Radim Jiroušek

Solution of the marginal problem and decomposable distributions

Kybernetika, Vol. 27 (1991), No. 5, 403--412

Persistent URL: http://dml.cz/dmlcz/125857

KYBERNETIKA — VOLUME 27 (1991), NUMBER 5

SOLUTION OF THE MARGINAL PROBLEM AND DECOMPOSABLE DISTRIBUTIONS[1]

RADIM JIROUŠEK

A well-known Iterative Proportional Fitting Procedure proposed to construct a probability distribution with given marginals is of exponential space and time algorithmical complexity. Nevertheless, study of the case in question may make possible to reduce its complexity by decomposition of the problem as it is showed in [6]. Further reduction of the time complexity is probably impossible but, in some cases, the space-consumption can be reduced by taking advantage of possibility to represent decomposable distributions in an economical way.

1. INTRODUCTION

In probabilistic models of expert systems partial (input, or expert) knowledge is, usually beforehand, transformed into a form of probability distributions of small dimensions. Total knowledge of the area of interest is then represented by a multidimensional distribution whose marginal distributions coincide with the input distributions, or, in other words, by a distribution which is a solution of the given marginal problem. As a rule, there exist infinitely many of distributions meeting this condition and thus it is quite natural to seek for the distribution which is the best one in a sense. Construction of this optimal distribution is called the knowledge integration process.

Probably the most often used criterion for a comparison of multidimensional distributions coming into consideration is the maximum entropy principle. According to it, the distribution with the highest value of the Shannon entropy is preferred. Though there are serious objections against this criterion (cf. e.g. [4]), it will be considered also in this paper as the studied *Iterative Proportional Fitting Procedure* yields distributions optimal in this sense.

The method was published by Deming and Stephan as early as in 1940 [2]. It is

[1] This is an extended version of the lecture presented at the Symposium on Distributions with Given Marginals, Rome, April 4—7, 1990.

known that it is of extremely high algorithmical complexity and so it can be used for a limited number of dimensions only. Nevertheless, in some cases it is possible to reduce this complexity in the way proposed by Malvestuto [6]. The presented paper shows another possibility how to decrease the space complexity of the algorithm.

2. ITERATIVE PROPORTIONAL FITTING PROCEDURE

In accordance with most of papers dealing with probabilistic approaches to expert systems, a finite model will be considered.

Let $X_1, \ldots, X_N$ be random variables taking their values from finite sets $\mathbf{X}_1, \ldots, \mathbf{X}_N$ each of them containing at least two elements. Input (expert) knowledge is assumed to have been transformed into a system θ of marginal distributions $\theta = \{P_{S_1}, \ldots, P_{S_K}\}$, where each the subset $S_k \subset \{X_1, \ldots, X_N\}$ $(k = 1, \ldots, K)$ determines variables for which the distribution P_{S_k} is defined. If P is an N-dimensional probability distribution of variables $X_1, \ldots, X_N$ and $S \subset \{X_1, \ldots, X_N\}$ then P^S denotes the marginal distribution of P defined for variables from S only.

In other words, P is defined on

$$\mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_N = \mathop{\times}_{i=1}^{N} \mathbf{X}_i$$

whereas P^S is defined on

$$\mathop{\times}_{i: X_i \in S} \mathbf{X}_i \, .$$

For $S = \emptyset$ let $P^S = 1$ for any distribution P.

