Regular finite Markov chains with interval probabilities

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

Abstract

In Markov chain theory a stochastic matrix P is regular if some matrix power P^n contains only strictly positive elements. Regularity of transition matrix of a Markov chain guarantees the existence of a unique invariant distribution which is also the limiting distribution. In the present paper a similar result is shown for the generalized Markov chain model that replaces classical probabilities with interval probabilities. We generalize the concept of regularity and show that for a regular interval transition matrix sets of probabilities corresponding to consecutive steps of a Markov chain converge to a unique limiting set of distributions that only depends on transition matrix and is independent of the initial distribution. A similar convergence result is also shown for approximations of the invariant set.

Keywords. Markov chains, interval probabilities

1 Introduction

Markov chains are one of the most important tools to model random phenomena evolving in time. They are enough simple to allow detailed description but also enough general to allow many possibilities for applications (see [6]). A weak point of the most widely used model is that transition probabilities have to be constant and precisely known.

An attempt to relax this restriction was proposed in [7], where classical probabilities are replaced by interval probabilities. The approach presented there extends a previous approach given in [5], where the assumption of precisely known initial and transition probabilities is relaxed so that probability intervals are used instead of precise probabilities. Their model is based on the assumption that constant classical probabilities rule the process but only approximations are known instead of precise values. Several estimates based on this model are also given in [2] and [4].

Our approach presented in [7] uses the more general model of interval probabilities based on Weichselberger's theory (see [10] or [9]) instead of simple probability intervals, and omits the assumption that transition probabilities that rule the process are constant in time. In the sequel we refer to this model as *Markov chains with interval probabilities (MCIP)*. The model allows computation of possible probability distributions at consecutive steps and estimation of invariant distributions, which are of great importance in Markov chain theory. But there is a fundamental problem of those estimations that the sets of distributions corresponding to further steps become much more complicated than sets representable by interval probabilities. A way to overcome this problem is the use of approximations.

In this paper we examine the relationship between invariant sets of distributions and long term behaviour of generalized Markov chains. In the classical theory an important class of Markov chains, so called *regular chains*, has the property that its unique invariant distribution is also the limiting distribution to which probabilities converge after long time. Here we generalize the concept of regularity to MCIP and show that generalized regular Markov chains have a similar convergence property. Moreover, we show a similar result for a class of approximations with interval probabilities.

The paper has the following structure. In Section 2 we introduce basic concepts of the theory of interval probabilities and MCIP. In Section 3 we give our main results on convergence for MCIP.

2 Markov chains with interval probabilities

2.1 Interval probabilities

First we introduce basic elements of interval probability due to Weichselberger ([10]), some of them in a simplified form. Let Ω be a non-empty set and $\mathcal{A}$ a σ-algebra of its subsets. The term *classical probability* or *additive probability* will denote any set function $p\colon \mathcal{A} \to \mathbb{R}$ satisfying Kolmogorov's axioms. Let L and U be set functions on $\mathcal{A}$, such that $L \le U$ and $L(\Omega) = U(\Omega) = 1$. The interval valued function $P(\,.\,) = [L(\,.\,), U(\,.\,)]$ is called an *interval probability.*

To each interval probability P we associate the set $\mathcal{M}$ of all additive probability measures on the measurable space $(\Omega, \mathcal{A})$ that lie between L and U. This set is called the *structure* of the interval probability P. The basic class of interval probabilities are those whose structure is non-empty. Such an interval probability is denoted as R-field. The most important subclass of interval probabilities, F-fields, additionally assumes that both lower bound L and upper bound U are strict according to the structure $\mathcal{M}$:

$$L(A) = \inf_{p\in\mathcal{M}} p(A) \quad \text{and} \quad U(A) = \sup_{p\in\mathcal{M}} p(A) \tag{1}$$

for every $A \in \mathcal{A}$.

The above property is in a close relation to *coherence* in Walley's sense (see [8]). The difference is that the definition of coherence allows finitely additive probabilities while Weichselberger's model only allows σ-additive probabilities. However, in the case of finite probability spaces, both terms coincide, because finite additivity and σ-additivity then coincide. The requirement (1) implies the relation $U(A) = 1 - L(\neg A)$ for every $A \in \mathcal{A}$, and therefore, only one of the bounds L and U is needed. Usually we only take the lower one. Thus, an F-field is sufficiently determined by the triple $(\Omega, \mathcal{A}, L)$.

MCIP require several approximations involving lower expectations with respect to sets of probabilities. Let $\mathcal{C}$ be a set of probability measures on $(\Omega, \mathcal{A})$ and let a random variable $X\colon \Omega \to \mathbb{R}$ be given. The *lower* and the *upper expectation* $\underline{E}_{\mathcal{C}}X$ and $\overline{E}_{\mathcal{C}}X$ of X with respect to $\mathcal{C}$ are defined as the infimum and supremum of mathematical expectations of X with respect to members of $\mathcal{C}$:

$$\underline{E}_{\mathcal{C}}X = \inf_{p\in\mathcal{C}} E_pX$$
$$\overline{E}_{\mathcal{C}}X = \sup_{p\in\mathcal{C}} E_pX.$$

An important class of interval probabilities are those whose lower bounds L are 2-*monotone* (*convex, supermodular*), i.e. for every $A, B \subseteq \Omega$

$$L(A \cup B) + L(A \cap B) \ge L(A) + L(B). \tag{2}$$

If equality holds in the above equation the set function L is said to be *modular*, which in the case where $L(\emptyset) = 0$ is equivalent to finite additivity.

