and the last follows from Equation $(5.57)_{128}$.

Similarly, for the global upper probability of turning on the server at time $n+1$, we find that

$$\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1})) = \overline{Q}_{\square}\big(\overline{T}^{n-1}(\mathbb{I}_0\overline{T}\mathbb{I}_1)\big). \tag{7.28}$$

For the lower and upper average probability of turning on the server, we use recursive functions that are based on lower and upper transition operators $\underline{T}$ and $\overline{T}$. For any $k \in \mathbb{N}$ and any $f \in \mathscr{L}(\mathscr{X})$, these real-valued functions on $\mathscr{X}$ are denoted by $\underline{\psi}_k(f)$ and $\overline{\psi}_k(f)$ and defined recursively as

$$\underline{\psi}_k(f) := \sum_{y\in\mathscr{X}} \mathbb{I}_y\underline{T}\Big(\mathbb{I}_0(y)\mathbb{I}_1 + \underline{\psi}_{k-1}(f)\Big)(y) \tag{7.29}$$

and

$$\overline{\psi}_k(f) := \sum_{y\in\mathscr{X}} \mathbb{I}_y\overline{T}\Big(\mathbb{I}_0(y)\mathbb{I}_1 + \overline{\psi}_{k-1}(f)\Big)(y), \tag{7.30}$$

with initial value $\underline{\psi}_0(f) = \overline{\psi}_0(f) = f$.

We can now compute the lower and upper average probability of turning on the server under epistemic irrelevance, using the following lemma.

Lemma 87. *Consider a homogeneous imprecise Markov chain under epistemic irrelevance and any $n \in \mathbb{N}$, then*

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1})\right) = \frac{1}{n}\underline{Q}_{\square}\big(\underline{\psi}_n(0)\big);$$
$$\overline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1})\right) = \frac{1}{n}\overline{Q}_{\square}\big(\overline{\psi}_n(0)\big).$$

Proof. We will only provide the proof for the global lower expectations; the proof for the global upper ones is completely analogous.

We first prove by induction that for all $k \in \mathbb{N}$ and all $f \in \mathscr{L}(\mathscr{X})$ the following holds:

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}) + f(X_{k+1})\right) = \underline{Q}_{\square}\big(\underline{\psi}_k(f)\big). \tag{7.31}$$

Consider first any $x_1 \in \mathscr{X}$ and observe that

$$\begin{aligned}\underline{Q}\big(\mathbb{I}_0(X_1)\mathbb{I}_1(X_2) + f(X_2)\big|x_1\big) &= \underline{Q}\big(\mathbb{I}_0(x_1)\mathbb{I}_1(X_2) + f(X_2)\big|x_1\big)\\ &= \underline{T}\big(\mathbb{I}_0(x_1)\mathbb{I}_1 + f\big)(x_1),\end{aligned} \tag{7.32}$$

where the first equality follows from Equation $(3.31)_{82}$ and the second follows from Equation $(5.54)_{127}$. Therefore, for $k = 1$, we find that

$$\begin{aligned}&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\big(\mathbb{I}_0(X_1)\mathbb{I}_1(X_2) + f(X_2)\big)\\ &\quad= \underline{Q}_{\square}\big(\underline{Q}\big(\mathbb{I}_0(X_1)\mathbb{I}_1(X_2) + f(X_2)\big|X_1\big)\big)\\ &\quad= \underline{Q}_{\square}\left(\sum_{x_1\in\mathscr{X}}\mathbb{I}_{x_1}(X_1)\underline{Q}\Big(\mathbb{I}_0(x_1)\mathbb{I}_1(X_2) + f(X_2)\Big|x_1\Big)\right)\\ &\quad= \underline{Q}_{\square}\left(\sum_{x_1\in\mathscr{X}}\mathbb{I}_{x_1}\underline{T}\big(\mathbb{I}_0(x_1)\mathbb{I}_1 + f\big)(x_1)\right) = \underline{Q}_{\square}\big(\underline{\psi}_1(f)\big),\end{aligned}$$

where the first equality follows from Theorem 21_{82}, the third follows from Equation (7.32) and the last from Equation $(7.29)_{\curvearrowleft}$.

Now consider any $k \geq 2$ and assume that Equation (7.31) is true for $k-1$. It follows from Theorem 21_{82} that

$$\begin{aligned}&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}) + f(X_{k+1})\right)\\ &\quad= \underline{Q}_{\square}\left(\dots\underline{Q}\left(\sum_{i=1}^{k}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}) + f(X_{k+1})\middle|X_{1:k}\right)\dots\right)\\ &\quad= \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\underline{Q}\left(\sum_{i=1}^{k}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}) + f(X_{k+1})\middle|X_{1:k}\right)\right).\end{aligned} \tag{7.33}$$

Consider now any $x_{1:k} \in \mathscr{X}^k$ and observe that

$$\begin{aligned}
&\underline{Q}\left(\sum_{i=1}^{k} \mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}) + f(X_{k+1}) \middle| x_{1:k}\right) \\
&\quad = \underline{Q}\left(\sum_{i=1}^{k-1} \mathbb{I}_0(x_i)\mathbb{I}_1(x_{i+1}) + \mathbb{I}_0(x_k)\mathbb{I}_1(X_{k+1}) + f(X_{k+1}) \middle| x_{1:k}\right) \\
&\quad = \sum_{i=1}^{k-1} \mathbb{I}_0(x_i)\mathbb{I}_1(x_{i+1}) + \underline{Q}\big(\mathbb{I}_0(x_k)\mathbb{I}_1(X_{k+1}) + f(X_{k+1})\big)\big| x_{1:k}\big) \\
&\quad = \sum_{i=1}^{k-1} \mathbb{I}_0(x_i)\mathbb{I}_1(x_{i+1}) + \underline{Q}\big(\mathbb{I}_0(x_k)\mathbb{I}_1(X_{k+1}) + f(X_{k+1})\big| x_k\big) \\
&\quad = \sum_{i=1}^{k-1} \mathbb{I}_0(x_i)\mathbb{I}_1(x_{i+1}) + \underline{T}\big(\mathbb{I}_0(x_k)\mathbb{I}_1 + f\big)(x_k),
\end{aligned} \tag{7.34}$$

where the first equality follows from Equation $(3.31)_{82}$, the second equality holds because, since $\sum_{i=1}^{k-1} \mathbb{I}_0(x_i)\mathbb{I}_1(x_{i+1})$ is a constant and $\underline{Q}(\cdot|x_{1:k-1})$ is an infimum of expectations, it follows from Lemmas 102_{240} and 106_{241} that we can take $\sum_{i=1}^{k-1} \mathbb{I}_0(x_i)\mathbb{I}_1(x_{i+1})$ out of the infimum, the third equality follows from Equation $(5.53)_{126}$ and the last equality follows from Equation $(5.54)_{127}$.

Furthermore, observe that

$$\begin{aligned}
&\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k} \mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}) + f(X_{k+1})\right) \\
&\quad = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\underline{Q}\left(\sum_{i=1}^{k} \mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}) + f(X_{k+1}) \middle| X_{1:k}\right)\right) \\
&\quad = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{k-1} \mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}) + \sum_{x_k \in \mathscr{X}} \mathbb{I}_{x_k}(X_k)\underline{T}\big(\mathbb{I}_0(x_k)\mathbb{I}_1 + f\big)(x_k)\right) \\
&\quad = \underline{Q}_{\square}\left(\underline{\psi}_{k-1}\left(\sum_{x_k \in \mathscr{X}} \mathbb{I}_{x_k}\underline{T}\big(\mathbb{I}_0(x_k)\mathbb{I}_1 + f\big)(x_k)\right)\right) = \underline{Q}_{\square}\big(\underline{\psi}_k(f)\big),
\end{aligned}$$

where the first equality follows from Equation $(7.33)_{\frown}$, the second follows from Equation (7.34), the third follows from the induction hypothesis since $\sum_{x_k \in \mathscr{X}} \mathbb{I}_{x_k}\underline{T}(\mathbb{I}_0(x_k)\mathbb{I}_1 + f)(x_k)$ is a function on $\mathscr{X}$ and the last follows from Equation $(7.29)_{214}$.

It now follows from Equation $(7.31)_{\frown}$ by letting $k = n$, $f = 0$ and multiplying both sides with $\frac{1}{n}$ that

$$\frac{1}{n}\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\sum_{i=1}^{n} \mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}))\right) = \frac{1}{n}\underline{Q}_{\square}\big(\underline{\psi}_n(0)\big)$$

and since $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\cdot)$ is an infimum of expectations, it follows from Lemma 102_{240} that

$$\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1})\right) = \frac{1}{n}\underline{Q}_{\square}\left(\underline{\psi}_n(0)\right). \qquad \square$$

Since the function of turning on the server at time $n+1$ and the corresponding one for the average are both $(n+1)$-measurable, it follows from Theorem 29_{96} that the measure-theoretic approach coincides with the martingale-theoretic approach. Hence, we will not refer to the martingale-theoretic approach for the remainder of this section.

For the lower and upper (average) probability of turning on the server under complete independence, we can use the approach that was described in Section $5.6.4_{138}$. More specifically, for the lower probability of turning on the server at time $n+1$, with $n \in \mathbb{N}$, we can construct a transition operator T_i for each $i \in \{1,\dots,n\}$, where each T_i is defined by Equation $(5.6)_{105}$ and the local models associated with it are derived from some arrival and departure probabilities of the selected ones presented in Section 7.4_{200}. Each T_i can be regarded as a transition matrix of the form in Equation $(6.1)_{152}$ of which each row $k \in \mathscr{X}$ is now constructed from a probability $a_{i,k} \in [\underline{a},\overline{a}]$ and a probability $d_{i,k} \in [\underline{d},\overline{d}]$. This implies that at each time point and given a state value we can choose an arrival and a departure probability from the selected ones regardless of the choice of arrival and departure probabilities at the same time point but given another state value or at any other time point and given any state value. This yields a collection of transition operators $\{T_1,T_2,\dots,T_n\}$ and we can then consider all such possible collections of transition operators that can be constructed from all possible combinations of the selected arrival and departure probabilities. Then we use a version of Equation $(7.27)_{214}$ adapted to the transition operators T_i and the selected initial models presented in Section 7.4_{200}, i.e. the extreme points of $\Sigma_{\mathscr{X}}$, in order to compute the global expectations of turning on the server, among which we consider the smallest to be the approximation for the global lower probability of turning on the server. In a similar way we can compute an approximation of the global upper probability of turning on the server and also the lower and upper average ones. In the latter case, we can again use each collection of transition operators $\{T_1,T_2,\dots,T_n\}$ and calculate for each such collection the global average probability and finally choose the smallest among these averages to be our approximation for the global lower average probability of turning on the server. For each collection of transition operators, we can use the formulas presented in Lemma 87_{214} adapted to the transition operators T_i.

Similar considerations apply for repetition independence, where we compute lower and upper (average) probabilities of turning on the server according to the method described in Section $5.7.2_{143}$. In particular, we now construct homogeneous transition operators T, as defined by Equation $(5.16)_{111}$, for each possible selection of the arrival and departure probabilities presented in Sec-

$\lim_{n\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1}))$	0.0743
$\lim_{n\to+\infty} \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}))$	0.0866
$\lim_{n\to+\infty} \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1}))$	0.2250
$\lim_{n\to+\infty} \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}))$	0.2000
$\lim_{n\to+\infty} \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1}))$	0.0892
$\lim_{n\to+\infty} \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}))$	0.0892
$\lim_{n\to+\infty} \overline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1}))$	0.1875
$\lim_{n\to+\infty} \overline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}))$	0.1875

Table 7.4: Global lower and upper (average) probability of of turning on the server under epistemic irrelevance and fixed-parameter repetition independence for $n \to +\infty$.

tion 7.4$_{200}$. This time, T can be regarded as transition matrix of the form in Equation (6.1)$_{152}$ of which each row $k \in \mathscr{X}$ is derived from an arrival probability $a_k \in [\underline{a}, \overline{a}]$ and a departure probability $d_k \in [\underline{d}, \overline{d}]$. We then compute global expectations using (adapted versions of) Equation (7.27)$_{214}$ and Lemma 87$_{214}$ for each possible homogeneous transition operator T that can be derived from the selected probabilities. Finally, we choose the smallest value to be the approximation for the global lower (average) probability of turning on the server. An approximation for the global upper (average) probability of turning on the server is computed similarly.

