Coefficients of ergodicity for imprecise Markov chains

Damjan Škulj
University of Ljubljana
Faculty of Social Sciences
damjan.skulj@fdv.uni-lj.si

Robert Hable
University of Bayreuth
Department of Mathematics
robert.hable@uni-bayreuth.de

Abstract

Coefficients of ergodicity are an important tool in measuring convergence of Markov chains. We explore possibilities to generalise the concept to imprecise Markov chains. We find that this can be done in at least two different ways, which both have interesting implications in the study of convergence of imprecise Markov chains. Thus we extend the existing definition of the uniform coefficient of ergodicity and define a new so-called weak coefficient of ergodicity. The definition is based on the endowment of a structure of a metric space to the class of imprecise probabilities. We show that this is possible to do in some different ways, which turn out to coincide.

Keywords. Markov chain, imprecise Markov chain, coefficient of ergodicity, lower expectation, upper expectation

1 Introduction

Markov chains are a very popular mathematical model used to describe various dynamical systems. Their properties have been studied in great detail. The modelling of a Markov chain requires estimating a relatively large number of parameters, which is in many practical situations very difficult to achieve precisely. Thus sometimes parameters are estimated with high imprecision, and the theory provides virtually no better answer than regarding the most likely estimates as precise, leading to seemingly precise results that do not reflect the lack of certainty in the input data.

The rapid development of the methods of imprecise probabilities has allowed the study of Markov chains where the imprecision in input data can be incorporated in the results. A detailed study in this topic has been presented by Hartfiel [8] who considered the model where precise initial and transition probability matrices are replaced by sets of possible initial probabilities and transition matrices. This model is known under the name Markov set-chains (see also Hartfiel and Seneta [9]). He pays special attention to the case where the sets can be described using probability intervals. This basically means that every probability of an elementary event is bounded by a lower and upper bound. A similar model was studied from the perspective of the theory of interval probabilities by Kozine and Utkin [11]. The more general interval probabilities based on the Weichselberger's model [20] were involved in the study of Markov chains by Škulj [16, 17]. A more recent approach by de Cooman et al. [2] further generalises the way imprecision is involved into Markov chains, taking an approach based on upper expectation operators. This approach is known from the study of the related field of Markov decision processes used by Satia and Lave [14], followed by [7, 10, 12, 21].

In this paper we follow the approach of de Cooman et al. The topic we study here is the convergence of imprecise Markov chains. The most common result in the classical theory is the Perron-Frobenius theorem that implies unique convergence for the case of regular Markov chains. In [17] the concept of regularity was generalised to imprecise Markov chains and a similar theorem was proved. However, it turns out that weaker conditions than regularity are sufficient to ensure convergence of Markov chains, both in precise and imprecise case. In both cases coefficients of ergodicity prove to be very useful tools. They have been widely used in the precise case (see e.g. Seneta [15]), while Hartfiel [8] generalises them to imprecise Markov chains.

Recently, de Cooman et al. give conditions for convergence of imprecise Markov chains that are substantially weaker than those used by Hartfiel [8], although in the precise case they seem to be very similar. The different generalisations of the conditions for convergence suggest that there may be different possibilities to define coefficients of ergodicity for the case of imprecise Markov chains. In this paper we show that indeed a generalisation different from the one used by Hartfiel is possible. We also believe, although we have not yet explored this relation, that conditions implied by our new generalised coefficients are closely related to those found by de Cooman et al. The definition of the new coefficient of ergodicity is based on endowing the set of imprecise probabilities with a structure of a metric space.

The paper has the following structure. In the next section we review some theory on lower expectation operators that form a basis for the model of imprecise Markov chains. Further, in Section 3 we explore some possibilities to endow the family of imprecise probabilities with the structure of a metric space, and in Section 4 we describe the model of imprecise Markov chains that we use. Finally, in Section 5 we study the generalisations of coefficients of ergodicity and compare them to the existing generalisations.

2 Lower expectation operators

Let Ω be a finite set of states and let $\mathcal{F}$ be the set of real-valued maps on Ω. Further let $\mathcal{F}_1$ denote the subset of all non-negative real-valued maps with $f(\omega) \leq 1$ for every $\omega \in \Omega$. We denote by 1_Ω, or sometimes just 1, the constant map on Ω such that $f(\omega) = 1$ for all $\omega \in \Omega$. For a pair of maps f and g such that $f(\omega) \geq g(\omega)$ for every $\omega \in \Omega$ we write $f \geq g$, and if at least one of the inequalities is strict we write $f > g$.

The set $\mathcal{F}$ can be equipped with the *maximum norm* given by

$$||f||_\infty = \max_{\omega\in\Omega} |f(\omega)|,$$

which induces the *Chebyshev distance*:

$$d_c(f, g) = \max_{\omega\in\Omega} |f(\omega) - g(\omega)|.$$

We can write $\mathcal{F}_1 = \{f \in \mathcal{F} \mid f \geq 0, ||f||_\infty \leq 1\}$.

We characterise a *probability measure* or a *probability* p as a real valued map on Ω such that

$$\sum_{\omega\in\Omega} p(\omega) = 1$$

and

$$p(\omega) \geq 0 \quad \text{for every } \omega \in \Omega.$$

Therefore $p(A) = \sum_{\omega\in A} p(\omega)$ for every $A \subseteq \Omega$. Thus every probability can be considered to belong to the set $\mathcal{F}_1$. We also consider sets of probabilities, which we usually assume to be closed and convex. Sometimes we assume an enumeration of elements of Ω and for short denote, for instance, $f_i = f(\omega_i)$.

There is a one-to-one correspondence between closed convex sets of probabilities and the corresponding *lower* and *upper expectation operators*. We denote the lower expectation operator of a closed convex set of probabilities $\mathcal{M}$ by $\underline{P}$ and the upper expectation operator by $\overline{P}$. So for any $f \in \mathcal{F}$ we define:

$$\underline{P}(f) = \min_{p\in\mathcal{M}} E_p f \tag{1}$$

and

$$\overline{P}(f) = \max_{p\in\mathcal{M}} E_p f. \tag{2}$$

The min and max in the above equations can be written because of the finiteness of the probability space which assures that all closed sets of probabilities are compact and therefore all minima and maxima exist. In the case of the above correspondence between a set of probabilities and an expectation operator we say that $\mathcal{M}$ is a *credal set* of $\underline{P}$ and we may denote

$$\mathcal{M} = \mathcal{M}(\underline{P}).$$

Every lower expectation operator $\underline{P}$ has the following properties. Let f, f_1, f_2 be arbitrary elements from $\mathcal{A}$. Then:

superadditivity: $\underline{P}(f_1 + f_2) \geq \underline{P}(f_1) + \underline{P}(f_2)$;

non-negative homogeneity: $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ for every $\lambda \geq 0$;

constant additivity: $\underline{P}(f + \mu 1_\Omega) = \underline{P}(f) + \mu$ for every real μ.