The goal of the knowledge integration process (under the assumption of acceptance of the maximum entropy principle) is to construct the N-dimensional distribution

$$P_\theta \in \Pi_\theta = \{P: P^{S_k} = P_{S_k} \text{ for all } k = 1, \ldots, K\}$$

for which

$$H(P_\theta) = - \sum_{(x_1, \ldots, x_N) \in \mathbf{X}_1 \times \ldots \times \mathbf{X}_n} P_\theta(x_1, \ldots, x_N) \log P_\theta(x_1, \ldots, x_N) = \max_{P \in \Pi_\theta} (H(P)) \, .$$

The distribution P_θ can be obtained by the Iterative Proportional Fitting Procedure which is based upon the recurrent computation of N-dimensional distributions $P_1, P_2, \ldots$ according to the following formula

$$P_i = P_{i-1} \frac{P_{S_j}}{P_{i-1}^{S_j}} \quad \text{for} \quad j = ((i - 1) \bmod K) + 1 \, .$$

Remark. Whenever a ratio of distributions P/Q appears in this paper, the distribution P is always absolutely continuous with respect to Q $(P \ll Q)$. Therefore expressions $a/0$ for $a \neq 0$ are out of question. The expression $0/0$ is always assumed to be 1.

Csiszár in [1] showed that if $\Pi_\theta \neq \emptyset$ then beginning with the uniform distribution

$$P_0 = \frac{1}{|\mathbf{X}_1| \dots |\mathbf{X}_N|}$$

this recurrent procedure converges and its limit distribution $P^* = P_\theta$. (More precisely, it follows immediately from his result that P^* is an I-projection of P_0 on Π_θ and from the fact that this specific choice of P_0 implies $I(P \parallel P_0) = H(P_0) - H(P)$.)

3. DECOMPOSABLE DISTRIBUTIONS

Let ζ be a system of $J(J \geqq 1)$ subsets $R_j \subset \{X_1, \dots, X_N\}$ such that they can be ordered in such a way that the sequence $(R_1, \dots, R_J)$ fulfill the following *running intersection property*:

$$\forall j = 2, \dots, J \quad \exists i, 1 \leqq i < j \left(R_j \cap \left(\bigcup_{l=1}^{j-1} R_l\right) \subset R_i\right).$$

Notice that any system consisting of one or two subsets meets the running intersection property.

Definition. A probability distribution P (defined for all variables $\{X_1, \dots, X_N\}$) is called *decomposable with respect to the system* $\zeta = \{R_1, \dots, R_J\}$ if

$$P = P^{R_1}\left[P^{R_2}/P^{R_2 \cap R_1}\right] \dots \left[P^{R_J}/P^{R_J \cap (R_1 \cup \dots \cup R_{J-1})}\right].$$

The decomposability is defined only for systems meeting (after reordering, if necessary) the running intersection property. Evidently, the above equality implies

$$\bigcup_{j=1}^{J} R_j = \{X_1, \dots, X_N\}.$$

Moreover, if there exists one ordering of subsets from ζ meeting the running intersection property such that this equality holds then it holds also for any other ordering with this property [5].

In addition to some theoretical reasons, decomposable distributions are of great practical importance. They are namely advantageous from the point of view of economical representation in a computer memory. To store a general N-dimensional distribution one needs

$$|\mathbf{X}_1| \cdot |\mathbf{X}_2| \dots |\mathbf{X}_N| \geqq 2^N$$

numbers whereas when representing a decomposable distribution, storage demands depend mainly on dimensionality of the sets $R_j \in \zeta$. Such a distribution can be easily recorded with the help of

$$\sum_{j=1}^{J} \prod_{i: X_i \in R_j} |\mathbf{X}_i|$$

numbers.

Some of theoretical characteristics of the decomposable distributions, which will be used in the sequel, are summarized in the following assertion.

Lemma 1. Let $\zeta = \{R_1, \ldots, R_J\}$ be a system of subsets of $\{X_1, \ldots, X_N\}$. Let $\{P_1, \ldots, P_J\}$ be a system of probability distributions such that for each $j = 1, \ldots, J$ P_j is a $|R_j|$-dimensional distribution defined for variables from R_j and for each couple of indices $i, j \in \{1, \ldots, J\}$

$$P_i^{R_i \cap R_j} = P_j^{R_i \cap R_j}.$$

Let Π_ζ denote the set of all N-dimensional distributions having $\{P_1, \ldots, P_J\}$ as marginals, i.e.

