Sum-Product Laws and Efficient Algorithms for Imprecise Markov Chains

Jasper De Bock[1] **Alexander Erreygers**[1] **Thomas Krak**[2]

[1]Foundations Lab for imprecise probabilities, ELIS, Ghent University, Belgium
[2]Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands

Abstract

We propose two sum-product laws for imprecise Markov chains, and use these laws to derive two algorithms to efficiently compute lower and upper expectations for imprecise Markov chains under complete independence and epistemic irrelevance. These algorithms work for inferences that have a corresponding sum-product decomposition, and we argue that many well-known inferences fit their scope. We illustrate our results on a simple epidemiological example.

1 INTRODUCTION

Imprecise Markov chains are a generalisation of Markov chains that allows them to deal with numerical parameters that are only partially specified, as well as possible violations of structural assumptions like Markovianity [Hermans and Škulj, 2014]. While different interpretations are possible, we will view an imprecise Markov chain as a set of stochastic processes. Depending on what kind of processes are included in this set, one obtains a different type of imprecise Markov chain; we will consider two of them.

The first type is basically a set of (traditional) Markov chains. These are now known as imprecise Markov chains under complete independence,[1] and were studied by e.g. Škulj [2009]. The second type that we consider are imprecise Markov chains under epistemic irrelevance, as introduced by De Cooman and Hermans [2008]. Unlike the first type, these sets of stochastic processes also contain non-Markovian ones. The set as a whole does satisfy a Markov property though, and this is why this second type of model is called an imprecise Markov chain as well.

[1]Most authors refer to it as an imprecise Markov chain under strong independence [Hermans and Škulj, 2014], but it should be clear from [Cozman, 2012] that this is a misnomer.

The merits of both these models lie in their ability to model the same kind of problems as (traditional) Markov chains, but under much more general assumptions; in cases where the dynamics under study are not completely known, they can still provide meaningful, non-trivial bounds on inferences of interest. In particular, this approach goes beyond what is feasible using more classical sensitivity analysis methods; not only can they capture the effect of perturbing the numerical parameters that specify the model dynamics, but they can also explicitly incorporate uncertainty about, and evaluate the effect of, assumptions of homogeneity and Markovianity. Specifically, the possibility of relaxing the assumption of Markovianity becomes apparent by considering the distinction between imprecise Markov chains under complete independence and epistemic irrelevance.

Perhaps surprisingly, it was the second of these two types that led to the development of many efficient inference algorithms; see e.g. the work of T'Joens et al. [2019] for a recent efficient algorithm that encompasses most previous algorithms as a special case, and that can be used to solve a wide class of practically relevant inference problems. For imprecise Markov chains under complete independence, however, almost no algorithms are available. We find this unfortunate, because that model is arguably more natural to conceptualise from a practitioner's point of view.

This brings us to the contributions of this work: we derive two efficient algorithms to compute (different types of) inferences for sets of stochastic processes, and we provide sufficient conditions for the applicability of these algorithms in terms of what we call sum-product laws. Crucially, we show that under some relatively mild conditions on their model parameters, imprecise Markov chains under complete independence and imprecise Markov chains under epistemic irrelevance both satisfy these sum-product laws, which implies that our algorithms can be applied to either of them. We illustrate the practical relevance of our algorithms by sketching how they cover, as special cases, many well-known inference problems, and apply them to a toy example in epidemiology to demonstrate their performance.

Accepted for the 37th Conference on Uncertainty in Artificial Intelligence (UAI 2021).

2 IMPRECISE MARKOV CHAINS

An imprecise Markov chain is a Markov chain whose local (transition) probabilities are partially specified, and whose Markov property may not hold exactly. Since we will formally define these as specific sets of stochastic processes, we start with a brief introduction to stochastic process, and the special case of (traditional) Markov chains.

2.1 STOCHASTIC PROCESSES

A stochastic process is a representation of the uncertain behaviour of some dynamical system of interest, as it moves through some state space $\mathscr{X}$. Throughout this work, we will assume that $\mathscr{X}$ is a fixed, non-empty, and finite set, whose elements we refer to as states. To parametrise a stochastic process, we will use a probability tree $p_\bullet$. This is simply a collection of probability mass functions on $\mathscr{X}$ that, for every sequence of initial states $x_{1:n} := (x_1,\ldots,x_n) \in \mathscr{X}^n$, specifies a probability mass function $p_{x_{1:n}}$ on $\mathscr{X}$. We denote the set of all such probability trees by $\mathbb{P}$.

A stochastic process, then, is essentially just a conditional probability distribution P over a space of events that represents an infinite sequence $X_1, X_2, \ldots, X_n, \ldots$ of random states. In particular, if a stochastic process is parametrised by a probability tree $p_\bullet$, then for any $x_{n+1} \in \mathscr{X}$, $p_{x_{1:n}}(x_{n+1})$ provides the probability that the state of the system is x_{n+1} at time $n+1$ given that its previous states were $x_{1:n}$:

$$P(X_{n+1} = x_{n+1} | X_{1:n} = x_{1:n}) = p_{x_{1:n}}(x_{n+1}). \quad (1)$$

For $n = 0$, the empty sequence $x_{1:0} = ()$ will also be denoted by $\square$. In that particular case, for every $x_1 \in \mathscr{X}$, the probability that the process starts in the state x_1 is given by

$$P(X_1 = x_1) := P(X_1 = x_1 | X_{1:0} = x_{1:0}) = p_\square(x_1).$$

Other probabilities are implied by the laws of probability. For every $x_{1:n} \in \mathscr{X}^n$ and $z_{1:m} \in \mathscr{X}^m$, we have that

$$P(X_{1:n} = z_{1:n} | X_{1:m} = x_{1:m}) := \begin{cases} \prod_{i=m}^{n-1} p_{z_{1:i}}(z_{i+1}) & \text{if } n > m \text{ and } z_{1:m} = x_{1:m} \\ 1 & \text{if } n \leq m \text{ and } z_{1:n} = x_{1:n} \\ 0 & \text{otherwise} \end{cases}$$

and, for every $A \subseteq \mathscr{X}^n$, it then follows that

$$P(X_{1:n} \in A | X_{1:m} = x_{1:m}) = \sum_{z_{1:n} \in A} P(X_{1:n} = z_{1:n} | X_{1:m} = x_{1:m}).$$

More involved probabilities—that depend on the state at more than a finite number of time points—can be obtained by considering the sigma-additive extension of these probabilities, but this will not be needed here.

What we will mainly be interested in are expectations. In particular, expectations of real-valued functions on $\mathscr{X}^n$. We call such functions gambles and use $\mathscr{G}(\mathscr{X}^n)$ to denote the set of all of them. For any $f \in \mathscr{G}(\mathscr{X}^n)$ and any $x_{1:m} \in \mathscr{X}^m$, the conditional expectation of $f(X_{1:n})$ is given by

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

Conditional probabilities correspond to a special case of expectations. For any $A \subseteq \mathscr{X}^n$, we have that

$$P(X_{1:n} \in A | X_{1:m} = x_{1:m}) = E(\mathbb{I}_A(X_{1:n}) | X_{1:m} = x_{1:m}), \quad (2)$$

where the gamble $\mathbb{I}_A$ is the indicator of A, defined by $\mathbb{I}_A(z_{1:n}) := 1$ if $z_{1:n} \in A$ and $\mathbb{I}_A(z_{1:n}) := 0$ otherwise.

2.2 MARKOV CHAINS

A stochastic process is called a Markov chain if it satisfies the Markov property. That is, if $P(X_{n+1} = x_{n+1} | X_{1:n} = x_{1:n})$ only depends on n and x_n—and hence not on $x_{1:n-1}$. In other words: given the current state X_n of the system, the uncertainty model for the next state X_{n+1} does not depend on the previous states $X_{1:n-1}$. What is particularly convenient about Markov chains is that we do not need to specify a complete probability tree to parametrise it. Instead, as we will see, we can use transition matrices.

