Reliable Computing **8:** 97–113, 2002.

Interval-Valued Finite Markov Chains

IGOR O. KOZINE
Systems Analysis Department, Risø National Laboratory, P.O. Box 49, 4000 Roskilde, Denmark, e-mail: igor.kozine@risoe.dk

and

LEV V. UTKIN
Department of Computer Science, Forest Technical Academy, Institutski per. 5, St.Petersburg, 194021, Russia, e-mail: lvu@utkin.usr.etu.spb.ru

(Received: 24 November 2000; accepted: 6 July 2001)

Abstract. The requirement that precise state and transition probabilities be available is often not realistic because of cost, technical difficulties or the uniqueness of the situation under study. Expert judgements, generic data, heterogeneous and partial information on the occurrences of events may be sources of the probability assessments. All this source information cannot produce precise probabilities of interest without having to introduce drastic assumptions often of quite an arbitrary nature. In this paper the theory of interval-valued coherent previsions is employed to generalise discrete Markov chains to interval-valued probabilities. A general procedure of interval-valued probability elicitation is analysed as well. In addition, examples are provided.

1. Introduction

The importance of Markov chain theory to many sciences and applications has been widely recognised. This theory is employed in the social and biological sciences, in reliability theory to assess the probabilities of states of reparable systems, by decision-makers in different areas, etc. However, there are some obstacles in applying Markov chains in practical situations. The major problem is that a large number of precise probability estimates is necessary to fully specify a model. The requirement that precise state and transition probabilities be available is often not realistic. In practice there are often few samples of occurrences to produce confident precise estimates. In addition, the evidence about the likelihood of the occurrence of an event may be subjective, or the results of comparative judgements, or a conclusion drawn from generic data. Moreover, some information provides evidence about the moments of a random value, while other information gives evidence about mean values like average losses or gains. All the evidence is chance related and partial. From these considerations, two questions arise: How can all available sources of information be used to quantify the present state of knowledge, preferably without having to introduce new assumptions; and how can the degree of uncertainty be preserved while propagating source knowledge to the likelihood of state occurrences in the future?

The recent theory of interval-valued coherent previsions provides tools to elicit lower and upper probabilities from heterogeneous and partial information on the occurrence of events and extend the probabilities in a consistent way to other domains. Precise probabilities are a particular case in the framework of this theory when the amount of data on the occurrence is exhaustive. The theory allows constructing interval-valued probabilities without having to assume a specific precise prior probability distribution. Any probability distribution satisfying some constraints defined by available information is considered possible if no evidence exist that a specific distribution is more relevant. A comprehensive presentation of the theory is given in [2] and [4].

This theory is employed here to generalise discrete Markov chains to coherent interval-valued probabilities. A general procedure for the elicitation of the interval-valued probabilities from heterogeneous and partial information is also described.

2. Basic Concepts of Interval-Valued Coherent Previsions

Throughout the paper we will adhere to terminology used by P. Walley (see, for example, [4]) and which experts in the field of imprecise probabilities are used to.

Suppose that $X(\omega)$ is a real-valued mapping on a set Ω called the possibility space or the universe of discourse, $\omega \in \Omega$. Often X is referred as a gamble (see, for example, [4]). In particular, X can be regarded as a function of a random variable. Assume that there exist some upper $\overline{P}(\omega)$ and lower $\underline{P}(\omega)$ distribution functions of the random variable ω. Then the upper and lower previsions (expectations) are

$$\overline{M}(X) = \sum_{\omega \in \Omega} X(\omega)\overline{P}(\omega), \qquad \underline{M}(X) = \sum_{\omega \in \Omega} X(\omega)\underline{P}(\omega).$$

The coherent interval-valued probabilities are a particular case of interval-valued coherent previsions and based on three fundamental principles: *avoiding sure loss, coherence* and *natural extension*. All these principles as well as lower and upper previsions and probabilities have behavioural interpretations that can be found in [4]. This paper addresses technical issues of constructing interval-valued coherent models and any interpretations of those are not discussed.

The principle of *avoiding sure loss* for the lower and upper probabilities is equivalent to holding the following inequalities:

$$\left.\begin{array}{ll} \underline{P}'(\Omega) = \sum_{i=1}^{n} \underline{P}'(A_i) \leq 1, & \overline{P}'(\Omega) = \sum_{i=1}^{n} \overline{P}'(A_i) \geq 1, \\ 0 \leq \underline{P}'(A_i) \leq \overline{P}'(A_i) \leq 1, & i = 1, \ldots, n, \end{array}\right\} \tag{2.1}$$

where A_i, $i = 1, \ldots, n$, form a set of all pairwise-disjoint subsets whose union is the possibility space Ω. The designation of the probabilities with the prime notation indicates that the probabilities should be considered initial and not necessarily coherent.

The construction of coherent interval-valued statistics and probabilities of events different from A_i is performed through the *natural extension*. The natural extension is a general mathematical procedure for calculating new previsions from the initial judgements. It produces a coherent overall model from an arbitrary collection of imprecise probability judgements and may be seen as the basic constructive step in statistical reasoning. The natural extension can have many forms of representation. One of them (discrete case) can be written as two linear programming problems

$$\underline{M}g = \inf_{\{P(A_i)\}} \sum_{i=1}^{n} g(A_i)P(A_i), \tag{2.2}$$

$$\overline{M}g = \sup_{\{P(A_i)\}} \sum_{i=1}^{n} g(A_i)P(A_i), \tag{2.3}$$

subject to

$$\sum_{i=1}^{n} P(A_i) = 1, \qquad \underline{P}'(A_i) \leq P(A_i) \leq \overline{P}'(A_i). \tag{2.4}$$

