International Journal of Approximate Reasoning 147 (2022) 78–124

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Markovian imprecise jump processes: Extension to measurable variables, convergence theorems and algorithms

Alexander Erreygers *, Jasper De Bock

Foundations Lab for imprecise probabilities, ELIS Department, Ghent University, Technologiepark 125, 9052 Gent, Belgium

ARTICLE INFO

Article history:
Received 12 February 2022
Received in revised form 13 May 2022
Accepted 19 May 2022
Available online 23 May 2022

Keywords:
Jump process
Markov process
Monotone convergence
Dominated convergence
Imprecision
Algorithm

ABSTRACT

The existing framework of Markovian imprecise jump processes, also known as imprecise continuous-time Markov chains, is limited to bounded real variables that depend on the state of the system at a finite number of (future) time points. This is an issue in many applications, because typically the variables of interest depend on the state of the system at *all* time points in some – possibly unbounded – (future) interval, and they can be unbounded or even extended real valued; examples of such variables are temporal averages, the number of (selected) jumps in some interval and hitting times. To eliminate this shortcoming, we assume that the sample paths are càdlàg and use measure theory to extend the domain of Markovian imprecise jump processes to extended real-valued variables that may depend on the state of the system at all (future) time points – that is, the extended real variables that are bounded below or above and are measurable with respect to the σ-algebra generated by the cylinder events. We investigate the continuity properties of the extended lower and upper expectations with respect to pointwise convergent sequences, and this yields generalisations of the Monotone Convergence Theorem and Lebesgue's Dominated Convergence Theorem. For two particular classes of variables, we strengthen these convergence theorems and present an iterative scheme to approximate their lower and upper expectations. The first class is the number of selected jumps in some interval, and the second class are real variables that take the form of a Riemann integral over some interval; this second class includes temporal averages and occupancy times.

1. Introduction

Recently, several authors have independently proposed generalisations of Markovian jump processes – also called continuous-time Markov chains or Markov processes – that provide an elegant way of dealing with parameter uncertainty [1–3]. Whereas a (homogeneous) Markovian jump process is uniquely defined by its rate matrix and initial probability mass function, these 'imprecise' generalisations allow for partially specified parameters: they are defined through sets of rate matrices and/or sets of initial probability mass functions.

There are two frameworks that obtain similar, and to some extent even equivalent, results. The first framework is the one advanced by Škulj [1] and Krak et al. [2], who use the theory of imprecise probabilities [4]. Škulj [1] starts from a set $\mathcal{Q}$ of rate matrices (one that is convex and compact and has separately specified rows, to be exact), defines an 'imprecise

* Corresponding author.
E-mail address: alexander.erreygers@ugent.be (A. Erreygers).

https://doi.org/10.1016/j.ijar.2022.05.006
0888-613X/

continuous-time Markov chain' as a 'random process' whose (time-dependent) 'rate matrix' is an unspecified function of time such that it belongs to the set $\mathcal{Q}$ at all times – although he never formalises what he means with a 'random process' and its (time-dependent) 'rate matrix' – and explains how one can compute lower and upper bounds on the corresponding expectation of variables that depend on the state of the system at a single time point. Krak et al. [2] put this work on a more sound theoretical footing, and significantly extend it. First, they formally define 'continuous-time stochastic processes' as coherent conditional probabilities [5–7] on a specific domain, and for these 'continuous-time stochastic processes' they introduce the notions of consistency with a set $\mathcal{M}$ of initial probability mass functions and a (bounded) set $\mathcal{Q}$ of rate matrices. Second, they define 'imprecise continuous-time Markov chains' as sets of 'continuous-time stochastic processes' that are consistent with sets $\mathcal{M}$ and $\mathcal{Q}$. In particular, they consider three such sets: the set of all consistent 'homogeneous continuous-time Markov chains', the set of all consistent (not necessarily homogeneous) 'continuous-time Markov chains' and the set of all consistent (not necessarily Markovian) 'continuous-time stochastic processes'. Third, they provide algorithms to determine (tight) lower and upper bounds on the corresponding expectations of variables that depend on the state of the system at a finite number of (future) time points. Nendel's [3] approach, on the other hand, is situated in the theory of non-linear (or convex) expectations – see also [8]. Leaning on a type of 'Kolmogorov Extension Theorem' for convex expectations [9, Theorems 4.6 and 5.6], he shows that any 'convex Q-operator' corresponds to a 'convex Markov chain', which is a convex expectation on the bounded measurable variables with respect to the product σ-algebra generated by the canonical process. However, both of these frameworks have crucial shortcomings: that of Škulj [1] and Krak et al. [2] only deals with lower and upper expectations of variables that depend on the state of the system at a single time point or at a finite number of time points, respectively, while that of Nendel [3] only deals with bounded variables that are measurable with respect to the product σ-algebra. For applications, this implies that for both of these frameworks, key inferences like (lower and upper) until probabilities, expected temporal averages, expected occupancy times, expected hitting times – also called expected first-passage times – and the expected number of (selected) jumps are not included in the domain.

To the best of our knowledge, this shortcoming of the theory of imprecise jump processes has only been circumvented by Troffaes et al. [10] – although Erreygers and De Bock [11] do something similar in the context of lumping. Troffaes et al. [10] use an imprecise jump process to assess the reliability of a power network, and use the corresponding 'limit lower/upper expectation' to heuristically obtain conservative bounds on the – theoretically undefined – lower/upper expectation of two classes of variables that depend on the state of the system at more than a finite number of time points: (i) occupancy times, or variables that represent the amount of time spent in a (set of) state(s) during some time period; and (ii) variables that represent the number of jumps – sometimes also called visits or transitions – to a (set of) state(s) during some time period. In contrast, we take it upon us to get rid of this shortcoming in a theoretically sound manner. More precisely, we extend the domain of imprecise jump processes so that the lower/upper expectation of more general variables is well defined, and we investigate the theoretical properties of this extension. Furthermore, we provide computational methods for the two aforementioned classes of variables, and show that these methods outperform the heuristics of Troffaes et al. [10], at least when it comes to tightness of the bounds.

The remainder of this contribution is structured as follows. In Section 2, we introduce jump processes in general and Markovian jump processes in particular, and we briefly introduce imprecise jump processes in Section 3. With these preliminaries out of the way, we set out to extend the domain of (imprecise) jump processes in Section 4. We start in Section 4.2 with extending the domain of a single countably additive jump process through Lebesgue integration, and show in Section 4.3 that this extended domain includes two important classes of variables: the number of (selected) jumps over a finite time period in Lemma 28 and integrals of a function of the state over a finite time period in Lemma 32. A crucial difference between our approach to jump processes and the predominant (measure-theoretic) one is that ours takes the càdlàg paths as a starting point, whereas the latter usually starts from the set of all paths to end up with the set of all càdlàg paths through a so-called modification. Next, Section 4.4 extends the domain of (the lower and upper expectations corresponding to) the three types of Markovian imprecise jump processes that we consider. The reason why we can do this is that by Theorem 33, any jump process that is consistent with a (bounded) set of rate matrices is countably additive. Hence, we can extend the domain of each jump process in the Markovian imprecise jump process through the aforementioned extension method for countably additive jump processes, and subsequently take the lower and upper envelopes of these extensions. In Section 5, we investigate the convergence properties of the lower and upper expectations corresponding to imprecise jump processes. Initially, we generalise two well-known convergence theorems to the context of imprecise jump processes: Theorem 36 generalises the Monotone Convergence Theorem – see Theorem 34 – while Theorem 37 is a generalisation of Lebesgue's Dominated Convergence Theorem – see Theorem 35. While the limits in these convergence theorems provide conservative bounds in general, Corollaries 45 and 48 establish that these bounds are actually tight for the two aforementioned classes of variables. We then use these convergence properties in Section 6 to propose iterative computational methods – see Theorems 65 and 66 – to determine the lower and upper expectations of variables in the two aforementioned classes. Section 7 concludes this contribution.

This contribution is the significantly extended journal version of [12], and contains many results that are also available in the first author's doctoral dissertation [13]. That said, the results in Sections 5.1 and Theorem 65 in Section 6.2 are entirely new. In order not to make this contribution unnecessarily long, we have chosen to omit our proof for some of the results that are not novel – most notably those of Theorems 33 and 39 – and to refer to the relevant result in [13].

2. Jump processes

A *stochastic process* is a model of someone's uncertainty about (the evolution of) the state of some system over time. In this contribution, we consider a generic system that evolves over continuous time whose state takes on values in a finite set; following Gikhman and Skorokhod [14] and Le Gall [15] – to name just a few – we call a stochastic process for such a system a *jump process*. We denote the state space of the system by $\mathcal{X}$; throughout this contribution, except in the examples, $\mathcal{X}$ is a fixed non-empty and finite set.

Example 1. Throughout this contribution, we turn to the (imprecise) jump process model used in [10] to illustrate our results. Troffaes et al. [10] construct this model to assess the reliability of a power network; they follow up on their earlier work [16] and consider a power network that consists of two power lines, called A and B. The reason why there are two is redundancy: the network works as long as at least one of the two power lines is working. Thus, an independent failure of one of the two power lines is not that much of an issue, because it does not cause a power outage as long as the other power line is in operation. If both power lines fail due to the same cause, this does result in a power outage; whenever this occurs, we speak of a *common cause failure*. Knowing this, the relevant state space is clearly $\mathcal{X} \coloneqq \{\mathtt{AB}, \mathtt{A}, \mathtt{B}, \mathtt{F}\}$, where the state F corresponds to a failure of both power lines and where the other state labels indicate the power lines that are working.

In the measure-theoretical tradition – see, for example, [14,15,17–19] – a jump process is made up of four ingredients: a sample space $\mathfrak{S}$, a σ-algebra of events $\mathcal{S}$ over this sample space, a probability measure P on this σ-algebra and a family of $\mathcal{S}/2^{\mathcal{X}}$-measurable variables $(X_t)_{t\in\mathbb{R}_{\geq 0}}$, where $2^{\mathcal{X}}$ denotes the power set of $\mathcal{X}$. Krak et al. [2] use a slightly different structure: they use a coherent conditional probability [5–7] on the specific domain of 'cylinder events' over some sufficiently rich subset of $\mathcal{X}^{\mathbb{R}_{\geq 0}}$, where we denote the set of all maps from $\mathbb{R}_{\geq 0}$ to $\mathcal{X}$ by $\mathcal{X}^{\mathbb{R}_{\geq 0}}$. We set out to marry both of these approaches so that we can keep the best of both worlds, and we will get down to this in Section 4 further on. In the present section, we will briefly recapitulate Krak et al. [2] their approach to jump processes.

We start in Section 2.1 by introducing the set of càdlàg paths, which will be our sample space. Next, Section 2.2 defines the necessary notation and terminology regarding cylinder events, which are those events that depend on the state of the system at a (finite) number of time points. After introducing general coherent conditional probabilities in Section 2.3, we can define jump processes as coherent conditional probabilities with a specific domain in Section 2.4, and explain in Section 2.5 how a jump process corresponds to a conditional expectation operator whose domain is the set of real-valued variables that depend on the state of the system at a finite number of (future) time points. Finally, Section 3.2 deals with the important special case of homogeneous Markovian jump processes.

2.1. Càdlàg paths

Because the system evolves in continuous time, an outcome in the sample space is a *path* $\omega\colon \mathbb{R}_{\geq 0} \to \mathcal{X}$, where $\omega(t)$ is the state of the system at the time point t in $\mathbb{R}_{\geq 0}$.[1] In general, a path ω can display some pretty erratic behaviour; take, for example, the path ω_{e} that assumes the state $\omega_{\mathrm{e}}(t) \coloneqq x$ whenever the time point t is a rational number and the state $\omega_{\mathrm{e}}(t) \coloneqq y$ otherwise, with x, y in $\mathcal{X}$ such that $x \neq y$. In many applications, including in the setting of Example 1, this erratic behaviour is clearly infeasible. Hence, we choose to exclude this erratic behaviour, and we do so as follows: we only include the path ω in our sample space if (i) it stays in the new state for some time directly after it changes states, and (ii) it only changes states a finite number of times over any finite time horizon. Mathematically, this translates to requiring that (i) the path ω is continuous from the right at all time points t in $\mathbb{R}_{\geq 0}$, and (ii) the path ω has a limit from the left at all time points t in $\mathbb{R}_{>0}$ – both with respect to the discrete topology on $\mathcal{X}$. Whenever the path ω satisfies these two requirements, it is called a *càdlàg path*; see, for example, [20, Section 12] or [15, p. 54]. We collect all càdlàg paths in the set Ω, which is sometimes called the *Skorokhod space*, and this will be our sample space.

Krak et al. [2] do not assume càdlàg paths a priori, for the simple reason that this assumption is not relevant in their setting. Instead, they consider any set of paths Ω' – that is, any (non-empty) subset of $\mathcal{X}^{\mathbb{R}_{\geq 0}}$ – such that

$$(\forall n \in \mathbb{N})(\forall t_1, \ldots, t_n \in \mathbb{R}_{\geq 0}\colon t_1 < \cdots < t_n)(\forall x_1, \ldots, x_n \in \mathcal{X})(\exists \omega \in \Omega')\ \omega(t_1) = x_1, \ldots, \omega(t_n) = x_n. \tag{1}$$

However, when extending the domain to extended real variables that depend on the state of the system at all (future) time points, which we will do in Section 4 further on, it is absolutely essential that all paths in the sample space are càdlàg. It is almost trivial to verify that the set of càdlàg paths Ω satisfies the requirement in Eqn. (1) [13, Lemma 3.5]; hence, our more restricted set up falls squarely in the scope of Ref. [2].

[1] We denote the set of real numbers, non-negative real numbers and positive real numbers by $\mathbb{R}$, $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$, respectively. Furthermore, $\mathbb{N}$ denotes the natural numbers (or positive integers) and $\overline{\mathbb{R}} \coloneqq \mathbb{R} \cup \{-\infty, +\infty\}$ the extended real numbers.

2.2. Cylinder events

Let us identify some events – that is, subsets of the sample space Ω – that play an essential role in our analysis. The most elementary events are those that depend on the state of the system at a single point in time. For any time point t in $\mathbb{R}_{\geq 0}$ and any state x in $\mathcal{X}$, we denote the event that 'the state of the system at time t is x' by

$$\{X_t = x\} \coloneqq \{\omega \in \Omega \colon \omega(t) = x\}.$$

In a similar fashion, we let

$$\{X_t \in B\} \coloneqq \{\omega \in \Omega \colon \omega(t) \in B\} = \bigcup_{x \in B} \{X_t = x\}$$

for any subset B of $\mathcal{X}$; in line with this notation, we let $X_t \colon \Omega \to \mathcal{X}$ be the projector defined by

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

To simplify the notation regarding events that depend on more than a single time point, we use notational conventions similar to those introduced by Krak et al. [2, Section 2.1]. A sequence of time points is a finite sequence of increasing time points, that is, a sequence $(t_1, \ldots, t_n)$ in $\mathbb{R}_{\geq 0}$ of arbitrary length – with n in $\mathbb{N}$ – such that $t_1 < \cdots < t_n$. For the sake of brevity, we denote a generic sequence by u, v or w. We collect all sequences of time points in $\mathcal{U}_{\text{ne}}$, and let $\mathcal{U} \coloneqq \mathcal{U}_{\text{ne}} \cup \{()\}$, where $()$ denotes the empty sequence. We denote the first and last time points of a non-empty sequence of time points $u = (t_1, \ldots, t_n)$ by $\min u \coloneqq t_1$ and $\max u \coloneqq t_n$, respectively; in order to conveniently deal with the edge case that u is the empty sequence of time points $()$, we let $\min() \coloneqq 0 =: \max()$. For u and v in $\mathcal{U}$, we write $u \preccurlyeq v$ whenever v only contains time points in or succeeding u, in the sense that every time point t in v belongs to u or to $[\max u, +\infty[$; note that $() \preccurlyeq v$ for all v in $\mathcal{U}$. Similarly, for all u, v in $\mathcal{U}_{\text{ne}}$, we write $u \prec v$ whenever v only contains time points succeeding those in u, in the sense that $\max u < \min v$; out of convenience, we let $() \prec v$ for all v in $\mathcal{U}$. With this convention, for any t in $\mathbb{R}_{\geq 0}$, we let $\mathcal{U}_{\prec t} \coloneqq \{u \in \mathcal{U} \colon u \prec (t)\}$ be the set of all sequences of time points of which the last time point precedes t; if $t = 0$, then there is no such non-empty sequence, so $\mathcal{U}_{\prec t} = \{()\}$. Because a sequence of time points is an ordered set, we may use set-theoretic operations on sequences of time points, in the understanding that the result of such an operation is again a sequence of time points; for example, for all u, v in $\mathcal{U}$, $u \cup v$ denotes the sequence of time points that is made up of the time points in u and v. Finally, for any sequence of time points $u = (t_1, \ldots, t_n)$ in $\mathcal{U}_{\text{ne}}$, we let $\mathcal{X}_u \coloneqq \bigtimes_{t \in u} \mathcal{X}$ be the set of all n-tuples $x_u = (x_{t_1}, \ldots, x_{t_n})$ of states in $\mathcal{X}$, and we always index these tuples with the time points t_1, …, t_n. If u is the empty sequence $()$, then we let $\mathcal{X}_u = \mathcal{X}_{()}$ denote the singleton containing the empty tuple, denoted by $x_{()}$.

Fix some $v = (t_1, \ldots, t_n)$ in $\mathcal{U}_{\text{ne}}$. Then we let $X_v \colon \Omega \to \mathcal{X}_v$ be the projector defined by

$$X_v(\omega) \coloneqq \big(\omega(t_1), \ldots, \omega(t_n)\big) \quad \text{for all } \omega \in \Omega.$$

Furthermore, for any $B \subseteq \mathcal{X}_v$, we define the corresponding event

$$\{X_v \in B\} \coloneqq \{\omega \in \Omega \colon (\omega(t_1), \ldots, \omega(t_n)) \in B\};$$

an event of this form is called a *cylinder event* [17, Section 36]. In order to reduce the number of edge cases, we also let $\{X_{()} = x_{()}\} \coloneqq \Omega =: \{X_{()} \in \mathcal{X}_{()}\}$.

For any u in $\mathcal{U}$, Krak et al. [2, Section 4.2] let $\mathcal{A}_u$ be the set of events that consists of the cylinder events for all sequences v with time points in or succeeding u:

$$\mathcal{A}_u \coloneqq \big\{\{X_v \in B\} \colon v \in \mathcal{U}, u \preccurlyeq v, B \subseteq \mathcal{X}_v\big\}.$$

Crucially, $\mathcal{A}_u$ is an algebra of events[2] [13, Lemma 3.10]. In order to shorten our notation, we will leave out the subscript u whenever u is the empty sequence of time points: $\mathcal{A} \coloneqq \mathcal{A}_{()}$.

2.3. Coherent conditional probabilities

In order to deal with conditioning in an unambiguous manner, Krak et al. [2] resort to the framework of coherent conditional probabilities. What follows is a brief introduction to coherent conditional probabilities; we refer to Dubins [7] and Regazzini [6] for a more detailed exposition. Throughout this section, we let $\mathfrak{S}$ be a generic non-empty set.

First, let us revisit the issue of conditioning for probability charges. Given an algebra $\mathcal{S}$ on $\mathfrak{S}$, a *probability charge P on $\mathcal{S}$* is a non-negative real-valued function P on $\mathcal{S}$ that is normed and additive, in the sense that $P(\mathfrak{S}) = 1$ and $P(A \cup B) =$

[2] For any generic sample space $\mathfrak{S}$, an *algebra (of events) on* $\mathfrak{S}$ – also called a field (of events) on $\mathfrak{S}$ – is a collection $\mathcal{S}$ of subsets of $\mathfrak{S}$ which includes the empty set $\varnothing$ and is closed under taking complements and unions; see [21, Definition 1.1.1], [17, p. 19] or [4, Definition 1.6].

$P(A) + P(B)$ for all disjoint events A, B in $\mathcal{S}$. For such a probability charge P, conditioning is dealt with through Bayes's rule: for any event C such that $P(C) > 0$, the corresponding map

$$P(\bullet \mid C) \colon \mathcal{S} \to \mathbb{R}_{\geq 0} \colon A \mapsto P(A \mid C) := \frac{P(A \cap C)}{P(C)}$$

is again a probability charge on the domain $\mathcal{S}$ of P. For any non-empty event C in $\mathcal{S}$ with probability zero, the corresponding conditional probabilities are left undefined. Measure theory has one way to mitigate this issue with conditioning on events with probability zero, through the Radon-Nikodym Theorem [18, Chapter 21], but the resulting conditional probabilities essentially suffer from the same issue, as they are not uniquely defined on a set with probability zero.

We mitigate this issue of conditioning on events with probability zero in a different way, by turning to the notion of full conditional probabilities – see [7, Section 3] or [6, Definition 2]. Henceforth, we denote the set of all events – that is, subsets of $\mathfrak{S}$ – by $2^{\mathfrak{S}}$ and let $2^{\mathfrak{S}}_{\text{ne}} := 2^{\mathfrak{S}} \setminus \{\varnothing\}$ be the set of all events that are non-empty.

Definition 2. Consider two algebras of events $\mathcal{G}, \mathcal{H} \subseteq 2^{\mathfrak{S}}$ such that $\mathcal{G}$ includes $\mathcal{H}$. A *full conditional probability* P *on* $\mathcal{K} := \mathcal{G} \times (\mathcal{H} \setminus \{\varnothing\})$ is a real-valued map on $\mathcal{K}$ such that, for all A, B in $\mathcal{G}$ and C, D in $\mathcal{H} \setminus \{\varnothing\}$,

CP1. $P(A \mid C) \geq 0$;
CP2. $P(A \mid C) = 1$ whenever A includes C;
CP3. $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C)$ whenever A and B are disjoint;
CP4. $P(A \cap D \mid C) = P(A \mid D \cap C) P(D \mid C)$ whenever C and D are not disjoint.

It should be clear that the notion of a full conditional probability generalises the notion of a probability charge in such a way that conditioning on events with probability zero is dealt with in a unique manner. Indeed, properties (CP1)–(CP3) ensure that $P(\bullet \mid C)$ is a probability charge on $\mathcal{G}$ for every conditioning event C in $\mathcal{H} \setminus \{\varnothing\}$, while property (CP4) is a multiplicative version of Bayes's rule.

Next, we deal with conditional probabilities whose domain does not have the structure as in Definition 2, and we do so through the so-called coherence condition. Regazzini [6, Definition 1] gives several equivalent forms of this condition, but we repeat the simpler form of Krak et al. [2, Definition 4.2]. In it, we let $\mathbb{I}_A \colon \mathfrak{S} \to \mathbb{R}$ denote the indicator of an event A in $2^{\mathfrak{S}}$, which is the real map that takes on the value 1 on A and 0 elsewhere.

Definition 3. Consider a non-empty subset $\mathcal{K}$ of $2^{\mathfrak{S}} \times 2^{\mathfrak{S}}_{\text{ne}}$. A *coherent conditional probability* P *on* $\mathcal{K}$ is a real-valued map on $\mathcal{K}$ such that for all n in $\mathbb{N}$, $(A_1, C_1), \ldots, (A_n, C_n)$ in $\mathcal{K}$ and $\mu_1, \ldots, \mu_n$ in $\mathbb{R}$,

$$\max\left\{ \sum_{k=1}^{n} \mu_k \mathbb{I}_{C_k}(\mathfrak{s}) \big(\mathbb{I}_{A_k}(\mathfrak{s}) - P(A_k \mid C_k) \big) \colon \mathfrak{s} \in \bigcup_{k=1}^{n} C_k \right\} \geq 0.$$

This coherence condition might seem daunting at first, but it has an intuitive betting interpretation: we can think of $P(A \mid C)$ as the 'called-off fair price' for the uncertain reward $\mathbb{I}_A$ contingent on C, and the coherence condition in Definition 3 then ensures that these called-off fair prices do not permit a 'Dutch book', that is, do not lead to a partial – or sure – loss. For a more thorough explanation of this betting interpretation, we refer the interested reader to [13, Section 2.4.1] – and also to Williams's [22,23] more general work or the more recent treatment by Troffaes and De Cooman [4, Chapter 13]. De Finetti [24, Section 2.3.2] cautions that this betting interpretation arguably only makes sense whenever for all (A, C) in $\mathcal{K}$, the events A and C are 'well-determined', in the sense that they 'should be specified in such a way that a possible bet based upon [them] can be decided without question'. For this reason, it is only in this specific case that we will impose the coherence condition of Definition 3.

A second – and perhaps more important – argument for using the coherence condition is that it ensures that the conditional probability P on $\mathcal{K}$ has desirable properties: Regazzini [6, Section 2] states that P is a non-negative real-valued function that satisfies (CP1)–(CP4) on its domain $\mathcal{K}$. Note that the converse does not hold in general: demanding that (CP1)–(CP4) hold on the domain $\mathcal{K}$ is *not* sufficient to guarantee that a real-valued function P on $\mathcal{K}$ is a coherent conditional probability – Krak et al. [2, Example 4.1] give a concise counterexample. However, Regazzini [6, Theorem 3] shows that (CP1)–(CP4) suffice for coherence whenever the domain $\mathcal{K}$ has the structure required in Definition 2.

Theorem 4. *Consider two algebras of events* $\mathcal{G}, \mathcal{H} \subseteq 2^{\mathfrak{S}}$ *such that* $\mathcal{G}$ *includes* $\mathcal{H}$*, and let* $\mathcal{K} := \mathcal{G} \times (\mathcal{H} \setminus \{\varnothing\})$*. Then a real-valued map* P *on* $\mathcal{K}$ *is a full conditional probability on* $\mathcal{K}$ *if and only if it is a coherent conditional probability on* $\mathcal{K}$*.*

Another strong argument for using coherent conditional probabilities is that they can always be coherently extended to a larger domain [see 6, Theorem 4].

Theorem 5. *Consider a coherent conditional probability P on a non-empty domain $\mathcal{K} \subseteq 2^{\mathfrak{S}} \times 2^{\mathfrak{S}}_{\text{ne}}$. Then for any larger domain $\mathcal{K}^\star$ such that $\mathcal{K} \subseteq \mathcal{K}^\star \subseteq 2^{\mathfrak{S}} \times 2^{\mathfrak{S}}_{\text{ne}}$, there is a coherent conditional probability $P^\star$ on $\mathcal{K}^\star$ that extends P, or equivalently, that coincides with P on $\mathcal{K}$.*

Together with Theorem 4, this result implies that coherent conditional probabilities are simply restrictions of full conditional probabilities to domains that are not of the form in Definition 2.

2.4. Jump processes as coherent conditional probabilities

Krak et al. [2, Definition 4.3] define a jump process as a coherent conditional probability with domain

$$\mathcal{D} \coloneqq \big\{(A \mid X_u = x_u) \colon u \in \mathcal{U}, x_u \in \mathcal{X}_u, A \in \mathcal{A}_u\big\}, \tag{2}$$

where we write $(A \mid X_u = x_u)$ instead of $(A, \{X_u = x_u\})$.

Definition 6. A *jump process* P is a coherent conditional probability on $\mathcal{D}$. We let $\mathbb{P}$ be the set of all jump processes.

Note that the domain $\mathcal{D}$ is *not* of the form in Definition 2, but Theorems 4 and 5 ensure that there is a (coherent) conditional probability on $\mathcal{A} \times (\mathcal{A} \setminus \{\varnothing\})$ that extends P. For every $(A \mid X_u = x_u)$ in $\mathcal{D}$, the conditioning event $\{X_u = x_u\}$ and the finitary event A in the algebra $\mathcal{A}_u$ depend on the state of the system at a finite number of time points, so a bet on these events can be decided without question – at least if we agree to the idealisation that for any given time point, we can observe the state of the system at precisely that time point, or alternatively, that we can measure time with arbitrary precision. Because the pairs of events in the domain $\mathcal{D}$ are 'well-determined', it arguably makes sense to impose coherence.

2.5. Expectation corresponding to a jump process

Every jump process P in $\mathbb{P}$ induces a (conditional) expectation operator E_P. To define E_P, we fix some conditioning event $\{X_u = x_u\}$, with u in $\mathcal{U}$ and x_u in $\mathcal{X}_u$. Recall from right before Theorem 4 that, because P is a coherent conditional probability on $\mathcal{D}$, $P(\bullet \mid X_u = x_u)$ satisfies (CP1)–(CP3) on $\mathcal{A}_u$, meaning that $P(\bullet \mid X_u = x_u)$ is a probability charge. This probability charge $P(\bullet \mid X_u = x_u)$ corresponds to an expectation operator $E_P(\bullet \mid X_u = x_u)$ that is defined on the set of $\mathcal{A}_u$-simple variables in the usual way: through the Dunford integral [4, Definition 8.13], or equivalently, through natural extension [4, Theorem 8.15]. Let us define this (conditional) expectation operator formally.

For the sake of generality – we also need the following concepts in Section 4.2 further on – we consider a generic sample space $\mathfrak{S}$ and an algebra $\mathcal{S}$ over this sample space. A variable is a map on $\mathfrak{S}$, and we call a variable (extended) real if it takes values in the (extended) reals. For example, for any event A in $2^{\mathfrak{S}}$, its indicator $\mathbb{I}_A \colon \mathfrak{S} \to \mathbb{R}$ is a (real) variable. A (real) variable $f \colon \mathfrak{S} \to \mathbb{R}$ is called *$\mathcal{S}$-simple* if it has a representation of the form $f = \sum_{k=1}^{n} a_k \mathbb{I}_{A_k}$ for some n in $\mathbb{N}$, $a_1, \ldots, a_n$ in $\mathbb{R}$ and $A_1, \ldots, A_n$ in $\mathcal{S}$ [4, Definition 1.16]. It is relatively easy to check that the set of $\mathcal{S}$-simple variables, which we will denote by $\mathbb{S}(\mathcal{S})$, constitutes a real vector space that includes all constant maps from $\mathfrak{S}$ to $\mathbb{R}$.

A probability charge P on $\mathcal{S}$ corresponds to an expectation operator $E_P \colon \mathbb{S}(\mathcal{S}) \to \mathbb{R}$ through Dunford integration: for all f in $\mathbb{S}(\mathcal{S})$,

$$E_P(f) \coloneqq \sum_{k=1}^{n} a_k P(A_k), \tag{3}$$

where $\sum_{k=1}^{n} a_k \mathbb{I}_{A_k}$ is any representation of f of the aforementioned type; this representation of the $\mathcal{S}$-simple variable f need not be unique, but Troffaes and De Cooman [4, Definition 1.16] show that this expectation does not depend on the specific representation used. This expectation has some convenient properties [4, Lemma 8.14 and Corollary 4.14]:

E1. $\min f \leq E_P(f) \leq \max f$ for all f in $\mathbb{S}(\mathcal{S})$;
E2. $E_P(\mu f) = \mu E_P(f)$ for all f in $\mathbb{S}(\mathcal{S})$ and μ in $\mathbb{R}$;
E3. $E_P(f + g) = E_P(f) + E_P(g)$ for all f, g in $\mathbb{S}(\mathcal{S})$;
E4. $E_P(f) \leq E_P(g)$ for all f, g in $\mathbb{S}(\mathcal{S})$ such that $f \leq g$;
E5. $E_P(f + \mu) = E_P(f) + \mu$ for all f in $\mathbb{S}(\mathcal{S})$ and μ in $\mathbb{R}$.

In the particular setting of jump processes, this specialises as follows. For any sequence of time points u in $\mathcal{U}$, we denote the set of $\mathcal{A}_u$-simple variables by $\mathbb{S}_u \coloneqq \mathbb{S}(\mathcal{A}_u)$. Fix any jump process P. Then for any u in $\mathcal{U}$ and x_u in $\mathcal{X}_u$, we denote the expectation on $\mathbb{S}_u$ corresponding to the probability charge $P(\bullet \mid X_u = x_u)$ on $\mathcal{A}_u$ by $E_P(\bullet \mid X_u = x_u) \coloneqq E_{P(\bullet \mid X_u = x_u)}$. This way, we have defined the expectation E_P on

$$\mathbb{JS} \coloneqq \{(f \mid X_u = x_u) \colon u \in \mathcal{U}, x_u \in \mathcal{X}_u, f \in \mathbb{S}_u\},$$

where – as in Eqn. (2) – we favour writing $(f \mid X_u = x_u)$ instead of $(f, \{X_u = x_u\})$.

Because every event in $\mathcal{A}_u$ depends on the state of the system at a finite number of time points in or succeeding u, it is rather obvious that the same holds for every $\mathcal{A}_u$-simple variable. For this reason, these variables have a convenient representation, for which we introduce some additional notation. For any v in $\mathcal{U}$ and any real-valued function g on $\mathcal{X}_v$, we let

$$g(X_v) \coloneqq g \circ X_v = \sum_{y_v \in \mathcal{X}_v} g(y_v) \mathbb{I}_{\{X_v = y_v\}}$$

denote the function composition of g after X_v. It is obvious that $g(X_v)$ is a real variable, which is $\mathcal{A}_u$-simple whenever $v \succcurlyeq u$; in this case Eqn. (3) specialises to

$$E_P(g(X_v) \mid X_u = x_u) = \sum_{y_v \in \mathcal{X}_v} g(y_v) P(X_v = y_v \mid X_u = x_u). \tag{4}$$

Conversely, it is not difficult to verify that every $\mathcal{A}_u$-simple variable f is of the form $g(X_v)$ with $v \succcurlyeq u$ [13, Lemma 3.15].