Further we note that any expectation operator is completely determined by its values on the space $\mathcal{F}_1$. To see this take any map $f \in \mathcal{F}$ and define the corresponding $\tilde{f} \in \mathcal{F}_1$ with

$$\tilde{f} = \frac{f}{2||f||_\infty} + \frac{1}{2}1_\Omega,$$

if $||f||_\infty > 0$, and $\tilde{f} = \frac{1}{2}1_\Omega$ otherwise. The value $\tilde{a} = \underline{P}(\tilde{f})$ then determines

$$\underline{P}(f) = \left(\tilde{a} - \frac{1}{2}\right) \cdot 2||f||_\infty,$$

as follows from non-negative homogeneity and constant additivity.

3 Distance measures between imprecise probabilities

The set of probability measures on a measurable space $(\Omega, \mathcal{A})$ can be metricised using the following metric:

$$d(p, p') = \max_{A\in\mathcal{A}} |p(A) - p'(A)| = \frac{1}{2} \sum_{\omega\in\Omega} |p(\omega) - p'(\omega)|, \tag{3}$$

for every pair of probability measures p and p'.

Given a metric space M and non-empty compact subsets $X, Y \subset M$ the *Hausdorff metric* (see e.g. [1]) is defined as

$$d_H(X,Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\right\}. \tag{4}$$

This metric makes the set of non-empty compact sets a metric space denoted by $F(M)$. Moreover, if M is a compact space, so is $F(M)$. Note also that every compact metric space is complete. The Hausdorff distance can be applied to the family of compact sets of probabilities using the distance function (3) in (4), making it, in the case of a finite space, a complete metric space.

Let $\underline{P}$ and $\underline{P}'$ be lower expectation operators. Then we define the following distance between them:

$$\tilde{d}(\underline{P}, \underline{P}') = \max_{f\in\mathcal{F}_1} |\underline{P}(f) - \underline{P}'(f)|. \tag{5}$$

Because of the finiteness of Ω the max in the above equation exists. If f is any non-negative real map on Ω then we have that $\tilde{f} = \dfrac{f}{||f||_\infty} \in \mathcal{F}_1$. Because of positive homogeneity of lower expectation operator we conclude that

$$|\underline{P}_1(f) - \underline{P}_2(f)| \leq \tilde{d}(\underline{P}_1, \underline{P}_2)||f||_\infty. \tag{6}$$

The next proposition shows that the metrics (5) and (3) coincide for probability measures. Therefore, from now on we denote both distances with d.

Proposition 1. *Let p and p' be probability measures on $(\Omega, \mathcal{A})$. Then we have that*

$$\max_{f\in\mathcal{F}_1} |E_p f - E_{p'} f| = d(p, p').$$

Proof. Define the function

$$F(\omega) = \begin{cases} 1, & p(\omega) \geq p'(\omega); \\ 0, & \text{otherwise.} \end{cases}$$

For any real function $f \in \mathcal{F}_1$ we have

$$\begin{aligned} |E_p f - E_{p'} f| &= \left|\sum_i (p_i - p'_i) f_i\right| \\ &\leq \left|\sum_i (p_i - p'_i) F_i\right| \\ &= \max_{A\subset\Omega} |p(A) - p'(A)| \\ &= d(p, p'). \end{aligned}$$

□

The following theorem shows that the metric (5) between lower expectation operators coincides with the Hausdorff metric between their credal sets. (A similar result can be found in [6], Lemma 6.7.)

Theorem 1. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be closed convex sets of probabilities and let $\underline{P}_1$ and $\underline{P}_2$ be their lower expectation operators. Then we have that*

$$d(\underline{P}_1, \underline{P}_2) = d_H(\mathcal{M}_1, \mathcal{M}_2). \tag{7}$$

Proof. First we show that for any probabilities p_1 and p_2 we have that

$$\max_{f\in\mathcal{F}_1} |E_{p_1} f - E_{p_2} f| = \max_{f\in\mathcal{F}_1} E_{p_1} f - E_{p_2} f. \tag{8}$$

This follows from the fact that $f \in \mathcal{F}_1$ implies $1_\Omega - f \in \mathcal{F}_1$ and $E_{p_1} f - E_{p_2} f = -(E_{p_1}(1-f) - E_{p_2}(1-f))$ which implies

$$\begin{aligned} \max_{f\in\mathcal{F}_1} |E_{p_1} f - E_{p_2} f| &= \max_{f\in\mathcal{F}_1} \max\{E_{p_1} f - E_{p_2} f, \\ &\qquad E_{p_1}(1-f) - E_{p_2}(1-f)\} \\ &= \max_{f\in\mathcal{F}_1} E_{p_1} f - E_{p_2} f. \end{aligned}$$

The definition of the Hausdorff distance and the equation (8) implies that

$$d_H(\mathcal{M}_1, \mathcal{M}_2) = \max_{p_1\in\mathcal{M}_1} \min_{p_2\in\mathcal{M}_2} \max_{f\in\mathcal{F}_1} E_{p_1} f - E_{p_2} f \tag{9}$$

or in the last expression the roles of $\mathcal{M}_1$ and $\mathcal{M}_2$ can be exchanged, and that case would be treated equally because of symmetry. Now fix any $p_1 \in \mathcal{M}_1$ and consider the map:

$$\Gamma\colon \mathcal{M}_2 \times \mathcal{F}_1 \to \mathbb{R}$$

where

$$(p_2, f) \mapsto E_{p_1} f - E_{p_2} f.$$

Now the set $\mathcal{M}_2$ is compact by definition, and the mapping $p_2 \mapsto \Gamma(p_2, f)$ is continuous and linear, therefore also convex, for any fixed $f \in \mathcal{F}_1$. Furthermore for a fixed p_2 the mapping $f \mapsto \Gamma(p_2, f)$ is also linear, and therefore concave. Now we can use the minimax theorem (see [5]: Theorem 2) to obtain:

$$\min_{p_2\in\mathcal{M}_2} \max_{f\in\mathcal{F}_1} \Gamma(p_2, f) = \max_{f\in\mathcal{F}_1} \min_{p_2\in\mathcal{M}_2} \Gamma(p_2, f).$$