In the finite case, 2-monotonicity implies the F-property, which is then equivalent to coherence that is always implied by 2-monotonicity. Moreover, in the case of a 2-monotone coherent lower probability L on a finite measurable space, the lower and the upper expectation operators with respect to the corresponding structure can be found in terms of Choquet integral with respect to L and the corresponding upper probability U respectively, where *Choquet integral* with respect to a set function L is defined as

$$\int_\Omega X\,dL = \int_{-\infty}^{0} (L(X > t) - L(\Omega))\,dt + \int_0^\infty L(X > t)\,dt.$$

The right hand side integrals are both Riemann integrals. Further, if L is an additive measure, Choquet integral coincides with Lebesgue integral.

Let Ω be a finite set. If $\mathcal{M}$ is the structure of an F-field $P = (\Omega, \mathcal{A}, L)$ with L 2-monotone, we have

$$\underline{E}_{\mathcal{M}}X = \int_\Omega X\,dL \tag{3}$$

for every random variable X. (For the proof see e.g. [3], and note that for an infinite Ω, instead of the structure $\mathcal{M}$, the set of all finitely additive measures dominating L would be required for the above equality.) In fact, the equality in (3) for every X is equivalent to 2-monotonicity if the lower expectation is taken with respect to the set of all finitely additive measures dominating L. For a non-2-monotone L Choquet integral is in general lower than the lower expectation.

2.2 Markov chains with interval probabilities

Now we introduce the framework of MCIP model proposed in [7]. Let Ω be a finite set with elements $\{\omega_1, \ldots, \omega_m\}$ and 2^Ω the algebra of its subsets. Further let

$$X_0, X_1, \ldots, X_n, \ldots \tag{4}$$

be a sequence of random variables such that

$$P(X_0 = \omega_i) = q^{(0)}(\omega_i) =: q_i^0,$$

where $q^{(0)}$ is a classical probability measure on $(\Omega, 2^\Omega)$ such that

$$L^{(0)} \le q^{(0)}, \tag{5}$$

where $Q^{(0)} = (\Omega, 2^\Omega, L^{(0)})$ is an F-probability field. Thus, $q^{(0)}$ belongs to the structure $\mathcal{M}^{(0)}$ of $Q^{(0)}$. This means that initial probability distribution is not known precisely, but only a set of possible distributions is given as a structure of an F-field.

Transition probabilities in a classical finite Markov chain can be given by a matrix whose (i, j)-th entry represents the probability that the process that is in the state ω_i at time n will be in the state ω_j at time $n + 1$. Each row of a transition probability matrix is then a probability distribution on $(\Omega, 2^\Omega)$.

The idea of the generalized transition matrix is to replace classical probability distributions in rows with interval probabilities. Thus, suppose that

$$P\big(X_{n+1} = \omega_j \mid X_n = \omega_i, \\ X_{n-1} = \omega_{k_{n-1}}, \dots, X_0 = \omega_{k_0}\big) \\ = p_i^{n+1}(\omega_j) =: p_{ij}^{n+1}, \quad (6)$$

where p_{ij}^{n+1} is independent of $X_0, \dots, X_{n-1}$ for all $n \geq 1$, and

$$L_i \leq p_i^{n+1}, \quad (7)$$

where $P_i = (\Omega, 2^\Omega, L_i)$, for $1 \leq i \leq m$, is an F-probability field. Thus, p_{ij}^{n+1} are transition probabilities at time $n + 1$, and they are not assumed to be constant in n. Instead, on each step they are only supposed to satisfy the inequality (7), where L_i are constant in time.

The above generalization of transition matrices suggests the following generalization of the concept of stochastic matrix to interval probabilities. Let $P = [P_1 \dots P_m]^T$ where P_i are F-fields for $i = 1, \dots, m$. We will call such P an *interval stochastic matrix*. The *lower bound* of an interval stochastic matrix is simply $P_L := [L_1 \dots L_m]^T$, where L_i is the lower bound of P_i and the *structure* of an interval stochastic matrix is the set $\mathcal{M}(P)$ of stochastic matrices $p = (p_{ij})$ such that $p_i \geq L_i$, where p_i, for $i = 1, \dots, m$, is the classical probability distribution on $(\Omega, 2^\Omega)$ generated by $p_i(\omega_j) = p_{ij}$ for $j = 1, \dots, m$.

To represent an F field on a given probability space, one value has to be given for each event A; usually, this is the lower probability $L(A)$ of A. Thus, a row of an interval stochastic matrix can be represented as a row of $2^m - 2$ values, where $\emptyset$ and Ω, whose lower probabilities are always 0 and 1 respectively, are excluded. All other events correspond to each column in a given order. In general, this requires $m(2^m - 2)$ values for the transition matrix and $2^m - 2$ values for the initial distribution. The (i, j)-th entry of the transition matrix is then the lower probability of transition from the state ω_i to the set A_j.

We demonstrate this by the following example.