$$\Pi_\zeta = \{Q\colon Q^{R_j} = P_j \text{ for all } j = 1, \ldots, J\}.$$

If there exists a permutation $\sigma(1), \ldots, \sigma(J)$ of integers $1, \ldots, J$ such that $(R_{\sigma(1)}, \ldots, R_{\sigma(J)})$ meets the running intersection property then

(1) there are many of such permutations; namely, for any $j \in \{1, \ldots, J\}$ there exists a permutation $\sigma_j(1), \ldots, \sigma_j(J)$ such that $(R_{\sigma_j(1)}, \ldots, R_{\sigma_j(J)})$ meets the running intersection property and $\sigma_j(1) = j$,

(2) for all permutations σ, for which $(R_{\sigma(1)}, \ldots, R_{\sigma(J)})$ meet the running intersection property, the expressions

$$P_\sigma = P_{\sigma(1)}[P_{\sigma(2)}/P_{\sigma(2)}^{R_{\sigma(2)} \cap R_{\sigma(1)}}] \ldots [P_{\sigma(J)}/P_{\sigma(J)}^{R_{\sigma(J)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(J-1)})}]$$

define the same probability distribution from Π_ζ for which

$$H(P_\sigma) = \max_{Q \in \Pi_\zeta} (H(Q)).$$

Proof. Lemma 1 is a summarization of well-known facts.

The existence of the permutation σ_j follows from the fact that a system with the running intersection property can be represented by a triangulated graph, or equivalently, by a hypertree. The problem of finding a respective ordering can be solved with the help of *a restricted maximum cardinality search* on an acyclic hypergraph [8].

The statement that $P_\sigma \in \Pi_\zeta$ was proved in [5].

Suppose $H(P_\sigma) < H(Q)$ for some $Q \in \Pi_\zeta$. Then, using conditional entropy as introduced e.g. in [3], one can express the Shannon entropy in the form of a sum

$$H(P_\sigma) = H(P_\sigma^{R_{\sigma(1)}}) + H(P_\sigma^{R_{\sigma(2)}|R_{\sigma(1)}}) + \ldots + H(P_\sigma^{R_{\sigma(J)}|(R_{\sigma(1)} \cup \ldots \cup R_{\sigma(J-1)})})$$

and also

$$H(Q) = H(Q^{R_{\sigma(1)}}) + H(Q^{R_{\sigma(2)}|R_{\sigma(1)}}) + \ldots + H(Q^{R_{\sigma(J)}|(R_{\sigma(1)} \cup \ldots \cup R_{\sigma(J-1)})}).$$

Therefore, for at least one j, $1 \leqq j < J$

$$H(P_\sigma^{R_{\sigma(j)}|(R_{\sigma(1)} \cup \ldots \cup R_{\sigma(j-1)})}) < H(Q^{R_{\sigma(j)}|(R_{\sigma(1)} \cup \ldots \cup R_{\sigma(j-1)})})$$

which contradicts with

$$H\big(P_\sigma^{R_{\sigma(j)}|(R_{\sigma(1)}\cup\ldots\cup R_{\sigma(j-1)})}\big) = H\big(P_\sigma^{R_{\sigma(j)}|R_{\sigma(j)}\cap(R_{\sigma(1)}\cup\ldots\cup R_{\sigma(j-1)})}\big) =$$

$$= H\big(Q^{R_{\sigma(j)}|R_{\sigma(j)}\cap(R_{\sigma(1)}\cup\ldots\cup R_{\sigma(j-1)})}\big) \geqq H\big(Q^{R_{\sigma(j)}|(R_{\sigma(1)}\cup\ldots\cup R_{\sigma(j-1)})}\big),$$

where the former equality holds because of the definition of P_σ and the latter one follows from $Q \in \Pi_\zeta$. Thus, $H(P_\sigma) \geqq H(Q)$ for any $Q \in \Pi_\zeta$ and therefore

$$H(P_\sigma) = \max_{Q \in \Pi_\zeta} (H(Q)).$$

Being a strictly convex function, Shannon entropy takes its maximum value in one point of Π_ζ and therefore all the considered permutations σ define the same probability distribution. □

4. SOLUTION OF THE PROBLEM

The goal of this section is to show that a result of the knowledge integration process, i.e. the distribution P_θ, can be sought in a class of probability distributions decomposable with respect to some systems of subsets.

