Metrika (2013) 76:107–133
DOI 10.1007/s00184-011-0378-0

Coefficients of ergodicity for Markov chains with uncertain parameters

D. Škulj · R. Hable

Received: 4 June 2010 / Published online: 20 December 2011

Abstract One of the central considerations in the theory of Markov chains is their convergence to an equilibrium. Coefficients of ergodicity provide an efficient method for such an analysis. Besides giving sufficient and sometimes necessary conditions for convergence, they additionally measure its rate. In this paper we explore coefficients of ergodicity for the case of imprecise Markov chains. The latter provide a convenient way of modelling dynamical systems where parameters are not determined precisely. In such cases a tool for measuring the rate of convergence is even more important than in the case of precisely determined Markov chains, since most of the existing methods of estimating the limit distributions are iterative. We define a new coefficient of ergodicity that provides necessary and sufficient conditions for convergence of the most commonly used class of imprecise Markov chains. This so-called *weak coefficient of ergodicity* is defined through an endowment of the structure of a metric space to the class of imprecise probabilities. Therefore we first make a detailed analysis of the metric properties of imprecise probabilities.

Keywords Markov chain · Imprecise Markov chain · Coefficient of ergodicity · Weak coefficient of ergodicity · Uniform coefficient of ergodicity · Lower expectation · Upper expectation · Hausdorff metric · Convergence of Markov chains

An earlier version of this paper, Škulj and Hable (2009), was presented at the ISIPTA'09 conference in Durham, UK, 2009.

D. Škulj (✉)
Faculty of Social Sciences, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia
e-mail: damjan.skulj@fdv.uni-lj.si

R. Hable
Department of Mathematics, 95440 Bayreuth, Germany
e-mail: robert.hable@uni-bayreuth.de

1 Introduction

Modelling Markov chains usually requires estimating a 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 classical 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 methods of imprecise probabilities has allowed the imprecision in parameters to be incorporated into the models and reflected in the results.

Several approaches to modelling Markov chains with uncertain parameters have been proposed in the literature by now. A very thorough study of the so-called Markov set-chains has been presented by Hartfiel (1998) (see also Hartfiel and Seneta 1994). A Markov set-chain is essentially a Markov chain where sets of probabilities and transition matrices are considered as possible candidates for an unknown true probability distribution or transition matrix. Special attention is paid 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 (2002). The more general interval probabilities based on Weichselberger's model Weichselberger (2001) are used in the study of Markov chains by Škulj (2006, 2007, 2009). A more recent approach by De Cooman et al. (2009) 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 (1973), followed by Harmanec (2002), Itoh and Nakamura (2007), Nilim and Ghaoui (2005), White and Eldeib (1994).

One of the central questions in the theory of Markov chains, also more or less explored in the above listed literature, is whether the probability distributions at consecutive steps converge to a limit distribution and whether such a limit distribution is independent of the initial state or distribution on the set of states. One of the most commonly known results is the Perron-Frobenius theorem which shows that the probability distributions corresponding to a finite regular Markov chain converge to a unique limit distribution independently of the initial distribution. However, regularity is not always a necessary condition for convergence. A more general approach is to use various coefficients of ergodicity (see e.g. Seneta 1979), which additionally to giving necessary or sufficient conditions for convergence also measure the rate of the convergence.

In the literature listed above the question of convergence of imprecise Markov chains and the related problem of invariant sets of probability distributions receive a great part of attention. Thus, Hartfiel (1998) generalises a coefficient of ergodicity to characterise the convergence of Markov set-chains. Further focus on the properties of invariant sets of distributions with a generalised concept of regularity was provided by Škulj (2009), where a convergence result is shown for regular imprecise Markov chains and some properties of invariant sets of distributions are analysed in the case of not necessarily regular imprecise Markov chains. The most general result by now seems to be the one given by De Cooman et al. (2009) who provide necessary and sufficient conditions for convergence in terms of accessibility relations. Their conditions are substantially weaker than those given by Hartfiel (1998), which is due to their restriction to models being able to be described using the expectation operators, which are somewhat less general than the models allowing completely arbitrary sets of probabilities and transition matrices used by Hartfiel; namely, expectation operators can only describe closed sets of distributions that are convex. The models using the expectation operators approach also assume an independence for rows of transition operators which does not necessarily hold for sets of transition matrices, which can result in not necessarily convex sets of probabilities. We explain this independence condition and its possible violations in Sect. 4.

In this paper we extend one of the most commonly used coefficients of ergodicity to the case of imprecise Markov chains. The main idea behind the coefficients of ergodicity is to measure the distances between rows of transition matrices, which in the precise case are probability distributions and in the imprecise case are, similarly, imprecise probability distributions. Thus the desired generalisation of the coefficients of ergodicity is possible through an appropriate endowment of the structure of a metric space to the set of imprecise probabilities. The main contribution of this paper is the introduction of a new so-called *weak coefficient of ergodicity* that, in an alternative way, characterises the type of convergence studied by De Cooman et al. (2009). We show that the conditions for convergence provided by this coefficient are necessary and sufficient if convex sets of probabilities are used to describe the imprecision of Markov chains, and if the independence condition for rows of transition operators is assumed. Moreover, additionally to stating necessary and sufficient conditions, the weak coefficient of ergodicity gives estimates for the rate of convergence, which is especially useful because the methods for estimating imprecise Markov chains are usually iterative. The coefficient of ergodicity then allows estimating an upper bound for the error of an estimate.

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 Sect. 3, we explore possibilities to endow the family of imprecise probabilities with the structure of a metric space, and in Sect. 4 we describe the model of imprecise Markov chains that we use. In Sect. 5 we give an alternative characterisation of the uniform coefficient of ergodicity and, as the main contribution of this paper, we define a new so-called weak coefficient of ergodicity which provides necessary and sufficient conditions for convergence for an imprecise Markov chain and measures its rate. Finally we give a numerical example.

2 Lower expectation operators

Imprecision in probability distributions is often described by sets of possible probability distributions, which are usually assumed to be convex. Such sets can equivalently be described using lower or upper expectation functionals. We explore the duality between both representations from the point of view of metric spaces.

Let Ω be a finite set and let $\mathscr{F}$ be the set of all real-valued maps on Ω. Further let $\mathscr{F}_1$ denote the subset of all real-valued maps with $0 \leq f(\omega) \leq 1$ for every $\omega \in \Omega$. We denote by 1_Ω, or sometimes just 1, the constant map on Ω such that $1_\Omega(\omega) = 1$

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 $\mathscr{F}$ 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 $\mathscr{F}_1 = \{f \in \mathscr{F} \mid f \geq 0, \|f\|_\infty \leq 1\}$.