Example 1. Take $\Omega = \{\omega_1, \omega_2, \omega_3\}$. The algebra 2^Ω contains six non-trivial subsets, which we denote by $A_1 = \{\omega_1\}, A_2 = \{\omega_2\}, A_3 = \{\omega_3\}, A_4 = \{\omega_1, \omega_2\}, A_5 = \{\omega_1, \omega_3\}, A_6 = \{\omega_2, \omega_3\}$. Thus, besides $L(\emptyset) = 0$ and $L(\Omega) = 1$ we have to give the values $L(A_i)$ for $i = 1, \dots, 6$. Let the lower probability L of an interval probability Q be represented through the n-tuple

$$L = (L(A_1), L(A_2), L(A_3), L(A_4), L(A_5), L(A_6)) \quad (8)$$

and take $L = (0.1, 0.3, 0.4, 0.5, 0.6, 0.7)$. Further we represent the interval transition matrix P by a matrix with three rows and six columns, each row representing an element ω_i of Ω and the values in the row representing the interval probability P_i through its lower probability L_i. Take for example the following matrix:

$$P_L = \begin{pmatrix} 0.5 & 0.1 & 0.1 & 0.7 & 0.7 & 0.4 \\ 0.1 & 0.4 & 0.3 & 0.6 & 0.5 & 0.8 \\ 0.2 & 0.2 & 0.4 & 0.5 & 0.7 & 0.7 \end{pmatrix}. \quad (9)$$

The probability of transition from ω_1 to A_2 is thus at least 0.1, and to A_5 at least 0.7. Since $A_2 = \Omega - A_5$, the corresponding upper probability of transition from ω_1 to A_2 is $1 - 0.7 = 0.3$.

Note that the case where $|\Omega| = 3$ is somewhat specific, because every non-trivial subset is either atomic or a complement of an atomic set. Therefore, lower probabilities of the non-atomic sets can be obtained from the upper probabilities corresponding to atomic sets using $L(A) = 1 - U(\neg A)$. However, in general, lower probabilities given for all non-trivial subsets carry more information than probability intervals on atomic sets alone. Another specific feature of the case with $|\Omega| \leq 3$ is that the lower probability corresponding to any F-field is 2-monotone.

2.3 Computing distributions at further steps

The main advantage of Markov chains is that knowing the probability distribution at time n we can easily compute the distribution at time $n+1$. This is done by multiplying the given distribution with the transition matrix.

In the case of MCIP, where initial distribution as well as transition matrix are interval valued, we would want the probability distribution at the next step to be of a similar form. Thus, in an ideal case, the next step probability distribution would be an interval probability or even an F-field. But this is in general not possible. According to MCIP model, the actual distribution at each step is a classical probability distribution which is assumed to be a member of some set of distributions forming a structure of an interval probability. Similarly, the transition matrix is a classical transition matrix belonging to a set of matrices, also given in terms of interval probabilities.

Let $q^{(0)}$ be an initial distribution, thus satisfying (5), and p^1 a transition probability, satisfying (7). According to the classical theory, the probability at the next step is $q^{(1)} = q^{(0)}p^1$. Thus, the corresponding set of possible probability distributions at the next step must contain all the probability distributions of this form. Consequently, in the most general form, the set of probability distributions corresponding to X_k would be

$$\mathcal{C}_k := \{q^{(0)}p^1 \dots p^k \mid q^{(0)} \in \mathcal{M}(Q^{(0)}), \\ p^i \in \mathcal{M}(P) \text{ for } i = 1, \dots, k\}. \quad (10)$$

But these sets in general cannot be represented as structures of interval probabilities. Thus, they cannot be observed in terms of interval probabilities, or even in terms of convex sets. However, a possible approach using interval probabilities is to calculate the lower and the upper envelope of the set of probabilities obtained at each step and do further calculations with this interval probability and its structure. The resulting set of possible distributions at n-th step is then in general larger than $\mathcal{C}_k$, and could only be regarded as an approximate to the true set of distributions. In a similar way also more general convex envelopes of sets $\mathcal{C}_k$ can be constructed.

Approximation with interval probabilities

Here we describe how to compute approximations of the sets $\mathcal{C}_n$ with interval probabilities. We define a sequence $(Q^{(n)})_{n\geq 0}$ of F-fields, where $Q^{(0)}$ denotes the initial interval probability distribution, such that the structure $\mathcal{M}^{(n)}$ of each member of the sequence contains the set $\mathcal{C}_n$.

For every n let $Q^{(n+1)}$ be the F-field generated by the set of all products of the form $q^{(n)}p^{n+1}$ where $q^{(n)}$ belongs to the structure $\mathcal{M}(Q^{(n)})$ and p^{n+1} is a member of $\mathcal{M}(P)$. Such $Q^{(n+1)}$ is thus the narrowest F-field whose structure contains all the products $q^{(n)}p^{n+1}$. The products $q^{(n)}p^{n+1}$ would be the possible distributions at time $n+1$ if every $q^{(n)} \in \mathcal{M}(Q^{(n)})$ was a possible distribution at time n. Clearly, the inclusions $\mathcal{C}_n \subseteq \mathcal{M}(Q^{(n)}) = \mathcal{M}^{(n)}$ hold, but the intervals are in general wider than necessary to bound the sets $\mathcal{C}_n$. However, finding exact intervals is a computationally difficult problem.