That is

$$\min_{p_2\in\mathcal{M}_2} \max_{f\in\mathcal{F}_1} E_{p_1} f - E_{p_2} f = \max_{f\in\mathcal{F}_1} \min_{p_2\in\mathcal{M}_2} E_{p_1} f - E_{p_2} f.$$

Using the above equality we obtain:

$$\begin{aligned}
&\max_{p_1\in\mathcal{M}_1}\min_{p_2\in\mathcal{M}_2} d(p_1,p_2)\\
&\quad= \max_{p_1\in\mathcal{M}_1}\min_{p_2\in\mathcal{M}_2}\max_{f\in\mathcal{F}_1} E_{p_1}f - E_{p_2}f\\
&\quad= \max_{p_1\in\mathcal{M}_1}\max_{f\in\mathcal{F}_1}\min_{p_2\in\mathcal{M}_2} E_{p_1}f - E_{p_2}f\\
&\quad= \max_{f\in\mathcal{F}_1}\max_{p_1\in\mathcal{M}_1}\min_{p_2\in\mathcal{M}_2} E_{p_1}f - E_{p_2}f\\
&\quad= \max_{f\in\mathcal{F}_1} \overline{P}_1(f) - \overline{P}_2(f)\\
&\quad= \max_{f\in\mathcal{F}_1} \overline{P}_1(1-f) - \overline{P}_2(1-f)\\
&\quad= \max_{f\in\mathcal{F}_1} \underline{P}_2(f) - \underline{P}_1(f).
\end{aligned}$$

Finally, using this and the symmetry between $\mathcal{M}_1$ and $\mathcal{M}_2$, we get

$$\begin{aligned}
d_H(\mathcal{M}_1,\mathcal{M}_2) &= \max\{\max_{p_1\in\mathcal{M}_1}\min_{p_2\in\mathcal{M}_2} d(p_1,p_2),\\
&\qquad \max_{p_2\in\mathcal{M}_2}\min_{p_1\in\mathcal{M}_1} d(p_1,p_2)\}\\
&= \max_{f\in\mathcal{F}_1}\{\underline{P}_2(f)-\underline{P}_1(f), \underline{P}_1(f)-\underline{P}_2(f)\}\\
&= \max_{f\in\mathcal{F}_1} |\underline{P}_1(f)-\underline{P}_2(f)|\\
&= d(\underline{P}_1,\underline{P}_2),
\end{aligned}$$

which completes the proof. □

We will also need the maximal distance between probability measures belonging to a pair of credal sets $\mathcal{M}_1$ and $\mathcal{M}_2$ with the corresponding lower and upper expectation operators $\underline{P}_1,\overline{P}_1$ and $\underline{P}_2,\overline{P}_2$ respectively. Using Proposition 1 we have that

$$\begin{aligned}
\max_{\substack{p_1\in\mathcal{M}_1\\ p_2\in\mathcal{M}_2}} d(p_1,p_2) &= \max_{\substack{p_1\in\mathcal{M}_1\\ p_2\in\mathcal{M}_2}}\max_{f\in\mathcal{F}_1} |E_{p_1}f - E_{p_2}f|\\
&= \max_{f\in\mathcal{F}_1}\max_{\substack{p_1\in\mathcal{M}_1\\ p_2\in\mathcal{M}_2}} |E_{p_1}f - E_{p_2}f|\\
&= \max_{f\in\mathcal{F}_1}\max\{\overline{P}_1(f)-\underline{P}_2(f),\\
&\qquad \overline{P}_2(f)-\underline{P}_1(f)\}.
\end{aligned}$$

However, instead of taking the maxima over the whole $\mathcal{F}_1$ in the above equation it would be enough to only consider characteristic functions of subsets of Ω, as follows from Proposition 1. Therefore

$$\max_{\substack{p_1\in\mathcal{M}_1\\ p_2\in\mathcal{M}_2}} d(p_1,p_2) = \max_{A\subset\Omega}\max\{\overline{P}_1(1_A)-\underline{P}_2(1_A), \overline{P}_2(1_A)-\underline{P}_1(1_A)\}.$$

It follows that for any pair of lower and upper expectation operators $\underline{P}_1$ and $\overline{P}_2$ we have that

$$\max_{f\in\mathcal{F}_1}\{\overline{P}_2(f)-\underline{P}_1(f)\} = \max_{A\subset\Omega}\{\overline{P}_2(1_A)-\underline{P}_1(1_A)\}. \tag{10}$$

We will also need some results on convergence of lower expectation operators. We study the convergence in the metric (5). In proving the convergence results we will use the result that any decreasing sequence of non-empty compact sets is non-empty (see [4]: Lemma I.5.6).

Proposition 2. *Let $\{\underline{P}_n\}_{n\in\mathbb{N}}$ be an increasing sequence of lower expectation operators and $\{\mathcal{M}_n\}_{n\in\mathbb{N}}$ the sequence of the corresponding credal sets. Then the sequence $\{\mathcal{M}_n\}_{n\in\mathbb{N}}$ is decreasing with respect to set inclusion and the limit*

$$\underline{P}_\infty = \lim_{n\to\infty}\underline{P}_n$$

exists and

$$\mathcal{M}(\underline{P}_\infty) = \bigcap_{n\in\mathbb{N}}\mathcal{M}_n.$$

Moreover, the above credal set is non-empty.

Proof. For every $f\in\mathcal{F}_1$ we have that the sequence $\{\underline{P}_n(f)\}$ is an increasing sequence bounded from above by 1 and is therefore convergent. Now take any $p\in\bigcap_{n\in\mathbb{N}}\mathcal{M}_n$. Then by definition, for every $f\in\mathcal{F}_1$ we have that $E_pf\ge\underline{P}_\infty(f)$, so $\bigcap_{n\in\mathbb{N}}\mathcal{M}_n\subseteq\mathcal{M}(\underline{P}_\infty)$. To see the converse inclusion take any probability p such that $E_pf\ge\underline{P}_\infty(f)\ge\underline{P}_n$ for every $n\in\mathbb{N}$. Therefore $p\in\mathcal{M}_n$ for every $n\in\mathbb{N}$ and every $f\in\mathcal{F}_1$, which implies that $p\in\bigcap_{n\in\mathbb{N}}\mathcal{M}_n$. Thus, $\mathcal{M}(\underline{P}_\infty)\subseteq\bigcap_{n\in\mathbb{N}}\mathcal{M}_n$. As follows from the above remark, the set $\bigcap_{n\in\mathbb{N}}\mathcal{M}_n$ is non-empty. □