We characterise a *probability measure* or *probability* on the measurable space $(\Omega, 2^\Omega)$ through its *probability mass function* p which is a real valued map on Ω such that

$$\sum_{\omega\in\Omega} p(\omega) = 1 \quad \text{and} \quad p(\omega) \geq 0 \quad \text{for every } \omega \in \Omega.$$

Therefore $p(A) = \sum_{\omega\in A} p(\omega)$ for every $A \subseteq \Omega$. Thus every probability mass function can be considered to belong to the set $\mathscr{F}_1$. Sometimes we enumerate the 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* (see Walley 1991, Chaps. 2, 3 for the proofs of the following results of this section). We denote the lower expectation operator of a closed convex set of probabilities $\mathscr{M}$ by $\underline{P}$ and the upper expectation operator by $\overline{P}$. So for any $f \in \mathscr{F}$ we define:

$$\underline{P}(f) = \min_{p\in\mathscr{M}} E_p f \quad \text{and} \quad \overline{P}(f) = \max_{p\in\mathscr{M}} E_p f, \tag{1}$$

where $E_p f = \sum_{\omega\in\Omega} f(\omega)p(\omega)$. 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 a lower expectation operator we say that $\mathscr{M}$ is a *credal set* of $\underline{P}$ and we may denote

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

Since the lower and the upper expectation operator are conjugate, i.e. $\overline{P}(f) = -\underline{P}(-f)$, and therefore the upper expectation is determined by the lower one, in the rest of the paper we will only use lower expectation operators. Every lower expectation operator $\underline{P}$ has the following properties. Let f, f_1, f_2, f_n be arbitrary elements from $\mathscr{F}$. Then:

boundedness $\min_{\omega\in\Omega} f(\omega) \leq \underline{P}(f) \leq \max_{\omega\in\Omega} f(\omega)$;

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;
monotonicity if $f_1 \leq f_2$ then $\underline{P}(f_1) \leq \underline{P}(f_2)$;
continuity if $f_n \to f$ point-wise then $\underline{P}(f_n) \to \underline{P}(f)$;
upper-lower consistency $\underline{P}(f) \leq -\underline{P}(-f) = \overline{P}(f)$.

Further we note that any expectation operator is completely determined by its values on the space $\mathscr{F}_1$. To see this take any map $f \in \mathscr{F}$ and define the corresponding $\tilde{f} \in \mathscr{F}_1$ 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 $(\Omega, \mathscr{A})$ can be made a metric space using the following metric:

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

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. Beer 1993, p. 85) is defined as

$$d_H(X, Y) = \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right\}. \tag{3}$$

This metric makes the set of non-empty compact sets a metric space denoted by $F(M)$. Moreover, if M is a compact space then 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 (2) in (3), 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:

$$\tilde{d}(\underline{P}, \underline{P}') = \max_{f \in \mathscr{F}_1} |\underline{P}(f) - \underline{P}'(f)|. \tag{4}$$

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

$$|\underline{P}(f) - \underline{P}'(f)| \leq \tilde{d}(\underline{P}, \underline{P}')\|f\|_\infty. \tag{5}$$

If a credal set consists of a singleton probability p then the corresponding lower or upper expectation operator is E_p, the expectation operator with respect to p. The next proposition shows that, given this correspondence, the metrics (4) and (2) coincide for probability measures. Therefore, from now on we denote both distances with d.

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

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

Proof Define the function

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

For any $f \in \mathscr{F}_1$ we have

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

Since for F we have the equality, this proves the proposition.

3.1 Equivalence between the Hausdorff distance and the maximal distance between lower expectation operators

The definition of a coefficient of ergodicity is based on distances between credal sets of rows of transition operators, or, equivalently, between the corresponding lower expectation operators. The following theorem shows that the metric (4) between lower expectation operators coincides with the Hausdorff metric between their credal sets. (A similar result also for infinite sets Ω can be found in Hable (2010, Lemma 3.2).)

Theorem 2 *Let $\mathscr{M}_1$ and $\mathscr{M}_2$ be closed convex sets of probabilities and let $\underline{P}_1$ and $\underline{P}_2$ be their lower expectation operators. Then we have that*

$$d(\underline{P}_1, \underline{P}_2) = d_H(\mathscr{M}_1, \mathscr{M}_2). \tag{6}$$

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

$$\max_{f\in\mathscr{F}_1} |E_{p_1}f - E_{p_2}f| = \max_{f\in\mathscr{F}_1} E_{p_1}f - E_{p_2}f. \tag{7}$$

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

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

The definition of the Hausdorff distance and the Eq. (7) implies that

$$d_H(\mathscr{M}_1, \mathscr{M}_2) = \max_{p_1\in\mathscr{M}_1} \min_{p_2\in\mathscr{M}_2} \max_{f\in\mathscr{F}_1} E_{p_1}f - E_{p_2}f \tag{8}$$

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

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

where

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

Now the set $\mathscr{M}_2$ is compact by definition, and the mapping $p_2 \mapsto \Gamma(p_2, f)$ is continuous and affine, therefore also convex, for any fixed $f \in \mathscr{F}_1$. Furthermore, for a fixed p_2, the mapping $f \mapsto \Gamma(p_2, f)$ is linear, and therefore concave. Now we can use the minimax theorem (see Fan 1953, Theorem 2) to obtain:

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

That is

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

Using the above equality we obtain:

$$\begin{aligned}&\max_{p_1\in\mathscr{M}_1} \min_{p_2\in\mathscr{M}_2} d(p_1, p_2)\\ &\quad= \max_{p_1\in\mathscr{M}_1} \min_{p_2\in\mathscr{M}_2} \max_{f\in\mathscr{F}_1} E_{p_1}f - E_{p_2}f\\ &\quad= \max_{p_1\in\mathscr{M}_1} \max_{f\in\mathscr{F}_1} \min_{p_2\in\mathscr{M}_2} E_{p_1}f - E_{p_2}f\end{aligned}$$

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

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

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

which completes the proof.

