# A Pointwise Ergodic Theorem for Imprecise Markov Chains

**Gert de Cooman** and **Jasper De Bock** and **Stavros Lopatatzidis**
Ghent University, SYSTeMS Research Group
{gert.decooman,jasper.debock,stavros.lopatatzidis}@UGent.be

## Abstract

We prove a game-theoretic version of the strong law of large numbers for submartingale differences, and use this to derive a pointwise ergodic theorem for discrete-time Markov chains with finite state sets, when the transition probabilities are imprecise, in the sense that they are only known to belong to some convex closed set of probability measures.

**Keywords.** Imprecise probabilities, lower expectation, pointwise ergodic theorem, imprecise Markov chain, game-theoretic probability

## 1 Introduction

In Ref. [2], de Cooman and Hermans made a first attempt at laying the foundations for a theory of discrete-event (and discrete-time) stochastic processes that are governed by sets of, rather than single, probability measures. They showed how this could be done by connecting Walley's [1991] theory of coherent lower previsions with ideas and results from Shafer and Vovk's [2001] game-theoretic approach to probability theory. In later papers, de Cooman et al. [5] applied these ideas to finite-state discrete-time Markov chains, inspired by the work of Hartfiel [6]. They showed how to do efficient inferences in, and proved a Perron–Frobenius-like theorem for, so-called imprecise Markov chains, which are finite-state discrete-time Markov chains whose transition probabilities are imprecise, in the sense that they are only known to belong to a convex closed set of probability measures—typically due to partial assessments involving probabilistic inequalities. This work was later refined and extended by Hermans and de Cooman [7] and Škulj and Hable [15].

The Perron–Frobenius-like theorems in these papers give equivalent necessary and sufficient conditions for the uncertainty model—a set of probabilities—about the state $X_n$ to converge, for $n \to +\infty$, to an uncertainty model that is independent of the uncertainty model for the initial state $X_1$.

In Markov chains with 'precise' transition probabilities, this convergence behaviour is sufficient for a pointwise ergodic theorem to hold, namely that:

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} f(X_k) = E_\infty(f) \text{ almost surely}$$

for all real functions $f$ on the finite state set $\mathscr{X}$, where $E_\infty$ is the limit expectation operator that the expectation operators $E_n$ for the state $X_n$ at time $n$ converge to pointwise, independently of the initial model $E_1$ for $X_1$, according to the classical Perron–Frobenius Theorem.[1]

The aim of the present paper is to extend this result to a version for imprecise Markov chains; see Theorem 11.

How do we mean to go about this? In Section 2, we explain what we mean by imprecise probability models: we extend the notion of an expectation operator to so-called lower (and upper) expectation operators, and explain how these can be associated with (convex and closed) sets of expectation operators.

In Section 3, we explain how these generalised uncertainty models can be combined with event trees to form so-called imprecise probability trees, to produce a simple theory of discrete-time stochastic processes. We show in particular how to combine local uncertainty models associated with the nodes in the tree into global uncertainty models (global conditional lower expectations) about the paths in the tree, and how this procedure is related to sub- and supermartingales. We also indicate how it extends and subsumes the (precise-)probabilistic approach.

In Section 4 we prove a very general strong law of large numbers for submartingale differences in our imprecise probability trees. Our pointwise ergodic theorem will turn out to be a consequence of this in the particular context of imprecise Markov chains. We briefly explain what imprecise Markov chains are in Section 5: how they are special

[1] Actually, much more general results can be proved, for functions $f$ that do not depend on a single state only, but on the entire sequence of states; see for instance Ref. [8, Chapter 20]. In this paper, we will focus on the simpler version.

cases of imprecise probability trees, how to do efficient inference for them, and how to define Perron–Frobenius-like behaviour. We also explore the influence of time shifts on the global (conditional) lower expectations, and discuss stationarity and its relation with Perron–Frobenius-like behaviour.

In Section 6 we show that there is an interesting identity between the time averages that appear in our strong law of large numbers, and the ones that appear in the pointwise ergodic theorem. The discussion in Section 7 first focusses on a number of terms in this identity, and investigates the convergence of these terms for Perron–Frobenius-like imprecise Markov chains. This allows us to use the identity to prove our version of the pointwise ergodic theorem, whose significance we discuss briefly in Section 8.

## 2 Basic Notions from Imprecise Probabilities

Let us begin with a brief sketch of a few basic definitions and results about imprecise probabilities. For more details, we refer to Walley's [16] seminal book, as well as more recent textbooks [1, 13].

Suppose a subject is uncertain about the value that a variable $Y$ assumes in a non-empty set of possible values $\mathscr{Y}$. He is therefore also uncertain about the value $f(Y)$ a so-called *gamble*—a bounded real-valued function—$f\colon \mathscr{Y} \to \mathbb{R}$ on the set $\mathscr{Y}$ assumes in $\mathbb{R}$. We will also call such an $f$ a gamble *on $Y$* when we want to make explicit what variable $Y$ the gamble $f$ is intended to depend on. The subject's uncertainty is modelled by a *lower expectation*[2] $\underline{E}$, which is a real functional defined on the set $\mathscr{G}(\mathscr{Y})$ of all gambles on the set $\mathscr{Y}$, satisfying the following basic so-called *coherence axioms*:

LE1. $\underline{E}(f) \geq \inf f$ for all $f \in \mathscr{G}(\mathscr{Y})$;

LE2. $\underline{E}(f+g) \geq \underline{E}(f)+\underline{E}(g)$ for all $f,g \in \mathscr{G}(\mathscr{Y})$;

LE3. $\underline{E}(\lambda f) = \lambda \underline{E}(f)$ for all $f \in \mathscr{G}(\mathscr{Y})$ and real $\lambda \geq 0$.

One—but by no means the only[3]—way to interpret $\underline{E}(f)$ is as a lower bound on the expectation $E(f)$ of the gamble $f(Y)$. The corresponding upper bounds are given by the *conjugate upper expectation* $\overline{E}$, defined by $\overline{E}(f) := -\underline{E}(-f)$ for all $f \in \mathscr{G}(\mathscr{Y})$. It follows from the coherence axioms LE1–LE3 that

LE4. $\inf f \leq \underline{E}(f) \leq \overline{E}(f) \leq \sup f$ for all $f \in \mathscr{G}(\mathscr{Y})$;

LE5. $\underline{E}(f) \leq \underline{E}(g)$ and $\overline{E}(f) \leq \overline{E}(g)$ for all $f,g \in \mathscr{G}(\mathscr{Y})$ with $f \leq g$;

LE6. $\underline{E}(f+\mu) = \underline{E}(f)+\mu$ and $\overline{E}(f+\mu) = \overline{E}(f)+\mu$ for all $f \in \mathscr{G}(\mathscr{Y})$ and real $\mu$.

Lower and upper expectations will be the basic uncertainty models we consider in this paper.

The *indicator* $\mathbb{I}_A$ of an *event* $A$—a subset of $\mathscr{Y}$—is the gamble on $Y$ that assumes the value 1 on $A$ and 0 outside $A$. It allows us to introduce the *lower* and *upper probabilities* of $A$ as $\underline{P}(A) := \underline{E}(\mathbb{I}_A)$ and $\overline{P}(A) := \overline{E}(\mathbb{I}_A)$, respectively. They can be seen as lower and upper bounds on the probability $P(A)$ of $A$, and satisfy the conjugacy relation $\overline{P}(A) = 1 - \underline{P}(\mathscr{Y} \setminus A)$.

