Extending the Domain of Imprecise Jump Processes from Simple Variables to Measurable Ones

Alexander Erreygers ALEXANDER.ERREYGERS@UGENT.BE
Jasper De Bock JASPER.DEBOCK@UGENT.BE
Foundations Lab for imprecise probabilities, ELIS, Ghent University, Belgium

Abstract

We extend the domain of imprecise jump processes, also known as imprecise continuous-time Markov chains, from inferences that depend on a finite number of time points to inferences that can depend on the state of the system at *all* time points. We also investigate the continuity properties of the resulting lower and upper expectations with respect to point-wise convergent sequences that are monotone or dominated. For two particular inferences, integrals over time and the number of jumps to a subset of states, we strengthen these continuity properties and present an iterative scheme to approximate their lower and upper expectations.

Keywords: continuous time, jump process, Markov chain, imprecise, integral over time, occupancy time, number of jumps, monotone convergence, dominated convergence

1. Introduction

Recently, several authors have independently proposed generalisations of Markovian jump processes—often also called continuous-time Markov chains—that provide an elegant way to deal with parameter uncertainty. Whereas a 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—to some extent even equivalent—results. Škulj [19] and Krak et al. [11] adhere to the framework of imprecise probabilities [22], while Nendel [15] follows the framework of non-linear (or convex) expectations—see also [16]. However, both of these frameworks have crucial shortcomings: that of Škulj [19] Krak et al. [11] only deals with lower and upper expectations of variables that depend on the state of the system at a single time point or a finite number of time points, respectively, while that of Nendel [15] 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) expected temporal averages, expected occupancy times, expected hitting times—also called expected first-passage times—and the expected number of jumps (to a set) are not included in the domain. Instead of resorting to heuristics to circumvent this issue [9, 23], our aim here is to get rid of this problem in a theoretically sound manner, by suitably extending the domain of imprecise jump processes.

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. In Section 5, we take a closer look at integrals over time—including occupancy times—and the number of jumps to a set of states, two classes of variables that belong to the extended domain, and establish methods to approximate their lower and upper expectations. We put these methods to the test in Section 6, where we compare our methods to those of Troffaes et al. [23]. Section 7 concludes this contribution. To adhere to the page limit, we state our results without proof; the proofs for most of the results—except for those in Section 5.2—can be found in [7].

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 assumes values in a finite set; following Gikhman and Skorokhod [10] and Le Gall [13]—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 $\mathscr{X}$; throughout this contribution, except in Section 6, $\mathscr{X}$ can be any non-empty and finite set.

2.1. Càdlàg Paths

Because the system evolves in continuous time, an outcome in the sample space is a *path* $\omega\colon \mathbb{R}_{\geq 0} \to \mathscr{X}$, 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,

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,

for example, the path that assumes the state x whenever the time point t is a rational number and the state $y \neq x$ otherwise, with x, y in $\mathscr{X}$. In this contribution, 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 $\mathbb{R}_{>0}$.[2] Whenever the path ω satisfies these two requirements, it is called a *càdlàg path*; see, for example, [3, Section 12] or [13, p. 54]. We collect all càdlàg paths in the set Ω,[3] and this will be our sample space.

Krak et al. [11] do not (need to) assume càdlàg paths, but that is simply because this assumption is not relevant to their setting. However, it is absolutely essential when extending the domain that all paths in the sample space are càdlàg, as will become clear in Sections 4 and 5 further on.

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 $\mathscr{X}$, we denote the event that 'the state of the system at time t is x' by

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

In a similar fashion, we let

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

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

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

To simplify the notation regarding events that depend on more than a single time point, we avail ourselves of the notational conventions used by Krak et al. [11, 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 or v. We collect all sequences of time points in $\mathscr{U}_{\text{ne}}$, and let $\mathscr{U} := \mathscr{U}_{\text{ne}} \cup \{()\}$, 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 $\min u := t_1$ and $\max u := t_n$, respectively. A statement of the form $\max u < t$ is taken to be trivially true whenever u is the empty sequence $()$. For u and v in $\mathscr{U}$, 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 $\mathscr{U}$. For any sequence of time points $u = (t_1, \ldots, t_n)$ in $\mathscr{U}_{\text{ne}}$, we let $\mathscr{X}_u$ be the set of all n-tuples $x_u = (x_{t_1}, \ldots, x_{t_n})$ of states in $\mathscr{X}$. If u is the empty sequence $()$, then we let $\mathscr{X}_u = \mathscr{X}_{()}$ denote the singleton containing the empty tuple, denoted by $x_{()}$.

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

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

Furthermore, for any $B \subseteq \mathscr{X}_v$, we let

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

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

For any u in $\mathscr{U}$, Krak et al. [11, Section 4.2] let $\mathscr{A}_u$ be the set of events that consists of the cylinder events for all sequences v with time points in or succeeding u:

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

Crucially, $\mathscr{A}_u$ is an algebra of events.

2.3. Coherent Conditional Probabilities

In order to deal with conditioning in an unambiguous manner, Krak et al. [11] resort to the framework of coherent conditional probabilities. What follows is a brief introduction to coherent conditional probabilities; we refer to [1, 6, 17] for a more detailed exposition. To keep our exposition simple, we will define coherent conditional probabilities through the related notion of full conditional probabilities. In the following definition, and throughout the remainder of this contribution, we denote the set of all events by 2^Ω and let $2^\Omega_{\text{ne}} := 2^\Omega \setminus \{\varnothing\}$ be the set of all events that are non-empty.

Definition 1 *A* full conditional probability P *is a real-valued map on* $2^\Omega \times 2^\Omega_{\text{ne}}$ *such that, for all* A, B *in* 2^Ω *and* C, D *in* 2^Ω_{ne},

P1. $P(A|C) \geq 0$;

P2. $P(A|C) = 1$ *if* $C \subseteq A$;

P3. $P(A \cup B|C) = P(A|C) + P(B|C)$ *if* $A \cap B = \varnothing$;

P4. $P(A \cap D|C) = P(A|D \cap C)P(D|C)$ *if* $D \cap C \neq \varnothing$.

Note that (P1)–(P3) simply state that $P(\bullet|C)$ is a probability charge—a 'finitely additive probability measure' [see

$\mathbb{N}$ denotes the natural numbers (or positive integers) and $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ the extended real numbers.

