KYBERNETIKA — VOLUME 42 (2006), NUMBER 4, PAGES 405 – 422

MARGINALIZATION IN MULTIDIMENSIONAL COMPOSITIONAL MODELS

Vladislav Bína and Radim Jiroušek

Efficient computational algorithms are what made graphical Markov models so popular and successful. Similar algorithms can also be developed for computation with compositional models, which form an alternative to graphical Markov models. In this paper we present a theoretical basis as well as a scheme of an algorithm enabling computation of marginals for multidimensional distributions represented in the form of compositional models.

Keywords: compositional models, marginalization, Bayesian network

AMS Subject Classification: 65C50, 60E99, 68T37

1. INTRODUCTION

Representation and processing of multidimensional probability distributions were made possible by success achieved in the field of graphical Markov models (see e. g. [7]) during the last twenty years. Here we have in mind not only ample theoretical background but also thoroughly elaborated algorithmic apparatus, which enabled design of very efficient software packages (e. g. HUGIN [2]). As an alternative to graphical models, during approximately the past eight years we have been studying the non-graphical approach of *compositional models*, which is based on the idea that multidimensional distributions can be assembled – *composed* – from a system of low-dimensional ones.

In this paper we present a theoretical background supporting a possible solution of one hard problem, which has not been implemented even in such systems as HUGIN: the marginalization of multidimensional distributions. For Bayesian networks a solution of this problem was proposed by Ross Shachter in [8, 9]. His famous procedure is based on two rules: *node deletion* and *edge reversal*. Roughly speaking, the efficiency of his approach corresponds to the efficiency of our process if we did not employ the speed-up theoretically supported by Theorem 2 presented below. This theorem, namely, takes advantage of the main difference between Bayesian networks [3] and compositional models revealed in [6]. This advantage consists of the fact that compositional models, when represented by perfect sequences, express explicitly some marginals, whose computation in a Bayesian network may be computationally expensive. However, it should be stressed up front that in cases where Theorem 2 cannot be applied (i. e. there does not exist a respective decomposition) the proposed procedure only increases the computational time because we do not have an efficient procedure recognizing such a disadvantageous situation.

2. OPERATORS OF COMPOSITION

In this paper we will consider a system of finite-valued random variables with indices from a non-empty finite set N. All the probability distributions discussed in the paper will be denoted by Greek letters. For $K \subset N$, $\kappa(x_K)$ denotes a distribution of variables $X_K = \{X_i\}_{i \in K}$, which is defined on all subsets of a Cartesian product $\times_{i \in K} \mathbf{X}_i$.

Having a distribution $\kappa(x_K)$ and $L \subset K$, we will denote its corresponding marginal distribution either $\kappa(x_L)$, or, using the notation used by Glenn Shafer and Prakash Shenoy (see e. g. [10]), $\kappa^{\downarrow L}$. These symbols are used when we want to highlight the variables for which the marginal distribution is defined. If we want to specify variables which are eliminated in the process of marginalization, we will use the symbol κ^{-M}, where M is a set of indices of the variables which do not appear among the arguments of the resulting marginal distribution. In our case, $M = K \setminus L$.

In order to describe how to compose low-dimensional distributions to get a distribution of a higher dimension we will use the following operator of composition.

Definition 1. For two arbitrary distributions $\kappa(x_K)$ and $\lambda(x_L)$ their *composition* is given by the formula

$$\kappa \triangleright \lambda = \begin{cases} \frac{\kappa \cdot \lambda}{\lambda^{\downarrow K \cap L}} & \text{when} \quad \kappa^{\downarrow K \cap L} \ll \lambda^{\downarrow K \cap L}, \\ \text{undefined} & \text{otherwise}, \end{cases}$$

where the symbol $\kappa^{\downarrow M} \ll \lambda^{\downarrow M}$ denotes that $\kappa^{\downarrow M}$ is *dominated* by $\lambda^{\downarrow M}$, which means (in the considered finite setting)

$$\forall x_M \in \times_{i \in M} \mathbf{X}_i \quad (\lambda(x_M) = 0 \implies \kappa(x_M) = 0).$$

Remark. If the marginal $\lambda^{\downarrow K \cap L}$ dominates $\kappa^{\downarrow K \cap L}$ then the formula in the definition is evaluated point-wise, i. e., for each $x \in \mathbf{X}_{K \cup L}$ value

$$(\kappa \triangleright \lambda)(x) = \frac{\kappa(x_K) \cdot \lambda(x_L)}{\lambda(x_{K \cap L})}$$

is computed (in case that $\lambda(x_{K \cap L}) = 0$ we define $\frac{0 \cdot 0}{0} = 0$).

Since the outcome of the composition (if it is defined) is a new distribution, we can iteratively repeat the application of this operator, composing a multidimensional model. This is why these multidimensional distributions are called *compositional models*. To describe such a model it is enough to introduce an ordered system of low-dimensional distributions $\kappa_1, \kappa_2, \ldots, \kappa_n$. We will refer to this as to a *generating sequence*, to which the operator is applied from left to right:

$$\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3 \triangleright \ldots \triangleright \kappa_{n-1} \triangleright \kappa_n = (\ldots((\kappa_1 \triangleright \kappa_2) \triangleright \kappa_3) \triangleright \ldots \triangleright \kappa_{n-1}) \triangleright \kappa_n.$$

Then we say that a generating sequence defines (or represents) a multidimensional compositional model.

In the process of marginalization we will also need another important operator.

Definition 2. For two arbitrary distributions $\kappa(x_K)$, $\lambda(x_L)$ and a set of indices of variables $M \subset N$, by application of an *anticipating operator* parameterized by the index set M, we understand computation of the following distribution

$$\kappa \circledtriangleright_M \lambda = \left(\lambda^{\downarrow(M\setminus K)\cap L} \cdot \kappa\right) \triangleright \lambda.$$

Remark. Notice that $\kappa \circledtriangleright_M \lambda$ is undefined only if $\kappa \triangleright \lambda$ is undefined. Analogously to Definition 1, if the composition is defined, computation of the expression

$$(\kappa \circledtriangleright_M \lambda)(x_{K\cup L}) = \left(\lambda(x_{(M\setminus K)\cap L}) \cdot \kappa(x_K)\right) \triangleright \lambda(x_L)$$

is performed point-wise.

3. BASIC PROPERTIES

In the following text we will need three simple lemmas which follow almost immediately from the definition of the operator of composition (their proofs can also be found in our previous papers).

Lemma 1. Consider two distributions $\kappa(x_K)$ and $\lambda(x_L)$. If the composition $\kappa \triangleright \lambda$ is defined then

$$(\kappa \triangleright \lambda)^{\downarrow K} = \kappa.$$

Lemma 2. Let for two distributions $\kappa(x_K)$ and $\lambda(x_L)$ their composition $\kappa \triangleright \lambda$ is defined and $L \subseteq M \subseteq K \cup L$. Then

$$\kappa \triangleright \lambda = \kappa \triangleright (\kappa \triangleright \lambda)^{\downarrow M}.$$

Lemma 3. Let M be such that $K \cap L \subseteq M \subseteq L$; then

$$\kappa \triangleright \lambda = (\kappa \triangleright \lambda) \triangleright \lambda.$$

The remaining four lemmas are more complex and therefore we present them with their proofs.