The method for computing the lower and upper (average) probability under fixed-parameter repetition independence, as defined in Section 7.3$_{197}$, is similar to the one under repetition independence. The difference is that our transition operators now are only allowed to include a single arrival and a single departure probability from the selected ones, that is, T can now be represented by a transition matrix of the form in Equation (7.4)$_{192}$. In other words, we consider all possible Geo/Geo/1/L queues that can be constructed from the selected arrival and departure probabilities presented in Section 7.4$_{200}$ and for each of the queues we compute the probability of turning on the server. From all the calculated probabilities of turning on the server, we consider the smallest to be the approximation for the global lower probability of turning on the server and the largest to be the approximation for the upper one. Similarly we calculate the respective average probabilities.

In Figures 7.9$_{\curvearrowright}$ and 7.10$_{220}$, we depict the lower and upper (average) probability of turning on the server, under both epistemic irrelevance and fixed-parameter repetition independence. The results obtained for $n \to +\infty$ are given in Table 7.4.

Our observations are similar to the ones for the queue lengths 1 to 6.

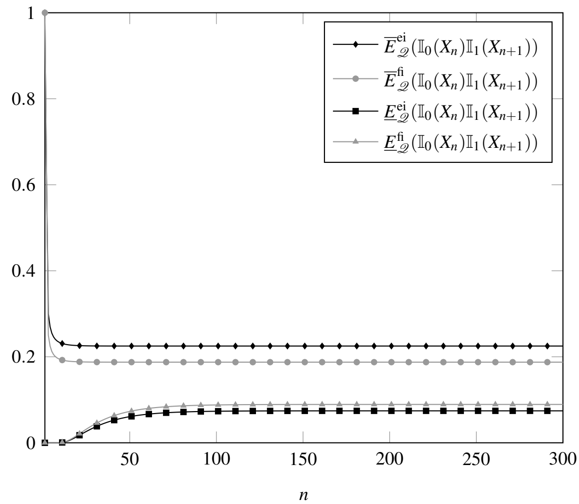

Figure 7.9: Global lower and upper probability of turning on the server.

First of all, the lower (and upper) probabilities of turning on the server depend on the chosen independence concept. It follows from Lemma 85_{198} that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1})) \leq \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\mathbb{I}_0(X_n)\mathbb{I}_1(X_{n+1}))$, and for $n \to +\infty$, we see that the inequality can in fact be strict. Due to Lemma 85_{198}, the lower probability of turning on the server under complete independence or repetition independence will lie between the respective ones under epistemic irrelevance and fixed-parameter repetition independence. Similar conclusions hold for the upper probability of turning on the server. Moreover, under fixed-parameter repetition independence, our example seems to suggest that the bounds on the average probability converge to the respective bounds on the probability at a single time point. Under epistemic irrelevance, however, the "worst-case" scenario in the limit is again at least as bad as the "average worst-case" scenario. We also observe that $\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1})) \leq \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_0(X_i)\mathbb{I}_1(X_{i+1}))$, and for $n \to +\infty$, we further see that the inequality is strict.

Finally, Lemma 85_{198} implies that the lower average probability of turning on the server under complete independence or repetition independence will lie between the respective ones under epistemic irrelevance and fixed-parameter repetition independence, and similarly for the upper.

7.8 EXPECTED FIRST-PASSAGE AND RETURN TIMES

The last performance measures to be discussed are expected first-passage and return times. It follows from Theorem 80_{180} that lower and upper expected first-passage and return times defined by the martingale-theoretic approach coincide with the respective ones defined by the measure-theoretic approach

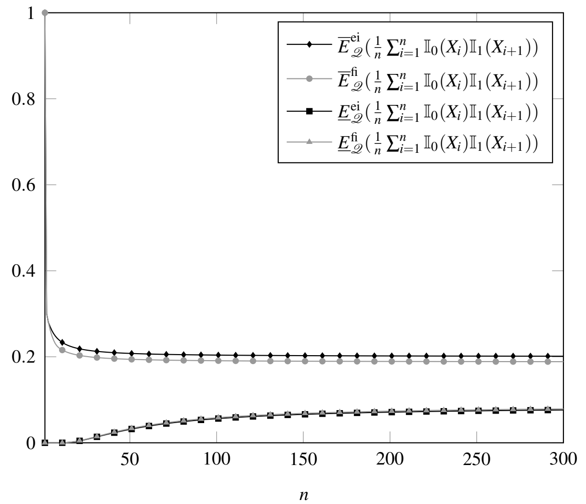

Figure 7.10: Global lower and upper average probability of turning on the server.

and also that epistemic irrelevance coincides with complete independence and repetition independence. In the next two sections, we compare the coinciding lower and upper expected first-passage and return times obtained under the aforementioned independence concepts with those obtained under fixed-parameter repetition independence.

7.8.1 *Expected first-passage times*

In Sections 6.4$_{161}$ and 6.5$_{165}$, we have described a method for computing lower and upper expected first-passage times. Regarding lower and upper upward expected first-passage times, for any $i, j \in \mathscr{X}$ such that $i < j$, $\underline{\tau}_{i\to j}$ and $\overline{\tau}_{i\to j}$ are given by the expressions in Corollary 62$_{164}$ and Proposition 65$_{165}$ respectively. These expressions include lower and upper expected first-passage times $\underline{\tau}_{k\to k+1}$ and $\overline{\tau}_{k\to k+1}$ for all $k \in \{i, \ldots, j-1\}$, which can be calculated recursively using a bisection method; see Propositions 60$_{163}$ and 59$_{162}$, and Propositions 63$_{164}$ and 64$_{164}$. Similarly, in the case of lower and upper downward expected first-passage times, for any $i, j \in \mathscr{X}$ such that $i > j$, $\underline{\tau}_{i\to j}$ and $\overline{\tau}_{i\to j}$ are given by the expressions in Proposition 68$_{167}$. These expressions include lower and upper expected first-passage times $\underline{\tau}_{\ell\to\ell-1}$ and $\overline{\tau}_{\ell\to\ell-1}$ for all $\ell \in \{j+1, \ldots, i\}$, which can be calculated recursively using a bisection method; see Propositions 66$_{166}$ and 67$_{166}$.

In an imprecise Geo/Geo/1/L queue, things become even simpler when it comes to lower and upper first-passage times. In order to show that, we start by presenting a property that is satisfied by lower and upper upward expected first-passage times when the local models are defined as in Section 7.2.2$_{195}$.

Theorem 88. *Consider an imprecise Geo/Geo/1/L queue with parameters in intervals* $[\underline{a},\overline{a}]$ *and* $[\underline{d},\overline{d}]$ *such that* $0<\underline{a}<\overline{a}<1$ *and* $0<\underline{d}<\overline{d}<1$*, a Geo/Geo/1/L queue with transition matrix* M *of the form in Equation* (7.4)$_{192}$ *such that* $a=\overline{a}$ *and* $d=\underline{d}$*, and a Geo/Geo/1/L queue with transition matrix* M' *of the form in Equation* (7.4)$_{192}$ *such that* $a=\underline{a}$ *and* $d=\overline{d}$*. Then for all* $i,j\in\mathscr{X}$ *such that* $i<j$*,* $\underline{\tau}_{i\to j}=\tau^{M}_{i\to j}$ *and* $\overline{\tau}_{i\to j}=\tau^{M'}_{i\to j}$*.*

Proof. For the statement about lower expected first-passage times, it follows from from Equations (6.1)$_{152}$ and (7.4)$_{192}$, Theorem 72$_{171}$ and Selection Methods LU$_L$ 1$_{171}$ and 2$_{171}$ that all we need to prove is that $\overline{w}_0=\overline{a}$ and that for all $\ell\in\{1,\dots,L-1\}$

$$\big(\underline{d}(1-\overline{a}),\underline{d}\overline{a}+(1-\underline{d})(1-\overline{a}),\overline{a}(1-\underline{d})\big)\in\operatorname*{argmin}_{\pi_\ell\in\mathscr{R}_\ell}\{b_\ell\underline{\tau}_{\ell-1\to\ell}-w_\ell\underline{\tau}_{\ell\to\ell+1}\}.$$

The fact that $\overline{w}_0=\overline{a}$ is trivial, since we have that $[\underline{w}_0,\overline{w}_0]=[\underline{a},\overline{a}]$. For all $\ell\in\{1,\dots,L-1\}$, we find that

$$\begin{aligned}\min_{\pi_\ell\in\mathscr{R}_\ell}\{b_\ell\underline{\tau}_{\ell-1\to\ell}-w_\ell\underline{\tau}_{\ell\to\ell+1}\}&=\min_{\substack{a\in[\underline{a},\overline{a}],\\ d\in[\underline{d},\overline{d}]}}\{d(1-a)\underline{\tau}_{\ell-1\to\ell}-a(1-d)\underline{\tau}_{\ell\to\ell+1}\}\\&=\underline{d}(1-\overline{a})\underline{\tau}_{\ell-1\to\ell}-\overline{a}(1-\underline{d})\underline{\tau}_{\ell\to\ell+1},\end{aligned}$$

where the second equality holds because it follows from Theorem 58$_{156}$ that $\underline{\tau}_{\ell-1\to\ell}$ and $\underline{\tau}_{\ell\to\ell+1}$ are positive and real-valued which implies that the minimum is obtained for the probability mass function that minimises $d(1-a)$ and maximises $a(1-d)$.

For upper expected first-passage times, it follows from Equations (6.1)$_{152}$ and (7.4)$_{192}$, Theorem 73$_{172}$ and Selection Methods UU$_L$ 1$_{172}$ and 2$_{172}$ that all we need to prove is that $\underline{w}_0=\underline{a}$ and that for all $\ell\in\{1,\dots,L-1\}$

$$\big(\overline{d}(1-\underline{a}),\overline{d}\underline{a}+(1-\overline{d})(1-\underline{a}),\underline{a}(1-\overline{d})\big)\in\operatorname*{argmax}_{\pi_\ell\in\mathscr{R}_\ell}\{b_\ell\overline{\tau}_{\ell-1\to\ell}-w_\ell\overline{\tau}_{\ell\to\ell+1}\}.$$

The fact that $\underline{w}_0=\underline{a}$ is trivial, since we have that $[\underline{w}_0,\overline{w}_0]=[\underline{a},\overline{a}]$. For all $\ell\in\{1,\dots,L-1\}$, we find that

$$\begin{aligned}\max_{\pi_\ell\in\mathscr{R}_\ell}\{b_\ell\overline{\tau}_{\ell-1\to\ell}-w_\ell\overline{\tau}_{\ell\to\ell+1}\}&=\max_{\substack{a\in[\underline{a},\overline{a}],\\ d\in[\underline{d},\overline{d}]}}\{d(1-a)\overline{\tau}_{\ell-1\to\ell}-a(1-d)\overline{\tau}_{\ell\to\ell+1}\}\\&=\overline{d}(1-\underline{a})\overline{\tau}_{\ell-1\to\ell}-\underline{a}(1-\overline{d})\overline{\tau}_{\ell\to\ell+1},\end{aligned}$$

where the second equality holds because it follows from Theorem 58$_{156}$ that $\overline{\tau}_{\ell-1\to\ell}$ and $\overline{\tau}_{\ell\to\ell+1}$ are positive and real-valued which implies that the maximum is obtained for the probability mass function that maximises $d(1-a)$ and minimises $a(1-d)$. □

Theorem 88↶ tells us that lower and upper expected upward first-passage times in an imprecise Geo/Geo/1/L queue are obtained by precise Geo/Geo/1/L queues. In particular, for any two intervals of parameters $[\underline{a},\overline{a}]$ and $[\underline{d},\overline{d}]$ such that $0<\underline{a}<\overline{a}<1$ and $0<\underline{d}<\overline{d}<1$, any lower expected upward first-passage time is obtained by the Geo/Geo/1/L queue with homogeneous parameters $\overline{a}$ and $\underline{d}$ and any upper expected upward first-passage time is obtained by the Geo/Geo/1/L queue with homogeneous parameters $\underline{a}$ and $\overline{d}$. From a computational point of view, Theorem 88↶ implies that we can directly compute lower and upper upward expected first-passage times by computing the expected upward first-passage times $\tau^{M}_{i\to j}$ and $\tau^{M'}_{i\to j}$ according to Equation $(6.48)_{170}$ and Proposition 70_{169}; see Tables 7.5_{225} and 7.6_{226} for some numerical results.