When the lower bound $\underline{E}$ coincides with the upper bound $\overline{E}$, the resulting functional $E := \underline{E} = \overline{E}$ satisfies the defining axioms of an *expectation*:

E1. $E(f) \geq \inf f$ for all $f \in \mathscr{G}(\mathscr{Y})$;

E2. $E(f+g) = E(f)+E(g)$ for all $f,g \in \mathscr{G}(\mathscr{Y})$;

E3. $E(\lambda f) = \lambda E(f)$ for all $f \in \mathscr{G}(\mathscr{Y})$ and real $\lambda$.

When $\mathscr{Y}$ is finite, $E$ is trivially the expectation associated with a (probability) mass function $p$ defined by $p(y) := \underline{P}(\{y\}) = \overline{P}(\{y\})$ for all $y \in \mathscr{Y}$, because it follows from the expectation axioms that then $E(f) = \sum_{y \in \mathscr{Y}} f(y)p(y)$; see for instance also the detailed discussion in Ref. [13].

With any lower expectation $\underline{E}$, we can always associate the following convex and closed[4] set of *compatible* expectation operators:

$$\mathfrak{M}(\underline{E}) := \{E\colon (\forall f \in \mathscr{G}(\mathscr{Y}))\underline{E}(f) \leq E(f) \leq \overline{E}(f)\}, \quad (1)$$

and the properties LE1–LE3 then guarantee that

$$\left.\begin{aligned} \underline{E}(f) &= \min\{E(f)\colon E \in \mathfrak{M}(\underline{E})\} \\ \overline{E}(f) &= \max\{E(f)\colon E \in \mathfrak{M}(\underline{E})\} \end{aligned}\right\} \text{ for all } f \in \mathscr{G}(\mathscr{Y}). \quad (2)$$

In this sense, an imprecise probability model $\underline{E}$ can always be identified with a closed convex set $\mathfrak{M}(\underline{E})$ of compatible 'precise' probability models $E$.

## 3 Discrete-Time Finite-State Imprecise Stochastic Processes

We consider a discrete-time process as a sequence of variables, henceforth called *states*, $X_1, X_2, \dots, X_n, \dots$, where each state $X_k$ is assumed to take values in a non-empty *finite* set $\mathscr{X}_k$.

[2] In the literature [16, 1, 13], other names, such as coherent lower expectation, or coherent lower prevision, have also been given to this concept.

[3] See Refs. [16, 10, 13] for other interpretations.

[4] The 'closedness' is associated with the weak* topology of pointwise convergence [16, Section 3.6].

### 3.1 Event Trees, Situations, Paths and Cuts

We will use, for any natural $k \leq \ell$, the notation $X_{k:\ell}$ for the tuple $(X_k, \dots, X_\ell)$, which can be seen as a variable assumed to take values in the product set $\mathscr{X}_{k:\ell} := \times_{r=k}^{\ell} \mathscr{X}_r$. We denote the set of all natural numbers (without 0) by $\mathbb{N}$, and let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

We call any $x_{1:n} \in \mathscr{X}_{1:n}$ for $n \in \mathbb{N}_0$ a *situation* and we denote the set of all situations by $\Omega^\Diamond$. So any situation is a finite string of possible values for the consecutive states, and if we denote the empty string by $\square$, then in particular, $\mathscr{X}_{1:0} = \{\square\}$. $\square$ is called the *initial situation*. We also use the generic notations $s$, $t$ or $u$ for situations.

An infinite sequence of state values is called a *path*, and we denote the set of all paths—also called the *sample space*—by $\Omega$. Hence

$$\Omega^\Diamond := \bigcup_{n \in \mathbb{N}_0} \mathscr{X}_{1:n} \text{ and } \Omega := \times_{r=1}^{\infty} \mathscr{X}_k.$$

We will denote generic paths by $\omega$. For any path $\omega \in \Omega$, the initial sequence that consists of its first $n$ elements is a situation in $\mathscr{X}_{1:n}$ that is denoted by $\omega^n$. Its $n$-th element belongs to $\mathscr{X}_n$ and is denoted by $\omega_n$. As a convention, we let its 0-th element be the initial situation $\omega^0 = \omega_0 = \square$. The possible realisations $\omega$ of a process can be represented graphically as paths in a so-called *event tree*, where each node is a situation; see Figure 1.

We write that $s \sqsubseteq t$, and say that $s$ *precedes* $t$ or that $t$ *follows* $s$, when every path that goes through $t$ also goes through $s$. The binary relation $\sqsubseteq$ is a partial order, and we write $s \sqsubset t$ whenever $s \sqsubseteq t$ but not $s = t$. We say that $s$ and $t$ are *incomparable* when neither $s \sqsubseteq t$ nor $t \sqsubseteq s$.

A (partial) *cut* $U$ is a collection of mutually incomparable situations, and represents a stopping time. For any two cuts $U$ and $V$, we define the following sets of situations:

$$\begin{aligned}
[U,V] &:= \{s \in \Omega^\Diamond \colon (\exists u \in U)(\exists v \in V) u \sqsubseteq s \sqsubseteq v\} \\
[U,V) &:= \{s \in \Omega^\Diamond \colon (\exists u \in U)(\exists v \in V) u \sqsubseteq s \sqsubset v\} \\
(U,V] &:= \{s \in \Omega^\Diamond \colon (\exists u \in U)(\exists v \in V) u \sqsubset s \sqsubseteq v\} \\
(U,V) &:= \{s \in \Omega^\Diamond \colon (\exists u \in U)(\exists v \in V) u \sqsubset s \sqsubset v\}.
\end{aligned}$$

When a cut $U$ consists of a single element $u$, then we will identify $U = \{u\}$ and $u$. This slight abuse of notation will for instance allow us to write $[u,v] = \{s \in \Omega^\Diamond \colon u \sqsubseteq s \sqsubseteq v\}$ and also $(U,v) = \{s \in \Omega^\Diamond \colon (\exists u \in U) u \sqsubset s \sqsubset v\}$. We also write $U \sqsubset V$ if $(\forall v \in V)(\exists u \in U) u \sqsubset v$. Observe that in that case $U \cap V = \emptyset$. In particular, $s \sqsupset U$ when there is some $u \in U$ such that $s \sqsupset u$, or in other words if $[U,s) \neq \emptyset$.

A *process* $\mathscr{F}$ is a map defined on $\Omega^\Diamond$. A *real process* is a real-valued process: it associates a real number $\mathscr{F}(x_{1:n}) \in \mathbb{R}$ with any situation $x_{1:n}$. It is called *bounded below* if there is some real $B$ such that $\mathscr{F}(s) \geq B$ for all situations $s \in \Omega^\Diamond$.

