On an Interval-Valued Solution of the Marginal Problem

J. Vejnarová
Inst. of Information Theory and Automation
Acad. Sci. of the Czech Republic
vejnar@vse.cz

R. Jiroušek
Inst. of Information Theory and Automation
Acad. Sci. of the Czech Republic
radim@utia.cas.cz

Vl. Bína
Faculty of Management
Jindřichův Hradec
University of Economics, Prague
bina@fm.vse.cz

Abstract

A version of a marginal problem considered in this paper was connected, in the 80ies of the last century, with the necessity to cope with a task of knowledge integration in probabilistic expert systems. In practical situations marginal distributions representing pieces of local knowledge were often inconsistent. The present paper introduces an interval-valued solution of a marginal problem, which always exists, even in cases when, because of inconsistencies, one cannot get a solution in a classical sense.

The paper shows how famous Iterative Proportional Fitting Procedure, which is often used to get a classical solution of a consistent problem, can be exploited also when constructing an interval-valued solution, and how any solution can be improved (tightened).

Keywords. Multidimensional distributions, marginal problem, consistency, IPFP, imprecise probabilities, coherence.

1 Introduction

This contribution goes back to 80ies of the last century, when several papers promoting probabilistic methods for knowledge representation in expert systems were published [11, 2, 13, 14]. These papers proposed to represent pieces of local knowledge as low-dimensional (oligodimensional) probability distributions, and the global knowledge of the area of interest by a multidimensional distribution [7] expressed in a form of a Bayesian network [6], or other type of (usually graphical) Markov model [10]. The problems, which have never been satisfactorily solved, arise when a system of local knowledge, i.e. a system of oligodimensional distributions, appear to be inconsistent. This happens quite often when pieces of local knowledge are received from different sources (it can happen even in the situation, when oligodimensional distributions are estimated from one data file with missing values). Then there does not exist a distribution having all the oligodimensional distributions for its marginals and therefore the process of *knowledge integration* could not been successfully accomplished. This situation, when the corresponding marginal problem did not have a solution, was thoroughly studied in the thesis of Jiří Vomlel [15], where algorithms reaching *approximate solutions* were designed.

Taking advantage of the results achieved in the cited thesis [15], in the present paper we shall study a classical marginal problem with respect to a solution in a class of *interval-valued* probability distributions. It means we shall study a possibility to represent a system of (precise) oligodimensional distributions by a multidimensional distribution whose values are imprecise, whose values are intervals from $[0, 1]$. It should be stressed, however, that a solution of an *inconsistent* marginal problem requires quite a different approach than that employed for description of a set of all solutions of a consistent problem.

2 Notation

In this text, we will deal with a finite system of finite-valued random variables. Let N be a finite index set, $N \neq \emptyset$. Each variable from $\{X_i\}_{i \in N}$ is assumed to have a finite (non-empty) set of values $\mathbf{X}_i$. As said above, in the paper we shall deal with both precise and interval-valued probability distributions. All these distributions will be denoted by Greek letters. Nevertheless, to make the presentation unambiguous we will call the precise distributions *densities* and denote them with "normal" symbols π, κ, μ. Thus $\pi(x)$, for a specific $x \in \mathbf{X}$, will always be a number from the interval $[0, 1]$. In what follows, the word *distribution* will be reserved for an *interval-valued distribution*, which will always be denoted by bold characters $\boldsymbol{\pi}, \boldsymbol{\kappa}, \boldsymbol{\mu}$. Therefore, a value of $\boldsymbol{\pi}(x)$ will always be an interval from $[0, 1]$. In the paper we will use a simpli-

fied analogy of the notation used by Peter Walley [16]: $\boldsymbol{\pi}(x) = [\underline{\boldsymbol{\pi}}(x), \overline{\boldsymbol{\pi}}(x)]$. In the paper we always assume that $0 \leq \underline{\boldsymbol{\pi}}(x) \leq \overline{\boldsymbol{\pi}}(x) \leq 1$, and

$$\sum_{x \in \mathbf{X}} \underline{\boldsymbol{\pi}}(x) \leq 1 \leq \sum_{x \in \mathbf{X}} \overline{\boldsymbol{\pi}}(x).$$

Most of the densities and distributions considered in the paper will be defined only for a subset of the whole set of variables $\{X_i\}_{i \in N}$. For $K \subseteq N$, we can consider both a density $\pi(\{X_i\}_{i \in K})$ and a distribution $\boldsymbol{\pi}(\{X_i\}_{i \in K})$. To make the formulae more lucid, the following simplified notation will be used. Symbol $\pi(x_K)$ ($\boldsymbol{\pi}(x_K)$) will denote both a $|K|$-dimensional density (distribution) and a value of a probability density π (distribution $\boldsymbol{\pi}$), which is defined for variables $(X_i)_{i \in K}$ at a combination of values x_K; x_K thus represents a $|K|$-dimensional vector of values of variables $\{X_i\}_{i \in K}$. Analogously, we shall also denote the set of all these vectors $\mathbf{X}_K$:

$$\mathbf{X}_K = \times_{i \in K} \mathbf{X}_i.$$

Naturally, when several densities (distributions) are considered, we shall distinguish between them by indices.

Since a distribution $\boldsymbol{\pi}(x_K)$ determines a subset of $[0,1]^{\mathbf{X}_K}$, we can introduce a natural partial ordering of distributions defined for the same set of variables $\{X_i\}_{i \in K}$:

$$\boldsymbol{\pi} \subseteq \boldsymbol{\kappa} \overset{def}{\iff} \forall x_K \in \mathbf{X}_K \ (\boldsymbol{\pi}(x_K) \subseteq \boldsymbol{\kappa}(x_K)).$$

For a distribution $\pi(x_K)$ symbol[1] $\mathcal{M}(\boldsymbol{\pi})$ will denote the set of all the densities complying with the limits given by $\boldsymbol{\pi}$:

$$\mathcal{M}(\boldsymbol{\pi}) = \{\pi(x_K) : \underline{\boldsymbol{\pi}}(x_K) \leq \pi(x_K) \leq \overline{\boldsymbol{\pi}}(x_K) \text{ for all } x_K \in \mathbf{X}_K\}.$$