Theorem 1. Let a system $\theta = \{P_{S_1}, \ldots, P_{S_K}\}$ $(S_k \subset \{X_1, \ldots, X_N\})$ be such that $\Pi_\theta \neq \emptyset$ and P_θ be the distribution from Π_θ maximizing the Shannon entropy. Then P_θ is decomposable with respect to any system $\{R_1, \ldots, R_J\}$ $(R_j \subset \{X_1, \ldots, X_N\}$, $j = 1, \ldots, J)$ fulfilling the following two conditions:

(1) there exists a permutation σ such that $(R_{\sigma(1)}, \ldots, R_{\sigma(J)})$ meets the running intersection property,

(2) for every $k = 1, \ldots, K$ there exists $j \in \{1, \ldots, J\}$ such that $S_k \subset R_j$.

Proof. Let a system $\{R_1, \ldots, R_J\}$ fulfill the conditions (1) and (2). Without loosing generality one can assume that the subsets are ordered in the way that $(R_1, \ldots, R_J)$ meets the running intersection property. Consider $P_\theta \in \Pi_\theta$ for which

$$H(P_\theta) = \max_{P \in \Pi_\theta} H(P).$$

The goal is to show that $P_\theta = P_\xi$ where

$$P_\xi = P_\theta^{R_1}[P_\theta^{R_2}/P_\theta^{R_2 \cap R_1}] \ldots [P_\theta^{R_J}/P_\theta^{R_J \cap (R_1 \cup \ldots \cup R_{J-1})}].$$

The system $(R_1, \ldots, R_J)$ and the system if distributions $(P_\theta^{R_1}, \ldots, P_\theta^{R_J})$ fulfill all assumptions of Lemma 1 and therefore the distribution P_ξ maximizes Shannon entropy within the class

$$\Pi_\xi = \{Q\colon Q^{R_j} = P_\theta^{R_j} \text{ for all } j = 1, \ldots, J\}.$$

From the condition (2) it follows immediately that $\Pi_\xi \subset \Pi_\theta$. It is selfevident that

P_θ belongs to Π_ξ and since the distribution P_θ maximizes Shannon entropy within Π_θ the same must hold also for Π_ξ. Thus

$$H(P_\theta) = \max_{Q \in \Pi_\xi} H(Q)) = H(P_\xi)$$

and, therefore $P_\xi = P_\theta$. □

Remark. Notice that if the system θ is rich enough then it may happen that conditions (1) and (2) are met only by the system containing the set $R = \{X_1, \ldots, X_N\}$. It happens for example when θ contains all two-dimensional distributions. In these cases, the presented method cannot be effectively used.

5. IPFP AND DECOMPOSABLE DISTRIBUTIONS

Being decomposable, the multidimensional distribution which is a result of the knowledge integration process can be economically stored in a computer memory. Does it also mean that all distributions $P_1, P_2, \ldots$ which are computed when performing the Iterative Proportional Fitting Procedure are also decomposable and that it makes possible to reduce the space complexity of the process? The answer to this question is a subject of the present paragraph.

Lemma 2. The uniform distribution

$$P_0 = \frac{1}{|\mathbf{X}_1| \ldots |\mathbf{X}_N|}$$

is decomposable with respect to any system meeting the running intersection property.

