Finite Discrete Time Markov Chains with Interval Probabilities

Damjan Škulj

Faculty of Social Sciences, University of Ljubljana
Kardeljeva ploščad 5, 1000 Ljubljana, Slovenia
Damjan.Skulj@Fdv.Uni-Lj.Si

Summary. A Markov chain model in generalised settings of interval probabilities is presented. Instead of the usual assumption of constant transitional probability matrix, we assume that at each step a transitional matrix is chosen from a set of matrices that corresponds to a structure of an interval probability matrix. We set up the model and show how to obtain intervals corresponding to sets of distributions at consecutive steps. We also state the problem of invariant distributions and examine possible approaches to their estimation in terms of convex sets of distributions, and in a special case in terms of interval probabilities.

1 Introduction

Interval probabilities present a generalised probabilistic model where classical single valued probabilities of events are replaced by intervals. In our paper we refer to Weichselberger's theory [4]; although, several other models also allow interval interpretation of probabilities.

An approach to involve interval probabilities to the theory of Markov chains was proposed by Kozine and Utkin [1]. They assume a model where transitional probability matrix is constant but unknown. Instead of that, only intervals belonging to each transitional probability are known.

In this paper we attempt to relax this model. We do this in two directions. First, we omit the assumption of the transitional probability matrix being constant, and second, instead of only allowing intervals to belong to single atoms, we allow them to belong to all subsets.

Allowing non-constant transitional probability matrix makes Markov chain model capable of modeling real situations where in general it is not reasonable to expect exactly the same transitional probabilities at each step. They can, however, be expected to belong to some set of transitional probabilities. In interval probability theory such sets are usually obtained as structures of interval probabilities. Our assumption is thus that transitional probability at each step is an arbitrary member of a set of transitional probability matrices generated by an interval probability matrix.

D. Škulj: *Finite Discrete Time Markov Chains with Interval Probabilities*, Advances in Soft Computing **6**, 299–306 (2006)
www.springerlink.com

A similar relaxation is also made to the initial distribution. Instead of a single distribution, we allow a set of distributions forming a structure of an interval probability.

Our goal is to estimate the interval probabilities after a number of steps and to find an invariant set of distributions. Unfortunately, it turns out that interval probabilities are not always sufficient to represent distributions obtained after some steps, which can form very general sets of distributions, that may not be easy to represent. The method based on interval probabilities can thus only approximate the true sets of distributions. To overcome this drawback, we provide a method to at least in principle approximate the corresponding sets of distributions with convex sets of probability distributions with arbitrary precision. In those settings, interval probabilities only present an easy to handle special case.

2 Basic Definitions and Model Setup

First we introduce basic elements of interval probability theory due to Weichselberger [4], but some of them are used here in a simplified form. Let Ω be a non-empty set and $\mathscr{A}$ a σ-algebra of its subsets. The term *classical probability* or *additive probability* will denote any set function $p\colon \mathscr{A} \to \mathbb{R}$ satisfying Kolmogorov's axioms. Let L and U be set functions on $\mathscr{A}$, 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 $\mathscr{M}$ of all additive probability measures on the measurable space $(\Omega, \mathscr{A})$ 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 an *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:

$$L(A) = \inf_{p \in \mathscr{M}} p(A) \qquad \text{and} \qquad U(A) = \sup_{p \in \mathscr{M}} p(A) \quad \text{for every } A \in \mathscr{A}. \tag{1}$$

The above property is in a close relation to *coherence* in Walley's sense (see [3]), in fact, in the case of finite probability spaces both terms coincide. Because of the requirement (1) 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 $(\Omega, \mathscr{A}, L)$, and therefore, we will from now on denote F-fields in this way.

Now we introduce the framework of our Markov chain model. Let Ω be a finite set with elements $\{\omega_1, \ldots, \omega_m\}$ and $\mathscr{A} := 2^\Omega$ be the algebra of its subsets. Further let

$$X_0, X_1, \ldots, X_n, \ldots \tag{2}$$

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 $(\Omega, \mathscr{A})$ such that

$$L^{(0)} \leq q^{(0)}, \tag{3}$$

where $Q^{(0)} = (\Omega, \mathscr{A}, L^{(0)})$ is an F-probability field. Thus $q^{(0)}$ belongs to the structure $\mathscr{M}^{(0)}$ of $Q^{(0)}$.