2. We bestow the state space $\mathscr{X}$ with the discrete topology.

3. This is known as the *Skorokhod space*.

22, Definition 1.15]—and that (P4) is a multiplicative version of Bayes's rule. We use the adjective *full* because the domain of P is $2^\Omega \times 2^\Omega_{\text{ne}}$. Next, we move to conditional probabilities whose domain is a subset of $2^\Omega \times 2^\Omega_{\text{ne}}$.

Definition 2 *A* coherent conditional probability *P is a real-valued map on $\mathscr{C} \subseteq 2^\Omega \times 2^\Omega_{\text{ne}}$ that can be extended to a full conditional probability on $2^\Omega \times 2^\Omega_{\text{ne}}$.*

Important to emphasise here is that simply demanding that (P1)–(P4) hold on the domain $\mathscr{C}$ is in general *not* sufficient to guarantee that the conditional probability P can be extended to a full conditional probability. A necessary and sufficient condition for the existence of such an extension can be found in [17, Theorem 3], but we refrain from stating it here because of its technical nature. It suffices to understand that this so-called *coherence* condition has an intuitive betting interpretation, hence explaining the use of the adjective coherent, and that checking this condition is usually more convenient than explicitly constructing a full conditional extension. Another strong argument for using coherent conditional probabilities is that they can always be coherently extended to a larger domain [17, Theorem 4].

2.4. Jump Processes as a Special Case

Krak et al. [11, Definition 4.3] define a jump process as a coherent conditional probability with domain

$$\mathscr{D} := \big\{(A \,|\, X_u = x_u) \colon u \in \mathscr{U}, A \in \mathscr{A}_u, x_u \in \mathscr{X}_u\big\}, \qquad (1)$$

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

Definition 3 *A* jump process *P is a coherent conditional probability on $\mathscr{D}$. We let $\mathbb{P}$ be the set of all jump processes.*

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 $\mathscr{U}$ and x_u in $\mathscr{X}_u$. Because $P(\bullet \,|\, X_u = x_u)$ is a probability charge on $\mathscr{A}_u$ by (P1)–(P3), it corresponds to an expectation operator $E_P(\bullet \,|\, X_u = x_u)$ that is defined on the set of $\mathscr{A}_u$-simple variables in the usual way: through the Dunford integral [22, Definition 8.13], or equivalently, through natural extension [22, Theorem 8.15].

A variable is a map on Ω that takes values in the (extended) reals. To characterise the $\mathscr{A}_u$-simple variables, we let $\mathbb{I}_A \colon \Omega \to \mathbb{R}$ denote the indicator of an event A in 2^Ω, which is a variable that assumes the value 1 on A and 0 elsewhere. A variable $f \colon \Omega \to \mathbb{R}$ is then *$\mathscr{A}_u$-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 $\mathscr{A}_u$; whenever this is the case, the expectation of f conditional on $\{X_u = x_u\}$ is

$$E_P(f \,|\, X_u = x_u) := \sum_{k=1}^n a_k P(A_k \,|\, X_u = x_u).$$

This way, the expectation E_P has domain

$$\mathbb{JS} := \{(f \,|\, X_u = x_u) \colon u \in \mathscr{U}, x_u \in \mathscr{X}_u, f \in \mathbb{S}_u\},$$

where $\mathbb{S}_u$ denotes the set of real variables that are $\mathscr{A}_u$-simple and where we favour writing $(f \,|\, X_u = x_u)$ instead of $(f, \{X_u = x_u\})$, as in Equation (1).

Because every event in $\mathscr{A}_u$ depends on the state of the system at a finite number of time points in or succeeding u, it is easy to see that the same holds for every $\mathscr{A}_u$-simple variable. Consequently, these variables have a convenient representation, given that we introduce some additional notation. For any real-valued function f on $\mathscr{X}_v$, we let $f(X_v) := f \circ X_v$ denote the function composition of f after X_v; thus, $f(X_v)$ is a real variable.

Lemma 4 *Consider some u in $\mathscr{U}$ and some real variable $f \colon \Omega \to \mathbb{R}$. Then f is $\mathscr{A}_u$-simple if and only if there is some v in $\mathscr{U}_{\text{ne}}$ with $u \preccurlyeq v$ and a real-valued function g on $\mathscr{X}_v$ such that $f = g(X_v)$.*

2.5. Markovianity and Homogeneity

In general, specifying a jump process is a non-trivial task. For this reason, it is customary to assume the following two simplifying properties.

Definition 5 *A jump process P is* Markovian*—or, alternatively, has the Markov property—if for all t in $\mathbb{R}_{\geq 0}$, all Δ in $\mathbb{R}_{>0}$, all x, y in $\mathscr{X}$, all u in $\mathscr{U}$ such that $\max u < t$ and all x_u in $\mathscr{X}_u$,*

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

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

Definition 6 *A Markovian jump process P is* homogeneous *if for all t in $\mathbb{R}_{\geq 0}$, Δ in $\mathbb{R}_{>0}$ and x, y in $\mathscr{X}$,*

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

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

Under a mild continuity condition, a homogeneous Markovian jump process P is uniquely characterised by two parameters: its initial probability mass function π_P and its rate matrix Q_P [11, Theorem 5.4]. The initial probability mass function $\pi_P \colon \mathscr{X} \to [0,1]$ is defined by

$$\pi_P(x) := P(X_0 = x) \quad \text{for all } x \in \mathscr{X}; \qquad (2)$$

the rate matrix Q_P is the $|\mathscr{X}| \times |\mathscr{X}|$ real-valued matrix whose (x,y)-component is defined by

$$Q_P(x,y) := \lim_{\Delta \searrow 0} \frac{P(X_\Delta = y \,|\, X_0 = x) - I(x,y)}{\Delta}, \qquad (3)$$

where I denotes the identity matrix. Conversely, any combination of a probability mass function π and a rate matrix Q—that is, a matrix with non-negative off-diagonal components and rows that sum to zero—characterises a unique homogeneous Markovian jump process $P_{\pi,Q}$ [11, Corollary 5.3].