Infimum and supremum in the above problems are taken over all possible probability distributions $\{P(A_i)\}$ over A_i, $i = 1, \ldots, n$ satisfying the constraints (2.4). Function g can be, for example, $g = x_i$, where x_i is a real number assigned to A_i, and then $\underline{M}g$ and $\overline{M}g$ are the lower and upper mean values of a random variable defined on Ω, or it can be that $g = x_i^2$ and then, $\underline{M}g$ and $\overline{M}g$ are the lower and upper second moments of the same random variable. If g is a characteristic function of an event $B = A_j \cup \cdots \cup A_k$, $j < k \leq n$, i.e. $g = I_B(A_i) = 1$ if $A_i \in B$ and $I_B(A_i) = 0$ if $A_i \notin B$, then $\underline{M}g = \underline{P}(B)$ and $\overline{M}g = \overline{P}(B)$.

The lower and upper mean values $\underline{M}g$ and $\overline{M}g$ or $\underline{P}(B)$ and $\overline{P}(B)$ obtained as the solutions of linear programming problems (2.2) and (2.3) subject to (2.4) are referred to as *coherent*. In [2] and [4] more general definitions of the natural extension can be found.

It is unrealistic to expect that the initial probabilities $\underline{P}'(A_i)$ and $\overline{P}'(A_i)$, $i = 1, \ldots, n$, elicited from heterogeneous and partial source information be coherent. Therefore, through the natural extension they can be corrected and made coherent as follows:

$$\underline{P}(A_i) = \inf_{\{P(A_i)\}} P(A_i), \qquad i = 1, \ldots, n, \tag{2.5}$$

$$\overline{P}(A_i) = \sup_{\{P(A_i)\}} P(A_i), \qquad i = 1, \ldots, n, \tag{2.6}$$

subject to (2.4).

Here equations (2.5) and (2.6) are obtained from (2.2) and (2.3) by substituting $g(A_i) = I_{A_i}(A_i) = 1$ into (2.2) and (2.3).

The sense of the natural extension in precise mathematical terms is to estimate the interval $[\underline{M}g, \overline{M}g]$ of possible values of Mg for all probability distributions for which $\underline{P}'(A_i) \leq P(A_i) \leq \overline{P}'(A_i)$, $i = 1, \ldots, n$. That is, we assume that any probability distribution, consistent with the initial judgements $\underline{P}'(A_i) \leq P(A_i) \leq \overline{P}'(A_i)$ for $i = 1, \ldots, n$, is possible, and we base our inferences on this assumption without preferring a particular distribution.

It can be inferred from the problem (2.5) and (2.6) subject to (2.4) that the coherent lower and upper probabilities are obtained from the initial probabilities avoiding sure loss according to the formulas [2]:

$$\left.\begin{aligned} \overline{P}(A_i) &= \min\left\{\overline{P}'(A_i),\ 1 - \sum_{j=1, j\neq i}^{n} \underline{P}'(A_i)\right\}, \\ \underline{P}(A_i) &= \max\left\{\underline{P}'(A_i),\ 1 - \sum_{j=1, j\neq i}^{n} \overline{P}'(A_i)\right\}, \qquad i = 1, \ldots, n. \end{aligned}\right\} \tag{2.7}$$

3. Interval-Valued Discrete Markov Chains

The basic result of precise discrete Markov chains can be written as follows [1]: $b_j(k) = \sum_{i=1}^{n} b_i(k-1)a_{ij}(k)$, where $b_j(k)$, $k = 1, 2, \ldots$ is the probability that at the k-th trial the chain is in the state s_j, $j = 1, \ldots, n$, and $a_{ij}(k)$ are one-step transition probabilities from state s_i to state s_j for each pair $i, j = 1, 2, \ldots, n$, and $b_j(0)$ and $a_{ij}(k)$ are known for any $i, j = 1, 2, \ldots, n$. For the sake of simplicity throughout the paper we will write a_{ij} instead of $a_{ij}(k)$, since we assume a_{ij} are constants.

In the light of the interval-valued probabilities we consider that probabilities $b_j(0)$ and a_{ij} are not known precisely but rather belong to intervals. That is, $\underline{b}'_i(0) \leq b_i(0) \leq \overline{b}'_i(0)$ and $\underline{a}'_{ij} \leq a_{ij} \leq \overline{a}'_{ij}$ for $i, j = 1, \ldots, n$, and $\underline{b}'_i(0)$, $\overline{b}'_i(0)$, $\underline{a}'_{ij}$ and $\overline{a}'_{ij}$ are known for any $i, j = 1, \ldots, n$.

The problem of obtaining the coherent interval-valued probabilities $\underline{b}_j(1)$ and $\overline{b}_j(1)$ for states $j = 1, \ldots, n$ at trial 1 consistent with the initial interval-valued probabilities can be formulated as follows:

$$\underline{b}_j(1) = \inf_{b_i, a_{ij}} \sum_{i=1}^{n} b_i(0)a_{ij}, \qquad j = 1, \ldots, n, \tag{3.1}$$

$$\overline{b}_j(1) = \sup_{b_i, a_{ij}} \sum_{i=1}^{n} b_i(0)a_{ij}, \qquad j = 1, \ldots, n, \tag{3.2}$$

subject to

$$\left.\begin{aligned} &\sum_{i=1}^{n} b_i(0) = 1, && \sum_{j=1}^{n} a_{ij} = 1, && i = 1, \ldots, n, \\ &\underline{b}'_i(0) \leq b_i(0) \leq \overline{b}'_i(0), && \underline{a}'_{ij} \leq a_{ij} \leq \overline{a}'_{ij}, && i, j = 1, \ldots, n. \end{aligned}\right\} \tag{3.3}$$

Thus, in order to find the lower coherent probabilities $\underline{b}_j(1)$ for $j = 1, \dots, n$ we have to solve the non-linear programming problems (3.1) under the conditions (3.3), and in order to find the upper coherent probabilities $\overline{b}_j(1)$ we have to solve the non-linear programming problems (3.2) under the conditions (3.3). These tasks can be simplified to linear programming problems.