Lemma 4. Let $K, L, M \subseteq N$. If $K \cup L \supseteq M \supseteq K \cap L$ then for any probability distributions $\kappa(x_K)$ and $\lambda(x_L)$

$$(\kappa \triangleright \lambda)^{\downarrow M} = \kappa^{\downarrow K\cap M} \triangleright \lambda^{\downarrow L\cap M}.$$

Proof. Let us first mention that $\kappa \triangleright \lambda$ is not defined only if $\kappa^{\downarrow K\cap L} \not\ll \lambda^{\downarrow K\cap L}$. However, because of the assumption laid on M, $K \cap L = (K \cap M) \cap (L \cap M)$, and therefore it holds true if and only if $\kappa^{\downarrow K\cap M} \triangleright \lambda^{\downarrow L\cap M}$ is not defined, too. Therefore, if one composition is not defined then neither is the other composition defined. To prove the assertion in case that $\kappa \triangleright \lambda$ is defined, let us first compute

$$\begin{aligned}(\kappa \triangleright \lambda)(x_{K\cup M}) &= \sum_{x_{L\setminus M}\in \mathbf{X}_{L\setminus M}} \frac{\kappa(x_K)\lambda(x_{L\cap M}, x_{L\setminus M})}{\lambda(x_{L\cap K})} \\ &= \frac{\kappa(x_K)\lambda(x_{L\cap M})}{\lambda(x_{L\cap K})} \sum_{x_{L\setminus M}\in \mathbf{X}_{L\setminus M}} \lambda(x_{L\setminus M}|x_{L\cap M}) = \kappa(x_K) \triangleright \lambda(x_{L\cap M}).\end{aligned}$$

Now we can compute the required marginal distribution

$$\begin{aligned}(\kappa \triangleright \lambda)(x_M) &= ((\kappa \triangleright \lambda)(x_{K\cup M}))^{\downarrow M} = (\kappa(x_K) \triangleright \lambda(x_{L\cap M}))^{\downarrow M} \\ &= \sum_{x_{K\setminus M}\in \mathbf{X}_{K\setminus M}} \frac{\kappa(x_{K\cap M}, x_{K\setminus M})\lambda(x_{L\cap M})}{\lambda(x_{L\cap K})} \\ &= \frac{\kappa(x_{K\cap M})\lambda(x_{L\cap M})}{\lambda(x_{L\cap K})} \sum_{x_{K\setminus M}\in \mathbf{X}_{K\setminus M}} \kappa(x_{K\setminus M}|x_{K\cap M}) = \kappa(x_{K\cap M}) \triangleright \lambda(x_{L\cap M}). \qquad \square\end{aligned}$$

Let us emphasize that when describing a generating sequence it is necessary to explain that the operator of composition is always applied from left to right. This is because the operator is neither commutative nor associative. So, generally $\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3 \neq \kappa_1 \triangleright (\kappa_2 \triangleright \kappa_3)$. Situations under which it is possible to exchange the ordering of operators are described in the following two assertions.

Lemma 5. If $\kappa_1(x_{K_1})$, $\kappa_2(x_{K_2})$ and $\kappa_3(x_{K_3})$ are such that $K_1 \supseteq (K_2 \cap K_3)$ then

$$\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3 = \kappa_1 \triangleright \kappa_3 \triangleright \kappa_2. \tag{1}$$

Proof. First, let us show that the left hand side expression in (1) is not defined *iff* the right hand side of this formula is not defined. From the definition of the operators we know that $\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3$ is not defined *iff*

$$\kappa_1^{\downarrow K_1\cap K_2} \not\ll \kappa_2^{\downarrow K_1\cap K_2} \quad \text{or} \quad (\kappa_1 \triangleright \kappa_2)^{\downarrow (K_1\cup K_2)\cap K_3} \not\ll \kappa_3^{\downarrow (K_1\cup K_2)\cap K_3}.$$

Analogously, $\kappa_1 \triangleright \kappa_3 \triangleright \kappa_2$ is not defined *iff*

$$\kappa_1^{\downarrow K_1\cap K_3} \not\ll \kappa_3^{\downarrow K_1\cap K_3} \quad \text{or} \quad (\kappa_1 \triangleright \kappa_3)^{\downarrow (K_1\cup K_3)\cap K_2} \not\ll \kappa_2^{\downarrow (K_1\cup K_3)\cap K_2}.$$

Under the given assumption $K_1 \supseteq (K_2 \cap K_3)$, these two conditions coincide because

$$(K_1 \cup K_2) \cap K_3 = K_1 \cap K_3 \ \ \& \ \ (K_1 \cup K_3) \cap K_2 = K_1 \cap K_2, \tag{2}$$

and

$$(\kappa_1 \triangleright \kappa_2)^{\downarrow (K_1 \cup K_2) \cap K_3} = \kappa_1^{\downarrow K_1 \cap K_3} \ \ \& \ \ (\kappa_1 \triangleright \kappa_3)^{\downarrow (K_1 \cup K_3) \cap K_2} = \kappa_1^{\downarrow K_1 \cap K_2}.$$

Now, let us assume that both the expressions in formula (1) are defined. Because of (2) the expressions

$$\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3 = \frac{\kappa_1 \kappa_2 \kappa_3}{\kappa_2^{\downarrow K_1 \cap K_2} \kappa_3^{\downarrow K_3 \cap (K_1 \cup K_2)}},$$

$$\kappa_1 \triangleright \kappa_3 \triangleright \kappa_2 = \frac{\kappa_1 \kappa_2 \kappa_3}{\kappa_3^{\downarrow K_1 \cap K_3} \kappa_2^{\downarrow K_2 \cap (K_1 \cup K_3)}}$$

are equivalent to each other, which finishes the proof. □

Lemma 6. If $\kappa_1(x_{K_1})$, $\kappa_2(x_{K_2})$ and $\kappa_3(x_{K_3})$ are such that $\kappa_1 \triangleright (\kappa_2 \circledtriangleright_{K_1} \kappa_3)$ is defined, then

$$\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3 = (\kappa_1 \triangleright \kappa_2) \triangleright \kappa_3 = \kappa_1 \triangleright (\kappa_2 \circledtriangleright_{K_1} \kappa_3). \tag{3}$$

Proof. Assume that $\kappa_1 \triangleright (\kappa_2 \circledtriangleright_{K_1} \kappa_3)$ is defined. It means that

$$\kappa_1^{\downarrow K_1 \cap (K_2 \cup K_3)} \ll (\kappa_2 \circledtriangleright_{K_1} \kappa_3)^{\downarrow K_1 \cap (K_2 \cup K_3)}, \tag{4}$$

and, as a consequence of the fact that dominance holds also for the respective marginal distributions, $\kappa_1^{\downarrow K_1 \cap K_2} \ll \kappa_2^{\downarrow K_1 \cap K_2}$. This guarantees that $\kappa_1 \triangleright \kappa_2$ is defined. Let us now show by contradiction that $(\kappa_1 \triangleright \kappa_2) \triangleright \kappa_3$ must also be defined. Assume it is not defined. It means that there exists $x \in \mathbf{X}_{K_1 \cup K_2 \cup K_3}$ such that in the expression

$$(\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3)(x) = \frac{\kappa_1(x_{K_1}) \cdot \kappa_2(x_{K_2}) \cdot \kappa_3(x_{K_3})}{\kappa_2(x_{K_2 \cap K_1}) \cdot \kappa_3(x_{K_3 \cap (K_1 \cup K_2)})}$$