A transition matrix T is a $|\mathscr{X}| \times |\mathscr{X}|$ matrix whose rows are probability mass functions. We will denote the set of all such transition matrices by

$$\mathbb{T} := \{T \in \mathbb{R}^{\mathscr{X} \times \mathscr{X}} : T(x,\cdot) \in \Sigma_{\mathscr{X}} \text{ for all } x \in \mathscr{X}\},$$

where $\Sigma_{\mathscr{X}}$ is the set of all probability mass functions on $\mathscr{X}$, and $T(x,\cdot)$ is the function on $\mathscr{X}$ whose value in $y \in \mathscr{X}$ is given by $T(x,y)$. As usual, we can also identify a matrix T with a linear operator from $\mathscr{G}(\mathscr{X})$ to $\mathscr{G}(\mathscr{X})$, defined for all $f \in \mathscr{G}(\mathscr{X})$ by $Tf(x) := \sum_{y \in \mathscr{X}} T(x,y) f(y)$.

To parametrise a Markov chain, the only thing we need to specify is an initial probability mass function $p_\square \in \Sigma_{\mathscr{X}}$ and, for all $n \in \mathbb{N}$,[2] a transition matrix $T_n \in \mathbb{T}$. The initial model $p_\square$ serves the same function as it does in general stochastic processes. The transition matrices provide transition probabilities. In particular, for every $n \in \mathbb{N}$, T_n provides the probabilities for the state transitions from time n to $n+1$:

$$p_{x_{1:n}} := T_n(x_n,\cdot) \text{ for all } x_{1:n} \in \mathscr{X}^n.$$

In this way, together with the initial model $p_\square$, the transition matrices $(T_n)_{n \in \mathbb{N}}$ specify a complete probability tree $p_\bullet$, and hence a stochastic process. That this stochastic process is a Markov chain, and that any Markov chain can be specified in this way, follows from the fact that—due to Equation (1)—the Markov property is equivalent to the requirement that $p_{x_{1:n}}$ only depends on n and x_n.

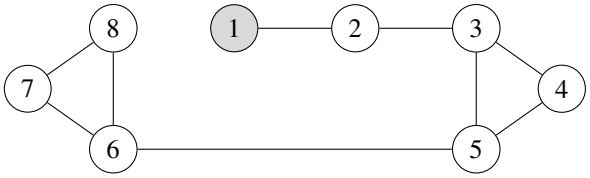

Figure 1: A Network of 8 People and Their Connections

Example 1. *Throughout this work, we illustrate our methods using a simple running toy example. To this end, we consider the evolution of a disease as it spreads within a small group of 8 people, depicted in Figure 1; every node is an individual and an edge between people indicates that they are regularly in contact with one another. The state of the system keeps track, for every person, whether they are sick or healthy; hence this system has* $|\mathscr{X}| = 2^8$ *states.*

This system can be modelled using a Markov chain, which we will assume is homogeneous. This means that there is a single transition matrix T *such that the transition matrices* $(T_n)_{n\in\mathbb{N}}$ *that parametrise the model satisfy* $T_n = T$ *for all* $n \in \mathbb{N}$*. The matrix* T *is obtained as follows. For any given current state of the system—so any configuration of who is sick and who is healthy—and any sick person* k*, the probability that they become healthy after one time step is* $p_{\mathrm{H}} = 0.2$ *and, for any of their neighbours in Figure 1, the probability that they are infected by person* k *is* $p_{\mathrm{I}} = 0.5$*. All these events are furthermore independent. For example, if state* x *corresponds to only person* 1 *and* 3 *being sick, and state* y *corresponds to only person* 2 *being sick, then*

$$T(x,y) = p_{\mathrm{H}}^2(1-p_{\mathrm{I}})^2\big(1-(1-p_{\mathrm{I}})^2\big) = 0.0075.$$

Moreover, for the initial model $p_\square$*, we consider the probability mass function that assigns all mass to the state where person 1 is sick and the others are healthy.* ◇

2.3 A RANGE OF IMPRECISE MARKOV CHAINS

Imprecise Markov chains come in different types, two of which we will focus on here. Both of them are sets of stochastic processes, but since stochastic processes are in one-to-one correspondence with probability trees, we can—and will—equivalently regard them as sets of probability trees. For the two types that we will consider, the starting point is a non-empty set $\mathscr{M}_\square$ of probability mass functions and, for all $n \in \mathbb{N}$, a non-empty set $\mathscr{T}_n$ of transition matrices. Throughout this work, we will take these sets to be fixed.

The first type of imprecise Markov chain that we consider, and arguably the most simple and intuitive one, consists of all Markov chains whose initial model $p_\square$ belongs to $\mathscr{M}_\square$ and whose transition matrix T_n, for all $n \in \mathbb{N}$, belongs to $\mathscr{T}_n$. We denote the set of all such Markov chains—or more formally, the corresponding set of probability trees—by

$$\mathscr{P}^{\mathrm{M}} := \big\{p_\bullet \in \mathbb{P}\colon\ p_\square \in \mathscr{M}_\square \text{ and, for all } n \in \mathbb{N},\ (\exists T_n \in \mathscr{T}_n)(\forall x_{1:n} \in \mathscr{X}^n)\, p_{x_{1:n}} = T_n(x_n,\cdot)\big\}.$$

We will refer to this type of imprecise Markov chain as an imprecise Markov chain under complete independence.

The second type of imprecise Markov chain that we consider is similar, but a bit more involved. The difference is that the transition matrix T_n can now depend on the states $x_{1:n-1}$ at previous time points, leading us to denote it by $T_{n,x_{1:n-1}}$. In particular, we consider the set of probability trees

$$\mathscr{P}^{\mathrm{EI}} := \big\{p_\bullet \in \mathbb{P}\colon\ p_\square \in \mathscr{M}_\square \text{ and, for all } n \in \mathbb{N},\ (\forall x_{1:n-1} \in \mathscr{X}^{n-1})(\exists T_{n,x_{1:n-1}} \in \mathscr{T}_n)\ (\forall x_n \in \mathscr{X})\, p_{x_{1:n}} = T_{n,x_{1:n-1}}(x_n,\cdot)\big\}.$$

This type of imprecise Markov chain is known as an imprecise Markov chain under epistemic irrelevance [De Cooman and Hermans, 2008, Hermans and Škulj, 2014].

An important observation here is that the elements of $\mathscr{P}^{\mathrm{EI}}$ are not required to satisfy the Markov property. Some of them will, and those are exactly the elements of $\mathscr{P}^{\mathrm{M}}$, but $\mathscr{P}^{\mathrm{EI}}$ will typically also contain many stochastic processes that are not Markov chains. That this set of processes is nevertheless called an imprecise *Markov* chain has several reasons. The first is that it is a generalisation of a Markov chain. In particular, if $\mathscr{M}_\square = \{p_\square\}$ and, for all $n \in \mathbb{N}$, $\mathscr{T}_n = \{T_n\}$, then $\mathscr{P}^{\mathrm{M}}$ and $\mathscr{P}^{\mathrm{EI}}$ coincide and consist of a single stochastic process, which is the unique Markov chain that is parameterised by $p_\square$ and $(T_n)_{n\in\mathbb{N}}$. Second, while the probability trees $p_\bullet$ in $\mathscr{P}^{\mathrm{EI}}$ may not satisfy the Markov property, meaning that $p_{x_{1:n}}$ may depend on $x_{1:n-1}$, the constraints that are imposed on $p_{x_{1:n}}$ do satisfy a Markov property, in the sense that $\mathscr{T}_n$ does not depend on $x_{1:n-1}$. In this particular sense, $\mathscr{P}^{\mathrm{EI}}$—and $\mathscr{P}^{\mathrm{M}}$—satisfies an *imprecise Markov property*.

Example 2. *Consider again the disease transition model from Example 1, and the associated (homogeneous) Markov chain with transition matrix* T *and initial model* $p_\square$*. One simple way to capture parameter uncertainty about this model, of the form described above, is to consider sets of transition matrices* $(\mathscr{T}_n)_{n\in\mathbb{N}}$ *that are defined as neighbourhoods of* T*. This enables us to study the sensitivity of traditional inferences to changes in this matrix. So, for the remainder of this running example, for all* $n \in \mathbb{N}$*, we let*

$$\mathscr{T}_n := \mathscr{T} := \{(1-\varepsilon)T + \varepsilon T' \colon T' \in \mathbb{T}\}, \tag{3}$$

with $\varepsilon = 0.01$*. For the initial model we will simply consider the singleton set* $\mathscr{M}_\square := \{p_\square\}$*, meaning that we do not introduce any additional uncertainty there. The choice between the associated imprecise Markov chains* $\mathscr{P}^{\mathrm{EI}}$ *and* $\mathscr{P}^{\mathrm{M}}$ *now depends on whether we want to carry out the remaining analysis under Markovian assumptions or not.* ◇

[2] We take $\mathbb{N}$ to be the set of natural numbers *without* zero.