The behaviour of lower and upper expected downward first-passage times in imprecise Geo/Geo/1/L queues is analogous. This time, we find that any lower expected downward first-passage time is obtained by the Geo/Geo/1/L queue with homogeneous parameters $\underline{a}$ and $\overline{d}$ and any upper expected downward first-passage time is obtained by the Geo/Geo/1/L queue with homogeneous parameters $\overline{a}$ and $\underline{d}$.

Theorem 89. *Consider an imprecise Geo/Geo/1/L queue with parameters in intervals $[\underline{a},\overline{a}]$ and $[\underline{d},\overline{d}]$ such that $0<\underline{a}<\overline{a}<1$ and $0<\underline{d}<\overline{d}<1$, a Geo/Geo/1/L queue with transition matrix M of the form in Equation $(7.4)_{192}$ such that $a=\underline{a}$ and $d=\overline{d}$, and a Geo/Geo/1/L queue with transition matrix M' of the form in Equation $(7.4)_{192}$ such that $a=\overline{a}$ and $d=\underline{d}$. Then for all $i,j\in\mathscr{X}$ such that $i>j$, $\underline{\tau}_{i\to j}=\tau^{M}_{i\to j}$ and $\overline{\tau}_{i\to j}=\tau^{M'}_{i\to j}$.*

Proof. Starting with lower expected first-passage times, it follows from Equations $(6.1)_{152}$ and $(7.4)_{192}$, Theorem 74_{173} and Selection Methods LD_0 1_{173} and 2_{173} that all we need to prove is that $\overline{b}_L=\overline{d}(1-\underline{a})$ and that for all $\ell\in\{1,\dots,L-1\}$

$$\big(\overline{d}(1-\underline{a}),\overline{d}\underline{a}+(1-\overline{d})(1-\underline{a}),\underline{a}(1-\overline{d})\big)\in\operatorname*{argmin}_{\pi_\ell\in\mathscr{R}_\ell}\{w_\ell\underline{\tau}_{\ell+1\to\ell}-b_\ell\underline{\tau}_{\ell\to\ell-1}\}.$$

Since $b_L\in\{d(1-a)\colon a\in[\underline{a},\overline{a}]\text{ and }d\in[\underline{d},\overline{d}]\}$, we infer that $\overline{b}_L=\overline{d}(1-\underline{a})$. For all $\ell\in\{1,\dots,L-1\}$, we find that

$$\begin{aligned}\min_{\pi_\ell\in\mathscr{R}_\ell}\{w_\ell\underline{\tau}_{\ell+1\to\ell}-b_\ell\underline{\tau}_{\ell\to\ell-1}\}&=\min_{\substack{a\in[\underline{a},\overline{a}],\\ d\in[\underline{d},\overline{d}]}}\{a(1-d)\underline{\tau}_{\ell+1\to\ell}-d(1-a)\underline{\tau}_{\ell\to\ell-1}\}\\&=\underline{a}(1-\overline{d})\underline{\tau}_{\ell+1\to\ell}-\overline{d}(1-\underline{a})\underline{\tau}_{\ell\to\ell-1},\end{aligned}$$

where the second equality holds because it follows from Theorem 58_{156} that $\underline{\tau}_{\ell\to\ell-1}$ and $\underline{\tau}_{\ell+1\to\ell}$ are positive and real-valued, and therefore the minimum is obtained for the probability mass function that maximises $d(1-a)$ and minimises $a(1-d)$.

For the upper expectations, it follows from Equations $(6.1)_{152}$ and $(7.4)_{192}$, Theorem 75_{174}, Selection Methods UD_0 1_{174} and 2_{174} that all we need to prove is that $\underline{b}_L = \underline{d}(1-\overline{a})$ and that for all $\ell \in \{1,\ldots,L-1\}$

$$\big(\underline{d}(1-\overline{a}), \underline{d}\overline{a} + (1-\underline{d})(1-\overline{a}), \overline{a}(1-\underline{d})\big) \in \underset{\pi_\ell \in \mathscr{R}_\ell}{\operatorname{argmax}} \{w_\ell \overline{\tau}_{\ell+1\to\ell} - b_\ell \overline{\tau}_{\ell\to\ell-1}\}.$$

Since $b_L \in \{d(1-a) \colon a \in [\underline{a},\overline{a}] \text{ and } d \in [\underline{d},\overline{d}]\}$, we have that $\underline{b}_L = \underline{d}(1-\overline{a})$. For all $\ell \in \{1,\ldots,L-1\}$, we infer that

$$\begin{aligned}\max_{\pi_\ell \in \mathscr{R}_\ell} \{w_\ell \overline{\tau}_{\ell+1\to\ell} - b_\ell \overline{\tau}_{\ell\to\ell-1}\} &= \max_{\substack{a\in[\underline{a},\overline{a}],\\ d\in[\underline{d},\overline{d}]}} \{a(1-d)\overline{\tau}_{\ell+1\to\ell} - d(1-a)\overline{\tau}_{\ell\to\ell-1}\} \\ &= \overline{a}(1-\underline{d})\overline{\tau}_{\ell+1\to\ell} - \underline{d}(1-\overline{a})\overline{\tau}_{\ell\to\ell-1},\end{aligned}$$

where the second equality holds because it follows from Theorem 58_{156} that $\overline{\tau}_{\ell\to\ell-1}$ and $\overline{\tau}_{\ell+1\to\ell}$ are positive and real-valued which implies that the maximum is obtained for the probability mass function that minimises $d(1-a)$ and maximises $a(1-d)$. □

Similarly to the case of lower and upper expected upward first-passage times, we can directly compute lower and upper expected downward first-passage times by computing the expected downward first-passage times $\tau^M_{i\to j}$ and $\tau^{M'}_{i\to j}$ according to Equation $(6.49)_{170}$ and Proposition 71_{170}. Numerical results can again be found in Tables 7.5_{225} and 7.6_{226}.

Finally, due to Theorem 80_{180}, we infer that Theorems 88_{221} and $89_{\curvearrowleft}$ apply also for lower and upper expected first-passage times defined by the measure-theoretic approach, provided that we adopt epistemic irrelevance, complete independence or repetition independence. Moreover, since for all $i,j \in \mathscr{X}$ such that $i \neq j$, $\underline{\tau}_{i\to j}$ and $\overline{\tau}_{i\to j}$ are obtained by Geo/Geo/1/L queues, we expect that they will also coincide with the respective ones under fixed-parameter repetition independence. The following theorems clarify these statements.

Theorem 90. *Consider a Geo/Geo/1/L queue with probability tree $p \in \mathscr{T}^{\mathrm{O}}_{\mathscr{Q}}$, of which the transition matrix M is given by Equation* $(7.4)_{192}$. *Consider as well any $n \in \mathbb{N}$ and any $P \in \mathbb{P}_p$. For all $i,j \in \mathscr{X}$, it then holds that*

$$\tau^M_{i\to j} = E_P(\tau^n_{i\to j} | X_n = i).$$

Proof. Since $\mathscr{T}^{\mathrm{O}}_{\mathscr{Q}} \subseteq \mathscr{T}^{\mathrm{HM}}_{\mathscr{Q}}$, the result follows directly from Theorem 79_{179}. □

Theorem 91. *Consider the local models as defined in Section 7.2.2*$_{195}$. *Consider as well any $n \in \mathbb{N}$ and any $i, j \in \mathscr{X}$ such that $i \neq j$. Then*

$$\begin{aligned}\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) =&\underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)\\ &=\underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \underline{\tau}_{i\to j};\\ \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) =&\overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)\\ &=\overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \overline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) = \overline{\tau}_{i\to j}.\end{aligned}$$

Proof. We will only provide the proof for the lower expectations; the proof for the upper ones is completely analogous.

Due to Lemmas 85$_{198}$ and 78$_{178}$ and Theorem 33$_{98}$, it suffices to prove that $\underline{\tau}_{i\to j} = \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)$.

If $i < j$, we let M be the transition matrix of the form in Equation (7.4)$_{192}$ with $a = \overline{a}$ and $d = \underline{d}$, otherwise if $i > j$, we let M be the transition matrix of the form in Equation (7.4)$_{192}$ with $a = \underline{a}$ and $d = \overline{d}$. It then follows from Theorems 88$_{221}$ and 89$_{222}$ that

$$\underline{\tau}_{i\to j} = \tau^M_{i\to j}. \tag{7.35}$$

It also follows from Equation (7.14)$_{198}$ that there is some $q_{a,d} \in \mathscr{T}^{\mathrm{O}}_{\mathscr{Q}}$ such that $q_{a,d}(y|s,x) = M(x,y)$ for all $x, y \in \mathscr{X}$ and all $s \in \mathscr{X}^*$, where $M(x,y)$ is the element of M at row x and column y. Consider any such probability tree $q_{a,d}$, then there is a Geo/Geo/1/L queue with unique probability tree $q_{a,d}$, of which the transition matrix is M. It then follows from Theorem 90$_{\curvearrowleft}$ that for any $n \in \mathbb{N}$ and any $P \in \mathbb{P}_{q_{a,d}}$, it holds that $\tau^M_{i\to j} = E_P(\tau^n_{i\to j}|X_n = i)$, and therefore, due to Equation (7.35), that

$$\underline{\tau}_{i\to j} = E_P(\tau^n_{i\to j}|X_n = i). \tag{7.36}$$

Since $q_{a,d} \in \mathscr{T}^{\mathrm{O}}_{\mathscr{Q}}$ and $P \in \mathbb{P}_{q_{a,d}}$, we infer that $P \in \mathbb{P}^{\mathrm{O}}_{\mathscr{Q}}$. Hence, it follows from Equation (7.15)$_{198}$ that $E_P(\tau^n_{i\to j}|X_n = i) \geq \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i)$, and due to Equation (7.36), we find that

$$\underline{\tau}_{i\to j} \geq \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i).$$

For the converse inequality, we infer that

$$\underline{\tau}_{i\to j} = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i) \leq \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to j}|X_n = i),$$

where the first equality follows from Theorem 80$_{180}$ and the second inequality follows from Lemma 85$_{198}$. □

k	0	1	2	3
$\underline{\tau}_{0\to k}$	2.6666	1.6667	9.8148	28.0453
$\underline{\tau}_{1\to k}$	3.3331	1.9332	8.1481	26.3786
$\underline{\tau}_{2\to k}$	6.6656	3.3325	3.8809	18.2305
$\underline{\tau}_{3\to k}$	9.9957	6.6626	3.3301	6.7022
$\underline{\tau}_{4\to k}$	13.3160	9.9829	6.6504	3.3203
$\underline{\tau}_{5\to k}$	16.5973	13.2642	9.9316	6.6016
$\underline{\tau}_{6\to k}$	19.7223	16.3892	13.0566	9.7266
$\underline{\tau}_{7\to k}$	22.2223	18.8892	15.5566	12.2266
$\overline{\tau}_{0\to k}$	6.7278	2	20	102
$\overline{\tau}_{1\to k}$	9.5463	3.2330	18	100
$\overline{\tau}_{2\to k}$	18.8405	9.2942	9.0902	82
$\overline{\tau}_{3\to k}$	27.7425	18.1963	8.9021	34.6292
$\overline{\tau}_{4\to k}$	36.0347	26.4884	17.1942	8.2921
$\overline{\tau}_{5\to k}$	43.3779	33.8317	24.5375	15.6354
$\overline{\tau}_{6\to k}$	49.2453	39.6990	30.4048	21.5028
$\overline{\tau}_{7\to k}$	52.8167	43.2705	33.9763	25.0742

Table 7.5: Lower and upper expected first-passage and return times $\underline{\tau}_{i\to k}$ and $\overline{\tau}_{i\to k}$ for all $i \in \mathscr{X}$ and all $k \in \{0,\dots,3\}$.