$\kappa_3(x_{K_3 \cap (K_1 \cup K_2)}) = 0$ and simultaneously $\kappa_1(x_{K_1}) \cdot \kappa_2(x_{K_2}) > 0$. This, however, contradicts to our assumption that $\kappa_2 \circledtriangleright_{K_1} \kappa_3$ is defined: as we can see from the respective formula

$$(\kappa_2 \circledtriangleright_{K_1} \kappa_3)(x) = \frac{\kappa_3(x_{(K_1 \setminus K_2) \cap K_2}) \cdot \kappa_2(x_{K_2}) \cdot \kappa_3(x_{K_3})}{\kappa_3(x_{K_3 \cap (K_1 \cup K_2)})},$$

$$\kappa_3(x_{K_3 \cap (K_1 \cup K_2)}) = 0 \implies \kappa_2(x_{K_2}) = 0.$$

Now, assuming $\kappa_1 \triangleright (\kappa_2 \circledtriangleright_{K_1} \kappa_3)$ is defined let us compute (using the definition of the operator $\circledtriangleright$ and Lemma 4):

$$\kappa_1 \triangleright (\kappa_2 \circledtriangleright_{K_1} \kappa_3) = \frac{\kappa_1 \frac{\kappa_3^{\downarrow (K_1 \setminus K_2)\cap K_3} \kappa_2 \kappa_3}{\kappa_3^{\downarrow (K_1 \cup K_2)\cap K_3}}}{\left(\frac{\kappa_3^{\downarrow (K_1 \setminus K_2)\cap K_3} \kappa_2 \kappa_3}{\kappa_3^{\downarrow (K_1 \cup K_2)\cap K_3}} \right)^{\downarrow (K_2 \cup K_3)\cap K_1}}$$

$$= \frac{\kappa_3^{\downarrow (K_1 \setminus K_2)\cap K_3} \frac{\kappa_1 \kappa_2 \kappa_3}{\kappa_3^{\downarrow (K_1 \cup K_2)\cap K_3}}}{\kappa_3^{\downarrow (K_1 \setminus K_2)\cap K_3} \left(\frac{\kappa_2 \kappa_3}{\kappa_3^{\downarrow (K_1 \cup K_2)\cap K_3}} \right)^{\downarrow (K_2 \cup K_3)\cap K_1}} = \frac{\frac{\kappa_1 \kappa_2 \kappa_3}{\kappa_3^{\downarrow (K_1 \cup K_2)\cap K_3}}}{\left(\frac{\kappa_2 \kappa_3}{\kappa_3^{\downarrow (K_1 \cup K_2)\cap K_3}} \right)^{\downarrow (K_2 \cup K_3)\cap K_1}},$$

where the second modification is feasible because

$$(K_1 \setminus K_2) \cap K_3 \subseteq (K_2 \cup K_3) \cap K_1.$$

Let us focus our attention on the denominator of the last fraction. It is a marginal of a product of κ_2 with a conditional distribution

$$\kappa_3(x_{K_3 \setminus (K_1 \cup K_2)} | x_{K_3 \cap (K_1 \cup K_2)}).$$

When computing this marginal, we have to sum up over all combinations of values of variables $X_{(K_2 \cup K_3)\setminus K_1}$. In the following computations we will separate these variables into two groups: $X_{K_2 \setminus K_1}$ and $X_{K_3 \setminus (K_1 \cup K_2)}$. $x_{K_2} \in \mathbf{X}_{K_2}$ is thus a vector of values of variables X_{K_2} which can be split into two parts: $x_{K_2} = (x_{K_2 \setminus K_1}, x_{K_2 \cap K_1})$. Analogously, for $x_{K_3 \cap (K_1 \cup K_2)} \in \mathbf{X}_{K_3 \cap (K_1 \cup K_2)}$ we will consider parts

$$x_{K_3 \cap (K_1 \cup K_2)} = (x_{K_3 \cap K_1}, x_{(K_3 \cap K_2)\setminus K_1}).$$

Using this notation, we can compute:

$$\begin{aligned}
&\left(\kappa_2(x_{K_2})\kappa_3(x_{K_3 \setminus (K_1 \cup K_2)} | x_{K_3 \cap (K_1 \cup K_2)})\right)^{\downarrow (K_2 \cup K_3)\cap K_1} \\
&= \sum_{x_{K_2 \setminus K_1} \in \mathbf{X}_{K_2 \setminus K_1}} \sum_{x_{K_3 \setminus (K_1 \cup K_2)} \in \mathbf{X}_{K_3 \setminus (K_1 \cup K_2)}} \kappa_2(x_{K_2 \cap K_1}, x_{K_2 \setminus K_1}) \\
&\qquad \cdot \kappa_3(x_{K_3 \setminus (K_1 \cup K_2)} | x_{(K_3 \cap K_2)\setminus K_1}, x_{K_3 \cap K_1}) \\
&= \kappa_2(x_{K_2 \cap K_1}) \sum_{x_{K_2 \setminus K_1}} \kappa_2(x_{K_2 \setminus K_1} | x_{K_2 \cap K_1}) \\
&\qquad \sum_{x_{K_3 \setminus (K_1 \cup K_2)}} \kappa_3(x_{K_3 \setminus (K_1 \cup K_2} | x_{(K_3 \cap K_2)\setminus K_1}, x_{K_3 \cap K_1}) \\
&= \kappa_2(x_{K_2 \cap K_1}).
\end{aligned}$$

Substituting this result back into the denominator of the fraction, we get

$$\kappa_1 \triangleright (\kappa_2 \circledtriangleright_{K_1} \kappa_3) = \frac{\frac{\kappa_1\kappa_2\kappa_3}{\kappa_3^{\downarrow(K_1\cup K_2)\cap K_3}}}{\kappa_2^{\downarrow K_2\cap K_1}} = \frac{\kappa_1\kappa_2\kappa_3}{\kappa_2^{\downarrow K_2\cap K_1}\kappa_3^{\downarrow(K_1\cup K_2)\cap K_3}} = \kappa_1 \triangleright \kappa_2 \triangleright \kappa_3.$$

which completes the proof. □

Remark. Notice that Lemma 6 does not claim that equality (3) holds true when $\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3$ is defined.

4. MARGINALIZATION IN COMPOSITIONAL MODEL

Now we will focus our attention on possibilities of marginalization of distributions given by generating sequences. From now on, we will consider a generating sequence

$$\kappa_1(x_{K_1}) \triangleright \kappa_2(x_{K_2}) \triangleright \ldots \kappa_n(x_{K_n}).$$

Therefore whenever we use distribution κ_j, we assume it is defined for variables X_{K_j}. A direct corollary of the following important assertion formulates rules which make it possible to decrease dimensionality of compositional models by one. By iterative application of these rules we may obtain any required marginal.

Theorem 1. Let $\kappa_1, \kappa_2, \ldots, \kappa_n$ be a generating sequence, and $\ell \in N$ and $I \subseteq \{1, \ldots, n\}$ be such that[1]

$$\ell \in \bigcap_{i \in I} K_i \quad \& \quad \ell \notin \bigcup_{i \notin I} K_i.$$

For all $i \in I, i \neq \min(I)$ denote the maximal preceding index from I by $a(i)$:

$$a(i) = \max\left(I \cap \{1, \ldots, i-1\}\right),$$

and $M_i = (K_1 \cup \ldots \cup K_{i-1}) \setminus \{\ell\}$. Further denote for $i = \min(I)$ $\pi_i = \kappa_i$ and for all other $i \in I, i \neq \min(I)$

$$\pi_i = \pi_{a(i)} \circledtriangleright_{M_i} \kappa_i,$$

and

$$\lambda_i = \pi_i^{-\{\ell\}}, \forall i \in I \;\; \& \;\; \lambda_i = \kappa_i, \forall i \notin I.$$

If all the distributions π_i as well as distribution $\lambda_1 \triangleright \lambda_2 \triangleright \ldots \triangleright \lambda_n \triangleright \pi_{\max(I)}$ are defined then

$$(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n) = \lambda_1 \triangleright \lambda_2 \triangleright \ldots \triangleright \lambda_n \triangleright \pi_{\max(I)},$$

and therefore

$$(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)^{-\{\ell\}} = \lambda_1 \triangleright \lambda_2 \triangleright \ldots \triangleright \lambda_n. \tag{5}$$

[1] $i \notin I$ stands for $i \in \{1, \ldots, n\} \setminus I$.