2.4 LOWER AND UPPER EXPECTATIONS

With any non-empty set of processes $\mathscr{P} \subseteq \mathbb{P}$—and hence also with $\mathscr{P}^{\mathrm{M}}$ and $\mathscr{P}^{\mathrm{EI}}$—we can associate a lower and upper expectation operator $\underline{E}$ and $\overline{E}$. For any $f \in \mathscr{G}(\mathscr{X}^n)$ and any $x_{1:m} \in \mathscr{X}^m$, the conditional lower expectation of $f(X_{1:n})$ with respect to $\mathscr{P}$ is given by

$$\underline{E}(f(X_{1:n})|X_{1:m}=x_{1:m}) := \inf_{p\in\mathscr{P}} E(f(X_{1:n})|X_{1:m}=x_{1:m}),$$

and similarly for $\overline{E}(f(X_{1:n})|X_{1:m}=x_{1:m})$, with sup instead of inf. It suffices to focus on either one of them though because they are related by the following conjugacy property:

$$\overline{E}(f(X_{1:n})|X_{1:m}=x_{1:m}) = -\underline{E}(-f(X_{1:n})|X_{1:m}=x_{1:m}).$$

We will therefore, without loss of generality, focus on lower expectations. Lower and upper probabilities can be defined analogously, as infima and suprema of probabilities. These too do not need to be dealt with seperately though, because they correspond to special cases of lower and upper expectations: for any $A \subseteq \mathscr{X}^n$ and $x_{1:m} \in \mathscr{X}^m$, it follows from Equation (2) that

$$\underline{P}(X_{1:n} \in A|X_{1:m}=x_{1:m}) = \underline{E}(\mathbb{I}_A(X_{1:n})|X_{1:m}=x_{1:m}), \quad (4)$$

and similarly for $\overline{P}(X_{1:n} \in A|X_{1:m}=x_{1:m})$.

For the particular cases where $\mathscr{P}=\mathscr{P}^{\mathrm{M}}$ or $\mathscr{P}=\mathscr{P}^{\mathrm{EI}}$, we will provide the corresponding lower and upper expectations and probabilities with the same superscript. For example, $\underline{E}^{\mathrm{M}}$ will denote the lower expectation of an imprecise Markov chain under complete independence, whereas $\overline{P}^{\mathrm{EI}}$ will denote the upper probability of an imprecise Markov chain under epistemic irrelevance.

The main goal of this paper is to demonstrate that there is a large class of functions for which the lower expections $\underline{E}^{\mathrm{M}}$ and $\underline{E}^{\mathrm{EI}}$—and hence also the corresponding upper expectations and lower and upper probabilities—can be efficiently computed. Furthermore, since the same algorithm will apply to both types of lower expectations, we also find that for that class of inferences, it does not matter whether we adopt complete independence or epistemic irrelevance—or equivalently, whether we assume Markovianity.

3 SUM-PRODUCT LAWS

The basic idea behind our algorithms is to recursively decompose the inference problems at hand into much smaller optimisation problems. In particular, we will reduce the problem to that of evaluating the following local operators. On the one hand, the lower expectation $\underline{E}_\Box$ with respect to the set $\mathscr{M}_\Box$ of initial probability mass functions, defined by

$$\underline{E}_\Box(f) := \inf_{p_\Box\in\mathscr{M}_\Box} \sum_{x\in\mathscr{X}} p_\Box(x)f(x) \text{ for all } f \in \mathscr{G}(\mathscr{X}).$$

On the other hand, for any $n \in \mathbb{N}$, the so-called lower transition operator $\underline{T}_n : \mathscr{G}(\mathscr{X}) \to \mathscr{G}(\mathscr{X})$, which maps any gamble $f \in \mathscr{G}(\mathscr{X})$ to a new gamble $\underline{T}_n f$ defined by

$$\underline{T}_n f(x) := \inf_{T\in\mathscr{T}_n} \sum_{y\in\mathscr{X}} T(x,y)f(y) \text{ for all } x \in \mathscr{X}. \quad (5)$$

It is easily seen that evaluating any of these local operators for a given $f \in \mathscr{G}(\mathscr{X})$ amounts to solving a constrained linear optimisation problem; for $\underline{E}_\Box(f)$ we need to solve a linear optimisation problem over the set $\mathscr{M}_\Box$, and for any $\underline{T}_n f$, $n \in \mathbb{N}$, we require $|\mathscr{X}|$ linear optimisations over $\mathscr{T}_n$—one for each $x \in \mathscr{X}$, to obtain $\underline{T}_n f(x)$. The complexity of solving these optimisation problems depends on the specification of $\mathscr{M}_\Box$ and $(\mathscr{T}_n)_{n\in\mathbb{N}}$; for example, if these sets are described using a finite number of linear (in)equality constraints, then it is well-known that the corresponding optimisation problems can be rewritten to linear programming problems that can then be solved using any of the available methods from the literature. However, specifics will be problem dependent, with some specifications allowing for straightforward and very efficient evaluation methods.

Example 3. *Consider again the parameters $\mathscr{M}_\Box$ and $(\mathscr{T}_n)_{n\in\mathbb{N}}$ from Example 2, and fix any $f \in \mathscr{G}(\mathscr{X})$. Clearly, evaluating $\underline{E}_\Box(f)$ is trivial since $\mathscr{M}_\Box = \{p_\Box\}$, so it follows that $\underline{E}_\Box(f) = \sum_{x\in\mathscr{X}} p_\Box(x)f(x)$. Moreover, for any $n \in \mathbb{N}$, it easily follows from Equation* (3) *that*

$$\underline{T}_n f = (1-\varepsilon)Tf + \varepsilon \min_{x\in\mathscr{X}} f(x).$$

Hence, evaluating $\underline{T}_n f$ has the same complexity as evaluating Tf in this case. ◇

To reduce our inferences to the evaluation of these simple operators, we will impose two conditions on a set of processes $\mathscr{P}$. The first one simply asks that

$$\underline{E}_\Box(f) = \underline{E}(f(X_1)) \text{ for all } f \in \mathscr{G}(\mathscr{X}). \quad (6)$$

Whenever this is the case, we say that $\mathscr{P}$ is *compatible* with $\mathscr{M}_\Box$. This is clearly true for both $\mathscr{P}^{\mathrm{M}}$ and $\mathscr{P}^{\mathrm{EI}}$. The second condition, which lies at the heart of our algorithms, is that $\mathscr{P}$ should satisfy what we call a sum-product law. Our first algorithm is based on the following simple version.

Definition 1. *A set of stochastic processes $\mathscr{P} \subseteq \mathbb{P}$ satisfies the* sum-product law for $(\mathscr{T}_n)_{n\in\mathbb{N}}$ *if for all $n \in \mathbb{N}$, $f \in \mathscr{G}(\mathscr{X})$ and $h, g \in \mathscr{G}(\mathscr{X}^n)$ with $h \geq 0$,*

$$\underline{E}\big(g(X_{1:n}) + h(X_{1:n})f(X_{n+1})\big) = \underline{E}\big(g(X_{1:n}) + h(X_{1:n})\underline{T}_n f(X_n)\big).$$

Crucially, $\mathscr{P}^{\mathrm{M}}$ and $\mathscr{P}^{\mathrm{EI}}$ both satisfy this law whenever $(\mathscr{T}_n)_{n\in\mathbb{N}}$ complies with the following fairly mild condition.