3. Imprecise Jump Processes

One is not always able to or willing to specify precise values for the characterising parameters of a homogeneous Markovian jump process. In this contribution, we suppose that instead of a single probability mass function π and rate matrix Q, we have a non-empty set $\mathscr{M}$ of probability mass functions and a non-empty and bounded set $\mathscr{Q}$ of rate matrices. It then makes sense to consider the set

$$\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}} := \{P_{\pi,Q}\colon \pi \in \mathscr{M}, Q \in \mathscr{Q}\}$$

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

Besides $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, Krak et al. [11] propose two additional imprecise jump processes $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ that are fully defined by $\mathscr{M}$ and $\mathscr{Q}$, and they do so by including the non-homogeneous and then the non-Markovian jump processes that are 'consistent' with $\mathscr{M}$ and $\mathscr{Q}$. A jump process P in $\mathbb{P}$ is *consistent with* $\mathscr{M}$ if there is a probability mass function π in $\mathscr{M}$ such that

$$P(X_0 = x) = \pi(x) \quad \text{for all } x \in \mathscr{X};$$

note that this condition is inspired by Equation (2). Consistency with $\mathscr{Q}$ is motivated by Equation (3), but the formal definition is rather involved [11, Definition 6.1]. For our present purposes, it suffices to understand that, essentially, a process P in $\mathbb{P}$ is consistent with $\mathscr{Q}$ if, for all t in $\mathbb{R}_{\geq 0}$, u in $\mathscr{U}$ such that $\max u < t$ and x_u in $\mathscr{X}_u$, the rate matrix $Q^{t+\Delta}_{t,x_u}$, defined for all Δ in $\mathbb{R}_{>0}$ and x, y in $\mathscr{X}$ by

$$Q^{t+\Delta}_{t,x_u} := \frac{P(X_{t+\Delta} = y \,|\, X_t = x, X_u = x_u) - I(x,y)}{\Delta},$$

comes arbitrarily close to—or is eventually contained in—$\mathscr{Q}$ as Δ approaches 0. The imprecise jump process $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ then consists of all—not necessarily homogeneous—Markovian jump processes that are consistent with $\mathscr{M}$ and $\mathscr{Q}$; similarly, $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ consists of all—not necessarily Markovian—jump processes that are consistent with $\mathscr{M}$ and $\mathscr{Q}$. Note that, by construction,

$$\mathbb{P}_{\mathscr{M},\mathscr{Q}} \supseteq \mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}} \supseteq \mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}. \tag{4}$$

3.1. Lower and Upper Expectations

Every jump process P in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ corresponds to an expectation operator E_P, and we are interested in (tight) lower and upper bounds on these expectations. For this reason, we consider the lower expectation $\underline{E}_{\mathscr{M},\mathscr{Q}}\colon \mathbb{JS} \to \mathbb{R}$, defined for all $(f \,|\, X_u = x_u)$ in $\mathbb{JS}$ by

$$\underline{E}_{\mathscr{M},\mathscr{Q}}(f \,|\, X_u = x_u) := \inf\{E_P(f \,|\, X_u = x_u)\colon P \in \mathbb{P}_{\mathscr{M},\mathscr{Q}}\}$$

and the upper expectation $\overline{E}_{\mathscr{M},\mathscr{Q}}\colon \mathbb{JS} \to \mathbb{R}$, defined similarly but with a supremum. Note that $\underline{E}_{\mathscr{M},\mathscr{Q}}$ and $\overline{E}_{\mathscr{M},\mathscr{Q}}$ are conjugate, in the sense that, for all $(f \,|\, X_u = x_u)$ in $\mathbb{JS}$,

$$\overline{E}_{\mathscr{M},\mathscr{Q}}(f \,|\, X_u = x_u) = -\underline{E}_{\mathscr{M},\mathscr{Q}}(-f \,|\, X_u = x_u).$$

It therefore suffices to study one of the two; we will focus on the lower envelope $\underline{E}_{\mathscr{M},\mathscr{Q}}$. In a similar fashion, Krak et al. [11, Definition 6.5] define the lower (and conjugate upper) envelopes of the expectations with respect to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, denoted by $\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, respectively; due to Equation (4),

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

and conversely for the upper expectations.

In general, determining these tight lower and upper bounds on the expectations corresponding to $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ or $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ is intractable if not impossible, as one would have to explicitly construct these sets in order to optimise over them. However, there are particular cases in which tight lower and upper bounds are relatively easy to obtain. Crucial is that the set $\mathscr{Q}$ has separately specified rows, essentially meaning that we can select the rows of the rate matrices in $\mathscr{Q}$ independently [11, Definition 7.3].

Definition 7 *A set $\mathscr{Q}$ of rate matrices has* separately specified rows *if*

$$\mathscr{Q} = \{Q \in \mathfrak{Q}\colon (\forall x \in \mathscr{X})\ Q(x,\bullet) \in \mathscr{Q}_x\},$$

where $\mathfrak{Q}$ denotes the set of all $|\mathscr{X}| \times |\mathscr{X}|$ rate matrices and where, for all x in $\mathscr{X}$, we let $\mathscr{Q}_x := \{Q(x,\bullet)\colon Q \in \mathscr{Q}\}$.

Krak et al. [11, Corollary 8.3] show that whenever $\mathscr{Q}$ is convex and has separately specified rows, $\underline{E}_{\mathscr{M},\mathscr{Q}}$ satisfies a 'law of iterated lower expectations'—but $\underline{E}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ or $\underline{E}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ may not. This implies that for any $(f \,|\, X_u = x_u)$ in $\mathbb{JS}$, the lower expectation $\underline{E}_{\mathscr{M},\mathscr{Q}}(f \,|\, X_u = x_u)$ can be computed up to arbitrary precision through backwards recursion [see 11, Section 9]. This recursive scheme requires repeated evaluation of the 'lower envelope' $\underline{Q}$ of $\mathscr{Q}$, which is defined for any real-valued function f on $\mathscr{X}$ and any x in $\mathscr{X}$ by

$$[\underline{Q}f](x) := \inf\left\{\sum_{y \in \mathscr{X}} Q(x,y) f(y)\colon Q \in \mathscr{Q}\right\}.$$

The conjugate upper envelope $\overline{Q}$ is defined by $\overline{Q}f := -\underline{Q}f$ for any real-valued function f on $\mathscr{X}$.

4. Extending the Domain

The domain $\mathbb{JS}$ of the lower expectations corresponding to $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ 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 inferences are occupancy times. In essence, an *occupancy time* is the length of time that the system is in some subset A of $\mathscr{X}$ over some interval $[s,r]$; we will formally define this variable in Section 5 further on.

So how can we extend the domain of the lower and upper expectations corresponding to $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ to include these general variables? We propose the following approach: we first extend the domain of the expectation E_P corresponding to every jump process P in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ to a larger domain, and subsequently take the lower (and upper) envelope over these extended expectations.

4.1. Extending the Domain of a Jump Process

Consider any jump process P, and recall from Definition 3 that P is a coherent conditional probability on $\mathscr{D}$. From Section 2.3, we know that for every conditioning event $\{X_u = x_u\}$, $P(\bullet\,|\,X_u = x_u)$ is a probability charge. We set out to extend the domain of E_P by extending the domain of $E_P(\bullet\,|\,X_u = x_u)$ for every conditioning event $\{X_u = x_u\}$, and we will do so with the help of 'countable additivity'.