This allows us to define an equivalence relation on the set of all distributions. We say that two distributions $\boldsymbol{\pi}$ and $\boldsymbol{\kappa}$ defined for the same set of variables X_K are *structure equivalent* (in symbol $\boldsymbol{\pi} \simeq \boldsymbol{\kappa}$) if $\mathcal{M}(\boldsymbol{\pi}) = \mathcal{M}(\boldsymbol{\kappa})$. Notice that it does not necessarily mean that $\boldsymbol{\pi} = \boldsymbol{\kappa}$. For example, as the reader can immediately see, all three 1-dimensional distributions in Table 1 are different, pairwise structure equivalent, and

$$\boldsymbol{\pi}_1 \supset \boldsymbol{\pi}_2 \subset \boldsymbol{\pi}_3.$$

[1] $\mathcal{M}(\boldsymbol{\pi})$ is called by Weichselberger *a structure* of $\boldsymbol{\pi}$ ([17]), Cozman calls it *a credal set* ([]).

	$\boldsymbol{\pi}_1$	$\boldsymbol{\pi}_2$	$\boldsymbol{\pi}_3$
$x = 0$	$[0.2, 0.7]$	$[0.4, 0.7]$	$[0.4, 1.0]$
$x = 1$	$[0.2, 0.6]$	$[0.3, 0.6]$	$[0.3, 1.0]$

Table 1: Structure equivalent probability distributions $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3$

3 Coherent distributions

In analogy with the Walley's terminology [16] we will call a distribution $\boldsymbol{\pi}(x_K)$ *coherent*[2] if $\underline{\boldsymbol{\pi}}$ is a lower envelope of $\mathcal{M}(\boldsymbol{\pi})$ and $\overline{\boldsymbol{\pi}}$ is an upper envelope of $\mathcal{M}(\boldsymbol{\pi})$, i.e.

$$\forall x_K \in \mathbf{X}_K \ \exists \pi, \kappa \in \mathcal{M}(\boldsymbol{\pi}) \\ (\pi(x_K) = \underline{\boldsymbol{\pi}}(x_K) \ \& \ \kappa(x_K) = \overline{\boldsymbol{\pi}}(x_K)).$$

Such a distribution is "minimal" in the sense of the following assertion.

Lemma 1 *A distribution $\boldsymbol{\pi}(x_K)$ is coherent* iff

$$\boldsymbol{\kappa}(x_K) \simeq \boldsymbol{\pi}(x_K) \ \& \ \boldsymbol{\kappa} \subseteq \boldsymbol{\pi} \implies \boldsymbol{\kappa} = \boldsymbol{\pi}. \tag{1}$$

Proof. Assume first that the expression (1) does not hold for $\boldsymbol{\pi}$. It means that there exists a distribution $\boldsymbol{\kappa}(x_K)$ such that for some (at least one) $x_K \in \mathbf{X}_K$ either

$$\underline{\boldsymbol{\pi}}(x_K) < \underline{\boldsymbol{\kappa}}(x_K) \leq \overline{\boldsymbol{\kappa}}(x_K) \leq \overline{\boldsymbol{\pi}}(x_K), \tag{2}$$

or

$$\underline{\boldsymbol{\pi}}(x_K) \leq \underline{\boldsymbol{\kappa}}(x_K) \leq \overline{\boldsymbol{\kappa}}(x_K) < \overline{\boldsymbol{\pi}}(x_K), \tag{3}$$

and $\boldsymbol{\kappa}(x_K) \simeq \boldsymbol{\pi}(x_K)$. Without loss of generality assume the system of inequalities (2) holds true. We can immediately see that there cannot be a density $\pi(x_K) \in \mathcal{M}(\boldsymbol{\pi})$, for which $\pi(x_K) = \underline{\boldsymbol{\pi}}(x_K)$. If such a density existed then it could not be from $\mathcal{M}(\boldsymbol{\kappa}(x_K))$ and therefore $\boldsymbol{\kappa}(x_K) \not\simeq \boldsymbol{\pi}(x_K)$. So, we have proven that $\boldsymbol{\pi}$ is not coherent.

Now, assume that $\boldsymbol{\pi}$ is not coherent, i.e., there is $x_K \in \mathbf{X}_K$ such that none of the densities from $\mathcal{M}(\boldsymbol{\pi})$ achieves for x_K one of the limit values of $\boldsymbol{\pi}$; say $\underline{\boldsymbol{\pi}}(x_K)$. It means that for all $\pi \in \mathcal{M}(\boldsymbol{\pi})$

$$\pi(x_K) > \underline{\boldsymbol{\pi}}(x_K).$$

Based on this assumption we shall show that even

$$\inf_{\pi \in \mathcal{M}(\boldsymbol{\pi})} (\pi(x_K)) > \underline{\boldsymbol{\pi}}(x_K). \tag{4}$$

[2] Let us note that in the cited Walley's book one finds a different (but equivalent) definition of coherence.

If the equality in (4) held true, it would be possible to construct an infinite sequence of densities $\pi_1, \pi_2, \pi_3, \ldots$ from $\mathcal{M}(\boldsymbol{\pi})$ such that

$$\lim_{j \to +\infty} \pi(x_K) = \underline{\boldsymbol{\pi}}(x_K).$$

Considering another vector $y_K \in \mathbf{X}_K$ ($y_K \neq x_K$) one could always find an infinite subsequence of $\pi_1, \pi_2, \pi_3, \ldots$ such that this subsequence would converge also for y_K (it follows from the fact that all values $\pi_j(y_K)$ are from the interval $[0, 1]$). Since we assume that $\mathbf{X}_K$ is finite, one could repeat this selection of a convergent subsequence for all vectors from $\mathbf{X}_K$ getting a convergent infinite sequence of densities from $\mathcal{M}(\boldsymbol{\pi})$. Its limit would also be from $\mathcal{M}(\boldsymbol{\pi})$ and its value at x_K would equal $\underline{\boldsymbol{\pi}}(x_K)$, which contradicts our assumption. So we have got that inequality (4) holds true. Therefore, if defining a distribution $\boldsymbol{\kappa}$ such that it differs from $\boldsymbol{\pi}$ only in one limit value:

$$\underline{\boldsymbol{\kappa}}(x_K) = \inf_{\pi \in \mathcal{M}(\boldsymbol{\pi})} (x_K)$$

we see that $\boldsymbol{\kappa} \neq \boldsymbol{\pi}$, $\boldsymbol{\kappa} \subset \boldsymbol{\pi}$ and $\boldsymbol{\kappa} \simeq \boldsymbol{\pi}$, which contradicts to (1). ∎

The reader certainly noticed that when proving inequality (4) in the previous proof we took advantage of the fact that we consider a finite system of finite valued variables. The same property is basic also for the following trivial characterization lemma, which is presented without a proof.