Proposition 3. *Let $\{\underline{P}_n\}_{n\in\mathbb{N}}$ be any convergent sequence of lower expectation operators and $\{\mathcal{M}_n\}_{n\in\mathbb{N}}$ the sequence of the corresponding credal sets. Then the set*

$$\mathcal{M}_\infty = \bigcap_{n\in\mathbb{N}}\mathrm{co}\Big(\overline{\bigcup_{m\ge n}\mathcal{M}_m}\Big),$$

where co *denotes the convex hull, is the credal set of the limit lower expectation operator $\underline{P}_\infty = \lim_{n\to\infty}\underline{P}_n$. Moreover, the set $\mathcal{M}_\infty$ is non-empty and therefore the lower expectation operator $\underline{P}_\infty$ is well defined.*

Proof. First we define the following sequence of lower expectation operators:

$$\tilde{\underline{P}}_n = \inf_{m\ge n}\underline{P}_m.$$

Clearly, the convergence of the sequence $\{\underline{P}_n\}$ implies the convergence of $\{\tilde{\underline{P}}_n\}$ with the same limit. We only need to see that the credal set of $\tilde{\underline{P}}_n$ is $\mathrm{co}(\overline{\bigcup_{m\ge n}\mathcal{M}_m})$.

To see this take any convergent sequence $\{p_r\}$ in $\bigcup_{m\ge n}\mathcal{M}_m$. For every $f\in\mathcal{F}$ we have that $E_{p_r}f\ge\tilde{\underline{P}}_n(f)$ and therefore $\lim_{r\to\infty}E_{p_r}f = E_{\lim_{r\to\infty}p_r}f\ge$

$\tilde{\underline{P}}_n(f)$, and thus $\lim_{r\to\infty} p_r$ belongs to the credal set of $\tilde{\underline{P}}_n(f)$. Further, given any $f \in \mathcal{F}_1$ there is some $p_r \in \mathcal{M}_m$, for $m \geq n$ so that $E_{p_r} f \leq \tilde{\underline{P}}_n(f) + \frac{1}{r}$. Since the set of all probabilities on a finite set is compact, the sequence $\{p_r\}$ has a convergent subsequence converging to a probability p and $E_p f = \tilde{\underline{P}}_n(f)$. Thus, $\tilde{\underline{P}}_n$ is the lower expectation operator of the set $\bigcup_{m\geq n} \mathcal{M}_m$ which implies that its closure is the credal set of $\tilde{\underline{P}}_n$.

To finish the proof we apply Proposition 2 to the increasing sequence $\{\tilde{\underline{P}}_n\}$ and the corresponding credal sets $\mathrm{co}(\bigcup_{m\geq n} \mathcal{M}_n)$. □

Corollary 1. *The set of all lower expectation operators is complete in the metric* (5)*.*

4 Imprecise Markov chains

One of the most natural ways to involve imprecision in a probabilistic model is to allow a set of possible probability distributions instead of a single one. In the case of Markov chains such sets can be allowed in place of transition probabilities as well as initial probability distributions. Additionally, we usually assume such sets are closed and convex. This assumption is particularly useful because, as described in Section 2, the sets can be equivalently described using lower or upper expectation operators. There are of course many models that allow description of sets of probabilities, such as *interval probabilities* (see e.g. [20]) or *lower* and *upper previsions* (see e.g. [18, 19]).

The most basic form used in most of the approaches taken until now is to put constraints, usually in the form of intervals, on the probabilities belonging to the elementary sets (see [8], [11]). The imprecision concerning the initial distribution is thus presented through the intervals $[p_i, q_i]$ which are supposed to contain the unknown initial probability $P(X_0 = i)$. Similarly, the probabilities of transition from the state i to j are given in the form of intervals $[p_{ij}, q_{ij}]$ supposed to contain the unknown true transition probability $P(X_{n+1} = j|X_n = i)$. Even though the true probabilities are unknown, it is certain that the sum of all probabilities is 1. Thus the values within the intervals must be taken so that they sum to 1, or in the case of transition interval matrices, all rows must sum to 1. An additional assumption that is usually made about the intervals is that all values within the interval are reachable or, in particular, that the interval bounds are reachable. In the common terminology of imprecise probabilities this requirement is named *coherence*. To each set of intervals, the set of probabilities assuming their values within those intervals can be assigned.

One of the crucial differences between precise and imprecise probabilities is that a precise probability can be fully determined by far less information than an imprecise probability. Thus to determine any precise probability, only its values on elementary sets are needed to be found, while the sets of probabilities able to be represented via simple intervals described above is fairly limited. (Many examples can be found e.g. in [20], [19], [18].) Another difference compared to the classical model is that transition probabilities that govern transitions of a Markov chain in the imprecise case may change in time. Thus, we are dealing with possibly non-homogeneous chains, which consequently require considering non-homogeneous matrix products.

Now we introduce the terminology used to describe imprecise Markov chains. We will assume a non-empty set Ω whose elements are called *states*. For simplicity we will assume they are the consecutive integers $1, \ldots, m$, since in the basic model their values have no special consequences. We will follow the approach similar to the one taken by de Cooman et al. [2] to describe the sets of probabilities using the corresponding expectation operators, usually this will mean lower expectation operators.

We will thus assume a set $\mathcal{M}_0$ of *initial probability distributions* and let $\underline{P}_0$ be its lower expectation operator (see (1)). Further, we assume a set of transition matrices $\mathcal{P}$, whose rows are *separately specified*, i.e. for any two transition matrices p and p' with rows p_i and p'_i replacing the ith row of p with p'_i results in a matrix that still belongs to $\mathcal{P}$. By adopting this property we can associate row sets of distributions $\mathcal{P}_i$ to $\mathcal{P}$ so that any independent choice of rows from the row sets gives a transition matrix in $\mathcal{P}$. If additionally we assume that row sets are closed and convex, we have the following important property.

Lemma 1. *Let $\mathcal{P}$ be a convex set of transition matrices with separtely specified rows and let $\mathcal{M}$ be a convex set of probabilities. Then the set of probability distributions at the next step $\mathcal{M} \cdot \mathcal{P}$ is a convex set.*

We slightly modify the proof of [8]: Lemma 2.5.

Proof. We prove the lemma by showing that given the probabilities q and $q' \in \mathcal{M}$ and transition matrices p and $p' \in \mathcal{P}$ then, whenever $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$,

$$(\alpha q \cdot p + \beta q' \cdot p') = (\alpha q + \beta q')r \tag{11}$$

with $r \in \mathcal{P}$.