Another approach to coefficients of ergodicity applies the maximal distance between probability measures belonging to a pair of credal sets $\mathscr{M}_1$ and $\mathscr{M}_2$ 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

$$
\begin{aligned}
\max_{\substack{p_1\in\mathscr{M}_1\\ p_2\in\mathscr{M}_2}} d(p_1, p_2) &= \max_{\substack{p_1\in\mathscr{M}_1\\ p_2\in\mathscr{M}_2}} \max_{f\in\mathscr{F}_1} |E_{p_1} f - E_{p_2} f| \\
&= \max_{f\in\mathscr{F}_1} \max_{\substack{p_1\in\mathscr{M}_1\\ p_2\in\mathscr{M}_2}} |E_{p_1} f - E_{p_2} f| \\
&= \max_{f\in\mathscr{F}_1} \max\{\overline{P}_1(f) - \underline{P}_2(f), \overline{P}_2(f) - \underline{P}_1(f)\}. \qquad (9)
\end{aligned}
$$

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

$$
\max_{\substack{p_1\in\mathscr{M}_1\\ p_2\in\mathscr{M}_2}} d(p_1, p_2) = \max_{A\subset\Omega} \max\{\overline{P}_1(1_A) - \underline{P}_2(1_A), \overline{P}_2(1_A) - \underline{P}_1(1_A)\}. \qquad (10)
$$

Now notice again that $\mathscr{F}_1 = \{1 - f : f \in \mathscr{F}_1\}$ and use conjugacy and constant additivity of coherent lower previsions: $\underline{P}(f) = -\overline{P}(-f) = 1 - \overline{P}(1 - f)$ to obtain

$$
\begin{aligned}
\max_{f\in\mathscr{F}_1} \{\overline{P}_2(f) - \underline{P}_1(f)\} &= \max_{f\in\mathscr{F}_1} \{(1 - \underline{P}_2(1 - f)) - (1 - \overline{P}_1(1 - f))\} \\
&= \max_{g\in\mathscr{F}_1} \{\overline{P}_1(g) - \underline{P}_2(g)\}.
\end{aligned}
$$

Note also that $1_\Omega - 1_A = 1_{A^c}$, making the above consideration valid also when the maps f are restricted to the characteristic functions of the subsets of Ω. Now, combining (9), (10) and the last equality gives, for any pair of lower and upper expectation operators $\underline{P}_1$ and $\overline{P}_2$,

$$\max_{f\in\mathscr{F}_1}\{\overline{P}_2(f) - \underline{P}_1(f)\} = \max_{A\subset\Omega}\{\overline{P}_2(1_A) - \underline{P}_1(1_A)\}. \tag{11}$$

3.2 Convergence of lower expectation operators

The convergence of imprecise Markov chains is studied in terms of convergence of the underlying imprecise probabilities in the metric (4). Here we give some preliminary results.

We will need the following result (see Dunford and Schwartz 1988, Lemma I.5.6):

Lemma 3 *A topological space is compact if and only if every family of closed sets, with the property that the intersection of every finite subfamily is non-empty, has a non-empty intersection.*

Corollary 4 *Every decreasing sequence with respect to set inclusion of non-empty closed compact sets has non-empty intersection.*

We use the above corollary to show the following:

Proposition 5 *Let* $\{\underline{P}_n\}_{n\in\mathbb{N}}$ *be an increasing sequence of lower expectation operators and* $\{\mathscr{M}_n\}_{n\in\mathbb{N}}$ *the sequence of the corresponding credal sets. Then the sequence* $\{\mathscr{M}_n\}_{n\in\mathbb{N}}$ *is decreasing with respect to set inclusion and the limit*

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

exists and

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

Moreover, the above credal set is non-empty.

Proof For every $f\in\mathscr{F}_1$ 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 $p\in\bigcap_{n\in\mathbb{N}}\mathscr{M}_n$. Then, by definition, for every $f\in\mathscr{F}_1$ we have that $E_pf\geq\underline{P}_\infty(f)$, so $\bigcap_{n\in\mathbb{N}}\mathscr{M}_n\subseteq\mathscr{M}(\underline{P}_\infty)$. To see the converse inclusion take any probability p such that $E_pf\geq\underline{P}_\infty(f)\geq\underline{P}_n(f)$ for every $n\in\mathbb{N}$. Therefore $p\in\mathscr{M}_n$ for every $n\in\mathbb{N}$ and every $f\in\mathscr{F}_1$, which implies that $p\in\bigcap_{n\in\mathbb{N}}\mathscr{M}_n$. Thus, $\mathscr{M}(\underline{P}_\infty)\subseteq\bigcap_{n\in\mathbb{N}}\mathscr{M}_n$. As follows from Corollary 4, the set $\bigcap_{n\in\mathbb{N}}\mathscr{M}_n$ is non-empty.

Proposition 6 *Let* $\{\underline{P}_n\}_{n\in\mathbb{N}}$ *be any convergent sequence of lower expectation operators and* $\{\mathscr{M}_n\}_{n\in\mathbb{N}}$ *the sequence of the corresponding credal sets. Then the set*

$$\mathscr{M}_\infty = \bigcap_{n\in\mathbb{N}}\text{co}\left(\overline{\bigcup_{m\geq n}\mathscr{M}_m}\right),$$

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* $\mathscr{M}_\infty$ *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:

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

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 $\mathrm{co}(\overline{\bigcup_{m\geq n} \mathscr{M}_m})$.

To see this take any $f \in \mathscr{F}$. We have

$$\tilde{\underline{P}}_n(f) = \inf_{m\geq n} \underline{P}_m(f) = \inf_{m\geq n} \inf_{p\in\mathscr{M}_m} E_p f = \inf_{p\in\bigcup_{m\geq n} \mathscr{M}_m} E_p f. \tag{12}$$

This implies that the convex closure of $\bigcup_{m\geq n} \mathscr{M}_m$ is the credal set of $\tilde{\underline{P}}_n$ (c.f. Hable 2009, Proposition 2.15). Hence, the credal set of $\tilde{\underline{P}}_n$ is also equal to the convex closure of $\overline{\bigcup_{m\geq n} \mathscr{M}_m}$. Since it is a compact set, this implies that it contains every extreme point of the credal set of $\tilde{\underline{P}}_n$ (c.f. Dunford and Schwartz 1988, Lemma V.8.5). Hence, the credal set is equal to the convex hull of $\overline{\bigcup_{m\geq n} \mathscr{M}_m}$ (c.f. Holmes 1975, p. 36).

To finish the proof we apply Proposition 5 to the increasing sequence $\{\tilde{\underline{P}}_n\}$ and the corresponding credal sets $\mathrm{co}(\overline{\bigcup_{m\geq n} \mathscr{M}_n})$.

Corollary 7 *The set of all lower expectation operators is complete in the metric* (4).

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 that such sets are closed and convex. This assumption is particularly useful because, as described in Sect. 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. Weichselberger 2001) or *lower* and *upper previsions* (see e.g. Walley 1991, 2000).

The most basic form involves placing constraints, usually in the form of intervals, on the probabilities belonging to the elementary sets (see Hartfiel 1998; Kozine and Utkin 2002). The imprecision concerning the initial distribution is thus presented through the intervals $[\underline{q}_i, \overline{q}_i]$ 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 $[\underline{p}_{ij}, \overline{p}_{ij}]$ 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. 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 which can be represented via simple intervals described above is fairly limited (Many examples can be found e.g. in Weichselberger 2001; Walley 1991, 2000). 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.