Lemma 2 *A distribution $\boldsymbol{\pi}(x_K)$ is coherent* iff *there exists a finite system of densities $\{\pi_i\}_{i=1}^m \subset \mathcal{M}(\boldsymbol{\pi})$, such that*

$$\begin{aligned} \underline{\boldsymbol{\pi}}(x_K) &= \min_{i=1,\ldots,m} (\pi_i(x_K)), \\ \overline{\boldsymbol{\pi}}(x_K) &= \max_{i=1,\ldots,m} (\pi_i(x_K)), \end{aligned} \tag{5}$$

4 Marginalization of interval-valued distributions

Consider a probability density $\pi(x_N) \in \mathcal{M}(\boldsymbol{\pi}(x_N))$. From $\pi(x_N)$ one can easily (at least theoretically - at this moment we are not interested in computational complexity of the respective procedures) compute all its marginal densities. For $K \subset N$ it is described by the formula

$$\begin{aligned} \pi(x_K) &= \sum_{x_{N \setminus K} \in \mathbf{X}_{N \setminus K}} \pi(x_N) \\ &= \sum_{x_{N \setminus K} \in \mathbf{X}_{N \setminus K}} \pi(x_{N \setminus K}, x_K), \end{aligned} \tag{6}$$

which must be computed for all $x_K \in \mathbf{X}_K$. In this expression we have introduced a notation used throughout this section: a vector x_N is composed of two subvectors $x_{N \setminus K}$ and x_K, where x_K is a *projection* of x_N into $\mathbf{X}_K$, and, analogously $x_{N \setminus K}$ is a projection of x_N into $\mathbf{X}_{N \setminus K}$. For computation of marginal densities we need not exclude situations when $K = \emptyset$. In accordance with the above-introduced formula we get $\pi(x_\emptyset) = 1$.

In general, marginalization of (interval-valued) distributions can be introduced in several different ways. In this paper we adopt that is the simplest from the computational point of view. It simply repeats the formula (6) for both lower and upper limits of intervals separately:

$$\begin{aligned} \underline{\boldsymbol{\pi}}(x_K) &= \sum_{x_{N \setminus K} \in \mathbf{X}_{N \setminus K}} \underline{\boldsymbol{\pi}}(x_{N \setminus K}, x_K), \\ \overline{\boldsymbol{\pi}}(x_K) &= \min \left(1, \sum_{x_{N \setminus K} \in \mathbf{X}_{N \setminus K}} \overline{\boldsymbol{\pi}}(x_{N \setminus K}, x_K) \right). \end{aligned}$$

Notice that we could not afford this simple type of marginalization if we did not admit also incoherent distributions. This is because for this type of marginalization it easily happens that $\boldsymbol{\pi}(x_N)$ is coherent and its marginal distribution $\boldsymbol{\pi}(x_K)$ is incoherent; for an example see Table 2, where $\boldsymbol{\pi}(x_1, x_2)$ is coherent and $\boldsymbol{\pi}(x_1) \simeq \boldsymbol{\kappa}(x_1)$ and therefore, due to Lemma 1, $\boldsymbol{\pi}(x_1)$ cannot be coherent. (To show that for a marginal coherent distribution $\boldsymbol{\pi}(x_K)$ all its extensions $\boldsymbol{\pi}(x_N)$ must also be coherent is left to the reader.)

	$\boldsymbol{\pi}(x_1, x_2)$		$\boldsymbol{\pi}(x_1)$	$\boldsymbol{\kappa}(x_1)$
	x_2			
x_1	$0.3 - 0.4$	$0.1 - 0.2$	$0.4 - 0.6$	$0.4 - 0.6$
	$0.2 - 0.3$	$0.1 - 0.4$	$0.3 - 0.7$	$0.4 - 0.6$

Table 2: Coherent distribution $\boldsymbol{\pi}(x_1, x_2)$, whose marginal $\boldsymbol{\pi}(x_1)$ is not coherent

With respect to the above mentioned fact that coherence of a distribution does not extends to its marginals, the following assertion is of great importance.

Lemma 3 *For any two distributions $\boldsymbol{\pi}(x_N)$ and $\boldsymbol{\kappa}(x_N)$*

$$\boldsymbol{\pi}(x_N) \simeq \boldsymbol{\kappa}(x_N) \Longrightarrow \boldsymbol{\pi}(x_K) \simeq \boldsymbol{\kappa}(x_K)$$

holds true for any $K \subseteq N$.