Take $j \in \Omega$. We have

$$\begin{aligned}(\alpha q \cdot p + \beta q' \cdot p')_j &= \alpha \sum_{i=1}^{m} q_i p_{ij} + \beta \sum_{i=1}^{m} q'_i p'_{ij} \\ &= \sum_{i=1}^{m} (\alpha q_i p_{ij} + \beta q'_i p'_{ij}) \\ &= \sum_{i=1}^{m} (\alpha q_i + \beta q'_i) \left(\frac{\alpha q_i}{\alpha q_i + \beta q'_i} p_{ij} + \frac{\beta q'_i}{\alpha q_i + \beta q'_i} p'_{ij} \right).\end{aligned}$$

Thus taking r with $r_{ij} = \frac{\alpha q_i}{\alpha q_i + \beta q'_i} p_{ij} + \frac{\beta q'_i}{\alpha q_i + \beta q'_i} p'_{ij}$ satisfies (11). Notice that ith row of r is a convex combination of some elements of $\mathcal{P}_i$ and therefore itself a member of $\mathcal{P}_i$ too. Now, because rows are separately specified the resulting matrix is also a member of $\mathcal{P}$. □

To each row set of probabilities we associate the lower expectation operator $\underline{T}_i$. Let $\underline{T}$ then be the matrix lower expectation operator whose ith row is $\underline{T}_i$. We will say that the set $\mathcal{P}$ is the *credal set* of $\underline{T}$.

Let $X_0, X_1, \ldots, X_n, \ldots$ be a sequence of random variables assuming the values in Ω. According to the given assumptions we have

$$P(X_0 = i) = q_i^0,$$

where $q^0 \in \mathcal{M}_0$. The role of the transition matrices is given by

$$P(X_{n+1} = j | X_n = i) = p_{ij}^n,$$

where $p^n \in \mathcal{P}$.

A basic feature of the theory of Markov chains is the ability to calculate the probability of being in some state j at time n given an initial probability. Of course, since the initial and transition probabilities are imprecise, the answer will also be given in the form of an imprecise probability, that is, in the form of a set of probabilities. Previous works such as Hartfiel's [8] provide the general answer to this question based on the classical theory. The set of possible probability distributions at step n is equal to the set of all possible initial distributions multiplied by all possible sequences of transition matrices. Let $\mathcal{M}_n$ denote the set of possible probability distributions at step n given the initial distribution $\mathcal{M}_0$. Then we have

$$\begin{aligned}\mathcal{M}_n = \{ q \cdot p_1 \cdot \ldots \cdot p_n \mid q \in \mathcal{M}_0, p_i \in \mathcal{P} \\ \text{for every } i = 1, \ldots, n\} = \mathcal{M}_{n-1} \cdot \mathcal{P}. \quad (12)\end{aligned}$$

It follows from Lemma 1 that in the case where the set of transition matrices $\mathcal{P}$ has closed convex separately specified row sets, every $\mathcal{M}_n$ is also a closed convex set of probabilities. Therefore, they can be equivalently represented using lower expectation operators. The lower expectation operator corresponding to the set $\mathcal{M}_n$ is denoted by $\underline{P}_n$.

To calculate the values of $\underline{P}_n$ on real functions on Ω we follow the approach proposed in [2]. They first calculate the nth power of the transition operator $\underline{T}$ using so-called backwards recursion. This method can be described in the following way. Let f be any real valued map on Ω. Every expectation operator assigns to it a real number corresponding to the lower expectation. In particular, every row lower expectation operator $\underline{T}_i$ assigns to it the value $\underline{T}_i(f)$. A transition operator $\underline{T}$ thus assign to every f a vector of values

$$\underline{T}(f) = \begin{pmatrix} \underline{T}_1(f) \\ \underline{T}_2(f) \\ \vdots \\ \underline{T}_m(f) \end{pmatrix}. \quad (13)$$

Now $\underline{T}(f)$ is another real valued function on Ω to which a new instance of T can be applied to obtain $\underline{T}^2(f)$ and so on. Finally, applying $\underline{P}_0$ to $T^n(f)$ gives exactly the lower expectation of the lower expectation operator $\underline{P}_n$ corresponding to the set $\mathcal{M}_n$. For the proof see [2].

Once probabilities of states on different steps are calculated, we are often interested in the limiting behaviour of these probabilities. Thus, the question is what can be said about the probability $P(X_n = i)$ for a large n and how does it depend on the initial distribution? In the classical theory, Perron-Frobenius theorem assures convergence for the class of regular Markov chains (a Markov chain with the transition matrix p is *regular* if for some positive integer r the power p^r has only strictly positive entries). The Perron-Frobenius theorem states that the probabilities $p_i^{(n)} = P(X_n = i)$ converge to some unique limit probabilities independently on the initial distribution.

Regularity is therefore a sufficient condition for unique convergence of a Markov chain, but not also a necessary one. This is true already in the case of precise Markov chains, where a more general criteria are derived using *coefficients of ergodicity* that besides telling whether a chain is convergent also measure the rate of convergence (see e.g. Seneta [15]). Hartfiel [8] then applies a generalised coefficient of ergodicity to study the convergence of Markov set-chains. Recently, de Cooman et al. [2] find that the conditions applied by Hartfiel are in general too strong to assure the convergence of imprecise Markov chains. They define a class of *regularly absorbing* imprecise Markov chains, based on the accessibility between states, for which they show unique convergence.

5 Coefficients of ergodicity

Coefficients of ergodicity or *contraction coefficients* measure the rate of convergence of Markov chains. Seneta in his paper [15] defines a general coefficient of ergodicity for a stochastic matrix p with no zero columns to be

$$\tau(p) = \sup_{x,y} \frac{d(xp, yp)}{d(x, y)}$$

where d is some metric on the set of vectors with positive coordinates and whose components sum to 1 and x, y are such vectors. The value of $\tau(p)$ is between 0 and 1 and further τ has the following properties:

(i) $\tau(p_1p_2) \le \tau(p_1)\tau(p_2)$ for every pair of stochastic matrices with no zero columns p_1 and p_2;

(ii) $\tau(p) = 0$ whenever rank of p is 1 i.e. $p = \mathbf{1}v$ for some vector v.

Depending on the metrics, different coefficients of ergodicity are used. In this paper we are concerned with the coefficient generated by the metric (3). This coefficient was introduced by Dobrushin [3] and its direct evaluation is derived by Paz [13]:

$$\tau(p) = \frac{1}{2} \max_{i,j} \sum_{s=1}^{m} |p_{is} - p_{js}|.$$

In view of (3), the above can be stated as

$$\tau(p) = \max_{i,j} d(p_i, p_j). \tag{14}$$

where p_i and p_j denote the ith and jth row of p respectively.