Proof. This simple assertion follows immediately from the definition of the uniform distribution:

$$P_0^R = \frac{1}{\prod\limits_{l:X_l \in R} |\mathbf{X}_l|}$$

for any $R \subset \{X_1, \ldots, X_N\}$ and therefore

$$P_0^{R_1}[P_0^{R_2}/P_0^{R_2 \cap R_1}] \ldots [P_0^{R_J}/P_0^{R_J \cap (R_1 \cup \ldots \cup R_{J-1})}] =$$

$$= \frac{1}{\prod\limits_{l:X_l \in R_1} |\mathbf{X}_l| \prod\limits_{l:X_l \in R_2 - R_1} |\mathbf{X}_l| \ldots \prod\limits_{l:X_l \in R_J - R_1 \cup \ldots \cup R_{J-1}} |\mathbf{X}_l|} = \frac{1}{\prod\limits_{i=1}^{N} |\mathbf{X}_i|} P_0 \,.$$ □

Theorem 2. Let $\theta = \{P_{S_1}, \ldots, P_{S_K}\}$ $(S_k \subset \{X_1, \ldots, X_N\})$ be a system of distributions such that $\Pi_\theta \neq \emptyset$. Let $\{R_1, \ldots, R_J\}$ meet the conditions (1) and (2) from Theorem 1 and let P_0 be such an N-dimensional distribution that $P_{S_k} \ll P_0^{S_k}$ for all $k = 1, \ldots$ $\ldots, K$. If P_0 is decomposable with respect to the system $\{R_1, \ldots, R_J\}$ then all the distributions P_i, $i = 1, 2, \ldots$ defined by the formula

$$P_i = P_{i-1} \frac{P_{S_j}}{P_{i-1}^{S_j}} \quad \text{for} \quad j = ((i-1) \bmod K) + 1$$

are decomposable with respect to $\{R_1, \ldots, R_J\}$, too.

Proof. In the sequel, let $\{R_1, \ldots, R_J\}$ be a system meeting the required conditions (1) and (2) from Theorem 1. The proof will be performed by the induction for $i = 1, 2, \ldots$. Assume that P_{i-i} is decomposable with respect to $\{R_1, \ldots, R_J\}$. According to Lemma 1 there exists a permutation $\sigma(1), \ldots, \sigma(J)$ such that $(R_{\sigma(1)}, \ldots, R_{\sigma(J)})$ meets the running intersection property and $R_{\sigma(1)} \subset S_j$ for $j = ((i-1) \bmod K) + 1$.

$$P_i = P_{i-1} \frac{P_{S_j}}{P_{i-1}^{S_j}} =$$

$$= \left(P_{i-1}^{R_{\sigma(1)}} \left[P_{i-1}^{R_{\sigma(2)}} / P_{i-1}^{R_{\sigma(2)} \cap R_{\sigma(1)}}\right] \ldots \left[P_{i-1}^{R_{\sigma(J)}} / P_{i-1}^{R_{\sigma(J)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(J-1)})}\right]\right) \frac{P_{S_j}}{P_{i-1}^{S_j}} =$$

$$= \left[P_{S_j} \frac{P_{i-1}^{R_{\sigma(1)}}}{P_{i-1}^{S_j}}\right] \left[P_{i-1}^{R_{\sigma(2)}} / P_{i-1}^{R_{\sigma(2)} \cap R_{\sigma(1)}}\right] \ldots \left[P_{i-1}^{R_{\sigma(J)}} / P_{i-1}^{R_{\sigma(J)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(J-1)})}\right].$$

Since $R_{\sigma(1)} \subset S_j$ it is clear that

$$P_i^{R_{\sigma(1)}} = P_{S_j} \left[P_{i-1}^{R_{\sigma(1)}} / P_{i-1}^{S_j}\right]$$

and for all $l = 2, \ldots, J$

$$P_i^{R_{\sigma(l)}} / P_i^{R_{\sigma(l)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(l-1)})} = P_{i-1}^{R_{\sigma(l)}} / P_{i-1}^{R_{\sigma(l)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(l-1)})}$$

which finishes the proof. □

The last two assertions bring immediately a positive answer to the question asked at the beginning of this section. Moreover, the computations performed in the proof give also an instruction how to realize a respective algorithm. A brief description of it is a content of the next section. But first, a simple condition is presented which makes possible to decompose the knowledge integration process in some cases. In graphical terms, this condition is also presented in [6] where it is used as a basis of *a decomposition algorithm.*