Proof. Consider a density $\pi(x_K) \in \mathcal{M}(\boldsymbol{\pi}(x_K))$, and let $\boldsymbol{\pi}(x_K)$ be a marginal of $\boldsymbol{\pi}(x_N)$. Then

one can always find a density $\kappa(x_N) \in \mathcal{M}(\boldsymbol{\pi}(x_N))$ such that[3] $\kappa(x_K) = \pi(x_K)$. Since we assume that $\boldsymbol{\pi}(x_N) \simeq \boldsymbol{\kappa}(x_N)$, $\mathcal{M}(\boldsymbol{\pi}(x_N)) = \mathcal{M}(\boldsymbol{\kappa}(x_N))$ and therefore $\kappa(x_N) \in \mathcal{M}(\boldsymbol{\kappa}(x_N))$. However, it is obvious that when marginalizing both a distribution $\boldsymbol{\kappa}(x_N)$ and a density $\kappa(x_N) \in \mathcal{M}(\boldsymbol{\kappa}(x_N))$, we get density $\kappa(x_K)$ from $\mathcal{M}(\boldsymbol{\kappa}(x_K))$. Since $\kappa(x_K) = \pi(x_K)$, we got $\mathcal{M}(\boldsymbol{\pi}(x_K)) \subseteq \mathcal{M}(\boldsymbol{\kappa}(x_K))$. The proof is finished by realizing that roles of $\boldsymbol{\pi}$ and $\boldsymbol{\kappa}$ are exchangeable. ∎

5 Marginal problem

Consider a cover $K_1, K_2, \ldots, K_m$ of N (i.e. $K_1 \cup \ldots \cup K_m = N$), and a finite system

$$\Xi = \{\pi_1(x_{K_1}), \pi_2(x_{K_2}), \ldots, \pi_m(x_{K_m})\}$$

of probability densities π_i. It is well known (see e.g. Section 5.2 in [5]) that exactly one of the following three possibilities occurs:

(i) system Ξ of densities $\pi_1, \ldots, \pi_m$ is inconsistent, i.e. there does not exist a density $\kappa(x_N)$ such that all π_i $(i = 1, \ldots, m)$ are its marginals

$$\pi_i(x_{K_i}) = \kappa(x_{K_i});$$

(ii) there exists a unique solution of the marginal problem, i.e. there exists one and only one density $\kappa(x_N)$, for which $\pi_i(x_{K_i}) = \kappa(x_{K_i})$ holds true for all $i = 1, \ldots, m$;

(iii) there exist an infinitely many densities, which are solutions of the given marginal problem and the solutions form a convex set in the set of all densities of variables X_N.

Denoting the set of all extensions of π_i by $\Pi(\pi_i)$:

$$\Pi(\pi_i) = \{\kappa(x_N) : \kappa(x_{K_i}) = \pi_i(x_{K_i})\},$$

and analogously

$$\begin{aligned}\Pi(\Xi) &= \{\kappa(x_N) : \kappa(x_{K_i}) = \pi_i(x_{K_i}) \ \ \forall i = 1, \ldots, m\} \\ &= \bigcap_{i=1}^{m} \Pi(\pi_i),\end{aligned}$$

the above presented points (i), (ii) and (iii) say that $\Pi(\Xi) = \emptyset$, $|\Pi(\Xi)| = 1$ and $\Pi(\Xi)$ is a convex set, respectively.

6 Iterative proportional fitting procedure

As in the previous paragraph, consider a system

$$\Xi = \{\pi_1(x_{K_1}), \pi_2(x_{K_2}), \ldots, \pi_m(x_{K_m})\}$$

of probability densities and let us briefly introduce a simple version of an iterative procedure usually called *Iterative Proportional Fitting Procedure* (IPFP), which is connected with the names of Deming and Stephan [4]. Starting with a uniform probability density $\kappa_0(x_N)$, this procedure computes an infinite sequence of densities $\kappa_1(x_N), \kappa_2(x_N), \ldots$ according to the following process

$$\begin{aligned}
&\kappa_1(x_N) = \pi_1(x_{K_1}) \tfrac{\kappa_0(x_N)}{\kappa_0(x_{K_1})}, \\
&\kappa_2(x_N) = \pi_2(x_{K_2}) \tfrac{\kappa_1(x_N)}{\kappa_1(x_{K_2})}, \\
&\quad\vdots \\
&\kappa_m(x_N) = \pi_m(x_{K_m}) \tfrac{\kappa_{m-1}(x_N)}{\kappa_{m-1}(x_{K_m})}, \\
&\kappa_{m+1}(x_N) = \pi_1(x_{K_1}) \tfrac{\kappa_m(x_N)}{\kappa_m(x_{K_1})}, \\
&\quad\vdots \\
&\kappa_j(x_N) = \pi_i(x_{K_i}) \tfrac{\kappa_{j-1}(x_N)}{\kappa_{j-1}(x_{K_i})}, \\
&\quad\vdots
\end{aligned} \tag{7}$$

where in the last equation $i = ((j-1) \bmod m) + 1$.

From the point of view of this paper, the most important properties of this process are the following ones, proven by Csiszár in [3].

- If $\Pi(\Xi) \neq \emptyset$ then there exists

  $$\kappa^*(x_N) = \lim_{j \to +\infty} \kappa_j(x_N),$$

  $\kappa^*(x_N) \in \Pi(\Xi)$, and

  $$H(\kappa^*(x_N)) = \max_{\mu \in \Pi(\Xi)} (H(\mu(x_N))),$$

  where $H(\mu(x_N))$ denotes the Shannon entropy of density μ:

  $$H(\mu(x_N)) = -\sum_{x_N \in \mathbf{X}_N} \mu(x_N) \log(\mu(x_N)).$$

[3] Let us present a hint of the proof for the reader, who does not consider this to be obvious. Consider an arbitrary vector $x_K \in \mathbf{X}_K$. For this vector

$$\begin{aligned}\underline{\boldsymbol{\pi}}(x_K) &= \sum_{x_{N\setminus K} \in \mathbf{X}_{N\setminus K}} \underline{\boldsymbol{\pi}}(x_{N\setminus K}, x_K) \leq \pi(x_K) \\ &\leq \sum_{x_{N\setminus K} \in \mathbf{X}_{N\setminus K}} \overline{\boldsymbol{\pi}}(x_{N\setminus K}, x_K) = \overline{\boldsymbol{\pi}}(x_K).\end{aligned}$$

Therefore, taking, for example,

$$\begin{aligned}&\kappa(x_{N\setminus K}, x_K) \\ &\quad = \underline{\boldsymbol{\pi}}(x_{N\setminus K}, x_K) + s(x_K) \cdot \left(\overline{\boldsymbol{\pi}}(x_{N\setminus K}, x_K) - \underline{\boldsymbol{\pi}}(x_{N\setminus K}, x_K)\right),\end{aligned}$$

where $s(x_K) = (\pi(x_K) - \underline{\boldsymbol{\pi}}(x_K)) / (\overline{\boldsymbol{\pi}}(x_K) - \underline{\boldsymbol{\pi}}(x_K))$, one gets

$$\sum_{x_{N\setminus K} \in \mathbf{X}_{N\setminus K}} \kappa(x_{N\setminus K}, x_K) = \pi(x_K)$$

and $\underline{\boldsymbol{\pi}}(x_{N\setminus K}, x_K) \leq \kappa(x_{N\setminus K}, x_K) \leq \overline{\boldsymbol{\pi}}(x_{N\setminus K}, x_K)$.