Lemma 7. *Consider some u in $\mathcal{U}$ and some real variable $f : \Omega \to \mathbb{R}$. Then f is $\mathcal{A}_u$-simple if and only if there is some v in $\mathcal{U}_{\text{ne}}$ with $u \preccurlyeq v$ and a real-valued function g on $\mathcal{X}_v$ such that $f = g(X_v)$.*

2.6. Markovianity and homogeneity

In general, specifying a jump process is a non-trivial task: one has to specify *all* probabilities of the form $P(A \mid X_u = x_u)$ with A in $\mathcal{A}_u$ in such a way that the resulting real-valued map P on $\mathcal{D}$ is a coherent conditional probability – that is, satisfies the non-trivial coherence condition in Definition 3. For this reason, it is customary to assume the following two simplifying properties. As we will presently see, these two properties – as well as a mild continuity property – ensure that a jump process is completely and uniquely determined by two parameters: its initial probability mass function and its rate matrix. The first property is the all-important Markov property, which simplifies the so-called transition probabilities: given the present state, the probability of a future state does not depend on (a finite specification of) the past states.

Definition 8. A jump process P is *Markovian* – or, alternatively, has the Markov property – if for all t, Δ in $\mathbb{R}_{\geq 0}$, x, y in $\mathcal{X}$, u in $\mathcal{U}_{\prec t}$ and x_u in $\mathcal{X}_u$,

$$P(X_{t+\Delta} = y \mid X_t = x, X_u = x_u) = P(X_{t+\Delta} = y \mid X_t = x).$$

We denote the set of all Markovian jump processes by $\mathbb{P}^{\text{M}}$.

The second property simplifies the transition probabilities of Markovian jump processes even more: it demands that the probability of going from the present state to some future state only depends on the duration of the time period and not on the present time point.

Definition 9. A Markovian jump process P is *homogeneous* if for all t, Δ in $\mathbb{R}_{\geq 0}$ and x, y in $\mathcal{X}$,

$$P(X_{t+\Delta} = y \mid X_t = x) = P(X_\Delta = y \mid X_0 = x).$$

We denote the set of all homogeneous Markovian jump processes by $\mathbb{P}^{\text{HM}}$.

To appreciate just how much these two properties simplify the jump process P, we need to introduce some additional mathematical machinery. The set of real-valued functions on $\mathcal{X}$, which we will denote by $\mathbb{R}^{\mathcal{X}}$, is a real vector space. One important subset of $\mathbb{R}^{\mathcal{X}}$ are the *probability mass functions*: those real-valued functions p on $\mathcal{X}$ which are non-negative and normalised, meaning that $p(x) \geq 0$ for all x in $\mathcal{X}$ and $\sum_{y \in \mathcal{X}} p(y) = 1$.

We will repeatedly use transformations on $\mathbb{R}^{\mathcal{X}}$ – or maps from $\mathbb{R}^{\mathcal{X}}$ to $\mathbb{R}^{\mathcal{X}}$ – which we will simply call operators. One example of an operator is the identity operator I, which maps any f in $\mathbb{R}^{\mathcal{X}}$ to itself: $If \coloneqq f$. An operator M is *non-negatively homogeneous* if $M(\lambda f) = \lambda M f$ for all λ in $\mathbb{R}_{\geq 0}$ and f in $\mathbb{R}^{\mathcal{X}}$. Furthermore, we call an operator M *superadditive* if $M(f + g) \geq Mf + Mg$ for all f, g in $\mathbb{R}^{\mathcal{X}}$, and *additive* if this relation holds with equality instead of inequality. An operator M that is non-negatively homogeneous and additive is called *linear*, and in that case M is homogeneous, in the sense that $M(\mu f) = \mu M f$ for all μ in $\mathbb{R}$ and f in $\mathbb{R}^{\mathcal{X}}$. Although it is technically incorrect, we will usually refer to a linear operator M as a *matrix*, because a linear operator M is completely determined by its 'components': for all x, y in $\mathcal{X}$, the (x, y)-component of M is $M(x, y) \coloneqq [M\mathbb{I}_y](x)$, where $\mathbb{I}_y$ is the real-valued function on $\mathcal{X}$ that takes the value 1 in y and 0 elsewhere. Then for all f in $\mathbb{R}^{\mathcal{X}}$ and x in $\mathcal{X}$,

$$[Mf](x) = \left[M\left(\sum_{y\in\mathcal{X}} f(y)\mathbb{I}_y \right) \right](x) = \sum_{y\in\mathcal{X}} [M\mathbb{I}_y](x) f(y) = \sum_{y\in\mathcal{X}} M(x,y) f(y). \tag{5}$$

Clearly, the identity operator I is linear. For homogeneous Markovian jump processes, an important class of matrices are the rate matrices.

Definition 10. A *rate matrix* Q is an operator $Q : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ such that

R1. $Q(\mu f) = \mu Q f$ for all μ in $\mathbb{R}$ and f in $\mathbb{R}^{\mathcal{X}}$; [homogeneity]
R2. $Q(f+g) = Qf + Qg$ for all f, g in $\mathbb{R}^{\mathcal{X}}$; [additivity]
R3. $[Q\mathbb{I}_y](x) \geq 0$ for all x, y in $\mathcal{X}$ such that $x \neq y$; [non-negative off-diagonal components]
R4. $Q\mu = 0$ for any constant real-valued function μ on $\mathcal{X}$. [zero on constant functions]

We denote the set of all rate matrices by $\mathfrak{Q}$.

It might not be immediately clear that our definition of a rate matrix coincides with the usual one, for example that of Norris [25, Section 2.1]. To see that this is the case, it suffices to realise that (i) (R1) and (R2) ensure that Q is a matrix; and (ii) due to Eqn. (5), (R4) is equivalent to requiring that the matrix Q has zero row sums, in the sense that $\sum_{y\in\mathcal{X}} Q(x,y) = 0$ for all x in $\mathcal{X}$.

We equip the real vector space $\mathbb{R}^{\mathcal{X}}$ with the maximum norm $\|\bullet\|$ [26, Section 23.3], defined by

$$\|f\| := \max\{|f(x)| : x \in \mathcal{X}\} \quad \text{for all } f \in \mathbb{R}^{\mathcal{X}}.$$

The maximum norm $\|\bullet\|$ on $\mathbb{R}^{\mathcal{X}}$ induces an operator norm on the real vector space of non-negatively homogeneous operators [2, Eqn. (1)], which we will also denote by $\|\bullet\|$: for any non-negatively homogeneous operator M on $\mathbb{R}^{\mathcal{X}}$,

$$\|M\| := \sup\{\|Mf\| : f \in \mathbb{R}^{\mathcal{X}}, \|f\| = 1\};$$

this is a straightforward generalisation of the induced operator norm for matrices [see 26, Section 23.1] to non-negatively homogeneous operators. It is well-known [see 26, Section 23.3] that for any matrix M,

$$\|M\| = \max\left\{ \sum_{y\in\mathcal{X}} |M(x,y)| : x \in \mathcal{X} \right\},$$

and – see for example [11, Eqn. (5)] or [13, (R5) on p. 81] – that for any rate matrix Q,

$$\|Q\| = 2\max\{-Q(x,x) : x \in \mathcal{X}\}. \tag{6}$$

With this machinery in place, let us now get back to homogeneous Markovian jump processes. Fix a probability mass function $p : \mathcal{X} \to [0,1]$ and a rate matrix $Q : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$. Then it is essentially well-known – see for example [13, Theorem 3.37] or [2, Corollary 5.3] – that there is a unique jump process such that its so-called *initial probabilities* are given by p, in the sense that

$$P(X_0 = x) = p(x) \quad \text{for all } x \in \mathcal{X}, \tag{7}$$

and its so-called *transition probabilities* are given by the matrix exponential[3] of Q, in the sense that

$$P(X_{t+\Delta} = y \,|\, X_u = x_u, X_t = x) = \left[e^{\Delta Q}\mathbb{I}_y\right](x) \quad \text{for all } t, \Delta \in \mathbb{R}_{\geq 0}, u \in \mathcal{U}_{\prec t}, x_u \in \mathcal{X}_u, x, y \in \mathcal{X}. \tag{8}$$

Clearly, this unique jump process P is Markovian and homogeneous. Because it is fully characterised by p and Q, we denote it by $P_{p,Q}$. It follows from Eqns. (4), (5) and (8) that for all t, Δ in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{\prec t}$, x_u in $\mathcal{X}_u$, x in $\mathcal{X}$ and f in $\mathbb{R}^{\mathcal{X}}$,

$$E_{P_{p,Q}}(f(X_{t+\Delta}) \,|\, X_u = x_u, X_t = x) = \sum_{y\in\mathcal{X}} f(y) P_{p,Q}(X_{t+\Delta} = y \,|\, X_u = x_u, X_t = x) = \left[e^{\Delta Q} f\right](x). \tag{9}$$

[3] For any matrix M and real number μ, the matrix exponential $e^{\mu M}$ of μM is

$$e^{\mu M} := \lim_{n\to+\infty} \left(I + \frac{\mu}{n} M\right)^n = \lim_{n\to+\infty} \sum_{k=0}^{n} \frac{\mu^k M^k}{k!}.$$

It is now natural to wonder whether the converse is true as well: is there, for every homogeneous Markov jump process P, a probability mass function p and rate matrix Q such that $P = P_{p,Q}$? The answer is yes, provided that we focus on homogeneous Markovian jump process that are continuous in zero, in the sense that

$$\lim_{t \searrow 0} P(X_t = x \mid X_0 = x) = 1 \quad \text{for all } x \in \mathcal{X};$$

see for example [13, Theorem 3.35]. Since this condition is extremely mild – it basically requires that the probability of jumping instantaneously is zero – we see that, in practice, homogeneous Markovian jump process can be thought of as being of the form $P_{p,Q}$.

Suppose now that P is a general – not necessarily homogeneous nor Markovian – jump process. It is then again natural to wonder whether there is an initial probability mass function p and a rate matrix Q such that $P = P_{p,Q}$. As far as the initial probability mass function p is concerned, it is clear from Eqn. (7) that the only possibility is the initial probability mass function $p_P : \mathcal{X} \to [0,1]$ corresponding to P, defined by

$$p_P(x) := P(X_0 = x) \quad \text{for all } x \in \mathcal{X}.$$

For the transition probabilities, things are a bit more intricate. Due to the properties of the matrix exponential – see, for example, [13, Corollary 3.31] – it suffices to look at the dynamics of the transition probabilities, or more exactly, the right-sided derivative of $P(X_\bullet = y \mid X_u = x_u, X_t = x)$ in t and, if $t > 0$, the left-sided derivative of $P(X_t = y \mid X_u = x_u, X_\bullet = x)$ in t. To elegantly deal with these derivatives, we introduce some additional notation. For all t, r in $\mathbb{R}_{\geq 0}$ such that $t < r$, u in $\mathcal{U}_{\prec t}$ and x_u in $\mathcal{X}_u$, the corresponding *history-dependent rate matrix* $Q_{t,r}^{\{X_u = x_u\}} : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ is defined by

$$Q_{t,r}^{\{X_u = x_u\}}(x, y) := \frac{P(X_r = y \mid X_t = x, X_u = x_u) - P(X_t = y \mid X_t = x, X_u = x_u)}{r - t} \quad \text{for all } x, y \in \mathcal{X}.$$

We are now interested in the right-sided limit of $Q_{t,\bullet}^{\{X_u = x_u\}}$ in t, and if $t > 0$, in the left-sided limit of $Q_{\bullet,t}^{\{X_u = x_u\}}$. In general, these limits need not exist, but whenever they do and are all equal, say to the matrix $Q_P : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$, the transition probabilities of the jump process P are given by the matrix exponential of Q_P, as in Eqn. (9); the following result, taken from [13, Proposition 3.42], formalises this.

Proposition 11. *Consider a rate matrix Q. Then a jump process P is a homogeneous Markovian jump process with rate matrix Q – in the sense that $P = P_{p,Q}$ with $p := p_P = P(X_0 = \bullet)$ – if and only if for all t in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{\prec t}$ and x_u in $\mathcal{X}_u$,*

$$\lim_{r \searrow t} Q_{t,r}^{\{X_u = x_u\}} = Q \quad \text{and, if } t > 0, \quad \lim_{s \nearrow t} Q_{s,t}^{\{X_u = x_u\}} = Q.$$

Example 12. Troffaes et al. [10, Section 2.3] explain that the power network could be modelled by a homogeneous Markovian jump process $P_{p,Q}$ with rate matrix

$$Q = \begin{pmatrix} Q(\mathtt{AB},\mathtt{AB}) & Q(\mathtt{AB},\mathtt{A}) & Q(\mathtt{AB},\mathtt{B}) & Q(\mathtt{AB},\mathtt{F}) \\ Q(\mathtt{A},\mathtt{AB}) & Q(\mathtt{A},\mathtt{A}) & Q(\mathtt{A},\mathtt{B}) & Q(\mathtt{A},\mathtt{F}) \\ Q(\mathtt{B},\mathtt{AB}) & Q(\mathtt{B},\mathtt{A}) & Q(\mathtt{B},\mathtt{B}) & Q(\mathtt{B},\mathtt{F}) \\ Q(\mathtt{F},\mathtt{AB}) & Q(\mathtt{F},\mathtt{A}) & Q(\mathtt{F},\mathtt{B}) & Q(\mathtt{F},\mathtt{F}) \end{pmatrix} = \begin{pmatrix} \diamond & q_1^{\mathtt{B}} & q_1^{\mathtt{A}} & q_2 \\ r_{\mathtt{B}} & \diamond & 0 & q_1^{\mathtt{A}} + q_2 \\ r_{\mathtt{A}} & 0 & \diamond & q_1^{\mathtt{B}} + q_2 \\ 0 & r_{\mathtt{A}} & r_{\mathtt{B}} & \diamond \end{pmatrix},$$

where the non-negative real numbers $q_1^{\mathtt{A}}$ and $q_1^{\mathtt{B}}$ are the rates of an independent failure of $\mathtt{A}$ and $\mathtt{B}$, the non-negative real numbers $r^{\mathtt{A}}$ and $r^{\mathtt{B}}$ are the repair rates of $\mathtt{A}$ and $\mathtt{B}$, the non-negative real number q_2 is the rate of a common cause failure of both $\mathtt{A}$ and $\mathtt{B}$, and the diagonal elements are such that the rows sum to zero. The initial probability mass function p plays no role in their analysis, so we can just take it to be any arbitrary probability mass function on $\mathcal{X}$.

3. Imprecise jump processes

One is not always able to or willing to specify precise values for the characterising parameters p and Q of a homogeneous Markovian jump process. One application where this is the case is the power network of our running example.

Example 13. Troffaes et al. [10, Section 3.1] argue that in the setting of Example 12, estimating the rates exactly is difficult due to a lack of data and because these rates may not be constant over time, for example due to seasonal effects. Their estimation method for the rates gives the following lower and upper bounds for the rates listed in Table 1. To take this into account, they do not consider a single rate matrix, but consider the set $\mathcal{Q}$ of rate matrices of the form given in Example 12, where for every row separately, the rates should be in the bounds of Table 1. In other words [see 10, Eqn. (58)], they consider the set $\mathcal{Q}$ of rate matrices specified through lower and upper bounds on the off-diagonal components of the rate matrices:

Table 1
Lower and upper bounds for the rates of Example 12.

	q_1^{A}	q_1^{B}	q_2	r_{A}	r_{B}
lower bound	0.32	0.32	0.19	730	730
upper bound	0.37	0.37	0.24	1460	1460

$$\mathcal{Q} \coloneqq \left\{Q \in \mathfrak{Q} \colon (\forall x, y \in \mathcal{X}, x \neq y)\ Q_{\mathrm{L}}(x, y) \leq Q(x, y) \leq Q_{\mathrm{U}}(x, y)\right\},$$

where the matrices

$$Q_{\mathrm{L}} \coloneqq \begin{pmatrix} \diamond & 0.32 & 0.32 & 0.19 \\ 730 & \diamond & 0 & 0.51 \\ 730 & 0 & \diamond & 0.51 \\ 0 & 730 & 730 & \diamond \end{pmatrix} \quad \text{and} \quad Q_{\mathrm{U}} \coloneqq \begin{pmatrix} \diamond & 0.37 & 0.37 & 0.24 \\ 1460 & \diamond & 0 & 0.61 \\ 1460 & 0 & \diamond & 0.61 \\ 0 & 1460 & 1460 & \diamond \end{pmatrix}$$

collect the bounds on the off-diagonal components. Because every rate matrix has rows that sum to zero, the constraints on the diagonal elements of Q are implied by the others.

While they talk at great length about rate matrices, Troffaes et al. [10] do not mention anything about the initial probability mass function p. The reason for this is simple: the initial probability mass function does not play a role in their analysis. Here too, however, if p cannot be accurately estimated, this can be dealt with by considering a set $\mathcal{M}$ of probability mass functions instead of a single such function. For example, if we want to make it explicit that no information is given about p, we can let $\mathcal{M}$ be the set of all probability mass functions on $\mathcal{X}$.

To deal with these kinds of situations, Krak et al. [2] allow that instead of a single probability mass function p and rate matrix Q, we have a non-empty set $\mathcal{M}$ of probability mass functions and a non-empty and bounded set $\mathcal{Q}$ of rate matrices, where a set $\mathcal{Q}$ of rate operators is bounded if

$$\|\mathcal{Q}\| \coloneqq \sup\left\{\|Q\| \colon Q \in \mathcal{Q}\right\} < +\infty.$$

Example 14. It is easy to verify with Eqn. (6) that for the set of rate matrices $\mathcal{Q}$ defined in Example 13,

$$\|\mathcal{Q}\| = 2(1460 + 1460) = 5840,$$

so this set is bounded.

Naturally, it makes sense to consider the set

$$\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}} \coloneqq \left\{P_{p,Q} \colon p \in \mathcal{M}, Q \in \mathcal{Q}\right\}$$

of homogeneous Markovian jump processes that are characterised by an initial probability mass function p in $\mathcal{M}$ and a rate matrix Q in $\mathcal{Q}$. This is a first example of an *imprecise jump process*, with which we generally mean a non-empty set $\mathcal{P} \subseteq \mathbb{P}$ of jump processes.

There is an alternative way to look at the definition of $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$ though, and that is through the notion of 'consistency' with $\mathcal{M}$ and $\mathcal{Q}$. The notion of consistency with $\mathcal{M}$ is inspired by Eqn. (7): a jump process P in $\mathbb{P}$ is *consistent with* $\mathcal{M}$, denoted by $P \sim \mathcal{M}$, if there is a probability mass function p in $\mathcal{M}$ such that

$$P(X_0 = x) = p(x) \quad \text{for all } x \in \mathcal{X}.$$

The notion of consistency with $\mathcal{Q}$ is motivated by Proposition 11, from which we know that the rate matrix Q of a homogeneous Markovian jump process P uniquely characterises the dynamics of the transition probabilities of P. So how do we extend this 'characterising the dynamics' from a single rate matrix Q to a set of rate matrices $\mathcal{Q}$? The first idea that springs to mind is to say that a jump process P is consistent with $\mathcal{Q}$ if for all t in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{\prec t}$ and x_u in $\mathcal{X}_u$, the right-sided limit $\lim_{r \searrow t} Q_{t,r}^{\{X_u = x_u\}}$ belongs to $\mathcal{Q}$ and, if $t > 0$, so does the left-sided limit $\lim_{s \nearrow t} Q_{s,t}^{\{X_u = x_u\}}$. However, these limits are not guaranteed to exist, so this naive definition would rule out a lot of jump processes. In fact, $Q_{t,r}^{\{X_u = x_u\}}$ can have more than one accumulation point as r decreases to t, and similarly for $Q_{s,t}^{\{X_u = x_u\}}$ as s increases to t. Intuitively, the idea behind the actual definition of consistency with $\mathcal{Q}$ is that if for a jump process P all these accumulation points belong to $\mathcal{Q}$, then this jump process P is consistent with $\mathcal{Q}$. Formally – see Krak et al. [2, Definition 6.1] and Erreygers [13, Definition 3.50][4] – a jump process P is *consistent with* $\mathcal{Q}$, and we denote this by $P \sim \mathcal{Q}$, if for all t in $\mathbb{R}_{\geq 0}$, u in $\mathcal{U}_{\prec t}$ and x_u in $\mathcal{X}_u$, the set $\partial_{r \searrow t} T_{t,r}^{\{X_u = x_u\}}$[5] of 'right-sided accumulation points of $\mathcal{Q}_{t,\bullet}^{\{X_u = x_u\}}$' belongs to $\mathcal{Q}$ and, if $t > 0$, so does the

[4] Because the set of rate operators $\mathcal{Q}$ is bounded, these two notions of consistency with $\mathcal{Q}$ are in fact equivalent; for a proof, see Lemma 3.55 and Proposition 3.57 in [13].

[5] We do not use 'standard' notation here: Krak et al. [2, Definition 4.8] use $\overline{\partial}_+ T_{t,x_u}^t$ while Erreygers [13, Eqn. (3.55)] uses $\partial_+ T_{t,t}^{\{X_u = x_u\}}$, and similarly for the set of left-sided accumulation points.

set $\partial_{s\nearrow t}T^{\{X_u=x_u\}}_{s,t}$ of 'left-sided accumulation points of $\mathcal{Q}^{\{X_u=x_u\}}_{\bullet,t}$'; the formal definition of these sets of accumulations points does not really matter for the remainder, but the interested reader is invited to consult [2, Definition 4.8 and Definition 6.1] or [13, Definition 3.46 and Definition 3.50]. One thing that is important for the remainder though, and which clearly follows from this definition of consistency, is that if a jump process P is consistent with $\mathcal{Q}$, then it is also consistent with any larger bounded set $\mathcal{R} \supset \mathcal{Q}$ of rate matrices.

Our reason for introducing the notions of consistency with $\mathcal{M}$ and $\mathcal{Q}$ was to rewrite the definition of $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$. With the help of Proposition 11, it is not all too difficult to show – see Erreygers [13, Proposition 3.56] for a formal proof – that this is indeed possible:

$$\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}} = \{P \in \mathbb{P}^{\mathrm{HM}} \colon P \sim \mathcal{M}, P \sim \mathcal{Q}\}. \tag{10}$$

At this point, it is important to reiterate that one usually makes the Markovianity and homogeneity assumptions out of convenience – that is, because they make specifying the jump process a matter of specifying the initial probability mass function p and rate matrix Q – and not because one is absolutely convinced that these are justified; for example, this is the case in our running example [see 10, Section 3.1]. It is for this reason that Krak et al. [2, Definition 6.4] consider two additional imprecise jump processes characterised by $\mathcal{M}$ and $\mathcal{Q}$ by relaxing the homogeneity and Markovianity assumptions in Eqn. (10):

$$\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \coloneqq \{P \in \mathbb{P}^{\mathrm{M}} \colon P \sim \mathcal{M}, P \sim \mathcal{Q}\} \quad \text{and} \quad \mathbb{P}_{\mathcal{M},\mathcal{Q}} \coloneqq \{P \in \mathbb{P} \colon P \sim \mathcal{M}, P \sim \mathcal{Q}\}.$$

Note that, by construction,

$$\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}. \tag{11}$$

Example 15. From their – somewhat informal – exposition in [10, Section 3.2], it is clear that Troffaes et al. model the power network with the imprecise jump process $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$, with $\mathcal{M}$ and $\mathcal{Q}$ as defined in Example 13.

The three imprecise jump processes $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ all generalise the 'classical' notion of the homogeneous Markovian jump process $P_{p,Q}$ characterised by a single probability mass function p and a single rate matrix Q: Erreygers [13, Eqn. (3.66)] shows that

$$\mathbb{P}^{\mathrm{HM}}_{\{p\},\{Q\}} = \mathbb{P}^{\mathrm{M}}_{\{p\},\{Q\}} = \mathbb{P}_{\{p\},\{Q\}} = \{P_{p,Q}\}.$$

As we will see in Section 3.2 further on, they also generalise the notion of a homogeneous Markovian jump process in a different way, in the sense that they are 'Markovian' and 'homogeneous'. This becomes clear when looking at their corresponding lower and upper expectations.

3.1. Lower and upper expectations

Consider some imprecise jump process $\mathcal{P} \subseteq \mathbb{P}$. Every jump process P in $\mathcal{P}$ corresponds to an expectation operator E_P with domain $\mathbb{JS}$. This means that for any $(f \mid X_u = x_u)$ in this shared domain $\mathbb{JS}$, there is not a single value for the expectation of f conditional on $\{X_u = x_u\}$ but a range $\{E_P(f \mid X_u = x_u) \colon P \in \mathcal{P}\}$. We could set out to determine this range, but it often suffices to determine its lower and upper bounds – for example if we are interested in the worst-case and/or best-case scenario. For this reason, we consider the *lower expectation* $\underline{E}_{\mathcal{P}} \colon \mathbb{JS} \to \mathbb{R}$, defined by

$$\underline{E}_{\mathcal{P}}(f \mid X_u = x_u) \coloneqq \inf\{E_P(f \mid X_u = x_u) \colon P \in \mathcal{P}\} \quad \text{for all } (f \mid X_u = x_u) \text{ in } \mathbb{JS},$$

and the *upper expectation* $\overline{E}_{\mathcal{P}} \colon \mathbb{JS} \to \mathbb{R}$, defined by

$$\overline{E}_{\mathcal{P}}(f \mid X_u = x_u) \coloneqq \sup\{E_P(f \mid X_u = x_u) \colon P \in \mathcal{P}\} \quad \text{for all } (f \mid X_u = x_u) \text{ in } \mathbb{JS}.$$

Note that $\underline{E}_{\mathcal{P}}$ and $\overline{E}_{\mathcal{P}}$ are real valued due to (E1), and that they are conjugate, in the sense that for all $(f \mid X_u = x_u)$ in $\mathbb{JS}$,

$$\begin{aligned}\overline{E}_{\mathcal{P}}(f \mid X_u = x_u) &= \sup\{E_P(f \mid X_u = x_u) \colon P \in \mathcal{P}\} \\ &= \sup\{-E_P(-f \mid X_u = x_u) \colon P \in \mathcal{P}\} \\ &= -\inf\{-E_P(f \mid X_u = x_u) \colon P \in \mathcal{P}\} = -\underline{E}_{\mathcal{P}}(-f \mid X_u = x_u),\end{aligned}$$

where we used (E2) for the second equality. Due to this conjugacy, it suffices to study one of the two; we will focus on the lower envelope $\underline{E}_{\mathcal{P}}$. Furthermore, the lower expectation $\underline{E}_{\mathcal{P}}$ also captures lower and upper probabilities: for all $(A \mid X_u = x_u)$ in $\mathcal{D}$,

$$\underline{E}_{\mathcal{P}}(\mathbb{I}_A \mid X_u = x_u) = \inf\big\{E_P(\mathbb{I}_A \mid X_u = x_u) \colon P \in \mathcal{P}\big\} = \inf\big\{P(A \mid X_u = x_u) \colon P \in \mathcal{P}\big\},$$

and similarly for the upper probability but with $-\underline{E}_{\mathcal{P}}(-\mathbb{I}_A \mid X_u = x_u)$.

It follows more or less immediately from (E1)–(E5) and the definition of the lower and upper expectation – see also [4, Proposition 4.20 and Theorem 4.13] – that, for all u in $\mathcal{U}$ and x_u in $\mathcal{X}_u$,

LD1. $\min f \leq \underline{E}_{\mathcal{P}}(f \mid X_u = x_u) \leq \overline{E}_{\mathcal{P}}(f \mid X_u = x_u) \leq \max f$ for all f in $\mathbb{S}_u$;
LD2. $\underline{E}_{\mathcal{P}}(\mu f \mid X_u = x_u) = \mu \underline{E}_{\mathcal{P}}(f \mid X_u = x_u)$ for all f in $\mathbb{S}_u$ and μ in $\mathbb{R}_{\geq 0}$;
LD3. $\underline{E}_{\mathcal{P}}(f + g \mid X_u = x_u) \geq \underline{E}_{\mathcal{P}}(f \mid X_u = x_u) + \underline{E}_{\mathcal{P}}(g \mid X_u = x_u)$ for all f, g in $\mathbb{S}_u$;
LD4. $\underline{E}_{\mathcal{P}}(f \mid X_u = x_u) \leq \underline{E}_{\mathcal{P}}(g \mid X_u = x_u)$ for all f, g in $\mathbb{S}_u$ such that $f \leq g$;
LD5. $\underline{E}_{\mathcal{P}}(f + \mu \mid X_u = x_u) = \underline{E}_{\mathcal{P}}(f \mid X_u = x_u) + \mu$ for all f in $\mathbb{S}_u$ and μ in $\mathbb{R}$.

To somewhat shorten our notation, we follow Krak et al. [2, Definition 6.5] in denoting the lower expectations corresponding to the imprecise jump processes $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ by $\underline{E}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$, $\underline{E}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\underline{E}_{\mathcal{M},\mathcal{Q}}$, respectively, and similarly for their conjugate upper expectations; due to Eqn. (11),

$$\underline{E}_{\mathcal{M},\mathcal{Q}}(\bullet \mid \bullet) \leq \underline{E}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}(\bullet \mid \bullet) \leq \underline{E}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}(\bullet \mid \bullet),$$

and conversely for the conjugate upper expectations.

In general, determining these tight lower and upper bounds on the expectations corresponding to the imprecise jump processes $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ is intractable if not impossible, as one would have to explicitly construct these sets in order to optimise over them. However, as we will see in Section 6.1 further on, there are particular cases in which tight lower and upper bounds can be determined by repeatedly solving a more straightforward optimisation problem.

3.2. Markovianity and homogeneity

The three imprecise jump processes $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ all generalise the 'classical' notion of the homogeneous Markovian jump process, in the sense that they are characterised by a set $\mathcal{M}$ of initial probability mass functions and a set $\mathcal{Q}$ of rate matrices instead of a single probability mass function p and a single rate matrix Q. This raises the question whether these imprecise jump processes also have similar properties, so whether they are Markovian and homogeneous. For this, we need to generalise these two properties from jump processes to imprecise jump processes.

While we could generalise Definitions 8 and 9 directly, it makes more sense to generalise their equivalent 'expectation-centred' statements. With the help of Eqn. (4), it is easy to verify that for the degenerate imprecise jump process $\mathcal{P} = \{P\}$, the following two definitions reduce to Definitions 8 and 9.

Definition 16. An imprecise jump process $\mathcal{P}$ is *Markovian* – or alternatively, has the Markov property – if for all t, Δ in $\mathbb{R}_{\geq 0}$, x in $\mathcal{X}$, f in $\mathbb{R}^{\mathcal{X}}$, u in $\mathcal{U}_{\prec t}$ and x_u in $\mathcal{X}_u$,

$$\underline{E}_{\mathcal{P}}(f(X_{t+\Delta}) \mid X_t = x, X_u = x_u) = \underline{E}_{\mathcal{P}}(f(X_{t+\Delta}) \mid X_t = x).$$

Definition 17. A Markovian imprecise jump process $\mathcal{P}$ is *homogeneous* if for all t, Δ in $\mathbb{R}_{\geq 0}$, x in $\mathcal{X}$, and f in $\mathbb{R}^{\mathcal{X}}$,

$$\underline{E}_{\mathcal{P}}(f(X_{t+\Delta}) \mid X_t = x) = \underline{E}_{\mathcal{P}}(f(X_\Delta) \mid X_0 = x).$$

It is easy to show that $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$ is a homogeneous Markovian jump process, and that $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ is a Markovian – but not necessarily homogeneous[6] – jump process.

Proposition 18. *Any imprecise jump process $\mathcal{P} \subseteq \mathbb{P}^{\mathrm{M}}$ is Markovian, and any (Markovian) imprecise jump process $\mathcal{P} \subseteq \mathbb{P}^{\mathrm{HM}}$ is homogeneous. Consequently, $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$ is a homogeneous Markovian imprecise jump process, and $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ is a Markovian jump process.*

Proof. First, we assume that $\mathcal{P}$ is a subset of $\mathbb{P}^{\mathrm{M}}$. To verify that $\mathcal{P}$ satisfies the condition in Definition 16, we fix some t, Δ in $\mathbb{R}_{\geq 0}$, x in $\mathcal{X}$, f in $\mathbb{R}^{\mathcal{X}}$, u in $\mathcal{U}_{\prec t}$ and x_u in $\mathcal{X}_u$. Observe that for any jump process P in $\mathcal{P} \subseteq \mathbb{P}^{\mathrm{M}}$,

$$\begin{aligned} E_P(f(X_{t+\Delta}) \mid X_t = x, X_u = x_u) &= \sum_{y \in \mathcal{X}} f(y) P(X_{t+\Delta} \mid X_t = x, X_u = x_u) \\ &= \sum_{y \in \mathcal{X}} f(y) P(X_{t+\Delta} \mid X_t = x) = E_P(f(X_{t+\Delta}) \mid X_t = x), \end{aligned}$$

[6] Actually, we conjecture that $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ is homogeneous too, but providing a formal proof would lead us to far astray.

where for the first and last equality we used Eqn. (4) and for the second equality we used the Markovianity of P. It follows immediately from the preceding equality and the definition of the lower envelope $\underline{E}_{\mathcal{P}}$ that

$$\begin{aligned}\underline{E}_{\mathcal{P}}(f(X_{t+\Delta}) \mid X_t = x, X_u = x_u) &= \inf\{E_P(f(X_{t+\Delta}) \mid X_t = x, X_u = x_u) \colon P \in \mathcal{P}\} \\ &= \inf\{E_P(f(X_{t+\Delta}) \mid X_t = x) \colon P \in \mathcal{P}\} \\ &= \underline{E}_{\mathcal{P}}(f(X_{t+\Delta}) \mid X_t = x),\end{aligned}$$

as required.