Further, suppose that

$$P\left(X_{n+1} = \omega_j \mid X_n = \omega_i, X_{n-1} = \omega_{k_{n-1}}, \ldots, X_0 = \omega_{k_0}\right) = p_i^{n+1}(\omega_j) =: p_{ij}^{n+1}, \tag{4}$$

where p_{ij}^{n+1} is independent of $X_0, \ldots, X_{n-1}$ and

$$L_i \leq p_i^{n+1}, \tag{5}$$

where $P_i = (\Omega, \mathscr{A}, L_i)$, for $1 \leq i \leq m$, is an F-probability field. Thus p_{ij}^{n+1} are transitional probabilities at time $n+1$; however, they do not need to be constant, but instead, on each step they only satisfy (5), where L_i are constant. Thus, the transitional probabilities are not constant in the usual sense but only in the sense of interval probabilities.

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

Thus, the transitional probabilities are given in terms of interval stochastic matrices. Under the above conditions, the probability distribution of each X_n will be given in terms of an F-field $Q^{(n)} = (\Omega, \mathscr{A}, L^{(n)})$. Thus

$$P(X_n = \omega_i) = q^{(n)}(\omega_i) =: q_i^n,$$

where $q^{(n)}$ is a classical probability measure on $(\Omega, \mathscr{A})$ such that

$$L^{(n)} \leq q^{(n)}.$$

We will call a sequence (2) with the above properties an *interval Markov chain*. An advantage of presenting sets of probability measures with interval probabilities is that only one value has to be given for each set to determine an interval probability. Usually, this is the lower probability $L(A)$ of an event A. In general this requires $m(2^m - 2)$ values for the transitional matrix and $2^m - 2$ values for the initial distribution. We demonstrate this by the following example.

Example 1. Take $\Omega = \{\omega_1, \omega_2, \omega_3\}$. The algebra $\mathscr{A} = 2^\Omega$ 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

$$L = (L(A_1), L(A_2), L(A_3), L(A_4), L(A_5), L(A_6)), \tag{6}$$

and take $L = (0.1, 0.3, 0.4, 0.5, 0.6, 0.7)$. Further we represent the interval transitional 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:

$$P_L = \begin{pmatrix} 0.5 & 0.1 & 0.1 & 0.7 & 0.7 & 0.4 \\ 0.1 & 0.4 & 0.3 & 0.6 & 0.5 & 0.8 \\ 0.2 & 0.2 & 0.4 & 0.5 & 0.7 & 0.7 \end{pmatrix}. \tag{7}$$

In the next section we will show how how to obtain the lower probability at the second step, given the lower bounds for Q and P.

3 Calculating Distributions at n-th Step

The main advantage of Markov chains is that knowing the probability distribution at time n we can easily calculate the distribution at the next time. This is done simply by multiplying the given distribution with the transitional matrix.

In the generalised case we consider a set of probability distributions and a set of transitional matrices, given as structures of the corresponding interval probabilities. The actual distribution as well as the actual transitional probability matrix can be any pair of members of the two sets. Let $q^{(0)}$ be an initial distribution, thus satisfying (3), and p^1 a transitional probability, satisfying (5). According to the classical theory, the probability at the next step is $q^{(1)} = q^{(0)}p^1$. Thus, the corresponding set of probability distributions on 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

$$\mathscr{M}_k := \{q^{(0)}p^1 \dots p^k \mid q^{(0)} \in \mathscr{M}(Q^{(0)}), p^i \in \mathscr{M}(P) \text{ for } i = 1,\dots,k\}. \tag{8}$$

But these sets of probability distributions are in general not structures of interval probability measures. Thus, they can not be observed in terms of interval probabilities. 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 the 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 $\mathscr{M}_k$, and could only be regarded as an approximate to the true set of distributions.

The advantage of the approach in terms of interval probabilities is that the calculations are in general computationally less difficult and that some calculations, such as the calculation of invariant distributions, can be done directly through systems of linear equations. As we shall see, the level of precision of estimates is very flexible and can be adjusted depending on our needs and the imprecision of the data.

Now we give such a procedure for a direct calculation of the lower bound $L^{(n+1)}$ under the assumption that the set of probabilities at n-th step is given in terms of an interval probability $Q^{(n)}$. Let π_A be a permutation on the set $\{1,\dots,m\}$ such that $L_{\pi_A(i)}(A) \geq L_{\pi_A(i+1)}(A)$ for $1 \leq i < m-1$ and denote $A_i := \bigcup_{k=1}^{i}\{\omega_{\pi_A(k)}\}$ where $A_0 = \emptyset$. Define the probability measure

$$q^{\pi_A}_{\pi_A(i)} = q^{\pi_A}(\omega_{\pi_A(i)}) := L^{(n)}(A_i) - L^{(n)}(A_{i-1}). \tag{9}$$