Proof. Let us start proving the theorem for $|I| = 1$ and denote $i \in I$. Since $K_1 \cup \ldots \cup K_{i-1}$ does not contain ℓ, we can apply Lemma 3, which yields

$$\kappa_1 \triangleright \ldots \triangleright \kappa_{i-1} \triangleright \kappa_i = \kappa_1 \triangleright \ldots \triangleright \kappa_{i-1} \triangleright \kappa_i^{-\{\ell\}} \triangleright \kappa_i = \lambda_1 \triangleright \ldots \triangleright \lambda_i \triangleright \kappa_i.$$

Distribution $(\lambda_1 \triangleright \ldots \triangleright \lambda_i)$ is defined for variables with indices from $(K_1 \cup \ldots \cup K_i) \setminus \{\ell\}$; this contains $K_i \cap K_j$ for all $j = i+1, \ldots, n$, because none of these K_j contain ℓ. Therefore, applying Lemma 5 $(n-i)$-times, we get

$$\begin{aligned}\kappa_1 \triangleright \ldots \triangleright \kappa_i \triangleright \kappa_{i+1} \triangleright \ldots \triangleright \kappa_n &= \lambda_1 \triangleright \ldots \triangleright \lambda_i \triangleright \kappa_i \triangleright \kappa_{i+1} \triangleright \ldots \triangleright \kappa_n \\ &= \lambda_1 \triangleright \ldots \triangleright \lambda_i \triangleright \kappa_{i+1} \triangleright \ldots \triangleright \kappa_n \triangleright \kappa_i = \lambda_1 \triangleright \ldots \triangleright \lambda_n \triangleright \pi_i.\end{aligned}$$

Now, we will prove the assertion for a general I assuming that it has been proven for the situations when ℓ is contained in a smaller number of sets than $|I|$. Denote $i = \max(I)$ and $a(i) = \max(I \cap \{1, \ldots, i-1\})$. In the following computations we will first use Lemma 3, then Lemma 6, and finally $(n - i_m)$-times Lemma 5.

$$\begin{aligned}\kappa_1 \triangleright \ldots \triangleright \kappa_{i-1} \triangleright \kappa_i \triangleright \ldots \triangleright \kappa_n &= \lambda_1 \triangleright \ldots \triangleright \lambda_{i-1} \triangleright \pi_{a(i)} \triangleright \kappa_i \triangleright \ldots \triangleright \kappa_n \\ &= \lambda_1 \triangleright \ldots \triangleright \lambda_{i-1} \triangleright \left(\pi_{a(i)} \textcircled{\triangleright}_{M_i} \kappa_i\right) \triangleright \kappa_{i+1} \triangleright \ldots \triangleright \kappa_n \\ &= \lambda_1 \triangleright \ldots \triangleright \lambda_{i-1} \triangleright \pi_i^{-\{\ell\}} \triangleright \pi_i \triangleright \kappa_{i+1} \triangleright \ldots \triangleright \kappa_n \\ &= \lambda_1 \triangleright \ldots \triangleright \lambda_i \triangleright \pi_i \triangleright \kappa_{i+1} \triangleright \ldots \triangleright \kappa_n \\ &= \lambda_1 \triangleright \ldots \triangleright \lambda_i \triangleright \kappa_{i+1} \triangleright \ldots \triangleright \kappa_n \triangleright \pi_i = \lambda_1 \triangleright \ldots \triangleright \lambda_n \triangleright \pi_i.\end{aligned}$$

Validity of Equation (5) follows immediately from the preceding formula and the fact that X_ℓ appears among the argument of no λ_i. Since we assume that all distributions π_i as well as $\lambda_\triangleright \ldots \triangleright \lambda_n$ are defined, all the expressions in the preceding computations are defined due to Lemma 6 (application of Lemmas 3 and 5 cannot cause any problems). □

Iterative application of Theorem 1 always leads to the desired marginal distribution and fully corresponds to the Shachter's marginalization procedure. In fact, application of the anticipating operator in a way corresponds to the *inheritance of parents* in his *edge reversal rule*. So one should not be surprised that the computational complexity of this process strongly depends on the number of occurrences of the variable ℓ among the arguments of the distributions in the considered generating sequence (it can be to some extent controlled by a proper ordering of variables which are to be eliminated). Beginning from the second occurrence of this variable we should replace distribution κ_i with an expression containing one or more anticipating operators. However, when the variable which is to be deleted is contained in the argument of only one of the distributions, Theorem 1 simplifies into the following Corollary. In this case, it is sufficient to marginalize only one distribution and all the others remain unchanged. This describes situations when Shachter's *deletion rule* may be applied either directly (the node is terminal), or when application of the *edge reversal* rule does not introduce new edges in the considered Bayesian network.

Corollary. Let $\kappa_1, \kappa_2, \ldots, \kappa_n$ be a generating sequence. If $\ell \in K_j$ for some $j \in \{1, \ldots, n\}$ and $\ell \notin K_i$ for all $i \neq j$ then

$$(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)^{-\{\ell\}} = \kappa_1 \triangleright \ldots \triangleright \kappa_{j-1} \triangleright \kappa_j^{-\{\ell\}} \triangleright \kappa_{j+1} \triangleright \ldots \triangleright \kappa_n.$$

More effective marginalizing procedures are, however, based on the following assertion, which is a generalization of Theorem 11 from [5]. It describes conditions under which a number of variables may be deleted in one computationally simple step. But first let us define the auxiliary notion of a *reduction of a generating sequence*, which will simplify formulations in the following text.

Definition 3. Let $\kappa_1, \kappa_2, \ldots, \kappa_n$ be a generating sequence, and $s \in Z \subsetneq \{1, \ldots, n\}$ be such that

$$\left(\bigcup_{j \in Z} K_j\right) \cap \left(\bigcup_{j \notin Z} K_j\right) \subseteq K_s.$$

Then we say that s and Z *determine a reduction* of generating sequence $\kappa_1, \ldots, \kappa_n$ (or simply that (s, Z) *is a reduction*).

Theorem 2. Let $s \in Z$ and $Z \subsetneq \{1, \ldots, n\}$ determine a reduction of a generating sequence $\kappa_1, \kappa_2, \ldots, \kappa_n$, and let all distributions κ_i be positive. Let $J = \bigcup_{j \in Z} K_j$, $\mu = (\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)^{\downarrow K_s}$, and distributions λ_i be defined

$$\begin{array}{lll} \lambda_j = \kappa_j & \text{for} & j \in Z, \\ \lambda_j = \mu^{\downarrow J \cap O_j} & \text{for} & j \notin Z \text{ and } \quad O_j = \bigcup_{i \in \{1, \ldots, j\} \setminus Z} K_i. \end{array}$$

Then the marginal distribution $(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)^{\downarrow J}$ can be expressed as a compositional model

$$(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)^{\downarrow J} = \lambda_1 \triangleright \lambda_2 \triangleright \ldots \triangleright \lambda_n.$$

Proof. Let $\{\ell_1, \ell_2, \ldots, \ell_m\} = (K_1 \cup \ldots \cup K_n) \setminus J$ be any ordering of the indices to be eliminated. Let $\nu_1^1, \nu_2^1, \ldots, \nu_n^1$ be a generating sequence received by application of Theorem 1 to the sequence $\kappa_1, \kappa_2, \ldots, \kappa_n$ and the index ℓ_1. What can be said about the generating sequence $\nu_1^1, \nu_2^1, \ldots, \nu_n^1$?