4.1 Sets of probabilities

Now we introduce the terminology used to describe imprecise Markov chains in this paper. For a more detailed treatment see Škulj (2009). 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 thus assume a set $\mathscr{M}_0$ of *initial probability distributions* and let $\underline{P}_0$ be its lower expectation operator (c.f. (1)). Further, we assume a set of transition matrices $\mathscr{P}$, whose rows are *separately specified*, i.e. for any two transition matrices p and p' with ith rows p_i and p'_i replacing the ith row of p with p'_i results in a matrix that still belongs to $\mathscr{P}$. By adopting this property we can associate row sets of distributions $\mathscr{P}_i$ to $\mathscr{P}$ so that any independent choice of rows from the row sets gives a transition matrix in $\mathscr{P}$. If additionally we assume that row sets are closed and convex, we have the following important property.

Lemma 8 *Let $\mathscr{P}$ be a convex set of transition matrices with separately specified rows and let $\mathscr{M}$ be a convex set of probabilities. Then the set of probability distributions at the next step $\mathscr{M} \cdot \mathscr{P} = \{q \cdot p \mid q \in \mathscr{M}, p \in \mathscr{P}\}$ is a convex set.*

We slightly modify the proof in Hartfiel (1998, Lemma 2.5).

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

$$(\alpha q \cdot p + \beta q' \cdot p') = (\alpha q + \beta q')r \tag{13}$$

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

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 (13). Notice that ith row of r is a convex combination of some elements of $\mathscr{P}_i$ and therefore itself a member of $\mathscr{P}_i$ too. Now, because rows are separately specified the resulting matrix is also a member of $\mathscr{P}$. □

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

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

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

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

where $p^{(n)} \in \mathscr{P}$.

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 (1998) or Škulj (2009) 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 $\mathscr{M}_n$ denote the set of possible probability distributions at step n given the initial distribution $\mathscr{M}_0$. Then we have

$$\mathscr{M}_n = \{q^{(0)} \cdot p^{(1)} \cdot \cdots \cdot p^{(n)} \mid q^{(0)} \in \mathscr{M}_0, p^{(i)} \in \mathscr{P} \quad \text{for } i = 1, \ldots, n\} = \mathscr{M}_{n-1} \cdot \mathscr{P}. \tag{14}$$

It follows from Lemma 8 that in the case where the set of transition matrices $\mathscr{P}$ has closed convex separately specified row sets, every $\mathscr{M}_n$ is also a closed convex set of probabilities. Therefore, they can be equivalently represented using lower expectation operators.

4.2 Expectation operators

Now we turn to expectation operators corresponding to both sets of probabilities and sets of transition operators. The initial set of probabilities $\mathscr{M}_0$ can equivalently be represented using its lower expectation operator $\underline{P}_0 = \underline{E}_{\mathscr{M}_0}$. Therefore, $\mathscr{M}_0$ is the credal set of $\underline{P}_0$. Similarly, 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 $\mathscr{P}$ is the *credal set* of $\underline{T}$. For every $f \in \mathscr{F}$ we clearly have that $\min_{p\in\mathscr{P}} pf = \underline{T}(f)$.

To calculate the values of $\underline{P}_n$ on real functions on Ω we follow the approach proposed by De Cooman et al. (2009). 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 assigns to every f the vector of values

$$\underline{T}(f) = \begin{pmatrix} \underline{T}_1(f) \\ \underline{T}_2(f) \\ \vdots \\ \underline{T}_m(f) \end{pmatrix}. \tag{15}$$

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, we can apply $\underline{P}_0$ to $T^n(f)$ to obtain $\underline{P}_n(f)$. We continue to prove that $P_n(f)$ is equal to $\underline{E}_{\mathscr{M}_n} f$ for every $f \in \mathscr{F}$.

4.3 Correspondence between sets of probabilities and expectation operators

Now we show that calculations using closed convex sets of probabilities using (14) and those using the corresponding expectation operators produce identical results under the assumption that row sets of sets of transition matrices are separately specified. We have the following result.

Proposition 9 *Let $\mathscr{M}_0$ be an initial set of probabilities and $\underline{P}_0$ the corresponding lower expectation operator. Further let $\mathscr{P}$ be a set of transition matrices with separately specified rows and $\underline{T}$ the corresponding lower transition operator. Then, for every $f \in \mathscr{F}$, we have that*

$$\underline{P}\,\underline{T}^n(f) = \underline{E}_{\mathscr{M}_0\mathscr{P}^n} f. \tag{16}$$

An analogous result would follow if the lower expectations were replaced with upper expectations.

Proof The equation (16) essentially says that for any initial probability $q^{(0)} \in \mathscr{M}_0$ and any sequence of transition matrices $p^{(1)}, \ldots, p^{(n)} \in \mathscr{P}$ we have that, for every $f \in \mathscr{F}$

$$\underline{P}\,\underline{T}^n(f) \le q^{(0)} \cdot p^{(1)} \cdot \dots \cdot p^{(n)} f, \tag{17}$$

and that for any given f, $q^{(0)}$ and $p^{(1)}, \dots, p^{(n)}$ can be chosen so that equality holds in (17). The inequality follows immediately from the fact that $p^{(i)} f \ge \underline{T}(f)$.

To show the existence of $q^{(0)}$ and $p^{(1)}, \dots, p^{(n)}$ that yield equality in (17) we proceed by induction on n where the case $n = 0$ follows directly from the definitions. We have

$$\underline{T}^n(f) = \underline{T}^{n-1}\underline{T}(f). \tag{18}$$

Since rows of $\underline{T}$ are separately specified we can find probability vectors $p_i^{(n)} \in \mathscr{P}_i$ such that $E_{p_i^{(n)}} f = \underline{T}_i(f)$. The matrix $p^{(n)}$ then satisfies $p^{(n)} f = \underline{T}(f)$.

Now we use the induction assumption to show the existence of matrices $p^{(1)}, \dots, p^{(n-1)} \in \mathscr{M}(\underline{T})$ and $q^{(0)} \in \mathscr{M}(\underline{P})$ so that $\underline{P}\,\underline{T}^{n-1}(\underline{T}(f)) = q^{(0)} \cdot p^{(1)} \cdot \dots \cdot p^{(n-1)} p^{(n)} f$.