Second, we assume that the jump process $\mathcal{P}$ is a subset of $\mathbb{P}^{\mathrm{HM}}$. Because $\mathbb{P}^{\mathrm{HM}} \subseteq \mathbb{P}^{\mathrm{M}}$, we know that $\mathcal{P}$ is Markovian. To prove that $\mathcal{P}$ satisfies the condition in Definition 17, it suffices to add one extra step – where we use that every P in $\mathcal{P}$ is homogeneous – to the argument in the first part of our proof. □

One case in which we do have that the Markovian imprecise jump process $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ is homogeneous, is when $\mathcal{Q}$ has 'separately specified rows', meaning that we can select the 'rows' of the rate matrices in $\mathcal{Q}$ independently [2, Definition 7.3].

Definition 19. A set $\mathcal{R}$ of rate matrices has *separately specified rows* if for any selection $(Q_x)_{x\in\mathcal{X}}$ of rate matrices in $\mathcal{R}$, there is a rate matrix Q in $\mathcal{R}$ such that

$$Q(x, y) = Q_x(x, y) \quad \text{for all } x, y \in \mathcal{X}.$$

The following result states that if $\mathcal{Q}$ has separately specified rows, then the Markovian imprecise jump process $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ is indeed homogeneous. Furthermore, quite remarkably, the imprecise jump process $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ turns out to be Markovian and homogeneous as well.

Corollary 20. *If $\mathcal{Q}$ has separately specified rows, then any imprecise jump process $\mathcal{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$ is a homogeneous Markovian imprecise jump process.*

Proof. Follows immediately from Lemma 57 further on, which repeats Corollary 8.3 in [2]. □

Example 21. Obviously, the (non-empty and bounded) set $\mathcal{Q}$ of rate matrices defined in Example 13 has separately specified rows, since the conditions are given row by row. Hence, our model $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ is a homogeneous Markovian imprecise jump process.

4. Extending the domain

The domain $\mathbb{JS}$ of the (conditional) lower expectation $\underline{E}_{\mathcal{P}}$ corresponding to an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}$ – or simply the expectation E_P corresponding to a jump process P – is not rich enough for many applications: the domain $\mathbb{JS}$ only contains variables that depend on the state of the system at a finite number of time points, and many practically relevant inferences correspond to variables that depend on the state of the system at *all* time points in some – possibly unbounded – interval. Examples of such variables are abundant, but let us give two examples relevant to our running example.

Example 22. As we mentioned previously in Section 1, Troffaes et al. [10, Eqns. (16) and (17)] are specifically interested in two types of variables: (i) the amount of time spent in a particular state during a particular time period, and (ii) the number of transitions to a particular state during a particular time period. The first type generalises to what is known as an 'occupancy time': in general, it is the length of time that the system is in some subset A of $\mathcal{X}$ during some interval $[s, r]$; we will formally define this type of variable in Section 4.3.2 further on. The second type generalises to what we call 'the number of selected jumps': in general, it is the number of times that during some interval $]s, r]$ the system jumps from some state x to state y, with (x, y) in a particular set of couples of distinct states; we will formally define this type of variable in Section 4.3.1 further on.

So how can we extend the domain of the lower (and upper) expectation corresponding to $\mathcal{P}$ to include these general variables? We could extend this lower expectation directly, for example through 'natural extension' [4, Section 13.7], which relies on a notion of coherence for lower expectations, or through the notion of 'previsibility' [4, Chapter 15], which basically extends the notion of convergence in probability to (unconditional) coherent lower expectations. Our reasons for not using these methods are the following. The natural extension is motivated through a betting interpretation that need not make sense for these general variables and is often overly conservative – as argued by Troffaes and De Cooman [4, Section 13.11] and Erreygers [13, Section 5.1.1]. The notion of previsibility does not line up well with our setting either: it starts from a coherent lower expectation on the set of all bounded real variables, while – for every conditioning event $\{X_u = x_u\}$ – we

seek to extend a coherent lower expectation on the set of all $\mathcal{A}_u$-simple variables. Furthermore, both of these approaches are limited to real variables, and we also desire to deal with extended real variables. Hence, we propose the following alternative approach: we extend the domain of the expectation E_P corresponding to every jump process P in $\mathcal{P}$ to a larger domain through the 'classical' approach in measure-theoretic probability theory – albeit with a twist – and subsequently take the lower (and upper) envelope over these extended expectations.

In this section, we will rely heavily on some standard definitions and results from measure-theoretical probability theory. While most reference works – for example [17], [18] or [27] – agree on the basics, there are some subtle technical points that matter in our context. For this reason, we repeat the specific definitions and results that we will need as we go; for a more gentle and thorough introduction to measure theory in general and its application to probability theory in particular, we refer the reader to Fristedt and Gray [18] and Schilling [27].

4.1. Extending a jump process through Caratheodory's Theorem

Consider a jump process P. We set out to extend the domain of E_P by extending the domain of $E_P(\bullet \mid X_u = x_u)$ for every conditioning event $\{X_u = x_u\}$, and we will do so with the help of 'countable additivity'. Before we do so, let us briefly reconsider the notion of countable additivity in the general setting of a generic sample space $\mathfrak{S}$. A probability charge P on an algebra $\mathcal{S}$ on $\mathfrak{S}$ is *countably additive* if for any sequence $(A_n)_{n\in\mathbb{N}}$ of pair-wise disjoint events in $\mathcal{S}$ such that $A := \bigcup_{n\in\mathbb{N}} A_n$ belongs to $\mathcal{S}$,

$$P(A) = \sum_{n\in\mathbb{N}} P(A_n).$$

A collection $\mathcal{S}$ of subsets of $\mathfrak{S}$ is a *σ-algebra (of events) on* $\mathfrak{S}$ – sometimes also called a σ-field – if it is an algebra of events over $\mathfrak{S}$ that is furthermore closed under countable unions, meaning that for any sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{S}$, $\bigcup_{n\in\mathbb{N}} A_n$ belongs to $\mathcal{S}$ – see [18, Definition 1 in Chapter 1] or [27, Definition 3.1]. If $\mathcal{S}$ is an algebra (on $\mathfrak{S}$), then there is a unique smallest σ-algebra (again on $\mathfrak{S}$) that includes $\mathcal{S}$, and this is the intersection of all σ-algebras that include $\mathcal{S}$ – see [18, Section 1.3] or [27, Theorem 3.4]; we denote this σ-algebra by $\sigma(\mathcal{S})$, and call it the *σ-algebra generated by* $\mathcal{S}$.

Given a σ-algebra $\mathcal{S}$ on $\mathfrak{S}$, a probability measure P on $\mathcal{S}$ is a countably additive probability charge on $\mathcal{S}$ [18, Definition 2 in Chapter 1], or equivalently, a non-negative real map on $\mathcal{S}$ that is normalised and countably additive, in the sense that $P(\mathfrak{S}) = 1$ and that for any sequence $(A_n)_{n\in\mathbb{N}}$ of pair-wise disjoint events in $\mathcal{S}$,

$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n).$$

Crucial to our extension method is that a countably additive probability charge P on some algebra $\mathcal{S}$ can be extended to a probability measure on the σ-algebra $\sigma(\mathcal{S})$ generated by $\mathcal{S}$: Caratheodory's Theorem – see [18, Theorem 14 in Chapter 7] or [27, Theorem 6.1] – ensures that there is a unique probability measure P_σ on $\sigma(\mathcal{S})$ that extends P, or equivalently, that coincides with P on $\mathcal{S}$.

Now that we have refreshed the notion of countable additivity, we can find out what it can do for us in the setting of jump processes. Recall from Section 2.3 that $P(\bullet \mid X_u = x_u)$ is a probability charge on the algebra of cylinder events $\mathcal{A}_u$; whenever this probability charge is countably additive for any conditioning event, we call the jump process P countably additive.

Definition 23. A jump process P is *countably additive* if for all u in $\mathcal{U}$ and x_u in $\mathcal{X}_u$, $P(\bullet \mid X_u = x_u)$ is countably additive.

Suppose P is a countably additive jump process. Then for any u in $\mathcal{U}$ and x_u in $\mathcal{X}_u$, $P(\bullet \mid X_u = x_u)$ is a countably additive probability charge on $\mathcal{A}_u$, so by Caratheodory's Theorem, there is a unique probability measure $P_\sigma(\bullet \mid X_u = x_u)$ on the σ-algebra $\sigma(\mathcal{A}_u)$ generated by $\mathcal{A}_u$ that extends $P(\bullet \mid X_u = x_u)$. The real-valued map P_σ on[7]

$$\mathcal{D}_\sigma := \left\{(A \mid X_u = x_u) \colon u \in \mathcal{U}, x_u \in \mathcal{X}_u, A \in \sigma(\mathcal{A}_u)\right\}$$

thus defined coincides with P on $\mathcal{D}$, so one might wonder whether this real-valued map P_σ is a coherent extension of P to $\mathcal{D}_\sigma$ – that is, whether P_σ is a coherent conditional probability on $\mathcal{D}_\sigma$. We have not been able to prove that this is the case, nor have we found a counterexample. Yet, we do not perceive this as an issue, and the reason for this is that the extended domain $\mathcal{D}_\sigma$ contains events that are *not* well-determined, so as explained right after Definition 3, there is no compelling argument to demand coherence on this extended domain. For example, fix some state x in $\mathcal{X}$, and consider the event

[7] Because the state space $\mathcal{X}$ equipped with the discrete topology is a Polish space, it follows from [18, Chapter 31, Theorem 2] that $\sigma(\mathcal{A}_u)$ is the σ-algebra of events that are measurable with respect to the Skorokhod topology on Ω – at least if $u = ()$ or $u = (0)$. For more information, we refer to Billingsley [20, Section 16].

$$A_\sigma := \bigcap_{n\in\mathbb{N}} \{X_n = x\}.$$

Because every event on the right-hand side belongs to $\mathcal{A} \subseteq \sigma(\mathcal{A})$ and the generated σ-algebra $\sigma(\mathcal{A})$ is closed under countable intersections, A_σ belongs to $\sigma(\mathcal{A})$. However, it is clear that there is no point in time after which we can unequivocally determine whether or not the event has occurred, so a bet based upon it cannot be decided without question; consequently, the event A_σ is not well-determined.

4.2. Extending the expectation corresponding to a countably additive jump process through Lebesgue integration

Consider a countably additive jump process P, and fix some u in $\mathcal{U}$ and x_u in $\mathcal{X}_u$. The extension $P_\sigma(\bullet \mid X_u = x_u)$ of the countably additive probability charge $P(\bullet \mid X_u = x_u)$ is a (countably additive) probability charge on $\sigma(\mathcal{A}_u)$. Hence, as we know from Section 2.5, there is a corresponding expectation operator $E_{P_\sigma(\bullet|X_u=x_u)}$ on the set $\mathbb{S}_u^\sigma := \mathbb{S}(\sigma(\mathcal{A}_u))$ of $\sigma(\mathcal{A}_u)$-simple variables. As the generated σ-algebra $\sigma(\mathcal{A}_u)$ includes the algebra $\mathcal{A}_u$ of cylinder events, it is clear that every $\mathcal{A}_u$-simple variable is $\sigma(\mathcal{A}_u)$-simple, so $\mathbb{S}_u \subseteq \mathbb{S}_u^\sigma$. Even more, $E_{P_\sigma(\bullet|X_u=x_u)}$ coincides with $E_P(\bullet \mid X_u = x_u)$ on $\mathbb{S}_u$: for any $\mathcal{A}_u$-simple variable f, it follows immediately from Eqn. (3) that

$$E_{P_\sigma(\bullet|X_u=x_u)}(f) = \sum_{k=1}^n a_k P_\sigma(A_k \mid X_u = x_u) = \sum_{k=1}^n a_k P(A_k \mid X_u = x_u) = E_P(f \mid X_u = x_u), \tag{12}$$

with n in $\mathbb{N}$, $a_1, \ldots, a_n$ in $\mathbb{R}$ and $A_1, \ldots, A_n$ in $\mathcal{A}_u$ such that $f = \sum_{k=1}^n a_k \mathbb{I}_{A_k}$. However, this extended domain $\mathbb{S}_u^\sigma$ is not sufficiently large for our purposes. For example, the number of jumps in some interval and the occupancy time of some set of states during some interval, two real-valued variables which we informally introduced in Example 22 and will formally introduce in Section 4.3 further on, are not $\mathbb{S}_u^\sigma$-simple: the former is not bounded, while the latter is bounded but does not have a finite range. What we need is an extension to the so-called measurable variables, and we can achieve this through Lebesgue integration.

To introduce Lebesgue integration, we briefly return to the general setting: as before, we consider a generic sample space $\mathfrak{S}$ and a σ-algebra $\mathcal{S}$ on this sample space. An extended real-variable $f : \mathfrak{S} \to \overline{\mathbb{R}}$ is *$\mathcal{S}$-measurable*[8] if

$$\{f > \alpha\} := \{\mathfrak{s} \in \mathfrak{S} : f(\mathfrak{s}) > \alpha\} \in \mathcal{S} \quad \text{for all } \alpha \in \mathbb{R}, \tag{13}$$

and we collect all $\mathcal{S}$-measurable variables in $\mathbb{M}(\mathcal{S})$. This set $\mathbb{M}(\mathcal{S})$ of $\mathcal{S}$-measurable variables is closed under several operations. In the remainder, we will rely on the following operations; whenever the proof of a property is not immediate from the condition in Eqn. (13), a proof can be found in [27, Chapter 8].

M1. Any $\mathcal{S}$-simple variable is $\mathcal{S}$-measurable.
M2. For any $\mathcal{S}$-measurable variable f and any real number λ, λf is $\mathcal{S}$-measurable.
M3. For any $\mathcal{S}$-measurable variables f and g, their sum $f + g$ is $\mathcal{S}$-measurable whenever it is defined.[9]
M4. For any $\mathcal{S}$-measurable variable f, $|f|$ is $\mathcal{S}$-measurable.
M5. For any sequence $(f_n)_{n\in\mathbb{N}}$ of $\mathcal{S}$-measurable variables that converges point-wise, its point-wise limit, denoted by $\lim_{n\to+\infty} f_n$, is $\mathcal{S}$-measurable.

Another important property of $\mathcal{S}$-measurable variables is that an extended real variable f is $\mathcal{S}$-measurable if and only if its non-negative part $f^+ := f \vee 0$ and its non-positive part $f^- = -(f \wedge 0)$ are $\mathcal{S}$-measurable. This is used implicitly in the definition of the Lebesgue integral, and here we repeat the one given by Fristedt and Gray [18, Definition 8]. Consider a probability measure P on $\mathcal{S}$. Then for any $\mathcal{S}$-measurable variable f such that at least one of the suprema

$$\sup\left\{E_P(g) : g \in \mathbb{S}(\mathcal{S}), 0 \leq g \leq f^+\right\} \quad \text{and} \quad \sup\left\{E_P(g) : g \in \mathbb{S}(\mathcal{S}), 0 \leq g \leq f^-\right\}$$

is real, the (Lebesgue) integral of f with respect to P exists and is

$$E_P^{\mathrm{L}}(f) := \sup\left\{E_P(g) : g \in \mathbb{S}(\mathcal{S}), 0 \leq g \leq f^+\right\} - \sup\left\{E_P(g) : g \in \mathbb{S}(\mathcal{S}), 0 \leq g \leq f^-\right\};$$

otherwise, the (Lebesgue) integral does not exist. We denote the set of all $\mathcal{S}$-measurable variables for which the Lebesgue integral with respect to P exists by $\operatorname{dom} E_P^{\mathrm{L}}$. In general, this domain need not be the same for different probability measures. However, it is relatively easy to prove that $\operatorname{dom} E_P^{\mathrm{L}}$ always contains the $\mathcal{S}$-measurable variable f if it is either bounded below or bounded above, meaning that $\inf f > -\infty$ or $\sup f < +\infty$, respectively.

[8] Technically, we should say $\mathcal{S}/\overline{\mathcal{B}}$-measurable, where $\overline{\mathcal{B}}$ denotes the Borel σ-algebra on $\overline{\mathbb{R}}$, being the σ-algebra generated by the open sets of the topology on $\overline{\mathbb{R}}$ that is the product of the usual topology on $\mathbb{R}$ and the discrete topology on $\{-\infty, +\infty\}$. For more details, we refer to Chapter 2 in [18] and Chapter 8 in [27].

[9] We use the standard extensions of the binary operations of addition, subtraction and multiplication from $\mathbb{R}$ to $\overline{\mathbb{R}}$ – see [4, Appendix D] or [27, p. 58]. This means that we leave $(+\infty) + (-\infty)$, $(+\infty) - (+\infty)$, $(-\infty) + (+\infty)$ and $(-\infty) - (-\infty)$ undefined.

Lemma 24. *Consider a σ-algebra $\mathcal{S}$ over $\mathfrak{S}$, a probability measure P on $\mathcal{S}$ and a $\mathcal{S}$-measurable variable f. If f is bounded below or bounded above, then the (Lebesgue) integral of f exists, or equivalently, f belongs to* $\operatorname{dom} E_P^{\mathrm{L}}$.

Proof. Assume that f is bounded below. Note that for any $\mathcal{S}$-simple variable g such that $0 \leq g \leq f^-$, $0 \leq \min g \leq \max g \leq \sup f^-$. It follows from this inequality and (E1) that

$$0 \leq \sup\left\{E_P(g) \colon g \in \mathbb{S}(\mathcal{S}), 0 \leq g \leq f^-\right\} \leq \sup f^-.$$

Either $\sup f^- = 0$ if $f \geq 0$, or $\sup f^- = \sup(-f) = -\inf f$ otherwise. Because $\inf f > -\infty$ by assumption, we infer from this that $\sup f^- < +\infty$, and this proves that the Lebesgue integral of f exists.

The proof for the case that f is bounded above is analogous. □

It is well-known – see [18, Problem 14 in Chapter 4] – that the (Lebesgue) expectation E_P^{L} on $\operatorname{dom} E_P^{\mathrm{L}}$ extends the (Dunford) expectation E_P on $\mathbb{S}(\mathcal{S})$: any $\mathcal{S}$-simple variable f belongs to $\operatorname{dom} E_P^{\mathrm{L}}$, and $E_P^{\mathrm{L}}(f) = E_P(f)$. For future reference, we list some additional well-known properties of the Lebesgue integral E_P^{L}; for a proof, we refer to Fristedt and Gray [18, Theorem 9 in Chapter 4]. For all f, g in $\operatorname{dom} E_P^{\mathrm{L}}$ and μ in $\mathbb{R}$,

LI1. $\inf f \leq E_P^{\mathrm{L}}(f) \leq \sup f$;
LI2. $\mu f \in \operatorname{dom} E_P^{\mathrm{L}}$ and $E_P^{\mathrm{L}}(\mu f) = \mu E_P^{\mathrm{L}}(f)$;
LI3. if $f + g$ and $E_P^{\mathrm{L}}(f) + E_P^{\mathrm{L}}(g)$ are defined, then $f + g \in \operatorname{dom} E_P^{\mathrm{L}}$ and $E_P^{\mathrm{L}}(f + g) = E_P^{\mathrm{L}}(f) + E_P^{\mathrm{L}}(g)$;
LI4. $E_P^{\mathrm{L}}(f) \leq E_P^{\mathrm{L}}(g)$ whenever $f \leq g$;
LI5. $|f| \in \operatorname{dom} E_P^{\mathrm{L}}$ and $|E_P^{\mathrm{L}}(f)| \leq E_P^{\mathrm{L}}(|f|)$.

Furthermore, for all f in $\operatorname{dom} E_P^{\mathrm{L}}$, its Lebesgue expectation is given by the Choquet integral [28, Chapter 5]:

$$E_P^{\mathrm{L}}(f) = \int_0^{+\infty} P(\{f^+ > \alpha\})\,\mathrm{d}\alpha - \int_0^{+\infty} P(\{f^- > \alpha\})\,\mathrm{d}\alpha = \int_0^{+\infty} P(\{f > \alpha\})\,\mathrm{d}\alpha - \int_{-\infty}^{0} P(\{f < \alpha\})\,\mathrm{d}\alpha,$$

where the integrals are – possibly improper – Riemann integrals that always exists because their integrand is monotone.

Let us now return to the setting of jump processes. As before, we consider a countably additive jump process P. For all u in $\mathcal{U}$ and x_u in $\mathcal{X}_u$, we denote the Lebesgue integral corresponding to the probability measure $P_\sigma(\bullet \mid X_u = x_u)$ on $\sigma(\mathcal{A}_u)$ by $E_P^\sigma(\bullet \mid X_u = x_u) \coloneqq E^{\mathrm{L}}_{P_\sigma(\bullet \mid X_u = x_u)}$. While the domain of this Lebesgue integral may depend on the specific jump process P, Lemma 24 guarantees that this domain includes all $\sigma(\mathcal{A}_u)$-measurable variables that are bounded below or bounded above. Therefore, we choose to restrict the domain of the Lebesgue integral $E_P^\sigma(\bullet \mid X_u = x_u)$ to these variables, so to the set $\mathbb{M}_u \coloneqq \mathbb{M}(\sigma(\mathcal{A}_u)) \cap \overline{\mathbb{V}}_{\mathrm{b}}$, where $\overline{\mathbb{V}}_{\mathrm{b}}$ is the set of all extended real variables $f \colon \Omega \to \overline{\mathbb{R}}$ that are either bounded below or bounded above. This ensures that the domain of $E_P^\sigma(\bullet \mid X_u = x_u)$ is the same for any countably additive jump process P. In Section 4.4 further on, it will become clear that this choice is especially convenient in the setting of imprecise jump processes. Furthermore, to the best of our knowledge, this somewhat non-standard restriction of the domain does not hinder the applicability of our framework: we are yet to encounter an application where all of the 'conventional' (extended) real variables that may be of interest are not bounded below or above.

To summarise, we have defined an expectation operator E_P^σ with domain

$$\mathbb{J}\mathbb{M} \coloneqq \left\{(f \mid X_u = x_u) \colon u \in \mathcal{U}, x_u \in \mathcal{X}_u, f \in \mathbb{M}_u\right\}.$$

But does E_P^σ extend E_P to a domain that includes interesting (extended) real variables that are not $\sigma(\mathcal{A}_u)$-simple? The following result establishes that E_P^σ extends E_P, which partially answers this question.

Lemma 25. *Consider a countably additive jump process P. Then $\mathbb{J}\mathbb{S} \subseteq \mathbb{J}\mathbb{M}$, and*

$$E_P^\sigma(f \mid X_u = x_u) = E_P(f \mid X_u = x_u) \quad \textit{for all } (f \mid X_u = x_u) \in \mathbb{J}\mathbb{S}.$$

Proof. Fix any u in $\mathcal{U}$ and x_u in $\mathcal{U}$. Recall from right before (LI1) that $E_P^\sigma(\bullet \mid X_u = x_u) = E^{\mathrm{L}}_{P_\sigma(\bullet \mid X_u = x_u)}$ coincides with $E_{P_\sigma(\bullet \mid X_u = x_u)}$ on $\mathbb{S}(\sigma(\mathcal{A}_u))$, and from around Eqn. (12) that $E_{P_\sigma(\bullet \mid X_u = x_u)}$ coincides with $E_P(\bullet \mid X_u = x_u)$ on $\mathbb{S}_u \subseteq \mathbb{S}(\sigma(\mathcal{A}_u))$. Consequently, $E_P^\sigma(\bullet \mid X_u = x_u)$ coincides with $E_P(\bullet \mid X_u = x_u)$ on $\mathbb{S}_u$. □

4.3. Two classes of measurable variables

It remains for us to show that the extended domain $\mathbb{J}\mathbb{M}$ contains many – if not most or all – of the non-simple variables that may be of interest in applications. We cannot possibly give an exhaustive list, so we will formally introduce two classes

of such variables: Section 4.3.1 deals with the number of (selected) jumps in an interval $]s,r]$, while Section 4.3.2 takes a closer look at real variables that take the form of a Riemann integral of $f(X_t)$ as t ranges over the interval $[s,r]$. We have chosen these two classes because they cover the two types of variables which Troffaes et al. [10] consider, and which we have informally introduced in Example 22. That said, we would like to emphasise that the extended domain $\mathbb{JM}$ includes a lot more variables that appear in applications. For example, this domain also includes indicators of until events [13, Section 6.2] – which play an important role in model checking [29,30] – and hitting times [13, Section 6.3].

Our exposition simplifies due to the following notation and terminology. Fix two time points s,r in $\mathbb{R}_{\geq 0}$ such that $s<r$. A *grid* over $[s,r]$ is a sequence of time points $v=(t_0,\ldots,t_n)$ in $\mathcal{U}_{\text{ne}}$ that starts in $t_0=s$ and ends in $t_n=r$. For any such grid $v=(t_0,\ldots,t_n)$ over $[s,r]$, we call

$$\Delta(v) \coloneqq \max\{t_k - t_{k-1} \colon k \in \{1,\ldots,n\}\}$$

the *maximum grid width of* v. For grids v,w over $[s,r]$, we say that w *refines* v, and denote this by $v \subseteq w$, if w includes all the time points in v.

4.3.1. The number of selected jumps

Fix some càdlàg path ω in Ω. Recall from Section 2.1 that then, by definition, for all t in $\mathbb{R}_{>0}$, ω has a left-sided limit in t and ω is continuous from the right in t. Whenever the value of ω in t differs from the left-sided limit of ω in t, meaning that

$$\lim_{\Delta \searrow 0} \omega(t-\Delta) \neq \omega(t),$$

we say that *a jump occurs at time* t. For any time points s,r in $\mathbb{R}_{\geq 0}$ such that $s<r$, we collect the jump times in $]s,r]$ in the set

$$\mathcal{J}_{]s,r]}(\omega) \coloneqq \left\{ t \in]s,r] \colon \lim_{\Delta \searrow 0} \omega(t-\Delta) \neq \omega(t) \right\}; \tag{14}$$

it is essentially well known – see for example Lemma 5.20 in [13] – that this set of jump times is always finite. Hence, for all s,r in $\mathbb{R}_{\geq 0}$ such that $s<r$, the number of jumps in $]s,r]$, which we denote by

$$\eta_{]s,r]} \colon \Omega \to \mathbb{Z}_{\geq 0} \colon \omega \mapsto \eta_{]s,r]}(\omega) \coloneqq \left| \mathcal{J}_{]s,r]}(\omega) \right|,$$

is a non-negative real-valued variable. Sometimes we are not interested in all jumps, but only in jumps between specific couples of states, which we then collect in some subset A of $\mathcal{X}^2_{\neq} \coloneqq \{(x,y) \in \mathcal{X}^2 \colon x \neq y\}$. To deal with this more general case, for any such subset A of $\mathcal{X}^2_{\neq}$ and any time points s,r in $\mathbb{R}_{\geq 0}$ such that $s<r$, we let $\eta^A_{]s,r]} \colon \Omega \to \mathbb{Z}_{\geq 0}$ be the non-negative real variable that is defined for any ω in Ω by

$$\eta^A_{]s,r]}(\omega) \coloneqq \left| \left\{ t \in]s,r] \colon \left(\lim_{\Delta \searrow 0} \omega(t-\Delta), \omega(t) \right) \in A \right\} \right|,$$

and we refer to this variable as the *number of selected jumps*. It follows immediately from this definition that

$$0 \leq \eta^A_{]s,r]} \leq \eta_{]s,r]}, \tag{15}$$

and that for $A = \mathcal{X}^2_{\neq}$, $\eta^A_{]s,r]} = \eta_{]s,r]}$.

In order to show that the number of (selected) jumps belongs to our domain, we set out to show that we can approximate the number of (selected) jumps by means of a specific type of simple variables. Fix time points s,r in $\mathbb{R}_{\geq 0}$ such that $s<r$, and a subset A of $\mathcal{X}^2_{\neq}$. The crucial idea is that we can approximate the number of selected jumps $\eta^A_{]s,r]}$ in $]s,r]$ by looking at the number of selected jumps along some grid v over $[s,r]$. More precisely, for any grid $v=(t_0,\ldots,t_n)$ over $[s,r]$, we use the corresponding approximation

$$\eta^A_v \colon \Omega \to \mathbb{Z}_{\geq 0} \colon \omega \mapsto \eta^A_v(\omega) \coloneqq \left| \left\{ k \in \{1,\ldots,n\} \colon \left(\omega(t_{k-1}), \omega(t_k)\right) \in A \right\} \right|;$$

to simplify our notation, we write η_v instead of η^A_v whenever $A = \mathcal{X}^2_{\neq}$. It is easy to see that

$$\eta^A_v = \sum_{k=1}^{n} \mathbb{I}_A\left(X_{t_{k-1}}, X_{t_k}\right), \tag{16}$$

where here and in the remainder, we follow the convention that the empty sum is zero, and where we let $\mathbb{I}_A$ be the function on $\mathcal{X}^2$ defined by

$$\mathbb{I}_A(x,y) \coloneqq \begin{cases} 1 & \text{if } (x,y) \in A \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } x,y \in \mathcal{X}.$$

It follows immediately from Eqn. (16) that

$$0 \leq \eta_v^A \leq \eta_v. \tag{17}$$

Because η_v^A clearly only depends on the state of the system at the time points in v, we have the following corollary of Lemma 7.

Corollary 26. *Consider some u in $\mathcal{U}$ and some s, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Then for any subset A of $\mathcal{X}_{\neq}^2$ and any grid v over $[s, r]$, the corresponding approximation η_v^A is a non-negative $\mathcal{A}_u$-simple variable.*

Because the number of jumps of a càdlàg path ω in Ω in the interval $]s, r]$ is finite, there is a grid width $\Delta_{]s,r]}^{\omega}$ such that every grid v with a grid width $\Delta(v)$ smaller than $\Delta_{]s,r]}^{\omega}$ captures all (selected) jumps in $]s, r]$. The following result, which is a generalisation of Lemma 5.25 in [13], establishes this.

Lemma 27. *Consider time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$ and a subset A of $\mathcal{X}_{\neq}^2$. For any càdlàg path ω in Ω, there is a positive real number δ such that if v is a grid over $[s, r]$ with $\Delta(v) < \delta$, then*

$$\eta_v^A(\omega) = \eta_{]s,r]}^A(\omega).$$

Proof. Fix some path ω in Ω. Recall from right after Eqn. (14) that the set $\mathcal{J}_{]s,r]}(\omega)$ of the jump times of ω in $]s, r]$ is finite, so we can order this finite set. This way, we obtain a grid $(t_0, \ldots, t_n)$ over $[s, r]$, where we always add $t_0 = s$ and only add $t_n = r$ if it is not a jump time of ω. Note that for all k in $\{1, \ldots, n\}$, $\omega(t) = \omega(t_{k-1})$ for all t in $[t_{k-1}, t_k[$, so ω is constant over $[t_{k-1}, t_k[$. Let

$$\delta := \min\{t_k - t_{k-1} : k \in \{1, \ldots, n\}\},$$

and take any grid $v = (t_0^v, \ldots, t_m^v)$ over $[s, r]$ such that $\Delta(v) < \delta$. It follows from this condition on v that, for all k in $\{1, \ldots, n\}$, v contains at least one time point t_ℓ^v in the subinterval $[t_{k-1}, t_k[$ where ω is constant. In other words, for every jump time t_k of ω in $]s, r]$, there is an index ℓ in $\{1, \ldots, m\}$ such that $t_{\ell-1}^v$ is a time point in $[t_{k-1}, t_k[$ and t_ℓ^v is a time point in $[t_k, t_{k+1}[$ if $k < n$ or $t_\ell^v = t_k$ if $k = n$. It is not difficult to see that this implies that $\eta_{[s,r]}^A(\omega) = \eta_v^A(\omega)$, as required. □

Due to Lemma 27, the approximation η_v^A converges point-wise to $\eta_{]s,r]}^A$ as the grid width $\Delta(v)$ of the grid v over $[s, r]$ vanishes; in combination with Corollary 26, this implies that the number of (selected) jumps belongs to our domain.

Lemma 28. *Consider some u in $\mathcal{U}$ and s, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Fix some subset A of $\mathcal{X}_{\neq}^2$ and let $(v_n)_{n\in\mathbb{N}}$ be a sequence of grids over $[s, r]$ such that $\lim_{n\to+\infty} \Delta(v_n) = 0$. Then $(\eta_{v_n}^A)_{n\in\mathbb{N}}$ is a sequence of $\mathcal{A}_u$-simple variables that converges point-wise to $\eta_{]s,r]}^A$. Consequently, $\eta_{]s,r]}^A$ belongs to $\mathbb{M}_u$.*

Proof. For all n in $\mathbb{N}$, $\eta_{v_n}^A$ is a non-negative $\mathcal{A}_u$-simple variable due to Corollary 26. Furthermore, because $\lim_{n\to+\infty} \Delta(v_n) = 0$ by assumption, it follows from Lemma 27 that the sequence $(\eta_{v_n}^A)_{n\in\mathbb{N}}$ converges point-wise to $\eta_{]s,r]}^A$. This proves the first part of the statement. The second part of the statement follows immediately from the first part because (i) every $\mathcal{A}_u$-simple variable is trivially $\sigma(\mathcal{A}_u)$-simple and therefore $\sigma(\mathcal{A}_u)$-measurable due to (M1); (ii) the point-wise limit of a sequence of $\sigma(\mathcal{A}_u)$-measurable variables is again $\sigma(\mathcal{A}_u)$-measurable due to (M5); and (iii) $\eta_{]s,r]}^A$ is non-negative and therefore trivially bounded below. □

In general, the convergence of η_v^A to $\eta_{]s,r]}^A$ need not be monotone. However, it can be in the case of all jumps, so if $A = \mathcal{X}_{\neq}^2$. The reason for this is that for any two grids v and w over $[s, r]$ such that w refines v, the number of jumps along w is (point-wise) greater than or equal to the number of jumps along v; Erreygers [13, Lemma 5.23] gives a formal proof.