From (2.7) the coherence conditions for the state and transfer probabilities are the following:

$$\underline{b}_j(k) \geq 1 - \sum_{i=1,\, i \neq j}^{n} \overline{b}_i(k), \quad \overline{b}_j(k) \leq 1 - \sum_{i=1,\, i \neq j}^{n} \underline{b}_i(k), \qquad j = 1, \dots, n, \tag{3.4}$$

$$\underline{a}_{ij} \geq 1 - \sum_{m=1,\, m \neq j}^{n} \overline{a}_{im}, \qquad \overline{a}_{ij} \leq 1 - \sum_{m=1,\, m \neq j}^{n} \underline{a}_{im}, \quad i, j = 1, \dots, n. \tag{3.5}$$

Now a theorem can be formulated.

THEOREM 3.1. *Given the transition probabilities* $\underline{a}_{ij}$ *and* $\overline{a}_{ij}$, $i, j = 1, \dots, n$ *are coherent and the initial state probabilities* $\underline{b}'_i(0)$, $\overline{b}'_i(0)$, $i = 1, \dots, n$ *avoid sure loss, the coherent lower and upper state probabilities can be found as solutions of* $2n$ *separate linear programming problems*

$$\underline{b}_j(k) \;=\; \inf_{b_i} \sum_{i=1}^{n} b_i(k-1)\underline{a}_{ij}, \qquad j = 1, \dots, n, \tag{3.6}$$

$$\overline{b}_j(k) \;=\; \sup_{b_i} \sum_{i=1}^{n} b_i(k-1)\overline{a}_{ij}, \qquad j = 1, \dots, n, \tag{3.7}$$

subject to

$$\left.\begin{array}{ll} \displaystyle\sum_{i=1}^{n} b_i(k-1) = 1, & i = 1, \dots, n, \\ \underline{b}_i(k-1) \leq b_i(k-1) \leq \overline{b}_i(k-1), & i = 1, \dots, n. \end{array}\right\} \tag{3.8}$$

Note. In (3.8) the prime notation is omitted. This can be done because even though the initial probabilities $\underline{b}'_i(0)$ and $\overline{b}'_i(0)$ are not necessarily coherent, the condition $\sum_{i=1}^{n} b_i(0) = 1$ imposes the coherency constraint and only the coherent $b_i(0)$ will be chosen when solving the problem. In the following trials, $k = 1, 2, \dots$, $\underline{b}_i(k)$ and $\overline{b}_i(k)$ will be coherent according to the theorem and the primes have to be omitted.

Proof. Let us prove first that problems (3.1)–(3.3) are equivalent to (3.6)–(3.8). Indeed, the objective functions (3.1) and (3.2) contain only the columns of the matrix $\{a_{ij}\}$, that is $i = 1, \dots, n$, and j is fixed. The condition of normalisation $\sum_{j=1}^{n} a_{ij} = 1$, $i = 1, \dots, n$, is imposed only on the rows rather than on the columns. Therefore, this condition can be removed from the list of constraints without affecting the results. Thus, we obtain the problems

$$\underline{b}_j(k) = \inf_{b_i, a_{ij}} \sum_{i=1}^{n} b_i(k-1)a_{ij}, \qquad j = 1, \dots, n,$$

$$\overline{b}_j(k) = \sup_{b_i, a_{ij}} \sum_{i=1}^{n} b_i(k-1)a_{ij}, \qquad j = 1, \dots, n,$$

subject to

$$\left.\begin{array}{l} \sum_{i=1}^{n} b_i(k-1) = 1, \\ \underline{b}'_i(k-1) \le b_i(k-1) \le \overline{b}'_i(k-1), \qquad \underline{a}_{ij} \le a_{ij} \le \overline{a}_{ij}, \qquad i = 1, \dots, n. \end{array}\right\}$$

Since the condition of normalisation is not valid for the columns, we are free to choose values a_{ij} from the range $\underline{a}_{ij} \le a_{ij} \le \overline{a}_{ij}$. If the objective function is to achieve the supremum, we choose a_{ij} as large as possible from the range $\underline{a}_{ij} \le a_{ij} \le \overline{a}_{ij}$. Such values are $\overline{a}_{ij}$. In order to achieve the infimum of the objective function, we choose a_{ij} as small as possible. Such values are $\underline{a}_{ij}$. Having this done, the conditions $\underline{a}_{ij} \le a_{ij} \le \overline{a}_{ij}$, $i = 1, \dots, n$ can also be removed from the list of the constraints since they have already been taken into consideration. Thus, having started from problems (3.1)–(3.3), we arrive at problems (3.6)–(3.8).