7.8.2 *Expected return times*

We close this chapter by discussing lower and upper expected return times in imprecise Geo/Geo/1/L queues under the different independence concepts. Having calculated all possible lower and upper expected first-passage times in Section 7.8.1$_{220}$, we now calculate lower and upper expected return times using Equations (6.40)$_{167}$—(6.45)$_{168}$. The obtained lower and upper expected first-passage and return times are given in Tables 7.5 and 7.6$_{\curvearrowright}$. As we know from Theorem 80$_{180}$, this approach is valid for the independence concepts of epistemic irrelevance, complete and repetition independence.

Regarding fixed-parameter repetition independence, we know from Equations (7.15)$_{198}$ and (7.15)$_{198}$ that global lower and upper expected return times are obtained by precise Geo/Geo/1/L queues. Furthermore, it follows from Theorem 90$_{223}$ that for any precise Geo/Geo/1/L queue we can compute the expected return time using its transition matrix instead of its corresponding conditional probability measure. Therefore, for each selection of arrival and departure probabilities from the sets presented in Section 7.4$_{200}$ we calculate the expected return time using Equations (6.50)$_{170}$ and (6.51)$_{170}$. Among the calculated expected return times, the smallest value serves as an approximation for the global lower expected return time and the largest as an approximation for the global upper expected return time. The obtained lower and upper ex-

k	4	5	6	7
$\underline{\tau}_{0\to k}$	61.9593	120.2700	216.5312	371.8262
$\underline{\tau}_{1\to k}$	60.2926	118.6034	214.8645	370.1596
$\underline{\tau}_{2\to k}$	52.1445	110.4552	206.7163	362.0114
$\underline{\tau}_{3\to k}$	33.9140	92.2248	188.4859	343.7810
$\underline{\tau}_{4\to k}$	11.0866	58.3107	154.5719	309.8669
$\underline{\tau}_{5\to k}$	3.2813	17.8895	96.2611	251.5562
$\underline{\tau}_{6\to k}$	6.4063	3.125	28.4031	155.2951
$\underline{\tau}_{7\to k}$	8.9063	5.625	2.5	44.4826
$\overline{\tau}_{0\to k}$	440	1802	7260	29102
$\overline{\tau}_{1\to k}$	438	1800	7258	29100
$\overline{\tau}_{2\to k}$	420	1782	7240	29082
$\overline{\tau}_{3\to k}$	338	1700	7158	29000
$\overline{\tau}_{4\to k}$	136.9343	1362	6820	28662
$\overline{\tau}_{5\to k}$	7.3433	546.3867	5458	27300
$\overline{\tau}_{6\to k}$	13.2106414	5.8673	2184.5571	21842
$\overline{\tau}_{7\to k}$	16.7821	9.4388	3.5714	8737.8

Table 7.6: Lower and upper expected first-passage and return times $\underline{\tau}_{i\to k}$ and $\overline{\tau}_{i\to k}$ for all $i \in \mathscr{X}$ and all $k \in \{4,\ldots,7\}$.

pected return times under fixed-parameter repetition independence are given in Table 7.7$_{\curvearrowright}$. We also provide—between parentheses—the probabilities of arrival and departure (a,d), for which the lower or upper expectation was obtained.

For $k \in \{0,7\}$, by comparing the results in Tables 7.5$_{\curvearrowleft}$ and 7.6 with these in Table 7.7$_{\curvearrowright}$, we see that the results under epistemic irrelevance coincide with the respective ones under fixed-parameter independence. This is due to the fact that $\underline{\tau}_{0\to 0}$ and $\overline{\tau}_{L\to L}$ are always obtained by the Geo/Geo/1/L queue whose transition matrix consists of probability mass functions that use the smallest probability of arrival and the largest probability of departure—and vice versa for $\overline{\tau}_{0\to 0}$ and $\underline{\tau}_{L\to L}$.

Theorem 92. *Consider an imprecise Geo/Geo/1/L queue with parameters in intervals* $[\underline{a},\overline{a}]$ *and* $[\underline{d},\overline{d}]$ *such that* $0<\underline{a}<\overline{a}<1$ *and* $0<\underline{d}<\overline{d}<1$*, a Geo/Geo/1/L queue with transition matrix* M *of the form in Equation* (7.4)$_{192}$ *such that* $a=\underline{a}$ *and* $d=\overline{d}$*, and a Geo/Geo/1/L queue with transition matrix* M' *of the form in Equation* (7.4)$_{192}$ *such that* $a=\overline{a}$ *and* $d=\underline{d}$*. Then* $\underline{\tau}_{0\to 0}=\tau^{M}_{0\to 0}$ *and* $\overline{\tau}_{0\to 0}=\tau^{M'}_{0\to 0}$.

Proof. Observe that

$$\underline{\tau}_{0\to 0} = 1+\underline{a}\,\underline{\tau}_{1\to 0} = 1+\underline{a}\,\tau^{M}_{1\to 0} = \tau^{M}_{0\to 0},$$

$\underline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{0\to 0}\|X_n = 0)$	2.6666 (0.5, 0.8)	$\overline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^1_{0\to 0}\|X_n = 0)$	6.7278 (0.6, 0.7)
$\underline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{1\to 1}\|X_n = 1)$	2.0942 (0.55, 0.8)	$\overline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{1\to 1}\|X_n = 1)$	3.1396 (0.6, 0.7)
$\underline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{2\to 2}\|X_n = 2)$	4.8426 (0.6, 0.72)	$\overline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{2\to 2}\|X_n = 2)$	8.5330 (0.5, 0.8)
$\underline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{3\to 3}\|X_n = 3)$	7.5971 (0.6, 0.7)	$\overline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{3\to 3}\|X_n = 3)$	34.1320 (0.5, 0.8)
$\underline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{4\to 4}\|X_n = 4)$	11.8177 (0.6, 0.7)	$\overline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{4\to 4}\|X_n = 4)$	136.5281 (0.5, 0.8)
$\underline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{5\to 5}\|X_n = 5)$	18.3831 (0.6, 0.7)	$\overline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{5\to 5}\|X_n = 5)$	546.1125 (0.5, 0.8)
$\underline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{6\to 6}\|X_n = 6)$	28.5960 (0.6, 0.7)	$\overline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{6\to 6}\|X_n = 6)$	2184.45 (0.5, 0.8)
$\underline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{7\to 7}\|X_n = 7)$	44.4826 (0.6, 0.7)	$\overline{E}^{\text{fi}}_{\mathscr{Q}}(\tau^n_{7\to 7}\|X_n = 7)$	8737.8 (0.5, 0.8)

Table 7.7: Lower and upper expected return times under fixed-parameter repetition independence.

where the first equality follows from Equation (6.40)$_{167}$ combined with the fact that $[\underline{w}_0, \overline{w}_0] = [\underline{a}, \overline{a}]$, the second equation follows from Theorem 89$_{222}$ and the third equality follows from Equation (6.50)$_{170}$.

Similarly, we find that

$$\overline{\tau}_{0\to 0} = 1 + \overline{a}\,\overline{\tau}_{1\to 0} = 1 + \overline{a}\,\tau^{M'}_{1\to 0} = \tau^{M'}_{0\to 0},$$

where the first equality follows from Equation (6.43)$_{168}$ combined with the fact that $[\underline{w}_0, \overline{w}_0] = [\underline{a}, \overline{a}]$, the second equation follows from Theorem 89$_{222}$ and the third equality follows from Equation (6.50)$_{170}$. □

Theorem 93. *Consider an imprecise Geo/Geo/1/L queue with parameters in intervals* $[\underline{a}, \overline{a}]$ *and* $[\underline{d}, \overline{d}]$ *such that* $0 < \underline{a} < \overline{a} < 1$ *and* $0 < \underline{d} < \overline{d} < 1$*, a Geo/Geo/1/L queue with transition matrix M of the form in Equation* (7.4)$_{192}$ *such that* $a = \overline{a}$ *and* $d = \underline{d}$*, and a Geo/Geo/1/L queue with transition matrix* M' *of the form in Equation* (7.4)$_{192}$ *such that* $a = \underline{a}$ *and* $d = \overline{d}$*. Then* $\underline{\tau}_{L\to L} = \tau^M_{L\to L}$ *and* $\overline{\tau}_{L\to L} = \tau^{M'}_{L\to L}$.

Proof. Observe first that

$$\underline{\tau}_{L\to L} = 1 + \underline{d}(1-\overline{a})\underline{\tau}_{L-1\to L} = 1 + \underline{d}(1-\overline{a})\tau^M_{L-1\to L} = \tau^M_{L\to L},$$

where the first equality follows from Equation $(6.41)_{167}$ combined with the fact that $\min\{d(1-a)\colon a \in [\underline{a},\overline{a}] \text{ and } d \in [\underline{d},\overline{d}]\} = \underline{d}(1-\overline{a})$, the second equality follows from Theorem 88_{221} and the third equality follows from Equation $(6.50)_{170}$.

For the upper case, we infer that

$$\overline{\tau}_{L\to L} = 1+\overline{d}(1-\underline{a})\overline{\tau}_{L-1\to L} = 1+\overline{d}(1-\underline{a})\tau^{M'}_{L-1\to L} = \tau^{M'}_{L\to L},$$

where the first equality follows from Equation $(6.44)_{168}$ combined the fact that $\max\{d(1-a)\colon a \in [\underline{a},\overline{a}] \text{ and } d \in [\underline{d},\overline{d}]\} = \overline{d}(1-\underline{a})$, the second equality follows from Theorem 88_{221} and the third equality follows from Equation $(6.50)_{170}$. □

Theorem 94. *Consider the local models as defined in Section $7.2.2_{195}$. Consider as well any $n \in \mathbb{N}$ and any $i \in \{0,L\}$. Then*

$$\begin{aligned}\underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i) =& \underline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i)\\ &= \underline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i) = \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i) = \underline{\tau}_{i\to i};\\ \overline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i) =& \overline{E}^{\mathrm{ci}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i)\\ &= \overline{E}^{\mathrm{ri}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i) = \overline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i) = \overline{\tau}_{i\to i}.\end{aligned}$$

Proof. We will only provide the proof for the lower expectations; the proof for the upper ones is completely analogous.