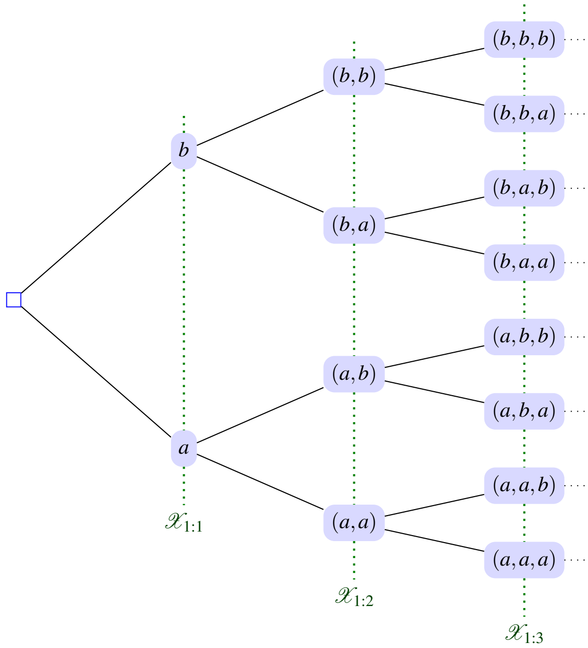


Figure 1: The (initial part of the) event tree for a process whose states can assume two values, $a$ and $b$, and can change at time instants $n = 1, 2, 3, \dots$ Each node in the tree corresponds to a situation. Also depicted are the respective sets of situations (cuts) $\mathscr{X}_{1:1}$, $\mathscr{X}_{1:2}$ and $\mathscr{X}_{1:3}$ where the states at times 1, 2 and 3 are revealed.

A *gamble process* $\mathscr{D}$ is a process that associates with any situation $x_{1:n}$ a gamble $\mathscr{D}(x_{1:n}) \in \mathscr{G}(\mathscr{X}_{n+1})$ on $X_{n+1}$. It is called *uniformly bounded* if there is some real $B$ such that $|\mathscr{D}(s)| \leq B$ for all situations $s \in \Omega^\Diamond$. With any real process $\mathscr{F}$, we can always associate a gamble process $\Delta\mathscr{F}$, called the *process difference*. For every situation $x_{1:n}$, the gamble $\Delta\mathscr{F}(x_{1:n}) \in \mathscr{G}(\mathscr{X}_{n+1})$ is defined by[5]

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

We will denote this more succinctly by $\Delta\mathscr{F}(x_{1:n}) = \mathscr{F}(x_{1:n}\,\cdot) - \mathscr{F}(x_{1:n})$, where the '$\cdot$' represents the generic value of the next state $X_{n+1}$.

Conversely, with a gamble process $\mathscr{D}$, we can associate a real process $\mathscr{I}^{\mathscr{D}}$, defined by

$$\mathscr{I}^{\mathscr{D}}(x_{1:n}) := \sum_{k=0}^{n-1} \mathscr{D}(x_{1:k})(x_{k+1}) \quad \text{for all } n \in \mathbb{N}_0 \text{ and } x_{1:n} \in \mathscr{X}_{1:n}.$$

Clearly, $\Delta\mathscr{I}^{\mathscr{D}} = \mathscr{D}$ and $\mathscr{F} = \mathscr{F}(\square) + \mathscr{I}^{\Delta\mathscr{F}}$.

[5]Our assumption that $\mathscr{X}_{n+1}$ is finite is crucial here because it guarantees that $\Delta\mathscr{F}(x_{1:n})$ is bounded, which in turn implies that it is indeed a gamble.

Also, with any real process $\mathscr{F}$ we can associate the *path-averaged process* $\langle\mathscr{F}\rangle$, which is the real process defined by:

$$\langle\mathscr{F}\rangle(x_{1:n}) := \begin{cases} 0 & \text{if } n=0 \\ \frac{1}{n}\mathscr{F}(x_{1:n}) & \text{if } n>0 \end{cases} \quad \text{for all } n \in \mathbb{N}_0 \text{ and } x_{1:n} \in \mathscr{X}_{1:n}.$$

### 3.2 Imprecise Probability Trees, Submartingales and Supermartingales

The standard way to turn an event tree into a *probability tree* is to attach to each of its nodes, or situations $x_{1:n}$, a *local* probability model $Q(\cdot|x_{1:n})$ for what will happen immediately afterwards, i.e. for the value that the next state $X_{n+1}$ will assume in $\mathscr{X}_{n+1}$. This local model $Q(\cdot|x_{1:n})$ is then an expectation operator on the set $\mathscr{G}(\mathscr{X}_{n+1})$ of all gambles $g(X_{n+1})$ on the next state $X_{n+1}$, conditional on observing $X_{1:n} = x_{1:n}$.

In a completely similar way, we can turn an event tree into an *imprecise probability tree* by attaching to each of its situations $x_{1:n}$ a local *imprecise* probability model $\underline{Q}(\cdot|x_{1:n})$ for what will happen immediately afterwards, i.e. for the value that the next state $X_{n+1}$ will assume in $\mathscr{X}_{n+1}$. This local model $\underline{Q}(\cdot|x_{1:n})$ is then a *lower* expectation operator on the set $\mathscr{G}(\mathscr{X}_{n+1})$ of all gambles $g(X_{n+1})$ on the next state $X_{n+1}$, conditional on observing $X_{1:n} = x_{1:n}$. This is represented graphically in Figure 2.

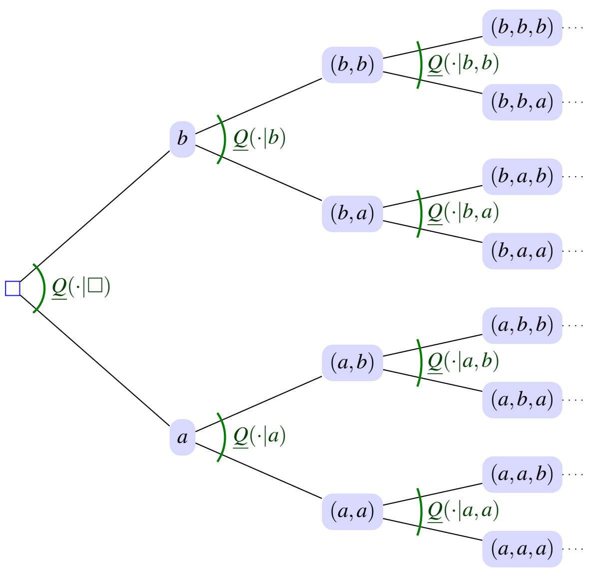


Figure 2: The (initial part of the) imprecise probability tree for a process whose states can assume two values, $a$ and $b$, and can change at time instants $n = 1, 2, 3, \ldots$

In a given imprecise probability tree, a *submartingale* $\mathscr{M}$ is a real process such that $\underline{Q}(\Delta\mathscr{M}(x_{1:n})|x_{1:n}) \geq 0$ for all $n \in \mathbb{N}_0$ and $x_{1:n} \in \mathscr{X}_{1:n}$: all submartingale differences have non-negative lower expectation. A real process $\mathscr{M}$ is a *supermartingale* if $-\mathscr{M}$ is a submartingale, meaning that $\overline{Q}(\Delta\mathscr{M}(x_{1:n})|x_{1:n}) \leq 0$ for all $n \in \mathbb{N}_0$ and $x_{1:n} \in \mathscr{X}_{1:n}$: all supermartingale differences have non-positive upper expectation. We denote the set of all submartingales for a given imprecise probability tree by $\underline{\mathbb{M}}$—whether a real process is a submartingale depends of course on the local uncertainty models. Similarly, the set $\overline{\mathbb{M}} := -\underline{\mathbb{M}}$ is the set of all supermartingales.