Theorem 3. Let $\theta = \{P_{S_1}, \ldots, P_{S_K}\}$ $(S_k \subset \{X_1, \ldots, X_N\})$ be a system of marginal distributions and $\theta_1 \subset \theta, \theta_2 \subset \theta$ such that $\theta_1 \cup \theta_2 = \theta$. Denote R_i $(i = 1, 2)$ the sets of variables included in the subsystems θ_i

$$R_i = \bigcup_{l : P_{S_l} \in \theta_i} S_l .$$

If $R_1 \cap R_2 \subset S_k$ for some $k \in \{1, \ldots, K\}$ then

$$P_\theta = \frac{P_{\theta_1} P_{\theta_2}}{P_{S_k}^{R_1 \cap R_2}} .$$

Proof. Denote

$$P = \frac{P_{\theta_1} P_{\theta_2}}{P_{S_k}^{R_1 \cap R_2}}.$$

P is a probability distribution (this is a trivial consequence of $R_1 \cap R_2 \subset S_k$ and $P_{\theta_1}^{R_1 \cap R_2} = P_{\theta_2}^{R_1 \cap R_2} = P_{S_k}^{R_1 \cap R_2}$) from Π_θ because any $P_{S_l} \in \theta$ belongs either to θ_1 or θ_2 and therefore $P_{\theta_1}^{S_l}$ or $P_{\theta_2}^{S_l}$ equals P_{S_l}.

Consider any distribution $Q \in \Pi_\theta$

$$H(Q) \leqq H\left(\frac{Q^{R_1} Q^{R_2}}{Q^{R_1 \cap R_2}}\right) = H\left(\frac{Q^{R_1} Q^{R_2}}{P_{S_k}^{R_1 \cap R_2}}\right) =$$

$$= H(Q^{R_1}) + H(Q^{R_2}) - H(P_{S_k}^{R_1 \cap R_2}) \leqq H(P_{\theta_1}) + H(P_{\theta_2}) - (P_{S_k}^{R_1 \cap R_2}) = H(P)$$

and therefore $H(P) = \max\limits_{Q \in \Pi_\theta} (H(Q))$ which implies $P = P_\theta$. □

6. DESIGN OF AN ALGORITHM

First step of any algorithm realizing the Iterative Proportional Fitting Procedure should be the test whether the problem can be simplified by the way proposed by Malvestuto [6], i.e. by repeated application of the following two operations:

(a) delete a variable that belongs to exactly one distribution $P_{S_k} \in \theta$,

(b) remove a distribution P_{S_k} that is a marginal of another (i.e. $\exists\, P_{S_l} \in \theta$: $S_k \subset S_l$),

and by application of *the decomposition algorithm*. This algorithm realizes the idea of decomposition of the problem which is presented here as Theorem 3.

Algorithm.

Input. A system of marginal distributions $\theta = \{P_{S_1}, \ldots, P_{S_K}\}$ which cannot be further decomposed.

Output. System of marginal distributions $\{P_{R_1}, \ldots, P_{R_J}\}$ such that $(R_1, \ldots, R_J)$ meets the running intersection property and

$$P_{R_1}[P_{R_2}/P_{R_2}^{R_2 \cap R_1}] \ldots [P_{R_J}/P_{R_J}^{R_J \cap (R_1 \cup \ldots \cup R_{J-1})}]$$

defines P_θ.