Definition 8 *A jump process P is* countably additive *if for all u in $\mathscr{U}$ and x_u in $\mathscr{X}_u$, $P(\bullet\,|\,X_u = x_u)$ is a countably-additive probability charge on $\mathscr{A}_u$, in the sense that for every sequence $(A_n)_{n\in\mathbb{N}}$ of pair-wise disjoint events in $\mathscr{A}_u$ such that $A := \bigcup_{n\in\mathbb{N}} A_n$ belongs to $\mathscr{A}_u$,*

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

Suppose P is a countably additive jump process, and fix some u in $\mathscr{U}$ and x_u in $\mathscr{X}_u$. Then by Caratheodory's Theorem [2, Theorem 3.1], there is a unique (countably additive) probability measure $P_\sigma(\bullet\,|\,X_u = x_u)$ on the σ-algebra $\sigma(\mathscr{A}_u)$ generated by $\mathscr{A}_u$ that extends $P(\bullet\,|\,X_u = x_u)$. Let $\overline{\mathbb{V}}_{\mathrm{b}}$ denote the set of all extended real variables $f\colon \Omega \to \overline{\mathbb{R}}$ that are either bounded below or bounded above, in the sense that $\inf f > -\infty$ or $\sup f < +\infty$, respectively.[4] A variable f in $\overline{\mathbb{V}}_{\mathrm{b}}$ is called $\sigma(\mathscr{A}_u)$-*measurable* if for all α in $\mathbb{R}$, the level set

$$\{f > \alpha\} := \{\omega \in \Omega\colon f(\omega) > \alpha\}$$

belongs to $\sigma(\mathscr{A}_u)$. We collect all $\sigma(\mathscr{A}_u)$-measurable variables in $\overline{\mathbb{V}}_{\mathrm{b}}$ in the set $\mathbb{M}_u$. For any variable f in $\mathbb{M}_u$, the expectation $E^\sigma_P(f\,|\,X_u = x_u)$ can now be defined through the Lebesgue integral with respect to $P_\sigma(\bullet\,|\,X_u = x_u)$ [2, Section 15]: if f is bounded below, then this integral is given by the Choquet integral [5, Chapter 5]

$$E^\sigma_P(f\,|\,X_u = x_u) := \inf f + \int_{\inf f}^{\sup f} P_\sigma(\{f > \alpha\}\,|\,X_u = x_u)\,\mathrm{d}\alpha,$$

where the integral is a—possibly improper—Riemann integral that always exists because $P_\sigma(\{f > \alpha\}\,|\,X_u = x_u)$ is a non-increasing function of α; if f is bounded above (but not below), then $E^\sigma_P(f\,|\,X_u = x_u) := -E^\sigma_P(-f\,|\,X_u = x_u)$, with $-f$ bounded below.

This way, we obtain an expectation E^σ_P with domain

$$\mathbb{JM} := \big\{(f\,|\,X_u = x_u)\colon u \in \mathscr{U}, x_u \in \mathscr{X}_u, f \in \mathbb{M}_u\big\}.$$

Note that $\mathbb{JS}$ is contained in $\mathbb{JM}$ and that E^σ_P coincides with E_P on $\mathbb{JS}$, so E^σ_P indeed extends E_P. The extended domain $\mathbb{JM}$ includes most of the 'conventional' (extended) real variables that are of interest; for example, we will see in Section 5 further on that $\mathbb{JM}$ includes occupancy times.

4.2. Extending the Domain of Imprecise Markovian Jump Processes

The extension method that we have laid out in Section 4.1 is only relevant to an imprecise jump processes $\mathscr{P} \subseteq \mathbb{P}$ if every jump processes P in $\mathscr{P}$ is countably additive. The following result establishes that this is the case for $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ and therefore, by Equation (4), also for $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$. Although it might seem a bit underwhelming, we believe that it is the single most important result in this contribution, and definitely the most difficult one to obtain.

Theorem 9 *Every jump process P in the imprecise jump process $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ 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 of all sample paths to obtain countable additivity, and then subsequently constructs a 'modification' of the projectors $(X_t)_{t\in\mathbb{R}_{\geq 0}}$ that has càdlàg sample paths—see, for example, [13, Lemma 3.16 and Theorem 3.18] or [2, Section 38]. Theorem 9 demonstrates that this is not needed—at least not for bounded $\mathscr{Q}$ and finite $\mathscr{X}$—as it is possible to immediately start with càdlàg paths. Ironically, our proof for Theorem 9 does make use of this 'modification' method, be it as an intermediate step under the hood.

Due to Theorem 9, we can extend the domain of the lower and upper expectations corresponding to $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$ as follows. First, for every jump process P in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, we use the method in Section 4.1 to extend the domain of the expectation E_P on $\mathbb{JS}$ by going over to the expectation E^σ_P on $\mathbb{JM}$. Second, we take the lower and upper envelopes over these extended expectations. For $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, this yields the lower and upper expectation $\underline{E}^\sigma_{\mathscr{M},\mathscr{Q}}$ and $\overline{E}^\sigma_{\mathscr{M},\mathscr{Q}}$ on $\mathbb{JM}$, defined for all $(f\,|\,X_u = x_u)$ in $\mathbb{JM}$ by

$$\underline{E}^\sigma_{\mathscr{M},\mathscr{Q}}(f\,|\,X_u = x_u) := \inf\big\{E^\sigma_P(f\,|\,X_u = x_u)\colon P \in \mathbb{P}_{\mathscr{M},\mathscr{Q}}\big\},$$

and similarly for $\overline{E}^\sigma_{\mathscr{M},\mathscr{Q}}(f\,|\,X_u = x_u)$. Note that $\underline{E}^\sigma_{\mathscr{M},\mathscr{Q}}$ and $\overline{E}^\sigma_{\mathscr{M},\mathscr{Q}}$ coincide with $\underline{E}_{\mathscr{M},\mathscr{Q}}$ and $\overline{E}_{\mathscr{M},\mathscr{Q}}$ on $\mathbb{JS}$, and that they are conjugate, in the sense that for all $(f\,|\,X_u = x_u)$ in $\mathbb{JM}$,

$$\overline{E}^\sigma_{\mathscr{M},\mathscr{Q}}(f\,|\,X_u = x_u) = -\underline{E}^\sigma_{\mathscr{M},\mathscr{Q}}(-f\,|\,X_u = x_u).$$

4. Requiring that they are either bounded below or bounded above is not essential, but it does facilitate a more elegant treatment. In most applications, the variables of interest satisfy this requirement.