In the present context of probability trees, we will also call *variable* any function defined on the so-called *sample space*—the set $\Omega$ of all paths. When this variable is real-valued and bounded, we will also call it a *gamble* on $\Omega$. When it is extended real-valued, meaning that it assumes values in the set $\mathbb{R}^* := \mathbb{R} \cup \{-\infty, +\infty\}$, we call in an *extended real variable*. An *event* $A$ in this context is a subset of $\Omega$, and its indicator $\mathbb{I}_A$ is a gamble on $\Omega$ assuming the value 1 on $A$ and 0 elsewhere. With any situation $x_{1:n}$, we can associated the so-called *exact event* $\Gamma(x_{1:n})$ that $X_{1:n} = x_{1:n}$, which is the set of all paths $\omega \in \Omega$ that go through $x_{1:n}$:

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

For a given $n \in \mathbb{N}_0$, we call a variable $\xi$ *n-measurable* if it is constant on the exact events $\Gamma(x_{1:n})$ for all $x_{1:n} \in \mathscr{X}_{1:n}$, or in other words, if it only depends on the values of the first $n$ states $X_{1:n}$. We then use the obvious notation $\xi(x_{1:n})$ for its constant value $\xi(\omega)$ on all paths $\omega$ in $\Gamma(x_{1:n})$.

With a real process $\mathscr{F}$, we can associate in particular the following extended real variables $\liminf\mathscr{F}$ and $\limsup\mathscr{F}$, defined for all $\omega \in \Omega$ by $\liminf\mathscr{F}(\omega) := \liminf_{n\to\infty}\mathscr{F}(\omega^n)$ and $\limsup\mathscr{F}(\omega) := \limsup_{n\to\infty}\mathscr{F}(\omega^n)$. If $\liminf\mathscr{F}(\omega) = \limsup\mathscr{F}(\omega)$ on some path $\omega$, then we also denote the common value there by $\lim\mathscr{F}(\omega) = \lim_{n\to\infty}\mathscr{F}(\omega^n)$.

### 3.3 Going from Local to Global Belief Models

So far, we have associated local uncertainty models with an imprecise probability tree. These represent, in any situation $x_{1:n}$, beliefs about what will happen immediately afterwards, or in other words about the step from $x_{1:n}$ to $x_{1:n}X_{n+1}$.

We now want to turn these local models into global ones: uncertainty models about which entire path $\omega$ is taken in the event tree, rather than which local steps are taken from one situation to the next. We use the following expression for the global lower expectation conditional on the situation $s$:

$$\underline{E}(f|s) := \sup\{\mathscr{M}(s) \colon \mathscr{M} \in \underline{\mathbb{M}}, \limsup\mathscr{M} \leq f \text{ on } \Gamma(s)\}, \tag{3}$$

and for the conjugate global upper expectation conditional on the situation $s$:

$$\overline{E}(f|s) := \inf\{\mathscr{M}(s) \colon \mathscr{M} \in \overline{\mathbb{M}}, \liminf\mathscr{M} \geq f \text{ on } \Gamma(s)\}, \tag{4}$$

where $f$ is any extented real variable on $\Omega$, and $s \in \Omega^\Diamond$ any situation. We use the simplified notations $\underline{E} = \underline{E}(\cdot|\Box)$ and $\overline{E} = \overline{E}(\cdot|\Box)$ for the (unconditional) global models, associated with the initial situation $\Box$.

Our reasons for using these so-called *Shafer–Vovk–Ville formulae*[6] are fourfold.

First of all, they are formally very closely related to the expressions for lower and upper prices in Shafer and Vovk's game-theoretic approach to probabilities, see for instance Refs. [11, Chapter 8.3], [12, Section 2] and [14, Section 6.3]. This allows us to import and adapt, with the necessary care, quite a number of powerful convergence results from that theory, as we shall see in Section 4. Moreover, Shafer and Vovk (see for instance Refs. [11, Proposition 8.8] and [14, Section 6.3]) have shown that they—or rather their restrictions to gambles—satisfy our defining properties for lower and upper expectations in Section 2, which is why we are calling them lower and upper expectations.

Secondly, we gather from Proposition 1 and Corollary 2 that the expressions (3) and (4) coincide for $n$-measurable gambles on $\Omega$ with the formulae derived in Ref. [2] as the most conservative[7] global lower and upper expectations that extend the local models—see Corollary 3.[8]

**Proposition 1.** *For any situation $x_{1:m} \in \Omega^\Diamond$ and any $n$-measurable extended real variable $f$, with $n, m \in \mathbb{N}_0$ such that $n \geq m$:*

$$\begin{aligned}\underline{E}(f|x_{1:m}) = \sup\{\mathscr{M}(x_{1:m})\colon \mathscr{M} \in \underline{\mathbb{M}} \text{ and} \\ (\forall x_{m+1:n} \in \mathscr{X}_{m+1:n})\mathscr{M}(x_{1:n}) \leq f(x_{1:n})\} \\ \overline{E}(f|x_{1:m}) = \inf\{\mathscr{M}(x_{1:m})\colon \mathscr{M} \in \overline{\mathbb{M}} \text{ and} \\ (\forall x_{m+1:n} \in \mathscr{X}_{m+1:n})\mathscr{M}(x_{1:n}) \geq f(x_{1:n})\}.\end{aligned}$$

**Corollary 2.** *For any situation $x_{1:m} \in \Omega^\Diamond$ and any $n$-measurable extended real variable $f$, with $n, m \in \mathbb{N}_0$ such that $n \geq m$:*

$$\begin{aligned}\underline{E}(f|x_{1:m}) = \sup\{\underline{E}(g|x_{1:m})\colon g \in \mathscr{G}(\mathscr{X}_{1:n}) \text{ and} \\ (\forall x_{m+1:n} \in \mathscr{X}_{m+1:n})g(x_{1:n}) \leq f(x_{1:n})\} \\ \overline{E}(f|x_{1:m}) = \inf\{\overline{E}(g|x_{1:m})\colon g \in \mathscr{G}(\mathscr{X}_{1:n}) \text{ and} \\ (\forall x_{m+1:n} \in \mathscr{X}_{m+1:n})g(x_{1:n}) \geq f(x_{1:n})\}.\end{aligned}$$

**Corollary 3.** *Consider $n \in \mathbb{N}_0$ and $x_{1:n} \in \Omega^\Diamond$. Then for any $(n+1)$-measurable gamble $g$ on $\Omega$: $\underline{E}(g|x_{1:n}) = \underline{Q}(g(x_{1:n}\,\cdot)|x_{1:n})$ and $\overline{E}(g|x_{1:n}) = \overline{Q}(g(x_{1:n}\,\cdot)|x_{1:n})$. Also, for any $(n+1)$-measurable extended real variable $f$:*

$$\begin{aligned}\underline{E}(f|x_{1:n}) = \sup\{\underline{Q}(h|x_{1:n})\colon h \in \mathscr{G}(\mathscr{X}) \text{ and } h \leq f(x_{1:n}\,\cdot)\} \\ \overline{E}(f|x_{1:n}) = \inf\{\overline{Q}(h|x_{1:n})\colon h \in \mathscr{G}(\mathscr{X}) \text{ and } h \geq f(x_{1:n}\,\cdot)\}.\end{aligned}$$

Thirdly, it is (essentially) the expressions in Proposition 1 that we have used in Refs. [5, 7, 15] for our studies of imprecise Markov chains, which we report in Section 5. The main result of the present paper, Theorem 11 in Section 7, will build on the ergodicity results proved in those papers.