Now let us prove the coherence of the state probabilities obtained from problems (3.6)–(3.8). By using the property of coherent probabilities (2.1), it follows:

$$\begin{aligned}
\underline{b}_m(k) &= \inf_{b_i(k-1)} \sum_{i=1}^{n} b_i(k-1)\underline{a}_{im} \ge \inf_{b_i(k-1)} \sum_{i=1}^{n} b_i(k-1) \left(1 - \sum_{j=1, j\ne m}^{n} \overline{a}_{ij}\right) \\
&= \inf_{b_i(k-1)} \left[\sum_{i=1}^{n} b_i(k-1) - \sum_{i=1}^{n} b_i(k-1) \sum_{j=1, j\ne m}^{n} \overline{a}_{ij}\right] \\
&= \inf_{b_i(k-1)} \left[1 - \sum_{i=1}^{n} b_i(k-1) \sum_{j=1, j\ne m}^{n} \overline{a}_{ij}\right] \\
&= 1 - \sup_{b_i(k-1)} \sum_{i=1}^{n} b_i(k-1) \sum_{j=1, j\ne m}^{n} \overline{a}_{ij} \\
&= 1 - \sup_{b_i(k-1)} \sum_{j=1, j\ne m}^{n} \sum_{i=1}^{n} b_i(k-1)\overline{a}_{ij}.
\end{aligned}$$

Then, it can be stated that $\sup \sum_i x_i \le \sum_i \sup x_i$. Indeed, if all the constraints are of the form $\overline{x}_i = \sup x_i \le x_i$ for $i = 1, \dots, n$, then $\sup \sum_i x_i = \sum_i \sup x_i$. On the other hand, if x_i, $i = 1, \dots, n$ or their functions are subject to some more constraints (this takes place in the case of interest: $x_i = b_i(k-1)\overline{a}_{ij}$ and $\sum_{i=1}^{n} b_i(k-1) = 1$), then $\sup \sum_i x_i$ is achieved not necessarily at points $x_i^0 = \sup x_i$ but at $x_i^0 \leq \sup x_i$. Therefore $\sup \sum_i x_i \leq \sum_i \sup x_i$. Thus, we arrive at the condition of coherence

$$\underline{b}_m(k) \geq 1 - \sum_{j=1, j\neq m}^{n} \sup_{b_i(k-1)} \sum_{i=1}^{n} b_i(k-1)\overline{a}_{ij} = 1 - \sum_{j=1, j\neq m}^{n} \overline{b}_j(k).$$

The case of the upper probabilities is proved similarly. □

It is clear from (3.6) and (3.7) that if the state and transition probabilities are precise we will get the conventional formula $b_j(k) = \sum_{i=1}^{n} b_i(k-1)a_{ij}$. In this case there is no need to solve the linear programming problems.

The existence of asymptotic solutions for some classes of Markov chains is a practically valued feature and it is worth considering whether the interval-valued discrete Markov chains converge to the limiting upper and lower state probabilities. Consider the class of ergodic stationary Markov chains. The ergodicity of a conventional (precise) chain implies $b_j(k) = b_j(k-1)$, $j = 1, \dots, n$ as $k \to \infty$ (see, for example, [1]).

Let us prove a theorem about the convergence of the lower and upper coherent state probabilities.

THEOREM 3.2. *If a Markov chain is ergodic, then the lower $\underline{b}_j(k)$ and upper $\overline{b}_j(k)$ state probabilities, obtained from Theorem 3.1, obey the conditions $\underline{b}_j(k) = \underline{b}_j(k-1)$ and $\overline{b}_j(k) = \overline{b}_j(k-1)$, $j = 1, \dots, n$ as $k \to \infty$.*

Proof. Let us return to the initial non-linear optimisation problem:

$$\underline{b}_j(k) = \inf \sum_{i=1}^{n} b_i(k-1)a_{ij}, \qquad \overline{b}_j(k) = \sup \sum_{i=1}^{n} b_i(k-1)a_{ij},$$

subject to

$$\left.\begin{array}{lll} \sum_{i=1}^{n} b_i(k-1) = 1, & \underline{b}_i(k-1) \leq b_i(k-1) \leq \overline{b}_i(k-1), & \\ \sum_{j=1}^{n} a_{ij} = 1, & \underline{a}_{ij} \leq a_{ij} \leq \overline{a}_{ij}, & i, j = 1, \dots, n. \end{array}\right\}$$

Let $b_i(0)$, $i = 1, \dots, n$ be an arbitrary distribution of the state probabilities compounded from the set

$$\mathcal{B} = \{\underline{b}_i(0) \leq b_i(0) \leq \overline{b}_i(0),\ i = 1, \dots, n\},$$

and $\|a_{ij}\|$ be an arbitrary transition matrix (with arbitrary distributions of the transition probabilities in rows) compounded from the set

$$\mathcal{A} = \{\underline{a}_{ij} \leq a_{ij} \leq \overline{a}_{ij},\ i,j = 1,\dots,n\}.$$

By the asymptotic properties of Markov ergodic chains in the classical probabilistic context

$$b_j(k) = \sum_{i=1}^{n} b_i(0)a_{ij}^{(k)} = \sum_{i=1}^{n} b_i(0)a_{ij}^{(k-1)} = b_j(k-1), \qquad k \to \infty,$$

where $a_{ij}^{(k)}$ is an element of the transition matrix to the power of k, it follows that

$$\underline{b}_j(k) = \inf_{\mathcal{B},\mathcal{A}} \sum_{i=1}^{n} b_i(0)a_{ij}^{(k)} = \inf_{\mathcal{B},\mathcal{A}} \sum_{i=1}^{n} b_i(0)a_{ij}^{(k-1)} = \underline{b}_j(k-1), \qquad k \to \infty.$$

The case of the upper state probabilities is proved similarly. □

4. Example 1

Consider a three-state Markov process with the initial assessments

$$\{\underline{b}_j'(0)\} = \{0.21; 0.27; 0.27\}, \qquad \{\overline{b}_j'(0)\} = \{0.31; 0.61; 0.4\},$$

$$\|\underline{a}_{ij}'\| = \left\| \begin{matrix} 0.7 & 0.05 & 0.01 \\ 0.15 & 0.6 & 0.08 \\ 0.02 & 0.7 & 0.1 \end{matrix} \right\|, \qquad \|\overline{a}_{ij}'\| = \left\| \begin{matrix} 0.9 & 0.4 & 0.3 \\ 0.3 & 0.8 & 0.2 \\ 0.1 & 0.9 & 0.2 \end{matrix} \right\|.$$