The set function $L^{(n+1)}$ is then the infimum of the set of all distributions from the structure of $Q^{(n)}$ multiplied by all members of $\mathscr{M}(P)$. It turns out that it can be directly calculated as

$$L^{(n+1)}(A) = \sum_{i=1}^{m} q_i^{\pi_A} L_i(A). \tag{10}$$

Example 2. Let us calculate the second step probability distribution on the data of Example 1. Let the lower bound $L^{(0)}$ of $Q^{(0)}$ be as in the previous example, $L^{(0)} = (0.1, 0.3, 0.4, 0.5, 0.6, 0.7)$ and let the transitional probability be given by its lower bound P_L from the same example. Further, let $L^{(1)}$ be the lower bound of the interval probability distribution at step 1, $Q^{(1)}$. By (10) we get

$$L^{(1)} = (0.19, 0.23, 0.28, 0.56, 0.62, 0.64).$$

4 Invariant Distributions

4.1 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 transitional probability matrix P is any distribution q such that $qP = q$. In the case of *ergodic* Markov chain an invariant distribution is also the limit distribution.

In our case, a single transitional probability matrix as well as initial distributions are replaced by sets of distributions given by structures of interval probabilities. Consequently, an invariant distribution has to be replaced by a set of distributions, which is invariant for the interval transitional probability matrix P. It turns out, that there always exists a set $\mathscr{M}$ such that

$$\mathscr{M} = \{qp \mid q \in \mathscr{M}, p \in \mathscr{M}(P)\} \tag{11}$$

and that for every initial set of probability distributions $\mathscr{M}_0$ late enough members of sequence (8) converge to $\mathscr{M}$.

For simplicity we may always assume the initial distribution to be the set of all probability measures on $(\Omega, \mathscr{A})$, which is equal to the structure of the interval probability $Q_0 = [0,1]$. Thus, from now on, let $\mathscr{M}_0 := \{q \mid q \text{ is a probability measure on } (\Omega, \mathscr{A})\}$. Clearly, the sequence (8) with initial set of distributions $\mathscr{M}_0$ includes all sequences with any other initial set of distributions.

Consider the following sequence of sets of probability measures:

$$\mathscr{M}_{i+1} := \{qp \mid q \in \mathscr{M}_i, p \in \mathscr{M}(P)\}, \tag{12}$$

z starting with $\mathscr{M}_0$. The above sequence corresponds to sequence (8) with initial set of distributions equal to $\mathscr{M}_0$.

It is easy to see that the sequence (12) is monotone: $\mathscr{M}_{i+1} \subseteq \mathscr{M}_i$, and thus we can define the limiting set of distributions by

$$\mathscr{M}_\infty := \bigcap_{i=1}^{\infty} \mathscr{M}_i. \tag{13}$$

The above set is clearly non-empty, since it contains all eigenvectors of all stochastic matrices from $\mathscr{M}(P)$ corresponding to eigenvalue 1. It is well known that such eigenvectors always exist. Besides, this set clearly satisfies the requirement (11). Thus, we will call the set (13) *the invariant set of distributions* of an interval Markov chain with the interval transitional probability matrix P.

The above definition of the invariant set of an interval Markov chain gives its construction only in terms of limits, but it says nothing about its nature, such as, whether it is representable in terms of the structure of some interval probability or in some other way. However, it is important that such a set always exists.

4.2 Approximating the Invariant Set of Distributions with Convex Sets of Distributions

Since the invariant set of distributions of an interval Markov chain in general does not have a representation in terms of a structure of an interval probability or maybe even in terms of a convex set, we have to find some methods to at least approximate it with such sets.

For every closed convex set $\mathscr{M}$ of probability distributions on $(\Omega, \mathscr{A})$ there exists a set of linear functionals $\mathscr{F}$ and a set of scalars $\{l_f \mid f \in \mathscr{F}\}$ such that

$$\mathscr{M} = \{p \mid p \text{ is a probability measure on } (\Omega, \mathscr{A}), f(p) \geq l_f\ \forall f \in \mathscr{F}\}. \tag{14}$$

Example 3. If the set of functionals is equal to the natural embedding of the algebra $\mathscr{A}$ then the resulting set of distributions forms the structure of an F-probability field: $f_A(p) := p(A)$, $l_{f_A} := L(A)$ and $P = (\Omega, \mathscr{A}, L)$.

Moreover, the set of functionals may correspond to even a smaller set, like a proper subset of $\mathscr{A}$, such as the set of atoms in $\mathscr{A}$ yielding a structure of an interval probability with additional properties.

Thus, every structure of an interval probability may be represented by a set of distributions of the form (14).