Let $L^{(n)}$ be the lower probability corresponding to $Q^{(n)}$ and $L^{(n+1)}$ the one corresponding to $Q^{(n+1)}$. Further, let $q^{(n)}$ be any member of the structure $\mathcal{M}(Q^{(n)})$ and $q^{(n+1)}$ the corresponding distribution at time $n+1$. For every $A \subseteq \Omega$ we have

$$\begin{aligned} q^{(n+1)}(A) &= \sum_{\omega_j \in A} \sum_{i=1}^{m} q_i^{(n)} p_{ij}^{n+1} \\ &= \sum_{i=1}^{m} q_i^{(n)} \sum_{\omega_j \in A} p_{ij}^{n+1} \\ &= \sum_{i=1}^{m} q_i^{(n)} p_i^{n+1}(A) \\ &\geq \sum_{i=1}^{m} q_i^{(n)} L_i(A). \end{aligned} \quad (11)$$

Since p_i^{n+1} can be chosen independently of each other and of $q^{(n)}$ and because L_i have the F-property, they can be chosen so that

$$p_i^{n+1}(A) = L_i(A) \quad \text{for every } 1 \leq i \leq m.$$

Therefore, equality can be achieved in (11). Consequently, we obtain:

$$L^{(n+1)}(A) = \inf_{q^{(n)} \geq L^{(n)}} \sum_{i=1}^{m} q_i^{(n)} L_i(A). \quad (12)$$

The above infimum can be viewed as a lower expectation with respect to $\mathcal{M}^{(n)}$ of the function $X_A(\omega_i) := L_i(A)$.

If the lower probability $L^{(n)}$ is 2-monotone, (12) can (because of finiteness) equivalently be expressed in terms of Choquet integral (see e.g. [3])

$$L^{(n+1)}(A) = \int L_i(A)\, dL^{(n)} = \int X_A\, dL^{(n)}. \quad (13)$$

The above expression is linear in $L^{(n)}$ and thus requires significantly less computation to evaluate than (12). But even if both $L^{(n)}$ and L_i, for $1 \leq i \leq m$, are 2-monotone, the resulting lower probability $L^{(n+1)}$ need not be 2-monotone. Therefore, the use of (13) would in general produce less accurate results.

2.4 Invariant distributions

The invariant set of distributions

One of the main concepts in the theory of Markov chains is the existence of an invariant distribution. In the classical theory, an invariant distribution of a Markov chain with transition probability matrix P is any distribution q such that $q\,P = q$. In the case of regular Markov chain an invariant distribution is also the limiting distribution.

In MCIP model, a single transition probability matrix as well as initial distributions are replaced by sets of distributions given by structures of interval probabilities. Consequently, an invariant distribution is replaced by a set of distributions, which is invariant for the interval transition probability matrix P. An invariant set of distributions is thus a set $\mathcal{C}$ satisfying the condition

$$\mathcal{C} = \{qp \mid q \in \mathcal{C}, p \in \mathcal{M}(P)\}. \quad (14)$$

Thus, the invariant set of probabilities is closed for multiplication with the set of possible transition matrices. Of course, this does not mean that all its members are invariant distributions corresponding to some matrices from $\mathcal{M}(P)$, but it will follow from the construction that the largest such set must contain all those invariant distributions.

Given an interval transition matrix P it is in principle easy to find its largest invariant set of distributions. We start with the set $\mathcal{C}_0$ of all probability distributions on $(\Omega, 2^\Omega)$ and construct the sequence of sets of probability measures:

$$\mathcal{C}_{i+1} := \{q\,p \mid q \in \mathcal{C}_i, p \in \mathcal{M}(P)\}, \quad (15)$$

starting with $\mathcal{C}_0$. The above sequence corresponds to sequence (10), where the initial set of distributions is equal to the set of all probability distributions. In this case the sequence is monotone and the limiting set of distributions

$$\mathcal{C}_\infty := \bigcap_{i=1}^{\infty} \mathcal{C}_i. \quad (16)$$

is the largest invariant set of distributions.

The set $\mathcal{C}_\infty$ is non-empty because it obviously contains all invariant distributions of the matrices in $\mathcal{M}(P)$, and in the finite case invariant distributions always exist, although are not necessarily unique. Even though the invariant set of distributions is easy to find in principle, its shape can be very complicated and therefore approximations may be useful for practical purposes.

We have defined the invariant set of distributions as the limiting set of the sequence (10) starting with the set of all probability distributions. But this does not say anything about limiting set if the initial set is different. In Section 3 we show that the limiting set is unique and independent of the initial set $\mathcal{C}_0$ if a regularity condition is satisfied, which is the main result of this paper.

Approximating invariant distributions with interval probabilities

To approximate the invariant set of distributions with interval probabilities we try to find the F-field $Q = (\Omega, 2^\Omega, L)$ such that

$$L(A) = \inf_{q \in \mathcal{M}} \sum_{i=1}^{m} q_i L_i(A) \quad (17)$$

or in terms of lower expectations

$$L(A) = \underline{E}_{\mathcal{M}} X_A$$

where $X_A(\omega_i) = L_i(A)$. If the approximation with Choquet integral is used instead, the conditions become

$$L(A) = \int X_A \, dL \quad (18)$$

which is a system of linear equations with unknowns $L(A)$.