1. $(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)^{-\{\ell_1\}} = (\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)^{\downarrow J \cup \{\ell_2, \ldots, \ell_m\}} = \nu_1^1 \triangleright \nu_2^1 \triangleright \ldots \triangleright \nu_n^1$;

2. For all $j \in Z$, $\nu_j^1 = \kappa_j$;

3. For each $j \notin Z$, ν_j^1 is a distribution of variables with indices from K_j and possibly some other indices from O_j but not ℓ_1. Therefore, the respective set of indices contains $K_j \setminus \{\ell_1\}$ and is contained in $O_j \setminus \{\ell_1\}$.

Now, iterative application of Theorem 1 to the generating sequences ν_1^i, $\nu_2^i, \dots \nu_n^i$ and the indices ℓ_{i+1} yields sequences $\nu_1^{i+1}, \nu_2^{i+1}, \dots, \nu_n^{i+1}$ for $i = 1, \dots, m-1$. Denote by N_j^i the sets of indices of variables in the argument of ν_j^i. Analogous to the first step, we can see that for all of these sequences:

4. $(\kappa_1 \triangleright \kappa_2 \triangleright \dots \triangleright \kappa_n)^{-\{\ell_1,\dots,\ell_i\}} = (\kappa_1 \triangleright \kappa_2 \triangleright \dots \triangleright \kappa_n)^{\downarrow J \cup \{\ell_{i+1},\dots,\ell_m\}} = \nu_1^i \triangleright \nu_2^i \triangleright \dots \triangleright \nu_n^i$;

5. For all $j \in Z$, $\nu_j^i = \kappa_j$, and thus $N_j^i = K_j$;

6. For each $j \notin Z$, $K_j \setminus \{\ell_1, \dots, \ell_i\} \subseteq N_j^i \subseteq O_j \setminus \{\ell_1, \dots, \ell_i\}$.

Since $(\kappa_1 \triangleright \kappa_2 \triangleright \dots \triangleright \kappa_n)^{\downarrow J} = \nu_1^m \triangleright \nu_2^m \triangleright \dots \triangleright \nu_n^m$, to finish the proof we have to show that $\nu_1^m \triangleright \dots \triangleright \nu_n^m = \lambda_1 \triangleright \dots \triangleright \lambda_n$. The elements with indices $j \in Z$ equal each other: $\kappa_j = \nu_j^m = \lambda_j$. Therefore, what has remained to be shown is that substituting ν_j^m with λ_j for $j \notin Z$ does not change the generated distribution.

From the relationships presented in the above items 4. and 6. (and using the fact that $O_j \setminus \{\ell_1, \dots, \ell_i\} = O_j \cap J$) we get

$$N_j^m \subseteq O_j \cap J \subseteq \bigcup_{i=1}^{j} N_i^m.$$

This enables us to apply Lemma 2 (where $M = J \cap O_j$), getting

$$(\nu_1^m \triangleright \nu_2^m \triangleright \dots \triangleright \nu_{j-1}^m) \triangleright \nu_j^m = (\nu_1^m \triangleright \nu_2^m \triangleright \dots \triangleright \nu_{j-1}^m) \triangleright (\nu_1^m \triangleright \dots \triangleright \nu_j^m)^{\downarrow J \cap O_j}.$$

Since both $(\nu_1^m \triangleright \dots \triangleright \nu_j^m)$ and μ are marginal distributions of $\kappa_1 \triangleright \dots \triangleright \kappa_n$, their common marginals must equal each other:

$$(\nu_1^m \triangleright \dots \triangleright \nu_j^m)^{\downarrow J \cap O_j} = \mu^{\downarrow J \cap O_j} = \lambda_j,$$

and therefore

$$(\nu_1^m \triangleright \nu_2^m \triangleright \dots \triangleright \nu_{j-1}^m) \triangleright \nu_j^m = (\nu_1^m \triangleright \nu_2^m \triangleright \dots \triangleright \nu_{j-1}^m) \triangleright \lambda_j.$$

Repeating these considerations for all $j \notin Z$, one can substitute ν_j^m by λ_j for all $j \notin Z$, which completes the proof. □

Remark. The reader familiar with our preceding papers knows that one of the most important notions of theory of compositional models is a so called *perfect sequence*. All distributions κ_i appearing in a perfect sequence are marginals of the represented distribution $\kappa_1 \triangleright \dots \triangleright \kappa_n$. Thus, if Theorem 2 is applied to a perfect sequence, $\mu = \kappa_s$, which further simplifies the necessary computations. So we see that application of this assertion to perfect sequences is trivial and computationally inexpensive.

Remark. The assumption of positivity for distributions $\kappa_1, \kappa_2, \ldots, \kappa_n$ was introduced just to avoid problems with application of Theorem 1 (positive distributions dominate each other and therefore any generating sequence of positive distributions defines a compositional model). The reader certainly noticed that even for general distributions κ_i (i. e., not necessarily positive) it holds that if $\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n$ is defined then $\lambda_1 \triangleright \lambda_2 \triangleright \ldots \triangleright \lambda_n$ is also defined. This follows from the fact that if there is a relationship of dominance between two distributions then the same relationships also holds between their respective marginal distributions. Therefore we are convinced that there must exist another proof which avoids application of Theorem 1.

In many cases Theorem 2 offers us a possibility to substantially reduce dimension of a considered compositional model in one step. Unfortunately, it gives us no instructions for how to find a set of indices Z (along with the index s) determining a reduction. Nevertheless, keeping in mind that in practical situations the process described in this assertion enables deleting of tens or hundreds of variables in one step, its realization will be discussed in the next section. For this, three simple lemmas will be useful. To formulate them in a transparent way we will use the following auxiliary symbol. Having a set $Z \subset \{1, \ldots, n\}$ and $j \notin Z$ the symbol $W(Z, j)$ denotes the following subset of indices:

$$W(Z,j) = \left\{ s \in \{1, \ldots, n\} : \left(\bigcup_{i \in Z} K_i \right) \cap K_j \subseteq K_s \right\}$$

(the reader should certainly keep in mind that sets $W(Z, j)$ depend not only on Z and j but naturally also on the considered generating sequence).

Lemma 7. If for Z ($\emptyset \neq Z \subsetneq \{1, \ldots, n\}$) there exists $s \in Z$, for which $s \in \bigcap_{i \notin Z} W(Z, i)$, then s and Z determine a reduction (of the considered generating sequence).