Lemma 29. *Consider some u in $\mathcal{U}$ and some s, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Then for all grids v, w over $[s, r]$ with $v \subseteq w$, $\eta_v \leq \eta_w$.*

Closely related to this, we also have that $\eta_{]s,r]}^A$ need not yield an upper bound for η_v^A. However, here too, it will if $A = \mathcal{X}_{\neq}^2$.

Lemma 30. *Consider some u in $\mathcal{U}$ and some s, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Then for all grids v over $[s, r]$, $\eta_v \leq \eta_{]s,r]}$.*

Proof. We fix some grid $v = (t_0, \ldots, t_n)$ over $[s, r]$. Furthermore, for all k in $\mathbb{N}$, we let v_k be the grid over $[s, r]$ that is the (ordered) union of the grids $v_{k,1}$, …, $v_{k,n}$, where for all ℓ in $\{1, \ldots, n\}$,

$$v_{k,\ell} := \left(t_{\ell-1}, t_{\ell-1} + \frac{1}{2^{k-1}}(t_\ell - t_{\ell-1}), \ldots, t_{\ell-1} + \frac{2^{k-1}}{2^{k-1}}(t_\ell - t_{\ell-1}) \right)$$

is the grid over $[t_{\ell-1}, t_\ell]$ that consists of $2^{k-1}+1$ evenly spaced time points. Then by construction, $v_1 = v$, $v_k \subseteq v_{k+1}$ for all k in $\mathbb{N}$, and $\lim_{k\to+\infty} \Delta(v_k) = \lim_{k\to+\infty} \Delta(v)/2^{k-1} = 0$. Hence, it follows from Lemmas 28 and 29 that the corresponding sequence $(\eta_{v_k})_{k\in\mathbb{N}}$ is a non-decreasing sequence of variables that converges point-wise to $\eta_{]s,r]}$. Because $v = v_1$, this proves that $\eta_v \leq \eta_{]s,r]}$. □

4.3.2. Integral of $f(X_t)$ over the interval $[s, r]$

As a second example of a class of measurable variables that is included in our extended domain, we treat real variables that take the form of (Riemann) integrals over time. As we will see momentarily, this class of variables includes occupancy times, which we have previously encountered in Example 22.

Fix some real-valued function f on $\mathcal{X}$ and any two time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$. Then for any path ω in Ω, the function composition $f \circ \omega$ is piece-wise constant over $[s, r]$ because ω is càdlàg; as every piece-wise constant real-valued function on $[s, r]$ is Riemann integrable [26, Section 24.26], it follows that $f \circ \omega$ is Riemann integrable over $[s, r]$. Hence,

$$\int_s^r f(X_t)\,\mathrm{d}t \colon \Omega \to \mathbb{R} \colon \omega \mapsto \int_s^r f\big(\omega(t)\big)\,\mathrm{d}t$$

is well-defined, and we call this real variable the *integral of $f(X_t)$ over* $[s, r]$. Without much extra effort, we could generalise this to families $(f_t)_{t\in[s,r]}$ of real-valued functions on $\mathcal{X}$ that are piece-wise constant. However, because the 'stationary' case already encompasses a lot of important types of measurable variables, and because it allows for a simpler exposition, we do not treat this more general set-up here; instead, we refer the interested reader to Section 6.4 in [13].

Before we continue, let us mention two types of measurable variables that take the form of an integral over time. For the first example, we take the real-valued function $f = \mathbb{I}_A$ for some subset A of $\mathcal{X}$. Then the integral $\int_s^r \mathbb{I}_A(X_t)\,\mathrm{d}t$ of $\mathbb{I}_A(X_t)$ over $[s, r]$ is the length of time that the system's state is in A between time points s and r; as explained in Example 22, we call this the *occupancy time of A over* $[s, r]$ [31, Section 4.5]. As a second example, for any real-valued function h on $\mathcal{X}$, the *temporal average of $h(X_t)$ over* $[s, r]$ is the integral

$$\frac{1}{r-s}\int_s^r h(X_t)\,\mathrm{d}t = \int_s^r \frac{1}{r-s} h(X_t)\,\mathrm{d}t = \int_s^r f(X_t)\,\mathrm{d}t,$$

of $f(X_t)$ over $[s, r]$, with $f := \frac{1}{r-s}h$.

For any path ω in Ω, the integral $\int_s^r f\big(\omega(t)\big)\,\mathrm{d}t$ is defined through a limit of Riemman sums [see 26, Definition 24.3], so it is natural to use these Riemann sums to construct a sequence of simple variables that converges point-wise to $\int_s^r f(X_t)\,\mathrm{d}t$. More precisely, for any grid $v = (t_0, \ldots, t_n)$ over $[s, r]$, we consider the corresponding Riemann sum

$$\langle f \rangle_v := \sum_{k=1}^{n} (t_k - t_{k-1}) f(X_{t_k}). \tag{18}$$

Note that in this expression, for every k in $\{1, \ldots, n\}$, we could replace $f(X_{t_k})$ with $f(X_{s_k})$ where s_k is any element of $[t_{k-1}, t_k]$; we choose to use $f(X_{t_k})$ for the sake of simplicity. By construction, $\langle f \rangle_v$ only depends on the state of the system at the time points in the grid v; for this reason, we have the following corollary of Lemma 7.

Corollary 31. *Consider some u in $\mathcal{U}$ and s, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Then for any grid v over $[s, r]$ and any real-valued function f on $\mathcal{X}$, the corresponding Riemann sum $\langle f \rangle_v$ is $\mathcal{A}_u$-simple, and*

$$(r-s)\min f \leq \langle f \rangle_v \leq (r-s)\max f.$$

Due to the definition of the Riemann integral [26, Definition 24.3], the Riemann sum $\langle f \rangle_v$ converges to the integral $\int_s^r f(X_t)\,\mathrm{d}t$ as the grid width $\Delta(v)$ of the grid v vanishes.

Lemma 32. *Consider some u in $\mathcal{U}$ and s, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Fix some real-valued function f on $\mathcal{X}$, and let $(v_n)_{n\in\mathbb{N}}$ be a sequence of grids over $[s, r]$ such that $\lim_{n\to+\infty} \Delta(v_n) = 0$. Then $(\langle f \rangle_{v_n})_{n\in\mathbb{N}}$ is a uniformly bounded sequence of $\mathcal{A}_u$-simple variables that converges point-wise to $\int_s^r f(X_t)\,\mathrm{d}t$. Consequently, $\int_s^r f(X_t)\,\mathrm{d}t$ belongs to $\mathbb{M}_u$.*

Proof. That $(\langle f \rangle_{v_n})_{n\in\mathbb{N}}$ is a uniformly bounded sequence of $\mathcal{A}_u$-simple variables follows immediately from Corollary 31. That $(\langle f \rangle_{v_n})_{n\in\mathbb{N}}$ converges point-wise to $\int_s^r f(X_t)\,\mathrm{d}t$ follows immediately from the definition of the Riemann integral – see, for example, Definition 24.3 in [26]. The second part of the statement follows from the preceding part due to the same argument as in the second part of our proof for Lemma 28: (i) every $\mathcal{A}_u$-simple variable is $\sigma(\mathcal{A}_u)$-simple and therefore $\sigma(\mathcal{A}_u)$-measurable; (ii) the point-wise limit of a sequence of $\sigma(\mathcal{A}_u)$-measurable variables is again $\sigma(\mathcal{A}_u)$-measurable; and (iii) the integral of $f(X_t)$ over $[s,r]$ is bounded below by $(r-s)\min f$ and above by $(r-s)\max f$. □

4.4. Extending the domain of Markovian imprecise jump processes

The extension method that we have laid out in Sections 4.1 and 4.2 allows us to extend the domain $\mathbb{JS}$ of the expectation E_P corresponding to any countably additive jump process P to the larger domain $\mathbb{JM}$ that includes many – if not all – relevant variables. Clearly, this method is only applicable to an imprecise jump processes $\mathcal{P}\subseteq\mathbb{P}$ if every jump processes P in $\mathcal{P}$ is countably additive. The following result establishes that this is the case for $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ and therefore, by Eqn. (11), also for $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$. Although it might seem a bit underwhelming, we believe that this is in fact the single most important result in this contribution – without it, we would not be able to state the others – and definitely the most difficult one to obtain; the interested reader can find a rather long proof in [13, Corollary 5.30].

Theorem 33. *Every jump process P in the imprecise jump process $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ is countably additive.*

To appreciate this result, it should be contrasted with the 'standard' measure-theoretical approach to continuous-time stochastic processes. On that approach, one starts off with the set $\mathcal{X}^{\mathbb{R}_{\geq 0}}$ of all paths – so not only those that are càdlàg – to ensure countable additivity of any probability charge on the algebra of cylinder events, subsequently extends this probability charge by means of Caratheodory's Theorem to a probability measure on the product σ-algebra, and finally constructs a 'modification' of the projectors $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ to obtain càdlàg sample paths – see, for example, [17, Section 38], [15, Lemma 3.16 and Theorem 3.18] or [19, Chapter 18]. Theorem 33 demonstrates that this somewhat convoluted method is not needed – at least not for bounded $\mathcal{Q}$ and finite $\mathcal{X}$ – as it is possible to immediately start with càdlàg paths. Ironically, our proof for Theorem 33 does make use of this 'modification' method, albeit as an intermediate step under the hood.

Due to Theorem 33, we can extend the domain of the lower and upper expectations corresponding to any imprecise jump process $\mathcal{P}\subseteq\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ – and so in particular $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$ – as follows. First, for every jump process P in $\mathcal{P}$, we use the method in Sections 4.1 and 4.2 to extend the domain of the expectation E_P on $\mathbb{JS}$ by going over to the expectation E_P^σ on $\mathbb{JM}$. Second, we take the lower and upper envelopes over these extended expectations: this yields the lower and upper expectation $\underline{E}^\sigma_{\mathcal{P}}$ and $\overline{E}^\sigma_{\mathcal{P}}$ on $\mathbb{JM}$, defined for all $(f\,|\,X_u=x_u)$ in $\mathbb{JM}$ by

$$\underline{E}^\sigma_{\mathcal{P}}(f\,|\,X_u=x_u) \coloneqq \inf\big\{E^\sigma_P(f\,|\,X_u=x_u)\colon P\in\mathcal{P}\big\} \quad\text{and}\quad \overline{E}^\sigma_{\mathcal{P}}(f\,|\,X_u=x_u) \coloneqq \sup\big\{E^\sigma_P(f\,|\,X_u=x_u)\colon P\in\mathcal{P}\big\}.$$

It follows from the properties of $(E^\sigma_P)_{P\in\mathcal{P}}$ that $\underline{E}^\sigma_{\mathcal{P}}$ and $\overline{E}^\sigma_{\mathcal{P}}$ extend $\underline{E}_{\mathcal{P}}$ and $\overline{E}_{\mathcal{P}}$ to $\mathbb{JM}$, and that they are conjugate, in the sense that for all $(f\,|\,X_u=x_u)$ in $\mathbb{JM}$,

$$\overline{E}^\sigma_{\mathcal{P}}(f\,|\,X_u=x_u) = -\underline{E}^\sigma_{\mathcal{P}}(-f\,|\,X_u=x_u).$$

In order not to burden our notation, we will implicitly extend the domain of the (conditional) lower and upper expectations corresponding to $\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$; that is, we will henceforth use $\underline{E}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$ to denote $\underline{E}^\sigma_{\mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}}$, and similarly for $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ and their corresponding lower expectation, making the superscript σ implicit.

5. Convergence theorems

For any countably additive jump process P, u in $\mathcal{U}$ and x_u in $\mathcal{X}_u$, the corresponding (conditional) expectation operator $E^\sigma_P(\bullet\,|\,X_u=x_u)$ has two well-known continuity properties. First and foremost, it is continuous with respect to monotone sequences of $\sigma(\mathcal{A}_u)$-measurable variables. The general result for the (Lebesgue) expectation with respect to a probability measure is known as the Monotone Convergence Theorem [18, Theorem 11 and Corollary 13 in Chapter 4]; in our specific setting of jump processes, it specialises to the following result.

Theorem 34. *Consider a countably additive jump process P, fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$ and f in $\mathbb{M}_u$, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{M}_u$ that converges point-wise to f. If the sequence $(f_n)_{n\in\mathbb{N}}$ is non-decreasing and $E^\sigma_P(f_1\,|\,X_u=x_u)>-\infty$, then*

$$\lim_{n\to+\infty} E^\sigma_P(f_n\,|\,X_u=x_u) = E^\sigma_P(f\,|\,X_u=x_u).$$

A similar statement holds if the sequence $(f_n)_{n\in\mathbb{N}}$ is non-increasing and $E^\sigma_P(f_1\,|\,X_u=x_u)<+\infty$.

The second continuity result is known as Lebesgue's Dominated Convergence Theorem [18, Theorem 9 in Chapter 8]. As with the Monotone Convergence Theorem, we only state this convergence result in the particular setting of countably additive jump processes.

Theorem 35. *Consider a countably additive jump process P, fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$ and f in $\mathbb{M}_u$, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{M}_u$ that converges point-wise to f. If there is some g in $\mathbb{M}_u$ such that $|f_n| \leq g$ for all n in $\mathbb{N}$ and $E_P^\sigma(g \mid X_u = x_u) < +\infty$, then*

$$\lim_{n\to+\infty} E_P^\sigma(f_n \mid X_u = x_u) = E_P^\sigma(f \mid X_u = x_u).$$

The conditions in Theorem 35 regarding g are trivially satisfied if there is a non-negative real number B such that $|f_n| \leq B$ for all n in $\mathbb{N}$. Fristedt and Gray [18, Theorem 10 in Chapter 8] call this well-known special case of Lebesgue's Dominated Convergence Theorem the Bounded Convergence Theorem.

Unfortunately, similar continuity results are not available for the (conditional) lower and upper expectations $\underline{E}_{\mathcal{P}}^\sigma(\bullet \mid X_u = x_u)$ and $\overline{E}_{\mathcal{P}}^\sigma(\bullet \mid X_u = x_u)$ corresponding to an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, at least not in general. This (potential) lack of continuity is not exclusive to imprecise jump processes. In a more general setting, Miranda and Zaffalon [32, Section 5.1] establish that for any given set of probability measures, the lower envelope of the corresponding expectations – in essence, the corresponding Lebesgue integrals – is always continuous with respect to monotone non-increasing sequences but may not be continuous with respect to non-decreasing sequences. In our setting, this translates to the following result.

Theorem 36. *Consider an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$ and f in $\mathbb{M}_u$, and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{M}_u$ that converges point-wise to f. If the sequence $(f_n)_{n\in\mathbb{N}}$ is non-decreasing and $\underline{E}_{\mathcal{P}}^\sigma(f_1 \mid X_u = x_u) > -\infty$, then*

$$\lim_{n\to+\infty} \underline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u) \leq \underline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u) \quad \textit{and} \quad \lim_{n\to+\infty} \overline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u) = \overline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u).$$

Similarly, if the sequence $(f_n)_{n\in\mathbb{N}}$ is non-increasing and $\overline{E}_{\mathcal{P}}^\sigma(f_1 \mid X_u = x_u) < +\infty$, then

$$\lim_{n\to+\infty} \underline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u) = \underline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u) \quad \textit{and} \quad \lim_{n\to+\infty} \overline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u) \geq \overline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u).$$

Proof. This proof is essentially that of Theorem 5.31 in [13], but with the necessary modifications to account for the difference between Lebesgue and Daniell integration. We will only prove the first part of the statement; the proof for the second part is analogous, but it also follows from the first part due to conjugacy.

Fix some P in $\mathcal{P}$. Then

$$E_P^\sigma(f_1 \mid X_u = x_u) \geq \underline{E}_{\mathcal{P}}^\sigma(f_1 \mid X_u = x_u) > -\infty,$$

where the strict inequality holds by assumption. Consequently, it follows from Theorem 34 that

$$E_P^\sigma(f \mid X_u = x_u) = \lim_{n\to+\infty} E_P^\sigma(f_n \mid X_u = x_u).$$

Because $(f_n)_{n\in\mathbb{N}}$ is non-decreasing by assumption, it follows from (LI4) that the corresponding sequence $(E_P^\sigma(f_n \mid X_u = x_u))_{n\in\mathbb{N}}$ is non-decreasing, and therefore

$$E_P^\sigma(f \mid X_u = x_u) = \lim_{n\to+\infty} E_P^\sigma(f_n \mid X_u = x_u) = \sup\left\{E_P^\sigma(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\right\}. \tag{19}$$

Because for every P in $\mathcal{P}$, $(E_P^\sigma(f_n \mid X_u = x_u))_{n\in\mathbb{N}}$ is a non-decreasing sequence of extended real numbers (bounded below by $\underline{E}_{\mathcal{P}}^\sigma(f_1 \mid X_u = x_u) > -\infty$), it is easy to see that

$$\left(\underline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u)\right)_{n\in\mathbb{N}} \quad \text{and} \quad \left(\overline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u)\right)_{n\in\mathbb{N}}$$

are non-decreasing sequences of extended real numbers (bounded below by $\underline{E}_{\mathcal{P}}^\sigma(f_1 \mid X_u = x_u) > -\infty$), so the limits of these sequences exist. Because these sequences are non-decreasing,

$$\lim_{n\to+\infty} \underline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u) = \sup\left\{\underline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\right\} \tag{20}$$

and

$$\lim_{n\to+\infty} \overline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u) = \sup\left\{\overline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\right\} \tag{21}$$

To verify the equality in the first part of the statement, we recall that by definition of the upper envelope $\overline{E}_{\mathcal{P}}^\sigma$,

$$\overline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u) = \sup\left\{E_P^\sigma(f \mid X_u = x_u) \colon P \in \mathcal{P}\right\}.$$

From this and Eqn. (19), it follows that

$$\begin{aligned}\overline{E}^{\sigma}_{\mathcal{P}}(f \mid X_u = x_u) &= \sup\left\{\sup\left\{E^{\sigma}_{P}(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\right\} \colon P \in \mathcal{P}\right\} \\ &= \sup\left\{\sup\left\{E^{\sigma}_{P}(f_n \mid X_u = x_u) \colon P \in \mathcal{P}\right\} \colon n \in \mathbb{N}\right\} \\ &= \sup\left\{\overline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\right\}.\end{aligned}$$

The equality in the first part of the statement follows immediately from the preceding equality and Eqn. (21):

$$\overline{E}^{\sigma}_{\mathcal{P}}(f \mid X_u = x_u) = \sup\left\{\overline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\right\} = \lim_{n \to +\infty} \overline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u).$$

The inequality in the first part of the statement follows from a similar argument. By definition of $\underline{E}^{\sigma}_{\mathcal{P}}$,

$$\underline{E}^{\sigma}_{\mathcal{P}}(f \mid X_u = x_u) = \inf\left\{E^{\sigma}_{P}(f \mid X_u = x_u) \colon P \in \mathcal{P}\right\}.$$

From this and Eqn. (19), it follows that

$$\begin{aligned}\underline{E}^{\sigma}_{\mathcal{P}}(f \mid X_u = x_u) &= \inf\left\{\sup\left\{E^{\sigma}_{P}(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\right\} \colon P \in \mathcal{P}\right\} \\ &\geq \sup\left\{\inf\left\{E^{\sigma}_{P}(f_n \mid X_u = x_u) \colon P \in \mathcal{P}\right\} \colon n \in \mathbb{N}\right\} \\ &= \sup\left\{\underline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\right\},\end{aligned}$$

where the inequality holds because we have changed the order of the supremum and the infimum – for a proof of this property, we refer to [4, Lemma 15.18]. The inequality in the first part of the statement follows immediately from the preceding inequality and Eqn. (20):

$$\underline{E}^{\sigma}_{\mathcal{P}}(f \mid X_u = x_u) \geq \sup\left\{\underline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u) \colon n \in \mathbb{N}\right\} = \lim_{n \to +\infty} \underline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u). \quad \square$$

We would like to emphasise that for non-decreasing sequences $(f_n)_{n \in \mathbb{N}}$ of variables in $\mathbb{M}_u$, the limit of $\underline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u)$ exists as n recedes to $+\infty$, but this limit is only guaranteed to be a – possibly conservative – lower bound; a similar remark holds for the limit of $\overline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u)$ for non-increasing sequences. For dominated sequences, we have a similar type of conservative limit behaviour.

Theorem 37. *Consider an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$ and f in $\mathbb{M}_u$, and let $(f_n)_{n \in \mathbb{N}}$ be a sequence in $\mathbb{M}_u$ that converges point-wise to f. If there is some g in $\mathbb{M}_u$ with $\overline{E}^{\sigma}_{\mathcal{P}}(g \mid X_u = x_u) < +\infty$ such that $|f_n| \leq g$ for all n in $\mathbb{N}$, then*

$$\limsup_{n \to +\infty} \underline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u) \leq \underline{E}^{\sigma}_{\mathcal{P}}(f \mid X_u = x_u) \quad \text{and} \quad \liminf_{n \to +\infty} \overline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u) \geq \overline{E}^{\sigma}_{\mathcal{P}}(f \mid X_u = x_u).$$

Proof. This proof is essentially that of Theorem 5.32 in [13], but with some obvious modifications to adapt it to the Lebesgue integral instead of the Daniell expectation. It clearly suffices to prove the inequality for the conditional lower expectation $\underline{E}^{\sigma}_{\mathcal{P}}$, because this implies the inequality for the conditional upper expectation $\overline{E}^{\sigma}_{\mathcal{P}}$ due to the conjugacy relation.

Observe that by assumption,

$$(\forall P \in \mathcal{P})\ E^{\sigma}_{P}(g \mid X_u = x_u) \leq \overline{E}^{\sigma}_{\mathcal{P}}(g \mid X_u = x_u) =: \beta < +\infty. \tag{22}$$

Due to Eqn. (22), and because $|f_n| \leq g$ for all n in $\mathbb{N}$ by assumption, it follows from Theorem 35 that

$$(\forall P \in \mathcal{P})\ E^{\sigma}_{P}(f \mid X_u = x_u) = \lim_{n \to +\infty} E^{\sigma}_{P}(f_n \mid X_u = x_u). \tag{23}$$

Furthermore, because $-g \leq f_n \leq g$ for all n in $\mathbb{N}$ by assumption, it follows from Eqn. (22), (LI4) and (LI2) that

$$(\forall n \in \mathbb{N})(\forall P \in \mathcal{P})\ -\beta \leq E^{\sigma}_{P}(f_n \mid X_u = x_u) \leq \beta. \tag{24}$$

From Eqns. (23) and (24), we infer that $-\beta \leq E^{\sigma}_{P}(f \mid X_u = x_u) \leq \beta$ for all P in $\mathcal{P}$, and therefore also $-\beta \leq \underline{E}^{\sigma}_{\mathcal{P}}(f \mid X_u = x_u) \leq \beta$.

Fix any ϵ in $\mathbb{R}_{>0}$. As $\underline{E}^{\sigma}_{\mathcal{P}}(f \mid X_u = x_u)$ is real, there is some P in $\mathcal{P}$ such that

$$\underline{E}^{\sigma}_{\mathcal{P}}(f \mid X_u = x_u) > E^{\sigma}_{P}(f \mid X_u = x_u) - \epsilon = \lim_{n \to +\infty} E^{\sigma}_{P}(f_n \mid X_u = x_u) - \epsilon.$$

Note that $E^{\sigma}_{P}(f_n \mid X_u = x_u) \geq \underline{E}^{\sigma}_{\mathcal{P}}(f_n \mid X_u = x_u)$ for all n in $\mathbb{N}$, and therefore

$$\lim_{n\to+\infty} E_P^\sigma(f_n \mid X_u = x_u) = \limsup_{n\to+\infty} E_P^\sigma(f_n \mid X_u = x_u) \geq \limsup_{n\to+\infty} \underline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u).$$

It follows immediately from the preceding two inequalities that

$$\underline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u) > \limsup_{n\to+\infty} \underline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u) - \epsilon.$$

Because ϵ was an arbitrary positive real number, we conclude that

$$\underline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u) \geq \limsup_{n\to+\infty} \underline{E}_{\mathcal{P}}^\sigma(f_n \mid X_u = x_u),$$

and this is what we set out to prove. □

While there is a (potential) lack of continuity in general, it turns out that for many practically relevant variables g in $\mathbb{M}_u$, the lower and upper expectations corresponding to $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$ are actually continuous for appropriately chosen sequences $(g_n)_{n\in\mathbb{N}}$ of $\mathcal{A}_u$-simple variables. In Section 5.1 we show that this is the case for the number of selected jumps $\eta^A_{]s,r]}$ and the sequence $(\eta^A_{\nu_n})_{n\in\mathbb{N}}$ with $(\nu_n)_{n\in\mathbb{N}}$ a sequence of grids over $[s,r]$ such that $\lim_{n\to+\infty}\Delta(\nu_n)=0$, and we do the same in Section 5.2 for $\int_s^r f(X_t)\,\mathrm{d}t$ and the sequence $(\langle f\rangle_{\nu_n})$ with $(\nu_n)_{n\in\mathbb{N}}$ again a sequence of grids over $[s,r]$ with vanishing grid width. In both of these cases, we make use of the following straightforward intermediary result to show that the lower/upper expectation of the simple variable g_n converges to that of the measurable variable g as n recedes to $+\infty$.

Lemma 38. *Consider an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$, f in $\mathbb{M}_u$ and g in $\mathbb{S}_u$. If ϵ is a non-negative real number such that*

$$(\forall P \in \mathcal{P})\ E_P^\sigma\big(|f-g| \,\big|\, X_u = x_u\big) \leq \epsilon,$$

then

$$\left|\underline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u) - \underline{E}_{\mathcal{P}}(g \mid X_u = x_u)\right| \leq \epsilon \quad \text{and} \quad \left|\overline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u) - \overline{E}_{\mathcal{P}}(g \mid X_u = x_u)\right| \leq \epsilon.$$

Proof. We only prove the inequality of the statement for the lower expectation. The inequality for the upper expectation can be proven in an analogous manner, but also follows from the one for the lower expectation due to conjugacy.

Observe that for any jump process P in $\mathcal{P}$,

$$\begin{aligned}
\left|E_P^\sigma(f \mid X_u = x_u) - E_P(g \mid X_u = x_u)\right| &= \left|E_P^\sigma(f \mid X_u = x_u) - E_P^\sigma(g \mid X_u = x_u)\right| \\
&= \left|E_P^\sigma(f \mid X_u = x_u) + E_P^\sigma(-g \mid X_u = x_u)\right| \\
&= \left|E_P^\sigma(f - g \mid X_u = x_u)\right| \\
&\leq E_P^\sigma(\left|f-g\right| \mid X_u = x_u),
\end{aligned}$$

where for the first equality we used that E_P^σ extends E_P, for the second equality we used (LI2), for the third equality we used (LI3) and the fact that $E_P^\sigma(-g \mid X_u = x_u) = -E_P^\sigma(g \mid X_u = x_u)$ is real (due to (LI1)), and for the inequality we used (LI5). It follows from this inequality and the assumption in the statement that

$$(\forall P \in \mathcal{P})\ \left|E_P^\sigma(f \mid X_u = x_u) - E_P(g \mid X_u = x_u)\right| \leq \epsilon.$$

This implies that

$$\begin{aligned}
\underline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u) = \inf\big\{E_P^\sigma(f \mid X_u = x_u) \colon P \in \mathcal{P}\big\} &\leq \inf\big\{E_P(g \mid X_u = x_u) + \epsilon \colon P \in \mathcal{P}\big\} \\
&= \inf\big\{E_P(g \mid X_u = x_u) \colon P \in \mathcal{P}\big\} + \epsilon \\
&= \underline{E}_{\mathcal{P}}(g \mid X_u = x_u) + \epsilon.
\end{aligned}$$

Similarly, we find that

$$\underline{E}_{\mathcal{P}}^\sigma(f \mid X_u = x_u) \geq \underline{E}_{\mathcal{P}}(g \mid X_u = x_u) - \epsilon.$$

The inequality of the statement follows immediately from these two inequalities because $\underline{E}_{\mathcal{P}}(g \mid X_u = x_u)$ is real valued due to (LD1). □

5.1. Convergence for the expected number of (selected) jumps

Consider an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, and fix a sequence of time points u in $\mathcal{U}$, a state instantiation x_u in $\mathcal{X}_u$, time points s, t in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$ and a subset A of $\mathcal{X}_{\neq}^2$. Our aim is to obtain a convergence result for $\underline{E}_{\mathcal{P}}(\eta^A_{]s,r]} \,|\, X_u = x_u)$ and/or $\overline{E}_{\mathcal{P}}(\eta^A_{]s,r]} \,|\, X_u = x_u)$. In particular, we want to specialise Theorems 36 and 37 for $f = \eta^A_{]s,r]}$ and $(f_n)_{n\in\mathbb{N}} = (\eta^A_{v_n})_{n\in\mathbb{N}}$, with $(v_n)_{n\in\mathbb{N}}$ a suitable sequence of grids over $[s, r]$. From Lemma 28, we know that $(\eta^A_{v_n})_{n\in\mathbb{N}}$ converges point-wise to $\eta^A_{]s,r]}$ if $\lim_{n\to+\infty} \Delta(v_n) = 0$. Furthermore, if $A = \mathcal{X}_{\neq}^2$, then we know from Lemma 29 that this sequence is non-decreasing if $v_n \subseteq v_{n+1}$ for all n in $\mathbb{N}$. Hence, if all these conditions are satisfied, then we can use Theorem 36 to determine conservative bounds on the lower/upper expectation of $\eta^A_{]s,r]} = \eta_{]s,r]}$. If we only have that $\lim_{n\to+\infty} \Delta(v_n) = 0$, then we need to resort to Theorem 37. In that case, we need to find some g in $\mathbb{M}_u$ with $\overline{E}^{\sigma}_{\mathcal{P}}(g \,|\, X_u = x_u) < +\infty$ such that $|\eta^A_{v_n}| \leq g$ for all n in $\mathbb{N}$. To this end, we observe that for all n in $\mathbb{N}$, $|\eta^A_{v_n}| \leq \eta_{]s,r]}$ by Eqn. (17) and Lemma 30. So all we need to apply Theorem 37 is a real-valued upper bound on the upper expectation of $\eta_{]s,r]}$. The following result, which will also play an essential role in our proofs for Theorem 40 – via Lemma 44 – and Theorem 46 further on, provides such a bound. The proof is rather lengthy, and can be found in [13, Corollary 5.18 and Theorem 5.27].

Theorem 39. *Consider any jump process P in $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$ and time points s, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Then for any grid v over $[s, r]$,*

$$E_P(\eta_v \,|\, X_u = x_u) \leq (r-s)\frac{\|\mathcal{Q}\|}{2} \quad \text{and} \quad E^{\sigma}_P(\eta_{]s,r]} \,|\, X_u = x_u) \leq (r-s)\frac{\|\mathcal{Q}\|}{2}.$$

Because $\overline{E}^{\sigma}_{\mathcal{P}}(\eta_{]s,r]} \,|\, X_u = x_u) < +\infty$ due to Theorem 39, this means that $g = \eta_{]s,r]}$ satisfies our needs, so we could invoke Theorem 37 to bound the lower/upper expectation of $\eta^A_{]s,r]}$. Remarkably, however, in the particular case of the number of (selected) jumps, the bounds in Theorems 36 and 37 turn out to be tight, and the limit superior and limit inferior in Theorem 37 are in fact limits. The reason for this is the following important result, which generalises Proposition 6.3 in [33].

Theorem 40. *Consider an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$ and time points s, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$, and fix any subset A of $\mathcal{X}_{\neq}^2$. Then for any grid v over $[s, r]$,*

$$\left|\underline{E}^{\sigma}_{\mathcal{P}}(\eta^A_{]s,r]} \,|\, X_u = x_u) - \underline{E}_{\mathcal{P}}(\eta^A_v \,|\, X_u = x_u)\right| \leq \frac{1}{4}\Delta(v)(r-s)\|\mathcal{Q}\|^2,$$

and the same inequality holds for the upper expectation.

For our proof, we use several intermediary results. The first one is a consequence of Theorem 39, and is taken from [13, Corollary 5.18 and Lemma 6.53].

Lemma 41. *Consider a jump process P in $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$ and time points s, t, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < t < r$. Then*

$$\{X_s \neq X_t \neq X_r\} := \left\{\omega \in \Omega \colon \omega(s) \neq \omega(t) \text{ and } \omega(t) \neq \omega(r)\right\}$$

belongs to $\mathcal{A}_u$, and

$$P(X_s \neq X_t \neq X_r \,|\, X_u = x_u) \leq \frac{1}{4}(t-s)(r-t)\|\mathcal{Q}\|^2.$$

The second intermediary result we need is a generalisation of Lemma 5.24 in [13] from η_v to η^A_v; the original statement of this result contained a small error, and we thank Arne Decadt for pointing this out to us.