Now fix a set of functionals $\mathscr{F}$ and an interval stochastic matrix P and define the following sequence of sets of probability distributions, where $\mathscr{M}_0$ is the set of all probability measures on $(\Omega, \mathscr{A})$:

$$\begin{aligned}
\mathscr{M}_{0,\mathscr{F}} &:= \mathscr{M}_0; \\
\mathscr{M}'_{i+1,\mathscr{F}} &:= \{qp \mid q \in \mathscr{M}_{i,\mathscr{F}}, p \in \mathscr{M}(P)\} \\
\mathscr{M}_{i+1,\mathscr{F}} &:= \{q \mid f(q) \geq \inf_{q' \in \mathscr{M}'_{i+1,\mathscr{F}}} f(q') \ \forall f \in \mathscr{F}\}.
\end{aligned}$$

The way the set $\mathscr{M}_{i+1,\mathscr{F}}$ is obtained from $\mathscr{M}'_{i+1,\mathscr{F}}$ is similar to the concept of *natural extension* for a set of lower previsions (see e.g. [3]).

The idea of the above sequence is to replace the sets $\mathscr{M}'_{i,\mathscr{F}}$, which are difficult to handle, with sets of distributions representable by linear functionals in $\mathscr{F}$. In the special case from Example 3 such a set is the structure of some interval probability.

The following properties are useful:

(i) If $\mathscr{F}' \subset \mathscr{F}$ then $\mathscr{M}_{i,\mathscr{F}'} \supseteq \mathscr{M}_{i,\mathscr{F}} \supseteq \mathscr{M}_i$ holds for every $i \in \mathbb{N} \cup \{0\}$, where $\mathscr{M}_i$ is a member of the sequence (12).

(ii) The inclusion $\mathscr{M}_{i+1,\mathscr{F}} \subseteq \mathscr{M}_{i,\mathscr{F}}$ holds for every $i \in \mathbb{N}$. Thus, the sequence $(\mathscr{M}_{i,\mathscr{F}})$ is monotone and this implies existence of a limiting set of distributions for every set of functionals $\mathscr{F}$:

$$\mathscr{M}_{\infty,\mathscr{F}} := \bigcap_{i \in \mathbb{N}} \mathscr{M}_{i,\mathscr{F}}.$$

The sets of distributions $\mathscr{M}_{\infty,\mathscr{F}}$ all comprise the set $\mathscr{M}_\infty$ and can be in some important cases found directly through a system of linear equations.

(iii) The set $\mathscr{M}_{\infty,\mathscr{F}}$ is a maximal set among all sets $\mathscr{M}$ with the property:

$$\inf_{q \in \mathscr{M}} f(q) = \inf_{\substack{q \in \mathscr{M} \\ p \in \mathscr{M}(P)}} f(q \cdot p) \ \forall f \in \mathscr{F}. \tag{15}$$

While the sets $\mathscr{M}_{\infty,\mathscr{F}}$ only approximate the invariant set of distributions $\mathscr{M}_\infty$ from below, it can clearly be approximated from above by the set $\mathscr{M}_e$ containing all eigenvectors of the stochastic matrices from the structure $\mathscr{M}(P)$.

4.3 Approximating Invariant Distributions with Interval Probabilities

The important special case of convex sets of probabilities is the case of structures of interval probabilities. For this case the conditions (15) translate to a system of linear equations with $2^m - 2$ unknowns. We obtain this case by considering the linear functionals on the set of probability measures on $(\Omega, \mathscr{A})$ of the form f_A, where $f_A(q) = q(A)$ for every probability measure q:

$$\mathscr{F}_{\mathscr{A}} = \{f_A \mid A \in \mathscr{A}\}.$$

The set of inequalities (15) can now be rewritten in terms of lower probabilities L and L_i to obtain:

$$L(A) = \sum_{i=1}^{m} q_i^{\pi_A} L_i(A) \ \forall A \in \mathscr{A}. \tag{16}$$

Recall that $q_i^{\pi_A}$ are expressible in terms of L as given by (9). The invariant set of distributions is then the structure of the F-field $Q_\infty = [L,U]$, where L is the minimal solution of the above system of linear equations, as follows from (iii).

Example 4. We approximate the invariant set of distributions of the Markov chain with interval transitional probability matrix given by the lower bound (7). We obtain the following solution to the system of equations (16):

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

where $L^{(\infty)}$ is of the form (6).

As we have pointed out earlier, the above lower bound is only an approximation of the true lower bound for the invariant set of distributions. For comparison we give the lower bound of the set of eigenvalues of 100,000 random matrices dominating P_L:

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

which is an approximation from above. Thus, the lower bound of the true invariant set of distributions lies between the above approximations.

References

[1] I. Kozine and L. V. Utkin. Interval-valued finite markov chains. *Reliable Computing*, 8(2):97–113, 2002.
[2] J. Norris. *Markov Chains.* Cambridge University Press, Cambridge, 1997.
[3] P. Walley. *Statistical reasoning with imprecise probabilities.* Chapman and Hall, London, New York, 1991.
[4] K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I – Intervallwarscheinlichkeit als umfassendes Konzept.* Physica-Verlag, Heidelberg, 2001.