A pointwise ergodic theorem for imprecise Markov chains

Gert de Cooman, Jasper De Bock, Stavros Lopatatzidis

SYSTeMS research group, Ghent University, Belgium

{Gert.deCooman,Jasper.DeBock,Stavros.Lopatatzidis}@UGent.be

Abstract

Ergodicity theorem Consider a Markov chain, with a finite state space $\mathscr{X}$. For such a system, we have proved various *Perron-Frobenius*-like theorems. They provide necessary and sufficient conditions for the uncertainty model about the state X_n to converge, as $n \to \infty$, to an uncertainty model independent of the initial state X_1. In Markov chains with precise probabilities, this convergence is sufficient for a pointwise ergodic theorem to hold:

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

Our result Applying the theory of imprecise probabilities to stochastic processes, we can define so-called imprecise Markov chains as special cases of imprecise probability trees. We introduce and study submartingales and supermartingales in such trees, for which we are able to prove a strong law of large numbers for submartingale differences. Combining this result with the Perron-Frobenius-like character of our model we can prove the following

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

Imprecise probabilities

Imprecise probabilities We present the basic axioms for the theory of imprecise probabilities introduced by Walley (1991). For a recent introductory book, see also Thomas Augustin, Frank P. A. Coolen, Gert de Cooman and Matthias C. M. Troffaes (2014).

Suppose a subject is uncertain about the value that a variable X assumes in a finite set of possible values $\mathscr{X}$. His uncertainty is modelled by a *lower expectation* $\underline{E}$, which is a real functional on the set $\mathscr{G}(\mathscr{X})$ of all real-valued functions (**gambles**) $f\colon \mathscr{X} \to \mathbb{R}$ on $\mathscr{X}$, satisfying the following the basic so-called coherence axioms:

1. $\underline{E}(f) \geq \min f$ for all $f \in \mathscr{G}(\mathscr{X})$;
2. $\underline{E}(f+g) \geq \underline{E}(f) + \underline{E}(g)$ for all $f,g \in \mathscr{G}(\mathscr{X})$;
3. $\underline{E}(\lambda f) = \lambda \underline{E}(f)$ for all $f \in \mathscr{G}(\mathscr{X})$ and real $\lambda \geq 0$.

The *conjugate upper expectation* $\overline{E}$ is defined by $\overline{E}(f) := -\underline{E}(-f)$ and it follows from the coherence axioms 1–3 that

4. $\min f \leq \underline{E}(f) \leq \overline{E}(f) \leq \max f$ for all $f \in \mathscr{G}(\mathscr{X})$;
5. $\underline{E}(f+g) \leq \underline{E}(f) + \overline{E}(g) \leq \overline{E}(f+g) \leq \overline{E}(f) + \overline{E}(g)$ for all $f,g \in \mathscr{G}(\mathscr{X})$;
6. $\underline{E}(f) \leq \underline{E}(g)$ and $\overline{E}(f) \leq \overline{E}(g)$ for all $f,g \in \mathscr{G}(\mathscr{X})$ with $f \leq g$;
7. $\underline{E}(f+\mu) = \underline{E}(f) + \mu$ and $\overline{E}(f+\mu) = \overline{E}(f) + \mu$ for all $f \in \mathscr{G}(\mathscr{X})$ and real μ.

Event trees and processes

Event trees We denote by $X_{k:\ell}$ the tuple $(X_k, \dots, X_\ell)$, taking values in set $\mathscr{X}_{k:\ell} := \times_{r=k}^{\ell} \mathscr{X}_r$, for any $k \leq \ell$ with $k, \ell \in \mathbb{N}$. A *situation* is an finite sequence of states $x_{1:n} \in \mathscr{X}_{1:n}$, with $n \in \mathbb{N}_0$, and the set of all situations is denoted by $\Omega^\lozenge$. Infinite sequences of states are called *paths*, and the set of all paths is denoted by Ω.

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

For any path $\omega \in \Omega$, the initial sequence of its first n elements, $\mathscr{X}_{1:n}$, is denoted by ω^n. A *variable* is a function defined on Ω. It is called n-*measurable* if it only depends on the value of $X_{1:n}$. An *event* is a subset of Ω. With any situation $x_{1:n}$, we can associate the so-called *exact event* $\Gamma(x_{1:n})$, which is the set of all paths $\omega \in \Omega$ that go through $x_{1:n}$.