- If $\Pi(\Xi) = \emptyset$ then $\lim_{j\to+\infty} \kappa_j(x_N)$ does not exist.

In spite of the fact that, according to our knowledge, it has not been proven, yet, based on the results presented in [15] and all the experiments performed by Jiří Vomlel and the authors, we conjecture that even in case that $\Pi(\Xi) = \emptyset$, the subsequences

$$\kappa_i(x_N), \kappa_{m+i}(x_N), \kappa_{2m+i}(x_N), \kappa_{3m+i}(x_N), \dots$$

converge for all $i = 1, 2, \dots, m$. Since all the densities $\kappa_i, \kappa_{m+i}, \kappa_{2m+i}, \dots$ are from $\Pi(\pi_i)$, it is obvious that, under the above mentioned conjecture, the distribution

$$\kappa_i^*(x_N) = \lim_{j\to+\infty} \kappa_{jm+i}(x_N)$$

is from $\Pi(\pi_i)$, too.

7 Interval-valued solution of a marginal problem

Considering a marginal problem given by a system of densities

$$\Xi = \{\pi_1(x_{K_1}), \pi_2(x_{K_2}), \dots, \pi_m(x_{K_m})\},$$

it is natural to ask whether it is possible to find an (in a sense optimal) interval-valued distribution $\boldsymbol{\kappa}(x_N)$, for which

$$\Pi(\Xi) \subseteq \mathcal{M}(\boldsymbol{\kappa}(x_N)).$$

This is a classical formulation of a marginal problem, which was studied by many authors and, in its general form, has never been satisfactorily solved. Famous is a solution of this problem in case that all the distributions π_i are one-dimensional; it is known as Fréchet bounds (see e.g. [12]).

In this paper we are discussing the marginal problem without any restrictions; it may be consistent or inconsistent, but primarily we have in mind an inconsistent situation, i.e. the situation when $\Pi(\Xi) = \emptyset$. For this case the above mentioned problem has no sense. So, our goal will be to specify a (smallest) possible area of densities such that all marginal conditions are met by at least one of the densities from the selected area. This idea can be formalized in the following two ways.

Definition 1 *A probability distribution $\boldsymbol{\pi}(x_N)$ is a solution of a marginal problem given by system Ξ if for each density $\pi_i(x_{K_i})$ from Ξ there exists a density $\kappa(x_N) \in \mathcal{M}(\boldsymbol{\pi}(x_N))$ such that $\kappa(x_{K_i}) = \pi_i(x_{K_i})$. Such a solution $\boldsymbol{\pi}$ is tight if it is minimal in the sense of ordering $\subseteq$; for any solution $\boldsymbol{\kappa}$ of the same marginal problem*

$$\boldsymbol{\kappa} \subseteq \boldsymbol{\pi} \implies \boldsymbol{\kappa} = \boldsymbol{\pi}.$$

Definition 2 *A probability distribution $\boldsymbol{\pi}(x_N)$ is a solution of a marginal problem given by system Ξ if each density $\pi_i(x_{K_i}) \in \Xi$ is from $\mathcal{M}(\boldsymbol{\pi}(x_{K_i}))$. Such a solution $\boldsymbol{\pi}$ is tight if it is minimal in the sense of ordering $\subseteq$; for any solution $\boldsymbol{\kappa}$ of the same marginal problem*

$$\boldsymbol{\kappa} \subseteq \boldsymbol{\pi} \implies \boldsymbol{\kappa} = \boldsymbol{\pi}.$$

Theorem 1 *Definitions 1 and 2 are equivalent.*

Proof. First, let us show that Definition 1 $\implies$ Definition 2. Let $\boldsymbol{\pi}(x_N)$ be a solution in the sense of Definition 1 and consider $\pi_i \in \Xi$. Therefore, there exists $\kappa(x_N) \in \mathcal{M}(\boldsymbol{\pi}(x_N))$ whose marginal equals π_i: $\kappa(x_{K_i}) = \pi_i(x_{K_i})$. However, since $\kappa \in \mathcal{M}(\boldsymbol{\pi})$ we know that its arbitrary marginal density $\kappa(x_K)$ is from the respective set $\mathcal{M}(\boldsymbol{\pi}(x_K))$ (see Section 4), and therefore

$$\pi_i(x_{K_i}) = \kappa(x_{K_i}) \in \mathcal{M}(\boldsymbol{\pi}(x_{K_i}))$$

for all $i = 1, 2, \dots, m$.

Now, let us show the reverse implication: Definition 2 $\implies$ Definition 1. Let now $\boldsymbol{\pi}(x_N)$ be a solution in the sense of Definition 2. Thus we know that $\pi_i(x_{K_i}) \in \mathcal{M}(\boldsymbol{\pi}(x_{K_i}))$, and therefore, following a hint in Footnote 3, we can construct $\kappa(x_N) \in \mathcal{M}(\boldsymbol{\pi}(x_N)$, which is an extension of $\pi_i(x_{K_i})$. ∎

Let us, now, answer a question, how to find an interval-valued solution of a marginal problem defined by a system of densities

$$\Xi = \{\pi_1(x_{K_1}), \pi_2(x_{K_2}), \dots, \pi_m(x_{K_m})\}.$$

If Ξ is consistent (i.e., if $\Pi(\Xi) \neq \emptyset$) then one can immediately see that any density $\pi \in \Pi(\Xi)$ can be used to define a solution $\boldsymbol{\pi}(x_N)$:

$$\underline{\boldsymbol{\pi}}(x_N) = \overline{\boldsymbol{\pi}}(x_N) = \pi(x_N) \quad \forall x_N \in \mathbf{X}_N. \tag{8}$$

Moreover, this solution is tight.