Due to Lemmas 85_{198} and 78_{178} and Theorem 33_{98}, it suffices to prove that $\underline{\tau}_{i\to i} = \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i)$.

If $i=0$, we let M be the transition matrix of the form in Equation $(7.4)_{192}$ with $a=\underline{a}$ and $d=\overline{d}$, otherwise if $i=L$, we let M be the transition matrix of the form in Equation $(7.4)_{192}$ with $a=\overline{a}$ and $d=\underline{d}$. It then follows from Theorems 92_{226} and $93_{\curvearrowleft}$ that

$$\underline{\tau}_{i\to i} = \tau^M_{i\to i}. \tag{7.37}$$

It also follows from Equation $(7.14)_{198}$ that there is some $q_{a,d} \in \mathscr{T}^{\mathrm{O}}_{\mathscr{Q}}$ such that $q_{a,d}(y|s,x) = M(x,y)$ for all $x,y \in \mathscr{X}$ and all $s \in \mathscr{X}^*$, where $M(x,y)$ is the element of M at row x and column y. Consider any such probability tree $q_{a,d}$, then there is a Geo/Geo/1/L queue with unique probability tree $q_{a,d}$, of which the transition matrix is M. It then follows from Theorem 90_{223} that for any $n \in \mathbb{N}$ and any $P \in \mathbb{P}_{q_{a,d}}$, it holds that $\tau^M_{i\to i} = E_P(\tau^n_{i\to i}|X_n=i)$, and therefore, due to Equation (7.37), that

$$\underline{\tau}_{i\to i} = E_P(\tau^n_{i\to i}|X_n=i). \tag{7.38}$$

Since $q_{a,d} \in \mathscr{T}^{\mathrm{O}}_{\mathscr{Q}}$ and $P \in \mathbb{P}_{q_{a,d}}$, we infer that $P \in \mathbb{P}^{\mathrm{O}}_{\mathscr{Q}}$. Hence, it follows from Equation $(7.15)_{198}$ that $E_P(\tau^n_{i\to i}|X_n=i) \geq \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i)$, and due to Equation (7.38), we find that

$$\underline{\tau}_{i\to i} \geq \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n=i).$$

For the converse inequality, we infer that

$$\underline{\tau}_{i\to i} = \underline{E}^{\mathrm{ei}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n = i) \leq \underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{i\to i}|X_n = i),$$

where the first equality follows from Theorem 80_{180} and the second inequality follows from Lemma 85_{198}. □

For all $k \in \{1,\ldots,6\}$, things are similar to what happened for the lower and upper probability of having queue length k. For example, judging by Table 7.7_{227}, $\underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{1\to 1}|X_n = 1)$ is obtained for $a = 0.55$ and $d = 0.8$. However, $\underline{\tau}_{1\to 1}$ is not necessarily obtained for homogeneous parameters a and d. For example, as we can see by comparing Table 7.5_{225} with Table 7.7_{227}, $\underline{\tau}_{1\to 1}$ is strictly smaller than $\underline{E}^{\mathrm{fi}}_{\mathscr{Q}}(\tau^n_{1\to 1}|X_n = 1)$. Hence, we conclude that fixed-parameter repetition independence does not necessarily coincide with the other independence concepts when it comes to lower and upper expected return times.

8

CONCLUSIONS

In this dissertation, we have discussed different computational methods for—making robust inferences using—lower and upper expectations in imprecise Markov chains, but also in more general imprecise stochastic processes. The imprecise stochastic processes that we considered were derived from sets of probability mass functions and typically, we can approximate lower and upper expectations by selecting a number of collections of probability mass functions and calculating the precise expectation for each of them. In most of the cases, however, there are more efficient ways for calculating lower and upper expectations, in which the sets of probability mass functions are directly used. This is because in such cases various properties were satisfied, such as the generalised law of iterated expectation, which turned out to be of significant importance for the methods we developed.

Our ability to develop methods for computing lower and upper expectations should be—at least partially—credited to the martingale-theoretic approach. Indeed, this approach allows us to prove various theoretical results which not only render sub- and supermartingales a simple and proper tool for defining lower and upper expectations, but also shed light on the similarities and differences with the standard measure-theoretic approach, something that gives us the freedom to choose our approach for building our global models, depending on the context. In many cases, the martingale-theoretic approach extends the measure-theoretic approach and offers technical advantages.

A particular feature of the martingale-theoretic approach, when applied to imprecise Markov chains, is that epistemic irrelevance is adopted by default. This is one of the important reasons why we could prove the properties that lead to efficient computations for lower and upper expectations, because they may not work under other independence concepts. On the other hand, in some cases, we were able to prove that the global models obtained under epistemic irrelevance coincide with the respective ones obtained under other independence concepts. For the rest of the cases, the global models under epistemic irrelevance can be considered as fairly good approximations for the lower and upper expectations obtained under other independence concepts; see for example Chapter 7_{189}. Another characteristic of epistemic irrelevance is that when it is assumed in an imprecise Markov process, it gives us the ability to consider stochastic processes that are not Markov chains and therefore, it opens research avenues to the field of optimisation in more general processes.

Our work has also shown that the theory of imprecise probabilities constitutes a coherent and well-defined framework with a lot of potential for developing—and also improving existing—computational methods for modelling uncertainty. Furthermore, we have seen that we can relax assumptions that frequently appear in this theory, such as closedness and convexity of the sets of probability mass functions. In this dissertation, we witnessed that by dropping these assumptions we can still prove existing powerful properties such as the generalised law of iterated expectations, and can still find cases where the measure-theoretic approach and the martingale-theoretic one coincide. Since the martingale-theoretic approach works only with lower and upper expectation operators, or equivalently, with closed and convex sets of probability mass functions, this implies that in these cases, the additional probability mass functions that need to be considered will not affect our inferences.

Our last conclusion is about the theory of stochastic processes: some things that are often taken for granted in the theory of stochastic processes should be handled with caution when imprecision takes place. We are mainly referring to our result that in the limit, for epistemic irrelevance and complete independence, the bounds on the expected value of the time average of a function can be included in the respective ones for the expected value of that function at a single point in time. From a practical point of view, since the 'worst-case' scenario in the limit may be worse than the 'average worst-case' scenario, this can lead to false predictions if—as in the precise case—we regard these scenarios as equivalent. This was illustrated in our queueing application when we calculated the (average) upper probability of 'turning on the server'.

Regarding future research and starting with discrete-time imprecise stochastic processes, there are quite a few challenges. One challenging problem is to find ways for modelling imprecise stochastic processes whose state spaces are infinite. Another challenging problem is the development of efficient methods for computing lower and upper expected first-passage and return times in imprecise birth-death chains, or other similar processes, while dropping the assumption that the local models are closed and consist of strictly positive probability mass functions. Of course, these are just some first ideas on future research on the topic of discrete-time imprecise stochastic processes. There are many more problems left that are of theoretical and/or practical interest.

The next natural thing is to see whether the results presented in this dissertation can be applied in a continuous-time regime. The first steps on imprecise continuous-time stochastic processes have been achieved, but still there are limitations on the type of inferences we can make—see References [22, 33, 48, 58, 69, 73]. Regarding imprecise continuous-time Markov chains, one particularly challenging problem could be the development of methods for efficient computations of lower and upper expectations of functions that depend on infinitely many time points. In such a problem, it would also be interesting to investigate whether we can develop a martingale-theoretic approach and also methods for computing lower and upper first-passage and return times.

Finally, it would be interesting to apply our results on imprecise stochastic processes to more complicated—discrete or continuous-time—queueing models than the imprecise Geo/Geo/1/L one. A starting point for such a research line would be imprecise Geo/Geo/c/L queueing models, where the number of servers in the queue can be more than one, or discrete-time single-server queueing models of which the arrivals and departures occur according to more general—so not necessarily geometrically distributed—probability distributions.

A

MEASURE-THEORETIC PROBABILITY

Consider a variable X whose exact value we do not know, though we know that it belongs to a set Ω. We do not impose any constraints on the size of Ω, in the sense that Ω can be either finite or countably infinite or even uncountably infinite. One common way to model our uncertainty about X is by using the framework of *measure-theoretic probability* [45], for which we provide some preliminaries in this appendix.

The essential tool for modelling uncertainty when using the framework of measure-theoretic probability is the so-called *probability space*, denoted by the triple $(\Omega, \mathscr{F}, P)$. The set Ω, in which our variable takes values, is called *sample space*. Any subset of Ω is called an event.[1] The set of all events is denoted by 2^{Ω} and $\mathscr{F}$ is a specific subset of 2^{Ω} of a type called *σ-algebra* or *σ-field*. The last element, that is P, of a probability space is a *σ-additive probability measure* and it is a function that assigns probabilities to the events in $\mathscr{F}$.

The outline of the appendix goes as follows. In Section A.1$_{\curvearrowright}$, we introduce the concepts of algebra and σ-algebra and we discuss how these can be generated from a set of events. In Section A.2$_{236}$, we introduce finitely additive and σ-additive probability measures and we discuss how we can extend a σ-additive probability measure on an algebra to a σ-additive probability measure on the generated σ-algebra. Then, in Section A.3$_{237}$, we present the types of functions for which we can compute expected values, the so-called *measurable functions*, and finally in Section A.4$_{239}$, we explain the procedure of how to compute the expected value of a measurable function.

The concepts and the ideas that are presented in this appendix are not new and they are part of the wider field of *measure theory*. Definitions, theorems,

[1]The empty set $\varnothing$ and Ω itself are also considered to be subsets of Ω.

propositions and other properties that are stated throughout this appendix can be found in References [9, 60, 82].

A.1 ALGEBRAS, σ-ALGEBRAS AND GENERATORS

We first introduce the concept of *algebra*, which is defined as follows.

Definition 11 ([9, Chapter 1, Section 2]). *Consider any non-empty sample space Ω. Then a set of events $\mathscr{F}_0 \subseteq 2^\Omega$ is called an algebra if*

A1. $\Omega \in \mathscr{F}_0$;

A2. *if $A \in \mathscr{F}_0$, then also $A^c \in \mathscr{F}_0$, with $A^c := \Omega \setminus A$;*

A3. *if $A, B \in \mathscr{F}_0$, then also $A \cup B \in \mathscr{F}_0$.*

It follows from (A3) that an algebra is closed under finite unions. We can also infer that an algebra is closed under finite intersections. For any A, B in $\mathscr{F}_0$, we have that

$$A^c, B^c \in \mathscr{F}_0 \Rightarrow A^c \cup B^c \in \mathscr{F}_0 \Rightarrow (A^c \cup B^c)^c \in \mathscr{F}_0 \Rightarrow A \cap B \in \mathscr{F}_0,$$

where the first and the third statement follows from (A2), the second from (A3) and the fourth from De Morgan's law, which says that $A \cap B = (A^c \cup B^c)^c$.

Next, we introduce the concept of a σ-algebra, which is defined as follows.

Definition 12 ([9, Chapter 1, Section 2]). *Consider any non-empty sample space Ω. Then, a set of events $\mathscr{F} \subseteq 2^\Omega$ is a σ-algebra if it is an algebra and if for all $\{A_n\}_{n\in\mathbb{N}} \in \mathscr{F}$, it holds that $\cup_{n=1}^{\infty} A_n \in \mathscr{F}$.*

From this definition, we understand that a σ-algebra is an algebra that is also closed under countably infinite unions. Using the property of countably infinite unions together with (A2), we can also show that a σ-algebra is closed under countably infinite intersections.