For the case of imprecise Markov chains, Hartfiel [8] extends the concept of a coefficient of ergodicity to Markov chains where sets of transition probabilities are considered. For a set of transition matrices $\mathcal{P}$ he defines the *uniform coefficient of ergodicity* as

$$\tau(\mathcal{P}) = \sup_{p\in\mathcal{P}} \tau(p).$$

If $\mathcal{P}$ is an interval $[P, Q]$, i.e. $\mathcal{P} = \{p \mid p \text{ is a stochastic matrix such that } P \le p \le Q\}$, then he finds that

$$\tau(\mathcal{P}) \le \frac{1}{2} \max_{i,j} \sum_{k=1}^{m} \max\{|q_{ik} - p_{jk}|, |q_{jk} - p_{ik}|\}.$$

where p_{ik} and q_{ik} are the components of P and Q respectively.

In our setting of lower and upper expectation operators, the calculation of the uniform coefficient of ergodicity is given by the following proposition.

Proposition 4. *Let $\mathcal{P}$ be a set of transition matrices and let $\underline{T}$ and $\overline{T}$ be its lower and upper expectation matrices. Then we have that*

$$\begin{aligned}\tau(\mathcal{P}) &= \max_{i,j} \max_{f\in\mathcal{F}_1} \overline{T}_i(f) - \underline{T}_j(f)\\ &= \max_{i,j} \max_{A\subset\Omega} \overline{T}_i(1_A) - \underline{T}_j(1_A).\end{aligned}$$

Proof. The second equality follows from (10). Let $p \in \mathcal{P}$ be arbitrary transition matrix. Then its ith and jth row are arbitrary probability distributions belonging to the credal sets of ith and jth row of $\mathcal{P}$. We have that

$$\begin{aligned}\tau(\mathcal{P}) &= \max_{p\in\mathcal{P}} \tau(p)\\ &= \max_{i,j} \max_{\substack{p_i\in\mathcal{M}_i\\ p_j\in\mathcal{M}_j}} d(p_i, p_j)\\ &= \max_{i,j} \max_{A\subset\Omega} \max\{\overline{T}_i(1_A) - \underline{T}_j(1_A),\\ &\qquad \overline{T}_j(1_A) - \underline{T}_i(1_A)\}\\ &= \max_{i,j} \max_{A\subset\Omega} \overline{T}_i(1_A) - \underline{T}_j(1_A),\end{aligned}$$

as required. □

Thus, we may define $\tau(\underline{T}) = \tau(\mathcal{M}(\underline{T}))$.

The uniform coefficient of ergodicity can be used as a contraction measure for a set of transition matrices. The following theorem holds ([8]: Theorem 3.3):

Theorem 2. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be non-empty compact sets of probabilities. Then*

$$d_H(\mathcal{M}_1 \cdot \mathcal{P}, \mathcal{M}_2 \cdot \mathcal{P}) \le \tau(\mathcal{P}) d_H(\mathcal{M}_1, \mathcal{M}_2).$$

A stochastic matrix p whose coefficient of ergodicity $\tau(p)$ is strictly smaller than 1 is called *scrambling* (see [15]). Further if $\mathcal{P}$ is a set of transition matrices such that $\tau(p_1 \cdot p_2 \cdots p_r) < 1$ for any matrices $p_i \in \mathcal{P}$ then such a set is called *product scrambling* (see [8]), and r is then called its *scrambling integer*. Thus we have that $\tau(\mathcal{P}^r) < 1$. Something very similar can be said about lower expectation matrices. We will say that a lower expectation matrix $\underline{T}$ is scrambling whenever $\tau(\underline{T}) < 1$ and if instead only $\tau(\underline{T}^r) < 1$ we will say that it is product scrambling with scrambling integer r.

Theorem 2 implies the following more general corollary ([8]: Theorem 3.4):

Corollary 2. *Let $\mathcal{P}$ be be product scrambling with scrambling integer r and let $\mathcal{M}_0$ be a non-empty compact set of probabilities. Then, for any positive integer h,*

$$d_H(\mathcal{M}_0\mathcal{P}^h, \mathcal{M}_\infty) \le K\beta^h$$

where $K = \tau(\mathcal{P}^r)^{-1} d_H(\mathcal{M}_0, \mathcal{M}_\infty)$ *and* $\beta = \tau(\mathcal{P}^r)^{\frac{1}{r}} < 1$ *and* $\mathcal{M}_\infty$ *is the unique compact set of probabilities such that*

$$\mathcal{M}_\infty \mathcal{P} = \mathcal{M}_\infty.$$

Thus,

$$\lim_{h\to\infty} \mathcal{M}_0 \mathcal{P}^h = \mathcal{M}_\infty.$$

Theorem 2 implies the convergence of a Markov set-chain in the Hausdorff metric. Moreover, if $\tau(\mathcal{P}) < 1$ for a set of transition matrices then given any initial probability distribution q_0 and a sequence of transition matrices $\{p_i\}_{i\in\mathbb{N}}$ such that every $p_i \in \mathcal{P}$ we have that the sequence $q_n = q_0 p_1 \cdots p_n$ converges to some p_∞. This is a consequence of the fact that $\tau(p_1 \cdots p_n) \to 0$ as n tends to infinity. Moreover, since clearly $\tau(\mathcal{P}') \le \tau(\mathcal{P})$ for every $\mathcal{P}' \subseteq \mathcal{P}$, it follows that given a convergent Markov chain with the set of transition probabilities $\mathcal{P}$ then a Markov chain with the set of transition probabilities $\mathcal{P}'$ is also convergent.

De Cooman et al. [2] show that it not necessary to require that every possible transition matrix is a contraction, but instead, what is needed is only that the corresponding upper (or lower) expectations are becoming more and more similar. As a simple demonstration consider the following example.

Example 1. Let a set of transition matrices on the set $\Omega = \{1, 2\}$ be given by the following lower and upper transition matrix

$$P = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Clearly this set contains the matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

which is not contractive. However, given any initial set of distributions the Markov chain with the above set of transition matrices converges to the set of all probability distributions on Ω.