Proposition 1. *The imprecise Markov chains $\mathscr{P}^{\mathrm{M}}$ and $\mathscr{P}^{\mathrm{EI}}$ both satisfy the sum-product law for $(\mathscr{T}_n)_{n\in\mathbb{N}}$ if for all $n \in \mathbb{N}$, $f \in \mathscr{G}(\mathscr{X})$ and $\varepsilon > 0$,*

$$(\exists T_n \in \mathscr{T}_n)(\forall x \in \mathscr{X})\ T_n f(x) \leq \underline{T}_n f(x) + \varepsilon.$$

Proof. Fix any $n \in \mathbb{N}$ and let $\mathscr{P} := \mathscr{P}^{\mathrm{M}}$ or $\mathscr{P} := \mathscr{P}^{\mathrm{EI}}$—the proof works for both cases. Observe that for any $p_\bullet \in \mathscr{P}$,

$$\begin{aligned}
&E\big(h(X_{1:n})f(X_{n+1})\big) \\
&\quad= \sum_{z_{1:n+1}\in\mathscr{X}^{n+1}} \prod_{i=0}^{n} p_{z_{1:i}}(z_{i+1})h(z_{1:n})f(z_{n+1}) \\
&\quad= \sum_{z_{1:n}\in\mathscr{X}^{n}} \prod_{i=0}^{n-1} p_{z_{1:i}}(z_{i+1})h(z_{1:n}) \sum_{z_{n+1}\in\mathscr{X}} f(z_{n+1})p_{z_{1:n}}(z_{n+1}).
\end{aligned}$$

Furthermore, for all $z_{1:n} \in \mathscr{X}^n$, since the definition of $\mathscr{P}$ implies that $p_{z_{1:n}} = T(z_n,\cdot)$ for some $T \in \mathscr{T}_n$, we know that

$$\begin{aligned}
\sum_{z_{n+1}\in\mathscr{X}} f(z_{n+1})p_{z_{1:n}}(z_{n+1}) &= \sum_{z_{n+1}\in\mathscr{X}} f(z_{n+1})T(z_n,z_{n+1}) \\
&\geq \underline{T}_n f(z_n),
\end{aligned}$$

where the inequality follows from Equation (5). Since $p_{z_{1:i}}$ and h are non-negative, it therefore follows that

$$\begin{aligned}
E\big(h(X_{1:n})f(X_{n+1})\big) &\geq \sum_{z_{1:n}\in\mathscr{X}^{n}} \prod_{i=0}^{n-1} p_{z_{1:i}}(z_{i+1})h(z_{1:n})\underline{T}_n f(z_n) \\
&= E\big(h(X_{1:n})\underline{T}_n f(X_n)\big),
\end{aligned}$$

The linearity of E therefore implies that

$$\begin{aligned}
&E\big(g(X_{1:n}) + h(X_{1:n})f(X_{n+1})\big) \\
&\qquad\geq E\big(g(X_{1:n}) + h(X_{1:n})\underline{T}_n f(X_n)\big).
\end{aligned}$$

Since this is true for every $p_\bullet \in \mathscr{P}$, we find that

$$\begin{aligned}
&\underline{E}\big(g(X_{1:n}) + h(X_{1:n})f(X_{n+1})\big) \\
&\qquad\geq \underline{E}\big(g(X_{1:n}) + h(X_{1:n})\underline{T}_n f(X_n)\big).
\end{aligned}$$

We now establish the converse inequality. Fix any $\varepsilon > 0$. It then follows from the definition of $\underline{E}$ that there is a probability tree $p_\bullet \in \mathscr{P}$ such that

$$\begin{aligned}
&E\big(g(X_{1:n}) + h(X_{1:n})\underline{T}_n f(X_n)\big) \\
&\qquad\leq \underline{E}\big(g(X_{1:n}) + h(X_{1:n})\underline{T}_n f(X_n)\big) + \varepsilon. \quad (7)
\end{aligned}$$

Furthermore, by assumption, there is some $T_n \in \mathscr{T}_n$ such that, for all $x \in \mathscr{X}$, $T_n f(x) \leq \underline{T}_n f(x) + \varepsilon$. Consider now the probability tree $p^*_\bullet$ defined by

$$p^*_{z_{1:m}} := \begin{cases} T_n(z_n,\cdot) & \text{if } m = n \\ p_{z_{1:m}} & \text{if } m \neq n \end{cases} \qquad \text{for all } z_{1:m} \in \mathscr{X}^m. \quad (8)$$

Note that $p^*_\bullet \in \mathscr{P}$ because $p_\bullet \in \mathscr{P}$ and $T_n \in \mathscr{T}_n$. Let E^* be the expectation operator that corresponds to p^*. Then

$$\begin{aligned}
&E^*\big(h(X_{1:n})f(X_{n+1})\big) \\
&\quad= \sum_{z_{1:n}\in\mathscr{X}^{n}} \prod_{i=0}^{n-1} p^*_{z_{1:i}}(z_{i+1})h(z_{1:n}) \sum_{z_{n+1}\in\mathscr{X}} f(z_{n+1})p^*_{z_{1:n}}(z_{n+1}) \\
&\quad= \sum_{z_{1:n}\in\mathscr{X}^{n}} \prod_{i=0}^{n-1} p_{z_{1:i}}(z_{i+1})h(z_{1:n})T_n f(z_n) \\
&\quad\leq \sum_{z_{1:n}\in\mathscr{X}^{n}} \prod_{i=0}^{n-1} p_{z_{1:i}}(z_{i+1})h(z_{1:n})\big(\underline{T}_n f(z_n) + \varepsilon\big) \\
&\quad= E\big(h(X_{1:n})\underline{T}_n f(X_n)\big) + \varepsilon E\big(h(X_{1:n})\big)
\end{aligned}$$

and, using a similar but simpler derivation, also that $E^*\big(g(X_{1:n})\big) = E\big(g(X_{1:n})\big)$. Since E^* and E are both linear operators, and since $E\big(h(X_{1:n})\big) \leq \max h$, it follows that

$$\begin{aligned}
&E^*\big(g(X_{1:n}) + h(X_{1:n})f(X_{n+1})\big) \\
&\qquad\leq E\big(g(X_{1:n}) + h(X_{1:n})\underline{T}_n f(X_n)\big) + \varepsilon \max h \\
&\qquad\leq \underline{E}\big(g(X_{1:n}) + h(X_{1:n})\underline{T}_n f(X_n)\big) + \varepsilon + \varepsilon \max h,
\end{aligned}$$

using Equation (7) for the second inequality. Furthermore, since $p^*_\bullet \in \mathscr{P}$, we also have that $E^* \geq \underline{E}$. Thus, we find that

$$\begin{aligned}
&\underline{E}\big(g(X_{1:n}) + h(X_{1:n})f(X_{n+1})\big) \\
&\qquad\leq \underline{E}\big(g(X_{1:n}) + h(X_{1:n})\underline{T}_n f(X_n)\big) + \varepsilon + \varepsilon \max h.
\end{aligned}$$

Since this is true for any $\varepsilon > 0$, we obtain the converse inequality that we were after. □

What makes this condition nontrivial is the order of the quantifiers: for any $f \in \mathscr{G}(\mathscr{X})$, we require that there is a *single* T_n such that $T_n f(x) \approx \underline{T}_n f(x)$ *for all* $x \in \mathscr{X}$. One way to satisfy this is to enforce it with equality. A second case occurs if for all $n \in \mathbb{N}$, the set $\mathscr{T}_n$ of transition matrices has *separately specified rows* [Hermans and Škulj, 2014, Def. 11.6]. In essence, a set $\mathscr{T}$ of transition matrices has separately specified rows if the rows of the transition matrices T in $\mathscr{T}$ can be selected independently; more formally, this property requires that for any selection $(T_x)_{x\in\mathscr{X}}$ in $\mathscr{T}$, there is a $T \in \mathscr{T}$ such that $T(x,\cdot) = T_x(x,\cdot)$ for all $x \in \mathscr{X}$.

Example 4. *By Equation* (3)*, the sets* $\mathscr{T}_n$*,* $n \in \mathbb{N}$*, of our running example all have separately specified rows.* ◇

Our second algorithm uses the following stronger form of the sum-product law, where we replace $f(X_{n+1})$—a function of the 'next' state—by $f(X_n, X_{n+1})$—a function of the 'current' and 'next' state.

Definition 2. *A set of stochastic processes* $\mathscr{P} \subseteq \mathbb{P}$ *satisfies the* second-order sum-product law for $(\mathscr{T}_n)_{n\in\mathbb{N}}$ *if for all* $n \in \mathbb{N}$*,* $f \in \mathscr{G}(\mathscr{X}^2)$ *and* $h, g \in \mathscr{G}(\mathscr{X}^n)$ *with* $h \geq 0$*,*

$$\begin{aligned}
&\underline{E}\big(g(X_{1:n}) + h(X_{1:n})f(X_n, X_{n+1})\big) \\
&\qquad= \underline{E}\big(g(X_{1:n}) + h(X_{1:n})\underline{T}_n f(X_n)\big),
\end{aligned}$$

where, for all $x \in \mathscr{X}$*, we let* $\underline{T}_n f(x) := [\underline{T}_n(f(x,\cdot))](x)$*.*

Both $\mathscr{P}^{\mathrm{M}}$ and $\mathscr{P}^{\mathrm{EI}}$ again satisfy this law. But the condition that we need to impose is now a bit stronger.