Here the rows of matrices correspond to different i, while the columns correspond to different j. It can be easily checked by (2.1) that all the probabilities $\underline{b}_i'(0)$, $\overline{b}_i'(0)$, $\underline{a}_{ij}'$ and $\overline{a}_{ij}'$ avoid sure loss, for example, $\sum_{j=1}^{3} \underline{a}_{1j}' = 0.76 \leq 1$. Avoiding sure loss for the matrixes is equivalent to avoiding sure loss for each of the rows. We can hardly expect the initial probabilities to be coherent. Let us make them coherent according to (2.7)

$$\{\underline{b}_j(0)\} = \{0.21; 0.29; 0.27\}, \qquad \{\overline{b}_j(0)\} = \{0.31; 0.52; 0.4\},$$

$$\|\underline{a}_{ij}\| = \left\| \begin{matrix} 0.7 & 0.05 & 0.01 \\ 0.15 & 0.6 & 0.08 \\ 0.02 & 0.7 & 0.1 \end{matrix} \right\|, \qquad \|\overline{a}_{ij}\| = \left\| \begin{matrix} 0.9 & 0.29 & 0.25 \\ 0.3 & 0.77 & 0.2 \\ 0.1 & 0.88 & 0.2 \end{matrix} \right\|.$$

Transformation from the non-coherent to the coherent probabilities is illustrated in Figures 1 and 2 in the simplex representation for the initial state probabilities $\underline{b}_i'(0)$ and $\overline{b}_i'(0)$, $i = 1, 2, 3$. Each assessment is represented by a line parallel to one side of the simplex. Probability $b_i(0)$ in the simplex is the perpendicular distance of the line from the side opposite vertex i.

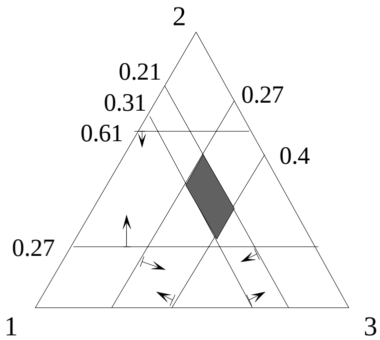

Figure 1. Model produced by incoherent (but avoiding sure loss) initial state probabilities $\underline{b}_i'(0)$ and $\overline{b}_i'(0)$, $i = 1, 2, 3$.

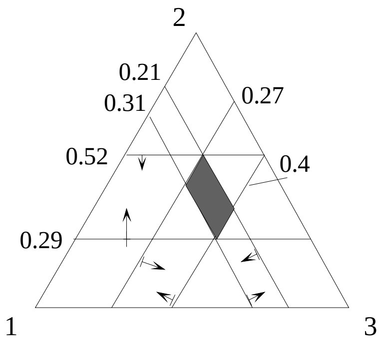

Figure 2. Model produced by coherent initial state probabilities $\underline{b}_i(0)$ and $\overline{b}_i(0)$, $i = 1, 2, 3$.

The propagation of the lower and upper state probabilities $\underline{b}_i(0)$ and $\overline{b}_i(0)$ for $i = 1, 2, 3$ and $k = 0, \ldots, 21$ is illustrated in Figures 3–5.

5. Construction of Interval-Valued Probabilities

The initial interval-valued probabilities are elicited from some source information. Source information may be from a small sample of the occurrences of an event A, analogies with known probabilities (precise or interval-valued), direct judgements of probability intervals by an expert, or some other probability-related pieces of evidence of the occurrences of event A. The analyst should be able to construct an integrating assessment which takes into account all the evidence rather than discard any information. Some pieces of evidence may be conflicting and there must be a mechanism for their allocation. Conflicting information should be subject to scrutiny, modification, disposal or combination.

The following definition can be formulated: *Conflicting pieces of evidence are those if considered separately produce (through the natural extension) disjoint probability intervals*.

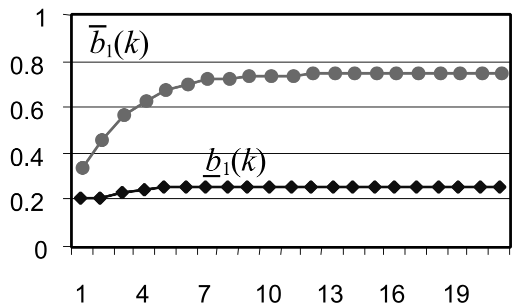

Figure 3. Propagation of lower and upper probabilities of state s_1.

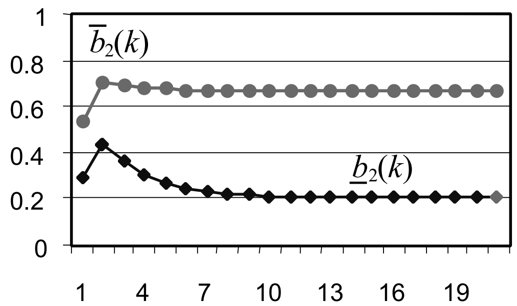

Figure 4. Propagation of lower and upper probabilities of state s_2.