Lemma 42. *Consider time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$ and a grid $v = (t_0, \ldots, t_n)$ over $[s, r]$ with $n \geq 2$. Then for any subset A of $\mathcal{X}_{\neq}^2$,*

$$\eta^A_{(s,r)} - \sum_{k=1}^{n-1} \mathbb{I}_{\{X_{t_{k-1}} \neq X_{t_k} \neq X_r\}} \leq \eta_v \leq \eta^A_{(s,r)} + 2\sum_{k=1}^{n-1} \mathbb{I}_{\{X_{t_{k-1}} \neq X_{t_k} \neq X_r\}}.$$

Proof. Crucial to our proof is the following observation. Let $w = (s_0, \ldots, s_m)$ be any grid over $[s, r]$, and fix some time point t in $]s_{m-1}, s_m[$. Then it follows from Eqn. (16) that

$$\eta^A_{w\cup(t)} - \eta^A_w = \mathbb{I}_A(X_{s_{m-1}}, X_t) + \mathbb{I}_A(X_t, X_{s_m}) - \mathbb{I}_A(X_{s_{m-1}}, X_{s_m}).$$

For all x, y, z in $\mathcal{X}$, $\mathbb{I}_A(x,y)+\mathbb{I}_A(y,z)-\mathbb{I}_A(x,z)=0$ if $x=y$ or $y=z$, and $-1\leq\mathbb{I}_A(x,y)+\mathbb{I}_A(y,z)-\mathbb{I}_A(x,z)\leq 2$ otherwise, so if $x\neq y\neq z$. Hence,

$$\eta^A_w-\mathbb{I}_{\{X_{s_{m-1}}\neq X_t\neq X_{s_m}\}}\leq\eta^A_{w\cup(t)}\leq\eta^A_w+2\mathbb{I}_{\{X_{s_{m-1}}\neq X_t\neq X_{s_m}\}}. \tag{25}$$

Let $v_0:=(t_0,t_n)$, and for all k in $\{1,\dots,n-1\}$, let $v_k:=(t_0,\dots,t_k,t_n)$; note that $v_{n-1}=v$. Then it follows from Eqn. (25) that for all k in $\{1,\dots,n-1\}$,

$$\eta^A_{v_{k-1}}-\mathbb{I}_{\{X_{t_{k-1}}\neq X_{t_k}\neq X_{t_n}\}}\leq\eta^A_{v_k}\leq\eta^A_{v_{k-1}}+2\mathbb{I}_{\{X_{t_{k-1}}\neq X_{t_k}\neq X_{t_n}\}}.$$

We repeatedly apply the second inequality, to yield

$$\begin{aligned}\eta^A_v=\eta^A_{v_{n-1}}\leq\eta^A_{v_{n-2}}+2\mathbb{I}_{\{X_{t_{n-2}}\neq X_{t_{n-1}}\neq X_{t_n}\}}\leq\cdots\leq\eta^A_{v_0}+2\sum_{k=1}^{n-1}\mathbb{I}_{\{X_{t_{k-1}}\neq X_{t_k}\neq X_{t_n}\}}\\=\eta^A_{(s,r)}+2\sum_{k=1}^{n-1}\mathbb{I}_{\{X_{t_{k-1}}\neq X_{t_k}\neq X_r\}};\end{aligned}$$

similarly, by repeated use of the first inequality we find that

$$\begin{aligned}\eta^A_v=\eta^A_{v_{n-1}}\geq\eta^A_{v_{n-2}}-\mathbb{I}_{\{X_{t_{n-2}}\neq X_{t_{n-1}}\neq X_{t_n}\}}\geq\cdots\geq\eta^A_{v_0}-\sum_{k=1}^{n-1}\mathbb{I}_{\{X_{t_{k-1}}\neq X_{t_k}\neq X_{t_n}\}}\\=\eta^A_{(s,r)}-\sum_{k=1}^{n-1}\mathbb{I}_{\{X_{t_{k-1}}\neq X_{t_k}\neq X_r\}}.\end{aligned}$$

This proves the inequalities in the statement. □

Next, we establish that the absolute value of the difference between the number of selected jumps in $]s,r]$ and the number of selected jumps along a grid v over $[s,r]$ is measurable and bounded below; this result is similar to – and, in some sense, is a corollary of – Lemma 28.

Lemma 43. *Consider some u in $\mathcal{U}$ and s,r in $\mathbb{R}_{\geq 0}$ such that $\max u\leq s<r$. Fix some subset A of $\mathcal{X}^2_{\neq}$ and a grid v over $[s,r]$, and let $(v_n)_{n\in\mathbb{N}}$ be a sequence of grids over $[s,r]$ such that $\lim_{n\to+\infty}\Delta(v_n)=0$. Then $(|\eta^A_{v_n}-\eta^A_v|)_{n\in\mathbb{N}}$ is a sequence of $\mathcal{A}_u$-simple variables that converges point-wise to $|\eta^A_{]s,r]}-\eta^A_v|$. Consequently, $|\eta^A_{]s,r]}-\eta^A_v|$ belongs to $\mathbb{M}_u$.*

Proof. Recall from Lemma 28 that $(\eta^A_{v_n})_{n\in\mathbb{N}}$ is a sequence of $\mathcal{A}_u$-simple variables that converges point-wise to $\eta^A_{]s,r]}$, and from Corollary 26 that η^A_v is a $\mathcal{A}_u$-simple variable. For any two $\mathcal{A}_u$-simple variables f and g, the non-negative real variable $|f-g|$ is $\mathcal{A}_u$-simple too. Hence, $(|\eta^A_{v_n}-\eta^A_v|)_{n\in\mathbb{N}}$ is a sequence of $\mathcal{A}_u$-simple variables, and it is obvious that this sequence converges point-wise to $|\eta^A_{]s,r]}-\eta^A_v|$. By the same argument as in the second part of the proof of Lemma 28, this verifies that $|\eta^A_{]s,r]}-\eta^A_v|$ belongs to $\mathbb{M}_u$. □

As a final intermediary result, we use Lemmas 41, 42 and 43 to bound the expectation of the absolute value of the difference between the number of selected jumps in some interval $]s,r]$ and the number of selected jumps along the trivial grid (s,r) over $[s,r]$.

Lemma 44. *Consider a jump process P in $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$ and time points s,r in $\mathbb{R}_{\geq 0}$ such that $\max u\leq s<r$. Then for any subset A of $\mathcal{X}^2_{\neq}$,*

$$E^\sigma_P\left(|\eta^A_{]s,r]}-\eta^A_{(s,r)}|\,\middle|\,X_u=x_u\right)\leq\frac{1}{4}(r-s)^2\|\mathcal{Q}\|^2.$$

Proof. For all n in $\mathbb{N}$, we let $v_n:=(t_{n,0},\dots,t_{n,n})$ be the grid over $[s,r]$ that consists of $n+1$ evenly spaced time points, so with

$$t_{n,k}:=s+\frac{k}{n}(r-s)\quad\text{for all } k\in\{0,\dots,n\}.$$

Our construction of the sequence $(v_n)_{n\in\mathbb{N}}$ guarantees that $\lim_{n\to+\infty}\Delta(v_n)=\lim_{n\to+\infty}\frac{r-s}{n}=0$. Therefore, it follows from Lemma 43 that $(|\eta^A_{v_n}-\eta^A_{(s,r)}|)_{n\in\mathbb{N}}$ is a sequence in $\mathbb{S}_u\subseteq\mathbb{M}_u$ that converges point-wise to $|\eta^A_{]s,r]}-\eta^A_{(s,r)}|\in\mathbb{M}_u$. Because this convergence is not monotone, we need to resort to Theorem 35. To this end, we observe that for all n in $\mathbb{N}$,

$$|\eta^A_{v_n} - \eta^A_{(s,r)}| \leq \max\{\eta^A_{v_n}, \eta^A_{(r,s)}\} \leq \max\{\eta_{v_n}, \eta_{(r,s)}\} \leq \eta_{]s,r]},$$

where for the first inequality we used that $\eta^A_{v_n}$ and $\eta^A_{(s,r)}$ are non-negative, for the second inequality we used Eqn. (17) and for the final inequality we used Lemma 30. Due to the preceding, and because $E^\sigma_P(\eta_{]s,r]} \,|\, X_u = x_u) < +\infty$ due to Theorem 39, it follows from Theorem 35 that

$$E^\sigma_P\big(|\eta^A_{]s,r]} - \eta^A_{(s,r)}| \,\big|\, X_u = x_u\big) = \lim_{n\to+\infty} E^\sigma_P\big(|\eta^A_{v_n} - \eta^A_{(s,r)}| \,\big|\, X_u = x_u\big). \tag{26}$$

In order to bound the right-hand side of Eqn. (26), we fix any n in $\mathbb{N}$. Because $|\eta^A_{v_n} - \eta^A_{(s,r)}|$ is $\mathcal{A}_u$-simple and because E^σ_P extends E_P,

$$E^\sigma_P\big(|\eta^A_{v_n} - \eta^A_{(s,r)}| \,\big|\, X_u = x_u\big) = E_P\big(|\eta^A_{v_n} - \eta^A_{(s,r)}| \,\big|\, X_u = x_u\big).$$

Next, we recall from Lemma 42 that

$$|\eta^A_{v_n} - \eta^A_{(s,r)}| \leq 2\sum_{k=1}^{n-1} \mathbb{I}_{\{X_{t_{n,k-1}} \neq X_{t_{n,k}} \neq X_r\}},$$

and from Lemma 41 that each of the events occurring on the right-hand side belongs to $\mathcal{A}_u$, which makes the expression on the right-hand side an $\mathcal{A}_u$-simple variable. Due to (E4), this implies that

$$\begin{aligned} E^\sigma_P\big(|\eta^A_{v_n} - \eta^A_{(s,r)}| \,\big|\, X_u = x_u\big) &\leq E_P\left(2\sum_{k=1}^{n-1} \mathbb{I}_{\{X_{t_{n,k-1}} \neq X_{t_{n,k}} \neq X_r\}} \,\middle|\, X_u = x_u\right) \\ &= 2\sum_{k=1}^{n-1} P(\{X_{t_{n,k-1}} \neq X_{t_{n,k}} \neq X_r\} \,|\, X_u = x_u), \end{aligned}$$

where for the equality we used Eqn. (3). We replace the probabilities on the right-hand side of the equality with the upper bound in Lemma 41 and execute some straightforward manipulations, to yield

$$\begin{aligned} E^\sigma_P\big(|\eta^A_{v_n} - \eta^A_{(s,r)}| \,\big|\, X_u = x_u\big) &\leq 2\sum_{k=1}^{n-1} \frac{1}{4}(t_{n,k} - t_{n,k-1})(r - t_{n,k})\|\mathcal{Q}\|^2 = 2\sum_{k=1}^{n-1} \frac{1}{4}\frac{r-s}{n}\left(\frac{n-k}{n}(r-s)\right)\|\mathcal{Q}\|^2 \\ &= 2\sum_{k=1}^{n-1} \frac{1}{4}\frac{(r-s)^2}{n^2}(n-k)\|\mathcal{Q}\|^2 = \frac{1}{2}\frac{(r-s)^2}{n^2}\|\mathcal{Q}\|^2 \sum_{k=1}^{n-1}(n-k). \end{aligned}$$

Because $\sum_{k=1}^{n-1}(n-k) = \frac{(n-1)n}{2}$, it follows from this inequality that

$$E^\sigma_P\big(|\eta^A_{v_n} - \eta^A_{(s,r)}| \,\big|\, X_u = x_u\big) \leq \frac{1}{4}\frac{n-1}{n}(r-s)^2\|\mathcal{Q}\|^2. \tag{27}$$

Finally, it follows from Eqns. (26) and (27) that

$$E^\sigma_P\big(|\eta^A_{]s,r]} - \eta^A_{(s,r)}| \,\big|\, X_u = x_u\big) \leq \lim_{n\to+\infty} \frac{1}{4}\frac{n-1}{n}(r-s)^2\|\mathcal{Q}\|^2 = \frac{1}{4}(r-s)^2\|\mathcal{Q}\|^2. \quad \square$$

At long last, we can get around to proving Theorem 40, and we do this with the help of Lemmas 38, 43 and 44 and some properties of the Lebesgue integral $E^\sigma_P(\bullet \,|\, X_u = x_u)$.

Proof of Theorem 40. Due to Lemma 38, it suffices to prove that for any jump process P in $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$,

$$E^\sigma_P\big(|\eta^A_{]s,r]} - \eta^A_v| \,\big|\, X_u = x_u\big) \leq \frac{1}{4}\Delta(v)(r-s)\|\mathcal{Q}\|^2. \tag{28}$$

Hence, we fix any jump process P in $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$.

Let us enumerate the time points in v as $(t_0, \ldots, t_n)$. It follows immediately from the definition of $\eta^A_{]s,r]}$ and η^A_v that

$$\eta^A_{]s,r]} = \sum_{k=1}^{n} \eta^A_{]t_{k-1},t_k]} \quad \text{and} \quad \eta^A_v = \sum_{k=1}^{n} \eta^A_{(t_{k-1},t_k)},$$

and therefore

$$|\eta^A_{]s,r]} - \eta^A_v| = \left|\sum_{k=1}^n \eta^A_{]t_{k-1},t_k]} - \sum_{k=1}^n \eta^A_{(t_{k-1},t_k)}\right| = \left|\sum_{k=1}^n \left(\eta^A_{]t_{k-1},t_k]} - \eta^A_{(t_{k-1},t_k)}\right)\right| \leq \sum_{k=1}^n |\eta^A_{]t_{k-1},t_k]} - \eta^A_{(t_{k-1},t_k)}|.$$

Due to Lemma 43, $|\eta^A_{]s,r]} - \eta^A_v|$ belongs to the domain $\mathbb{M}_u$, and so does $|\eta^A_{]s,r]} - \eta^A_{(t_{k-1},t_k)}|$ for all k in $\{1,\dots,n\}$. Furthermore, for all k in $\{1,\dots,n\}$, $E^\sigma_P(|\eta^A_{]t_{k-1},t_k]} - \eta^A_{(t_{k-1},t_k)}| \,|\, X_u = x_u) \geq 0$ due to (LI1). Hence, due to the monotonicity (LI4) and additivity (LI3) of E^σ_P, it follows from the preceding inequality that

$$\begin{aligned} E^\sigma_P\left(|\eta^A_{]s,r]} - \eta^A_v| \,\middle|\, X_u = x_u\right) &\leq E^\sigma_P\left(\sum_{k=1}^n |\eta^A_{]t_{k-1},t_k]} - \eta^A_{(t_{k-1},t_k)}| \,\middle|\, X_u = x_u\right) \\ &= \sum_{k=1}^n E^\sigma_P\left(|\eta^A_{]t_{k-1},t_k]} - \eta^A_{(t_{k-1},t_k)}| \,\middle|\, X_u = x_u\right). \end{aligned}$$

We substitute the expectations on the right-hand side of the equality by the upper bound in Lemma 44, to yield

$$E^\sigma_P\left(|\eta^A_{]s,r]} - \eta^A_v| \,\middle|\, X_u = x_u\right) \leq \sum_{k=1}^n \frac{1}{4}(t_k - t_{k-1})^2 \|\mathcal{Q}\|^2.$$

Finally, because $(t_k - t_{k-1}) \leq \Delta(v)$ for all k in $\{1,\dots,n\}$ and $\sum_{k=1}^n (t_k - t_{k-1}) = (t_n - t_0) = (r-s)$, it follows from this inequality that

$$E^\sigma_P\left(|\eta^A_{]s,r]} - \eta^A_v| \,\middle|\, X_u = x_u\right) \leq \frac{1}{4}\Delta(v)\|\mathcal{Q}\|^2 \sum_{k=1}^n (t_k - t_{k-1}) = \frac{1}{4}\Delta(v)(r-s)\|\mathcal{Q}\|^2,$$

and this is precisely the inequality in Eqn. (28). □

Finally, it follows immediately from Theorem 40 that in the particular case of the number of selected jumps, the convergence in Theorem 37 – and, if $A = \mathcal{X}^2_{\neq}$ and $v_n \subseteq v_{n+1}$ for all n in $\mathbb{N}$, also that in Theorem 36 – is not conservative but tight, and that we have limits instead of a limit superior and a limit inferior.

Corollary 45. *Consider an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$ and s, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$, and fix a subset A of $\mathcal{X}^2_{\neq}$. Then for any sequence $(v_n)_{n\in\mathbb{N}}$ of grids over $[s,r]$ such that $\lim_{n\to+\infty} \Delta(v_n) = 0$,*

$$\underline{E}^\sigma_{\mathcal{P}}\left(\eta^A_{]s,r]} \,\middle|\, X_u = x_u\right) = \lim_{n\to+\infty} \underline{E}_{\mathcal{P}}\left(\eta^A_{v_n} \,\middle|\, X_u = x_u\right) \quad \text{and} \quad \overline{E}^\sigma_{\mathcal{P}}\left(\eta^A_{]s,r]} \,\middle|\, X_u = x_u\right) = \lim_{n\to+\infty} \overline{E}_{\mathcal{P}}\left(\eta^A_{v_n} \,\middle|\, X_u = x_u\right).$$

Proof. Follows immediately from Theorem 40. □

Theorem 40 and Corollary 45 allow us to compute – or approximate – the lower/upper expected number of selected jumps with respect to an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Obviously, however, this requires that we can compute $\underline{E}_{\mathcal{P}}(\eta^A_v \,|\, X_u = x_u)$ and/or $\overline{E}_{\mathcal{P}}(\eta^A_v \,|\, X_u = x_u)$ for any grid v over $[s,r]$. In Section 6.1 further on, we will identify conditions on $\mathcal{P}$ under which this is possible as well as present other, more explicit methods for computing lower/upper expectations of the number of selected jumps.

5.2. Convergence for the expected Riemann sums

Next, we set out to prove a (tight) convergence result for the lower/upper expectation of $\int_s^r f(X_t)\,\mathrm{d}t$. Fix some sequence $(v_n)_{n\in\mathbb{N}}$ of grids over $[s,r]$ with $\lim_{n\to+\infty} \Delta(v_n) = 0$ – take, for example, the sequence with grids defined for all n in $\mathbb{N}$ by

$$v_n := \left(s, s + \frac{r-s}{n}, s + 2\frac{r-s}{n}, \dots, s + n\frac{r-s}{n}\right).$$

Then by Lemma 32, the corresponding sequence $(\langle f \rangle_{v_n})_{n\in\mathbb{N}}$ of Riemann sums is a uniformly-bounded sequence of simple variables that converges point-wise to $\int_s^r f(X_t)\,\mathrm{d}t$. Consequently, we could invoke Theorem 37 – with g the constant (and therefore bounded and measurable) real variable $(r-s)\|f\|$ – to conservatively bound the lower/upper expectation of $\int_s^r f(X_t)\,\mathrm{d}t$. Here too, however, the approximation in Theorem 37 is actually tight, and the limit superior and limit inferior in the statement are actually limits. The reason for this is the following important result; it is a special case of Proposition 6.38 in [13], but we provide a formal proof for the sake of completeness.

Theorem 46. *Consider an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$ and time points s, r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$, and fix a real-valued function f on $\mathcal{X}$. Then for any grid v over $[s, r]$,*

$$\left| \underline{E}^{\sigma}_{\mathcal{P}} \left(\int_s^r f(X_t)\,\mathrm{d}t \,\middle|\, X_u = x_u \right) - \underline{E}_{\mathcal{P}}(\langle f \rangle_v \mid X_u = x_u) \right| \leq \frac{1}{2}(\max f - \min f)\Delta(v)(r-s)\|\mathcal{Q}\|,$$

and the same inequality holds for the upper expectation.

Our proof for Theorem 46 uses the following rather obvious intermediary result, which is a special case of Lemma 6.37 in [33]; again, we provide a formal proof here for the sake of completeness.

Lemma 47. *Consider time points s, r in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, and a real-valued function f on $\mathcal{X}$. Then for any grid v over $[s, r]$,*

$$\left| \int_s^r f(X_t)\,\mathrm{d}t - \langle f \rangle_v \right| \leq (\max f - \min f)\Delta(v)\eta_{[s,r]}.$$

Proof. We enumerate the time points in v as $(t_0, \ldots, t_n)$. Then

$$\langle f \rangle_v = \sum_{k=1}^{n} f(X_{t_k})(t_k - t_{k-1}),$$

by definition, and it is a well-known property of Riemann integrals [see 26, Theorem 24.19] that

$$\int_s^r f(X_t)\,\mathrm{d}t = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} f(X_t)\,\mathrm{d}t.$$

It follows from these two equalities and the triangle inequality that

$$\begin{aligned} \left| \int_s^r f(X_t)\,\mathrm{d}t - \langle f \rangle_v \right| &= \left| \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} f(X_t)\,\mathrm{d}t - \sum_{k=1}^{n} f(X_{t_k})(t_k - t_{k-1}) \right| \\ &\leq \sum_{k=1}^{n} \left| \int_{t_{k-1}}^{t_k} f(X_t)\,\mathrm{d}t - f(X_{t_k})(t_k - t_{k-1}) \right|. \end{aligned} \tag{29}$$

We now investigate the terms of the sum on the right-hand side of this inequality individually. To this end, we fix some k in $\{1, \ldots, n\}$ and ω in Ω. Because the Riemann integral is additive,

$$\begin{aligned} \left| \int_{t_{k-1}}^{t_k} f\big(\omega(t)\big)\,\mathrm{d}t - f\big(\omega(t_k)\big)(t_{k+1} - t_k) \right| &= \left| \int_{t_{k-1}}^{t_k} f\big(\omega(t)\big)\,\mathrm{d}t - \int_{t_{k-1}}^{t_k} f\big(\omega(t_k)\big)\,\mathrm{d}t \right| \\ &= \left| \int_{t_{k-1}}^{t_k} f\big(\omega(t)\big) - f\big(\omega(t_k)\big)\,\mathrm{d}t \right|. \end{aligned} \tag{30}$$

First, we consider the case that ω is constant over $[t_{k-1}, t_k]$, meaning that $\eta_{]t_{k-1},t_k]}(\omega) = 0$. Then for all t in $[t_{k-1}, t_k]$, $\omega(t) = \omega(t_k)$, and hence $f\big(\omega(t)\big) = f\big(\omega(t_k)\big)$. Consequently, using Eqn. (30),

$$\left| \int_{t_{k-1}}^{t_k} f\big(\omega(t)\big)\,\mathrm{d}t - f\big(\omega(t_k)\big)(t_k - t_{k-1}) \right| = \left| \int_{t_{k-1}}^{t_k} 0\,\mathrm{d}t \right| = 0 = (\max f - \min f)\Delta(v)\eta_{]t_{k-1},t_k]}(\omega).$$

Second, we consider the case that ω is not constant over $[t_{k-1}, t_k]$, in the sense that $\eta_{]t_{k-1},t_k]}(\omega) > 0$. Observe that

$$\big(\forall t \in [t_{k-1}, t_k]\big)\ \big|f\big(\omega(t)\big) - f\big(\omega(t_k)\big)\big| \leq (\max f - \min f).$$

It follows from Eqn. (30), this inequality, the counterpart of (LI5) for the Riemann integral and the monotonicity of the Riemann integral that

$$\begin{aligned}\left|\int_{t_{k-1}}^{t_k} f\big(\omega(t)\big)\,\mathrm{d}t - f\big(\omega(t_k)\big)(t_{k+1}-t_k)\right| &= \left|\int_{t_{k-1}}^{t_k} f\big(\omega(t)\big) - f\big(\omega(t_k)\big)\,\mathrm{d}t\right| \\ &\leq \int_{t_{k-1}}^{t_k} \left|f\big(\omega(t)\big) - f\big(\omega(t_k)\big)\right|\mathrm{d}t \\ &\leq \int_{t_{k-1}}^{t_k} (\max f - \min f)\,\mathrm{d}t \\ &= (\max f - \min f)(t_k - t_{k-1}) \leq (\max f - \min f)\Delta(v),\end{aligned}$$

where for the final inequality we used that $(t_k - t_{k-1}) \leq \Delta(v)$. Hence, and because $\eta_{]t_{k-1},t_k]}(\omega) \geq 1$, it follows that

$$\left|\int_{t_{k-1}}^{t_k} f\big(\omega(t)\big)\,\mathrm{d}t - f\big(\omega(t_k)\big)(t_{k+1}-t_k)\right| \leq (\max f - \min f)\Delta(v)\eta_{]t_{k-1},t_k]}(\omega).$$

Thus, we have shown that for all k in $\{1,\ldots,n\}$,

$$\left|\int_{t_{k-1}}^{t_k} f(X_t)\,\mathrm{d}t - f(X_{t_k})(t_k - t_{k-1})\right| \leq (\max f - \min f)\Delta(v)\eta_{]t_{k-1},t_k]}.$$

It follows from this and Eqn. (29) that

$$\left|\int_{s}^{r} f(X_t)\,\mathrm{d}t - \langle f\rangle_v\right| \leq \sum_{k=1}^{n}(\max f - \min f)\Delta(v)\eta_{]t_{k-1},t_k]} = (\max f - \min f)\Delta(v)\eta_{]s,r]},$$

where for the equality we used that $\sum_{k=1}^{n}\eta_{]t_{k-1},t_k]} = \eta_{]s,r]}$. □

Proof of Theorem 46. For any jump process P in $\mathcal{P}$, it follows from Lemma 47, the monotonicity and homogeneity of E_P^σ and Theorem 39 that

$$\begin{aligned}E_P^\sigma\left(\left|\int_s^r f(X_t)\,\mathrm{d}t - \langle f\rangle_v\right| \,\middle|\, X_u = x_u\right) &\leq E_P^\sigma\big((\max f - \min f)\Delta(v)\eta_{]s,r]} \,\big|\, X_u = x_u\big) \\ &= (\max f - \min f)\Delta(v)E_P^\sigma(\eta_{]s,r]} \,|\, X_u = x_u) \\ &\leq \frac{1}{2}(\max f - \min f)\Delta(v)(r-s)\|\mathcal{Q}\|.\end{aligned}$$

Since this inequality holds for any jump process P in $\mathcal{P}$, it implies the inequalities in the statement due to Lemma 38. □

Clearly, Theorem 46 implies that in the case of a measurable variable that can be written as the integral of $f(X_t)$ over some interval $[s,r]$, the bounds in Theorem 37 are actually tight; even more, the lower/upper expectation of $\langle f\rangle_n$ converges to a limit.

Corollary 48. *Consider an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Fix some u in $\mathcal{U}$, x_u in $\mathcal{X}_u$, s,r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$ and f in $\mathbb{R}^{\mathcal{X}}$. Then for any sequence $(v_n)_{n\in\mathbb{N}}$ of grids over $[s,r]$ such that $\lim_{n\to+\infty}\Delta(v_n) = 0$,*

$$\underline{E}_{\mathcal{P}}^\sigma\left(\int_s^r f(X_t)\,\mathrm{d}t \,\middle|\, X_u = x_u\right) = \lim_{n\to+\infty}\underline{E}_{\mathcal{P}}\big(\langle f\rangle_{v_n} \,\big|\, X_u = x_u\big)$$

and

$$\overline{E}^{\sigma}_{\mathcal{P}}\left(\int_s^r f(X_t)\,\mathrm{d}t \,\middle|\, X_u = x_u\right) = \lim_{n\to+\infty} \overline{E}_{\mathcal{P}}\big(\langle f\rangle_{v_n} \,\big|\, X_u = x_u\big).$$

We can make a similar remark here as the one we made right after Corollary 45: Theorem 46 and Corollary 48 allow us to compute – or approximate – the lower/upper expectation of $\int_s^r f(X_t)\,\mathrm{d}t$ for an imprecise jump process $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, given that we can determine the lower/upper expectation of $\langle f\rangle_v$ for any sufficiently fine grid v over $[s,r]$. Here too, we leave a more thorough discussion, and improved algorithms, for Section 6.1 further on.

6. Computational methods

In Section 5.1, we learned that we can use Theorem 40 or Corollary 45 to determine the lower/upper expectation of the number of selected jumps $\eta^A_{]s,r]}$ if we can determine the lower/upper expectation of η^A_v for a suitably fine grid v over $[s,r]$, and we learned a similar thing for variables of the form of a Riemann integral in Section 5.2, provided we can determine the lower/upper expectation of $\langle f\rangle_v$. In this section, we go deeper into this idea of determining these lower/upper expectations. In Section 6.1, we investigate conditions on the set of rate operators $\mathcal{Q}$ under which we can compute the lower/upper expectation of η^A_v and $\langle f\rangle_v$ with respect to $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Quite surprisingly, this provides the ideal starting point for a computational method that works under fewer conditions on $\mathcal{Q}$ and for any imprecise jump process $\mathcal{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, as we will see in Section 6.2.

6.1. Computing the lower expectation of simple variables with a specific sum decomposition

Fix a sequence of time points u in $\mathcal{U}$, a state instantiation x_u in $\mathcal{X}_u$ and time points s,r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$, and let $v = (t_0,\ldots,t_n)$ be any grid over $[s,r]$. The aim of this section is to investigate under which conditions on the imprecise jump process $\mathcal{P}$ we can determine $\underline{E}_{\mathcal{P}}(f(X_v) \,|\, X_u = x_u)$ and/or $\overline{E}_{\mathcal{P}}(f(X_v) \,|\, X_u = x_u)$. We are particularly interested in the case that $f(X_v) = \eta^A_v$ or $f(X_v) = \langle f\rangle_v$; we cover both of these cases if we point our attention to a generic $\mathcal{A}_u$-simple variable $f(X_v)$ with a sum decomposition of the form

$$f(X_v) = \sum_{k=1}^{n} g_k\big(X_{t_{k-1}}, X_{t_k}\big), \tag{31}$$

where $g_1, \ldots, g_n$ are real-valued functions on $\mathcal{X}^2$. This specific sum decomposition suffices because both η^A_v and $\langle f\rangle_v$ are of this form: it follows from Eqn. (16) that η^A_v is of this form with $g_k = \mathbb{I}_A$ for all k in $\{1,\ldots,n\}$, and from Eqn. (18) that $\langle f\rangle_v$ is of this form with $g_k(X_{t_{k-1}}, X_{t_k}) = (t_k - t_{k-1}) f(X_{t_k})$ for all k in $\{1,\ldots,n\}$. Even more, both $-\eta^A_v$ and $-\langle f\rangle_v$ are obviously of this form as well; due to conjugacy, it therefore suffices to focus on determining the lower expectation of a generic $\mathcal{A}_u$-simple variable $f(X_v)$ with a sum decomposition of the form in Eqn. (31).

6.1.1. The law of iterated lower expectations

Krak et al. [2, Section 9] essentially argue – but see [13, Section 4.1] for a more generic and detailed explanation – that determining lower expectations becomes simpler whenever $\mathcal{P}$ satisfies the 'law of iterated lower expectations', which generalises the well-known 'law of iterated expectations'. To state this law of iterated expectations in the setting of jump processes, we need to introduce some additional notation: for all u in $\mathcal{U}$, v in $\mathcal{U}_{\mathrm{ne}}$ such that $u \preccurlyeq v$ and f in $\mathbb{R}^{\mathcal{X}_v}$, we define the $\mathbb{S}_u$-simple variable

$$E_P(f(X_v) \,|\, X_u) := \sum_{x_u \in \mathcal{X}_u} E_P(f(X_v) \,|\, X_u = x_u)\mathbb{I}_{\{X_u = x_u\}}. \tag{32}$$

Proposition 49. *Consider any jump process P. Then for all u in $\mathcal{U}$, v, w in $\mathcal{U}_{\mathrm{ne}}$ such that $u \prec v$ and $u \cup v \preccurlyeq w$, f in $\mathbb{R}^{\mathcal{X}_w}$ and x_u in $\mathcal{X}_u$,*

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

For imprecise jump processes, a result similar to Proposition 49 exists for the corresponding lower expectation, but it is a bit more involved. To see why, we let $\mathcal{P}$ be any imprecise jump process. First, we generalise the variable defined in Eqn. (32): for all u in $\mathcal{U}$, v in $\mathcal{U}_{\mathrm{ne}}$ such that $u \preccurlyeq v$ and f in $\mathbb{R}^{\mathcal{X}_v}$, we define the $\mathbb{S}_u$-simple variable

$$\underline{E}_{\mathcal{P}}(f(X_v) \,|\, X_u) := \sum_{x_u \in \mathcal{X}_u} \underline{E}_{\mathcal{P}}(f(X_v) \,|\, X_u = x_u)\mathbb{I}_{\{X_u = x_u\}}.$$

Fix some u in $\mathcal{U}$, some v, w in $\mathcal{U}_{\mathrm{ne}}$ such that $u \prec v$ and $u \cup v \preccurlyeq w$, some real-valued function f on $\mathcal{X}_w$ and some x_u in $\mathcal{X}_u$. It is obvious that

$$\underline{E}_{\mathcal{P}}(f(X_w) \mid X_{u\cup v}) \leq E_P(f(X_w) \mid X_{u\cup v}) \quad \text{for all } P \in \mathcal{P},$$

and we infer from this inequality, Proposition 49 and (E4) that for any jump process P in the imprecise jump process $\mathcal{P}$,

$$E_P(f(X_w) \mid X_u = x_u) = E_P(E_P(f(X_w) \mid X_{u\cup v}) \mid X_u = x_u) \geq E_P(\underline{E}_{\mathcal{P}}(f(X_w) \mid X_{u\cup v}) \mid X_u = x_u).$$

Inequalities are preserved under taking the infimum, so we can conclude that

$$\underline{E}_{\mathcal{P}}(f(X_w) \mid X_u = x_u) \geq \underline{E}_{\mathcal{P}}\big(\underline{E}_{\mathcal{P}}(f(X_w) \mid X_{u\cup v}) \,\big|\, X_u = x_u\big).$$

In the special case that this inequality always holds with equality, we say that $\mathcal{P}$ satisfies the law of iterated lower expectations.