Processes A *process* $\mathscr{F}$ is a map defined on $\Omega^\lozenge$. The *process difference* $\Delta\mathscr{F}(x_{1:n}) \in \mathscr{G}(\mathscr{X}_{n+1})$ is defined by

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

We can associate a real process $\mathscr{F}$ with extended real variables $\liminf \mathscr{F}$ and $\limsup \mathscr{F}$, defined for all $\omega \in \Omega$ by:

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

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:

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

Imprecise trees and martingales

Imprecise probability trees We turn the event tree into a *probability tree* by assigning to each situation $x_{1:n}$, a *local* probability model $Q(\cdot|x_{1:n})$. 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}, given that $X_{1:n} = x_{1:n}$. We can equally well attach to each situation $x_{1:n}$ a local *imprecise* probability model $\underline{Q}(\cdot|x_{1:n})$ for the next state 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}, given $X_{1:n} = x_{1:n}$.

Martingales 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 $n \in \mathbb{N}$ and $x_{1:n} \in \mathscr{X}_{1:n}$. A real process $\mathscr{M}$ is a *supermartingale* if $-\mathscr{M}$ is a submartingale, meaning that $\overline{Q}(\Delta\mathscr{M}(x_{1:n})|x_{1:n}) \leq 0$. We denote the set of all submartingales for a given imprecise probability tree by $\underline{\mathbb{M}}$. Similarly, we have $\overline{\mathbb{M}} := -\underline{\mathbb{M}}$.

Consider any submartingale $\mathscr{M}$ and any situation $s \in \Omega^\lozenge$, then:

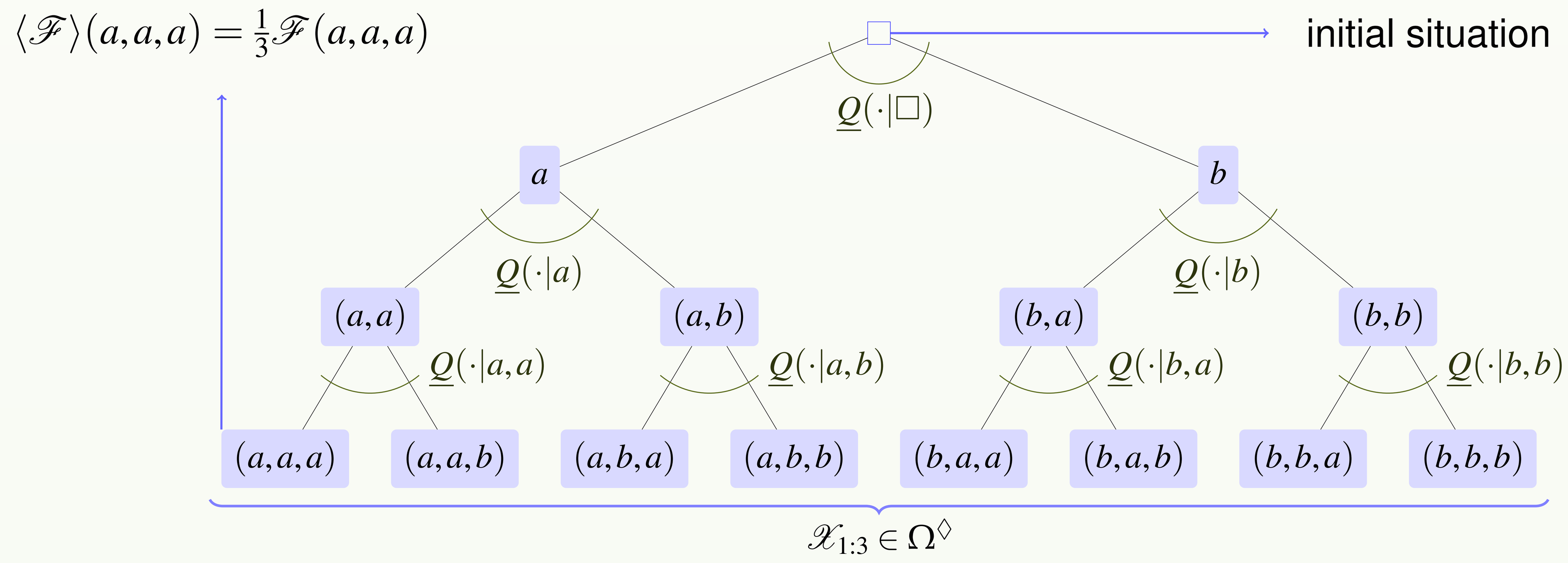
$$\mathscr{M}(s) \leq \sup_{\omega\in\Gamma(s)} \liminf \mathscr{M}(\omega) \leq \sup_{\omega\in\Gamma(s)} \limsup \mathscr{M}(\omega).$$