Fourthly, it was also shown in Ref. [2] that the expressions in Proposition 1 have an interesting interpretation in terms of (precise) probability trees. Indeed, we can associate with an imprecise probability tree a (usually infinite) collection of (so-called *compatible*) precise probability trees with the same event tree, by associating with each situation $s$ in the event tree some arbitrarily chosen precise local expectation $Q(\cdot|s)$ that belongs to the convex closed set $\mathfrak{M}(\underline{Q}(\cdot|s))$ of expectations that are compatible with the local lower expectation $\underline{Q}(\cdot|s)$. For any $n$-measurable gamble $f$ on $\Omega$, the global precise expectations in the compatible precise probability trees will then range over a closed interval whose lower and upper bounds are given by the expressions in Proposition 1.

And finally, Shafer and Vovk have shown [11, Chapter 8] that when the local models are precise probability models, these formulae (3) and (4) lead to global models that coincide with the ones found in measure-theoretic probability theory. *This implies that the results we shall prove below, subsume, as special cases, the classical results of measure-theoretic probability theory.*

## 4 A Strong Law of Large Numbers for Submartingale Differences

We now discuss and prove two powerful convergence results for the processes we have defined in the previous section.

We call an event $A$ *null* if $\overline{P}(A) = \overline{E}(\mathbb{I}_A) = 0$, and *strictly null* if there is some test supermartingale $\mathscr{T}$ that converges to $+\infty$ on $A$, meaning that:

$$\lim \mathscr{T}(\omega) = +\infty \quad \text{for all } \omega \in A.$$

Here, a *test supermartingale* is a supermartingale with $\mathscr{T}(\Box) = 1$ that is moreover non-negative in the sense that $\mathscr{T}(s) \geq 0$ for all situations $s \in \Omega^\Diamond$. Any strictly null event is null, but null events need not be strictly null [14].

**Proposition 4.** *Any strictly null event is null, but not vice versa.*[9]

[6] We give this name to these formulae because Glenn Shafer and Vladimir Vovk first suggested them, based on the ideas of Jean Ville; see the discussion of Ville's Theorem in Ref. [11, Appendix 8.5].

[7] By more conservative, we mean associated with a larger set of precise models, so pointwise smaller for lower expectations, and pointwise larger for upper expectations.

[8] We have also shown in recent, still unpublished work that in a more general context—where $X_k$ takes values in a possibly *infinite* set $\mathscr{X}_k$—for arbitrary gambles on $\Omega$ they are the most conservative global models that extend the local ones and satisfy additional conglomerability and continuity properties.

[9] For the null and strictly null events to be the same, it is necessary to consider supermartingales that may assume extended real values, as is done in Refs. [14, 12]. We see no need for doing so here.

In this paper, we shall use the ‘strict’ approach, and prove that events are strictly null—and therefore also null—by actually showing that there is a test supermartingale that converges to $+\infty$ there.

As usual, an inequality or equality between two variables is said to hold *(strictly) almost surely* when the event that it does not hold is (strictly) null. Shafer and Vovk [11, 14] have proved the following interesting result, which we shall have occasion to use a few times further on. It can be seen as a generalisation of Doob’s supermartingale convergence theorem [19, Sections 11.5–7] to imprecise probability trees.

**Theorem 5** ([14, Section 6.5] Supermartingale convergence theorem)**.** *Let $\mathscr{M}$ be a supermartingale that is bounded below. Then $\mathscr{M}$ converges strictly almost surely to a real variable.*

We now turn to a very general version of the strong law of large numbers. Weak (as well as less general) versions of this law were proven by one of us in Refs. [3, 2]. It is this law that will, in Section 7, be used to derive our version of the pointwise ergodic theorem. Its proof is based on a tried-and-tested method for constructing test supermartingales that goes back to an idea in Ref. [11, Lemma 3.3].

**Theorem 6** (Strong law of large numbers for submartingale differences)**.** *Let $\mathscr{M}$ be a submartingale such that $\Delta\mathscr{M}$ is uniformly bounded. Then* $\liminf\langle\mathscr{M}\rangle \geq 0$ *strictly almost surely.*

## 5 Imprecise Markov Chains

We are now ready to apply what we have learned in the previous sections to the special case of (time-homogeneous) imprecise Markov chains. These are imprecise probability trees where (i) all states $X_k$ assume values in the same *finite* set $\mathscr{X}_k = \mathscr{X}$, called the *state space*, and (ii) all local uncertainty models satisfy the so-called *Markov condition*:

$$\underline{Q}(\cdot|x_{1:n}) = \underline{Q}(\cdot|x_n) \text{ for all situations } x_{1:n} \in \Omega^{\Diamond}, \qquad (5)$$

meaning that these local models only depend on the last observed state; see Figure 3.

We refer to Refs. [5, 7, 15] for detailed studies of the behaviour of these processes. We restrict ourselves here to a summary of the existing material that is relevant for the present discussion of ergodicity.

From now on, we start using a convenient notational device often encountered in texts on stochastic processes: when we want to indicate which states a process or variable depends on, we indicate them explicitly in the notation. Thus, we use for instance the notation $\mathscr{F}(X_{1:n})$ to indicate the ‘uncertain’ value of the process $\mathscr{F}$ after the first $n$ time steps, and write $f(X_n)$ for a gamble that only depends on the value of the $n$-th state.

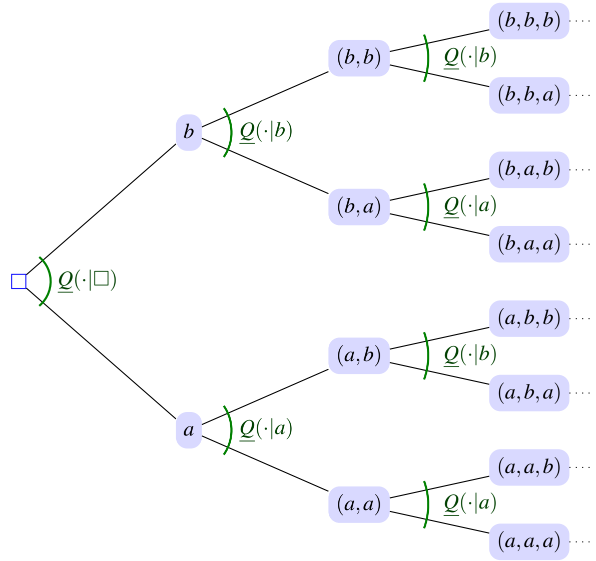


Figure 3: The (initial part of the) imprecise probability tree for an imprecise Markov process whose states can assume two values, $a$ and $b$, and can change at time instants $n = 1,2,3,\dots$