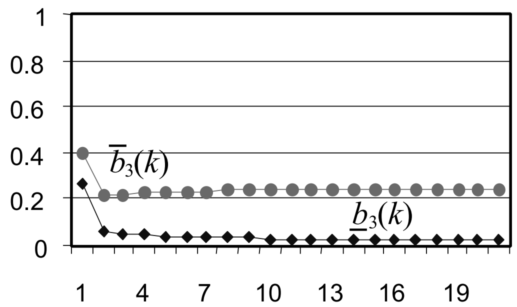

Figure 5. Propagation of lower and upper probabilities of state s_3.

Source information on the likelihood of the occurrences of the states and transitions is not necessarily given as interval valued but likelihood-related and partial. For example, regarding the probabilities of states, we may know the interval-valued probabilities of only some of the states or we do not know them at all. Some evidence might be available in the form of comparative judgements, like "being at state s_j is at least λ times as probable as being at state s_m" or "being at state s_j or s_l

is at least as probable as being at state s_m." Other evidence can be represented in the form of mean values on the possibility set or comparative judgements on them. All this source information, obtained prior to the modelling, defines the interval-valued state probabilities at 0 trial and imposes constraints on possible state probabilities at trial 1. Similarly, source information about the transition probabilities can also be partial.

In the case of partial information, there exist two approaches to assessing the interval-valued state probabilities at the next trial: 1) using partial source information as constraints for the direct assessment of the interval-valued state probabilities at the next trial, 2) using partial source information as constraints for the assessment of the interval-valued state and/or transition probabilities at the current trial, and, as the following step, calculating the interval-valued probabilities at the next trial.

Suppose that there are r judgements about state probabilities and h_i judgements about transition probability for the i-th row of the transition matrix. The first, one-step problem, is formulated as follows:

$$\underline{b}_j(k) = \inf \sum_{i=1}^{n} b_i(k-1)a_{ij}, \qquad \overline{b}_j(k) = \sup \sum_{i=1}^{n} b_i(k-1)a_{ij}, \tag{5.1}$$

subject to

$$\left. \begin{array}{lll} \sum_{i=1}^{n} b_i(k-1) = 1, \qquad \sum_{j=1}^{n} a_{ij} = 1, & & \\ \underline{B}_l(k-1) \leq \sum_{i=1}^{n} f_l(i) b_i(k-1) \leq \overline{B}_l(k-1), & & l = 1, \ldots, r, \\ \underline{A}_u(i) \leq \sum_{j=1}^{n} g_u(i,j) a_{ij} \leq \overline{A}_u(i), & u = 1, \ldots, h_i, & i = 1, \ldots, n, \\ \underline{a}_{ij} \leq a_{ij} \leq \overline{a}_{ij}, \qquad i,j = 1, \ldots, n. & & \end{array} \right\} \tag{5.2}$$

where $f_i(i)$ and $g_u(i,j)$ are gambles, and $\underline{B}_l(k-1)$ and $\overline{B}_l(k-1)$ and $\underline{A}_u(i)$ and $\overline{A}_u(i)$ are lower and upper previsions of the gambles. If $f_l(i) = x_i$, where $x \in \mathbf{R}$, then $\underline{B}_l(k-1)$ and $\overline{B}_l(k-1)$ can be regarded as the lower and upper mean values of x at time $k-1$. If $f_l(i)$ is the characteristic function of the state s_l, i.e. $f_l(i) = I_l(i) = 1$ for $i = l$ and $f_l(i) = I_l(i) = 0$ for $i \neq l$, then $\underline{B}_l(k-1)$ and $\overline{B}_l(k-1)$ are lower and upper probabilities of the state s_l. Functions $f_l(i)$ can model comparative judgements. So, the judgement "being at state s_j is at least λ times as probable as being at state s_m" is formalized as

$$\sum_{i=1}^{n} [I_j(i) - \lambda I_m(i)] b_i(k-1) \geq 0.$$

The judgement "being at state s_j and s_l is at least as probable as being at state s_m" is modelled as

$$\sum_{i=1}^{n}[I_j(i) + I_l(i) - \lambda I_m(i)]b_i(k-1) \geq 0.$$

There may be a number of comparative judgements on mean values. A vast variety of source information can be represented in the form of gambles and utilised through the natural extension (5.1)–(5.2).

The problem (5.1)–(5.2) is non-linear due to the non-linearity of the objective functions, but it can be simplified to a linear programming problem.

First, let us prove the following theorem assuming that the constraints are not conflicting:

THEOREM 5.1. *The lower and upper transition probabilities obtained as a solution of the linear programming problems*

$$\underline{a}_{ij} = \inf a_{ij}, \qquad \overline{a}_{ij} = \sup a_{ij}, \qquad i,j = 1,\ldots,n, \tag{5.3}$$

subject to

$$\underline{A}_u(i) \leq \sum_{j=1}^{n} g_u(i,j)a_{ij} \leq \overline{A}_u(i), \quad u = 1,\ldots,h_i,\; i = 1,\ldots,n,\; \sum_{j=1}^{n} a_{ij} = 1 \tag{5.4}$$

are coherent.