Proposition 2. *The imprecise Markov chains* $\mathscr{P}^{\mathrm{M}}$ *and* $\mathscr{P}^{\mathrm{EI}}$ *both satisfy the second order sum-product law for* $(\mathscr{T}_n)_{n\in\mathbb{N}}$ *if for any* $n \in \mathbb{N}$*, any* $(f_x)_{x\in\mathscr{X}}$ *such that* $f_x \in \mathscr{G}(\mathscr{X})$ *for all* $x \in \mathscr{X}$*, and any* $\varepsilon > 0$*,*

$$(\exists T_n \in \mathscr{T}_n)(\forall x \in \mathscr{X})\ T_n f_x(x) \leq \underline{T}_n f_x(x) + \varepsilon.$$

Proof. Since the proof is extremely similar to that of Proposition 1, we only explain the steps that are different. For the first part of the proof, which consists in proving that

$$\underline{E}\big(g(X_{1:n})+h(X_{1:n})f(X_n,X_{n+1})\big) \geq \underline{E}\big(g(X_{1:n})+h(X_{1:n})\underline{T}_n f(X_n)\big),$$

the only difference is that $f(z_{n+1})$ is replaced by $f(z_n,z_{n+1})$ and that $\underline{T}_n f(z_n)$ should be interpreted as in Definition 2. The rest of the argument is completely analogous. For the second part of the proof, which establishes the converse inequality, the first difference is that $T_n \in \mathscr{T}_n$ now satisfies the stronger property that, for all $x \in \mathscr{X}$,

$$[T_n(f(x,\cdot))](x) \leq [\underline{T}_n(f(x,\cdot))](x)+\varepsilon = \underline{T}_n f(x)+\varepsilon.$$

The second difference is that the subsequent derivation changes accordingly. It now goes as follows:

$$\begin{aligned}
&E^*\big(h(X_{1:n})f(X_n,X_{n+1})\big)\\
&\quad= \sum_{z_{1:n}\in\mathscr{X}^n}\prod_{i=0}^{n-1}p^*_{z_{1:i}}(z_{i+1})h(z_{1:n}) \sum_{z_{n+1}\in\mathscr{X}} f(z_n,z_{n+1})p^*_{z_{1:n}}(z_{n+1})\\
&\quad= \sum_{z_{1:n}\in\mathscr{X}^n}\prod_{i=0}^{n-1}p_{z_{1:i}}(z_{i+1})h(z_{1:n})[T_n((f(z_n,\cdot))](z_n)\\
&\quad\leq \sum_{z_{1:n}\in\mathscr{X}^n}\prod_{i=0}^{n-1}p_{z_{1:i}}(z_{i+1})h(z_{1:n})\big(\underline{T}_n f(z_n)+\varepsilon\big)\\
&\quad= E\big(h(X_{1:n})\underline{T}_n f(X_n)\big)+\varepsilon E\big(h(X_{1:n})\big). \qquad \square
\end{aligned}$$

A simple way to enforce this condition is to require it with equality, without the ε. It is also relatively straightforward to show that $(\mathscr{T}_n)_{n\in\mathbb{N}}$ satisfies the condition in Proposition 2 whenever $\mathscr{T}_n$ has separately specified rows for all $n \in \mathbb{N}$.

4 RECURSIVE ALGORITHMS

With the required conditions out of the way, we now arrive at the two main results of this contribution. Both of them are simple and straightforward backwards recursive schemes to compute lower—and, by conjugacy, upper—expectations for sets of stochastic processes $\mathscr{P} \subseteq \mathbb{P}$ that satisfy a sum-product law for $(\mathscr{T}_n)_{n\in\mathbb{N}}$ and are compatible with $\mathscr{M}_\square$. As we have seen in the previous section, these conditions are satisfied by $\mathscr{P}^{\mathrm{M}}$ and $\mathscr{P}^{\mathrm{EI}}$, under fairly mild conditions. The essential property that makes the recursion possible is the sum-product law. The reason why we are able to exploit that property is because we restrict attention to variables $f(X_{1:n})$ that have a 'sum-product' decomposition. In the following result, and also in Theorem 2, we denote the set of non-negative gambles on $\mathscr{X}$ by $\mathscr{G}_{\geq 0}(\mathscr{X})$.

Theorem 1. *Consider a set of stochastic processes $\mathscr{P} \subseteq \mathbb{P}$ that satisfies the sum-product law for $(\mathscr{T}_n)_{n\in\mathbb{N}}$ and is compatible with $\mathscr{M}_\square$. Consider $g_1,\dots,g_n$ in $\mathscr{G}(\mathscr{X})$ and $h_1,\dots,h_{n-1}$ in $\mathscr{G}_{\geq 0}(\mathscr{X})$, and define $f \in \mathscr{G}(\mathscr{X}^n)$ by*

$$f(x_{1:n}) := \sum_{k=1}^{n} g_k(x_k)\prod_{\ell=1}^{k-1}h_\ell(x_\ell) \text{ for all } x_{1:n}\in\mathscr{X}^n.$$

Then $\underline{E}(f(X_{1:n})) = \underline{E}_\square(\underline{\pi}_1)$, where $\underline{\pi}_k \in \mathscr{G}(\mathscr{X})$ is recursively defined by the initial condition $\underline{\pi}_n := g_n$ and, for all $k \in \{1,\dots,n-1\}$, by $\underline{\pi}_k := g_k + h_k\underline{T}_k\underline{\pi}_{k+1}$.

Proof. For all $m \in \{0,\dots,n-1\}$, let

$$\tilde{h}_m(X_{1:m}) := \prod_{\ell=1}^{m}h_\ell(X_\ell)$$

and

$$\tilde{g}_m(X_{1:m}) := \sum_{k=1}^{m} g_k(X_k)\prod_{\ell=1}^{k-1}h_\ell(X_\ell).$$

Then $\tilde{h}_0 = 1$ and $\tilde{g}_0 = 0$. Since the compatibility with $\mathscr{M}_\square$ implies that $\underline{E}_\square(\underline{\pi}_1) = \underline{E}(\underline{\pi}_1(X_1))$, it suffices to prove that for all $m \in \{1,\dots,n\}$, and for $m = 1$ in particular,

$$\underline{E}(f(X_{1:n})) = \underline{E}\big(\tilde{g}_{m-1}(X_{1:m-1})+\tilde{h}_{m-1}(X_{1:m-1})\underline{\pi}_m(X_m)\big).$$

We provide a proof by induction. That this equation holds for $m = n$ is straightforward. Indeed, we see that

$$\begin{aligned}
\underline{E}(f(X_{1:n})) &= \underline{E}\big(\tilde{g}_{n-1}(X_{1:n-1})+\tilde{h}_{n-1}(X_{1:n-1})g_n(X_n)\big)\\
&= \underline{E}\big(\tilde{g}_{n-1}(X_{1:n-1})+\tilde{h}_{n-1}(X_{1:n-1})\underline{\pi}_n(X_n)\big).
\end{aligned}$$

For the inductive step, we assume that it holds for $m = i+1$, with $i \in \{1,\dots,n-1\}$, and prove that it then also holds for $m = i$. Observe that

$$\begin{aligned}
\underline{E}(f(X_{1:n})) &= \underline{E}\big(\tilde{g}_i(X_{1:i})+\tilde{h}_i(X_{1:i})\underline{\pi}_{i+1}(X_{i+1})\big)\\
&= \underline{E}\big(\tilde{g}_i(X_{1:i})+\tilde{h}_i(X_{1:i})\underline{T}_i\underline{\pi}_{i+1}(X_i)\big)\\
&= \underline{E}\big(\tilde{g}_{i-1}(X_{1:i-1})+\tilde{h}_{i-1}(X_{1:i-1})g_i(X_i)\\
&\qquad +\tilde{h}_{i-1}(X_{1:i-1})h_i(X_i)\underline{T}_i\underline{\pi}_{i+1}(X_i)\big)\\
&= \underline{E}\big(\tilde{g}_{i-1}(X_{1:i-1})\\
&\qquad +\tilde{h}_{i-1}(X_{1:i-1})\big[g_i(X_i)+h_i(X_i)\underline{T}_i\underline{\pi}_{i+1}(X_i)\big]\big)\\
&= \underline{E}\big(\tilde{g}_{i-1}(X_{1:i-1})+\tilde{h}_{i-1}(X_{1:i-1})\underline{\pi}_i(X_i)\big),
\end{aligned}$$

where the first equality follows from the induction hypothesis and the second follows from the sum-product law. Hence, the statement holds for $m = i$ as well. $\square$