As follows from the above propositions the resulting sets of probabilities obtained using calculations with sets of probabilities and those using expectation operators coincide, however it is not necessarily true that $\mathscr{P}^n$ and $\mathscr{M}(\underline{T}^n)$ coincide. Although the lower expectation operators of the rows $\mathscr{P}_i^n$ coincide with the rows $\underline{T}_i^n$, they are not necessarily separately specified, which causes the mentioned difference. For an example see De Cooman et al. (2009, Example 5.1).

4.4 Metric properties of transition operators

The metric (4) can be extended to lower transition operators using the Chebyshev distance by defining

$$d(\underline{T}, \underline{T}') = \max_{f \in \mathscr{F}_1} d_c(\underline{T} f, \underline{T}' f), \tag{19}$$

since $\underline{T} f$ and $\underline{T}' f$ are real valued maps on Ω. Now

$$\begin{aligned} \max_{f \in \mathscr{F}_1} d_c(\underline{T} f, \underline{T}' f) &= \max_{f \in \mathscr{F}_1} \max_{\omega \in \Omega} |\underline{T} f(\omega) - \underline{T}' f(\omega)| \\ &= \max_{f \in \mathscr{F}_1} \max_i |\underline{T}_i f - \underline{T}_i' f| \\ &= \max_i d(\underline{T}_i, \underline{T}_i'), \end{aligned}$$

where $\underline{T}_i$ and $\underline{T}_i'$ denote the ith rows of $\underline{T}$ and $\underline{T}'$ respectively. Thus (19) is equivalent to

$$d(\underline{T}, \underline{T}') = \max_i d(\underline{T}_i, \underline{T}_i'). \tag{20}$$

4.5 Convergence

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 $q_i^{(n)} = P(X_n = i)$ 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 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 1979). Hartfiel (1998) then applies a generalised coefficient of ergodicity to study the convergence of Markov set-chains.

Recently, De Cooman et al. (2009) 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 relation between states. They show that the property of being regularly absorbing is necessary and sufficient for convergence. In the following section we construct a new coefficient of ergodicity for imprecise Markov chains which also gives a necessary and sufficient condition for convergence.

5 Coefficients of ergodicity

Coefficients of ergodicity or *contraction coefficients* measure the rate of convergence of Markov chains. In his paper Seneta (1979) 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) $\tau(p \cdot p') \leq \tau(p) \cdot \tau(p')$ for every pair of stochastic matrices with no zero columns p and p';
(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 (2). This coefficient was introduced by Dobrushin (1956) and its direct evaluation is derived by Paz (1970):

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

According to (2), the above can be stated as

$$\tau(p) = \max_{i,j} d(p_i, p_j), \tag{21}$$

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

Another possible approach is to use the projective distance between vectors $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$ which is defined by

$$d_p(x, y) = \max_{i,j} \ln \left(\frac{x_1 y_j}{x_j y_i} \right) \tag{22}$$

(see Seneta 2006). Let T be a transition matrix with positive entries. Using the above metric the Birkhoff's coefficient of ergodicity (Birkhoff 1957) is defined by

$$\tau_b(T) = \sup_{\substack{x,y>0 \\ x \neq \lambda y}} \frac{d_p(xT, yT)}{d_p(x, y)} = \frac{1 - \sqrt{\phi(T)}}{1 + \sqrt{\phi(T)}}, \tag{23}$$

where

$$\phi(T) = \min_{i,j,k,l} \frac{t_{ik} t_{jl}}{t_{jk} t_{il}}. \tag{24}$$

5.1 The uniform coefficient of ergodicity

For the case of imprecise Markov chains Hartfiel (1998) extends the concept of a coefficient of ergodicity to Markov chains where sets of transition probabilities are considered. For a set of transition matrices $\mathscr{P}$ he defines the *uniform coefficient of ergodicity* as

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

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

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

where p_{ik} and q_{ik} are the components of P and Q respectively. A related problem of convergence of inhomogeneous products of matrices using coefficients of ergodicity was studied by Hartfiel and Rothblum (1998).

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

Proposition 10 *Let $\mathscr{P}$ be a set of transition matrices with separately specified rows and let $\underline{T}$ and $\overline{T}$ be its lower and upper expectation operators. Then we have that*

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

Proof The second equality follows from (11). Let $p \in \mathscr{P}$ be an arbitrary transition matrix. Then its ith and jth rows are arbitrary probability distributions belonging to the credal sets of the ith and jth row of $\mathscr{P}$. We have that

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

where the third equation follows from (9).

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

Remark 1 The crucial assumption in the above proposition is that rows are separately specified. Thus, for instance, $\tau(\mathscr{P}^n)$ is not guaranteed to be equal to $\tau(\underline{T}^n)$, because, as explained earlier, $\mathscr{M}(\underline{T}^n)$ does not necessarily coincide with $\mathscr{P}^n$. In fact, $\mathscr{P}^n \subseteq \mathscr{M}(\underline{T}^n)$ always holds, which clearly implies that $\tau(\mathscr{P}^n) \leq \tau(\underline{T}^n)$.

The uniform coefficient of ergodicity can be used as a contraction measure for a set of transition matrices. The following theorem holds (Hartfiel 1998, Theorem 3.3):

Theorem 11 *Let $\mathscr{M}_1$ and $\mathscr{M}_2$ be non-empty compact sets of probabilities. Then*

$$d_H(\mathscr{M}_1 \cdot \mathscr{P}, \mathscr{M}_2 \cdot \mathscr{P}) \leq \tau(\mathscr{P}) d_H(\mathscr{M}_1, \mathscr{M}_2).$$

A stochastic matrix p whose coefficient of ergodicity $\tau(p)$ is strictly smaller than 1 is called *scrambling* (see Seneta 1979). Further if $\mathscr{P}$ is a set of transition matrices such that $\tau(p^{(1)} \cdot p^{(2)} \cdots p^{(r)}) < 1$ for any matrices $p^{(i)} \in \mathscr{P}$ then such a set is called *product scrambling* (see Hartfiel 1998), and r is then called its *scrambling integer*. Thus we have that $\tau(\mathscr{P}^r) < 1$.

Theorem 11 implies the following more general corollary (Hartfiel 1998, Theorem 3.4):

Corollary 12 *Let $\mathscr{P}$ be be product scrambling with scrambling integer r and let $\mathscr{M}_0$ be a non-empty compact set of probabilities. Then, for any positive integer h,*

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

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

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

Thus,

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

As follows from Remark 1 using $\tau(\underline{T}^r)$ instead of $\tau(\mathscr{P}^r)$ would produce a more conservative estimate.