De Cooman et al. further find sufficient conditions for unique convergence by studying the accessibility relation between states. Our aim here is to find a coefficient of ergodicity that would describe this type of convergence for imprecise Markov chains. We implement the following idea. Given a lower transition matrix $\underline{T}$, the backwards recursion allows the calculation of its powers $\underline{T}^n$ for every positive integer n. In the case of a precise transition matrix, the rows of its consequent powers get more and more similar, which is measured by the coefficient of ergodicity (14). In the case of a lower expectation matrix, the same effect will be achieved by measuring the distances between the row lower expectation operators corresponding to the powers of $\underline{T}$.

Definition 1. Let $\underline{T}$ be a transition lower expectation matrix. Then we define the *weak coefficient of ergodicity* as

$$\rho(\underline{T}) = \max_{\substack{f\in\mathcal{F}_1 \\ i,j}} |\underline{T}_i(f) - \underline{T}_j(f)|,$$

where $\underline{T}_i$ and $\underline{T}_j$ are ith and jth row lower expectation operators respectively.

The following proposition is an immediate consequence of the definitions.

Proposition 5. *Let* $\underline{T}$ *be a transition lower expectation matrix with rows* $\underline{T}_i$*. Then:*

$$\rho(\underline{T}) = \max_{i,j} d(\underline{T}_i, \underline{T}_j).$$

Proposition 6. *Let* $\underline{P}_1$ *and* $\underline{P}_2$ *be lower expectation operators and* $\underline{T}$ *a transition lower expectation matrix. Then we have that*

$$d(\underline{P}_1\underline{T}, \underline{P}_2\underline{T}) \le \rho(\underline{T}) d(\underline{P}_1, \underline{P}_2).$$

Proof. Denote $c_f = \underline{T}(f)$ (see (13)) and let $\underline{c}_f$ and $\overline{c}_f$ be its minimal and maximal element respectively. Further let $\tilde{\underline{P}}_1 = \underline{P}_1\underline{T}$ and $\tilde{\underline{P}}_2 = \underline{P}_2\underline{T}$. Then using constant additivity and (6) we obtain

$$\begin{aligned}
|\tilde{\underline{P}}_1(f) - \tilde{\underline{P}}_2(f)| &= |\underline{P}_1(c_f) - \underline{P}_2(c_f)| \\
&= |\underline{P}_1((c_f - \underline{c}_f) + \underline{c}_f) \\
&\qquad - \underline{P}_2((c_f - \underline{c}_f) + \underline{c}_f)| \\
&\le d(\underline{P}_1, \underline{P}_2) ||c_f - \underline{c}_f||_\infty \\
&= d(\underline{P}_1, \underline{P}_2)(\overline{c}_f - \underline{c}_f) \\
&\le d(\underline{P}_1, \underline{P}_2)\rho(\underline{T})
\end{aligned}$$

□

Corollary 3. *Let* $\underline{R}$ *and* $\underline{S}$ *be any transition lower expectation matrices. Then:*

$$\rho(\underline{RS}) \le \rho(\underline{R})\rho(\underline{S}).$$

Proof. Denote $\underline{T} = \underline{RS}$ and let $\underline{T}_i$ and $\underline{T}_j$ be the ith and jth row lower expectation operators. We have that, for instance,

$$\underline{T}_i(f) = \underline{R}_i\underline{S}(f).$$

Proposition 6 then yields

$$\begin{aligned}
|\underline{T}_i(f) - \underline{T}_j(f)| &= |\underline{R}_i\underline{S}(f), \underline{R}_j\underline{S}(f)| \\
&\le d(\underline{R}_i, \underline{R}_j)\rho(\underline{S}) \\
&\le \rho(\underline{R})\rho(\underline{S}),
\end{aligned}$$

as required. □

The next corollary is now immediate.

Corollary 4. *For any lower expectation operator $\underline{T}$ we have that*

$$\rho(\underline{T}^n) \leq \rho(\underline{T})^n.$$

Thus, it may happen that even if $\rho(\underline{T}) = 1$ it may be that $\rho(\underline{T}^n) < 1$.

The following proposition shows that the credal set of a contractive lower expectation operator contains at least one contractive transition matrix. The converse does not hold, as demonstrated by the example following the proposition.

Proposition 7. *Let $\underline{T}$ be a transition lower expectation matrix such that $\rho(\underline{T}) < 1$. Then there exists a precise transition matrix $p \in \mathcal{M}(\underline{T})$ such that $\tau(p) < 1$.*

Proof. Denote $\rho := \rho(\underline{T})$. Then for any pair of indices i and j we have $d(\underline{T}_i, \underline{T}_j) \leq \rho$. Coherence of $\underline{T}$ implies that for every set $A \subset \Omega$ we have a probability measure p^A such that $p_i^A(A) = \underline{T}(1_A)$ for every $1 \leq i \leq m$. Then $|p_i^A(A) - p_j^A(A)| < 1$ and $|p_i^A(A') - p_j^A(A')| \leq 1$ for any $A' \subset \Omega$. Let $\lambda_A > 0$ for every $A \subset \Omega$ and let $\sum_{A\subset\Omega} \lambda_A = 1$. Let $p = \sum_{A\subset\Omega} \lambda_A p^A$. Clearly then $p_i(A) - p_j(A) < 1$ for every $A \subset \Omega$ and thus $\tau(p) < 1$. □

Example 2. Let the lower expectation operator $\underline{T} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ be given. Thus the credal set of $\underline{T}$ contains all possible stochastic matrices with the first row equal to $(1, 0)$. Clearly, the weak coefficient of ergodicity of $\underline{T} = \underline{T}^n$, for every $n \in \mathbb{N}$, is equal to 1; however, the credal set contains, for instance, the matrix $p = \begin{pmatrix} 1 & 0 \\ 0.5 & 0.5 \end{pmatrix}$, whose coefficient of ergodicity is equal to 0.5.

Proposition 8. *Let $\underline{T}$ be a transition lower expectation matrix such that $\rho(\underline{T}) < 1$. Then there exists a lower expectation operator $\underline{P}_\infty$ satisfying the property:*

$$\underline{P}_\infty \underline{T} = \underline{P}_\infty. \tag{15}$$

We will call a lower expectation operator satisfying the property (15) an *invariant lower expectation operator* for a transition lower expectation matrix $\underline{T}$.