P r o o f. For s meeting the assumption of this Lemma $\left(\bigcup_{i \in Z} K_i\right) \cap K_j \subseteq K_s$ for all $j \notin Z$, and therefore $(\bigcup_{i \in Z} K_i) \cap (\bigcup_{i \notin Z} K_i) \subseteq K_s$. □

Lemma 8. If for $j \notin Z$ ($\emptyset \neq Z \subsetneq \{1, \ldots, n\}$) $W(Z, j) \cap Z = \emptyset$ then for any reduction (s, Z')

$$Z \subset Z' \implies W(Z, j) \cap Z' \neq \emptyset.$$

P r o o f. If $j \in Z'$ the assertion holds because j is always an element of $W(Z, j)$. If $j \notin Z' \supseteq Z$ then for a reduction (s, Z')

$$\left(\bigcup_{i \in Z} K_i \right) \cap K_j \subseteq \left(\bigcup_{i \in Z'} K_i \right) \cap \left(\bigcup_{i \notin Z'} K_i \right) \subseteq K_s,$$

and therefore $s \in W(Z, j) \cap Z'$. □

Lemma 9. Let for two nonempty disjoint subsets Z, Z_1 of $\{1,\dots,n\}$ there exist $s \in Z$ such that

$$\left(\bigcup_{j\in Z} K_j\right) \cap \left(\bigcup_{j\in Z_1} K_j\right) \subseteq K_s.$$

If for all pairs of indices $j_1 \in Z_1$, $j_2 \in Z_2 = \{1,\dots,n\} \setminus (Z \cup Z_1)$

$$K_{j_1} \cap K_{j_2} \subseteq \bigcup_{j\in Z} K_j,$$

then s and $Z \cup Z_2$ determine a reduction.

Proof. To show that s and $Z \cup Z_2$ determine a reduction we have just to show that

$$\left(\bigcup_{j\in Z_2} K_j\right) \cap \left(\bigcup_{j\in Z_1} K_j\right) \subseteq K_s.$$

We know that

$$K_{j_1} \cap K_{j_2} \cap \left(N \setminus \bigcup_{j\in Z} K_j\right) = \emptyset,$$

or equivalently

$$K_{j_1} \cap K_{j_2} \subseteq \left(\bigcup_{j\in Z} K_j\right),$$

for all $j_1 \in Z_1$, $j_2 \in Z_2$. From this one immediately gets

$$(\bigcup_{j\in Z_2} K_j) \cap \left(\bigcup_{j\in Z_1} K_j\right) \subseteq \left(\bigcup_{j\in Z} K_j\right),$$

and therefore also

$$\left(\bigcup_{j\in Z_2} K_j\right) \cap \left(\bigcup_{j\in Z_1} K_j\right) \subseteq \left(\bigcup_{j\in Z} K_j\right) \cap \left(\bigcup_{j\in Z_1} K_j\right) \subseteq K_s,$$

which finishes the proof. □

5. MARGINALIZATION PROCEDURE

In this section we will briefly formulate the main ideas of an efficient algorithm for marginalization of compositional models. The whole procedure is based on application of Lemma 1, Theorem 1 and its Corollary. However, our goal is to minimize use of Theorem 1. The whole process will be illustrated by an example in the following section.

1. If applicable, the simplest way of marginalization is a commutative employment of Lemma 1 and Corollary. The former makes it possible to delete one or several distributions from the generating sequence, the latter deletes variable(s) in a very efficient way. Therefore, we use them as the first step of the procedure, and then whenever the assumptions of one of these assertions are fulfilled. It is important to realize that application of Corollary decreases only the number of variables for which individual distributions are defined, while application of Lemma 1 may decrease both the number of variables and distributions. Therefore commutative application of these two assertions may be necessary.

2. When step 1 is no longer applicable, and there are still more than a few variables to be deleted, it is advisable to try to find a reduction and apply Theorem 2 (this option is discussed in more detail below). If this is successful, the idea of step 1 should be repeated.

3. If neither step 1 nor step 2 is applicable one should marginalize the resulting generating sequence using Theorem 1. After deletion of each variable one should try to apply procedures from step 1. (We do not discuss it in this paper but the order in which the variables are eliminated affects the time demands. To solve this problem efficiently one can, for example, take advantage of the heuristics proposed by R. Shachter for Bayesian network marginalization.)

The most computationally demanding part is step 2. Its implementation influences the total time requirements of the marginalization procedure. To find a reduction one can adapt one of the procedures proposed for decomposition of (hyper)graphs; see e. g. [1]. Here we present an alternative procedure based on Lemmas 7 − 9.

Consider a situation when we are to compute $(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)^{\downarrow M}$, and neither Lemma 1 nor Corollary can be applied. Let us start with a minimum Z for which

$$M \subseteq \left(\bigcup_{i \in Z} K_i \right),$$

and compute $W(Z, j)$ for all $j \notin Z$. Lemma 8 gives us instructions which indices $j \notin Z$ must be added to Z if one wants to have a chance to find a reduction (s, Z') for which $Z' \supseteq Z$. One has to consider all $j \notin Z$ for which $W(Z, j) \cap Z = \emptyset$. If $W(Z, j) = \{j\}$, then j must be added to set Z. If $W(Z, j)$ contains more indices, then we have to add (at least) one $k \in W(Z, j)$ (a heuristic recommendation saying which one should be selected may be a result of computational experiments).

If Lemma 8 cannot be applied to Z, i. e. $W(Z,j) \cap Z \neq \emptyset$ for all $j \notin Z$, then there are two possibilities. The best case occurs when there exists $s \in \left(\bigcap_{j \notin Z} W(Z,j) \right) \cap Z$; then s and Z determine a reduction due to Lemma 7 and the process is finished; no other reduction exists.

When neither Lemma 7 nor 8 can be applied there is still a chance that Lemma 9 helps us. Using the notation of this Lemma, it is obvious that if $j, k \notin Z$ are such that

$$K_j \cap K_k \not\subseteq \bigcup_{i \in Z} K_i,$$

then both j, k must be either in Z_1 or in Z_2. Therefore we recommend partitioning set $\{1, \ldots, n\} \setminus Z$ into subsets $\hat{Z}_1, \ldots, \hat{Z}_t$ in the way that two indices j_1, j_2 are in the same $\hat{Z}_i$ if and only if there exists an "outer connection" of K_{j_1} with K_{j_2}; i. e., a sequence of indices $k_1 = j_1, k_2, k_3, \ldots, k_r = j_2$ such that

$$K_{k_\ell} \cap K_{k_{\ell+1}} \not\subseteq \bigcup_{i \in Z} K_i,$$

for all $\ell = 1, \ldots, r-1$. (Notice that since the relation of having an outer connection is reflexive, symmetric and transitive, $\hat{Z}_1, \ldots, \hat{Z}_t$ is a partition of $\{1, \ldots, n\} \setminus Z$.) For each $\hat{Z}_i$ one of the following three situations becomes effective.

(a) There exists $s_i \in \bigcap_{j \in \hat{Z}_i} W(Z, j)$. In this case s_i and $\{1, \ldots, n\} \setminus \hat{Z}_i$ determine a reduction due to Lemma 9. Realization of marginalization according to Lemma 7 does not influence the other sets from the partition $\hat{Z}_1, \ldots, \hat{Z}_t$.

(b) The other possibility is that there are two indices $j_1, j_2 \in \hat{Z}_i$, for which $W(Z, j_1) \cap W(Z, j_2) = \emptyset$. It means that there does not exist Z_1 with the properties required by Lemma 9 containing both j_1 and j_2. Therefore, both these indices, along with the indices of the shortest "outer connection" must be added to Z. Then (after processing all other sets from the considered partition) one should go back in the process and again test whether Lemma 8 does not recommend further extension of Z.