Theorem 11 estimates the rate of convergence for a Markov set-chain in the Hausdorff metric. Moreover, if $\tau(\mathscr{P}) < 1$ 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 $p^{(i)} \in \mathscr{P}$ we have that the sequence $q^{(n)} = q^{(0)} p^{(1)} \cdots p^{(n)}$ converges to some q_∞ at the given rate. This is a consequence of the fact that $\tau(p^{(1)} \cdots p^{(n)}) \to 0$ as n tends to infinity. Moreover, since clearly $\tau(\mathscr{P}') \leq \tau(\mathscr{P})$ for every $\mathscr{P}' \subseteq \mathscr{P}$, it follows that given a convergent Markov chain with the set of transition probabilities $\mathscr{P}$ then a Markov chain with the set of transition probabilities $\mathscr{P}'$ is also convergent.

5.2 The weak coefficient of ergodicity

In the case where sets of probabilities are convex, the above requirements are clearly sufficient but not necessary. It has been shown by De Cooman et al. (2009) that it is 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.

Further sufficient conditions for unique convergence have been found by De Cooman et al. (2009) by studying the accessibility relation between states. They define that the state j is *accessible* from i if it is not impossible to reach i from j in some number of steps, meaning that the upper probability is non-zero. A stationary imprecise Markov chain is called *regularly absorbing* if there exists a subset of states called *top class* such that it is accessible from all other states and no other state is accessible from this class. If additionally every state in the top class is reachable from any other in any number of steps greater than some given r then such a chain is called *top class regular*. A top class regular Markov chain is *regularly absorbing* whenever it is top class regular and the lower probability of reaching its top class from any state is positive. One of the main results of the paper by De Cooman et al. is that an imprecise Markov chain is convergent if and only if it is regularly absorbing.

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 (21). 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

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

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 13 *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 14 *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 (15)) and let $\underline{c}_f$ and $\overline{c}_f$ be its minimal and maximal element respectively. Further let $\underline{\tilde{P}_1} = \underline{P}_1\underline{T}$ and $\underline{\tilde{P}_2} = \underline{P}_2\underline{T}$. Then using constant additivity and (5) we obtain

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

Corollary 15 *Let $\underline{R}$ and $\underline{S}$ be any transition lower expectation matrices. Then:*

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

Proof Denote $\underline{T} = \underline{R}\,\underline{S}$ 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 14 then yields

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

as required.

The next corollary is now immediate.

Corollary 16 *For any lower expectation operator $\underline{T}$ we have that*

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

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

Proof Denote $\rho := \rho(\underline{T})$. Then for any pair of indices i and j we have $d(\underline{T}_i, \underline{T}_j) \le \rho$. 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 $1 \le i \le m$. Then $|p_i^A(A) - p_j^A(A)| < 1$ and $|p_i^A(A') - p_j^A(A')| \le 1$ 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 18 *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:*

$$\underline{P}_\infty \underline{T} = \underline{P}_\infty. \tag{25}$$

We will call a lower expectation operator satisfying the property (25) 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 (4). 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 7), the sequence converges to some lower expectation operator $\underline{P}_\infty$.

Clearly, if $\underline{P}$ is an invariant lower expectation operator for $\underline{T}$ then it is also invariant for $\underline{T}^r$ for any $r > 0$. Thus, if $\underline{T}^r$ has a unique invariant lower expectation $\underline{P}$, then it is also the unique invariant lower expectation operator for $\underline{T}$. Therefore, if $\rho(\underline{T}^r) < 1$, for some $r > 0$, then $\underline{T}$ has a unique invariant lower expectation operator.

Theorem 19 *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 (25) and Proposition 14 and Corollary 16 we obtain

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

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

A corollary analogous to Corollary 12 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 that 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 20 *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,*

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

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

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

5.3 Necessary and sufficient conditions for convergence

The type of convergence measured by the weak coefficient of ergodicity is clearly closely related to that described by De Cooman et al. (2009). This suggests that regularly absorbing and weakly scrambling lower expectation matrices are closely related. In the following we show that indeed both give necessary and sufficient conditions, and are therefore equivalent.

Now we show that ρ is continuous with respect to the metric (20). Let $\underline{T}$ and $\underline{T}'$ be two lower transition operators. Then we have, for instance,

$$\begin{aligned}\rho(\underline{T}') &= \max_{i,j} d(\underline{T}'_i, \underline{T}'_j)\\ &\le \max_{i,j}\{d(\underline{T}'_i, \underline{T}_i) + d(\underline{T}'_j, \underline{T}_j) + d(\underline{T}_i, \underline{T}_j)\}\\ &\le 2d(\underline{T}, \underline{T}') + \rho(\underline{T}).\end{aligned}$$

It follows that $|\rho(\underline{T}) - \rho(\underline{T}')| \le 2d(\underline{T}, \underline{T}')$, and this implies continuity of ρ.

In the following we show that a lower transition operator being weakly product scrambling is not only a sufficient but also a necessary condition for convergence of the corresponding Markov chain.

Let $\underline{T}$ be a lower transition operator such that $\underline{P}_\infty = \lim_{n\to\infty} \underline{P}_0\underline{T}^n$ exists and is independent of $\underline{P}_0$. Then the limit $\lim_{n\to\infty} \underline{T}^n$ exists and is the lower transition operator whose rows are equal to $\underline{P}_\infty$. To see this, define the operator $\underline{P}_i$ with $\underline{P}_i f = f_i$. We have

$$\underline{P}_\infty f = \lim_{n\to\infty} \underline{P}_i\underline{T}^n f = \lim_{n\to\infty} (\underline{T}^n f)_i = \lim_{n\to\infty} (\underline{T}^n)_i f$$

for every $f \in \mathscr{F}$. We define the limit lower transition operator $\underline{T}_\infty$ with $(\underline{T}_\infty)_i f = \underline{P}_\infty f$. We have that $\underline{T}_\infty = \lim_{n\to\infty} \underline{T}^n$. Clearly then

$$\rho(\underline{T}_\infty) = 0. \tag{26}$$

Now we can prove the following theorem.

Theorem 21 *Let $\underline{T}$ be the lower transition operator for an imprecise Markov chain. Then the chain converges uniquely if and only if $\rho(\underline{T}^r) < 1$ for some $r > 0$.*

Proof Let $\underline{T}$ be the lower transition operator of a uniquely converging Markov chain. Then, as shown above, $\underline{T}_\infty$ exists and $\rho(\underline{T}_\infty) = 0$. Now we have, by continuity of ρ,

$$\rho(\lim_{n\to\infty} \underline{T}^n) = \lim_{n\to\infty} \rho(\underline{T}^n) = 0.$$

The second equality implies that for some r $\rho(\underline{T}^r)$ must be less than 1.

The converse follows from Theorem 19.