The minimal solution L of either of the sets of equations (17) or (18) approximates the largest invariant set of distributions $\mathcal{C}_\infty$ in the sense that all its members dominate L, or in other words, the set $\mathcal{C}_\infty$ is contained in the structure of the interval probability $(\Omega, 2^\Omega, L)$. This can be seen on the following way. Let $L^{(0)}$ be the lower probability with $L(A) = 0$ for every $A \subset \Omega$ and $L(\Omega) = 1$. It can be shown that both sequences of lower probabilities obtained through (12) and (13) starting with $L^{(0)}$ are monotone and therefore convergent. Clearly, their suprema are the minimal solutions of the equations (17) and (18) respectively. The inclusions $\mathcal{C}_n \subseteq \mathcal{M}^{(n)}$ for every $n \geq 0$ imply the required inclusion.

Example 2. We approximate the invariant set of distributions of the Markov chain with interval transition probability matrix given by the lower bound (9). We obtain the following solution to the system of equations (18):

$$L^{(\infty)} = (0.232, 0.2, 0.244, 0.581, 0.625, 0.6),$$

where $L^{(\infty)}$ is of the form (8). The intervals corresponding to the above solutions are then

$$P^{(\infty)} = ([0.232, 0.4], [0.2, 0.375], [0.244, 0.419], [0.581, 0.756], [0.625, 0.8], [0.6, 0.768]).$$

The above lower bound is of course only an approximation (from below) of the true lower bound for the invariant set of distributions. For comparison we include the lower bound of the set of invariant distributions corresponding to 100,000 randomly generated matrices dominating P_L:

$$(0.236, 0.223, 0.275, 0.587, 0.628, 0.608).$$

Since all invariant distributions of the members of the structure $\mathcal{M}(P)$ must belong to the set $\mathcal{C}_\infty$, the above lower bound is an approximation from above of the true lower bound and yields the intervals:

$$([0.236, 0.392], [0.223, 0.372], [0.275, 0.413], [0.587, 0.725], [0.628, 0.777], [0.608, 0.764]).$$

Thus, the lower bound of the true invariant set of distributions lies somewhere between the two approximations.

3 Convergence to equilibrium

3.1 Regular interval stochastic matrices

One of the main results of classical Markov chain theory is that chains with *irreducible* and *aperiodic* transition matrices always converge to a unique invariant distribution. Such transition matrices are sometimes called *regular*. In short, a transition matrix is regular if $p_{ij}^{(n)} > 0$ holds for all sufficiently large n, where $p_{ij}^{(n)}$ is the (i, j)-th entry of the matrix power P^n. Note that if all entries of P^r are strictly positive then also P^k, where $k > r$, has the same property. This follows from the properties of matrix multiplication and the fact that P has no zero rows. Therefore, a stochastic matrix is regular if all entries of P^r are strictly positive for at least one integer r.

If λ is any initial distribution and $(X_n)_{n\geq 0}$ is Markov (λ, P) with P regular then

$$P(X_n = j) \to \pi_j \quad \text{as } n \to \infty \text{ for all } j,$$

where π is the unique invariant distribution.

Regularity can similarly be defined for the case of Markov chains with interval probabilities. Let us first define the n-th power of an interval stochastic matrix P.

Definition 1. Let P be an interval stochastic matrix. We will call the set $\mathcal{P}^n = \{p_1 p_2 \dots p_n \mid p_i \in \mathcal{M}(P) \text{ for } i = 1, \dots, n\}$ the *n-th power of P*.

Note that the n-th power of an interval stochastic matrix is in general not an interval stochastic matrix, but a more general set of stochastic matrices, which are not easily tractable. Therefore, approximations in terms of interval probabilities will be useful. We also note that powers of interval stochastic matrices are associative in the sense that $\mathcal{P}^m \mathcal{P}^n = \mathcal{P}^{m+n}$, where the product of sets of matrices on the left hand side denotes the set of all products of matrices from corresponding sets.

Now we generalize the concept of regularity to interval stochastic matrices.

Definition 2. An interval stochastic matrix P is *regular* if there exists $n > 0$ such that $p_{ij} > 0$ for every $p \in \mathcal{P}^n$.

Clearly, this condition of regularity implies that every matrix in $\mathcal{M}(P)$ is regular, but inverse does not necessarily hold. In a similar way as in the classical case, it can be seen that if all matrices from $\mathcal{P}^n$ have strictly positive entries, then $\mathcal{P}^k$, where $k > n$, has the same property.

As we have pointed out before, powers of stochastic matrices as defined here are not easily tractable, thus checking regularity could be difficult in general. However, some simpler to check sufficient conditions easily follow from approximations presented before.

First we define two pseudo-powers for stochastic matrices that approximate powers from Definition 1. Both pseudo-powers are based on two operations similar to matrix multiplication, using approximations (12) and (13).

Definition 3. Let P be an interval stochastic matrix with lower probability matrix $P_L = [L_1 \dots L_m]^T$. Define $P_L^n = [L_1^n \dots L_m^n]^T$ where $L_i^1 = L_i$ and $L_i^n = \inf_{q \geq L_i^{n-1}} \sum_{i=1}^m q_j L_j(A)$ for $i = 1, \dots, m$ and $n \geq 2$.