Theorem 1 is related to—and, in fact, inspired by—a result of T'Joens et al. [2019, Thm. 2], but there are two key differences. The first is that T'Joens et al. [2019, § 3] limit themselves to 'homogeneous' imprecise Markov chains under epistemic irrelevance—that is, to $\mathscr{P}^{\mathrm{EI}}$ with $\mathscr{T}_n = \mathscr{T}_1$ for all $n \in \mathbb{N}$—whereas our results do not require homogeneity and also work for complete independence—that is, for $\mathscr{P}^{\mathrm{M}}$. The second is that they assume that each $\mathscr{T}_n$ has separately

specified rows, whereas our condition is weaker. That said, we do not get this more general setting for free: we need to require that the gambles $h_1, \ldots, h_{n-1}$ are non-negative, while T'Joens et al. [2019, Thm. 2] do not. This extra non-negativity condition is often satisfied in practice though, as we will see in Section 5.1 further on.

Our second main result is similar to the first one, but we now rely on the second-order sum-product law. More precisely, we show that we can also use a—slightly more involved—backwards recursive method to compute the lower expectation of inferences $f(X_{1:n})$ that have a 'second-order sum-product' decomposition; for an example of such an inference, we refer to Section 5.2 further on.

Theorem 2. *Consider a set of stochastic processes $\mathscr{P} \subseteq \mathbb{P}$ that satisfies the second-order sum-product law for $(\mathscr{T}_n)_{n\in\mathbb{N}}$ and is compatible with $\mathscr{M}_\Box$. Consider $g_2, \ldots, g_n$ in $\mathscr{G}(\mathscr{X}^2)$ and $h_1, \ldots, h_{n-1}$ in $\mathscr{G}_{\geq 0}(\mathscr{X})$, and define $f \in \mathscr{G}(\mathscr{X}^n)$ by*

$$f(x_{1:n}) := \sum_{k=2}^{n} g_k(x_{k-1}, x_k) \prod_{\ell=1}^{k-1} h_\ell(x_\ell) \text{ for all } x_{1:n} \in \mathscr{X}^n.$$

Then $\underline{E}(f(X_{1:n})) = \underline{E}_\Box(\underline{\pi}_1)$, where $\underline{\pi}_k \in \mathscr{G}(\mathscr{X})$ is recursively defined by the initial condition $\underline{\pi}_n := 0$ and, for all $k \in \{1, \ldots, n-1\}$ and $x_k \in \mathscr{X}$, by

$$\underline{\pi}_k(x) := h_k(x)\big[\underline{T}_k\big(g_{k+1}(x,\cdot) + \underline{\pi}_{k+1}\big)\big](x).$$

Proof. For all $m \in \{0, \ldots, n-1\}$, let

$$\tilde{h}_m(X_{1:m}) := \prod_{\ell=1}^{m} h_\ell(X_\ell)$$

and

$$\tilde{g}_m(X_{1:m}) := \sum_{k=2}^{m} g_k(X_{k-1}, X_k) \prod_{\ell=1}^{k-1} h_\ell(X_\ell).$$

Then $\tilde{h}_0 = 1$ and $\tilde{g}_0 = \tilde{g}_1 = 0$. Since the compatibility with $\mathscr{M}_\Box$ implies that $\underline{E}_\Box(\underline{\pi}_1) = \underline{E}(\underline{\pi}_1(X_1))$, it suffices to prove that for all $m \in \{1, \ldots, n\}$, and for $m = 1$ in particular,

$$\underline{E}(f(X_{1:n})) = \underline{E}\big(\tilde{g}_m(X_{1:m}) + \tilde{h}_{m-1}(X_{1:m-1})\underline{\pi}_m(X_m)\big). \quad (9)$$

We provide a proof by induction. That this equation holds for $m = n$ is straightforward. Indeed, we see that

$$\begin{aligned}\underline{E}(f(X_{1:n})) &= \underline{E}\big(\tilde{g}_n(X_{1:n}) + 0\big) \\ &= \underline{E}\big(\tilde{g}_n(X_{1:n}) + \tilde{h}_{n-1}(X_{1:n-1})\underline{\pi}_n(X_n)\big)\end{aligned}$$

For the inductive step, we assume that Equation (9) holds for $m = i+1$, with $i \in \{1, \ldots, n-1\}$, and will prove that it then also holds for $m = i$. Let $\tilde{f} \in \mathscr{G}(\mathscr{X}^2)$ be defined by

$$\tilde{f}(x,y) := g_{i+1}(x,y) + \underline{\pi}_{i+1}(y) \text{ for all } x, y \in \mathscr{X}.$$

We use here the notational convention of Definition 2, so

$$\underline{T}_i\tilde{f}(x) = \big[\underline{T}_i\big(\tilde{f}(x,\cdot)\big)\big](x) \text{ for all } x \in \mathscr{X}.$$

Thus, for all $x \in \mathscr{X}$, we find that

$$h_i(x)\underline{T}_i\tilde{f}(x) = h_i(x)\big[\underline{T}_i\big(g_{i+1}(x,\cdot) + \underline{\pi}_{i+1}\big)\big](x) = \underline{\pi}_i(x),$$

so, $h_i(X_i)\underline{T}_i\tilde{f}(X_i) = \underline{\pi}_i(X_i)$. It therefore follows that

$$\begin{aligned}\underline{E}(f(X_{1:n})) &= \underline{E}\big(\tilde{g}_{i+1}(X_{1:i+1}) + \tilde{h}_i(X_{1:i})\underline{\pi}_{i+1}(X_{i+1})\big) \\ &= \underline{E}\big(\tilde{g}_i(X_{1:i}) + \tilde{h}_i(X_{1:i})g_{i+1}(X_i, X_{i+1}) \\ &\qquad + \tilde{h}_i(X_{1:i})\underline{\pi}_{i+1}(X_{i+1})\big) \\ &= \underline{E}\big(\tilde{g}_i(X_{1:i}) + \tilde{h}_i(X_{1:i})\tilde{f}(X_i, X_{i+1})\big) \\ &= \underline{E}\big(\tilde{g}_i(X_{1:i}) + \tilde{h}_i(X_{1:i})\underline{T}_i\tilde{f}(X_i)\big) \\ &= \underline{E}\big(\tilde{g}_i(X_{1:i}) + \tilde{h}_{i-1}(X_{1:i-1})h_i(X_i)\underline{T}_i\tilde{f}(X_i)\big) \\ &= \underline{E}\big(\tilde{g}_i(X_{1:i}) + \tilde{h}_{i-1}(X_{1:i-1})\underline{\pi}_i(X_i)\big),\end{aligned}$$

where the first equality is the induction hypothesis and the fourth equality follows from the second-order sum-product law. Thus, Equation (9) holds for $m = i$, as required. □

Example 5. *Consider the simple disease transition model from Example 1, with the parameters $\mathscr{M}_\Box$ and $(\mathscr{T}_n)_{n\in\mathbb{N}}$ from Example 2. As was noted in Example 4, the sets $\mathscr{T}_n$, $n \in \mathbb{N}$, have separately specified rows, so the associated imprecise Markov chains $\mathscr{P}^{\mathrm{M}}$ and $\mathscr{P}^{\mathrm{EI}}$ satisfy both sum-product laws, which means that we can use the algorithms in Theorems 1 and 2 to compute lower expectations for them.* ◇