Definition 50. An imprecise jump process $\mathcal{P}$ satisfies the *law of iterated lower expectations* if for all u in $\mathcal{U}$, v, w in $\mathcal{U}_{\text{ne}}$ such that $u \prec v$ and $u \cup v \preccurlyeq w$, f in $\mathbb{R}^{\mathcal{X}_w}$ and x_u in $\mathcal{X}_u$,

$$\underline{E}_{\mathcal{P}}(f(X_w) \mid X_u = x_u) = \underline{E}_{\mathcal{P}}\big(\underline{E}_{\mathcal{P}}(f(X_w) \mid X_{u\cup v}) \,\big|\, X_u = x_u\big). \tag{33}$$

Unfortunately, many imprecise jump processes do not satisfy this law of iterated lower expectations; see for example [2, Example 9.2]. However, Krak et al. [2, Theorem 6.5] show that whenever the non-empty and bounded set $\mathcal{Q}$ of rate matrices is convex and has separately specified rows,[10] $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ satisfies this law – and it is important to stress that $\mathbb{P}^{\text{M}}_{\mathcal{M},\mathcal{Q}}$ or $\mathbb{P}^{\text{HM}}_{\mathcal{M},\mathcal{Q}}$ may not [2, Example 9.2].

Proposition 51. *If $\mathcal{Q}$ is convex and has separately specified rows, then $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ satisfies the law of iterated lower expectations.*

Example 52. It follows immediately from its definition in Example 13 that the set $\mathcal{Q}$ is convex, and we have seen in Example 21 that $\mathcal{Q}$ also has separately specified rows. Hence, our model $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ of the power network is a Markovian jump process that satisfies the law of iterated lower expectations.

Crucially, if an imprecise jump process $\mathcal{P}$ satisfies the law of iterated lower expectations, then we can compute conditional lower expectations of the general form $\underline{E}_{\mathcal{P}}(f(X_v) \mid X_u = x_u)$ – with $u \preccurlyeq v$ – through backwards recursion. This recursive method goes back to earlier work on imprecise Markov chains – see [35, Section 4.5] and [36, Section 3]. In this contribution, we will only give this backwards recursive method for a special case. First, we make the additional assumption that $\mathcal{P}$ is Markovian. Second, we only consider variables $f(X_v)$ that have a sum decomposition of the form in Eqn. (31). Third, we assume that the sequence of time points u in the conditioning event is non-empty, and that the last time point of u equals the first time point of v; formally, we do this by considering conditioning events of the form $\{X_u = x_u, X_s = x\}$ with $u \prec (s)$ and non-empty sequences of time points v such that $t_0 = s$. This third assumption is not strictly necessary, but it does allow for more elegant statements – we refer the interested reader to Lemma 6.44 in [13] for ideas on how to remove this condition. We prove in Appendix A that, under these three additional assumptions, Proposition 4.11 in [13] specialises to the following result.

Proposition 53. *Consider a Markovian imprecise jump process $\mathcal{P}$ that satisfies the law of iterated lower expectations. Fix time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$, a grid $v = (t_0, \ldots, t_n)$ over $[s, r]$ and real-valued functions $g_1, \ldots, g_n$ on $\mathcal{X}^2$. Let $f_0, \ldots, f_n$ be the real-valued functions on $\mathcal{X}$ defined by the initial condition $f_n := 0$ and, for all k in $\{0, \ldots, n-1\}$, by the recursive relation*

$$f_k \colon \mathcal{X} \to \mathbb{R} \colon x \mapsto \underline{E}_{\mathcal{P}}(g_{k+1}(x, X_{t_{k+1}}) + f_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = x).$$

Then for all x in $\mathcal{X}$, u in $\mathcal{U}_{\prec s}$ and x_u in $\mathcal{X}_u$,

$$\underline{E}_{\mathcal{P}}\left(\sum_{k=1}^{n} g_k(X_{t_{k-1}}, X_{t_k}) \,\middle|\, X_u = x_u, X_s = x\right) = f_0(x).$$

6.1.2. The lower envelope $\underline{Q}$ of $\mathcal{Q}$ and its exponential

Obviously, Proposition 53 is only useful if we can determine lower expectations of the form

$$\underline{E}_{\mathcal{P}}(h(X_{t+\Delta}) \mid X_u = x_u, X_t = x).$$

[10] The published result lacks the condition that $\mathcal{Q}$ has separately specified rows, but this is implicitly used in the proof given there; the interested reader can find the corrected statement and proof in [13, Theorem 3.88] or [34, Theorem 5.32].

As we will presently see, this turns out to be the case for any imprecise jump process $\mathcal{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, at least if $\mathcal{Q}$ has separately specified rows. To explain how this works, we need to introduce the operator exponential $e^{\Delta \underline{Q}}$ of the lower envelope $\underline{Q}$ of $\mathcal{Q}$.

Given a non-empty and bounded set $\mathcal{R} \subseteq \mathfrak{Q}$ of rate matrices, Krak et al. [2, Eqn. (38)] define the corresponding lower envelope $\underline{R}_{\mathcal{R}} : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ as the operator that maps any real-valued function f on $\mathcal{X}$ to

$$\underline{R}_{\mathcal{R}} f : \mathcal{X} \to \mathbb{R} : x \mapsto [\underline{R}_{\mathcal{R}} f](x) := \inf\big\{[Qf](x) : Q \in \mathcal{R}\big\}; \tag{34}$$

the upper envelope $\overline{R}_{\mathcal{R}} : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ is defined similarly, and is related to the lower envelope $\underline{R}_{\mathcal{R}}$ by conjugacy: $\overline{R}_{\mathcal{R}} f := -\underline{R}_{\mathcal{R}}(-f)$ for any real-valued function f on $\mathcal{X}$. Krak et al. [2, Proposition 7.5] show that the lower envelope $\underline{R}_{\mathcal{R}}$ is a so-called lower (transition) rate operator – a generalisation of the notion of a rate matrix that relaxes linearity to non-negative homogeneity and superadditivity.

Definition 54. A *lower rate operator* $\underline{R}$ is an operator $\underline{R} : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ such that

LR1. $\underline{R}(\lambda f) = \lambda \underline{R} f$ for all λ in $\mathbb{R}_{\geq 0}$ and f in $\mathbb{R}^{\mathcal{X}}$; [non-negative homogeneity]
LR2. $\underline{R}(f+g) \geq \underline{R} f + \underline{R} g$ for all f, g in $\mathbb{R}^{\mathcal{X}}$; [superadditivity]
LR3. $[\underline{R}\mathbb{I}_y](x) \geq 0$ for all x, y in $\mathcal{X}$ such that $x \neq y$; [non-negative off-diagonal components]
LR4. $\underline{R}\mu = 0$ for any constant function μ on $\mathcal{X}$. [zero on constant functions]

In order to shorten our notation, we will denote the lower envelope of $\mathcal{Q}$ by $\underline{Q} := \underline{R}_{\mathcal{Q}}$, and its conjugate upper envelope by $\overline{Q} := \overline{R}_{\mathcal{Q}}$.

Example 55. In Example 13, we defined a set of rate matrices through inequalities on the off-diagonal components. For this reason, the lower envelope $\underline{Q}$ of this set $\mathcal{Q}$ can be easily determined: for all f in $\mathbb{R}^{\mathcal{X}}$ and x in $\mathcal{X}$, $[\underline{Q} f](x)$ can be determined either through a straightforward case study or through linear programming [see also 10, p. 293].

One interesting property of the lower envelope $\underline{Q}$ of $\mathcal{Q}$ is that its operator norm $\|\underline{Q}\|$ is equal to the supremum of the norms of the rate matrices in $\mathcal{Q}$.

Lemma 56. *For any non-empty and bounded set $\mathcal{R}$ of rate matrices, $\|\underline{R}_{\mathcal{R}}\| = \|\mathcal{R}\|$.*

Proof. Let us denote the lower envelope of $\mathcal{R}$ by $\underline{R} := \underline{R}_{\mathcal{R}}$. For any rate operator Q in $\mathcal{R}$, it follows from Eqn. (6) and the non-positivity of the off-diagonal elements of Q that

$$\|Q\| = 2\max\big\{-Q(x,x) : x \in \mathcal{X}\big\} = -2\min\big\{Q(x,x) : x \in \mathcal{X}\big\} = -2\min\big\{[Q\mathbb{I}_x](x) : x \in \mathcal{X}\big\}.$$

Similarly, by (LR7) in [13, p. 111] or Proposition 4 in [33],

$$\|\underline{R}\| = 2\max\big\{-[\underline{R}\mathbb{I}_x](x) : x \in \mathcal{X}\big\} = -2\min\big\{[\underline{R}\mathbb{I}_x](x) : x \in \mathcal{X}\big\}.$$

To obtain the equality in the statement, we substitute the definition of $\underline{R}$ in the preceding equality, change the order of minimisations and substitute our expression for $\|Q\|$:

$$\begin{aligned}\|\underline{R}\| &= -2\min\Big\{\inf\big\{[Q\mathbb{I}_x](x) : Q \in \mathcal{R}\big\} : x \in \mathcal{X}\Big\} = -2\inf\Big\{\min\big\{[Q\mathbb{I}_x](x) : x \in \mathcal{X}\big\} : Q \in \mathcal{R}\Big\} \\ &= \sup\big\{-2\min\big\{[Q\mathbb{I}_x](x) : x \in \mathcal{X}\big\} : Q \in \mathcal{R}\big\} = \sup\big\{\|Q\| : Q \in \mathcal{R}\big\} = \|\mathcal{R}\|. \quad \square\end{aligned}$$

More importantly, the 'operator exponential' $e^{\Delta \underline{Q}}$ of the lower envelope $\underline{Q}$ of $\mathcal{Q}$ plays the same role for the imprecise jump processes $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ as the matrix exponential $e^{\Delta Q}$ of the rate matrix $\mathcal{Q}$ does for the homogeneous Markovian jump process $P_{p,Q}$ in Eqn. (9). We formalise this statement in Lemma 57 further on, but we need to introduce the operator exponential of a lower rate operator before we can do so.

Consider a lower rate operator $\underline{R}$. Krak et al. [2, Corollary 7.11 and Theorem 7.12] – see also Section 6 in [37] – prove that for all Δ in $\mathbb{R}_{\geq 0}$,

$$e^{\Delta \underline{R}} : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}} : f \mapsto e^{\Delta \underline{Q}} f := \lim_{n \to +\infty} \left(I + \frac{\Delta}{n}\underline{R}\right)^n f$$

is well-defined. If the lower rate operator $\underline{R}$ is linear – or equivalently, if $\underline{R}$ is a rate matrix – then $e^{\Delta \underline{R}}$ is the (matrix) exponential of $\underline{R}$; for this reason, we call $e^{\Delta \underline{R}}$ the *(operator) exponential of* $\underline{R}$.

Now recall from Eqn. (9) that the matrix exponential $e^{\Delta Q}$ of the rate matrix Q can be used to compute specific expectations for a homogeneous Markovian jump process with rate matrix Q. Interestingly, the (operator) exponential $e^{\Delta \underline{Q}}$ of the lower envelope $\underline{Q}$ of $\mathcal{Q}$ plays a similar role for the homogeneous Markovian imprecise jump processes $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$, in the sense that whenever $\mathcal{Q}$ has separately specified rows, the operator exponential of its lower envelope $\underline{Q}$ gives the (lower) expectation of any real variable that depends on the state of the system at a single future time point [2, Corollary 8.3].

Lemma 57. *If $\mathcal{Q}$ has separately specified rows, then for all t, Δ in $\mathbb{R}_{\geq 0}$, f in $\mathbb{R}^{\mathcal{X}}$, x in $\mathcal{X}$, u in $\mathcal{U}_{\prec t}$ and x_u in $\mathcal{X}_u$,*

$$\underline{E}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}(f(X_{t+\Delta}) \mid X_u = x_u, X_t = x) = \left[e^{\Delta \underline{Q}} f\right](x) \quad \text{and} \quad \underline{E}_{\mathcal{M},\mathcal{Q}}(f(X_{t+\Delta}) \mid X_u = x_u, X_t = x) = \left[e^{\Delta \underline{Q}} f\right](x),$$

and hence, more generally, for any imprecise jump process $\mathcal{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$,

$$\underline{E}_{\mathcal{P}}(f(X_{t+\Delta}) \mid X_u = x_u, X_t = x) = \left[e^{\Delta \underline{Q}} f\right](x).$$

Finally, combining Proposition 53 and Lemma 57 gives us the straightforward backwards-recursive method we are after, although it only works for $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ and in case $\mathcal{Q}$ is convex and has separately specified rows.

Corollary 58. *Fix time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$, a grid $v = (t_0, \ldots, t_n)$ over $[s, r]$ and real-valued functions $g_1, \ldots, g_n$ on $\mathcal{X}^2$. Let $\underline{h}_0, \ldots, \underline{h}_n$ be the real-valued functions on $\mathcal{X}$ defined by the initial condition $\underline{h}_n := 0$ and, for all k in $\{0, \ldots, n-1\}$, by the recursive relation*

$$\underline{h}_k \colon \mathcal{X} \to \mathbb{R} \colon x \mapsto \left[e^{(t_{k+1}-t_k)\underline{Q}}(g_{k+1}(x, \bullet) + \underline{h}_{k+1})\right](x).$$

If $\mathcal{Q}$ is convex and has separately specified rows, then for all x in $\mathcal{X}$, u in $\mathcal{U}_{\prec s}$ and x_u in $\mathcal{X}_u$,

$$\underline{E}_{\mathcal{M},\mathcal{Q}}\left(\sum_{k=1}^{n} g_k(X_{t_{k-1}}, X_{t_k}) \,\middle|\, X_u = x_u, X_s = x\right) = \underline{h}_0(x).$$

Proof. Because $\mathcal{Q}$ is convex and has separately specified rows, $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$ is Markovian due to Corollary 20 and satisfies the law of iterated lower expectations due to Proposition 51. Hence, the equality in the statement follows immediately from Lemma 57 and Proposition 53. □

At first sight, Corollary 58 simply shifts the problem we had with Proposition 53: we now need to determine $[e^{\Delta \underline{Q}} h](x)$ instead of $\underline{E}_{\mathcal{P}}(h(X_{t+\Delta}) \mid X_t = x)$ for h in $\mathbb{R}^{\mathcal{X}}$. This is not an issue though, because there are plenty of methods to determine $e^{\Delta \underline{Q}} h$. We will not treat these computational methods in detail, but the interested reader may refer to [2, Proposition 8.5], [33] or [13, Section 4.2]. Because it is relevant further on, we do mention that it is standard practice to use 'Euler's method' to approximate $e^{\Delta \underline{Q}}$ with $(I + \frac{\Delta}{n}\underline{Q})^n$ for some (sufficiently large) natural number n. Krak et al. [2, Lemma E.8] give an upper bound on the resulting error, but here we only need this bound for the special case that $n = 1$ – in fact, we repeat the slightly tighter bound from [13, Lemma 4.16].

Lemma 59. *Consider a lower rate operator $\underline{R}$. Then for any Δ in $\mathbb{R}_{\geq 0}$,*

$$\left\| e^{\Delta \underline{R}} - (I + \Delta \underline{R}) \right\| \leq \frac{1}{2}\Delta^2 \|\underline{R}\|^2.$$

If $\mathcal{Q}$ is not convex and/or does not have separately specified rows, we can still use the method in Corollary 58 to give a lower bound on the lower expectation, even for imprecise jump processes other than $\mathbb{P}_{\mathcal{M},\mathcal{Q}}$. To see this, consider an imprecise jump process $\mathcal{P}$ such that $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, and some bounded set $\mathcal{R}$ of rate matrices that includes $\mathcal{Q}$. Recall from Section 3 that every jump process P that is consistent with $\mathcal{Q} \subseteq \mathcal{R}$ is also consistent with $\mathcal{R}$. Consequently, $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{R}}$, and therefore

$$\underline{E}_{\mathcal{P}}(\bullet \mid \bullet) \geq \underline{E}_{\mathcal{M},\mathcal{Q}}(\bullet \mid \bullet) \geq \underline{E}_{\mathcal{M},\mathcal{R}}(\bullet \mid \bullet).$$

Hence, if $\mathcal{R}$ is convex and has separately specified rows, then we can use the method in Corollary 58 to compute a lower bound on the lower expectation of interest, provided we can evaluate the lower rate operator $\underline{R}$ associated with $\mathcal{R}$. Krak et al. [2, Propositions 7.6 to 7.8] prove that the set

$$\mathcal{R}_{\underline{Q}} := \{Q \in \mathfrak{Q} \colon (\forall f \in \mathbb{R}^{\mathcal{X}})\ Qf \geq \underline{Q} f\},$$

of rate matrices that dominate the lower envelope $\underline{Q}$ of $\mathcal{Q}$ has the required properties, and – quite conveniently – that it turns out to have the same lower envelope as $\mathcal{Q}$.

Lemma 60. *The set $\mathcal{R}_{\underline{Q}}$ of rate matrices that dominate the lower envelope $\underline{Q}$ of $\mathcal{Q}$ (i) includes $\mathcal{Q}$; (ii) is bounded and convex and has separately specified rows; and (iii) its lower envelope $\underline{R}_{\mathcal{R}_{\underline{Q}}}$, as defined by Eqn. (34), is $\underline{Q}$.*

Because $\mathcal{R}_{\underline{Q}}$ includes $\mathcal{Q}$, is convex, has separately specified rows and has $\underline{Q}$ as lower envelope, we end up with the following immediate consequence of Corollary 58 by applying it to $\mathbb{P}_{\mathcal{M},\mathcal{R}_{\underline{Q}}}$.

Corollary 61. *Fix time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$, a grid $v = (t_0, \ldots, t_n)$ over $[s, r]$ and real-valued functions $g_1, \ldots, g_n$ on $\mathcal{X}^2$. Let $\underline{h}_0, \ldots, \underline{h}_n$ be the real-valued functions on $\mathcal{X}$ as defined in Corollary 58. Then for all imprecise jump processes $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, x in $\mathcal{X}$, u in $\mathcal{U}_{\prec s}$ and x_u in $\mathcal{X}_u$,*

$$\underline{E}_{\mathcal{P}}\left(\sum_{k=1}^{n} g_k(X_{t_{k-1}}, X_{t_k}) \,\middle|\, X_u = x_u, X_s = x\right) \geq \underline{h}_0(x);$$

in particular, this holds for $\mathcal{P} = \mathbb{P}^{\mathrm{HM}}_{\mathcal{M},\mathcal{Q}}$, $\mathcal{P} = \mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ and $\mathcal{P} = \mathbb{P}_{\mathcal{M},\mathcal{Q}}$.

Our next result shows that if $\mathcal{Q}$ has separately specified rows – but isn't necessarily convex – we do not have to settle for a conservative lower bound; instead, we can replace it by an approximation, at least up to an error that is proportional to $\Delta(v)$. This approximation furthermore has the added benefit that instead of the operator exponential $e^{(t_{k+1}-t_k)\underline{Q}}$ that appears in the definition of $\underline{h}_k$, it uses its 'one-step Euler approximation' $I + (t_{k+1} - t_k)\underline{Q}$. Our proof for this result is rather technical and quite long, which is why we have relegated it to Appendix B.

Theorem 62. *Fix time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$, a grid $v = (t_0, \ldots, t_n)$ over $[s, r]$ and real-valued functions $g_1, \ldots, g_n$ on $\mathcal{X}^2$. Let $\underset{\sim}{h}_0, \ldots, \underset{\sim}{h}_n$ be the real-valued functions on $\mathcal{X}$ defined by the initial condition $\underset{\sim}{h}_n := 0$ and, for all k in $\{0, \ldots, n-1\}$, by the recursive relation*

$$\underset{\sim}{h}_k \colon \mathcal{X} \to \mathbb{R} \colon x \mapsto \left[(I + (t_{k+1} - t_k)\underline{Q})(g_{k+1}(x, \bullet) + \underset{\sim}{h}_{k+1})\right](x).$$

If (i) $\mathcal{Q}$ has separately specified rows, (ii) $\Delta(v)\|\mathcal{Q}\| \leq 2$ and (iii) there is a non-negative real number β such that

$$(\forall k \in \{0, \ldots, n-1\})(\forall x \in \mathcal{X})(\forall P \in \mathbb{P}_{\mathcal{M},\mathcal{R}})\ |E_P(g_{k+1}(X_{t_k}, X_{t_{k+1}}) \mid X_{t_k} = x)| \leq \beta(t_{k+1} - t_k),$$

with $\mathcal{R} := \mathcal{R}_{\underline{Q}}$, then for any imprecise jump process $\mathcal{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$ and all x in $\mathcal{X}$, u in $\mathcal{U}_{\prec s}$ and x_u in $\mathcal{X}_u$,

$$\left|\underline{E}_{\mathcal{P}}\left(\sum_{k=1}^{n} g_k(X_{t_{k-1}}, X_{t_k}) \,\middle|\, X_u = x_u, X_s = x\right) - \underset{\sim}{h}_0(x)\right| \leq \frac{1}{4}\Delta(v)\sum_{k=0}^{n-1}\Delta_k\|\mathcal{Q}\|^2\big(\|g_{k+1}\| + \beta(r-s)\big),$$

where for all k in $\{0, \ldots, n-1\}$, we let $\Delta_k := (t_{k+1} - t_k)$ and $\|g_{k+1}\| := \max\{|g_{k+1}(y, z)| \colon (y, z) \in \mathcal{X}^2\}$.

6.2. Two direct computational methods

Corollary 58 and Theorem 62 solve the issue we encountered at the end of Sections 5.1 and 5.2. To return to that setting, we fix an imprecise jump process $\mathcal{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$, a subset A of $\mathcal{X}^2_{\neq}$, a real-valued function f on $\mathcal{X}$, a sequence of time points u in $\mathcal{U}_{\prec s}$ and a tuple of states x_u in $\mathcal{X}_u$. The issue we raised there was that for a sufficiently fine grid v over $[s, r]$, we need to be able to determine the lower/upper expectation of η^A_v or $\langle f \rangle_v$ conditional on $\{X_u = x_u, X_s = s\}$. Clearly, Corollary 58 allows us to do so whenever $\mathcal{Q}$ is convex and has separately specified rows, at least if we can easily evaluate the operator exponential $e^{\Delta\underline{Q}}$ of the lower envelope $\underline{Q}$ of $\mathcal{Q}$. Alternatively, Theorem 62 also allows us to do so, although only up to some bounded error. This method furthermore has the advantage that it does not require $\mathcal{Q}$ to be convex, but only that it has separately specified rows, and that it does not require us to evaluate $e^{\Delta\underline{Q}}$, but only the lower envelope $\underline{Q}$. However, it does require that the condition regarding the non-negative real number β is satisfied.

Hence, let us deal with the condition in Theorem 62 regarding β. In the case of the number of selected jumps, this condition is satisfied due to Theorem 39.

Lemma 63. *Fix time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$, a grid $v = (t_0, \ldots, t_n)$ over $[s, r]$ and a subset A of $\mathcal{X}^2_{\neq}$. Then for any non-empty and bounded set of rate matrices $\mathcal{R}$,*

$$(\forall k \in \{0, \ldots, n-1\})(\forall x \in \mathcal{X})(\forall P \in \mathbb{P}_{\mathcal{M},\mathcal{R}})\ |E_P(\mathbb{I}_A(X_{t_k}, X_{t_{k+1}}) \mid X_{t_k} = x)| \leq \frac{1}{2}(t_{k+1} - t_k)\|\mathcal{R}\|.$$

Proof. Fix any index k in $\{0,\dots,n-1\}$ and x in $\mathcal{X}$. Then obviously $\min \mathbb{I}_A(X_{t_k}, X_{t_{k+1}}) = 0$ and $\mathbb{I}_A(X_{t_k}, X_{t_{k+1}}) = \eta^A_{(t_k,t_{k+1})}$. Hence, for all P in $\mathbb{P}_{\mathcal{M},\mathcal{R}}$, it follows from (E1) and Theorem 39 that

$$0 \leq E_P(\mathbb{I}_A(X_{t_k}, X_{t_{k+1}}) \,|\, X_{t_k} = x) \leq (t_{k+1} - t_k)\frac{\|\mathcal{R}\|}{2}.$$

Clearly, this implies the inequality in the statement. □

In the case of the Riemann sum $\langle f \rangle_v$ that approximates the integral of $f(X_t)$ over $[s,r]$, the condition regarding β in Theorem 62 is trivially satisfied due to (LI1).

Lemma 64. *Fix time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$, a grid $v = (t_0,\dots,t_n)$ over $[s,r]$ and a real-valued function f on $\mathcal{X}$. Then for any non-empty and bounded set of rate matrices $\mathcal{R}$,*

$$(\forall k \in \{0,\dots,n-1\})(\forall x \in \mathcal{X})(\forall P \in \mathbb{P}_{\mathcal{M},\mathcal{R}})\, |E_P((t_{k+1} - t_k) f(X_{t_{k+1}}) \,|\, X_{t_k} = x)| \leq (t_{k+1} - t_k)\|f\|.$$

Proof. Fix any index k in $\{0,\dots,n-1\}$ and x in $\mathcal{X}$. Then $-(t_{k+1} - t_k)\|f\| \leq \min(t_{k+1} - t_k) f(X_{t_{k+1}})$ and $\max(t_{k+1} - t_k) f(X_{t_{k+1}}) \leq (t_{k+1} - t_k)\|f\|$. Hence, for all P in $\mathbb{P}_{\mathcal{M},\mathcal{R}}$, it follows from (E1) that

$$-(t_{k+1} - t_k)\|f\| \leq E_P(\Delta_n f(X_{t_{k+1}}) \,|\, X_{t_k} = y) \leq (t_{k+1} - t_k)\|f\|,$$

and this verifies the inequality in the statement. □

At this point, it should be clear that together with the preceding results – Lemma 63 and either Theorem 40 or Corollary 45, or Lemma 64 and either Theorem 46 or Corollary 48 – Theorem 62 provides us with a method to compute the lower/upper expectation of $\eta^A_{]s,r]}$ or $\int_s^t f(X_t)\,dt$ up to arbitrary precision, and that this method will only require the (repeated) evaluation of $\underline{Q}$. If we were to go with Theorems 40 and 46, we would get a theoretical upper bound on the error made by the approximation, but we know from past experience – for example [33] and [13, Section 6.5.2] – that when replacing $e^{\Delta \underline{Q}}$ by $(I + \frac{\Delta}{n}\underline{Q})^n$, the theoretical upper bound on the error is usually overly – and often ridiculously – conservative. Hence, we choose to lean on Corollaries 45 and 48 with a sequence of 'uniform' grids $(v_n)_{n\in\mathbb{N}}$ over $[s,r]$, as the resulting theorems make clear that one can rely on empirical convergence. For the number of selected jumps, this yields the following result.

Theorem 65. *Consider an imprecise jump process $\mathcal{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$, and fix time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$ and a subset A of $\mathcal{X}^2_{\neq}$. For all n in $\mathbb{N}$, let $\Delta_n := (r-s)/n$ and let $f_{n,0}$ be recursively defined by the initial condition $f_{n,n} := 0$ and, for all k in $\{0,\dots,n-1\}$, by the recursive relation*

$$f_{n,k} : \mathcal{X} \to \mathbb{R} : x \mapsto \left[(I + \Delta_n \underline{Q})(\mathbb{I}_A(x,\bullet) + f_{n,k+1})\right](x). \tag{35}$$

If $\mathcal{Q}$ has separately specified rows, then for all x in $\mathcal{X}$, u in $\mathcal{U}_{\prec s}$ and x_u in $\mathcal{X}_u$,

$$\underline{E}^{\sigma}_{\mathcal{P}}\left(\eta^A_{]s,r]} \,\middle|\, X_u = x_u, X_s = x\right) = \lim_{n\to+\infty} f_{n,0}(x);$$

a similar equality holds for the upper expectation $\overline{E}^{\sigma}_{\mathcal{P}}$ if we replace $\underline{Q}$ by $\overline{Q}$ in Equation (35).

Proof. We only prove the statement for the lower expectation, the proof for the upper expectation is similar, although here and there one may have to use the conjugacy of $\overline{E}_{\mathcal{P}}$ and $\underline{E}_{\mathcal{P}}$ or the homogeneity of E_P with P a jump process. For all n in $\mathbb{N}$, we let v_n be the grid over $[s,r]$ that consists of $n+1$ evenly spaced time points, so

$$v_n := (t_{n,0},\dots,t_{n,n}) = (s, s+\Delta_n, \dots, s+n\Delta_n).$$

Because $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$ by assumption, it follows from Theorem 40 that for all n in $\mathbb{N}$

$$\left|\underline{E}^{\sigma}_{\mathcal{P}}(\eta^A_{]s,r]} \,|\, X_u = x_u, X_s = x) - \underline{E}_{\mathcal{P}}(\eta^A_{v_n} \,|\, X_u = x_u, X_s = x)\right| \leq \frac{1}{4}\Delta_n(r-s)\|\mathcal{Q}\|^2. \tag{36}$$

Next, we seek to apply Theorem 62. To this end, we fix any natural number n such that $\Delta(v_n)\|\mathcal{Q}\| \leq 2$, or equivalently, because $\Delta(v_n) = \Delta_n = (r-s)/n$ by construction, any n such that $(r-s)\|\mathcal{Q}\|/2 \leq n$. Recall from Eqn. (16) that

$$\eta^A_{v_n} = \sum_{k=1}^{n} \mathbb{I}_A(X_{t_{k-1}}, X_{t_k}).$$

Furthermore, we recall from Lemma 60 that the set $\mathcal{R} := \mathcal{R}_{\underline{Q}}$ of rate operators that dominate $\underline{Q}$ includes $\mathcal{Q}$, is bounded and has $\underline{Q}$ as lower envelope. Hence, it follows from Lemma 63 that

$$(\forall k \in \{0, \dots, n-1\})(\forall y \in \mathcal{X})(\forall P \in \mathbb{P}_{\mathcal{M},\mathcal{R}})\ |E_P(\mathbb{I}_A(X_{t_{n,k}}, X_{t_{n,k+1}}) \mid X_{t_{n,k}} = y)| \leq \frac{1}{2}(t_{n,k+1} - t_{n,k})\|\mathcal{R}\|.$$

Because $\underline{R}_{\mathcal{R}} = \underline{Q}$, it follows from Lemma 56 twice that $\|\mathcal{R}\| = \|\underline{R}_{\mathcal{R}}\| = \|\underline{Q}\| = \|\mathcal{Q}\|$. Hence, we conclude that the condition in Theorem 62 regarding $g_1 = \mathbb{I}_A, \dots, g_n = \mathbb{I}_A$ is satisfied for $\beta = \|\mathcal{Q}\|/2$. For this reason, and because $\mathcal{Q}$ has separately specified rows and $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$ by assumption, it follows from Theorem 62 that

$$\begin{aligned}\left|\underline{E}_{\mathcal{P}}(\eta^A_{v_n} \mid X_u = x_u, X_s = x) - f_{n,0}(x)\right| &\leq \frac{1}{4}\Delta_n \sum_{k=0}^{n-1} \Delta_n \|\mathcal{Q}\|^2 \left(\|\mathbb{I}_A\| + \frac{1}{2}(r-s)\|\mathcal{Q}\|\right) \\ &\leq \frac{1}{8}\Delta_n \|\mathcal{Q}\|^2 (2 + (r-s)\|\mathcal{Q}\|) \sum_{k=0}^{n-1} \Delta_n \\ &= \frac{1}{8}\Delta_n (r-s)\|\mathcal{Q}\|^2 (2 + (r-s)\|\mathcal{Q}\|), \end{aligned} \tag{37}$$

where for the second inequality we used that $\|\mathbb{I}_A\| \leq 1$ and for the first equality we used that $\sum_{k=0}^{n-1} \Delta_n = n\Delta_n = (r-s)$.

For all n in $\mathbb{N}$ such that $(r-s)\|\mathcal{Q}\|/2 \leq n$, it follows from the triangle inequality and Eqns. (36) and (37) that

$$\begin{aligned}\left|\underline{E}^{\sigma}_{\mathcal{P}}(\eta^A_{]s,r]} \mid X_u = x_u, X_s = x) - f_{n,0}(x)\right| &\leq \frac{1}{4}\Delta_n (r-s)\|\mathcal{Q}\|^2 + \frac{1}{8}\Delta_n (r-s)\|\mathcal{Q}\|^2 (2 + (r-s)\|\mathcal{Q}\|) \\ &= \frac{1}{8}\Delta_n (r-s)\|\mathcal{Q}\|^2 (4 + (r-s)\|\mathcal{Q}\|). \end{aligned} \tag{38}$$

Because $\Delta_n = (r-s)/n$ vanishes as n recedes to $+\infty$, this proves the limit in the statement. □

It is relatively straightforward to adapt the argument in our proof for Theorem 65 so that it works for $\int_s^r f(X_t)\,\mathrm{d}t$ instead of $\eta^A_{]s,r]}$. This gives the following result, which is a special case of Theorem 6.50 in [13] as well – but the proof given there differs from the one we give here.