5.4 Calculation of coefficients of ergodicity for 2-monotone models

The special case of imprecise Markov chains where transition probabilities are bounded by 2-monotone lower probabilities allows an especially convenient way of calculating the weak coefficients of ergodicity, based on the values of the lower probabilities on the subsets of Ω. We will use the relation between lower probabilities and the Choquet integral, which in addition allows a more convenient way of calculating expectation operators at future time steps.

Let $L\colon 2^\Omega \to \mathbb{R}$ be a monotone set function with $L(\emptyset) = 0$ and $L(\Omega) = 1$ such that

$$L(A \cup B) + L(A \cap B) \geq L(A) + L(B). \tag{27}$$

Then we say that L is a 2-monotone lower probability. Let $\mathscr{M}(L)$ be the set of all additive probabilities p that dominate L: $p(A) \geq L(A)$ for every $A \subseteq \Omega$. Then we have that

$$\min_{p\in\mathscr{M}(L)} p(A) = L(A) \quad \text{for every } A \subseteq \Omega. \tag{28}$$

Moreover, let f be any real valued map on Ω. The lower expectation $E_{\mathscr{M}(L)} f$ can now be calculated using the Choquet integral:

$$E_{\mathscr{M}(L)} f = \min f + \int_{\min f}^{\max f} L(\{\omega | f(\omega) \geq x\})\, \mathrm{d}x =: \int f\, \mathrm{d}L. \tag{29}$$

We define the following distance between lower probabilities:

$$d(L_1, L_2) = \max_{A\subseteq\Omega} |L_1(A) - L_2(A)| \tag{30}$$

and show that it coincides with the distance (4) between the corresponding lower expectation operators.

Proposition 22 *Let* $\underline{P}_1 f = \int f\, \mathrm{d}L_1$ *and* $\underline{P}_2 f = \int f\, \mathrm{d}L_2$ *be lower expectation operators corresponding to 2-monotone lower probabilities* L_1 *and* L_2 *respectively. Then*

$$d(\underline{P}_1, \underline{P}_2) = d(L_1, L_2). \tag{31}$$

Proof Take an $f \in \mathscr{F}_1$ and let $S_x = \{\omega | f(\omega) \geq x\}$ be its level sets. We first show that $d(\underline{P}_1, \underline{P}_2) \leq d(L_1, L_2)$. We have

$$\begin{aligned} |\underline{P}_1 f - \underline{P}_2 f| &= \left| \int_0^1 L_1(S_x)\,\mathrm{d}x - \int_0^1 L_2(S_x)\,\mathrm{d}x \right| \\ &= \left| \int_0^1 (L_1(S_x) - L_2(S_x))\,\mathrm{d}x \right| \\ &\leq \int_0^1 |L_1(S_x) - L_2(S_x)|\,\mathrm{d}x \\ &\leq \max_{A \subseteq \Omega} |L_1(A) - L_2(A)| \\ &= d(L_1, L_2). \end{aligned}$$

Now let $f = 1_A$ where $A = \text{argmax}_{A \subseteq \Omega} |L_1(A) - L_2(A)|$. We have $S_x = A$ for every $x \in (0, 1)$ and therefore

$$|\underline{P}_1 f - \underline{P}_2 f| = |L_1(A) - L_2(A)| = d(L_1, L_2),$$

which completes the proof.

Corollary 23 *Let a lower transition operator $\underline{T}$ be given such that each row T_i corresponds to a 2-monotone lower probability L_i. Then*

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

The most commonly used model that falls into the class of 2-monotone models is the model of imprecise Markov chains that uses so-called *probability intervals*. A lower and upper probability mass function $\underline{p}$ and $\overline{p}$ are given which determine the set of probability mass functions

$$\mathscr{M} = \{p \mid \underline{p}(\omega) \leq p(\omega) \leq \overline{p}(\omega) \text{ for every } \omega \in \Omega\}. \tag{32}$$

We assume that for each $\omega \in \Omega$ a probability mass function $p \in \mathscr{M}$ exists so that $p(\omega) = \underline{p}(\omega)$ and similarly for the upper mass function. The lower probability given by

$$L(A) = \max \left\{ \sum_{\omega \in A} \underline{p}(\omega), 1 - \sum_{\omega \in A^c} \overline{p}(\omega) \right\}, \text{ for every } A \subseteq \Omega, \tag{33}$$

is 2-monotone (see e.g. de Campos et al. 1994) and $\mathscr{M}(L) = \mathscr{M}$, where $\mathscr{M}$ is the set defined by (32).

5.5 Numerical example

Let an imprecise Markov chain on the set $\Omega = \{1, 2, 3\}$ be given with the lower transition matrix

$$P_L = \begin{pmatrix} 0.5 & 0.1 & 0.7 & 0.1 & 0.7 & 0.4 \\ 0.1 & 0.4 & 0.6 & 0.3 & 0.5 & 0.8 \\ 0.2 & 0.2 & 0.5 & 0.4 & 0.7 & 0.7 \end{pmatrix},$$

where the columns denote sets $A_1 = \{1\}$, $A_2 = \{2\}$, $A_3 = \{1, 2\}$, $A_4 = \{3\}$, $A_5 = \{1, 3\}$, $A_6 = \{2, 3\}$. The weak coefficient of ergodicity is equal to $0.4 = |L_1(A_1) - L_2(A_1)|$. Let the initial lower probability be 0. The following figure shows seven consecutive iterations and the distances to the limit distribution:

i	$L_i(A_1)$	$L_i(A_2)$	$L_i(A_3)$	$L_i(A_4)$	$L_i(A_5)$	$L_i(A_6)$	$d(L_i, L_\infty)$
0	0	0	0	0	0	0	0.6250
1	0.1000	0.1000	0.5000	0.1000	0.5000	0.4000	0.2000
2	0.1800	0.1600	0.5600	0.1900	0.6000	0.5300	0.0700
3	0.2140	0.1850	0.5740	0.2250	0.6200	0.5750	0.0250
4	0.2262	0.1945	0.5788	0.2375	0.6240	0.5910	0.0090
5	0.2303	0.1980	0.5805	0.2419	0.6248	0.5968	0.0032
6	0.2316	0.1993	0.5811	0.2435	0.6250	0.5988	0.0012
7	0.2320	0.1997	0.5813	0.2441	0.6250	0.5996	0.0004
∞	0.2321	0.2000	0.5813	0.2444	0.6250	0.6000	

Clearly, the distances $d(L_i, L_\infty)$ are bounded from above by $d(L_0, L_\infty)\rho(\underline{T})^i$.

