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.

Imprecise Markov chains

Markov chains

A Markov chain is a random process with the Markov property, which means that future states depend on the present state and not on the past states. This dependence is expressed through transition probabilities:

$$P(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j|X_n = i) = p_{ij}^n$$

for every $n \in \mathbb{N}$. The knowledge about first state is given by an initial probability $P(X_0 = j) = q_j$.

An imprecise Markov chain is a Markov chain where imprecise knowledge of parameters is built into the model and is reflected in results.

Representation of imprecise Markov chains

The uncertainty of parameters can be expressed through sets of probabilities, which means that instead of single precisely known initial and transition probabilities we take sets of possible candidates. We do not assume the transition probability to be constant but only that at every step it belongs to the same set.

The same is assumed for probability distributions on states at different time steps. Let $\mathcal{M}_n$ be the set of possible distributions at step n and $\mathcal{P}$ the set of possible transition matrices. The above assumptions imply $\mathcal{M}_{n+1} = \mathcal{M}_n \cdot \mathcal{P}$.

Since the sets $\mathcal{M}_n$ are usually convex, there is another equivalent way of expressing them using expectation operators (de Cooman et. al (2009)): The lower expectation operator corresponding to a set of probabilities $\mathcal{M}$ is given by $\underline{P}_n[f] = \min_{p\in\mathcal{M}_n} E_p[f]$, where f is a real valued map on the set of states.

Similarly, sets of transition matrices can be represented by (lower) transition operators:

$$\underline{T}[f] = \begin{pmatrix} \underline{T}_1[f] \\ \vdots \\ \underline{T}_m[f] \end{pmatrix}.$$

The lower expectation operator $\underline{P}_n$ is then evaluated as

$$\underline{P}_n[f] = \underline{P}_0\underline{T}^n[f].$$

Coefficients of ergodicity

Coefficients of ergodicity or contraction coefficients measure the rate of convergence of Markov chains. Let p be a stochastic matrix with no zero columns. Then a coefficient of ergodicity is defined as

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

If we take the metric

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

then it can be directly evaluated as

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

In terms of distances this expression can be written as

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

where p_i and p_j are the i-th and j-th row of p respectively.

In general, a coefficient of ergodicity $\tau(p)$ satisfies:

- $0 \le \tau(p) \le 1$;
- $\tau(p_1p_2) \le \tau(p_1)\tau(p_2)$;
- $\tau(p) = 0$ iff p has rank 1: $p = \mathbf{1}v$ for some vector v.

This clearly implies: If $\tau(p) < 1$ then the powers p^n converge to a matrix with rank 1, which is equivalent to unique convergence of the corresponding Markov chain.

Generalised coefficients of ergodicity

The idea of the generalisation to imprecise case is to use distances between imprecise probabilities; more precisely, between the rows of imprecise transition matrices.

Hartfiel (1998) defines the so called uniform coefficient of ergodicity which measures the maximal distance between rows of transition matrices:

$$\tau(\mathcal{P}) = \sup_{p\in\mathcal{P}} \tau(p) = \max_{i,j}\max_{A\subset\Omega} \overline{T}_i(1_A) - \underline{T}_j(1_A),$$

where $\overline{T}_i$ and $\underline{T}_j$ are the upper and the lower expectation operators corresponding to the rows $\mathcal{P}_i$ and $\mathcal{P}_j$ respectively.

Instead of measuring the maximal distance between rows, we can only measure the distance between corresponding expectation operators, which turns out to coincide with the Hausdorff metric. Thus we define the weak coefficient of ergodicity as:

$$\rho(\underline{T}) = \max_{0\le f\le 1}\max_{i,j} |\underline{T}_i(f) - \underline{T}_j(f)| = d_H(\mathcal{P}_i, \mathcal{P}_j),$$

where $\underline{T}_i$ and $\underline{T}_j$ are ith and jth row lower expectation operators respectively, f a real valued map on the set of states, and d_H is the Hausdorff metric between the rows $\mathcal{P}_i$ and $\mathcal{P}_j$.

It is easy to see that $\rho(\underline{T}) \le \tau(\mathcal{P})$ if $\underline{T}$ is the lower expectation operator corresponding to the set of matrices $\mathcal{P}$.

Convergence

A stochastic matrix p whose coefficient of ergodicity $\tau(p)$ is strictly smaller than 1 is called scrambling. Further if $\mathcal{P}$ is a set of probabilities 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, and r is then called its scrambling integer. Thus we have that $\tau(\mathcal{P}^r) < 1$.

Hartfiel proves the following theorem:

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,

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

A similar result is obtained using the weak coefficient of ergodicity. 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.

The following theorem holds:

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,

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

In the case where an imprecise Markov chain is represented using convex sets of probabilities, convergence is assured by the lower transition operator being weakly product scrambling. However, when the sets of probabilities are not convex, the stronger condition of product scrambling is needed.

References

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

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

[3] D.J. Hartfiel. *Markov set-chains*. Springer-Verlag, Berlin, Heidelberg, New York, 1998.

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

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

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

[7] 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.

[8] 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.