Let now $\mathscr{H}$ be a set of events in 2^Ω, we can then construct the smallest algebra containing $\mathscr{H}$. This algebra is denoted by $\langle \mathscr{H} \rangle$ and in fact, it is the intersection of all algebras containing $\mathscr{H}$. Note that $\langle \mathscr{H} \rangle$ is not necessarily a σ-algebra. Fortunately, we can also generate the smallest σ-algebra that contains all the events in $\mathscr{H}$. This σ-algebra is denoted by $\sigma(\mathscr{H})$ and $\sigma(\cdot)$ is called the *generator operator*. In the following example, we illustrate the concepts of $\langle \mathscr{H} \rangle$ and $\sigma(\mathscr{H})$.

Example 11. Consider the sample space consisting of infinite sequences of coin tosses, that is $\Omega = \{H, T\}^{\mathbb{N}}$. A generic element of Ω is denoted by ω. For all $n \in \mathbb{N}$, the n-th outcome of a sequence $\omega \in \Omega$ is denoted by ω_n, taking values in $\{H,T\}$, and its first n outcomes are denoted by ω^n, taking values in $\{H,T\}^n$. For all $n \in \mathbb{N}$ and all $x_{1:n} \in \{H,T\}^n$, consider now the event

$$\Gamma(x_{1:n}) := \left\{\omega \in \Omega \colon \omega^n = x_{1:n}\right\}. \tag{A.1}$$

Since there is one-to-one correspondance between $\Gamma(x_{1:n})$ and $x_{1:n}$, from now on we will use only the notation $x_{1:n}$ to indicate the event $\Gamma(x_{1:n})$. The set of all events defined by Equation (A.1) is denoted by $\{H,T\}^*$ and defined by

$$\{H,T\}^* := \left\{x_{1:n} \in \{H,T\}^n \colon n \in \mathbb{N}\right\}. \tag{A.2}$$

The algebra $\langle\{H,T\}^*\rangle$ is the set that contains all events that depend on a finite number of coin tosses. For instance, the event 'Heads at the second toss' is an event that depends on a finite number of coin tosses and it can be derived by $TH \cup HH$, where $TH, HH \in \{H,T\}^*$.

However, there are events that are not included in $\langle\{H,T\}^*\rangle$. Such an event, for example, is the event 'only Heads' denoted by H^∞. Therefore, we need the σ-algebra $\sigma(\{H,T\}^*)$, which allows us to consider events that depend on an infinite number of coin tosses. For instance, the event 'at least one occurrence of Heads' depends on an infinite number of coin tosses and is given by $\cup_{n=0}^{\infty} T^n H$ or alternatively by $\Omega \setminus T^\infty$, where $T^nH \in \{H,T\}^*$ for all $n \in \mathbb{N}$.$\Diamond$

The events defined by Equation (A.1) are known as *cylinder events* and the algebra $\langle\{H,T\}^*\rangle$ is known as the algebra generated by the cylinder events. Both the cylinder events and the algebra generated by them play a key role in the stochastic processes that we consider in Chapter 357. In general, for any finite set $\mathscr{X}$, such that $\Omega = \mathscr{X}^{\mathbb{N}}$, we denote by $\langle \mathscr{X}^* \rangle$ the algebra generated by the cylinder events of Ω, and as we will see in Section A.2↷, we can specify a probability measure on $\langle \mathscr{X}^* \rangle$, and then extend this measure to the generated σ-algebra $\sigma(\mathscr{X}^*)$.

We close this section by introducing a special type of σ-algebra that is common in measure-theoretic probability and turns out to be useful in later sections where we consider $\overline{\mathbb{R}}$-valued functions, where $\overline{\mathbb{R}}$ is the set of *extended real numbers*, defined by $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$. This σ-algebra is the so-called *Borel σ-algebra* and with respect to any topological space, it is the σ-algebra generated by the topology of that topological space. Here, we are only interested in the Borel σ-algebra on $\overline{\mathbb{R}}$, whose standard topology is the order topology associated with the natural ordering on $\overline{\mathbb{R}}$. The order topology of $\overline{\mathbb{R}}$ can also be seen as the set of all open subsets of $\overline{\mathbb{R}}$, where a set $U \subseteq \overline{\mathbb{R}}$ is an open set if for all $x \in U$

either $x \in \mathbb{R}$ and $\exists a, b \in \mathbb{R}$ such that $a < x < b$ and $(a,b) \subseteq U$
or $x = -\infty$ and $\exists b \in \mathbb{R}$ such that $x < b$ and $[-\infty, b) \subseteq U$
or $x = +\infty$ and $\exists a \in \mathbb{R}$ such that $a < x$ and $(a, +\infty] \subseteq U$.

The definition of the Borel σ-algebra on the extended real line goes as follows.

Definition 13. *Consider a sample space $\Omega = \overline{\mathbb{R}}$. Let now $\mathscr{O}$ be the set of all open subsets of $\overline{\mathbb{R}}$. Then the σ-algebra generated by $\mathscr{O}$, that is $\sigma(\mathscr{O})$, is called the* Borel σ-algebra *on $\overline{\mathbb{R}}$ and is denoted by $\mathbb{B}_{\overline{\mathbb{R}}}$.*

One useful property of $\mathbb{B}_{\overline{\mathbb{R}}}$ is that it can be generated from sets of extended real intervals. That is,

$$\begin{aligned}\mathbb{B}_{\overline{\mathbb{R}}} =& \sigma(\{[a,+\infty]\colon a \in \mathbb{R}\}) = \sigma(\{(a,+\infty]\colon a \in \mathbb{R}\}) \\ =& \sigma(\{[-\infty,b]\colon b \in \mathbb{R}\}) = \sigma(\{[-\infty,b)\colon b \in \mathbb{R}\}).\end{aligned} \tag{A.3}$$

A.2 PROBABILITY MEASURES

Given a sample space Ω and a σ-algebra $\mathscr{F}$ on it, we have a so-called *measurable space*, denoted by $(\Omega, \mathscr{F})$. In this section, we discuss assigning probabilities to σ-algebras through a so-called *σ-additive probability measure*, denoted by P, and then we will have a complete characterisation of the probability space $(\Omega, \mathscr{F}, P)$. We first introduce the concept of a σ-additive probability measure, which is defined as follows.

Definition 14 ([9, Chapter 1, Section 2])**.** *Consider any sample space Ω and an algebra $\mathscr{F}_0$ on it. Then a σ-additive probability measure is a function $P\colon \mathscr{F}_0 \to [0,1]$, which has the following properties*

Σ1. $0 \leq P(A) \leq 1$ *for all* $A \in \mathscr{F}_0$*;*

Σ2. $P(\Omega) = 1$ *and* $P(\varnothing) = 0$*;*

Σ3. *if $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of pairwise disjoint events in $\mathscr{F}_0$ with $\cup_{n=1}^{\infty} A_n$ also in $\mathscr{F}_0$, then $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.*

Another important definition, similar to Definition 14, is that of a *finitely additive probability measure*.

Definition 15 ([9, Chapter 1, Section 2])**.** *Consider any sample space Ω and an algebra $\mathscr{F}_0$ on it. Then, a finitely additive probability measure is a function $P_0\colon \mathscr{F}_0 \to [0,1]$, which has the following properties*

Σ4. $0 \leq P_0(A) \leq 1$ *for all* $A \in \mathscr{F}_0$*;*

Σ5. $P_0(\Omega) = 1$ *and* $P_0(\varnothing) = 0$*;*

Σ6. *if A, B are disjoint events in $\mathscr{F}_0$, then $P_0(A \cup B) = P_0(A) + P_0(B)$.*

Note that a σ-additive probability measure is always finitely additive but not necessarily vice versa. As it is sometimes natural to construct a finitely additive probability measure on an algebra, we might in those cases want to investigate whether this probability measure is also σ-additive. One interesting case is the algebra generated by the cylinder events, where finite additivity implies σ-additivity. This is due to the following theorem.

Theorem 95 ([9, Chapter 1, Section 2, Theorem 2.3]). *Any finitely additive probability measure on the algebra that is generated by the cylinder events is also σ-additive.*

The reason why we emphasise σ-additive probability measures on algebras is because they can always be extended to unique σ-additive probability measures on the corresponding generated σ-algebras. This is due to Carathéodory's theorem, which goes as follows.

Theorem 96 ([82, Chapter 1, Theorem 1.7]). *Consider any sample space Ω equipped with an algebra $\mathscr{F}_0$ and a σ-additive probability measure $P_0\colon \mathscr{F}_0 \to [0,1]$. Then there is a unique σ-additive probability measure P on $\sigma(\mathscr{F}_0)$ such that $P = P_0$ on $\mathscr{F}_0$.*

There are cases where the extension of a σ-additive probability measure on an algebra only allows us to compute the probabilities of a limited number of events. For instance, suppose we have the sample space $\Omega = \{H,T\}^{\mathbb{N}}$ and the algebra $\mathscr{F}_0 = \{\Omega, H, T, \varnothing\}$, where H stand for 'Heads at the first coin toss' and similarly T stands for 'Tails at the first coin toss'. Suppose as well that we have a σ-additive probability measure P_0 on $\mathscr{F}_0$. In this example, we have that $\sigma(\mathscr{F}_0) = \mathscr{F}_0$ and therefore the extension of P_0 to $\sigma(\mathscr{F}_0)$ is again P_0. On the other hand, if we consider the algebra generated by the cylinder events $\langle\{H,T\}^*\rangle$ and a σ-additive probability measure P_0 on $\langle\{H,T\}^*\rangle$, then we are able to uniquely extend this probability measure to a σ-additive probability measure on the σ-algebra $\sigma(\{H,T\}^*)$. In this way, we will be able to assign probabilities to events that belong to $\sigma(\{H,T\}^*)$ and not to $\langle\{H,T\}^*\rangle$, such as the event 'at least one occurrence of Heads'.

A.3 MEASURABLE FUNCTIONS

In the previous sections, we discussed specifying a σ-algebra $\mathscr{F}$ for a sample space Ω and a σ-additive probability measure P on $\mathscr{F}$, giving thus a clear image of what a probability space $(\Omega, \mathscr{F}, P)$ is. As our next goal is to use probability spaces for computing expectations, in this section we present some preliminaries on the type of functions for which we can compute their expectations. We first introduce the concept of a *measurable* function, which is defined as follows.

Definition 16. *Consider a measurable space $(\Omega, \mathscr{F})$ and a function $g\colon \Omega \to \overline{\mathbb{R}}$, then g is called a* measurable function *if*

$$g^{-1}(B) \in \mathscr{F} \ \text{ for all } B \in \mathbb{B}_{\overline{\mathbb{R}}},$$

where $g^{-1}(A) := \big\{\omega \in \Omega\colon\, g(\omega) \in A\big\}$ for any $A \subseteq \overline{\mathbb{R}}$.

According to Definition 16$_{\curvearrowleft}$, an extended real-valued function g is measurable with respect to a σ-algebra $\mathscr{F}$, if $g^{-1}(B) \in \mathscr{F}$ for all $B \in \mathbb{B}_{\overline{\mathbb{R}}}$. This condition can be be simplified because $\mathbb{B}_{\overline{\mathbb{R}}}$ is generated by a set of extended real intervals as shown in Equation (A.3)$_{236}$. The corresponding theorem goes as follows.

Theorem 97 ([82, Chapter 3, Section 3.2])**.** *Consider a measurable space $(\Omega, \mathscr{F})$ and any set $\mathscr{I}$ of extended real intervals such that $\sigma(\mathscr{I}) = \mathbb{B}_{\overline{\mathbb{R}}}$. Consider also a function $g \colon \Omega \to \overline{\mathbb{R}}$. Then g is a measurable function if and only if $g^{-1}(A) \in \mathscr{F}$ for all $A \in \mathscr{I}$.*

Common types of extended real-valued functions are $\inf_{n\in\mathbb{N}} g_n$, $\sup_{n\in\mathbb{N}} g_n$, $\lim_{n\to\infty} g_n$, $\liminf_{n\to\infty} g_n$ and $\limsup_{n\to\infty} g_n$, where $\{g_n\}_{n\in\mathbb{N}}$ is any sequence of real-valued measurable functions. According to the following theorem, these extended real-valued functions are measurable.