Let us briefly consider the computational complexity of the methods presented above. For the algorithm in Theorem 1, we see that the main effort lies in computing the $n-1$ functions $\underline{\pi}_k := g_k + h_k\underline{T}_k\underline{\pi}_{k+1}$, with $k \in \{1, \ldots, n-1\}$. For any such k, this requires us to evaluate $\underline{\pi}_k(x) := g_k(x) + h_k(x)\underline{T}_k\underline{\pi}_{k+1}(x)$ for all $x \in \mathscr{X}$, the complexity of which is dominated by the computation of $\underline{T}_k\underline{\pi}_{k+1}(x)$. As discussed in Section 3, the latter has a complexity that is problem dependent, say $O(c)$, where for simplicity we will assume that this complexity is independent of x and k. Repeating this for all $x \in \mathscr{X}$ to obtain the entire vector $\underline{\pi}_k$, we get a complexity of $O(c|\mathscr{X}|)$. Crucially, this complexity is independent of n, so the total complexity to obtain $\underline{\pi}_1$ by backwards recursion is simply $O(nc|\mathscr{X}|)$. The final step then requires evaluating $\underline{E}_\Box(\underline{\pi}_1)$, which as discussed in Section 3, has a problem dependent complexity, say $O(c_\Box)$. The total complexity of the algorithm in Theorem 1 therefore comes out to $O(c_\Box + nc|\mathscr{X}|)$, with the coefficients $c_\Box$ and c being determined by the specification of $\mathscr{M}_\Box$ and $(\mathscr{T}_k)_{k\in\mathbb{N}}$, respectively. The analysis of the algorithm in Theorem 2 proceeds similar to the above, and yields the exact same complexity. Notably, these complexities scale only *linearly* with the number of time points on which $\underline{E}(f(X_{1:n}))$ depends.

Example 6. *It follows from the analysis in Example 3 that for our simple disease transition model, the corresponding coefficients are $c_\Box = c = |\mathscr{X}|$. Hence, the total complexity of the algorithms from Theorem 1 and 2, when applied to inferences with this model, comes out to $O\big(n|\mathscr{X}|^2\big)$.* ◇

Finally, an interesting feature of the algorithms in Theorems 1 and 2 is that they do not depend on the specific set of stochastic processes $\mathscr{P}$. Indeed, observe that the final result is completely determined by $\underline{E}_\square$ and $(\underline{T}_n)_{n\in\mathbb{N}}$. This implies that whenever $(\mathscr{T}_n)_{n\in\mathbb{N}}$ satisfies the condition in Proposition 1 (or 2), the lower—and, by conjugacy, upper—expectations with respect to $\mathscr{P}^{\mathrm{M}}$ and $\mathscr{P}^{\mathrm{EI}}$ coincide for all inferences that have a 'sum-product' (or 'second-order sum-product') decomposition. As we are about to demonstrate, this class of inferences is surprisingly large.

5 SPECIAL CASES

Many practically relevant inferences are of the form $f(X_{1:n})$, with f as defined in Theorems 1 or 2. The following list provides various examples; it is by no means exhaustive, but does contain many common inferences. For any of these inferences, our theorems can be used to compute their lower *and* upper expectation. For the lower expectations, this is immediate. For the upper expectations, this follows from two observations: first, for any f of the required sum-product form, $-f$ is also of this form (simply replace g_k by $-g_k$); second, by conjugacy, $\overline{E}(f) = -\underline{E}(-f)$.

5.1 USING THE SUM-PRODUCT LAW

Let us start with inferences that are of the form as defined in Theorem 1. Recall from Section 4 that Theorem 1 is similar to [T'Joens et al., 2019, Theorem 2]. This is relevant because T'Joens et al. [2019, § 5] list five classes of inferences that fall in the scope of their recursive scheme. These five classes—or, in some cases, generalisations of them—also fall in the scope of Theorem 1.

Inferences that depend on a single time point The most basic example are inferences of the form $f(X_n)$, with $f \in \mathscr{G}(\mathscr{X})$. In Theorem 1, this corresponds to letting $g_n := f$ and, for all $k \in \{1,\ldots,n-1\}$, $g_k := 0$ and $h_k := 1$.

Example 7. *Consider the simple disease transition model from Example 1, with the parameters $\mathscr{M}_\square$ and $(\mathscr{T}_n)_{n\in\mathbb{N}}$ from Example 2. As we know from Example 5, we can use our algorithms to compute inferences for the associated imprecise Markov chains—as noted at the end of Section 4, the distinction between $\mathscr{P}^{\mathrm{M}}$ and $\mathscr{P}^{\mathrm{EI}}$ does not matter here. We will compute the lower and upper probability that person k is sick after n time steps. These are the lower and upper expectation of $\mathbb{I}_A(X_n)$, where $A \subseteq \mathscr{X}$ is the set of all states in which person k is sick. Since this inference depends on one time point, we know from the above that we can compute it with Theorem 1. The results are depicted in Figure 2, for different values of k; the dashed lines depict the probability interval and the solid line corresponds to the precise Markov chain with transition matrix T from Example 1.* ◇

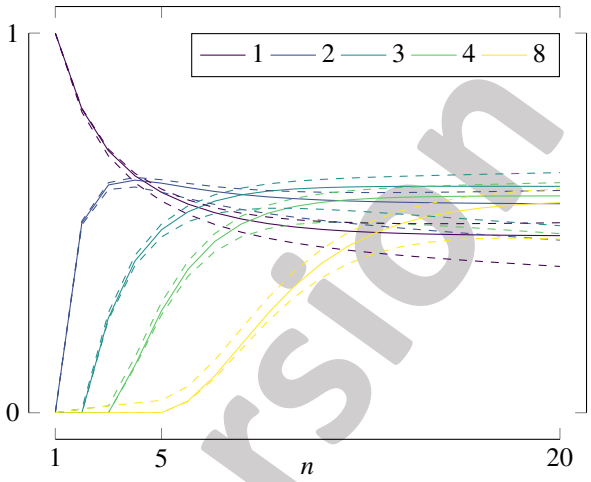

Figure 2: Probability of A Person Being Sick at Time n.

Pure sums A second, slightly more involved example are inferences of the form $\sum_{k=1}^{n} f_k(X_k)$, with $f_k \in \mathscr{G}(\mathscr{X})$ for all $k \in \{1,\ldots,n\}$. This corresponds to letting $g_k := f_k$ for all $k \in \{1,\ldots,n\}$ and $h_k := 1$ for all $k \in \{1,\ldots,n-1\}$ in Theorem 1. One important special case is the *occupancy time* $\sum_{k=1}^{n} \mathbb{I}_A(X_k)$ of the system in the set $A \subseteq \mathscr{X}$ over the finite horizon $\{1,\ldots,n\}$. A second important special case is the *temporal average* $\frac{1}{n}\sum_{k=1}^{n} f(X_k) = \sum_{k=1}^{n} \frac{1}{n} f(X_k)$ of $f(X_k)$ over the finite horizon $\{1,\ldots,n\}$. In the limit for n going to $+\infty$, T'Joens and De Bock [2021] discuss this inference in great depth for homogeneous imprecise Markov chains.

Pure products We can also deal with pure products instead of sums. Such inferences are of the form $\prod_{k=1}^{n} f_k(X_k)$, with $f_n \in \mathscr{G}(\mathscr{X})$ and $f_k \in \mathscr{G}_{\geq 0}(\mathscr{X})$ for all $k \in \{1,\ldots,n-1\}$. To deal with these, we invoke Theorem 1 with $g_n := f_n$ and, for all $k \in \{1,\ldots,n-1\}$, $g_k := 0$ and $h_k := f_k$.

Time-bounded until In the setting of model checking [Baier and Katoen, 2008], 'time-bounded until' events play an important role. Consider a time bound $n \in \mathbb{N}$, a set $G \subseteq \mathscr{X}$ of goal states and a set $S \subseteq \mathscr{X}$ of safe states. Katoen et al. [2012, § 2.2] define the corresponding 'time-bounded until' event as the event that the system visits a goal state in G at some time point k in the finite time horizon $\{1,\ldots,n\}$ while only visiting safe states in S at the time points before k; that is, they consider the event $\{S\mathscr{U}^{\leq n}G\} := \{X_{1:n} \in A\}$ with

$$A := \big\{x_{1:n} \in \mathscr{X}^n \colon \big(\exists k \in \{1,\ldots,n\}\big)\, x_k \in G, \quad (\forall \ell \in \{1,\ldots,k-1\})\, x_\ell \in S\big\}.$$

Note that $\mathbb{I}_A(X_{1:n}) = \sum_{k=1}^{n} \mathbb{I}_G(X_k) \prod_{\ell=1}^{k-1} \mathbb{I}_{S\cap G^{\mathrm{c}}}(X_\ell)$, so, due to Equation (4), we can compute the lower and upper probability of these 'time-bounded until' events with Theorem 1. When the set of safe states is $S = \mathscr{X}$, these are more commonly known as (time-bounded) *lower and upper hitting probabilities*; for homogenous imprecise Markov chais,

Krak et al. [2019] discuss the behaviour of these inferences in the limit for large n.