Using this inequality, and results from previous papers, we were able to prove the following formulas for the global conditional lower expectations (the so-called Shafer–Vovk–Ville formulas)

$$\underline{E}(f|s) := \sup\{\mathscr{M}(s)\colon \mathscr{M} \in \underline{\mathbb{M}} \text{ and } \limsup \mathscr{M}(\omega) \leq f(\omega) \text{ for all } \omega \in \Gamma(s)\}.$$

As a special case, for any situation $x_{1:m} \in \Omega^\lozenge$ and any n-measurable real variable f, with $n, m \in \mathbb{N}$ such that $n \geq m$:

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



Strong law of large numbers for submartingale differences

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

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

A *test supermartingale* is a supermartingale with $\mathscr{T}(\square) = 1$ that is non-negative for all situations in $\Omega^\lozenge$.

Using the definitions of null and strictly null event, Shafer and Vovk (2001) proved the following version of the *supermartingale convergence theorem*:

Theorem 1. *Let $\mathscr{M}$ be a supermartingale that is bounded below. Then $\mathscr{M}$ converges strictly almost surely to a real variable.*

The intuition behind it is that there exists a test supermartingale which is $+\infty$ on the paths where the process diverges. We were able to derive the following useful theorem:

Theorem 2 (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.*

An interesting result for imprecise Markov chains

Imprecise Markov chains Imprecise Markov chains are imprecise probability trees where all local uncertainty models satisfy the so-called *Markov condition*:

$$\underline{Q}(\cdot|x_{1:n}) = \underline{Q}(\cdot|x_n) \text{ for all situations } x_{1:n} \in \Omega^\lozenge.$$

The *lower transition operator* $\underline{\mathrm{T}}\colon \mathscr{G}(\mathscr{X}) \to \mathscr{G}(\mathscr{X})\colon f \mapsto \underline{\mathrm{T}}f$ is defined by

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

and the (global) lower expectation $\underline{E}_n(f) := \underline{E}(f(X_n))$ at time n is then given by

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

An imprecise Markov chain is *Perron–Frobenius-like* if for all $f \in \mathscr{G}(\mathscr{X})$, the sequence $\underline{\mathrm{T}}^n f$ converges pointwise to a constant real number, denoted by $\underline{E}_\infty(f)$. The $\underline{E}_\infty(f)$ is also $\underline{\mathrm{T}}$-invariant in the sense that $\underline{E}_\infty \circ \underline{\mathrm{T}} = \underline{E}_\infty$.

Towards an imprecise ergodic theorem For any $f \in \mathscr{G}(\mathscr{X})$, the *average gain* process is defined by:

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

and the *ergodic average* process by:

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

It can be proved that

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

Associate with $\underline{\mathrm{T}}$ the (weak) *coefficient of ergodicity* ρ:

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

where $\mathscr{G}_1(\mathscr{X}) := \{h \in \mathscr{G}(\mathscr{X})\colon 0 \leq h \leq 1\}$, and for any $h \in \mathscr{G}(\mathscr{X})$, $\|h\|_{\mathrm{v}} := \max h - \min h$. Then it can be shown that an imprecise Markov chain is Perron–Frobenius-like if and only if $\rho(\underline{\mathrm{T}}^r) < 1$ for some $r \in \mathbb{N}$. If we define the following distance :

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

then we derive [using 1, 3 and 7] that $0 \leq d(\underline{E}, \underline{F}) \leq 1$, and :

$$|\underline{E}(f) - \underline{F}(f)| \leq d(\underline{E}, \underline{F})\|f\|_{\mathrm{v}}. \tag{1}$$

Using (1) and the property $0 \leq \rho(\underline{\mathrm{T}}) \leq 1$, we get

$$|\underline{\mathrm{T}}^{\ell_1}f(X_{k_1}) - \underline{\mathrm{T}}^{\ell_2}f(X_{k_2})| \leq \|f\|_{\mathrm{v}}\rho^{\lfloor\frac{\min\{\ell_1,\ell_2\}}{r}\rfloor}. \tag{2}$$

Combining $\langle\mathscr{W}\rangle[f](X_{1:n})$ with (2), we have

$$|\langle\mathscr{W}\rangle[\underline{\mathrm{T}}^\ell f]| \leq \|f\|_{\mathrm{v}}\rho^{\lfloor\frac{\ell}{r}\rfloor}. \tag{3}$$

From (3) and $\mathscr{A}[f](X_{1:n})$ and using **Theorem 2**

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

and consequently our main result,

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