Theorem 98 ([9, Chapter 2, Section 13, Theorem 13.4])**.** *Consider any sequence $\{g_n\}_{n\in\mathbb{N}}$ of real-valued measurable functions on Ω. Then*

- *the extended real-valued functions* $\inf_{n\in\mathbb{N}} g_n$*,* $\sup_{n\in\mathbb{N}} g_n$*,* $\liminf_{n\to\infty} g_n$ *and* $\limsup_{n\to\infty} g_n$ *are measurable;*
- *if* $\lim_{n\to\infty} g_n(\omega)$ *exists for all $\omega \in \Omega$, then the extended real-valued function* $\lim_{n\to\infty} g_n$ *is measurable.*

The set of all extended real-valued measurable functions is denoted by $\mathbb{M}$ and the set of all non-negative extended real-valued measurable functions by $\mathbb{M}^+$.

We now introduce two special types of functions that belong to $\mathbb{M}$ and which will turn out to be useful when we compute expectations. The first one is the indicator already introduced for finite spaces in Chapter 2$_{36}$. We now extend its definition as follows.

Definition 17. *Consider any sample space Ω and any event $A \in 2^{\Omega}$. Then the function $\mathbb{I}_A$, defined by*

$$\mathbb{I}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases} \quad \text{for all } \omega \in \Omega$$

is called the indicator *of A.*

For the indicators, the following lemma holds.

Lemma 99 ([60, Chapter 8])**.** *Given any measurable space $(\Omega, \mathscr{F})$ and any $A \subseteq \Omega$, the indicator $\mathbb{I}_A$ is measurable if and only if $A \in \mathscr{F}$.*

Next, we introduce the so-called *simple functions*, which are defined as follows.

Definition 18. *Consider any measurable space $(\Omega, \mathscr{F})$. If there is some $n \in \mathbb{N}$, such that for all $i \in \{1, \ldots, n\}$, there are $c_i \in \mathbb{R}$ and $A_i \in \mathscr{F}$ such that*

$$h(\omega) = \sum_{i=1}^{n} c_i \mathbb{I}_{A_i}(\omega) \text{ for all } \omega \in \Omega$$

then h is called a simple function.

The set of all simple functions is denoted by $\mathbb{S}$ and the set of all non-negative simple functions by $\mathbb{S}^+$. By combining Lemma 99 with the following lemma, we find that simple functions are measurable.

Lemma 100 ([82, Chapter 3, Section 3.3]). *For all $g, g_1, g_2 \in \mathbb{M}$ and all $c \in \mathbb{R}$, it holds that*

$$g_1 + g_2 \in \mathbb{M}, \; g_1 g_2 \in \mathbb{M} \text{ and } cg \in \mathbb{M}.$$

The main advantage of simple functions is that they have a finite number of values. In addition, we can use non-negative simple functions to represent non-negative measurable functions. The following proposition clarifies this statement.

Theorem 101 ([60, Chapter 8, Theorem 8.8]). *Consider any measurable function g in $\mathbb{M}^+$. Then, there is a non-decreasing sequence of non-negative simple functions $\{h_n\}_{n \in \mathbb{N}}$, such that $\lim_{n \to +\infty} h_n(\omega) = g(\omega)$ for all $\omega \in \Omega$.*

Theorem 101 leads us to suspect that we will be able to prove interesting properties and results concerning extended real-valued measurable functions using simple functions.

A.4 EXPECTATIONS IN PROBABILITY SPACES

Now that we have presented all the essential information regarding probability spaces and measurable functions, we show how to compute unconditional expectations of measurable functions. For conditional expected values of measurable functions, we provide all the necessary information in Chapter 4_{86}, where many results are based on properties that are stated here.

In order to define the expectation of a measurable function, the standard way is to use the so-called *(Lebesgue) integral*. For any probability space $(\Omega, \mathscr{F}, P)$ and any measurable function $g \colon \Omega \to \overline{\mathbb{R}}_{\geq 0}$, where $\overline{\mathbb{R}}_{\geq 0} = \mathbb{R}_{\geq 0} \cup \{+\infty\}$, the expectation of g with respect to P is defined by

$$E_P(g) := \int_\Omega g(\omega) \mathrm{d}P(\omega) := \int_0^\infty g^*(t) \mathrm{d}t, \tag{A.4}$$

where the integral on the left-hand side is a Lebesgue integral, the integral on the right-hand side is a Riemann integral and $g^*(t) := P(\{\omega \in \Omega \colon g(\omega) > t\})$.

We move now to general measurable functions $g\colon \Omega \to \overline{\mathbb{R}}$, which—as shown in Reference [60, Chapter 10]—can be uniquely decomposed into the difference of the two non-negative measurable functions g^+ and g^- defined by

$$g^+(\omega) := \max\{g(\omega), 0\} = \begin{cases} g(\omega) & \text{if } g(\omega) > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \tag{A.5}$$
$$g^-(\omega) := \max\{-g(\omega), 0\} = \begin{cases} -g(\omega) & \text{if } g(\omega) < 0 \\ 0 & \text{otherwise} \end{cases}.$$

Under the condition that $\min\{E_P(g^+), E_P(g^-)\} < +\infty$, the expectation of g is given by

$$E_P(g) = \int_\Omega g^+(\omega)\,\mathrm{d}P(\omega) - \int_\Omega g^-(\omega)\,\mathrm{d}P(\omega). \tag{A.6}$$

Lebesgue integrals satisfy various properties and we present some basic ones adapted to the context of expectations.

Lemma 102 (Linearity; [60, Chapter 10])**.** *For any probability space $(\Omega, \mathscr{F}, P)$ and any pair of extended real-valued functions g_1, g_2 in $\mathbb{M}$ such that $g_1 + g_2$ and $g_1 - g_2$ are defined in $\overline{\mathbb{R}}$,*[2] *the following holds*

$$E_P(\alpha g_1 + \beta g_2) = \alpha E_P(g_1) + \beta E_P(g_2), \text{ for all } \alpha, \beta \in \mathbb{R}.$$

Lemma 103 (Monotonicity; [60, Chapter 10, Theorem 10.4 (iv)])**.** *Consider a probability space $(\Omega, \mathscr{F}, P)$ and a pair of extended real-valued functions g_1, g_2 in $\mathbb{M}$ such that $\min\{E_P(g_1^+), E_P(g_1^-)\} < +\infty$, $\min\{E_P(g_2^+), E_P(g_2^-)\} < +\infty$ and $g_1 \leq g_2$. Then $E_P(g_1) \leq E_P(g_2)$.*

Theorem 104 (Beppo Levi's Monotone Convergence Theorem; [60, Chapter 9, Theorem 9.6])**.** *Consider any probability space $(\Omega, \mathscr{F}, P)$, any $g \in \mathbb{M}^+$ and any non-decreasing sequence of non-negative real-valued measurable functions $\{h_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to+\infty} h_n = g$. Then*

$$\lim_{n\to+\infty} E_P(h_n) = E_P(g).$$

Since the sequence $\{h_n\}_{n\in\mathbb{N}}$ in Theorem 104 is non-decreasing and non-negative, then due to Lemma 103, we also find that

$$\lim_{n\to+\infty} E_P(h_n) = \sup_{n\in\mathbb{N}} E_P(h_n) = E_P(g). \tag{A.7}$$

Lemma 105 ([60, Chapter 9, Properties 9.3 (i)])**.** *Consider any probability space $(\Omega, \mathscr{F}, P)$, then*

$$E_P(\mathbb{I}_A) = P(A) \text{ for all } A \in \mathscr{F}.$$

[2] This means that the sum and the subtraction do not lead to $+\infty - \infty$.

Lemma 106. *Consider any probability space $(\Omega, \mathscr{F}, P)$, any $c \in \mathbb{R}$ and any g in $\mathbb{M}$ such that $g(\omega) = c$ for all $\omega \in \Omega$. It then holds that $E_P(g) = c$.*

Proof. This follows from the fact that

$$\begin{aligned} E_P(g) &= E_P(c\mathbb{I}_\Omega) = E_P(\frac{1}{2}c\mathbb{I}_\Omega + \frac{1}{2}c\mathbb{I}_\Omega) = \frac{1}{2}cE_P(\mathbb{I}_\Omega) + \frac{1}{2}cE_P(\mathbb{I}_\Omega) \\ &= cE_P(\mathbb{I}_\Omega) = cP(\Omega) = c, \end{aligned}$$

where the first equality follows from the fact that $g(\omega) = c$ for all $\omega \in \Omega$, the third from Lemma 102↶, the fifth from Lemma 105↶ and the last from property $\Sigma 2_{236}$ in Definition 14_{236}. □

We now provide a very brief explanation why it makes sense to define expected values of measurable functions by Lebesgue integrals. Given a probability space $(\Omega, \mathscr{F}, P)$, we have the following:

Step 1. For all $A \in \mathscr{F}$, it follows from Lemma 105↶ that $E_P(\mathbb{I}_A) = P(A)$;

Step 2. From Definition 18_{239}, we know that for all $h \in \mathbb{S}^+$, there is $n \in \mathbb{N}$ and for all $i \in \{1, \ldots, n\}$, there are $A_i \in \mathscr{F}$ and $c_i \in \mathbb{R}_{\geq 0}$, such that $h = \sum_{i=1}^n c_i \mathbb{I}_{A_i}$. Therefore, it follows from Lemma 102↶ and Step 1 that the expected value of h is given by

$$E_P(h) = \sum_{i=1}^n c_i E(\mathbb{I}_{A_i}) = \sum_{i=1}^n c_i P(A_i);$$

Step 3. For all $g \in \mathbb{M}^+$, it follows from Theorem 101_{239} and Equation (A.7) that the expected value $E(g)$ is given by

$$E_P(g) = \sup\left\{E_P(h) \colon h \in \mathbb{S}^+ \text{ and } h \leq g\right\}; \tag{A.8}$$

Step 4. For all $g \in \mathbb{M}$ such that $\min\{E_P(g^+), E_P(g^-)\} < +\infty$, the expected value $E(g)$ is then given by

$$E_P(g) = E_P(g^+) - E_P(g^-),$$

where each of the expectations on the right-hand side come from Equation (A.8) and g^+, g^- from Equation (A.5)↶.

For the expected value of a non-negative simple function, we just need Step 1 and Step 2. Similarly, if the function under study is a non-negative measurable function we need Step 1–Step 3 and finally if it is extended real-valued we need Step 1–Step 4.

Bibliography

[1] Attahiru S. Alfa. *Queueing Theory for Telecommunications: Discrete Time Modelling of a Single Node System*. Springer, 1st edition, 2010.

[2] Linda J.S. Allen. *An Introduction to Stochastic Processes with Applications to Biology*. Prentice Hall, 2003.

[3] Eitan Altman, Konstantin E. Avrachenkov, and Rudesindo Núñez-Queija. Perturbation analysis for denumerable Markov chains with application to queueing models. *Advances in Applied Probability*, 36:839–853, 2004.

[4] Robert B. Ash. *Real Analysis and Probability*. Academic Press, New York, 1972.

[5] Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors. *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[6] David Avis and Komei Fukuda. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete & Computational Geometry*, 8(3):295–313, 1992.

[7] Patrizia Berti, Eugenio Regazzini, and Pietro Rigo. Coherent statistical inference and Bayes theorem. *The Annals of Statistics*, pages 366–381, 1991.

[8] Patrizia Berti and Pietro Rigo. On coherent conditional probabilities and disintegrations. *Annals of Mathematics and Artificial Intelligence*, 35(1-4):71–82, 2002.