We can do precisely the same for the lower and upper expectations corresponding to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$. However, in order not to exceed the page limit, we will from here on only state results for the lower and upper expectation corresponding to $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$. Theorems 10 and 11 further on also hold for the lower and upper expectations corresponding to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$ and $\mathbb{P}^{\mathrm{HM}}_{\mathscr{M},\mathscr{Q}}$, and so do Propositions 14 and 19 further on; Theorem 15 also holds for the lower and upper expectations corresponding to $\mathbb{P}^{\mathrm{M}}_{\mathscr{M},\mathscr{Q}}$, and we conjecture that the same is true for Theorem 20.

4.3. Continuity Properties

For all P in $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$, u in $\mathscr{U}$ and x_u in $\mathscr{X}_u$, the corresponding expectation operator $E^{\sigma}_P(\bullet\,|\,X_u = x_u)$ has well-known continuity properties: it is continuous with respect to pointwise convergent sequences of $\sigma(\mathscr{A}_u)$-measurable variables if these sequences are monotone—due to the Monotone Convergence Theorem [2, Theorem 16.2]—or uniformly bounded—due to Lebesgue's Dominated Convergence Theorem [2, Theorem 16.4]. Unfortunately, however, similar results are not available for the lower and upper expectations $\underline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(\bullet\,|\,X_u = x_u)$ and $\overline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(\bullet\,|\,X_u = x_u)$, 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 [14, 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 10 *Consider any u in $\mathscr{U}$, x_u in $\mathscr{X}_u$ and f in $\mathbb{M}_u$. Then for any non-decreasing sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathbb{M}_u$ with $\inf f_1 > -\infty$ that converges point-wise to f,*

$$\lim_{n\to+\infty} \overline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(f_n\,|\,X_u = x_u) = \overline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(f\,|\,X_u = x_u) \text{ and}$$
$$\lim_{n\to+\infty} \underline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(f_n\,|\,X_u = x_u) \leq \underline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(f\,|\,X_u = x_u).$$

Due to conjugacy, similar results apply to non-increasing sequences with $\sup f_1 < +\infty$. In that case, equality is obtained for the lower expectation, but not for the upper one.

Note that for non-decreasing sequences, the limit of $\underline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(f_n\,|\,X_u = x_u)$ exists as n recedes to infinity; and while this limit may not give a tight bound, it always yields a conservative one. For dominated sequences, we have a similar type of conservative limit behaviour.

Theorem 11 *Consider any u in $\mathscr{U}$, x_u in $\mathscr{X}_u$ and f in $\mathbb{M}_u$. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{M}_u$ that converges point-wise to f. If $(f_n)_{n\in\mathbb{N}}$ is uniformly bounded, meaning that there is some B in $\mathbb{R}_{\geq 0}$ such that $|f_n| \leq B$ for all n in $\mathbb{N}$, then*

$$\liminf_{n\to+\infty} \overline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(f_n\,|\,X_u = x_u) \geq \overline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(f\,|\,X_u = x_u) \text{ and}$$
$$\limsup_{n\to+\infty} \underline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(f_n\,|\,X_u = x_u) \leq \underline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}(f\,|\,X_u = x_u).$$

5. Two Measurable Variables

While there is a (potential) lack of continuity in general, it turns out that for many practically relevant variables, the lower expectation $\underline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}$ is actually continuous. We discuss here two examples of such variables: occupancy times and the number of jumps to some set of states.

Our exposition simplifies due to the following notation and terminology. Fix two time points s,r in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. A *grid* over $[s,r]$ is a sequence of distinct time points $v = (t_0,\ldots,t_n)$ in $\mathscr{U}_{\mathrm{ne}}$ 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

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

the *maximum grid width of v*. Note that whenever $s = r$, there is only the degenerate grid $v = (s)$; in this case, we let $\Delta(v) := 0$. We say that a grid w over $[s,r]$ refines a grid v over $[s,r]$, and denote this by $v \sqsubseteq w$, if w includes all the time points in v.

5.1. Integral of $f(X_t)$ Over Time

Occupancy times belong to a larger class of real variables that are formally defined through a Riemann integral over time. Fix some real-valued function f on $\mathscr{X}$ and any two time points s,r in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. 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 [18, Section 24.26], it follows that $f \circ \omega$ is Riemann integrable over $[s,r]$. Hence,

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

is well-defined, and we call this real variable the *integral of $f(X_t)$ over $[s,r]$*.

Take, for example, the real-valued function $f = \mathbb{I}_A$ for some subset A of $\mathscr{X}$. 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 point s and r; as explained before, we call this the *occupancy time of A over $[s,r]$* [12, Section 4.5].

Since the integral $\int_s^r f(X_t)\,\mathrm{d}t$ is defined through Riemman sums [see 18, Definition 24.3], we can use these 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

$$\langle f\rangle_v := \sum_{k=0}^{n-1} (t_{k+1} - t_k) f(X_{t_k}).$$

Note that in this expression, for every k in $\{0,\dots,n-1\}$, we could replace $f(X_{t_k})$ with $f(X_{s_k})$ where s_k is any element of $[t_k,t_{k+1}]$; 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 4.

Corollary 12 *Consider some u in $\mathscr{U}$ and some s,r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Then for any grid v over $[s,r]$ and any real-valued function f on $\mathscr{X}$, the corresponding Riemann sum $\langle f\rangle_v$ is $\mathscr{A}_u$-simple, and*

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

Due to the definition of the Riemann integral [18, 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 13 *Consider some u in $\mathscr{U}$ and some s,r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Fix some real-valued function f on $\mathscr{X}$, 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 sequence of $\mathscr{A}_u$-simple variables that converges point-wise to $\int_s^r f(X_t)\,\mathrm{d}t$; hence, $\int_s^r f(X_t)\,\mathrm{d}t$ belongs to $\mathbb{M}_u$.*

Suppose $(v_n)_{n\in\mathbb{N}}$ is a sequence of grids over $[s,r]$ with $\lim_{n\to+\infty}\Delta(v_n)=0$. Then by Corollary 12 and Lemma 13, 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 the integral $\int_s^r f(X_t)\,\mathrm{d}t$. Consequently, we could invoke Theorem 11 to conservatively approximate the lower and upper expectation of $\int_s^r f(X_t)\,\mathrm{d}t$. In this particular case, however, the approximation in Theorem 11 is actually tight.