We can use the local uncertainty models to introduce a (generally non-linear) transformation $\underline{\mathrm{T}}$ of the set $\mathscr{G}(\mathscr{X})$ of all gambles on the state space $\mathscr{X}$. The so-called *lower transition operator* of the imprecise Markov chain is given by:

$$\underline{\mathrm{T}}\colon \mathscr{G}(\mathscr{X}) \to \mathscr{G}(\mathscr{X})\colon f \mapsto \underline{\mathrm{T}}f,$$

where $\underline{\mathrm{T}}f$ is the gamble on $\mathscr{X}$ defined by

$$\underline{\mathrm{T}}f(x) := \underline{Q}(f|x) \text{ for all } x \in \mathscr{X}.$$

The conjugate *upper transition operator* $\overline{\mathrm{T}}$ is defined by $\overline{\mathrm{T}}f := -\underline{\mathrm{T}}(-f)$ for all $f \in \mathscr{G}(\mathscr{X})$. In particular, $\underline{\mathrm{T}}\mathbb{I}_{\{y\}}(x)$ is the lower probability to go from state value $x$ to state value $y$ in one time step, and $\overline{\mathrm{T}}\mathbb{I}_{\{y\}}(x)$ the conjugate upper probability. This seems to suggest that the lower/upper transition operators $\underline{\mathrm{T}}$ are generalisations of the concept of a Markov transition matrix for ordinary Markov chains. This is confirmed by the following result, proved in Ref. [5, Corollary 3.3] as a special case of the so-called Law of Iterated (Lower) Expectations [2, 11]. If, for any $n \in \mathbb{N}$, we denote by $\underline{E}_n(f)$ the value of the (global) lower expectation $\underline{E}(f(X_n))$ of a gamble $f(X_n)$ on the state $X_n$ at time $n$, then

$$\underline{E}_n(f) = \underline{E}_1(\underline{\mathrm{T}}^{n-1}f), \text{ with } \underline{\mathrm{T}}^{n-1}f := \underbrace{\underline{\mathrm{T}}\,\underline{\mathrm{T}}\dots\underline{\mathrm{T}}}_{n-1 \text{ times}}f,$$

and where, of course, $\underline{E}_1 = \underline{Q}(\cdot|\square)$ is the marginal local model for the state $X_1$ at time 1. In a similar vein, for any $n \in \mathbb{N}_0$, $\underline{\mathrm{T}}^n\mathbb{I}_{\{y\}}(x)$ is the lower probability to go from state value $x$ to state value $y$ in $n$ time steps, and $\overline{\mathrm{T}}^n\mathbb{I}_{\{y\}}(x)$ the conjugate upper probability.

We can formally call *lower transition operator* any transformation $\underline{\mathrm{T}}$ of $\mathscr{G}(\mathscr{X})$ such that for any $x \in \mathscr{X}$, the real functional $\underline{\mathrm{T}}_x$ on $\mathscr{G}(\mathscr{X})$, defined by $\underline{\mathrm{T}}_x(f) := \underline{\mathrm{T}}f(x)$ for all $f \in \mathscr{G}(\mathscr{X})$, is a lower expectation—satisfies the coherence axioms LE1–LE3. The composition of any two lower transition operators is again a lower transition operator. See Ref. [5] for more details on the definition and properties of such lower transition operators, and Ref. [4] for a mathematical discussion of the general role of these operators in imprecise probabilities.

We call an imprecise Markov chain with lower transition operator $\underline{\mathrm{T}}$ *Perron–Frobenius-like* if for all $f \in \mathscr{G}(\mathscr{X})$, the sequence of gambles $\underline{\mathrm{T}}^n f$ converges pointwise to a constant real number, which we shall then denote by $\underline{E}_{\mathrm{PF}}(f)$.

The following result was proved in Ref. [5, Theorem 5.1], together with a simple sufficient (and quite weak) condition on $\underline{\mathrm{T}}$ for a Markov chain to be Perron–Frobenius-like: there is some $n \in \mathbb{N}$ such that $\min \overline{\mathrm{T}}^n \mathbb{I}_{\{y\}} > 0$ for all $y \in \mathscr{X}$, or in other words, all state values can be reached from any state value with positive upper probability in (precisely) $n$ time steps. More involved necessary *and* sufficient conditions were given later in Refs. [7, 15]; see also Theorem 8(iv) further on.

**Proposition 7** ([5])**.** *The imprecise Markov chain with lower transition operator $\underline{\mathrm{T}}$ is Perron–Frobenius-like if and only if there is some real functional $\underline{E}_\infty$ on $\mathscr{G}(\mathscr{X})$ such that for any initial model $\underline{E}_1$ and any $f \in \mathscr{G}(\mathscr{X})$, it holds that $\underline{E}_n(f) = \underline{E}_1(\underline{\mathrm{T}}^{n-1} f) \to \underline{E}_\infty(f)$. Moreover, in that case the functional $\underline{E}_\infty$ is a lower expectation on $\mathscr{G}(\mathscr{X})$, called the* stationary lower expectation*, it coincides with $\underline{E}_{\mathrm{PF}}$, and it is the only lower expectation that is $\underline{\mathrm{T}}$-invariant in the sense that $\underline{E}_\infty \circ \underline{\mathrm{T}} = \underline{E}_\infty$.*

## 6 An Interesting Equality in Imprecise Markov Chains

We now prove an interesting equality for imprecise Markov chains, which will be instrumental in proving our pointwise ergodic theorem in the next section.

Consider, for any $f \in \mathscr{G}(\mathscr{X})$, the corresponding *gain* process $\mathscr{W}[f]$, defined by, for any $n \in \mathbb{N}$:

$$\mathscr{W}[f](X_{1:n}) := [f(X_1) - \underline{E}_1(f)] + \sum_{k=2}^{n} [f(X_k) - \underline{\mathrm{T}} f(X_{k-1})], \quad (6)$$

the corresponding *average gain* process $\langle\mathscr{W}\rangle[f]$, defined by:

$$\langle\mathscr{W}\rangle[f](X_{1:n}) := \frac{1}{n}\left[[f(X_1) - \underline{E}_1(f)] + \sum_{k=2}^{n} [f(X_k) - \underline{\mathrm{T}} f(X_{k-1})]\right], \quad (7)$$

and the *ergodic average* process $\mathscr{A}[f]$, defined by:

$$\mathscr{A}[f](X_{1:n}) := \frac{1}{n} \sum_{k=1}^{n} [f(X_k) - \underline{E}_k(f)]. \quad (8)$$