Proof. Consider the sequence $\underline{P}_n = \underline{P}_0 \underline{T}^n$. We will show that it is a Cauchy sequence in the metric (5). To see this, take some positive integers m and n with $m > n$. Using the fact that $d(\underline{P}, \underline{P}') \leq 1$ for any pair of expectation operators, we have that

$$\begin{aligned} d(\underline{P}_n, \underline{P}_m) &= d(\underline{P}_0\underline{T}^n, \underline{P}_0\underline{T}^m) \\ &= d(\underline{P}_0\underline{T}^n, \underline{P}_0\underline{T}^{m-n}\underline{T}^n) \\ &\leq d(\underline{P}_0, \underline{P}_0\underline{T}^{m-n})\rho(\underline{T}^n) \\ &\leq \rho(\underline{T}^n) \\ &\leq \rho(\underline{T})^n, \end{aligned}$$

and since $\rho(\underline{T}) < 1$ it follows that, with n large enough, this distance can be arbitrarily small. Because of the completeness of the set of lower expectation operators (Corollary (1)), the sequence converges to some lower expectation operator $\underline{P}_\infty$. □

Clearly the invariant lower operators of $\underline{T}$ is the same as the one for $\underline{T}^n$, and thus the above result also holds for a transition lower expectation matrix $\underline{T}$ such that $\rho(\underline{T}^n) < 1$.

Theorem 3. *Let $\underline{T}$ be a transition lower expectation matrix with $\rho(\underline{T}) < 1$ and $\underline{P}_0$ an initial lower expectation operator and $\underline{P}_\infty$ the invariant lower expectation operator for $\underline{T}$. Then*

$$d(\underline{P}_0\underline{T}^n, \underline{P}_\infty) \leq d(\underline{P}_0, \underline{P}_\infty)\rho(\underline{T})^n.$$

Therefore,

$$\lim_{n\to\infty} \underline{P}_0\underline{T}^n = \underline{P}_\infty$$

independently of $\underline{P}_0$, and $\underline{P}_\infty$ is thus the unique invariant lower expectation operator for $\underline{T}$.

Proof. Using (15) and Proposition 6 and Corollary 4 we obtain

$$\begin{aligned} d(\underline{P}_0\underline{T}^n, \underline{P}_\infty) &= d(\underline{P}_0\underline{T}^n, \underline{P}_\infty\underline{T}^n) \\ &\leq d(\underline{P}_0, \underline{P}_\infty)\rho(\underline{T})^n. \end{aligned}$$

Now since $\rho(\underline{T}) < 1$ the right hand side converges to 0. □

A corollary analogous to Corollary 2 of the last theorem can also be stated. We extend the notion of scrambling lower expectation matrices to the case where the weak coefficient of ergodicity is used. We will say a lower expectation matrix $\underline{T}$ is weakly scrambling if $\rho(\underline{T}) < 1$ and if $\rho(\underline{T}) = 1$ but $\rho(\underline{T}^r) < 1$ for some positive integer r that it is *weakly product scrambling* with *scrambling integer* r.

Corollary 5. *Let $\underline{T}$ be weakly product scrambling with scrambling integer r and let $\underline{P}_0$ be a lower expectation operator. Then, for any positive integer h,*

$$d(\underline{P}_0\underline{T}^h, \underline{P}_\infty) \leq K\beta^h$$

where $K = \rho(\underline{T}^r)^{-1} d(\underline{P}_0, \underline{P}_\infty)$ and $\beta = \rho(\underline{T}^r)^{\frac{1}{r}}$. Thus,

$$\lim_{k\to\infty} \underline{P}_0\underline{T}^k = \underline{P}_\infty.$$

The type of convergence measured by the weak coefficient of ergodicity is clearly closely related to that described in [2]. This suggests that regularly absorbing and weakly scrambling lower expectation matrices are closely related, if not identical. One of the directions in our future research is therefore to clarify this relation.

Acknowledgement

We thank the referees for their helpful suggestions that helped us to improve our paper.

References

[1] G. Beer. *Topologies on closed and closed convex sets*. Kluwer Academic Publishers, Dordrecht, 1993.

[2] G. de Cooman, F. Hermans, and E. Quaeghebeur. Imprecise Markov chains and their limit behaviour. *Probability in the Engineering and Informational Sciences*, 2009.

[3] R.L. Dobrushin. Central limit theorem for nonstationary Markov chains, I, II. *Theory of Probability and its Applications*, 1(4):329–383, 1956.

[4] N. Dunford and J.T. Schwartz. *Linear Operators. Part I: General Theory*. Wiley, New York, 1988.

[5] Ky Fan. Minimax theorems. *Proceedings of the National Academy of Sciences of the United States of America*, 39:42–47, 1953.

[6] R. Hable. *Data-based decisions under complex uncertainty*. PhD thesis, Ludwig-Maximilians-Universität (LMU) Munich, 2009.

[7] D. Harmanec. Generalizing Markov decision processes to imprecise probabilities. *Journal of Statistical Planning and Inference*, 105:199–213, 2002.

[8] D.J. Hartfiel. *Markov Set-Chains*. Springer-Verlag, Berlin, 1998.

[9] D.J. Hartfiel and E. Seneta. On the theory of Markov set-chains. *Advances in Applied Probability*, 26(4):947–964, 1994.

[10] H. Itoh and K. Nakamura. Partially observable Markov decision processes with imprecise parameters. *Artificial Intelligence*, 171(8–9):453–490, 2007.

[11] I. Kozine and L.V. Utkin. Interval-valued finite Markov chains. *Reliable Computing*, 8(2):97–113, 2002.

[12] A. Nilim and L. El Ghaoui. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53:780–798, 2005.

[13] A. Paz. Ergodic theorems for infinite probabilistic tables. *Annals of Mathematical Statistics*, 41(2):539–550, 1970.

[14] J.K. Satia and R.E. Lave. Markovian decision processes with uncertain transition probabilities. *Operations Research*, 21(3):728–740, 1973.

[15] E. Seneta. Coefficients of ergodicity - structure and applications. *Advances in Applied Probability*, 11(2):270–271, 1979.

[16] D. Škulj. Finite discrete time Markov chains with interval probabilities. In J. Lawry, E. Miranda, A. Bugarín, S. Li, M. A. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *SMPS*, volume 37 of *Advances in Soft Computing*, pages 299–306. Springer, 2006.

[17] D. Škulj. Regular finite Markov chains with interval probabilities. In G. de Cooman, M. Zaffalon, and J. Vejnarová, editors, *ISIPTA'07 - Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications*, pages 405–413. SIPTA, 2007.

[18] P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London, New York, 1991.

[19] P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.

[20] K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung. I: Intervallwahrscheinlichkeit als umfassendes Konzept*. Physica-Verlag, Heidelberg, 2001.

[21] C.C. White and H.K. Eldeib. Markov decision processes with imprecise transition probabilities. *Operations Research*, 42(4):739–749, 1994.