In an inconsistent case we propose to proceed in the following way. First, apply IPFP to the densities from Ξ getting, according to formulae (7), a sequence of densities $\kappa_0(x_N), \kappa_1(x_N), \kappa_2(x_N), \dots$. From what was said in Section 6 we know that this sequence certainly does not converge. However, in agreement with our conjecture presented in that section, we can get m densities

$$\begin{aligned}
\kappa_1^* &= \lim_{j\to+\infty} \kappa_{jm+1}, \\
\kappa_2^* &= \lim_{j\to+\infty} \kappa_{jm+2}, \\
&\vdots \\
\kappa_m^* &= \lim_{j\to+\infty} \kappa_{jm}.
\end{aligned}$$

Since $\kappa_i^* \in \Pi(\pi_i)$ for all $i = 1, \ldots, m$, the distribution $\boldsymbol{\pi}(x_N)$ defined at each vector $x_N \in \mathbf{X}_N$

$$\begin{aligned} \underline{\boldsymbol{\pi}}(x_N) &= \min_{i=1,\ldots,m} (\kappa_i^*(x_N)), \\ \overline{\boldsymbol{\pi}}(x_N) &= \max_{i=1,\ldots,m} (\kappa_i^*(x_N)), \end{aligned} \tag{9}$$

must be a solution of the given inconsistent marginal problem, because all κ_i^* are from $\mathcal{M}(\boldsymbol{\pi})$. Moreover, it is obvious (due to Lemma 2) that $\boldsymbol{\pi}$ is coherent.

The reader certainly noticed that this procedure is applicable also for consistent Ξ, since, in this case,

$$\kappa_1^* = \kappa_2^* = \ldots = \kappa_m^* \in \Pi(\Xi)$$

and therefore formulae (9) give exactly the same solution as (8).

Example 1 *Let us consider three (inconsistent) 2-dimensional densities from Table 5. The computational process resulting from an application of IPFP to these three densities is recorded in Table 3. When observing the probabilities with the precision of 5 decimal digits, the computational process stabilizes after 19 steps when it starts cycling among 3 densities κ_{19}, κ_{20} and κ_{21}.*

$\pi_1(x_1, x_2)$	x_1	
x_2	0.28	0.13
	0.25	0.34

$\pi_2(x_1, x_3)$	x_1	
x_3	0.50	0.31
	0.03	0.16

$\pi_3(x_2, x_3)$	x_3	
x_2	0.32	0.09
	0.35	0.24

Table 5: Inconsistent 2-dimensional densities

As we said in the previous section

$$\begin{aligned} \lim_{j\to\infty} \kappa_{3j+1} &\doteq \kappa_{19} \in \Pi(\pi_1), \\ \lim_{j\to\infty} \kappa_{3j+2} &\doteq \kappa_{20} \in \Pi(\pi_2), \\ \lim_{j\to\infty} \kappa_{3j} &\doteq \kappa_{21} \in \Pi(\pi_3). \end{aligned}$$

which can easily be verified from Tables 5 and 3.

Since the densities π_1, π_2 and π_3 are inconsistent, $\Pi(\pi_1) \cap \Pi(\pi_2) \cap \Pi(\pi_3) = \emptyset$, however, the interval-valued solution of this inconsistent marginal problem is a coherent distribution defined by formulae (9); see Table 6.

Thanks to the presented approach we are always able to find an interval-valued solution of a marginal problem in a form of a coherent distribution. As it will be seen from the next example, this solution is not unique. Namely, different orderings of oligodimensional distributions entering IPFP may lead to different "limit" distributions appearing in the cycle $\lim_{j\to\infty} \kappa_{mj+1}, \lim_{j\to\infty} \kappa_{mj+2}, \ldots, \lim_{j\to\infty} \kappa_{mj}$.

$\boldsymbol{\pi}$		$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$x_3 = 0$	$0.232 - 0.272$	$0.073 - 0.102$
$x_2 = 0$	$x_3 = 1$	$0.012 - 0.026$	$0.036 - 0.068$
$x_2 = 1$	$x_3 = 0$	$0.183 - 0.229$	$0.150 - 0.209$
$x_2 = 1$	$x_3 = 1$	$0.017 - 0.035$	$0.123 - 0.211$

Table 6: Interval-valued solution resulting from the computational process recorded in Table 3

Example 2 *Consider again the same three inconsistent 2-dimensional densities from Table 5 as in Example 1. Table 4 presents selected steps of the computational process resulting from application of IPFP to these densities in the ordering π_1, π_3, π_2. This time the process starts "cycling" (when considering again the precision given by 5 decimal digits) after 16 steps. Notice that because of the different ordering of the input densities*

$$\begin{aligned} \lim_{j\to\infty} \kappa_{3j+1} &\doteq \kappa_{16} \in \Pi(\pi_1), \\ \lim_{j\to\infty} \kappa_{3j+2} &\doteq \kappa_{17} \in \Pi(\pi_3), \\ \lim_{j\to\infty} \kappa_{3j} &\doteq \kappa_{18} \in \Pi(\pi_2). \end{aligned}$$

The resulting interval-valued solution is in Table 7.

$\boldsymbol{\pi}$		$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$x_3 = 0$	$0.236 - 0.283$	$0.083 - 0.104$
$x_2 = 0$	$x_3 = 1$	$0.012 - 0.024$	$0.035 - 0.067$
$x_2 = 1$	$x_3 = 0$	$0.182 - 0.233$	$0.167 - 0.214$
$x_2 = 1$	$x_3 = 1$	$0.016 - 0.030$	$0.121 - 0.211$

Table 7: Interval-valued solution resulting from the computational process recorded in Table 4

8 Tightening of a solution

The method proposed in the previous section has one great disadvantage: the constructed solution of an inconsistent marginal problem is usually not tight. In the examples presented in this section we shall see that neither of the solutions from Tables 6 and 7 is tight. The goal of this section is to show that this disadvantage can easily be corrected.