6 Conclusions

The main contribution of this paper is the definition of the weak coefficient of ergodicity that gives necessary and sufficient conditions for convergence and additionally measures the rate of convergence of imprecise Markov chains. We have used two equivalent approaches to its derivation, through sets of probabilities and through the corresponding lower expectation operators. In the first part our result gives an alternative approach to the characterisation of convergence conditions, given by De Cooman et al. (2009), who derive them on the basis of accessibility relations. However, a more important use of this coefficient is its ability to estimate how close a distribution obtained at a certain step is to the limit distribution. The importance of such a measure is even greater in the imprecise case than in precise, since analytical methods to finding limit distributions are not known at the moment. Thus limit distributions are being calculated iteratively and the weak coefficient of ergodicity can then be used to measure how close a certain iteration is to the true distribution.

On the way to derive the coefficients of ergodicity we have studied metric properties of imprecise measures, and found some interesting relations between the representation with sets of probabilities and the corresponding expectation operators.

Another interesting problem where convergence of imprecise Markov chains is considered was described by Crossman et al. (2009a,b). They consider absorbing Markov chains and study conditioning of probability distributions on non-absorption. For the precise case this problem was studied by Darroch and Seneta (1965). It turns out that conditional on non-absorption probability distributions converge to some limit distribution if some regularity assumptions are made (Crossman and Škulj 2010). One of the challenges for future work would be to modify the coefficients of ergodicity to also work for this type of convergence. Possibly, generalising the Birkhoff's coefficient of ergodicity would be a step towards solving this problem.

A problem related to convergence of Markov chains is the problem of invariant distributions for Markov chains. For the case of imprecise Markov chains some results can be found in Škulj (2009), but the general structure of invariant sets of distributions has not yet been fully explored.

Acknowledgments We are grateful to Gert De Cooman, Filip Hermans, Erik Quaeghebeur and an anonymous referee for their suggestions and discussions that have led to many improvements to this paper.

References

Beer G (1993) Topologies on closed and closed convex sets. Kluwer, Dordrecht

Birkhoff G (1957) Extensions of Jentzsch's theorem. Trans Am Math Soc 85(1):219–227

Crossman RJ, Škulj D (2010) Imprecise Markov chains with absorption. Int J Approx Reason 51:1085–1099. doi:10.1016/j.ijar.2010.08.008

Crossman RJ, Coolen-Schrijner P, Coolen FPA (2009a) Time-homogeneous birth-death processes with probability intervals and absorbing state. J Stat Theory Practice 3(1):103–118

Crossman RJ, Coolen-Schrijner P, Škulj D, Coolen FPA (2009b) Imprecise Markov chains with an absorbing state. In: Augustin T, Coolen FPA, Moral S, Troffaes MCM (eds) ISIPTA'09: proceedings of the sixth international symposium on imprecise probability: theories and applications, SIPTA, Durham, UK, pp 119–128

Darroch JN, Seneta E (1965) On quasi-stationary distributions in absorbing discrete-time finite Markov chains. J Appl Probab 2(1):88–100, http://www.jstor.org/stable/3211876

de Campos LD, Huete J, Moral S (1994) Probability intervals: a tool for uncertain reasoning. Int J Uncertain Fuzz Knowl Based Syst 2(2):167–196

De Cooman G, Hermans F, Quaeghebeur E (2009) Imprecise Markov chains and their limit behavior. Probab Eng Inform Sci 23(4): 597–635. doi:10.1017/S0269964809990039

Dobrushin R (1956) Central limit theorem for non-stationary Markov chains, I, II. Theory Probab Appl 1(4):329–383

Dunford N, Schwartz J (1988) Linear operators. Part I: general theory. Wiley, New York

Fan K (1953) Minimax theorems. Proc Natl Acad Sci USA 39:42–47

Hable R (2009) Data-based decisions under complex uncertainty. PhD thesis, Ludwig-Maximilians-Universität (LMU) Munich, http://edoc.ub.uni-muenchen.de/9874/

Hable R (2010) Minimum distance estimation in imprecise probability models. J Stat Plan Inference 140:461–479

Harmanec D (2002) Generalizing Markov decision processes to imprecise probabilities. J Stat Plan Inference 105:199–213

Hartfiel D (1998) Markov set-chains. Springer, Berlin

Hartfiel D, Rothblum U (1998) Convergence of inhomogenous products of matrices and coefficients of ergodicity. Linear Algebra Appl 277:1–9

Hartfiel D, Seneta E (1994) On the theory of Markov set-chains. Adv Appl Probab 26(4):947–964

Holmes RB (1975) Geometric functional analysis and its applications. Springer, Berlin

Itoh H, Nakamura K (2007) Partially observable Markov decision processes with imprecise parameters. Artif Intell 171(8–9):453–490

Kozine I, Utkin L (2002) Interval-valued finite Markov chains. Reliable Comput 8(2):97–113
Nilim A, Ghaoui LE (2005) Robust control of Markov decision processes with uncertain transition matrices. Oper Res 53:780–798
Paz A (1970) Ergodic theorems for infinite probabilistic tables. Ann Math Stat 41(2):539–550
Satia J, Lave R (1973) Markovian decision processes with uncertain transition probabilities. Oper Res 21(3):728–740
Seneta E (1979) Coefficients of ergodicity—structure and applications. Adv Appl Probab 11(2):270–271
Seneta E (2006) Non-negative matrices and Markov chains. Springer, Berlin
Škulj D (2006) Finite discrete time Markov chains with interval probabilities. In: Lawry J, Miranda E, Bugarín A, Li S, Gil MA, Grzegorzewski P, Hryniewicz O (eds) SMPS. Advances in soft computing. Springer, Berlin, vol 37, pp 299–306
Škulj D (2007) Regular finite Markov chains with interval probabilities. In: De Cooman G, Zaffalon M, Vejnarová J (eds) ISIPTA'07—proceedings of the fifth international symposium on imprecise probability: theories and applications, SIPTA, pp 405–413
Škulj D (2009) Discrete time Markov chains with interval probabilities. Int J Approx Reason 50(8):1314–1329. doi:10.1016/j.ijar.2009.06.007
Škulj D, Hable R (2009) Coefficients of ergodicity for imprecise Markov chains. In: Augustin T, Coolen FPA, Moral S, Troffaes MCM (eds) ISIPTA'09: proceedings of the sixth international symposium on imprecise probability: theories and applications, SIPTA, Durham, UK, pp 377–386
Walley P (1991) Statistical reasoning with imprecise probabilities. Chapman and Hall, London
Walley P (2000) Towards a unified theory of imprecise probability. Int J Approx Reason 24:125–148
Weichselberger K (2001) Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung. I: Intervallwahrscheinlichkeit als umfassendes Konzept. Physica-Verlag, Heidelberg
White C, Eldeib H (1994) Markov decision processes with imprecise transition probabilities. Oper Res 42(4):739–749