[9] Patrick Billingsley. *Probability and Measure*. Wiley, third edition, 1995.

[10] Jasper De Bock and Gert de Cooman. Credal networks under epistemic irrelevance: The sets of desirable gambles approach. *International Journal of Approximate Reasoning*, 56, Part B(0):178 – 207, 2015.

[11] James A. Bucklew, Peter Ney, and John S. Sadowsky. Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *Journal of Applied Probability*, 27:44–59, 1990.

[12] Xi-Ren Cao and Han-Fu Chen. Perturbation realization, potentials, and sensitivity analysis of markov processes. *Automatic Control, IEEE Transactions on*, 42(10):1382–1393, 1997.

[13] Mohan L. Chaudhry. On numerical computations of some discrete-time queues. In *Computational probability*, chapter 10, pages 365–408. Springer, 2000.

[14] Inés Couso, Serafín Moral, and Peter Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5(02):165–181, 2000.

[15] Fabio G. Cozman. Sets of probability distributions, independence, and convexity. *Synthese*, 186:177–200, 2012.

[16] Fabio G. Cozman. Independence for full conditional probabilities: Structure, factorization, non-uniqueness, and bayesian networks. *International Journal of Approximate Reasoning*, 54(9):1261–1278, 2013.

[17] Richard J. Crossman, Pauline Coolen-Schrijner, and Frank P.A. Coolen. Time-homogeneous birth-death processes with probability intervals and absorbing state. *Journal of Statistical Theory and Practice*, 3(1):103–118, 2009.

[18] José da Rocha and Fabio Cozman. Evidence Propagation in Credal Networks: An Exact Algorithm Based on Separately Specified Sets of Probability. *Advances in Artificial Intelligence*, pages 376–385, 2002.

[19] George B. Dantzig. Maximization of a linear function of variables subject to linear inequalities, in activity analysis of production and allocation. 1951.

[20] George B. Dantzig. *Computational algorithm of the revised simplex method*. Rand Corporation, 1953.

[21] Jasper De Bock. *Credal Networks under Epistemic Irrelevance: Theory and Algorithms*. PhD thesis, Ghent University, Faculty of Engineering and Architecture, 2015.

[22] Jasper De Bock. The limit behaviour of imprecise continuous-time Markov chains. *Journal of Nonlinear Science*, 27:159–196, 2017.

[23] Jasper De Bock and Gert de Cooman. Allowing for probability zero in credal networks under epistemic irrelevance. In *ISIPTA'13 – Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications*, pages 109–118. SIPTA, 2013.

[24] Jasper De Bock and Gert de Cooman. Continuity of imprecise stochastic processes with respect to the pointwise convergence of monotone sequences. *arXiv preprint arXiv:1402.3056*, 2017.

[25] Gert de Cooman, Jasper De Bock, and Stavros Lopatatzidis. Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems. *International Journal of Approximate Reasoning*, 76:18–46, 2016.

[26] Gert de Cooman and Filip Hermans. Imprecise probability trees: Bridging two theories of imprecise probability. *Artificial Intelligence*, 172(11):1400–1427, 2008.

[27] Gert de Cooman, Filip Hermans, Alessandro Antonucci, and Marco Zaffalon. Epistemic irrelevance in credal nets: The case of imprecise Markov trees. *International Journal of Approximate Reasoning*, 51:1029–1052, 2010.

[28] Gert de Cooman, Filip Hermans, and Erik Quaeghebeur. Imprecise Markov chains and their limit behaviour. *Probability in the Engineering and Informational Sciences*, 23(4):597–635, 2009.

[29] Bruno de Finetti. *Teoria delle probabilità*. Einaudi, Turin, 1970. English translation: [30].

[30] Bruno de Finetti. Theory of probability. John Wiley & Sons, New York, 1974.

[31] J. L. Doob. *Stochastic Processes*. John Wiley & Sons, New York, 1953.

[32] Lester E. Dubins. Finitely additive conditional probabilities, conglomerability and disintegrations. *The Annals of Probability*, 3:88–99, 1975.

[33] Alexander Erreygers and Jasper De Bock. Imprecise continuous-time markov chains: efficient computational methods with guaranteed error bounds. *Accepted for publication in ISIPTA '17 – Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, 2017.

[34] Donald Gross, John F. Shortle, James M. Thompson, and Carl M. Harris. *Fundamentals of Queueing Theory*. Wiley-Interscience, 2008.

[35] Donald Gross, John F. Shortle, James M. Thompson, and Carl M. Harris. *Fundamentals of queueing theory*. John Wiley & Sons, 2013.

[36] Branko Grünbaum. *Convex Polytopes*. Springer, 1967.

[37] Darald J. Hartfiel. *Markov Set-Chains*. Number 1695 in Lecture Notes in Mathematics. Springer, Berlin, 1998.

[38] Filip Hermans and Gert de Cooman. Characterisation of ergodic upper transition operators. *International Journal of Approximate Reasoning*, 53(4):573–583, 2012.

[39] Peter J. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

[40] Christiaan Huygens. *Van Rekeningh in Spelen van Geluck*. 1656–1657.

[41] Cassius T. Ionescu Tulcea. Mesures dans les espaces produits. *Atti Accad. Naz. Lincei Rend*, 7:208–211, 1949.

[42] Narendra Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 302–311. ACM, 1984.

[43] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. Undergraduate Text in Mathematics. Springer-Verlag, New York, 1976.

[44] Alexander Khintchine. Korrelationstheorie der stationären stochastischen prozesse. *Mathematische Annalen*, 109:604–615, 1934.

[45] Andrey Nikolaevich Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. English translation: [46].

[46] Andrey Nikolaevich Kolmogorov. *Foundations of Probability*. Chelsea Publishing Co., New York, 1950.

[47] Igor O. Kozine and Lev V. Utkin. Interval-valued finite Markov chains. *Reliable Computing*, 8(2):97–113, 2002.

[48] Thomas Krak, Jasper De Bock, and Arno Siebes. Efficient computation of updated lower expectations for imprecise continuous-time hidden markov chains. *Accepted for publication in ISIPTA '17 – Proceedings of the Tenth International Symposium on Imprecise Probability: Theories and Applications*, 2017.

[49] Stavros Lopatatzidis, Jasper De Bock, and Gert de Cooman. First steps towards Little's law with imprecise probabilities. In *ISIPTA '13 – Proceedings of the Eight International Symposium on Imprecise Probability: Theories and Applications*, pages 392–393. SIPTA, 2013.

[50] Stavros Lopatatzidis, Jasper De Bock, and Gert de Cooman. Calculating bounds on expected return and first passage times in finite-state imprecise birth-death chains. In *ISIPTA '15 – Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*, pages 177–186. SIPTA, 2015.

[51] Stavros Lopatatzidis, Jasper De Bock, and Gert de Cooman. Computational methods for imprecise continuous-time birth-death processes: a preliminary study of flipping times. In *ISIPTA '15 – Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*. SIPTA, 2015.

[52] Stavros Lopatatzidis, Jasper De Bock, and Gert de Cooman. Computing lower and upper expected first-passage and return times in imprecise birth–death chains. *International Journal of Approximate Reasoning*, 80:137–173, 2017.

[53] Stavros Lopatatzidis, Jasper De Bock, Gert de Cooman, Stijn De Vuyst, and Joris Walraevens. Robust queueing theory: an initial study using imprecise probabilities. *Queueing Systems*, 82:75–101, 2016.

[54] David G. Luenberger. *Introduction to Dynamic Systems. Theory, Models & Applications*. John Wiley & Sons, New York, 1979.

[55] Andrey Andreyevich Markov. Extension of the law of large numbers to dependent quantities (in Russian). *Izvestiia Fiz.-Matem. Obsch. Kazan Univ. (2nd Ser)*, 15:135–156, 1906.

[56] James R. Norris. *Markov chains*. Cambridge university press, 1998.

[57] Eugenio Regazzini. Finitely additive conditional probabilities. *Rendiconti del Seminario Matematico e Fisico di Milano*, 55(1):69–89, 1985.

[58] Cristina Rottondi, Alexander Erreygers, Giacomo Verticale, and Jasper De Bock. Modelling Spectrum Assignment in a Two-Service Flexi-Grid Optical Link with Imprecise Continuous-Time Markov Chains. In *13th international conference on Design of Reliable Communication Networks (DRCN 2017)*, 2017.

[59] Reuven Y. Rubinstein. Optimization of computer simulation models with rare events. *European Journal of Operational Research*, 99(1):89–112, 1997.

[60] René L. Schilling. *Measures, integrals and martingales*. Cambridge University Press, 2005.

[61] Paul J. Schweitzer. Perturbation theory and finite Markov chains. *Journal of Applied Probability*, 5:401–413, 1968.

[62] Teddy Seidenfeld. Conditional independence, imprecise probabilities, null-events and graph-theoretic models. Seimour Geisser Distinguished Lecture, School of Statistics. *University of Minnesota, Minneapolis, MN*, 2007.

[63] Glenn Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.

[64] Glenn Shafer. *The Art of Causal Conjecture*. The MIT Press, 1996.

[65] Glenn Shafer, Vladimir Vovk, and Akimichi Takemura. Lévy's zero–one law in game-theoretic probability. *Journal of Theoretical Probability*, 25:1–24, 2012.

[66] Glenn Shafer and Volodya Vovk. *Probability and Finance: It's Only a Game!* Wiley, 2001.

[67] Albert N. Shiryaev. *Probability: Volume 1 (Graduate Texts in Mathematics)*. 1996.

[68] Damjan Škulj. Discrete time Markov chains with interval probabilities. *International Journal of Approximate Reasoning*, 50(8):1314–1329, 2009.

[69] Damjan Škulj. Efficient computation of the bounds of continuous time imprecise Markov chains. *Applied Mathematics and Computation*, 250:165–180, 2015.

[70] Daniel W Stroock. *An introduction to Markov processes*, volume 230. Springer Science & Business Media, 2013.

[71] Ushio Sumita and Yasushi Masuda. On first passage time structure of random walks. *Stochastic processes and their applications*, 20(1):133–147, 1985.

[72] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley, 2014.

[73] Matthias C. M. Troffaes, Jacob Gledhill, Damjan Škulj, and Simon Blake. Using imprecise continuous time Markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair. In *ISIPTA '15 – Proceedings of the Ninth International Symposium on Imprecise Probability: Theories and Applications*. SIPTA, 2015.

[74] Paolo Vicig, Marco Zaffalon, and Fabio G. Cozman. Notes on "Notes on conditional previsions". *International Journal of Approximate Reasoning*, 44(3):358–365, 2007.

[75] Damjan Škulj. Finite discrete time Markov chains with interval probabilities. In J. Lawry, E. Miranda, A. Bugarin, S. Li, M. A. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Soft Methods for Integrated Uncertainty Modelling*, pages 299–306. Springer, Berlin, 2006.

[76] Damjan Škulj. Regular finite Markov chains with interval probabilities. In *ISIPTA '07 – Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications*, pages 405–413. SIPTA, 2007.

[77] Damjan Škulj and Robert Hable. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, 76(1):107–133, 2013.

[78] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1991.

[79] Kurt Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24:149–170, 2000.

[80] John A. White. *Analysis of queueing systems*. Elsevier, 2012.

[81] Peter Whittle. *Probability via Expectation*. Springer, 2000.

[82] David Williams. *Probability with Martingales*. Cambridge University Press, 1991.

[83] Peter M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975.

[84] Peter M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44:366–383, 2007.

[85] Li Xia and Xi-Ren Cao. Performance optimization of queueing systems with perturbation realization. *European Journal of Operational Research*, 218:293–304, 2012.

[86] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.