Proposition 14 *Consider some u in $\mathscr{U}$ and some s,r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Fix a real-valued function f on $\mathscr{X}$, 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 for any x_u in $\mathscr{X}_u$,*

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

and a similar equality holds for the conjugate upper expectation $\overline{E}_{\mathscr{M},\mathscr{Q}}$.

Because $\langle f\rangle_{v_n}$ is an $\mathscr{A}_u$-simple variable by Corollary 12, we could use the backwards recursive method of Krak et al. [11, Section 9] to compute $\underline{E}_{\mathscr{M},\mathscr{Q}}\left(\langle f\rangle_{v_n} \,\middle|\, X_u = x_u\right)$ up to arbitrary precision, provided that $\mathscr{Q}$ is convex and has separately specified rows, and then refine the grid until we observe empirical convergence. However, to make this method computationally tractable, it would have to be combined with the decomposition ideas of T'Joens et al. [20]. It is therefore arguably more practical to use the following direct approximation method, which can also be used if $\mathscr{Q}$ is *not* convex, and which has these ideas—and the resulting efficiency—already built in.

Theorem 15 *Consider some s,r in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. Fix some real-valued function f on $\mathscr{X}$. For all n in $\mathbb{N}$, let $\Delta_n := (r-s)/n$ and let $f_{n,n}$ be recursively defined by the initial condition $f_{n,0} := 0$ and, for all k in $\{1,\dots,n\}$, by*

$$f_{n,k} := \Delta_n f + f_{n,k-1} + \Delta_n \underline{Q} f_{n,k-1}. \tag{5}$$

If $\mathscr{Q}$ has separately specified rows, then for any x in $\mathscr{X}$,

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

a similar equality holds for the upper expectation $\overline{E}_{\mathscr{M},\mathscr{Q}}$ if we replace $\underline{Q}$ by $\overline{Q}$ in Equation (5). If $\mathscr{Q}$ does not have separately specified rows, the obtained results provide conservative—outer—bounds.

5.2. Number of Jumps to A

As a second example of a measurable variable that depends on the state of the system at more than a finite number of time points, we consider 'the number of jumps' to some subset A of the state space $\mathscr{X}$, meaning the number of times that the system changes state from a state in A^{c} to a state in A. For any subset A of $\mathscr{X}$ and any time points s,r in $\mathbb{R}_{\geq 0}$ such that $s \leq r$, we let $\eta^A_{(s,r]}\colon \Omega \to \mathbb{R}$ be the real variable that is defined for all ω in Ω by

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

Note that $\eta^A_{(s,r]}$ is a real variable because any càdlàg path ω in Ω can have but a finite number of jumps—that is, discontinuity points—in the bounded interval $[s,r]$.

Here too, we use a grid over $[s,r]$ to approximate $\eta^A_{(s,r]}$. More precisely, for any grid $v=(t_0,\dots,t_n)$ over $[s,r]$, we use the corresponding approximation $\eta^A_v\colon \Omega \to \mathbb{R}$, which is defined for all ω in Ω by

$$\eta^A_v(\omega) := \left|\left\{k \in \{1,\dots,n\}\colon \omega(t_k) \in A, \omega(t_{k-1}) \notin A\right\}\right|.$$

It is easy to see that

$$\eta^A_v = \sum_{k=1}^{n} \mathbb{I}_{A^{\mathrm{c}}}(X_{t_{k-1}})\mathbb{I}_A(X_{t_k}).$$

Hence, we have the following corollary of Lemma 4.

Corollary 16 *Consider some u in $\mathscr{U}$ and some s,r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Then for any subset A of $\mathscr{X}$ and any grid v over $[s,r]$, the corresponding approximation η^A_v is $\mathscr{A}_u$-simple.*

Consider now a second grid w over $[s,r]$ that refines the grid v. Then it is not all too difficult to verify that the number of jumps to A over w is (point-wise) greater than or equal to the number of jumps to A over v.

Lemma 17 *Consider some u in $\mathscr{U}$ and some s,r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Then for any subset A of $\mathscr{X}$ and any two grids v and w over $[s,r]$ with $v \sqsubseteq w$, $\eta^A_v \leq \eta^A_w$.*

Furthermore, it follows almost immediately from the properties of càdlàg paths that the approximation η^A_v converges point-wise to $\eta^A_{(s,r]}$ as the grid width $\Delta(v)$ of the grid v vanishes.

Lemma 18 *Consider some u in $\mathscr{U}$ and some s,r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s < r$. Fix some subset A of $\mathscr{X}$ 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 $(\eta^A_{v_n})_{n\in\mathbb{N}}$ is a sequence of $\mathscr{A}_u$-simple variables that converges point-wise to $\eta^A_{(s,r]}$; hence, $\eta^A_{(s,r]}$ belongs to $\mathbb{M}_u$.*

Suppose $(v_n)_{n\in\mathbb{N}}$ is a sequence of grids over $[s,r]$ such that $\lim_{n\to+\infty}\Delta(v_n) = 0$ and $v_n \sqsubseteq v_{n+1}$ for all n in $\mathbb{N}$. Then by Corollary 16 and Lemmas 17 and 18, the corresponding sequence $(\eta^A_{v_n})_{n\in\mathbb{N}}$ is a non-decreasing sequence of simple variables that converges point-wise to $\eta^A_{(s,r]}$. Consequently, we could invoke Theorem 10 to approximate the lower and upper expectation of $\eta^A_{(s,r]}$. In this particular case though, these approximations are actually tight. Even more, we can drop the requirement that the next grid in the sequence should always refine the preceding one.

Proposition 19 *Consider some u in $\mathscr{U}$ and some s,r in $\mathbb{R}_{\geq 0}$ such that $\max u \leq s \leq r$. Fix some subset A of $\mathscr{X}$, 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 for any x_u in $\mathscr{X}_u$,*

$$\underline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}\left(\eta^A_{(s,r]} \,\middle|\, X_u = x_u\right) = \lim_{n\to+\infty} \underline{E}_{\mathscr{M},\mathscr{Q}}\left(\eta^A_{v_n} \,\middle|\, X_u = x_u\right),$$

and a similar equality holds for the conjugate upper expectation $\overline{E}_{\mathscr{M},\mathscr{Q}}$.

Here too, we could use the backwards recursive method of Krak et al. [11, Section 9] to compute the lower and upper expectation of $\eta^A_{(s,r]}$ up to arbitrary precision, provided that $\mathscr{Q}$ has separately specified rows and is convex. That said, it is more efficient to use a direct approximation method similar to that in Theorem 15, now taking inspiration from the decomposition ideas of De Bock et al. [4].