Corollary 1. *If P is an interval stochastic matrix with lower probability matrix P_L such that $P_L^n = [L_1^n \dots L_m^n]^T$ and $L_i^n(A) > 0$ for every $i = 1, \dots, m$ and $A \subseteq \Omega,\ A \neq \emptyset$ then P is regular.*

Definition 4. Let P be an interval stochastic matrix with lower probability matrix $P_L = [L_1 \dots L_m]^T$. Define $\underline{P}_L^n = [\underline{L}_1^n \dots \underline{L}_m^n]^T$ where $\underline{L}_i^1 = L_i$ and $\underline{L}_i^n = \int L_j(A) d\underline{L}_i^{n-1}$ for $i = 1, \dots, m$ and $n \geq 2$, and the integral used is Choquet integral (as in (13)) .

Corollary 2. *If P is an interval stochastic matrix with lower probability matrix P_L such that $\underline{P}_L^n = [\underline{L}_1^n \dots \underline{L}_m^n]^T$ and $\underline{L}_i^n(A) > 0$ for every $i = 1, \dots, m$ and $A \subseteq \Omega,\ A \neq \emptyset$ then P is regular.*

The above corollaries present sufficient conditions for regularity because each power $\mathcal{P}^n$ as a set of stochastic matrices is contained within the structure of the corresponding pseudo-power, which is representable in terms of interval probabilities. Since powers from Definition 1 have no such representation, the sufficient conditions should be easier to check for pseudo-powers. Clearly, the sufficient condition in Corollary 2 implies the one in Corollary 1, but is much easier to check.

Even though the operations used in Definitions 3 and 4 resemble matrix multiplication, such a multiplication has an important weakness that it is not associative. But associativity is crucial in most methods concerning Markov chains and there is no obvious way to define an associative matrix multiplication for interval stochastic matrices, which is one of the main problems of the model.

3.2 Convergence to equilibrium

The main result of this section states that there is a unique compact set corresponding to a MCIP with a regular interval transition matrix to which its sets of distributions converge. To prove the theorem we use Banach fixed point theorem on the multivalued mapping between compact sets of probabilities corresponding to the transition matrix in Hausdorff metric.

Let (M,d) be a metric space. A mapping $T\colon M \to M$ is a *contraction* if there exists a constant $0 \le k < 1$ such that $d(Tx,Ty) \le k\, d(x,y)$ for all $x,y \in M$. If $k=1$ is allowed in the above condition then the mapping T is said to be *non-expansive.*

An element $x \in M$ is a *fixed point* of an operator T if $T(x) = x$.

Theorem 1 (Banach fixed point theorem)**.** *Let (M,d) be a non-empty complete metric space and $T\colon M \to M$ a contraction. Then there exists a unique fixed point $x \in M$ of T. Furthermore, this fixed point is the limit of the sequence $\{x_n\}_{n\in\mathbb{N}}$ where $x_{i+1} = Tx_i$ and x_0 is an arbitrary element of M.*

Given a metric space M and non-empty compact subsets $X, Y \subset M$, *Hausdorff distance* is defined as

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

This distance makes the set of non-empty compact sets a metric space $F(M)$. Moreover, if M is a compact space, so is $F(M)$ (see e.g. [1], p. 87). Note also that every compact metric space is complete.

To justify the use of Hausdorff metric, we show that all sets used are indeed compact. As the set of all probability distributions on a finite space is compact, we only have to note that the sets are closed. We start with a set of probabilities forming structure of an interval probability Q with lower probability L. Such a set is of the form $\mathcal{M}(Q) = \{q \mid q \text{ is a probability measure on } (\Omega, 2^{\Omega}), q \ge L\}$ and thus clearly closed and consequently compact. To see that $\mathcal{M}(P)$ is compact too, note that in topological sense it is a direct product of m structures corresponding to each row of P.

All sets of distributions corresponding to further steps are of the form $\mathcal{C}P = \{qp \mid q \in \mathcal{C}, p \in \mathcal{M}(P)\}$. Those sets are images of the compact sets $\mathcal{C} \times \mathcal{M}(P)$ with the continuous mapping $(q,p) \mapsto qp$, and are therefore compact too.

Proposition 1. *Let p be a stochastic matrix. Then the mapping from the set of all probability distributions $q \mapsto qp$ is non-expansive in the metric*

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

Moreover, if $p_{ij} > 0$ for every $1 \le i,j \le m$ and $k = 1 - \inf_{1\le i,j\le m} p_{ij}$ then the mapping $q \mapsto qp$ is a contraction and

$$d(qp, q'p) \le k\, d(q,q').$$

Proof. Take arbitrary $A \subseteq \Omega$ and let $p_i(A) = \sum_{\omega_j\in A} p_{ij}$. Further let q and q' be probability distributions on Ω with $q \ne q'$, and denote $B = \{\omega \in \Omega \mid q(\omega) \ge q'(\omega)\} \subsetneq \Omega$. Clearly, $k = \sup_{A\subsetneq\Omega} p_i(A)$ where $1 \le i \le m$.