(c) Though it is rather improbable, in very special situations it may happen that neither of the above two situations is applicable (an example of such a situation is shown in Figure 1 – here Z contains the three distributions which are defined for the three "inner" variables). In this case one should add all those $j \in \hat{Z}_i$ to Z for which $W(Z, j) \cap \hat{Z}_i \subseteq \{k \in \hat{Z}_i : W(Z, j) \supseteq W(Z, k)\}$ (in the example these are the distributions "connecting" the "inner" triangle with the "outer" circle). Also in this case, after processing all other sets from the considered partition one should go back in the process and again test whether Lemma 8 does not recommend further extension of Z.

6. EXAMPLE

Let us consider distributions $\kappa_1, \kappa_2, \ldots, \kappa_{14}$ with the following sets of variables (as shown in Figure 2):

$$
\begin{array}{lll}
K_1 = \{1,3,7,8\}, & K_2 = \{1,2\}, & K_3 = \{3,4\}, \\
K_4 = \{4,5\}, & K_5 = \{5,6,7\}, & K_6 = \{8,9\}, \\
K_7 = \{9,10,11,12\}, & K_8 = \{12,13,19\}, & K_9 = \{10,14\}, \\
K_{10} = \{14,15,16\}, & K_{11} = \{16,17\}, & K_{12} = \{17,18\}, \\
K_{13} = \{11,18,19,20\}, & K_{14} = \{20,21\}. &
\end{array}
$$

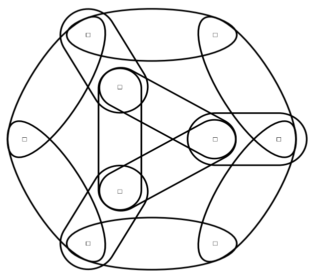

Fig. 1. Case (c) of marginalization procedure.

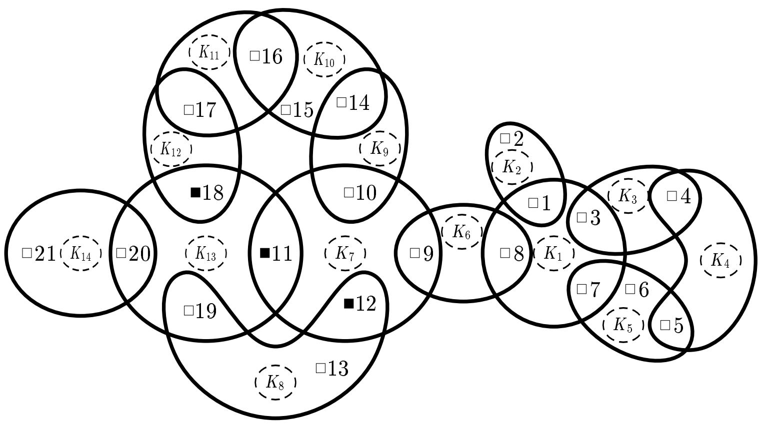

Fig. 2. Sets of variables, for which distributions $\kappa_1, \kappa_2, \ldots, \kappa_{14}$ are defined.

Those distributions, as a generating sequence, define a multidimensional compositional model $\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_{14}$ and assume that our goal is to compute

$$(\kappa_1 \triangleright \ldots \triangleright \kappa_{14})^{\downarrow\{11,12,18\}}.$$

We start with deletion of distribution κ_{14}; this is enabled by Lemma 1. Then, all the variables appearing only in one distribution may be marginalized out using Corollary. So, we get

$$\begin{aligned}(\kappa_1 \triangleright \ldots \triangleright \kappa_{14})^{-\{2,6,13,15,20,21\}} = \kappa_1 \triangleright \kappa_2^{-\{2\}} \triangleright \kappa_3 \triangleright \kappa_4 \triangleright \kappa_5^{-\{6\}} \triangleright \kappa_6 \triangleright \kappa_7 \\ \triangleright \kappa_8^{-\{13\}} \triangleright \kappa_9 \triangleright \kappa_{10}^{-\{15\}} \triangleright \kappa_{11} \triangleright \kappa_{12} \triangleright \kappa_{13}^{-\{20\}}.\end{aligned}$$

Now, Lemma 1 can be applied again to this expression; we may omit distribution $\kappa_2^{-\{2\}} = \kappa_2^{\downarrow\{1\}}$. (Actually, we do not need to calculate marginal $\kappa_2^{-\{2\}}$, instead we may simply leave κ_2 out.)

After this simplification we can also see that variable X_1 appears only among the arguments of one distribution (κ_1) and Corollary may be used once more. In this way we get the following simplified model

$$(\kappa_1 \triangleright \ldots \triangleright \kappa_{14})^{-\{1,2,6,13,15,20,21\}} = \kappa_1^{-\{1\}} \triangleright \kappa_3 \triangleright \kappa_4 \triangleright \kappa_5^{-\{6\}} \triangleright \kappa_6 \triangleright \kappa_7 \triangleright \kappa_8^{-\{13\}} \triangleright \kappa_9 \triangleright \kappa_{10}^{-\{15\}} \triangleright \kappa_{11} \triangleright \kappa_{12} \triangleright \kappa_{13}^{-\{20\}}.$$

Further application of neither Lemma 1 nor Corollary is possible. So we have to start considering application of Theorem 2 as recommended in step 2 of the previous section.

Table. Searching for a reduction.

Z	j	$W(Z,j)\cap Z$	j	$W(Z,j)\cap Z$
$7,13$	1	$7,13$	3	$7,13$
	4	$7,13$	5	$7,13$
	6	7	8	
	9	7	10	$7,13$
	11	$7,13$	12	13
$7,8,13$	1	$7,8,13$	3	$7,8,13$
	4	$7,8,13$	5	$7,8,13$
	6	7	9	7
	10	$7,8,13$	11	$7,8,13$
	12	13		

The task is to find a marginal distribution for variables X_{11}, X_{12}, X_{18}; therefore we start considering the smallest set of distributions covering these variables: $Z = \{7,13\}$. When computing all $W(Z,j)$ (see Table, where $W(Z,j)\cap Z$ is depicted) one can see that $W(Z,8) = \{8\}$. Therefore Lemma 8 may be applied with the conclusion that to find a reduction, Z must be extended by (at least) index 8. For $Z = \{7,8,13\}$ all $W(Z,j)$ (for $j \notin Z$) contains at least one index from Z. So, in accordance with the recommendation described in Section 5, we find a partition of $\{j : j \notin Z\}$ into subsets containing indices of distributions for which there exists an "outer connection":

$$\hat{Z}_1 = \{1,3,4,5,6\}, \quad \hat{Z}_2 = \{9,10,11,12\}.$$

Since $7 \in \bigcap_{j\in\hat{Z}_1} W(Z,j)$, Lemma 9 says that 7 and $\{7,8,9,10,11,12,13\}$ determine a reduction, and therefore application of Theorem 2 yields

$$\begin{aligned}(\kappa_1 \triangleright \ldots \triangleright \kappa_{14})^{-\{1,2,3,4,5,6,7,8,13,15,20,21\}} &= (\kappa_1 \triangleright \ldots \triangleright \kappa_{14})^{\downarrow\{9,10,11,12,14,16,17,18,19\}} \\ &= \mu^{\downarrow\{9\}} \triangleright \mu^{\downarrow\{9\}} \triangleright \mu^{\downarrow\{9\}} \triangleright \mu^{\downarrow\{9\}} \triangleright \mu^{\downarrow\{9\}} \triangleright \kappa_7 \triangleright \kappa_8^{-\{13\}} \triangleright \kappa_9 \\ &\qquad \triangleright \kappa_{10}^{-\{15\}} \triangleright \kappa_{11} \triangleright \kappa_{12} \triangleright \kappa_{13}^{-\{20\}} \\ &= \mu^{\downarrow\{9\}} \triangleright \kappa_7 \triangleright \kappa_8^{\downarrow\{12,19\}} \triangleright \kappa_9 \triangleright \kappa_{10}^{\downarrow\{14,16\}} \triangleright \kappa_{11} \triangleright \kappa_{12} \triangleright \kappa_{13}^{\downarrow\{11,18,19\}},\end{aligned}$$

where

$$\mu^{\downarrow\{9\}} = (\kappa_1 \triangleright \kappa_3 \triangleright \kappa_4 \triangleright \kappa_5 \triangleright \kappa_6)^{\downarrow\{9\}}.$$