Theorem 2 *A distribution $\boldsymbol{\pi}(x_N)$ is a tight solution of a marginal problem given by a system of densities*

$$\Xi = \{\pi_1(x_{K_1}), \pi_2(x_{K_2}), \ldots, \pi_m(x_{K_m})\}$$

	000	001	010	011	100	101	110	111
κ_0	.125	.125	.125	.125	.125	.125	.125	.125
κ_1	.14	.14	.125	.125	.065	.065	.17	.17
κ_2	.26415	.01585	.23585	.01415	.08574	.04426	.22426	.11574
κ_3	.24158	.02373	.17941	.02615	.07842	.06627	.17059	.21385
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
κ_{19}	.25485	.02515	.21511	.03489	.07356	.05644	.15036	.18964
κ_{20}	.27114	.01257	.22886	.01743	.10184	.03670	.20816	.12330
κ_{21}	.23263	.02296	.18329	.02973	.08737	.06704	.16671	.21027
κ_{22}	.25485	.02515	.21511	.03489	.07356	.05644	.15036	.18964
κ_{23}	.27114	.01257	.22886	.01743	.10184	.03670	.20816	.12330
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 3: Iterative Proportional Fitting Procedure applied to π_1, π_2, π_3

	000	001	010	011	100	101	110	111
κ_0	.125	.125	.125	.125	.125	.125	.125	.125
κ_1	.14	.14	.125	.125	.065	.065	.17	.17
κ_2	.21854	.06146	.14830	.10170	.10146	.02854	.20170	.13831
κ_3	.29786	.01130	.20214	.01870	.10375	.02737	.20625	.13263
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
κ_{16}	.26726	.01274	.23235	.01765	.09491	.03509	.21393	.12607
κ_{17}	.23614	.02397	.18222	.02948	.08386	.06603	.16778	.21052
κ_{18}	.28222	.01345	.21778	.01655	.10331	.03820	.20669	.12180
κ_{19}	.26726	.01274	.23235	.01765	.09491	.03509	.21393	.12607
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 4: Iterative Proportional Fitting Procedure applied to π_1, π_3, π_2

iff **(i)** *it is a solution, and*

(ii) $\forall x_N \in \mathbf{X}_N \ \exists i, j \in \{1, 2, \ldots, m\}$

$$\begin{aligned} \underline{\boldsymbol{\pi}}(x_{K_i}) &= \pi_i(x_{K_i}) \\ \overline{\boldsymbol{\pi}}(x_{K_j}) &= \pi_j(x_{K_j}). \end{aligned} \tag{10}$$

Proof. First, realize that if $\boldsymbol{\pi}$ is a solution then, due to Definition 2, for all $\ell = 1, 2, \ldots, m$

$$\underline{\boldsymbol{\pi}}(x_{K_\ell}) \leq \pi_\ell(x_{K_\ell}) \leq \overline{\boldsymbol{\pi}}(x_{K_\ell}). \tag{11}$$

If, in addition to this, condition (ii) holds then no $\underline{\boldsymbol{\pi}}(x_N)$ may be increased without violating condition (11) because it would have also increased all the respective marginal low boundaries $\underline{\boldsymbol{\pi}}(x_{K_\ell})$. Analogously, also no $\overline{\boldsymbol{\pi}}(x_N)$ may be decreased without breaking condition (11) and therefore $\boldsymbol{\pi}$ is tight.

Let $\boldsymbol{\pi}$ be a tight solution and assume condition (ii) is not fulfilled. It means there is a point x_N such that either

$$\underline{\boldsymbol{\pi}}(x_{K_\ell}) < \pi_\ell(x_{K_\ell})$$

or

$$\overline{\boldsymbol{\pi}}(x_{K_\ell}) > \pi_\ell(x_{K_\ell})$$

holds true for all $\ell = 1, 2, \ldots, m$. Without loss of generality assume the former inequality is valid and define a distribution $\boldsymbol{\kappa}$ in the way that it equals $\boldsymbol{\pi}$ at all vectors of $\mathbf{X}_N$ but the considered vector x_N. For this vector define:

$$\begin{aligned} \underline{\boldsymbol{\kappa}}(x_N) &= \underline{\boldsymbol{\pi}}(x_N) + \min_{\ell \in \{1,\ldots,m\}} (\pi_\ell(x_{K_\ell}) - \underline{\boldsymbol{\pi}}(x_{K_\ell})) \\ \overline{\boldsymbol{\kappa}}(x_N) &= \overline{\boldsymbol{\pi}}(x_N). \end{aligned}$$

Distribution $\boldsymbol{\kappa}$ is defined in the way that it is a solution of the marginal problem given by system Ξ (i.e., $\underline{\boldsymbol{\kappa}}(x_{K_\ell}) \leq \pi_\ell(x_{K_\ell}) \leq \overline{\boldsymbol{\kappa}}(x_{K_\ell})$ for all $\ell = 1, \ldots, m$) and $\boldsymbol{\kappa} \subset \boldsymbol{\pi}$, $\boldsymbol{\kappa} \neq \boldsymbol{\pi}$, which contradicts the definition of tightness. ∎

Theorem 2 may serve as a basis for a simple tightening procedure. Its simplest version recomputes values of a solution $\boldsymbol{\pi}$ in a cycle over all $x_N \in \mathbf{X}_N$:

$$\begin{aligned} \underline{\boldsymbol{\pi}}(x_N) &:= \underline{\boldsymbol{\pi}}(x_N) + \min_{\ell \in \{1,\ldots,m\}} (\pi_\ell(x_{K_\ell}) - \underline{\boldsymbol{\pi}}(x_{K_\ell})), \\ \overline{\boldsymbol{\pi}}(x_N) &:= \overline{\boldsymbol{\pi}}(x_N) - \min_{\ell \in \{1,\ldots,m\}} (\overline{\boldsymbol{\pi}}(x_{K_\ell}) - \pi_\ell(x_{K_\ell})). \end{aligned}$$

Notice, that using the previous "program-style" formula, which contains the same symbol $\boldsymbol{\pi}(x_N)$ on both sides, we stress that at each step of the cycle we have to compute the marginal distributions $\boldsymbol{\pi}(x_{K_\ell})$ from the last updated version of $\boldsymbol{\pi}$. Table 8 presents the result of the simplest version of the tightening process when applied to distribution from Table 6.