I. *Initial phase.* Find a system $(R_1, \ldots, R_J)$ fulfilling the conditions (1) and (2) from Theorem 1. This can be done very easily with the help of *a fill in algorithm* [8] by adding some edges to the graph $G(V, E)$

$$V = \{X_1, \ldots, X_N\}$$

$$E = \{\{X_i, X_j\} : \exists k \in \{1, \ldots, K\}\ (X_i, X_j \in S_k)\}$$

to make the graph triangulated; maximal cliques then define sets R_j. It is, however, desirable to construct such a system $(R_1, \ldots, R_J)$ with minimum possible value

$$\sum_{j=1}^{J} \prod_{l: X_l \in R_j} |\mathbf{X}_l|.$$

Some ideas how to achieve this optimal solution are in [7].

II. *Iterative phase.* Define

$$P_{0,R_j} = \frac{1}{\prod_{l:X_l \in R_j} |\mathbf{X}_l|}$$

for all $j = 1, \ldots, J$.

For $i = 1, 2, \ldots$ (until it converges) perform the following procedure:

(i) Choose any permutation $\sigma(1), \ldots, \sigma(J)$ such that $(R_{\sigma(1)}, \ldots, R_{\sigma(J)})$ meets the running intersection property and $R_{\sigma(1)} \supset S_j$ for $j = ((i-1) \bmod K) + 1$. As mentioned above, it can be done by *the restricted maximum cardinality search* cf. [8].

(ii) Compute

$$P_{i,R_{\sigma(1)}} = P_{S_j}[P_{i-1,R_{\sigma(1)}}/P^{S_j}_{i-1,R_{\sigma(1)}}]$$

(iii) For $l = 2, \ldots, J$ compute the following two distributions:

(a) marginal distributions of the distribution

$$Q_{i,l} = P_{i,R_{\sigma(1)}}[P_{i,R_{\sigma(2)}}/P^{R_{\sigma(2)} \cap R_{\sigma(1)}}_{i,R_{\sigma(2)}}] \ldots [P_{i,R_{\sigma(l-1)}}/P^{R_{\sigma(l-1)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(l-2)})}_{i,R_{\sigma(l-1)}}]$$

for variables $R_{\sigma(l)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(l-1)})$ only and

(b)

$$P_{i,R_{\sigma(l)}} = Q^{R_{\sigma(l)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(l-1)})}_{i,l}[P_{i-1,R_{\sigma(l)}}/P^{R_{\sigma(l)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(l-1)})}_{i-1,R_{\sigma(l)}}]$$

Remark. Notice that implementation of the last step of the process is not so difficult because

$$[P_{i,R_{\sigma(l)}}/P^{R_{\sigma(l)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(l-1)})}_{i,R_{\sigma(l)}}] = [P_{i-1,R_{\sigma(l)}}/P^{R_{\sigma(l)} \cap (R_{\sigma(1)} \cup \ldots \cup R_{\sigma(l-1)})}_{i-1,R_{\sigma(l)}}]$$

for all $l = 2, \ldots, J$.

ACKNOWLEDGEMENT

The author is grateful to F. M. Malvestuto for his comments and valuable suggestions which generalized some results and improved legibility of the paper.

(Received April 25, 1990.)

REFERENCES

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

[2] 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. Statist. *11* (1940), 427–444.

[3] A. Feinstein: Foundations of Information Theory. McGraw-Hill, New York–Toronto–London 1958.

[4] R. Jiroušek: A survey of methods used in probabilistic expert system for knowledge integration. Knowledge Based Systems *3* (1990), 1, 7–12.

[5] H. G. Kellerer: Verteilungsfunktionen mit gegeben Marginalverteilungen. Z. Warhsch. Verw. Gebiete *3* (1964), 247–270.

[6] F. M. Malvestuto: Computing the maximum-entropy extension of given discrete probability distributions. Comput. Statist. Data Anal. *8* (1989), 299–311.

[7] Lianwen Zhang: Studies on finding hypertree covers for hypergraphs. Working Paper No. 198, School of Business, The University of Kansas, Lawrence 1988.

[8] R. E. Tarjan and M. Yannakakis: Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. SIAM J. Comput. *13* (1984), 3, 566–579.

Radim Jiroušek, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.