We can let these processes be 0 in the initial situation $\square$—the choice is immaterial. Now observe that, for any $n \in \mathbb{N}$ and any $f \in \mathscr{G}(\mathscr{X})$:

$$\begin{aligned}\sum_{\ell=0}^{n-1} \langle\mathscr{W}\rangle[\underline{\mathrm{T}}^\ell f](X_{1:n}) &= \frac{1}{n} \sum_{\ell=0}^{n-1} \left[\underline{\mathrm{T}}^\ell f(X_1) - \underline{E}_1(\underline{\mathrm{T}}^\ell f)\right] \\ &\quad + \frac{1}{n} \sum_{\ell=0}^{n-1} \sum_{k=2}^{n} \left[\underline{\mathrm{T}}^\ell f(X_k) - \underline{\mathrm{T}}^{\ell+1} f(X_{k-1})\right], \quad (9)\end{aligned}$$

and moreover

$$\begin{aligned}&\sum_{\ell=0}^{n-1} \sum_{k=2}^{n} \left[\underline{\mathrm{T}}^\ell f(X_k) - \underline{\mathrm{T}}^{\ell+1} f(X_{k-1})\right] \\ &= \sum_{\ell=0}^{n-1} \sum_{k=2}^{n} \underline{\mathrm{T}}^\ell f(X_k) - \sum_{\ell=0}^{n-1} \sum_{k=2}^{n} \underline{\mathrm{T}}^{\ell+1} f(X_{k-1}) \\ &= \sum_{\ell=0}^{n-1} \sum_{k=2}^{n} \underline{\mathrm{T}}^\ell f(X_k) - \sum_{\ell=1}^{n} \sum_{k=1}^{n-1} \underline{\mathrm{T}}^\ell f(X_k) \\ &= \sum_{k=2}^{n} f(X_k) + \sum_{\ell=1}^{n-1} \left(\underline{\mathrm{T}}^\ell f(X_n) + \sum_{k=2}^{n-1} \underline{\mathrm{T}}^\ell f(X_k)\right) \\ &\qquad - \sum_{k=1}^{n-1} \underline{\mathrm{T}}^n f(X_k) - \sum_{\ell=1}^{n-1} \left(\underline{\mathrm{T}}^\ell f(X_1) + \sum_{k=2}^{n-1} \underline{\mathrm{T}}^\ell f(X_k)\right) \\ &= \sum_{k=2}^{n} f(X_k) + \sum_{\ell=1}^{n-1} \underline{\mathrm{T}}^\ell f(X_n) - \sum_{k=1}^{n-1} \underline{\mathrm{T}}^n f(X_k) - \sum_{\ell=1}^{n-1} \underline{\mathrm{T}}^\ell f(X_1) \\ &= \sum_{k=1}^{n} f(X_k) + \sum_{\ell=1}^{n} \underline{\mathrm{T}}^\ell f(X_n) - \sum_{k=1}^{n} \underline{\mathrm{T}}^n f(X_k) - \sum_{\ell=0}^{n-1} \underline{\mathrm{T}}^\ell f(X_1),\end{aligned}$$

and if we substitute this back into Equation (9), we find, after getting rid of the cancelling terms, recalling that $\underline{E}_1(\underline{\mathrm{T}}^\ell f) = \underline{E}_{\ell+1}(f)$, and reorganising a bit, that:

$$\mathscr{A}[f](X_{1:n}) = \sum_{\ell=0}^{n-1} \langle\mathscr{W}\rangle[\underline{\mathrm{T}}^\ell f](X_{1:n}) + \frac{1}{n} \sum_{k=1}^{n} \underline{\mathrm{T}}^n f(X_k) - \frac{1}{n} \sum_{\ell=1}^{n} \underline{\mathrm{T}}^\ell f(X_n). \quad (10)$$

This is an important relationship between the ergodic average and the average gain. We now intend to show that under certain conditions the remaining terms on the right-hand side essentially cancel out for large enough $n$.

# 7 Consequences of the Perron–Frobenius-like Character

Let us associate with a lower transition operator $\underline{\mathrm{T}}$ the following (weak) *coefficient of ergodicity* [15, 7]:

$$\rho(\underline{\mathrm{T}}) := \max_{x,y\in\mathscr{X}} \max_{h\in\mathscr{G}_1(\mathscr{X})} |\underline{\mathrm{T}}h(x) - \underline{\mathrm{T}}h(y)| = \max_{h\in\mathscr{G}_1(\mathscr{X})} \|\underline{\mathrm{T}}h\|_{\mathrm{v}},$$

where $\mathscr{G}_1(\mathscr{X}) := \{h \in \mathscr{G}(\mathscr{X}) \colon 0 \le h \le 1\}$, and where for any $h \in \mathscr{G}(\mathscr{X})$, its *variation (semi)norm* is given by $\|h\|_{\mathrm{v}} := \max h - \min h$. If we define the following distance between two lower expectation operators $\underline{E}$ and $\underline{F}$ [15]:

$$d(\underline{E},\underline{F}) = \max_{h\in\mathscr{G}_1(\mathscr{X})} |\underline{E}(h) - \underline{F}(h)|,$$

then it is not difficult to see [using LE3, LE4 and LE6] that $0 \le d(\underline{E},\underline{F}) \le 1$, and that for any $f \in \mathscr{G}(\mathscr{X})$:

$$|\underline{E}(f) - \underline{F}(f)| \le d(\underline{E},\underline{F})\|f\|_{\mathrm{v}}. \tag{11}$$

Škulj and Hable [15] have proved the following results, which will turn out to be crucial to our argument.

**Theorem 8** ([15])**.** *Consider lower transition operators $\underline{\mathrm{S}}$ and $\underline{\mathrm{T}}$, and two lower expectations $\underline{E}_a$ and $\underline{E}_b$ on $\mathscr{G}(\mathscr{X})$. Then the following statements hold:*

(i) $0 \le \rho(\underline{\mathrm{T}}) \le 1$.

(ii) $\rho(\underline{\mathrm{S}}\underline{\mathrm{T}}) \le \rho(\underline{\mathrm{S}})\rho(\underline{\mathrm{T}})$ *and therefore* $\rho(\underline{\mathrm{T}}^n) \le \rho(\underline{\mathrm{T}})^n$ *for all* $n \in \mathbb{N}$.

(iii) $d(\underline{E}_a\underline{\mathrm{T}}, \underline{E}_b\underline{\mathrm{T}}) \le d(\underline{E}_a,\underline{E}_b)\rho(\underline{\mathrm{T}})$.

(iv) *The lower transition operator $\underline{\mathrm{T}}$ is Perron–Frobenius-like if and only if there is some $r \in \mathbb{N}$ such that $\rho(\underline{\mathrm{T}}^r) < 1$.*

Indeed, they allow us to derive useful bounds for the various terms on the right-hand side of Equation (10). For any non-negative real number $a$ we denote by $\lfloor a \rfloor = \max\{n \in \mathbb{N}_0 \colon n \le a\}$ the largest natural number that it still dominates—its integer part.

**Proposition 9.** *Let $\underline{\mathrm{T}}$ be a Perron–Frobenius-like lower transition operator, with invariant lower expectation $\underline{E}_\infty$, and let $r$ be the smallest natural number such that $\rho := \rho(\underline{\mathrm{T}}^r) < 1$. Let $\underline{E}_a$ and $\underline{E}_b$ be any two lower expectations on $\mathscr{G}(\mathscr{X})$. Then for all $f \in \mathscr{G}(\mathscr{X})$, $\ell_1, \ell_2 \in \mathbb{N}_0$:*

$$\left|\underline{E}_a(\underline{\mathrm{T}}^{\ell_1} f) - \underline{E}_b(\underline{\mathrm{T}}^{\ell_2} f)\right| \le \|f\|_{\mathrm{v}}\rho^{\lfloor\frac{\min\{\ell_1,\ell_2\}}{r}\rfloor}. \tag{12}$$