$\boldsymbol{\pi}$		$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$x_3 = 0$	$0.247 - 0.271$	$0.073 - 0.101$
$x_2 = 0$	$x_3 = 1$	$0.013 - 0.022$	$0.037 - 0.068$
$x_2 = 1$	$x_3 = 0$	$0.200 - 0.229$	$0.150 - 0.209$
$x_2 = 1$	$x_3 = 1$	$0.017 - 0.029$	$0.123 - 0.211$

Table 8: Tight interval-valued solution received from the solution in Table 6

A little bit more sophisticated version of this process orders the points x_N in the cycle according to a decreasing value of $\overline{\boldsymbol{\pi}}(x_N) - \underline{\boldsymbol{\pi}}(x_N)$. This strategy was applied to solution from Table 7. Before this, however, all probabilities from the Table 7 were cut to 2 decimal points, since all the input densities were also given with this precision, so, the resulting tight distribution, which is contained in Table 9, have all the probabilities with two decimal digits.

$\boldsymbol{\pi}$		$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$x_3 = 0$	$0.25 - 0.27$	$0.07 - 0.11$
$x_2 = 0$	$x_3 = 1$	$0.01 - 0.03$	$0.03 - 0.06$
$x_2 = 1$	$x_3 = 0$	$0.18 - 0.23$	$0.17 - 0.20$
$x_2 = 1$	$x_3 = 1$	$0.02 - 0.04$	$0.13 - 0.20$

Table 9: Tight interval-valued solution received from the solution in Table 7

9 Conclusions

We introduced a new notion of an interval-valued solution of a marginal problem, which always exists, even in cases when input marginal distributions are inconsistent. We showed that even a tight solution is not unique. Perhaps, some tight solutions are better than the others. Therefore, it would be nice to have a criterion comparing tight solutions and enabling us to define optimal solution(s).

The tightening process presented in the last section could be applied to any solution, i.e., also to a trivial distribution, which is a solution to any marginal problem:

$$\underline{\boldsymbol{\pi}}(x_N) = 0, \quad \overline{\boldsymbol{\pi}}(x_N) = 1.$$

However, the resulting tight solution would rarely possess reasonable properties, so we propose to find a solution to be further tightened with the help of IPFP. In connection with this we want to recall that an effective implementation of this procedure was designed in [8, 9]. It is based on an idea that all the densities are represented in a form of decomposable models. This type of representation substantially increases the size of tractable problems. So it may happen (in special situations) that one finds a solution, which cannot be further tightened, since the computation in a cycle over all vectors x_N is not possible because of the size of $\mathbf{X}_N$.

Let us recollect that the process of finding a solution of an inconsistent marginal problem with the help of IPFP is based on a conjecture, whose proof remains a challenge for the future research.

Acknowledgements

The research was supported in part by GA AV ČR, under grant A2075302 (Jiroušek), by GA ČR, under grant 201/04/0393 (Vejnarová) and VŠE, under grant IGA ???? (Bína).

References

[1] T. Augustin, On the suboptimality of the generalized Bayes rule and robust Bayesian procedures from the decision theoretic point of view - a cautionary note on updating imprecise priors. In: *Proceedfings of the 3rd Int. Symposium on Imprecise Probabilities and their Applications.* (J.M. Bernard, T. Seidenfeld and M. Zaffalon, eds.) Carleton Scientific, Canada, 2003, pp. 31-45.

[2] P. Cheeseman, In defence of probability. In: *Proc. 8th Int. Conf. on AI* (IJCAI 85), Los Angeles, CA, USA, 1985, pp. 1002-1007.

[3] I. Csiszár, I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.*, 3:146-158, 1975.

[4] W.E. Deming and F.F. Stephan, On a least square adjustment of a sampled frequency table when the expected marginal totals are known, *Ann. Math. Stat.* 11:427-444, 1940.

[5] P. Hájek, T. Havránek, and R. Jiroušek, *Uncertain Information Processing in Expert Systems*. CRC Press, Inc., Boca Raton, 1992.

[6] F. V. Jensen, *Introduction to Bayesian Network*, UCL Press 1996.

[7] R. Jiroušek, A survey of methods used in probabilistic expert systems for knowledge integration, *Knowledge-Based Systems* 3:7-12, 1990.

[8] R. Jiroušek, Solution of the Marginal Problem and Decomposable Distributions. *Kybernetika* 27:403-412, 1991.

[9] R. Jiroušek and S. Přeučil, On the Effective Implementation of the Iterative Proportional Fitting Procedure. *Computational Statistics and Data Analysis*, 19:177-189, 1995.

[10] S.L. Lauritzen, *Graphical Models*. Clarendon Press, Oxford, 1996.

[11] A. Perez, A probabilistic approach to the integration of partial knowledge for medical decision making. In: *Proc. 1st Czechoslovak Cingress of Biomedical Engineering* (BMI'83), (1983), pp. 221–226.

[12] L. Rschendorf, Sharpness of Fréchet-Bounds. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 57:293-302, 1981.

[13] R.D. Shachter, Evaluating Influence Diagrams, *Operations Res.* 34:871-890, 1986.

[14] R.D. Shachter, Intelligent probabilistic inference. In: *Uncertainty in Artificial Intelligence* (L.N. Knal, J.F. Lemmer, eds.), North Holland, Amsterdam, 1986, pp. 371- ???.

[15] J. Vomlel, Methods of probabilistic knowledge integration. PhD. thesis. FEL ČVUT, Praha, 2000.

[16] P. Walley, *Statistical Reasoning with Inprecise Probabilities*. Chapman and Hall, London, etc., 1991.

[17] K. Weichselberger and T. Augustin, On the symbiosis of two concepts of conditional interval probability. In: *Proceedfings of the 3rd Int. Symposium on Imprecise Probabilities and their Applications.* (J.M. Bernard, T. Seidenfeld and M. Zaffalon, eds.) Carleton Scientific, Canada, 2003, pp. 606-629.