We have

$$\begin{aligned}
&|qp(A) - q'p(A)| \\
&\quad= \left|\sum_{i=1}^{m} q_i p_i(A) - \sum_{i=1}^{m} q'_i p_i(A)\right| \\
&\quad= \left|\sum_{i=1}^{m} p_i(A)(q_i - q'_i)\right| \\
&\quad= \Bigg|\sum_{\omega_i\in B} p_i(A)|q_i - q'_i| \\
&\qquad\qquad - \sum_{\omega_i\notin B} p_i(A)|q_i - q'_i|\Bigg| \\
&\quad\le \max\Bigg\{\sum_{\omega_i\in B} p_i(A)|q_i - q'_i|, \\
&\qquad\qquad \sum_{\omega_i\notin B} p_i(A)|q_i - q'_i|\Bigg\} \\
&\quad\le \max\left\{\sum_{\omega_i\in B} k|q_i - q'_i|, \sum_{\omega_i\notin B} k|q_i - q'_i|\right\} \\
&\quad= k\max\left\{\sum_{\omega_i\in B} |q_i - q'_i|, \sum_{\omega_i\notin B} |q_i - q'_i|\right\} \\
&\quad\le k\, d(q,q')
\end{aligned}$$

Since $k \le 1$, the mapping is non-expansive. Furthermore, if $p_{ij} > 0$ for every $1 \le i,j \le m$ then $k < 1$ and thus the mapping is a contraction. ☐

The next proposition shows that the mapping $\mathcal{C} \mapsto \mathcal{C}\mathcal{P}^n$ is a contraction if P is a regular interval stochastic matrix and n is large enough.

Proposition 2. *Let P be a regular interval stochastic matrix and $n > 0$ an integer such that $p_{ij} > 0$ for every $p \in \mathcal{P}^n$ where $1 \le i,j \le m$. Let $k = 1 - \inf_{\substack{1\le i,j\le m\\ p\in\mathcal{P}^n}} p_{ij}$. The mapping $\mathcal{C} \mapsto \mathcal{C}\mathcal{P}^n = \{qp_1 \dots p_n \mid q \in \mathcal{C}, p_i \in \mathcal{M}(P) \text{ for } i = 1,\dots,n\}$ is then a contraction and*

$$d_H(\mathcal{C}\mathcal{P}^n, \mathcal{C}'\mathcal{P}^n) \le k\, d_H(\mathcal{C}, \mathcal{C}').$$

Proof. By the assumption, $p_{ij} > 0$ for every $p \in \mathcal{P}^n$

and $1 \le i,j \le m$. We have

$$d_H(\mathcal{C}\mathcal{P}^n, \mathcal{C}'\mathcal{P}^n) = \max\left\{ \sup_{\substack{q\in\mathcal{C}\\ p\in\mathcal{P}^n}} \inf_{\substack{q'\in\mathcal{C}'\\ p'\in\mathcal{P}^n}} d(qp, q'p'), \sup_{\substack{q'\in\mathcal{C}'\\ p'\in\mathcal{P}^n}} \inf_{\substack{q\in\mathcal{C}\\ p\in\mathcal{P}^n}} d(qp, q'p') \right\}.$$

Take for instance

$$\begin{aligned}
\sup_{\substack{q\in\mathcal{C}\\ p\in\mathcal{P}^n}} \inf_{\substack{q'\in\mathcal{C}'\\ p'\in\mathcal{P}^n}} d(qp, q'p') &\le \sup_{p\in\mathcal{P}^n} \sup_{q\in\mathcal{C}} \inf_{q'\in\mathcal{C}'} d(qp, q'p) \\
&\le \sup_{q\in\mathcal{C}} \inf_{q'\in\mathcal{C}'} k\, d(q, q') \\
&\le k\, d_H(\mathcal{C}, \mathcal{C}'),
\end{aligned}$$

where the second inequality follows from Proposition 1. Finally, this clearly implies $d_H(\mathcal{C}\mathcal{P}^n, \mathcal{C}'\mathcal{P}^n) \le k\, d_H(\mathcal{C}, \mathcal{C}')$. □

Finally we prove the main convergence theorem.

Theorem 2. *Let P be a regular interval stochastic matrix and $\mathcal{C}$ a compact set of probability distributions on $(\Omega, 2^\Omega)$ where Ω is a finite set. Then the sequence $\{\mathcal{C}\mathcal{P}^n\}_{n\in\mathbb{N}}$ converges in Hausdorff metric to a unique compact invariant set $\mathcal{C}_\infty$ that only depends on P and coincides with* (16).

Proof. Let $n > 0$ be an integer such that every $p \in \mathcal{P}^n$ satisfies $p_{ij} > 0$ for every $1 \le i, j, \le m$. By Proposition 2, the mapping $\mathcal{C} \mapsto \mathcal{C}\mathcal{P}^n$ is a contraction, and so, by Banach fixed point theorem, the sequence $\{\mathcal{C}(\mathcal{P}^n)^k\}_{k\in\mathbb{N}}$ converges to $\mathcal{C}_\infty$.

To see that the sequence $\{\mathcal{C}\mathcal{P}^k\}_{k\in\mathbb{N}}$ converges to the same set $\mathcal{C}_\infty$, we use associativity of powers of P. Thus, we have $\mathcal{C}\mathcal{P}^k = \mathcal{C}\mathcal{P}^r(\mathcal{P}^n)^s$, where $r < n$ and s goes to infinity as k goes to infinity. Since $\mathcal{C}\mathcal{P}^r$ is a compact set, the sequence converges to $\mathcal{C}_\infty$. □

3.3 Convergence of approximations

The limiting set of probabilities is computationally very difficult to find directly; therefore, approximations would be very useful in practice. Now we show that also a family of approximations converges independently from the initial distribution.