For $\hat{Z}_2$ the set $\bigcap_{j\in\hat{Z}_2} W(Z,j)$ is empty and therefore, in accordance with point (b), we are checking whether there are two indices $j_1, j_2 \in \hat{Z}_2$, for which $W(Z,j_1) \cap W(Z,j_2) = \emptyset$. This property holds for $j_1 = 9, j_2 = 12$, and point (b) recommends[2] us to add all the indices from $\hat{Z}_2$ to Z. Thus we see that there does not exist another reduction and therefore to proceed further we have to start applying Theorem 1.

Let us apply Theorem 1 to delete variable X_9:

$$
\begin{aligned}
&(\kappa_1 \triangleright \ldots \triangleright \kappa_{14})^{\downarrow\{10,11,12,14,16,17,18,19\}} \\
&\quad = \mu^{\downarrow\{\emptyset\}} \triangleright (\mu^{\downarrow\{9\}} \circledtriangleright_{\emptyset} \kappa_7)^{-9} \triangleright \kappa_8^{\downarrow\{12,19\}} \triangleright \kappa_9 \triangleright \kappa_{10}^{\downarrow\{14,16\}} \triangleright \kappa_{11} \triangleright \kappa_{12} \triangleright \kappa_{13}^{\downarrow\{11,18,19\}} \\
&\qquad = (\mu^{\downarrow\{9\}} \circledtriangleright_{\emptyset} \kappa_7)^{-9} \triangleright \kappa_8^{\downarrow\{12,19\}} \triangleright \kappa_9 \triangleright \kappa_{10}^{\downarrow\{14,16\}} \triangleright \kappa_{11} \triangleright \kappa_{12} \triangleright \kappa_{13}^{\downarrow\{11,18,19\}}
\end{aligned}
$$

Let us show how to marginalize yet another variable out, for example, X_{16}. The rest will be left to the reader (as a rather trivial repetition of the described process).

$$
\begin{aligned}
&(\kappa_1 \triangleright \ldots \triangleright \kappa_{14})^{\downarrow\{10,11,12,14,17,18,19\}} = \\
&= (\mu^{\downarrow\{9\}} \circledtriangleright_{\emptyset} \kappa_7)^{-9} \triangleright \kappa_8^{\downarrow\{12,19\}} \triangleright \kappa_9 \triangleright \kappa_{10}^{\downarrow\{14\}} \triangleright \left(\kappa_{10}^{\downarrow\{14,16\}} \circledtriangleright_{\{10,11,12,14,19\}} \kappa_{11}\right)^{-\{16\}} \\
&\qquad \triangleright \kappa_{12} \triangleright \kappa_{13}^{\downarrow\{11,18,19\}} \\
&\qquad = \kappa_1 \triangleright \kappa_8^{\downarrow\{12,19\}} \triangleright \kappa_9 \triangleright \kappa_2 \triangleright \kappa_{12} \triangleright \kappa_{13}^{\downarrow\{11,18,19\}},
\end{aligned}
$$

where

$$
\begin{aligned}
\kappa_1(x_{10}, x_{11}, x_{12}) &= (\mu^{\downarrow\{9\}} \circledtriangleright_{\emptyset} \kappa_7)^{-9} \\
\kappa_2(x_{14,17}) &= \left(\kappa_{10}^{\downarrow\{14,16\}} \circledtriangleright_{\{10,11,12,14,19\}} \kappa_{11}\right)^{-\{16\}} = \left(\kappa_{10}^{\downarrow\{14,16\}} \triangleright \kappa_{11}\right)^{-\{16\}}.
\end{aligned}
$$

Let us mention that deletion of $\kappa_{10}^{\downarrow\{14\}}$ from the generating sequence was enabled by Lemma 1.

7. CONCLUSION

We introduced a theoretical background for efficient marginalization of multidimensional compositional models. The efficiency of our approach stems from a possibility to eliminate a set of variables in one computationally simple step. To find a required reduction we introduced a technique based on properties of sets $W(Z,j)$, which leads to a straightforward, algorithmically simple procedure.

Analysis of computational complexity as well as implementation of the whole process, which can take advantage of published algorithms for testing acyclicity of hypergraphs, still remain to be done. In any case, practical implementation of the process is a challenge for a sophisticated application of a number of heuristic steps controlling specific situations, in which, for example, direct marginalization by multiple application of Theorem 1 may be faster than searching for a reduction that would enable a deletion of only a small number of variables.

[2] Point (b) recommends to add the shortest "outer connection" of j_1, j_2 to Z, but in the considered situation it is the whole $\hat{Z}_2$.

ACKNOWLEDGEMENT

The authors want to express their sincere gratitude to both anonymous referees for their valuable comments and suggestions leading to a substantial improvement of the paper. This research was partially supported by the Grant Agency of the Academy of Sciences of the Czech Republic under grant A 2075302, by the University of Economics in Prague under grant IGA 611025 and by the Ministry of Education, Youth and Sports of the Czech Republic under project 1M0572.

(Received December 9, 2005.)

REFERENCES

[1] J. H. Badsberg: An Environment for Graphical Models. Ph.D. Thesis, Aalborg University 1995. http://www.math.aau.dk/~jhb/CoCo/documentation.
[2] Hugin Explorer, ver 6.0, Computer software, 2005. http://www.hugin.com
[3] F. V. Jensen: Bayesian Networks and Decision Graphs. Springer Verlag, New York 2001.
[4] R. Jiroušek: Marginalization in composed probabilistic models. In: Proc. 16th Conf. Uncertainty in Artificial Intelligence UAI'00 (C. Boutilier and M. Goldszmidt, eds.), Morgan Kaufmann, San Francisco 2000, pp. 301–308.
[5] R. Jiroušek: Decomposition of multidimensional distributions represented by perfect sequences. Ann. Math. and Artificial Intelligence *35* (2002), 215–226.
[6] R. Jiroušek: What is the difference between Bayesian networks and compositional models? In: Proc. 7th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty (H. Noguchi, H. Ishii, M. Inuiguchi, eds.), Awaji Yumebutai ICC 2004, pp. 191–196.
[7] S. L. Lauritzen: Graphical Models. Clarendon Press, Oxford 1996.
[8] R. D. Shachter: Evaluating influence diagrams. Oper. Res. *34* (1986), 871–890.
[9] R. D. Shachter: Probabilistic inference and influence diagrams. Oper. Res. *36* (1988), 589–604.
[10] G. Shafer: Probabilistic Expert Systems. SIAM, Philadelphia 1996.

Vladislav Bína, Faculty of Management, University of Economics, Jarošovská 1117/II, 377 01 Jindřichův Hradec. Czech Republic.

Radim Jiroušek, Institute of Information Theory and Automation — Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 09 Praha 8. Czech Republic.

e-mails: bina@fm.vse.cz, radim@utia.cas.cz