Theorem 20 *Consider some s,r in $\mathbb{R}_{\geq 0}$ such that $s \leq r$. Fix some subset A of $\mathscr{X}$. For all n in $\mathbb{N}$, let $\Delta_n := (r-s)/n$ and let $f_{n,n}$ be recursively defined by the initial condition $f_{n,0} := 0$ and, for all k in $\{1,\ldots,n\}$, by*

$$\begin{aligned} f_{n,k}\colon \mathscr{X} &\to \mathbb{R} \\ x &\mapsto f_{n,k-1}(x) + \Delta_n\left[\underline{Q}(\mathbb{I}_{A^c}(x)\mathbb{I}_A + f_{n,k-1})\right](x). \end{aligned} \tag{6}$$

If $\mathscr{Q}$ has separately specified rows, then for any x in $\mathscr{X}$,

$$\underline{E}_{\mathscr{M},\mathscr{Q}}\left(\eta^A_{(s,r]} \,\middle|\, X_s = x\right) = \lim_{n\to+\infty} f_{n,n}(x);$$

a similar equality holds for the upper expectation $\overline{E}_{\mathscr{M},\mathscr{Q}}$ if we replace $\underline{Q}$ by $\overline{Q}$ in Equation (6). If $\mathscr{Q}$ does not have separately specified rows, the obtained results provide conservative—outer—bounds.

6. Numerical Examples

Troffaes et al. [23] use an imprecise jump process to asses the reliability of a power network. They follow up on their earlier work [21] 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*.

Troffaes et al. [23, Sections 2.3 and 3.4] model this power network with an imprecise jump process as follows. The state space is $\mathscr{X} := \{\texttt{AB},\texttt{A},\texttt{B},\texttt{F}\}$, where the state `F` corresponds to a failure of both power lines and where the other state labels indicate the power lines that are working. The set $\mathscr{Q}$ of rate matrices is specified through lower and upper bounds on the off-diagonal components of the rate matrices:

$$\begin{aligned} \mathscr{Q} := \big\{Q \in \mathfrak{Q}\colon\ &(\forall x,y \in \mathscr{X}, x \neq y) \\ &Q_{\mathrm{L}}(x,y) \leq Q(x,y) \leq Q_{\mathrm{U}}(x,y)\big\}, \end{aligned}$$

where the matrices

$$Q_{\mathrm{L}} := \begin{pmatrix} \diamond & 0.32 & 0.32 & 0.19 \\ 730 & \diamond & 0 & 0.51 \\ 730 & 0 & \diamond & 0.51 \\ 0 & 730 & 730 & \diamond \end{pmatrix}$$

and

$$Q_{\mathrm{U}} := \begin{pmatrix} \diamond & 0.37 & 0.37 & 0.24 \\ 1460 & \diamond & 0 & 0.61 \\ 1460 & 0 & \diamond & 0.61 \\ 0 & 1460 & 1460 & \diamond \end{pmatrix}$$

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. Note that, by construction, $\mathscr{Q}$ has separately specified rows. Furthermore, evaluating the lower envelope $\underline{Q}$ of $\mathscr{Q}$ is almost trivial because of the specific structure of $\overline{\mathscr{Q}}$. Troffaes et al. [23] do not specify a set $\mathscr{M}$ of initial probability mass functions. As $\mathscr{M}$ plays no role in the following analysis, it can be chosen arbitrarily.

Troffaes et al. [23, Section 3.4] use—an informal version of—the imprecise jump process $\mathbb{P}_{\mathscr{M},\mathscr{Q}}$ to obtain lower

Table 1: Estimates for $\underline{\alpha}^x_{\mathtt{F}}$ and $\overline{\alpha}^x_{\mathtt{F}}$, in years.

x	$\underline{\alpha}^x_{\mathtt{F}}$	$\overline{\alpha}^x_{\mathtt{F}}$
AB	6.512×10^{-4}	1.647×10^{-3}
F	9.938×10^{-4}	2.332×10^{-3}
[23]	6.513×10^{-4}	1.647×10^{-3}

and upper bounds on two performance measures: the expected time that the power network is down over a period of 10 years, and the expected number of times that the power network is down over a period of 10 years.

Let us first consider the expected time that the power network is down over a period of 10 years. In our formalism, the time that the power network is down over a period of 10 years corresponds to the integral $\int_0^{10} \mathbb{I}_{\mathtt{F}}(X_t)\,\mathrm{d}t$ of $\mathbb{I}_{\mathtt{F}} := \mathbb{I}_{\{\mathtt{F}\}}$ over $[0,10]$. For any initial state x in $\mathscr{X}$, we use Theorem 15 to compute the lower expected value

$$\underline{\alpha}^x_{\mathtt{F}} := \underline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}\left(\int_0^{10} \mathbb{I}_{\mathtt{F}}(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right)$$

of this downtime, and similarly for the upper expected downtime, which we denote by $\overline{\alpha}^x_{\mathtt{F}}$. More precisely, we start with $n = 29\,200$ iterations—this to ensure that $(I + \Delta_n \underline{Q})$ is a so-called lower transition operator [see 8, Proposition 3]—and repeatedly increase the number of iterations by a factor 10 until we observe empirical convergence, up to four significant digits. Our results are reported in Table 1 for the initial states $x = \mathtt{AB}$ and $x = \mathtt{F}$, and we also report the lower and upper bounds that Troffaes et al. [23, Equation 60] find using their heuristic. If we start with two functioning power lines—so for $x = \mathtt{AB}$—our approximations are (almost) equal to those of Troffaes et al. [23], up to three significant digits. If we start in the state where both power lines are down—so for $x = \mathtt{F}$—our results are different. The heuristics of Troffaes et al. [23] 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 $x = \mathtt{F}$.

Second, we consider the expected number of times that the power network is down over a period of 10 years. In our setting, if we start in the state $x = \mathtt{AB}$, the number of times that the power network is down over a period of 10 years corresponds to $\eta^{\{\mathtt{F}\}}_{(0,10]}$, the number of jumps to $\{\mathtt{F}\}$ over $(s,r]$. Thus, we can use Theorem 20 to compute the lower expected number of downtimes

$$\underline{\beta}_{\mathtt{F}} := \underline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}\left(\eta^{\{\mathtt{F}\}}_{(0,10]} \,\middle|\, X_0 = \mathtt{AB}\right),$$

and similarly for the upper expected number of downtimes $\overline{\beta}_{\mathtt{F}}$. We report our results in Table 2. Here too, our results agree surprisingly well with those of Troffaes et al. [23].