Proof. Let a_{ij}^c be the probability of an event complementary to a_{ij}. Then, for any fixed i $\{a_{ij}, j = 1,\ldots,n\}$ is a discrete probability distribution that $\sum_{j=1}^{n} a_{ij} = 1$ and

$$\underline{a}_{ij} = \inf a_{ij} = \inf(1 - a_{ij}^c) = 1 - \sup a_{ij}^c = 1 - \overline{a}_{ij}^c.$$

Using the notation $s_i \to s_j$ for a transition from state s_i to state s_j and the property of the interval-valued probabilities for two events A and B of $\overline{P}(A \cup B) \leq \overline{P}(A)+\overline{P}(B)$ and $\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)$ [2] and [4] it follows:

$$\underline{a}_{ij} = 1 - \overline{a}_{ij}^c = 1 - \overline{P}\left\{\bigcup_{m=1, m\neq j}^{n} s_i \to s_m\right\} \geq 1 - \sum_{m=1, m\neq j}^{n} \overline{a}_{im}.$$

The last inequality is the condition of coherence (see (2.7)). The case of the upper probabilities is proved similarly. □

Assuming that the initial judgements are not conflicting, the following theorem can also be proved:

THEOREM 5.2. *Given matrixes* $\|\underline{a}_{ij}\|$ *and* $\|\overline{a}_{ij}\|$, $i,j = 1,\ldots,n$ *are coherent, the state probabilities obtained as the solutions of the following problem*

$$\underline{b}_j(k) = \inf \sum_{i=1}^{n} b_i(k-1)\underline{a}_{ij}, \qquad \overline{b}_j(k) = \sup \sum_{i=1}^{n} b_i(k-1)\overline{a}_{ij}, \tag{5.5}$$

subject to

$$\underline{B}_l(k-1) \leq \sum_{i=1}^{n} f_l(i) b_i(k-1) \leq \overline{B}_l(k-1), \qquad l = 1, \ldots, r, \ \sum_{i=1}^{n} b_i(k-1) = 1, \tag{5.6}$$

are coherent as well.

Proof. Similar to Theorem 3.1. □

As it stated at the beginning of this section, there exist two approaches to assess the interval-valued state probabilities. The first one is based on Theorem 5.2 and suggests to solve the problem in one step assuming that the transition probabilities have been made coherent (Theorem 5.1). On the other hand, it may be computationally easier to split problem (5.5)–(5.6) into two problems. So, the second approach would be to make the interval-valued state probabilities in the current step coherent (they are designated $\underline{b}_j^*(k-1)$ and $\overline{b}_j^*(k-1)$)

$$\underline{b}_j^*(k-1) = \inf b_j(k-1), \qquad \overline{b}_j^*(k-1) = \sup b_j(k-1), \tag{5.7}$$

subject to

$$\underline{B}_l(k-1) \leq \sum_{i=1}^{n} f_l(i) b_i(k-1) \leq \overline{B}_l(k-1), \qquad l = 1, \ldots, r, \ \sum_{i=1}^{n} b_i(k-1) = 1, \tag{5.8}$$

and then, to find coherent state probabilities in the next step assuming that the transition probabilities have been made coherent. That is

$$\underline{b}_j(k) = \inf \sum_{i=1}^{n} b_i(k-1) \underline{a}_{ij}, \qquad \overline{b}_j(k) = \sup \sum_{i=1}^{n} b_i(k-1) \overline{a}_{ij}, \tag{5.9}$$

subject to

$$\sum_{i=1}^{n} b_i(k-1) = 1, \qquad \underline{b}_j^*(k-1) \leq b_j(k-1) \leq \overline{b}_j^*(k-1). \tag{5.10}$$

Thus, if source information on the transition matrix and initial state probabilities is a set of evidence represented as the constraints (5.2), there exist two ways of obtaining the interval-valued state probabilities in the next step.

Approach 1.

a) Calculating the lower and upper coherent transition probabilities (solving the problem (5.3)–(5.4)),

b) calculating the lower and upper coherent state probabilities in the next step (solving the problem (5.5)–(5.6)).

Approach 2.

a) Calculating the lower and upper coherent transition probabilities (solving the problem (5.3)–(5.4)),

b) calculating the lower and upper coherent state probabilities in the current step (solving the problem (5.7)–(5.8)),

c) calculating the lower and upper coherent state probabilities in the next step (solving the problem (5.9)-(5.10)).

The second approach has an advantage of the traceability of the intermediate results despite the fact that it assumes the solving of three optimisation problems.

As it was mentioned above, the set of constraints and evidence available can be conflicting and in this case Theorems 5.1 and 5.2 are not valid. How can we combine conflicting evidence in the case where there are no arguments to neglect any of the judgements or change the evidence?

There is a method devised to address this issue: (1) to allocate conflicting bodies of evidence (in the simplex representation they are disjoint areas), (2) to calculate the lower and upper probabilities separately for each body of evidence (each set of conflicting evidence produces a set of the interval-valued probabilities of interest), and (3) to combine the calculated probabilities based on the theorem proved below.

THEOREM 5.3. *If there are two disjoint intervals produced by separately coherent lower and upper probabilities of the same event so that* $[\underline{q}_i, \overline{q}_i]$, $[\underline{p}_i, \overline{p}_i]$, *and* $\underline{p}_i > \overline{q}_i$ *for some of* $i = 1, \ldots, n$, *then the combination of these two intervals is* $[\underline{q}_i, \overline{p}_i]$.

Proof. To prove this theorem it is sufficient to prove that the probabilities defining the interval $[\underline{q}_i, \overline{p}_i]$ are coherent with both the intervals $[\underline{q}_i, \overline{q}_i]$ and $[\underline{p}_i, \overline{p}_i]$. Since the probabilities are coherent separately, it holds

$$\underline{q}_i \geq 1 - \sum_{j=1, j\neq i}^{n} \overline{q}_j, \qquad \overline{p}_i \leq 1 - \sum_{j=1, j\neq i}^{n} \underline{p}_i.$$

Then it also holds that

$$\underline{q}_i \ \geq \ 1 - \sum_{j=1, j\neq i}^{n} \overline{q}_j \geq 1 - \underline{q}_i > 1 - \sum_{j=1, j\neq i}^{n} \overline{p}_j,$$

$$\overline{p}_i \ \leq \ 1 - \sum_{j=1, j\neq i}^{n} \underline{p}_i < 1 - \sum_{j=1, j\neq i}^{n} \underline{q}_i.$$

The last inequalities prove that $\underline{q}_i$ is coherent with the set of the upper probabilities $\overline{p}_i$, $i = 1, \ldots, n$. Otherwise $\overline{p}_i$ is coherent with the set of the lower probabilities $\underline{q}_i$, $i = 1, \ldots, n$. □

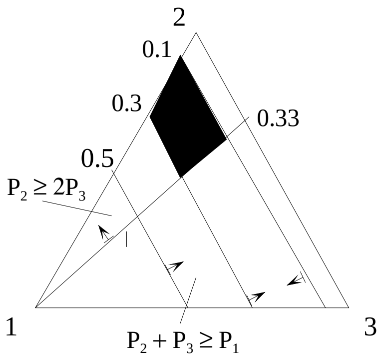

Figure 6. Model produced by evidence.