Theorem 66. *Consider an imprecise jump process $\mathcal{P}$ such that $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$. Fix time points s, r in $\mathbb{R}_{\geq 0}$ such that $s < r$ and a real-valued function f on $\mathcal{X}$. For all n in $\mathbb{N}$, let $\Delta_n := (r-s)/n$ and let $f_{n,0}$ be recursively defined by the initial condition $f_{n,n} := 0$ and, for all k in $\{0, \dots, n-1\}$, by the recursive relation*

$$f_{n,k} := (I + \Delta_n \underline{Q})(\Delta_n f + f_{n,k+1}). \tag{39}$$

If $\mathcal{Q}$ has separately specified rows, then for all x in $\mathcal{X}$, u in $\mathcal{U}_{\prec s}$ and x_u in $\mathcal{X}_u$,

$$\underline{E}^{\sigma}_{\mathcal{P}}\left(\int_s^r f(X_t)\,\mathrm{d}t \,\middle|\, X_u = x_u, X_s = x\right) = \lim_{n\to+\infty} f_{n,0}(x);$$

a similar equality holds for the upper expectation $\overline{E}^{\sigma}_{\mathcal{P}}$ if we replace $\underline{Q}$ by $\overline{Q}$ in Eqn. (39).

Proof. The proof is almost entirely the same as that for Theorem 65, so we only highlight the differences between both proofs. Clearly, we need to invoke Theorem 46 instead of Theorem 40. To verify that Theorem 62 is applicable here, we recall from Eqn. (18) that

$$\langle f \rangle_{v_n} = \sum_{k=1}^{n} \Delta_n f(X_{t_{n,k}}),$$

and we invoke Lemma 64 instead of Lemma 63. Hence, we may use Theorem 62 with $\beta = \|f\|$, to find that for all n in $\mathbb{N}$ such that $(r-s)\|\mathcal{Q}\|/2 \leq n$,

$$\left|\underline{E}_{\mathcal{P}}(\langle f \rangle_{v_n} \mid X_u = x_u, X_s = x) - f_{n,0}(x)\right| \leq \frac{1}{4}\Delta_n (r-s)\|\mathcal{Q}\|^2 (\Delta_n + (r-s))\|f\|.$$

Table 2
The lower/upper expected number of downtimes during the first 10-year period.

x	AB	A	B	F	[10, Eqn. (62)]
$\underline{\beta}_x$	1.902	1.902	1.902	1.902	1.900
$\overline{\beta}_x$	2.405	2.405	2.405	2.405	2.407

For all n in $\mathbb{N}$, we know from Theorem 46 that

$$\left|\underline{E}^{\sigma}_{\mathcal{P}}\left(\int_s^r f(X_t)\,\mathrm{d}t \,\middle|\, X_u = x_u, X_s = x\right) - \underline{E}^{\sigma}_{\mathcal{P}}(\langle f\rangle_{v_n} \mid X_u = x_u, X_s = x)\right| \leq \frac{1}{2}\Delta_n(r-s)\|\mathcal{Q}\|(\max f - \min f)$$
$$\leq \Delta_n(r-s)\|\mathcal{Q}\|\|f\|,$$

where for the second inequality we used that $(\max f - \min f) \leq 2\|f\|$. Hence, for all n in $\mathbb{N}$ such that $(r-s)\|\mathcal{Q}\|/2 \leq n$, we find that

$$\left|\underline{E}^{\sigma}_{\mathcal{P}}\left(\int_s^r f(X_t)\,\mathrm{d}t \,\middle|\, X_u = x_u, X_s = x\right) - f_{n,0}(x)\right| \leq \frac{1}{4}\Delta_n(r-s)\|\mathcal{Q}\|\big(4 + (\Delta_n + (r-s))\|\mathcal{Q}\|\big)\|f\|. \tag{40}$$

Because $\Delta_n = (r-s)/n$ vanishes as n recedes to $+\infty$, this proves the limit in the statement. □

Let us put these two methods to the test in the setting of our running example.

Example 67. Recall from Example 22 that Troffaes et al. [10] are interested in the expected number of jumps to a particular set. Specifically, they heuristically determine conservative bounds on the lower/upper expected number of downtimes – that is, the number of jumps to F – during a period of 10 years, and the lower/upper expected downtime – that is, the amount of time in F – during the same period. We do not have to resort to heuristics any more, because we can use our methods to determine these quantities exactly.

We start with the number of downtimes. Let $F := \{(\mathtt{AB}, \mathtt{F}), (\mathtt{A}, \mathtt{F}), (\mathtt{B}, \mathtt{F})\}$. Then in our more formal setting, the expected down period over the first 10-year period is tightly bounded by

$$\underline{\beta}_x := \underline{E}_{\mathcal{M},\mathcal{Q}}\big(\eta^F_{]0,10]} \,\big|\, X_0 = x\big) \quad \text{and} \quad \overline{\beta}_x := \overline{E}_{\mathcal{M},\mathcal{Q}}\big(\eta^F_{]0,10]} \,\big|\, X_0 = x\big),$$

where x in $\mathcal{X}$ is the initial state of the system. Because $\mathcal{Q}$ has separately specified rows, we can use Theorem 65 to compute the lower and upper expected number of downtimes. More precisely, we start with $n_{\min} = 29\,200$ iterations[11] and repeatedly double the number of iterations until we observe empirical convergence, up to four significant digits. Quite remarkably, in this case convergence aleardy occurs after the first doubling, and the first four significant digits remain the same even after 10 doublings. Even more, halving the minimum number of iterations $n_{\min}$ also gives the same results up to at least four significant digits; halving the number of iterations again gives 'NaN' values, which indicates that something is going wrong, as we would expect. Our results can be found in Table 2. Quite remarkably, the exact values we find are quite close to the conservative bounds that Troffaes et al. [10, Eqn. (62)] obtain with their heuristic. While they appear not to depend on the initial state x, this is only true for the first four significant digits, and we find that the lower/upper expected downtimes differ more for shorter time periods.

Next, we deal with the downtime. In our formalism, the downtime over a period of 10 years corresponds to

$$\underline{\alpha}_x := \underline{E}_{\mathcal{M},\mathcal{Q}}\left(\int_0^{10} \mathbb{I}_{\mathtt{F}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right) \quad \text{and} \quad \overline{\alpha}_x := \overline{E}_{\mathcal{M},\mathcal{Q}}\left(\int_0^{10} \mathbb{I}_{\mathtt{F}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right),$$

where x is the initial state of the system. For any such initial state x in $\mathcal{X}$, we use Theorem 66 to compute the lower and upper expected downtime. Here too, we start with $n_{\min} = 29\,200$ iterations and repeatedly increase the number of iterations by a factor 2 until we observe empirical convergence, up to four significant digits; this time around, we need to double the number of iterations about twelve times before we observe convergence. This yields the values reported in Table 3. Remarkably, for the initial states AB, A and B, the lower and upper bounds reported by Troffaes et al. [10, Eqn. 60] agree with the values we find up to three significant digits. However, this is *not* the case if we start in the state $x = \mathtt{F}$ where both power lines are down. The heuristics of Troffaes et al. [10] do not take into account this initial state because they are designed for a system that is in regime; for this reason, they do not pick up the transient effect caused by the initial state.

[11] The reason for this specific start is straightforward: it is the smallest value of n for which the condition in Theorem 62 regarding $\Delta(v)$ is satisfied, or equivalently, due to Lemma 70 in Appendix B, such that $(I + \Delta_n \underline{Q})$ is a so-called lower transition operator.

Table 3
The lower/upper expected downtime during the first 10-year period.

x	AB	A	B	F	[10, Eqn. (60)]
$\underline{\alpha}_x \cdot 10^4$	6.512	6.513	6.513	9.938	6.513
$\overline{\alpha}_x \cdot 10^3$	1.647	1.647	1.647	2.332	1.647

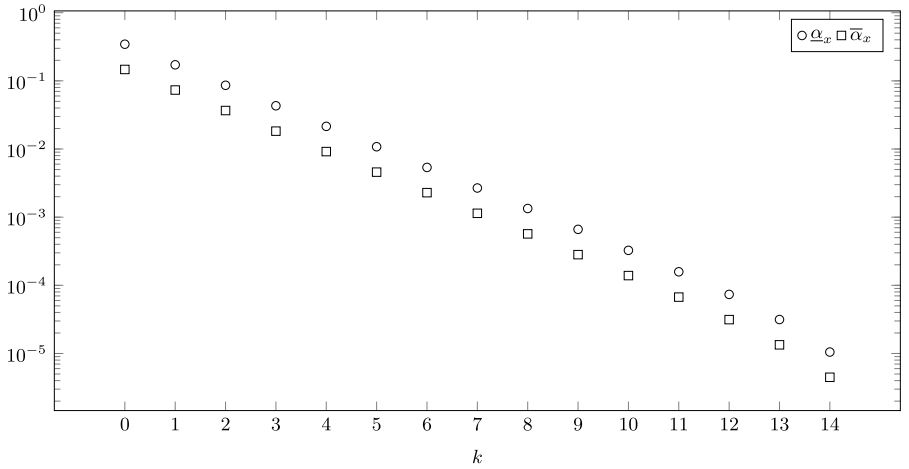

Fig. 1. Maximum relative difference between the approximation for $n = 2^k n_{\min}$ and that for $n = 2^{15} n_{\min}$ for $\underline{\alpha}_x$ and $\overline{\alpha}_x$.

In Fig. 1, we report the maximum relative difference

$$\max\left\{\left|\frac{f_{2^k n_{\min},0}(x) - f_{2^{15} n_{\min},0}(x)}{f_{2^{15} n_{\min},0}(x)}\right| : x \in \mathcal{X}\right\}$$

between the k-th estimate and the last estimate, with k ranging from 0 to 14. The maximum relative difference for $k = 0$ is 0.147 and then decreases as the number of iterations grows; the decrease is linear on the log-log graph, so the error decreases exponentially as the number of iterations increases. For $n = 2^k n_{\min}$, the theoretical upper bound on the error given by Eqn. (40) is

$$\frac{1}{4}\frac{10^2}{2^k n_{\min}}\|\mathcal{Q}\|\left(4 + \left(\frac{10}{2^k n_{\min}} + 10\right)\|\mathcal{Q}\|\right)\|\mathbb{I}_{\mathtt{F}}\| = \frac{292\,030}{2^k},$$

where we used that $\|\mathbb{I}_{\mathtt{F}}\| = 1$ and used the value of $\|\mathcal{Q}\|$ which we determined in Example 14. While this bound does decrease exponentially, it is clearly overly conservative; it overestimates the error by about five orders of magnitude!

For our running example, our limited experiments in Example 67 show that our method based on Theorems 65 and 66 gives similar results as the heuristics proposed in [10]. There are a couple of important differences though. First, their heuristics can only be used to conservatively bound the lower and upper expected 'number of jumps to some state' – that is, it works for $\eta^A_{]s,r]}$ with $A = \{(y, x) : y \in \mathcal{X}, y \neq x\}$ for some x in $\mathcal{X}$ – while our method works for the lower and upper expected 'number of selected jumps' – so with A any arbitrary subset of $\mathcal{X}^2_{\neq}$. In the example of the power network, this means that in contrast to Troffaes et al. [10], we can for example determine the lower and upper expected number of common cause failures – so the number of jumps from AB to F. Second, our method takes into account transient effects due to the initial state, while Troffaes et al. their heuristics are based around a system in regime.

That said, even if the system is in regime – so if the time period is long enough – our method can outperform that of Troffaes et al.; while this is not so in our running example, it is clear in the following example.

Example 68. Let us consider an imprecise jump process with ternary state space $\mathcal{X} := \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, set of rate matrices

$$\mathcal{Q} := \left\{\begin{pmatrix} -\lambda_{\mathtt{a}} & \lambda_{\mathtt{a}} & 0 \\ \mu_{\mathtt{b}} & -\mu_{\mathtt{b}} - \lambda_{\mathtt{b}} & \lambda_{\mathtt{b}} \\ 0 & \mu_{\mathtt{c}} & -\mu_{\mathtt{c}} \end{pmatrix} : \lambda_{\mathtt{a}} = 1, \mu_{\mathtt{b}} = 10, \lambda_{\mathtt{b}} \in [1, 100], \mu_{\mathtt{c}} \in [1, 100]\right\}$$

and arbitrary set $\mathcal{M}$ of initial probability mass functions. This time around, we are after the upper expected fraction of time that the system is in state b. For each initial state x in $\mathcal{X}$ and period of time T in $\mathbb{R}_{>0}$, this upper expected fraction is

$$\overline{E}_{\mathcal{M},\mathcal{Q}}\left(\frac{1}{T}\int_0^T \mathbb{I}_{\mathrm{b}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right) = \overline{E}_{\mathcal{M},\mathcal{Q}}\left(\int_0^T \frac{1}{T}\mathbb{I}_{\mathrm{b}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right),$$

and can thus be obtained by applying Theorem 66. We observe that for increasing values of T, this upper expected fraction converges to about 0.091, or 9.1%, and does not depend on the initial state. The heuristic of Troffaes et al. [10], on the other hand, comes down to determining the limit of $\overline{E}_{\mathcal{M},\mathcal{Q}}(\mathbb{I}_{\mathrm{b}}(X_t) \mid X_0 = x) = -[e^{t\underline{Q}}(-\mathbb{I}_{\mathrm{b}})](x)$ for t going to $+\infty$ – that this limit exists follows from the ergodicity of $\underline{Q}$, see [37]. This way, we find that an upper bound on the fraction is 0.703, or 70.3%. The fact that this bound differs a lot from the upper expected fraction may come as a surprise to the reader who is familiar with the (Point-Wise) Ergodic Theorem for homogeneous Markovian jump processes – see [38, Section 8.6.6] or [25, Theorem 3.8.1]. Indeed, in the case of a homogeneous Markovian jump process P with so-called ergodic rate operator Q, this Ergodic Theorem implies that

$$\lim_{T\to+\infty} E_P\left(\frac{1}{T}\int_0^T \mathbb{I}_{\mathrm{b}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right) = \lim_{t\to+\infty} E_P(\mathbb{I}_{\mathrm{b}}(X_t) \mid X_0 = x) = \lim_{t\to+\infty} [e^{tQ}\mathbb{I}_{\mathrm{b}}](x).$$

In our example, the 'upper expectation' counterpart of the first of these equalities turns out to be a strict inequality! This is in line with recent results regarding the Point-Wise Ergodic Theorem in the setting of imprecise Markov chains [39–41].

This example, together with Example 67, shows that while the heuristic of Troffaes et al. [10] performs surprisingly well in some instances, there are also instances where it yields approximations that are much too conservative compared to the exact results provided by our methods.

7. Conclusion

We have extended the domain of imprecise jump processes, so that these can formally deal with (extended) real variables that depend on the state of the system at more than a finite number of time points. Furthermore, we have investigated the continuity properties of the extended lower and upper expectations, similar to the Monotone Convergence Theorem and Lebesgue's Dominated Convergence Theorem. While the extended lower and upper expectations may not be continuous with respect to monotone and dominated convergence in general, we have identified two particular cases in which they are. For these two particular cases, being the number of selected jumps and integrals over time, we have also established recursive numerical methods to iteratively compute the lower and upper expectations exactly. Our experiments indicate that these methods can – significantly – outperform the heuristics of Troffaes et al. [10], at least with respect to transient behaviour and the tightness of the bounds.

As far as future research is concerned, we see two promising avenues. First, one could study other measurable variables in a similar fashion as we have studied the number of selected jumps and integrals over time here, so with the intent of obtaining a convergence theorem similar to Corollaries 45 and 48 and/or a recursive computational method as in Theorems 65 and 66. For so-called 'until events' and '(truncated) hitting times', this has already been done in [13], where the focus lies on approximating these measurable variables with simple variables $f(X_v)$ of the form

$$f(X_v) = \sum_{k=0}^{n} g_k(X_{t_k}) \prod_{\ell=0}^{k-1} h_\ell(X_{t_\ell}) = g_0(X_{t_0}) + h_0(X_{t_0})g_1(X_{t_1}) + \cdots + h_0(X_{t_0})\cdots h_{n-1}(X_{t_{n-1}})g_n(X_{t_n}),$$

with $g_1, \ldots, g_n$ real-valued functions on $\mathcal{X}$ and $h_1, \ldots, h_{n-1}$ non-negative real-valued functions on $\mathcal{X}$. However, it is possible to deal with much more intricate measurable variables, for example the number of visits to some state x over some period of time $[s, r]$ before you visit an 'unsafe' state – that is, a state outside of a given set $S \subseteq \mathcal{X}$ of 'safe' states. The reason for this is that as an intermediary step, one could generalise the results in Section 6.1 from $\mathcal{A}_u$-simple variables $f(X_v)$ that have a sum decomposition of the form in Eqn. (31) to those that have a 'sum-product' decomposition of the form

$$\begin{aligned} f(X_v) &= \sum_{k=1}^{n} g_k(X_{t_{k-1}}, X_{t_k}) \prod_{\ell=1}^{k-1} h_\ell(X_{t_{\ell-1}}, X_{t_\ell}) \\ &= g_1(X_{t_0}, X_{t_1}) + h_1(X_{t_0}, X_{t_1})g_2(X_{t_1}, X_{t_2}) + \cdots + h_1(X_{t_0}, X_{t_1})\cdots h_{n-1}(X_{t_{n-2}}, X_{t_{n-1}})g_n(X_{t_{n-1}}, X_{t_n}), \end{aligned}$$

where $g_1, \ldots, g_n$ are real-valued functions on $\mathcal{X}^2$ and $h_1, \ldots, h_{n-1}$ are non-negative real-valued functions on $\mathcal{X}^2$. Second, because temporal averages belong to the extended domain of imprecise jump processes, one can now set out to investigate if and how the Point-Wise Ergodic Theorem for homogeneous Markovian jump processes, which we mentioned in Example 68, generalises to imprecise jump processes. A good starting point is the second author's work on ergodicity [37], as well as the work done on ergodicity and the Point-Wise Ergodic Theorem in the setting of imprecise Markov chains [39–41].

CRediT authorship contribution statement

Alexander Erreygers: Conceptualization, Software, Visualization, Writing – original draft. **Jasper De Bock:** Conceptualization, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded through project number 3G028919 of the Research Foundation – Flanders (FWO).

Appendix A. Proof of Proposition 53

Proposition 53 follows from Proposition 4.11 in [13] and the combination of the sum decomposition of $f(X_v)$ and (LD5), but our formal proof is rather long.

Proof of Proposition 53. For all k in $\{1,\dots,n\}$, we let $t_{1:k} \coloneqq (t_1,\dots,t_k)$ and $t_{0:k} \coloneqq (t_0,\dots,t_k)$. Fix any x_s in $\mathcal{X}$, u in $\mathcal{U}_{\prec s}$ and x_u in $\mathcal{X}_u$. Then by Corollary 4.1 in [13],

$$\underline{E}_{\mathcal{P}}(f(X_v) \mid X_u = x_u, X_s = x_s) = \underline{E}_{\mathcal{P}}(f(x_s, X_{t_{1:n}}) \mid X_u = x_u, X_s = x_s).$$

Due to this equality, and because $\mathcal{P}$ satisfies the law of iterated lower expectations, it follows from Proposition 4.11 in [13] that

$$\underline{E}_{\mathcal{P}}(f(X_v) \mid X_u = x_u, X_s = x_s) = \underline{E}_{\mathcal{P}}(f_1'(X_{t_1}) \mid X_u = x_u, X_s = x_s), \tag{A.1}$$

where f_1' is the real-valued function on $\mathcal{X}$ that is defined recursively by the initial condition

$$f_n' \colon \mathcal{X}_{t_{1:n}} \to \mathbb{R} \colon x_{t_{1:n}} \mapsto f(x_v) = \sum_{k=1}^{n} g_k(x_{t_{k-1}}, x_{t_k})$$

and, for all k in $\{1,\dots,n-1\}$, by the recursive relation

$$f_k' \colon \mathcal{X}_{t_{1:k}} \to \mathbb{R} \colon x_{t_{1:k}} \mapsto \underline{E}_{\mathcal{P}}(f_{k+1}'(x_{t_{1:k}}, X_{t_{k+1}}) \mid X_u = x_u, X_{t_{0:k}} = x_{t_{0:k}});$$

in both of these expressions, we silently used the fact that $x_s = x_{t_0}$ has already been fixed.

We now claim that for all k in $\{1,\dots,n\}$ and $x_{t_{1:k}}$ in $\mathcal{X}_{t_{1:k}}$,

$$f_k'(x_{t_{1:k}}) = \sum_{\ell=1}^{k} g_\ell(x_{t_{\ell-1}}, x_{t_\ell}) + f_k(x_{t_k}). \tag{A.2}$$

Our proof for this claim will be one by induction. For the base case that $k = n$, this is obvious from the definition of f_n' and f_n. For the inductive step, we fix some ℓ in $\{1,\dots,n-1\}$, and assume that the equality in Eqn. (A.2) holds for $k = \ell + 1$. Fix any $x_{t_{1:\ell}}$ in $\mathcal{X}_{1:\ell}$. Then by the definition of f_ℓ' and the induction hypothesis,

$$f_\ell'(x_{t_{1:\ell}}) = \underline{E}_{\mathcal{P}}\left(\sum_{i=1}^{\ell} g_i(x_{t_{i-1}}, x_{t_i}) + g_{\ell+1}(x_{t_\ell}, X_{t_{\ell+1}}) + f_{\ell+1}(X_{t_{\ell+1}}) \,\middle|\, X_u = x_u, X_{t_{0:\ell}} = x_{t_{0:\ell}}\right).$$

From this and (LD5), it follows immediately that

$$f_\ell'(x_{t_{1:\ell}}) = \sum_{i=1}^{\ell} g_i(x_{t_{i-1}}, x_{t_i}) + \underline{E}_{\mathcal{P}}\left(g_{\ell+1}(x_{t_\ell}, X_{t_{\ell+1}}) + f_{\ell+1}(X_{t_{\ell+1}}) \,\middle|\, X_u = x_u, X_{t_{0:\ell}} = x_{t_{0:\ell}}\right).$$

Because $\mathcal{P}$ is Markovian, it follows that

$$\begin{aligned} f_\ell'(x_{t_{1:\ell}}) &= \sum_{i=1}^{\ell} g_i(x_{t_{i-1}}, x_{t_i}) + \underline{E}_{\mathcal{P}}\left(g_{\ell+1}(x_{t_\ell}, X_{t_{\ell+1}}) + f_{\ell+1}(X_{t_{\ell+1}}) \,\middle|\, X_{t_\ell} = x_{t_\ell}\right) \\ &= \sum_{i=1}^{\ell} g_i(x_{t_{i-1}}, x_{t_i}) + f_\ell(x_{t_\ell}). \end{aligned}$$

As this equality holds for all $x_{t_{1:\ell}}$ in $\mathcal{X}_{t_{1:\ell}}$, this proves that Eqn. (A.2) holds for $k=\ell$, and this finalises our proof of our claim in Eqn. (A.2).

Finally, we substitute Eqn. (A.2) for $k=1$ in Eqn. (A.1) and use the Markovianity of $\mathcal{P}$, to yield

$$\begin{aligned}\underline{E}_{\mathcal{P}}(f(X_v)\,|\,X_u=x_u, X_s=x_s) &= \underline{E}_{\mathcal{P}}(g_1(x_s,X_{t_1})+f_1(X_{t_1})\,|\,X_u=x_u, X_s=x_s)\\ &= \underline{E}_{\mathcal{P}}(g_1(x_s,X_{t_1})+f_1(X_{t_1})\,|\,X_s=x_s)\\ &= f_0(x_s),\end{aligned}$$

as required. □

Appendix B. Proof of Theorem 62

In our proof for Theorem 62, we make use of several intermediary results. In many of these intermediary results, we use that for any lower rate operator $\underline{R}$, its corresponding operator exponential $e^{\Delta\underline{R}}$ is a so-called 'lower transition operator' – see [2, Corollary 7.10] or [37, Proposition 8 and 10]. This notion of a lower transition operator generalises the notion of a (linear) transition operator – also known as a transition matrix [42, Theorem 9.2.3] or a stochastic matrix [25, p. 2] – by relaxing the linearity to non-linear homogeneity and superadditivity; see, for example, Definition 7.1 in [2].

Definition 69. A lower transition operator $\underline{T}$ is an operator $\underline{T}\colon \mathbb{R}^{\mathcal{X}}\to\mathbb{R}^{\mathcal{X}}$ such that

LT1. $\underline{T}(\lambda f)=\lambda\underline{T}f$ for all λ in $\mathbb{R}_{\geq 0}$ and f in $\mathbb{R}^{\mathcal{X}}$; [non-negative homogeneity]
LT2. $\underline{T}(f+g)\geq\underline{T}f+\underline{T}g$ for all f,g in $\mathbb{R}^{\mathcal{X}}$; [superadditivity]
LT3. $\underline{T}f\geq\min f$ for all f in $\mathbb{R}^{\mathcal{X}}$. [bounded below by the minimum]

For any lower transition operator $\underline{T}$, its conjugate upper transition operator $\overline{T}$ is the (non-negatively homogeneous, subadditive) operator defined by $\overline{T}f:=-\underline{T}(-f)$ for all f in $\mathbb{R}^{\mathcal{X}}$; in the particular case of the operator exponential $e^{\Delta\underline{R}}$ of a lower rate operator $\underline{R}$, it is not difficult to verify that the conjugate is

$$e^{\Delta\overline{R}}:=\lim_{n\to+\infty}\left(I+\frac{\Delta}{n}\overline{R}\right)^n,$$

where $\overline{R}$ is the conjugate of the lower rate operator $\underline{R}$ – so $\overline{R}f:=-\underline{R}(-f)$ for all f in $\mathbb{R}^{\mathcal{X}}$. Furthermore, De Bock [37, (L10), p. 165] proves that for any lower transition operator $\underline{T}$,

LT4. $\|\underline{T}f-\underline{T}g\|\leq\|f-g\|$ for all f,g in $\mathbb{R}^{\mathcal{X}}$.

The following result, which links lower rate operators and lower transition operators and is taken from [13, Lemma 3.72] – see also Proposition 5 in [37] – will also be of use in the remainder.

Lemma 70. *Consider a lower rate operator $\underline{R}$. Then for all Δ in $\mathbb{R}_{\geq 0}$ such that $\Delta\|\underline{R}\|\leq 2$, $(I+\Delta\underline{R})$ is a lower transition operator.*

Now that we have covered lower transition operators, we can start with establishing the necessary intermediary results. The first of these results deals with replacing $e^{\Delta_n\underline{Q}}$ by $(I+\Delta_n\underline{Q})$.

Lemma 71. *Consider a non-empty sequence of time points $v=(t_0,\ldots,t_n)$ in $\mathcal{U}$ with $n\geq 1$, real-valued functions $g_1,\ldots,g_n$ on $\mathcal{X}^2$ and lower rate operators $\underline{R}_0,\ldots,\underline{R}_{n-1}$. Let $\underline{h}_0,\ldots,\underline{h}_n$ and $\underline{\smash{h}}_0,\ldots,\underline{\smash{h}}_n$ be the real-valued functions on $\mathcal{X}$ defined by the initial condition $\underline{h}_n:=0=:\underline{\smash{h}}_n$ and for all k in $\{0,\ldots,n-1\}$ by the recursive relations*

$$\underline{h}_k\colon\mathcal{X}\to\mathbb{R}\colon x\mapsto\left[e^{\Delta_k\underline{R}_k}(g_{k+1}(x,\bullet)+\underline{h}_{k+1})\right](x)$$

and

$$\underline{\smash{h}}_k\colon\mathcal{X}\to\mathbb{R}\colon x\mapsto\left[(I+\Delta_k\underline{R}_k)(g_{k+1}(x,\bullet)+\underline{\smash{h}}_{k+1})\right](x),$$

with $\Delta_k:=(t_{k+1}-t_k)$. If $\Delta_k\|\underline{R}_k\|\leq 2$ for all k in $\{0,\ldots,n-1\}$, then for all k in $\{0,\ldots,n\}$,

$$\|\underline{h}_k-\underline{\smash{h}}_k\|\leq\frac{1}{4}\Delta(v)\sum_{\ell=k}^{n-1}\Delta_\ell\|\underline{R}_\ell\|^2\left(\|g_{\ell+1}\|+\|\underline{h}_{\ell+1}\|\right),$$

with $\|g_{\ell+1}\|:=\max\{|g_{\ell+1}(y,z)|\colon(y,z)\in\mathcal{X}^2\}$.

Proof. Our proof will be by induction. For the base case that $k = n$, we observe that $\|\underline{h}_n - \underline{\tilde{h}}_n\| = 0$ because $\underline{h}_n = 0 = \underline{\tilde{h}}_n$ by definition. Clearly, this agrees with the inequality in the statement.

For the inductive step, we fix some ℓ in $\{0, \dots, n-1\}$, assume that the inequality in the statement holds for $k = \ell + 1$, and set out to prove that it then also holds for $k = \ell$. To this end, we fix any x in $\mathcal{X}$. Then by definition of $\underline{h}_\ell$ and $\underline{\tilde{h}}_\ell$,

$$|\underline{h}_\ell(x) - \underline{\tilde{h}}_\ell(x)| = \left|\left[e^{\Delta_\ell \underline{R}_\ell}(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})\right](x) - \left[(I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{\tilde{h}}_{\ell+1})\right](x)\right|.$$

It follows from this equality and the triangle inequality that

$$\begin{aligned}|\underline{h}_\ell(x) - \underline{\tilde{h}}_\ell(x)| \leq{}& \left|\left[e^{\Delta_\ell \underline{R}_\ell}(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})\right](x) - \left[(I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})\right](x)\right| \\ &+ \left|\left[(I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})\right](x) - \left[(I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{\tilde{h}}_{\ell+1})\right](x)\right|. \quad \text{(B.1)}\end{aligned}$$

Let us deal with the two terms on the right-hand side of this inequality separately.

For the first term, it follows from the definition of the maximum norm $\|\bullet\|$, [37, (N1) on p. 163] – that is, that $\|Mf\| \leq \|M\|\|f\|$ for any non-negatively homogeneous operator M and f in $\mathbb{R}^{\mathcal{X}}$ – and Lemma 59 that

$$\begin{aligned}&\left|\left[e^{\Delta_\ell \underline{R}_\ell}(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})\right](x) - \left[(I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})\right](x)\right| \\ &\quad\leq \left\|e^{\Delta_\ell \underline{R}_\ell}(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1}) - (I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})\right\| \\ &\quad\leq \left\|e^{\Delta_\ell \underline{R}_\ell} - (I + \Delta_\ell \underline{R}_\ell)\right\| \|g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1}\| \\ &\quad\leq \frac{1}{2}\Delta_\ell^2 \|\underline{R}_\ell\|^2 \|g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1}\|.\end{aligned}$$

Due to the triangle inequality

$$\|g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1}\| \leq \|g_{\ell+1}(x,\bullet)\| + \|\underline{h}_{\ell+1}\| \leq \|g_{\ell+1}\| + \|\underline{h}_{\ell+1}\|,$$

and therefore

$$\left|\left[e^{\Delta_\ell \underline{R}_\ell}(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})\right](x) - \left[(I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})\right](x)\right| \leq \frac{1}{2}\Delta_\ell^2 \|\underline{R}_\ell\|^2 \big(\|g_{\ell+1}\| + \|\underline{h}_{\ell+1}\|\big). \quad \text{(B.2)}$$

For the second term, we observe that $(I + \Delta_\ell \underline{R}_\ell)$ is a lower transition operator due to Lemma 70 because $\Delta_\ell \|\underline{R}_\ell\| \leq 2$ by assumption. Hence, it follows from (LT4) that

$$\begin{aligned}&\left|\left[(I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})\right](x) - \left[(I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{\tilde{h}}_{\ell+1})\right](x)\right| \\ &\quad\leq \left\|(I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1}) - (I + \Delta_\ell \underline{R}_\ell)(g_{\ell+1}(x,\bullet) + \underline{\tilde{h}}_{\ell+1})\right\| \\ &\quad\leq \|(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1}) - (g_{\ell+1}(x,\bullet) + \underline{\tilde{h}}_{\ell+1})\| \\ &\quad= \|\underline{h}_{\ell+1} - \underline{\tilde{h}}_{\ell+1}\|. \quad \text{(B.3)}\end{aligned}$$

We substitute each of the two terms on the right-hand side of the inequality in Eqn. (B.1) with the upper bounds in Eqns. (B.2) and (B.3), to yield

$$\begin{aligned}|\underline{h}_\ell(x) - \underline{\tilde{h}}_\ell(x)| &\leq \frac{1}{4}\Delta_\ell^2 \|\underline{R}_k\|^2 \big(\|g_{\ell+1}\| + \|\underline{h}_{\ell+1}\|\big) + \|\underline{h}_{\ell+1} - \underline{\tilde{h}}_{\ell+1}\| \\ &\leq \frac{1}{4}\Delta(v)\Delta_\ell \|\underline{R}_\ell\|^2 \big(\|g_{\ell+1}\| + \|\underline{h}_{\ell+1}\|\big) + \|\underline{h}_{\ell+1} - \underline{\tilde{h}}_{\ell+1}\|,\end{aligned}$$

where for the second inequality we used that $\Delta_\ell \leq \Delta(v)$. Since this inequality holds for arbitrary x in $\mathcal{X}$, it follows that

$$\|\underline{h}_\ell - \underline{\tilde{h}}_\ell\| \leq \frac{1}{4}\Delta(v)\Delta_\ell \|\underline{R}_\ell\|^2 \big(\|g_{\ell+1}\| + \|\underline{h}_{\ell+1}\|\big) + \|\underline{h}_{\ell+1} - \underline{\tilde{h}}_{\ell+1}\|.$$

We substitute $\|\underline{h}_{\ell+1} - \underline{\tilde{h}}_{\ell+1}\|$ on the right-hand side of this inequality by the upper bound of the induction hypothesis – that is, by the right-hand side of the inequality in the statement for $k = \ell + 1$ – to yield

$$\begin{aligned}\|\underline{h}_\ell - \underline{\tilde{h}}_\ell\| &\leq \frac{1}{4}\Delta(v)\Delta_\ell \|\underline{R}_\ell\|^2 \big(\|g_{\ell+1}\| + \|\underline{h}_{\ell+1}\|\big) + \frac{1}{4}\Delta(v) \sum_{i=\ell+1}^{n-1} \Delta_i \|\underline{R}_i\|^2 \big(\|g_{i+1}\| + \|\underline{h}_{i+1}\|\big) \\ &= \frac{1}{4}\Delta(v) \sum_{i=\ell}^{n-1} \Delta_i \|\underline{R}_i\|^2 \big(\|g_{i+1}\| + \|\underline{h}_{i+1}\|\big);\end{aligned}$$

this shows that the inequality in the statement holds for $k = \ell$. □

The second technical result that we need is a construction method for Markovian jump processes starting from some rate matrices, and this is exactly what Proposition 5.6 in [2] provides us. Here, we give a version of this result using our notation and tailored to our needs.