Table 2: Estimates for $\underline{\beta}_{\mathtt{F}}$ and $\overline{\beta}_{\mathtt{F}}$.

	Theorem 15	[23]
$\underline{\beta}_{\mathtt{F}}$	1.902	1.900
$\overline{\beta}_{\mathtt{F}}$	2.405	2.407

As a final experiment, we consider an imprecise jump process with state space $\mathscr{X} := \{a,b,c\}$, set of rate matrices

$$\mathscr{Q} := \left\{ \begin{pmatrix} -\lambda_a & \lambda_a & 0 \\ \mu_b & -\mu_b - \lambda_b & \lambda_b \\ 0 & \mu_c & -\mu_c \end{pmatrix} : \begin{array}{l} \lambda_a = 1, \mu_b = 10, \\ \lambda_b \in [1,100], \\ \mu_c \in [1,100] \end{array} \right\}$$

and arbitrary set $\mathscr{M}$ of initial probability mass functions. In this case, we are after the upper expected fraction of time that the system is in state b. For each initial state x in $\mathscr{X}$, this upper expected fraction is given by

$$\overline{\gamma}^x_b := \lim_{T \to +\infty} \frac{1}{T} \overline{E}^{\sigma}_{\mathscr{M},\mathscr{Q}}\left(\int_0^T \mathbb{I}_a(X_t)\,\mathrm{d}t \,\middle|\, X_0 = x\right),$$

and can thus be obtained by applying Theorem 15 for increasingly larger T until empirical convergence is observed. The obtained result is $\overline{\gamma}_b := \overline{\gamma}^x_b = 0.091 = 9.1\%$ and does not depend on x, so there are no transient effects of the initial state here. Computing the same inference with the heuristic of Troffaes et al. [23], we find that $\overline{\gamma}_b = 0.703 = 70.3\%$. This shows that while the heuristic method of Troffaes et al. [23] 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 integrals over time and the number of jumps to a set of states, 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 heuristic of Troffaes et al. [23] with respect to the tightness of the bounds.

Acknowledgments

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

References

[1] Patrizia Berti, Eugenio Regazzini, and Pietro Rigo. Coherent statistical inference and Bayes theorem. *The Annals of Statistics*, 19(1):366–381, 1991. doi: 10.1214/aos/1176347988.

[2] Patrick Billingsley. *Probability and Measure*. Wiley, 3rd edition, 1995.

[3] Patrick Billingsley. *Convergence of Probability Measures*. Wiley, 2nd edition, 1999.

[4] Jasper De Bock, Alexander Erreygers, and Thomas Krak. Sum-product laws and efficient algorithms for imprecise Markov chains. Submitted for possible publication in the Proceedings of UAI 2021.

[5] Dieter Denneberg. *Non-Additive Measure and Integral*. Springer, 1994. doi: 10.1007/978-94-017-2434-0.

[6] Lester E. Dubins. Finitely additive conditional probabilities, conglomerability and disintegrations. *Annals of Probability*, 3(1):89–99, 1975. doi: 10.1214/aop/1176996451.

[7] Alexander Erreygers. *Markovian Imprecise Jump Processes: Foundations, Algorithms and Applications*. PhD thesis, Ghent University, 2021. Forthcoming, see https://users.ugent.be/~aerreyge/ for a preliminary version.

[8] Alexander Erreygers and Jasper De Bock. Imprecise continuous-time Markov chains: Efficient computational methods with guaranteed error bounds. In *Proceedings of ISIPTA'17*, pages 145–156. PMLR, 2017.

[9] Alexander Erreygers and Jasper De Bock. Bounding inferences for large-scale continuous-time Markov chains: A new approach based on lumping and imprecise Markov chains. *International Journal of Approximate Reasoning*, 115:96–133, 2019. doi: j.ijar.2019.09.003.

[10] Iosif I. Gikhman and Anatoly V. Skorokhod. *Introduction to the Theory of Random Processes*. W.B. Saunders Company, 1969.

[11] Thomas Krak, Jasper De Bock, and Arno Siebes. Imprecise continuous-time Markov chains. *International Journal of Approximate Reasoning*, 88:452–528, 2017. doi: 10.1016/j.ijar.2017.06.012.

[12] Vidyadhar G. Kulkarni. *Introduction to Modeling and Analysis of Stochastic Systems*. Springer, 2011. doi: 10.1007/978-1-4419-1772-0.

[13] Jean-François Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Springer, 2016. doi: 10.1007/978-3-319-31089-3.

[14] Enrique Miranda and Marco Zaffalon. Full conglomerability. *Journal of Statistical Theory and Practice*, 11(4):634–669, 2017. doi: 10.1080/15598608.2017.1295890.

[15] Max Nendel. Markov chains under nonlinear expectation. *Mathematical Finance*, 31(1):474–507, 2021. doi: 10.1111/mafi.12289.

[16] Shige Peng. Nonlinear expectations and nonlinear Markov chains. *Chinese Annals of Mathematics*, 26(02):159–184, 2005. doi: 10.1142/S0252959905000154.

[17] Eugenio Regazzini. Finitely additive conditional probabilities. *Rendiconti del Seminario Matematico e Fisico di Milano*, 55(1):69–89, 1985. doi: 10.1007/BF02924866.

[18] Eric Schechter. *Handbook of Analysis and Its Foundations*. Academic Press, 1997.

[19] Damjan Škulj. Efficient computation of the bounds of continuous time imprecise Markov chains. *Applied Mathematics and Computation*, 250:165–180, 2015. doi: 10.1016/j.amc.2014.10.092.

[20] Natan T'Joens, Thomas Krak, Jasper De Bock, and Gert de Cooman. A recursive algorithm for computing inferences in imprecise Markov chains. In *ECSQARU 2019: Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 455–465. Springer, 2019. doi: 10.1007/978-3-030-29765-7_38.

[21] Matthias C.M. Troffaes and Simon Blake. A robust data driven approach to quantifying common-cause failure in power networks. In *Proceedings of ISIPTA'13*, pages 311–317, 2013.

[22] Matthias C.M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley, 2014.

[23] Matthias C.M. Troffaes, Jacob Gledhill, Damjan Škulj, and Simon Blake. Using imprecise continuous time Markov chains for assessing the reliability of power networks with common cause failure and non-immediate repair. In *Proceedings of ISIPTA'15*, pages 287–294, 2015.