According to [2] and [4] the rule of the combination of conflicting probabilities is

$$\underline{P}(X) = \min\{\underline{P}_1(X), \underline{P}_2(X)\}, \qquad \overline{P}(X) = \max\{\overline{P}_1(X), \overline{P}_2(X)\},$$

which is a direct consequence of Theorem 5.3. According to Walley [4] this rule is called the unanimity rule.

6. Example 2

There are three states s_1, s_2 and s_3. Source information on the probabilities of the occurrences of these events is the following: (1) $P(s_3) \in [0.1, 0.3]$, (2) s_2 is at least two times as probable as s_3, and (3) s_2 and s_3 is at least as probable as s_1. What are the probabilities $P(s_2)$ and $P(s_3)$?

The source evidence can be rewritten in the form of inequalities (1) $0.1 \leq p_1 \leq 0.3$, (2) $p_2 \geq 2p_3$, and (3) $p_2 + p_3 \geq p_1$. These inequalities and the constraint $p_1 + p_2 + p_3 = 1$ define a constrained area which is shown in Figure 6. Figure 7 demonstrates how the interval-valued probabilities are constructed. The numerical calculated values of the probabilities are $\underline{P}(s_2) = 0.466$, $\overline{P}(s_2) = 0.9$, $\underline{P}(s_3) = 0$, $\overline{P}(s_3) = 0.3$.

It is seen from Figure 6 that the evidence $p_2 + p_3 \geq p_1$ does not contribute to the precision and can be discarded. That is, the black area, defining the lower and upper probabilities, does not change if this evidence is removed from the set of evidence.

Let us say there is more evidence: if s_1 occurs one will get the gain equal to 1, if s_2—the gain is 2, if s_3—the gain is 3. Assume that based on subsidiary evidence a rational expert is willing to bet on the event that the evarage gain is not less than 2. This judgement produces one more constraint of $2 \leq 1p_1 + 2p_2 + 3p_3 \leq 3$ and can be utilized and make the resulting probabilities more precise. So, it is seen from Figure 8 that the black rectangular defines more precise probabilities: $\underline{P}(s_2) = 0.5$,

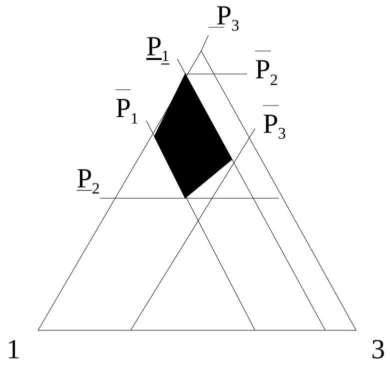

Figure 7. Imprecise probabilities produced by model.

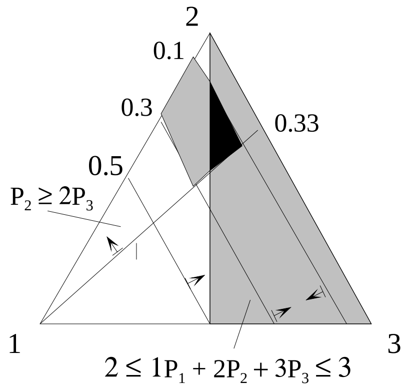

Figure 8. New model produced by extra judgement.

$\overline{P}(s_2) = 0.8$, $\underline{P}(s_3) = 0.1$, $\overline{P}(s_3) = 0.3$. Furthermore, the upper probability $\overline{P}(s_1)$ also changes and becomes more precise in the light of the new evidence, $\overline{P}(s_1) = 0.25$.

7. Concluding Remarks

This paper is an attempt to apply the theory of coherent interval-valued probabilities for computing probability characteristics of various systems and to generalize classical Markov models. Theoretical results and the results of numerical examples demonstrate that they are non-trivial and intuitively explainable. Their imprecision reflects a lack of available information. The results obtained have a strong mathematical sense and can be used in practice along with precise probabilities if they are available. The proposed generalization is not simply interval computation but has a deeper sense based on the sets of probabilities and their natural extensions.

It should be noted that further study is needed to develop efficient methods to analyze Markov models and to solve various problems including investigation of the balance equations, continuous Markov processes, chains with fuzzy states, and the generalization of known Markov models in reliability and other applied areas. Recent research by the authors and others in reliability analysis based on the interval-valued probabilities is encouraging that the interval-valued statistical reasoning is a promising approach for reliability engineering.

References

1. Kemeny, J. G. and Snell, J. L.: *Finite Markov Chains*, Van Nostrand Reinhold Company, New York, 1960.
2. Kuznetsov, V. P.: *Interval Statistical Models*, Radio and Communication, Moscow, 1991 (in Russian).
3. Walley, P.: Measures of Uncertainty in Expert Systems, *Artificial Intelligence* **83** (1996), pp.1–58.
4. Walley, P.: *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, 1991.