Example 8. *As a second example of using our disease model, we compute the probability that person* 6 *becomes sick within the first* $n = 40$ *time steps, without person* 4 *getting sick before him. This is a 'time-bounded until' event, where* G *collects the states where person* 6 *is sick and* S *the states where person* 4 *is not sick. So here as well, we can use Theorem 1 to compute the bounds that we are after. We find that—up to four significant digits—the lower probability is* 0.1374 *and the upper one is* 0.2553*; for the precise Markov chain that corresponds to* T*, the probability is* 0.1464. ◇

Hitting times Consider again a set $G \subseteq \mathscr{X}$ of goal states. Then for all $n \in \mathbb{N}$, we let τ_G^n be the gamble on $\mathscr{X}^n$ that is defined for all $x_{1:n} \in \mathscr{X}^n$ by

$$\tau_G^n(x_{1:n}) := \min\big(\{k \in \{1,\dots,n\} \colon x_k \in G\} \cup \{n+1\}\big).$$

Since $\tau_G^n(X_{1:n}) = 1 + \mathbb{I}_{G^c}(X_1) + \sum_{k=2}^n \mathbb{I}_{G^c}(X_k) \prod_{\ell=1}^{k-1} \mathbb{I}_{G^c}(X_\ell)$, it follows that we can use Theorem 1 to recursively compute $\underline{E}(\tau_G^n(X_{1:n}))$ and $\overline{E}(\tau_G^n(X_{1:n}))$. Krak et al. [2019] again discuss the limit for large n.

5.2 USING THE SECOND-ORDER SUM-PRODUCT LAW

Inferences that have a decomposition as in Theorem 2 but not as in Theorem 1 are perhaps a bit more difficult to conjure. We will treat one—yet important—example: the expected number of transitions of the system. To be as general as possible, we consider a subset A of $\mathscr{X}^2$, and we think of every couple $(x,y) \in A$ as a 'transition' that interests us; note that here, we do not exclude the case (x,x). Then for all $n \in \mathbb{N}$, we let $\eta_A^{\leq n}(X_{1:n})$ denote the number of 'interesting transitions' over the finite horizon $\{1,\dots,n\}$; that is, $\eta_A^{\leq n}$ is the gamble on $\mathscr{X}^n$ defined for all $x_{1:n} \in \mathscr{X}^n$ by

$$\eta_A^{\leq n}(x_{1:n}) := |\{k \in \{2,\dots,n\} \colon (x_{k-1}, x_k) \in A\}|.$$

It is quite easy to see that $\eta_A^{\leq n}(X_{1:n}) = \sum_{k=2}^n \mathbb{I}_A(X_{k-1}, X_k)$, which makes clear that we can use Theorem 2 to determine the lower (and conjugate upper) expectation of $\eta_A^{\leq n}(X_{1:n})$. An important example of a set $A \subseteq \mathscr{X}^2$ of 'interesting transitions' is $A = \{(x,y) \in \mathscr{X}^2 \colon x \neq y\}$. In that case, $\eta_A^{\leq n}(X_{1:n})$ corresponds to the number of transitions over $\{1,\dots,n\}$. Alternatively, if $A = S \times G$, with $G \subseteq \mathscr{X}$ a set of goal states and $S \subseteq \mathscr{X}$ a set of safe states, then $\eta_A^{\leq n}(X_{1:n})$ corresponds to the number of transitions from a safe state in S to a goal state in G over $\{1,\dots,n\}$.

Example 9. *As a final example that uses our simple disease transition model, in order to illustrate the algorithm from Theorem 2, we consider the expected number of times that person 3 becomes sick in the first 60 time steps. This is a specific case of the expected number of transitions, as discussed above, with* $n = 60$ *and* A *the set of all couples* (x,y) *where person 3 is healthy in state* x *and sick in state* y*. We find that—up to four significant digits—the lower expected number of times is* 5.455 *and the upper is* 8.093*; for the precise Markov chain that corresponds to* T*, the expected number of times is* 7.212. ◇

6 CONCLUSION

The main conclusion of this contribution is that there is a diverse range of practically relevant inferences that can be efficiently computed for imprecise Markov chains, and that for those inferences, it does not matter whether we adopt complete independence or epistemic irrelevance. To obtain tight bounds, mild conditions had to be imposed on the sets of transition matrices $(\mathscr{T}_n)_{n\in\mathbb{N}}$. Note however that we can drop these conditions at the cost of the tightness of our bounds. This can be seen by noticing that, for every $n \in \mathbb{N}$,

$$\mathscr{T}_n' := \{T \in \mathbb{T} \colon Tf \geq \underline{T}_n f \text{ for all } f \in \mathscr{G}(\mathscr{X})\}$$

is a superset of $\mathscr{T}_n$ whose lower transition operator is also equal to $\underline{T}_n$, but which always satisfies the conditions in Propositions 1 and 2. Our algorithms therefore yield tight bounds for the imprecise Markov chains corresponding to $(\mathscr{T}_n')_{n\in\mathbb{N}}$, and hence also—possibly non-tight—bounds for the ones corresponding to $(\mathscr{T}_n)_{n\in\mathbb{N}}$.

We see several future lines of research. First, to drop the restriction that the multipliers h_ℓ should be non-negative. Second, to study the limit behaviour for large n for the new types of inferences that we can now deal with. Third, to consider more special cases, besides the ones we discussed in Section 5. Fourth, to explore the implications of our results for Markov decision processes [Feinberg and Shwartz, 2012], by using the close relationship between Markov decision processes and imprecise Markov chains [Troffaes and Skulj, 2013]. And fifth, to test our methods on more examples, and to apply them to real-life problems.

Author Contributions

Jasper De Bock conceived the initial idea. Thomas Krak created the code for the example. Thomas Krak and Alexander Erreygers ran the experiments. Alexander Erreygers created the figures. All authors contributed to the writing, and took part in discussions that shaped the final form of the results.

Acknowledgements

The research of Jasper De Bock and Alexander Erreygers was funded by project number 3GO28919 of the FWO (Research Foundation - Flanders). We thank the anonymous reviewers for their constructive feedback.

References

Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. MIT Press, 2008.

Fabio G. Cozman. Sets of probability distributions, independence, and convexity. *Synthese*, 186(2):577–600, 2012.

Gert de Cooman and Filip Hermans. Imprecise probability trees: Bridging two theories of imprecise probability. *Artificial Intelligence*, 172(11):1400–1427, July 2008.

Eugene A. Feinberg and Adam Shwartz. *Handbook of Markov decision processes: methods and applications*, volume 40. Springer Science & Business Media, 2012.

Filip Hermans and Damjan Škulj. Stochastic processes. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, chapter 11. Wiley, 2014.

Joost-Pieter Katoen, Daniel Klink, Martin Leucker, and Verena Wolf. Three-valued abstraction for probabilistic systems. *The Journal of Logic and Algebraic Programming*, 81(4):356–389, 2012. doi: 10.1016/j.jlap.2012.03.007.

Thomas Krak, Natan T'Joens, and Jasper De Bock. Hitting times and probabilities for imprecise markov chains. In *Proceedings of ISIPTA 2019*, volume 103 of *Proceedings of Machine Learning Research*, pages 265–275. PMLR, 2019.

Damjan Škulj. Discrete time Markov chains with interval probabilities. *International Journal of Approximate Reasoning*, 50:1314–1329, 2009.

Natan T'Joens and Jasper De Bock. Average behaviour in discrete-time imprecise Markov chains: A study of weak ergodicity, 2021. arXiv:2102.04793 [math.PR].

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.

Matthias Troffaes and Damjan Skulj. Model checking for imprecise Markov chains. Society for Imprecise Probability: Theories and Applications (SIPTA), 2013.