Proposition 3. *Let $P_L = [L_1 \dots L_m]^T$ be a lower transition probability matrix such that $L_i(A) < 1$ for every $1 \le i \le m$ and $A \subsetneq \Omega$. Let the mapping*

$$L \mapsto LP_L = L^*$$

be given, where L^* is the lower probability such that $L^*(A) = \int L_i(A)\, dL$. This mapping is then a contraction in the maximum distance metric

$$d(L, L') = \max_{A\subseteq\Omega} |L(A) - L'(A)|.$$

Further, if $k = \sup_{\substack{1\le i\le m\\ A\subsetneq\Omega}} L_i(A)$ then

$$d(LP_L, L'P_L) \le k\, d(L, L').$$

Proof. Take an arbitrary set $A \subsetneq \Omega$. We have:

$$|LP_L(A) - L'P_L(A)| = \left| \int L_i(A)\, dL - \int L_i(A)\, dL' \right|.$$

Let π be a permutation such that $L_{\pi(i)}(A) \ge L_{\pi(i+1)}(A)$ for every $1 \le i \le m$ and denote $S_i = \{\pi(1), \dots, \pi(i)\}$ and $x_i = L_{\pi(i)}(A)$ where $x_{m+1} = 0$. The above Choquet integrals can then be transformed into (see [3])

$$\begin{aligned}
&\left| \int L_i(A)\, dL - \int L_i(A)\, dL' \right| \\
&= \left| \sum_{i=1}^m (x_i - x_{i+1}) L(S_i) - \sum_{i=1}^m (x_i - x_{i+1}) L'(S_i) \right| \\
&= \left| \sum_{i=1}^m (x_i - x_{i+1})(L(S_i) - L'(S_i)) \right| \\
&\le \sum_{i=1}^m (x_i - x_{i+1}) d(L, L') \\
&= x_1\, d(L, L') \\
&\le k\, d(L, L').
\end{aligned}$$

Thus, $d(LP_L, L'P_L) = \max_{A\subseteq\Omega} |LP_L(A) - L'P_L(A)| \le k\, d(L, L')$, which completes the proof. □

In previous sections we approximated sets of distributions corresponding to consecutive steps by lower probabilities $L^{(n)}$ where $L^{(0)} = L$ is the initial lower probability and $L^{(n)}(A) = \int L_i(A)\, dL^{(n-1)}$. A problem with this approximation is its non-associativity, but associativity was crucial in the proof of Theorem 2.

Because of this inconvenience we can only prove a convergence theorem for a slightly different approximation. The construction easily implies that for every $n > 0$ the pseudo-power $\underline{P}_L^n$ approximates the n-th power of P in the sense that every $p \in \mathcal{P}^n$

satisfies $p \geq \underline{P}_L^n$. Therefore, the sequence of lower probabilities $\{L^{(kn)}\}_{k\in\mathbb{N}}$ defined by $L^{(0)} = L$ and $L^{(kn)} = L^{((k-1)n)}\underline{P}_L^n$, where L is the initial lower probability, approximates the sets of distributions at kn-th steps in the sense that $p \geq L^{(kn)}$ for every $p \in \mathcal{C}_{kn}$.

Now we give a convergence theorem for those approximations.

Theorem 3. *Let P be an interval stochastic matrix with the lower bound P_L such that $\underline{P}_L^n = [\underline{L}_1^n \dots \underline{L}_m^n]^T$ satisfies $\underline{L}_i^n(A) < 1$ for every $1 \leq i \leq m$ and $A \subsetneq \Omega$. Further let L be any lower probability on $(\Omega, 2^\Omega)$. Then the sequence $\{L^{(kn)}\}_{k\in\mathbb{N}}$ converges to a unique lower probability $L^{(\infty)}$ that only depends on $\underline{P}_L^n$.*

Proof. By Proposition 3 the mapping $L \mapsto L\underline{P}_L^n$ is a contraction in the maximum distance metric. By Banach fixed point theorem the sequence $\{L^{(kn)}\}_{k\in\mathbb{N}}$ converges to a unique lower probability $L^{(\infty)}$ which is the fixed point for the mapping $L \mapsto L\underline{P}_L^n$. □

4 Conclusion

Results in the paper show that even if the assumptions of Markov chain model are substantially relaxed, the behaviour remains similar as in the most widely used model with constant precisely known initial and transition probabilities. However, several interesting questions still remain open. Especially those related to approximations of the intractable true sets of distributions with convex sets representable with interval probabilities.

Acknowledgements

I wish to thank the referees for their helpful comments and suggestions.

References

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

[2] M. A. Campos, G. P. Dimuro, A. C. da Rocha Costa, and V. Kreinovich. Computing 2-step predictions for interval-valued finite stationary Markov chains. Technical Report UTEP-CS-03-20a, University of Texas at El Paso, 2003.

[3] D. Denneberg. *Non-additive measure and integral.* Kluwer Academic Publishers, Dordrecht, 1997.

[4] O. Kosheleva, M. Shpak, M. A. Campos, G. P. Dimuro, and A. C. da Rocha Costa. Computing linear and nonlinear normal modes under interval (and fuzzy) uncertainty. *Proceedings of the 25th International Conference of the North American Fuzzy Information Processing Society NAFIPS'2006*, 2006.

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

[6] J. Norris. *Markov Chains.* Cambridge University Press, Cambridge, 1997.

[7] D. Škulj. Finite discrete time Markov chains with interval probabilities. *Soft methods for integrated uncertainty modelling, (Advances in soft computing)*, pages 299–306, 2006.

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

[9] K. Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24:149–170, 2000.

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