*As a consequence, for all $f \in \mathscr{G}(\mathscr{X})$, $\ell, \ell_1, \ell_2 \in \mathbb{N}_0$ and $k, k_1, k_2 \in \mathbb{N}$:*

$$\left|\underline{\mathrm{T}}^{\ell} f(X_k) - \underline{E}_\infty(f)\right| \le \|f\|_{\mathrm{v}}\rho^{\lfloor\frac{\ell}{r}\rfloor}, \tag{13}$$

$$\left|\underline{E}_a(\underline{\mathrm{T}}^{\ell} f) - \underline{E}_\infty(f)\right| \le \|f\|_{\mathrm{v}}\rho^{\lfloor\frac{\ell}{r}\rfloor}, \tag{14}$$

$$\left|\underline{\mathrm{T}}^{\ell} f(X_k) - \underline{E}_b(\underline{\mathrm{T}}^{\ell} f)\right| \le \|f\|_{\mathrm{v}}\rho^{\lfloor\frac{\ell}{r}\rfloor}, \tag{15}$$

$$\left|\underline{\mathrm{T}}^{\ell_1} f(X_{k_1}) - \underline{\mathrm{T}}^{\ell_2} f(X_{k_2})\right| \le \|f\|_{\mathrm{v}}\rho^{\lfloor\frac{\min\{\ell_1,\ell_2\}}{r}\rfloor}. \tag{16}$$

**Proposition 10.** *Consider an imprecise Markov chain with initial—or marginal—model $\underline{E}_1$ and lower transition operator $\underline{\mathrm{T}}$. Assume that $\underline{\mathrm{T}}$ is Perron–Frobenius-like, with invariant lower expectation $\underline{E}_\infty$, and let $r$ be the smallest natural number such that $\rho := \rho(\underline{\mathrm{T}}^r) < 1$. Then the following statements hold for all $f \in \mathscr{G}(\mathscr{X})$, $\ell \in \mathbb{N}_0$ and $n \in \mathbb{N}$:*

(i) $|\langle\mathscr{W}\rangle[\underline{\mathrm{T}}^{\ell} f](X_{1:n})| \le \|f\|_{\mathrm{v}}\rho^{\lfloor\frac{\ell}{r}\rfloor}$.

(ii) $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \underline{\mathrm{T}}^{n} f(X_k) = \underline{E}_\infty(f)$.

(iii) $\lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} \underline{\mathrm{T}}^{\ell} f(X_n) = \underline{E}_\infty(f)$.

(iv) $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \underline{E}_k(f) = \underline{E}_\infty(f)$.

We can now state our main result.

**Theorem 11** (Pointwise ergodic theorem)**.** *Consider an imprecise Markov chain with initial—or marginal—model $\underline{E}_1$ and lower transition operator $\underline{\mathrm{T}}$. Assume that $\underline{\mathrm{T}}$ is Perron–Frobenius-like, with invariant lower expectation $\underline{E}_\infty$. Then for all $f \in \mathscr{G}(\mathscr{X})$:*

$$\liminf \mathscr{A}[f] \ge 0 \text{ *strictly almost surely*},$$

*and consequently,*

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(X_k) \ge \underline{E}_\infty(f) \text{ *strictly almost surely*}.$$

# 8 Conclusions and Discussion

We have proved a version of the pointwise ergodic theorem for imprecise Markov chains involving functions of a single state. It does not seem very difficult to extend this result to involve functions of a finite number of states, but it is still a subject of current research whether it can be extended to gambles that depend on the entire state trajectory, and not just on a finite number of states.

Our version subsumes the one for (precise) Markov chains, because there $\underline{E}_\infty(f) = \overline{E}_\infty(f) = E_\infty(f)$ and therefore

$$\begin{aligned} E_\infty(f) = \overline{E}_\infty(f) &\ge \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(X_k) \\ &\ge \liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(X_k) \\ &\ge \underline{E}_\infty(f) = E_\infty(f) \\ &\qquad\text{strictly almost surely,} \end{aligned}$$

implying that $\frac{1}{n}\sum_{k=1}^{n} f(X_k)$ converges to $E_\infty(f)$ (strictly) almost surely. In our more general case, however, we cannot generally prove that there is almost sure *convergence*, and we retain only almost sure inequalities involving limits inferior and superior, as is also the case for our strong law of large numbers for submartingale differences. Indeed, that such convergence should not really be expected for imprecise probability models was already argued by Walley and Fine [17].

Ergodicity results for Markov chains are quite relevant for applications in queuing theory, where they are for instance used to prove Little's Law [18], or ASTA (Arrivals See Time Averages) properties [9]. We believe the discussion in this paper could be instrumental in deriving similar properties for queues where the probability models for arrivals and departures are imprecise.

## Acknowledgements

Research by Gert de Cooman and Stavros Lopatatzidis was funded through project number G012512N of the Research Foundation Flanders (FWO). Jasper De Bock is a PhD Fellow of the FWO and wishes to acknowledge its financial support. The authors would like to express their gratitude to three anonymous referees for their comments, and to Tom Ward and Volodya Vovk for taking the time to discuss some of the ideas behind this paper.

## References

[1] Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors. *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[2] Gert de Cooman and Filip Hermans. Imprecise probability trees: Bridging two theories of imprecise probability. *Artificial Intelligence*, 172(11):1400–1427, 2008. doi: 10.1016/j.artint.2008.03.001.

[3] Gert de Cooman and Enrique Miranda. Weak and strong laws of large numbers for coherent lower previsions. *Journal of Statistical Planning and Inference*, 138(8):2409–2432, 2008. doi: 10.1016/j.jspi.2007.10.020.

[4] Gert de Cooman and Enrique Miranda. Lower previsions induced by filter maps. *Journal of Mathematical Analysis and Applications*, 410(1):101–116, 2014.

[5] Gert de Cooman, Filip Hermans, and Erik Quaeghebeur. Imprecise Markov chains and their limit behaviour. *Probability in the Engineering and Informational Sciences*, 23(4):597–635, January 2009. doi: 10.1017/S0269964809990039. arXiv:0801.0980.

[6] Darald J. Hartfiel. *Markov Set-Chains*. Number 1695 in Lecture Notes in Mathematics. Springer, Berlin, 1998.

[7] Filip Hermans and Gert de Cooman. Characterisation of ergodic upper transition operators. *International Journal of Approximate Reasoning*, 53(4):573–583, 2012. doi: 10.1016/j.ijar.2011.12.008.

[8] Olav Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, New York, second edition, 2002.

[9] Armand Makowski, Benjamin Melamed, and Ward Whitt. On averages seen by arrivals in discrete time. In *28th IEEE Conference on Decision and Control*, pages 1084–1086. IEEE, 1989.

[10] Enrique Miranda and Gert de Cooman. Lower previsions. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[11] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley, New York, 2001.

[12] Glenn Shafer, Vladimir Vovk, and Akimichi Takemura. Lévy's zero–one law in game-theoretic probability. *Journal of Theoretical Probability*, 25: 1–24, 2012.

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

[14] Vladimir Vovk and Glenn Shafer. Game-theoretic probability. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[15] Damjan Škulj and Robert Hable. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, 76(1):107–133, 2013. doi: 10.1007/s00184-011-0378-0.

[16] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[17] Peter Walley and Terrence L. Fine. Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10:741–761, 1982.

[18] Ward Whitt. A review of $L = \lambda W$ and extensions. *Queueing Systems*, 9(3):235–268, 1991.

[19] David Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, UK, 1991.