Lemma 72. *Consider a non-empty sequence of time points $(t_0, \dots, t_n)$ in $\mathcal{U}_{\text{ne}}$ with $n \geq 1$ and rate matrices $Q_0, \dots, Q_{n-1}$ in $\mathcal{Q}$. Then there is a Markovian jump process P in $\mathbb{P}^{\text{M}}_{\mathcal{M},\mathcal{Q}}$ such that for all k in $\{0, \dots, n-1\}$, x in $\mathcal{X}$, f in $\mathbb{R}^{\mathcal{X}}$, u in $\mathcal{U}_{\prec t_k}$ and x_u in $\mathcal{X}_u$,*

$$E_P(f(X_{t_{k+1}}) \mid X_u = x_u, X_{t_k} = x) = \left[e^{(t_{k+1}-t_k)Q_k} f\right](x).$$

Proof. Krak et al. [2] actually prove a stronger result, but they use a lot of notation and terminology regarding so-called 'transition matrix systems' to do so. Because we have no need for these 'transition matrix systems' except in this lemma, we choose not to introduce them; instead, we will explain why this lemma is a special case of Proposition 5.6 in [2]. Let p be any mass function in $\mathcal{M}$. Then it follows from Proposition 5.6 – and Definitions 3.5, 3.6 and 4.5 – in [2] that there is a Markovian jump process P in $\mathbb{P}^{\text{M}}_{\mathcal{M},\mathcal{Q}}$ such that

(i) $P(X_0 = x) = p(x)$ for all x in $\mathcal{X}$;
(ii) for all indices k in $\{0, \dots, n-1\}$, time points s, r in $[t_k, t_{k+1}]$ such that $s \leq r$ and states x, y in $\mathcal{X}$,

$$P(X_r = y \mid X_s = x) = \left[e^{(r-s)Q_k}\mathbb{I}_y\right](x). \tag{B.4}$$

What remains for us is to verify the equality in the statement. To this end, we fix any k in $\{0, \dots, n-1\}$, x in $\mathcal{X}$, f in $\mathbb{R}^{\mathcal{X}}$, u in $\mathcal{U}_{\prec t_k}$ and x_u in $\mathcal{X}_u$. Then by Eqn. (4),

$$E_P(f(X_{t_{k+1}}) \mid X_u = x_u, X_{t_k} = x) = \sum_{y \in \mathcal{X}} f(y) P(X_{t_{k+1}} = y \mid X_u = x_u, X_{t_k} = x).$$

Because P is Markovian,

$$E_P(f(X_{t_{k+1}}) \mid X_u = x_u, X_{t_k} = x) = \sum_{y \in \mathcal{X}} f(y) P(X_{t_{k+1}} = y \mid X_{t_k} = x).$$

Finally, we substitute Eqns. (B.4) and (5) to yield the equality in the statement:

$$E_P(f(X_{t_{k+1}}) \mid X_u = x_u, X_{t_k} = x) = \sum_{y \in \mathcal{X}} f(y)\left[e^{(t_{k+1}-t_k)Q_k}\mathbb{I}_y\right](x) = \left[e^{(t_{k+1}-t_k)Q_k} f\right](x). \quad \square$$

The third and final technical result allows us to replace the lower rate operator $\underline{Q}$ by rate matrices $Q_0, \dots, Q_{n-1}$ in $\mathcal{Q}$. In essence, this result is a consequence of the definition of the lower envelope, Lemma 70 and of several properties of (lower) transition operators.

Lemma 73. *Consider a non-empty sequence of time points $(t_0, \dots, t_n)$ in $\mathcal{U}$ with $n \geq 1$ and real-valued functions $g_1, \dots, g_n$ on $\mathcal{X}^2$. Let $\underset{\sim}{h}_0, \dots, \underset{\sim}{h}_n$ be the real-valued functions on $\mathcal{X}$ defined by the initial condition $\underset{\sim}{h}_n := 0$ and for all k in $\{0, \dots, n-1\}$ by the recursive relation*

$$\underset{\sim}{h}_k \colon \mathcal{X} \to \mathbb{R} \colon x \mapsto \left[(I + \Delta_k \underline{Q})(g_{k+1}(x, \bullet) + \underset{\sim}{h}_{k+1})\right](x),$$

with $\Delta_k := (t_{k+1} - t_k)$. If $\mathcal{Q}$ has separately specified rows and $\Delta(v)\|\mathcal{Q}\| \leq 2$, then for any positive real number ϵ in $\mathbb{R}_{>0}$, there are rate matrices $Q_0, \dots, Q_{n-1}$ in $\mathcal{Q}$ such that for all k in $\{0, \dots, n\}$ and x in $\mathcal{X}$,

$$\|\underset{\sim}{h}_k - \tilde{h}_k\| \leq \frac{t_n - t_k}{t_n - t_0}\epsilon, \tag{B.5}$$

where $\tilde{h}_0, \dots, \tilde{h}_n$ are the real-valued functions on $\mathcal{X}$ defined by the initial condition $\tilde{h}_n := 0$ and, for all k in $\{0, \dots, n-1\}$, by the recursive relation

$$\tilde{h}_k \colon \mathcal{X} \to \mathbb{R} \colon x \mapsto \left[(I + \Delta_k Q_k)(g_{k+1}(x, \bullet) + \tilde{h}_{k+1})\right](x).$$

Proof. Let us determine the rate matrices $Q_0, \dots, Q_{n-1}$. To this end, we fix any k in $\{0, \dots, n-1\}$. Then for all x in $\mathcal{X}$, it follows from Eqn. (34), the definition of the lower envelope $\underline{Q}$, that there is a rate matrix $Q_{k,x}$ in $\mathcal{Q}$ such that

$$\left|\left[\underline{Q}(g_{k+1}(x, \bullet) + \underset{\sim}{h}_{k+1})\right](x) - \left[Q_{k,x}(g_{k+1}(x, \bullet) + \underset{\sim}{h}_{k+1})\right](x)\right| < \frac{1}{t_n - t_0}\epsilon.$$

Let Q_k be the matrix defined by $Q_k(x,y) \coloneqq Q_{k,x}(x,y)$ for all x, y in $\mathscr{X}$ – or more formally put, we let $Q_k \colon \mathbb{R}^{\mathscr{X}} \to \mathbb{R}^{\mathscr{X}}$ be the operator defined by $[Q_k f](x) \coloneqq [Q_{k,x} f](x)$ for all f in $\mathbb{R}^{\mathscr{X}}$ and x in $\mathscr{X}$. Then because $\mathcal{Q}$ has separately specified rows, it is clear that Q_k belongs to $\mathcal{Q}$. Furthermore, our definition of Q_k ensures that for all x in $\mathscr{X}$,

$$\left|\left[\underline{Q}(g_{k+1}(x,\bullet) + \underline{h}_{k+1})\right](x) - \left[Q_k(g_{k+1}(x,\bullet) + \underline{h}_{k+1})\right](x)\right| < \frac{1}{t_n - t_0}\epsilon. \tag{B.6}$$

Next, we set out to verify Eqn. (B.5), and we will do so using induction. For the base case that $k = n$, it follows immediately from the definition of $\underline{h}_n$ and $\tilde{h}_n$ that

$$\|\underline{h}_n - \tilde{h}_n\| = 0 = \frac{t_n - t_n}{t_n - t_0}\epsilon.$$

For the inductive step, we fix some ℓ in $\{0, \ldots, n-1\}$ assume that Eqn. (B.5) holds for $k = \ell + 1$, and set out to prove that Eqn. (B.5) holds for $k = \ell$. To this end, we fix any x in $\mathscr{X}$. Then by definition of $\underline{h}_\ell$ and $\tilde{h}_\ell$, the triangle inequality and the definition of the maximum norm $\|\bullet\|$,

$$\begin{aligned}
\left|\underline{h}_\ell(x) - \tilde{h}_\ell(x)\right| &= \left|[(I + \Delta_\ell \underline{Q})(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})](x) - [(I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \tilde{h}_{\ell+1})](x)\right| \\
&\leq \left|[(I + \Delta_\ell \underline{Q})(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})](x) - [(I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})](x)\right| \\
&\quad + \left|[(I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})](x) - [(I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \tilde{h}_{\ell+1})](x)\right| \\
&\leq \left|[(I + \Delta_\ell \underline{Q})(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})](x) - [(I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})](x)\right| \\
&\quad + \left\|(I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1}) - (I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \tilde{h}_{\ell+1})\right\|.
\end{aligned}$$

For the first term, it follows from some straightforward manipulations and Eqn. (B.6) that

$$\begin{aligned}
\left|\underline{h}_\ell(x) - \tilde{h}_\ell(x)\right| &\leq \Delta_\ell \left|[\underline{Q}(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})](x) - [Q_\ell(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1})](x)\right| \\
&\quad + \left\|(I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1}) - (I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \tilde{h}_{\ell+1})\right\| \\
&< \frac{t_{\ell+1} - t_\ell}{t_n - t_0}\epsilon + \left\|(I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1}) - (I + \Delta_\ell Q_\ell)(g_{\ell+1}(x,\bullet) + \tilde{h}_{\ell+1})\right\|.
\end{aligned}$$

For the second term, we observe that $(I + \Delta_\ell Q_\ell)$ is a (lower) transition operator due to Lemma 70 because (i) $\Delta(v)\|\mathcal{Q}\| \leq 2$ by assumption, and (ii) Q_ℓ belongs to $\mathcal{Q}$ by construction and therefore $\|Q_\ell\| \leq \|\mathcal{Q}\|$. Hence, it follows from the preceding inequality and (LT4) that

$$\begin{aligned}
\left|\underline{h}_\ell(x) - \tilde{h}_\ell(x)\right| &< \frac{t_{\ell+1} - t_\ell}{t_n - t_0}\epsilon + \|(g_{\ell+1}(x,\bullet) + \underline{h}_{\ell+1}) - (g_{\ell+1}(x,\bullet) + \tilde{h}_{\ell+1})\| \\
&= \frac{t_{\ell+1} - t_\ell}{t_n - t_0}\epsilon + \|\underline{h}_{\ell+1} - \tilde{h}_{\ell+1}\|.
\end{aligned}$$

Because this inequality holds for all x in $\mathscr{X}$, we infer that

$$\|\underline{h}_\ell - \tilde{h}_\ell\| < \frac{t_{\ell+1} - t_\ell}{t_n - t_0}\epsilon + \left\|\underline{h}_{\ell+1} - \tilde{h}_{\ell+1}\right\|.$$

Finally, it follows from the induction hypothesis – so from Eqn. (B.5) for $k = \ell + 1$ – that

$$\|\underline{h}_\ell - \tilde{h}_\ell\| < \frac{t_{\ell+1} - t_\ell}{t_n - t_0}\epsilon + \frac{t_n - t_{\ell+1}}{t_n - t_0}\epsilon = \frac{t_n - t_\ell}{t_n - t_0}\epsilon,$$

and this shows that Eqn. (B.5) holds for $k = \ell$ too. □

Finally, we combine these intermediary technical results and Corollary 61 in our proof for Theorem 62.

Proof of Theorem 62. Fix any x in $\mathscr{X}$, u in $\mathcal{U}_{\prec s}$ and x_u in $\mathscr{X}_u$. To simplify our notation, we let

$$f \coloneqq \sum_{k=1}^{n} g_k(X_{t_{k-1}}, X_{t_k}).$$

Recall from Lemma 60 that $\mathcal{R}$ includes $\mathcal{Q}$ – and is bounded and convex, has separately specified rows and has $\underline{Q}$ as lower envelope. Because $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}}$ by assumption, it follows that $\mathcal{P} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{Q}} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{R}}$, and therefore

$$\begin{aligned}\underline{E}_{\mathcal{P}}(f \mid X_u = x_u, X_s = x) &= \inf\{E_P(f \mid X_u = x_u, X_s = x) \colon P \in \mathcal{P}\} \\ &\geq \inf\{E_P(f \mid X_u = x_u, X_s = x) \colon P \in \mathbb{P}_{\mathcal{M},\mathcal{R}}\} \\ &= \underline{E}_{\mathcal{M},\mathcal{R}}(f \mid X_u = x_u, X_s = x). \end{aligned} \tag{B.7}$$

Because $\mathcal{R}$ is bounded and convex and has separately specified rows, it follows from Corollary 20 that $\mathbb{P}_{\mathcal{M},\mathcal{R}}$ is Markovian, and from Proposition 51 that $\mathbb{P}_{\mathcal{M},\mathcal{R}}$ satisfies the law of iterated lower expectations. Hence, it follows from Proposition 53 and Lemma 57 that

$$\underline{E}_{\mathcal{M},\mathcal{R}}(f \mid X_u = x_u, X_s = x) = \underline{h}_0(x), \tag{B.8}$$

where $\underline{h}_0$ is the real-valued function on $\mathcal{X}$ defined by the initial condition $\underline{h}_n := 0$ and, for all k in $\{0, \ldots, n-1\}$, by the recursive relation

$$\underline{h}_k \colon \mathcal{X} \to \mathbb{R} \colon y \mapsto \underline{E}_{\mathcal{M},\mathcal{R}}\big(g_{k+1}(y, X_{t_{k+1}}) + \underline{h}_{k+1}(X_{t_{k+1}}) \,\big|\, X_{t_k} = y\big) = \big[e^{\Delta_k \underline{Q}}(g_{k+1}(y, \bullet) + \underline{h}_{k+1})\big](y).$$

Recall from Lemma 56 that $\|\underline{Q}\| = \|\mathcal{Q}\|$. Because $(t_{k+1} - t_k) \leq \Delta(v)$ for all k in $\{0, \ldots, n-1\}$ and $\Delta(v)\|\mathcal{Q}\| \leq 2$ by assumption, it now follows from Lemma 71 – with $\underline{R}_0 = \underline{Q}$, …, $\underline{R}_{n-1} = \underline{Q}$ – that

$$\|\underline{h}_0 - \underline{\tilde{h}}_0\| \leq \frac{1}{4}\Delta(v) \sum_{k=0}^{n-1} \Delta_k \|\mathcal{Q}\|^2 \big(\|g_{k+1}\| + \|\underline{h}_{k+1}\|\big).$$

To obtain and expression similar to the one in the inequality in the statement, we set out to verify that for all k in $\{1, \ldots, n\}$, $\|\underline{h}_k\| \leq \beta(r - t_k)$. Our proof will be one by induction. For the base case that $k = n$, this is trivial because $\underline{h}_n = 0$ by definition, and therefore $\|\underline{h}_n\| = 0 = \beta(r - t_n)$. For the inductive step, we fix some k in $\{1, \ldots, n-1\}$ and assume that $\|\underline{h}_{k+1}\| \leq \beta(r - t_{k+1})$. Fix any y in $\mathcal{X}$. Then by definition of $\underline{h}_k$

$$\underline{h}_k(y) = \underline{E}_{\mathcal{M},\mathcal{R}}\big(g_{k+1}(y, X_{t_{k+1}}) + \underline{h}_{k+1}(X_{t_{k+1}}) \,\big|\, X_{t_k} = y\big).$$

For all P in $\mathbb{P}_{\mathcal{M},\mathcal{R}}$, it follows from (E3) and the triangle inequality that

$$\big|E_P(g_{k+1}(y, X_{t_{k+1}}) + \underline{h}_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = y)\big| \leq \big|E_P(g_{k+1}(y, X_{t_{k+1}}) \mid X_{t_k} = y)\big| + \big|E_P(\underline{h}_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = y)\big|.$$

The second term on the right-hand side of this expression can be bounded with (E1), and the first term is indirectly bounded by the condition on g_{k+1} in the statement because

$$E_P(g_{k+1}(y, X_{t_{k+1}}) \mid X_{t_k} = y) = E_P(g_{k+1}(X_{t_k}, X_{t_{k+1}}) \mid X_{t_k} = y)$$

due to Corollary 3.18 in [13]. Hence, for all P in $\mathbb{P}_{\mathcal{M},\mathcal{R}}$

$$\big|E_P(g_{k+1}(y, X_{t_{k+1}}) + \underline{h}_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = y)\big| \leq \beta(t_{k+1} - t_k) + \|\underline{h}_{k+1}\| \leq \beta(t_{k+1} - t_k) + \beta(r - t_{k+1}) = \beta(r - t_k),$$

where we used the induction hypothesis for the second inequality. Because this inequality holds for all P in $\mathbb{P}_{\mathcal{M},\mathcal{R}}$ and y in $\mathcal{X}$, we conclude that $\|\underline{h}_k\| \leq \beta(r - t_k)$, as required.

We substitute this upper bound on $\|\underline{h}_k\|$ in our upper bound on $\|\underline{h}_0 - \underline{\tilde{h}}_0\|$, to yield

$$\|\underline{h}_0 - \underline{\tilde{h}}_0\| \leq \frac{1}{4}\Delta(v) \sum_{k=0}^{n-1} \Delta_k \|\mathcal{Q}\|^2 (\|g_{k+1}\| + \beta(r - t_{k+1})) \leq \frac{1}{4}\Delta(v) \sum_{k=0}^{n-1} \Delta_k \|\mathcal{Q}\|^2 (\|g_{k+1}\| + \beta(r - s)),$$

where for the second inequality we used that $(r - t_{k+1}) \leq (r - s)$. Finally, it follows from Eqns. (B.7) and (B.8) and the preceding inequality that

$$\underline{E}_{\mathcal{P}}(f \mid X_u = x_u, X_s = x) \geq \underline{h}_0(x) \geq \underline{\tilde{h}}_0(x) - \frac{1}{4}\Delta(v) \sum_{k=0}^{n-1} \Delta_k \|\mathcal{Q}\|^2 (\|g_{k+1}\| + \beta(r - s)). \tag{B.9}$$

Eqn. (B.9) proves 'one side' of the inequality in the statement. In the second part of this proof, we will prove the 'other side'. To this end, we fix any positive real number ϵ. Then by Lemma 73 – which is applicable here because by assumption $\mathcal{Q}$ has separately specified rows and satisfies $\Delta(v)\|\mathcal{Q}\| \leq 2$ – there are rate matrices Q_0, …, Q_{n-1} in $\mathcal{Q}$ such that

$$\|\underline{\tilde{h}}_0 - \tilde{h}_0\| \leq \epsilon, \tag{B.10}$$

where $\tilde{h}_0$, …, $\tilde{h}_n$ are the real-valued functions on $\mathcal{X}$ defined by the initial condition $\tilde{h}_n := 0$ and, for all k in $\{0, \ldots, n-1\}$, by the recursive relation

$$\tilde{h}_k\colon \mathcal{X} \to \mathbb{R}\colon y \mapsto \big[(I + \Delta_k Q_k)(g_{k+1}(y, \bullet) + \tilde{h}_{k+1})\big](y).$$

For all k in $\{0, \dots, n-1\}$, $\|Q_k\| \leq \|\mathcal{Q}\|$ because the rate matrix Q_k belongs to $\mathcal{Q}$ by construction. Because $(t_{k+1} - t_k) \leq \Delta(v)$ for all k in $\{0, \dots, n-1\}$ and $\Delta(v)\|\mathcal{Q}\| \leq 2$ by assumption, it now follows from Lemma 71 – with $\underline{R}_0 = Q_0$, …, $\underline{R}_{n-1} = \mathcal{Q}_{n-1}$ – that

$$\|\underset{\sim}{h}_0 - h_0\| \leq \frac{1}{4}\Delta(v)\sum_{k=0}^{n-1}\Delta_k\|Q_k\|^2\big(\|g_{k+1}\| + \|h_{k+1}\|\big) \leq \frac{1}{4}\Delta(v)\sum_{k=0}^{n-1}\Delta_k\|\mathcal{Q}\|^2\big(\|g_{k+1}\| + \|h_{k+1}\|\big), \tag{B.11}$$

where h_1, …, h_n are the real-valued functions on $\mathcal{X}$ defined by the initial condition $h_n := 0$ and, for all k in $\{1, \dots, n-1\}$, by the recursive relation

$$h_k\colon \mathcal{X} \to \mathbb{R}\colon y \mapsto [e^{\Delta_k Q_k}(g_{k+1}(y, \bullet) + h_{k+1})](y).$$

By Lemma 72, there is a Markovian jump process P in $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P}$ such that for all k in $\{0, \dots, n-1\}$ and y in $\mathcal{X}$,

$$E_P(g_{k+1}(y, X_{t_{k+1}}) + h_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = y) = \big[e^{(t_{k+1}-t_k)Q_k}(g_{k+1}(y, \bullet) + h_{k+1})\big](y) = h_k(y). \tag{B.12}$$

With this equality, it is easy to verify that $\|h_k\| \leq \beta(r - t_k)$ for all k in $\{1, \dots, n\}$. As before, our proof for this claim will be one by induction. For the base case that $k = n$, this is trivial because $h_n = 0$ by definition and therefore $\|h_n\| = 0 = \beta(r - t_n)$. For the inductive step, we fix any k in $\{1, \dots, n-1\}$ and assume that $\|h_{k+1}\| \leq \beta(r - t_{k+1})$. We use Eqn. (B.12), the additivity of E_P – so (E3) – and the triangle inequality, to find that for all y in $\mathcal{X}$,

$$\begin{aligned}|h_k(y)| &= \big|E_P(g_{k+1}(y, X_{t_{k+1}}) + h_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = y)\big| \\ &\leq \big|E_P(g_{k+1}(y, X_{t_{k+1}}) \mid X_{t_k} = y)\big| + \big|E_P(h_{k+1}(X_{t_{k+1}}) \mid X_{t_k} = y)\big|.\end{aligned}$$

As before, it follows from this inequality, the condition on g_{k+1} in the statement – which is relevant due to [33, Corollary 3.18] and because P belongs to $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathbb{P}_{\mathcal{M},\mathcal{R}}$ by construction – (E1) and the induction hypothesis that $\|h_k\| \leq \beta(t_{k+1} - t_k) + \beta(r - t_{k+1}) = \beta(r - t_k)$, as required. Hence, it follows from Eqns. (B.10) and (B.11) that

$$\underset{\sim}{h}_0(x) \geq \tilde{h}_0(x) - \epsilon \geq h_0(x) - \frac{1}{4}\Delta(v)\sum_{k=0}^{n-1}\Delta_k\|\mathcal{Q}\|^2\big(\|g_{k+1}\| + \beta(r - s)\big) - \epsilon, \tag{B.13}$$

where for the last inequality we used that $\|h_{k+1}\| \leq \beta(r - t_{k+1}) \leq \beta(r - s)$ for all k in $\{0, \dots, n-1\}$.

It also follows from Eqn. (B.12) and Proposition 53 for the degenerate imprecise jump process $\{P\}$ – which trivially satisfies the law of iterated lower expectation due to Proposition 49 and is Markovian due to Proposition 18 because P is Markovian by construction – that

$$E_P(f \mid X_u = x_u, X_s = x) = h_0(x).$$

Because P belongs to $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}}$ by construction and $\mathbb{P}^{\mathrm{M}}_{\mathcal{M},\mathcal{Q}} \subseteq \mathcal{P}$ by assumption, we infer from this equality that

$$\underline{E}_{\mathcal{P}}(f \mid X_u = x_u, X_s = x) = \inf\{E_P(f \mid X_u = x_u, X_s = x)\colon P \in \mathcal{P}\} \leq h_0(x). \tag{B.14}$$

Finally, we combine Eqns. (B.13) and (B.14), to yield

$$\underset{\sim}{h}_0(x) \geq \underline{E}_{\mathcal{P}}(f \mid X_u = x_u, X_s = x) - \frac{1}{4}\Delta(v)\sum_{k=0}^{n-1}\Delta_k\|\mathcal{Q}\|^2\big(\|g_{k+1}\| + \beta(r - s)\big) - \epsilon.$$

Because this inequality holds for any positive real number ϵ, we conclude that

$$\underset{\sim}{h}_0(x) \geq \underline{E}_{\mathcal{P}}(f \mid X_u = x_u, X_s = x) - \frac{1}{4}\Delta(v)\sum_{k=0}^{n-1}\Delta_k\|\mathcal{Q}\|^2\big(\|g_{k+1}\| + \beta(r - s)\big).$$

Together with Eqn. (B.9), this implies the inequality in the statement. □

References

[1] D. Škulj, Efficient computation of the bounds of continuous time imprecise Markov chains, Appl. Math. Comput. 250 (2015) 165–180, https://doi.org/10.1016/j.amc.2014.10.092.
[2] T. Krak, J. De Bock, A. Siebes, Imprecise continuous-time Markov chains, Int. J. Approx. Reason. 88 (2017) 452–528, https://doi.org/10.1016/j.ijar.2017.06.012.
[3] M. Nendel, Markov chains under nonlinear expectation, Math. Finance 31 (2021) 474–507, https://doi.org/10.1111/mafi.12289.

[4] M.C.M. Troffaes, G. de Cooman, Lower Previsions, Wiley, 2014.
[5] P. Berti, E. Regazzini, P. Rigo, Coherent statistical inference and Bayes theorem, Ann. Stat. 19 (1991) 366–381, https://doi.org/10.1214/aos/1176347988.
[6] E. Regazzini, Finitely additive conditional probabilities, Rend. Semin. Mat. Fis. Milano 55 (1985) 69–89, https://doi.org/10.1007/BF02924866.
[7] L.E. Dubins, Finitely additive conditional probabilities, conglomerability and disintegrations, Ann. Probab. 3 (1975) 89–99, https://doi.org/10.1214/aop/1176996451.
[8] S. Peng, Nonlinear expectations and nonlinear Markov chains, Chin. Ann. Math. 26 (2005) 159–184, https://doi.org/10.1142/S0252959905000154.
[9] R. Denk, M. Kupper, M. Nendel, Kolmogorov-type and general extension results for nonlinear expectations, Banach J. Math. Anal. 12 (2018) 515–540, https://doi.org/10.1215/17358787-2017-0024.
[10] M.C.M. Troffaes, J. Gledhill, D. Škulj, S. Blake, Using imprecise continuous time Markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair, in: Proceedings of the 9th International Symposium on Imprecise Probability: Theories and Applications, ISIPTA '15, 2015, pp. 287–294.
[11] A. Erreygers, J. De Bock, Bounding inferences for large-scale continuous-time Markov chains: a new approach based on lumping and imprecise Markov chains, Int. J. Approx. Reason. 115 (2019) 96–133, https://doi.org/10.1016/j.ijar.2019.09.003.
[12] A. Erreygers, J. De Bock, Extending the domain of imprecise jump processes from simple variables to measurable ones, in: Proceedings of the 12th International Symposium on Imprecise Probability: Theories and Applications, ISIPTA 2021, in: Proceedings of Machine Learning Research, vol. 147, 2021, pp. 140–149.
[13] A. Erreygers, Markovian imprecise jump processes: foundations, algorithms and applications, Ph.D. thesis, Ghent University, 2021, Available at https://users.ugent.be/~aerreyge/.
[14] I.I. Gikhman, A.V. Skorokhod, Introduction to the Theory of Random Processes, W.B. Saunders Company, 1969.
[15] J.-F. Le Gall, Brownian Motion, Martingales, and Stochastic Calculus, Springer, 2016.
[16] M.C.M. Troffaes, S. Blake, A robust data driven approach to quantifying common-cause failure in power networks, in: Proceedings of the 8th International Symposium on Imprecise Probability: Theories and Applications, ISIPTA '13, 2013, pp. 311–317.
[17] P. Billingsley, Probability and Measure, 3rd ed., Wiley, 1995.
[18] B.E. Fristedt, L.F. Gray, A Modern Approach to Probability Theory, Birkhäuser, Basel, 1997.
[19] A.A. Borovkov, Probability Theory, Springer, 2013.
[20] P. Billingsley, Convergence of Probability Measures, 2nd ed., Wiley, 1999.
[21] K.P.S. Bhaskara Rao, M. Bhaskara Rao, Theory of Charges, Academic Press, 1983.
[22] P.M. Williams, Notes on conditional previsions, Technical Report, School of Mathematical and Physical Sciences, University of Sussex, 1975, reprinted as [23].
[23] P.M. Williams, Notes on conditional previsions, Int. J. Approx. Reason. 44 (2007) 366–383, https://doi.org/10.1016/j.ijar.2006.07.019, reprint of [22].
[24] B. de Finetti, Theory of Probability, Wiley, 2017, reprint of [43].
[25] J.R. Norris, Markov Chains, Cambridge University Press, 1997.
[26] E. Schechter, Handbook of Analysis and Its Foundations, Academic Press, 1997.
[27] R. Schilling, Measures, Integrals and Martingales, 2nd ed., Cambridge University Press, 2017.
[28] D. Denneberg, Non-Additive Measure and Integral, Springer, 1994.
[29] C. Baier, J.-P. Katoen, Principles of Model Checking, MIT Press, 2008.
[30] C. Baier, B. Haverkort, H. Hermanns, J.-P. Katoen, Model-checking algorithms for continuous-time Markov chains, IEEE Trans. Softw. Eng. 29 (2003) 524–541, https://doi.org/10.1109/TSE.2003.1205180.
[31] V.G. Kulkarni, Introduction to Modeling and Analysis of Stochastic Systems, Springer, 2011.
[32] E. Miranda, M. Zaffalon, Full conglomerability, J. Stat. Theory Pract. 11 (2017) 634–669, https://doi.org/10.1080/15598608.2017.1295890.
[33] A. Erreygers, J. De Bock, Imprecise continuous-time Markov chains: efficient computational methods with guaranteed error bounds, in: Proceedings of the 10th International Symposium on Imprecise Probability: Theories and Applications, ISIPTA '17, in: Proceedings of Machine Learning Research, vol. 62, 2017, pp. 145–156.
[34] T. Krak, Continuous-time imprecise-Markov chains: theory and algorithms, Ph.D. thesis, Ghent University, 2021, Available at https://users.ugent.be/~jdbock/publications.html.
[35] G. de Cooman, F. Hermans, Imprecise probability trees: bridging two theories of imprecise probability, Artif. Intell. 172 (2008) 1400–1427, https://doi.org/10.1016/j.artint.2008.03.001.
[36] G. de Cooman, F. Hermans, E. Quaeghebeur, Imprecise Markov chains and their limit behavior, Probab. Eng. Inf. Sci. 23 (2009) 597–635, https://doi.org/10.1017/S0269964809990039.
[37] J. De Bock, The limit behaviour of imprecise continuous-time Markov chains, J. Nonlinear Sci. 27 (2017) 159–196, https://doi.org/10.1007/s00332-016-9328-3.
[38] M. Iosifescu, Finite Markov Processes and Their Applications, Wiley, 1980.
[39] G. de Cooman, J. De Bock, S. Lopatatzidis, Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems, Int. J. Approx. Reason. 76 (2016) 18–46, https://doi.org/10.1016/j.ijar.2016.04.009.
[40] N. T'Joens, J. De Bock, Average behaviour in discrete-time imprecise Markov chains: a study of weak ergodicity, Int. J. Approx. Reason. 132 (2021) 181–205, arXiv:2102.04793.
[41] J. De Bock, N. T'Joens, Average behaviour of imprecise Markov chains: a single pointwise ergodic theorem for six different models, in: Proceedings of the 12th International Symposium on Imprecise Probability: Theories and Applications, ISIPTA 2021, in: Proceedings of Machine Learning Research, vol. 147, 2021, pp. 90–99.
[42] P. Whittle, Probability via Expectation, 4th ed., Springer, 2000.
[43] B. de Finetti, Theory of Probability, Wiley, 1974–1975, English translation of [44].
[44] B. de Finetti, Teoria Delle Probabilità, Giuliu Einaudi, 1970.