About an Effective Algorithm for Marginalization in Multidimensional Compositional Models

Vladislav Bína
Fac. of Management, Jindřichův Hradec
University of Economics, Prague
bina@fm.vse.cz

Radim Jiroušek
Inst. of Information Th. and Automat.
Academy of Sciences of the Czech Rep.
radim@utia.cas.cz

Abstract

1 Introduction

Representation and processing of multidimensional probability distributions were made possible by a succes achieved in the field of graphical Markov models (see e.g. [5]) during last about twenty years. Here we have in mind not only ample theoretical background but also thoroughly elaborated algorithmical apparatus, which enabled developing extremely efficient software packages like, for example, HUGIN [**?**]. As an alternative to graphical models, we have been elaborating (during last about eight years) 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 the presented paper we propose a solution of one hard problem, which has not been not solved even in such software systems like HUGIN: the problem of marginalization of multidimensional distribution. For Bayesian networks a solution of this problem was proposed by Ross Shachter in [6, 7]. His famous procedure is based on two rules: *node deletion* and *edge reversal*. Roughly speaking, the effectivity of his appraoch corresponds to the effectivity of the presented process in case we did not employ the speed-up theoretically supported by Theorem 3 presented in Section 4 of this paper. This theorem, namely, takes advantage of the main difference between Bayesian networks [1] and compositional models revealed in [4]. This advantage consists in the fact that compositional models have some marginal distributions, whose computation in Bayesian network may be computationally expensive, expressed explicitly.

2 Notation

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 $K \subset N$, $\pi(x_K)$ denotes a distribution of variables $\{X_i\}_{i\in K}$, which is defined on all subset of a Cartesian product $\mathbf{X}_K \stackrel{\text{def}}{=} \times_{i\in K}\mathbf{X}_i$.

Having a distribution $\pi(x_K)$ and $L \subset K$, we will denote its corresponding marginal distribution either $\pi(x_L)$, or, using the notation introduced by Glenn Shafer [8], $\pi^{\downarrow L}$. 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 deleted in the process of marginalization, we will use the symbol π^{-M}, where M is a set of indices of the variables, which do not appear among the arguments of the resulting marginal distribution. Thus, in our case, M is any set, for which $K \setminus M = L$.

Most of the time we will consider sequences of distributions. To shorten the notation, for an integer n, the set of all positive integers lower or equal to n will be denoted by $\hat{n} \stackrel{\text{def}}{=} \{1, 2, \ldots, n\}$.

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 arbitrary two distributions $\pi(x_K)$ and $\kappa(x_L)$ their *composition* is given by the formula

$$\pi(x_K) \triangleright \kappa(x_L) = \begin{cases} \frac{\pi(x_K)\kappa(x_L)}{\kappa(x_{K\cap L})} & \text{when} \quad \pi(x_{K\cap L}) \ll \kappa(x_{K\cap L}), \\ \text{undefined} & \text{otherwise}, \end{cases}$$

where the symbol $\pi(x_M) \ll \kappa(x_M)$ denotes that $\pi(x_M)$ is *dominated* by $\kappa(x_M)$, which means (in the considered finite setting)

$$\forall x_M \in \mathbf{X}_M \quad (\kappa(x_M) = 0 \implies \pi(x_M) = 0).$$

Since the outcome of the composition is a new distribution, we can iteratively repeat the application of this operator composing thus 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 $\pi_1, \pi_2, \ldots, \pi_n$, we will refer to it as to a *generating sequence*, to which the operator is applied from left to right:

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \ldots \triangleright \pi_{n-1} \triangleright \pi_n := (\ldots((\pi_1 \triangleright \pi_2) \triangleright \pi_3) \triangleright \ldots \triangleright \pi_{n-1}) \triangleright \pi_n.$$

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

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

3 Basic properties

In the following text we will need two simple lemmata which follow from the definition of the operator of composition (their proofs can also be found in out previous papers).

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

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

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

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

Let us emphasize that when describing a generating sequence it was necessary to explain that the operator of composition is always applied from left to right, since the operator is neither commutative nor associative. So, generally

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 \neq \pi_1 \triangleright (\pi_2 \triangleright \pi_3),$$
$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 \neq \pi_1 \triangleright \pi_3 \triangleright \pi_2.$$

This was also the reason why we introduced the anticipating operator $\circledtriangleright_K$. Namely, this operator allows us to change the ordering of compositions in the sense described in the following assertion (for its proof see [2]).

Lemma 3 *If $\pi_1(x_{K_1})$, $\pi_2(x_{K_2})$ and $\pi_3(x_{K_3})$ are such that the composition $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is defined then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = (\pi_1 \triangleright \pi_2) \triangleright \pi_3 = \pi_1 \triangleright (\pi_2 \circledtriangleright_{K_1} \pi_3).$$

4 Marginalization in compositional models

Now we will focus our attention on possibilities of marginalization of distributions given by generating sequences. Let us stress that (in a general case) marginal distribution of a compositional model is not a distrubution represented by a sequence of marginalized distributions. The exact meaning of this sentence will be clear from Theorem 2.

From now on, we will consider generating sequences

$$\pi_1(x_{K_1}) \triangleright \pi_2(x_{K_2}) \triangleright \ldots \pi_n(x_{K_n}).$$

Therefore whenever we use distribution π_j, we assume it is defined for variables $\{X_i\}_{i\in K_j}$.

First, we will formulate 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. First let us formulate very simple but useful assertion. Its proof, as well as the proof of Theorem 2 can be found in [2].

Theorem 1 *Let $\pi_1, \pi_2, \ldots, \pi_n$ be a generating sequence. If $\ell \in K_i$ for some $i \in \hat{n}$ and $\ell \notin K_j$ for all $j \in (\hat{n} \setminus \{i\})$ then the marginal of the distribution represented by the generating sequence may be easily got acording to the following simple formula:*

$$(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{-\{\ell\}} = \pi_1 \triangleright \ldots \triangleright \pi_{i-1} \triangleright \pi_i^{-\{\ell\}} \triangleright \pi_{i+1} \triangleright \ldots \triangleright \pi_n.$$

Hence, when the variable which is to be deleted is contained in an argument of only one of the distributions, it is sufficient to marginalize only this one distribution. The others remain unchanged. The reader familiar with the Shachter's marginalizing procedure [6, 7] certainly noticed, that Theorem 1 describes situations when his *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.

For general situations when marginalized variable is among arguments of more than one distribution, the following rather complicated theorem must be used.

Theorem 2 (Marginalization over one variable) *Let $\pi_1, \pi_2, \ldots, \pi_n$ be a generating sequence and*

$$\ell \in K_{i_1} \cap K_{i_2} \cap \ldots \cap K_{i_m}$$

for a subsequence $(i_1, i_2, \ldots, i_m)$ of $\hat{n}$ such that $\ell \notin K_j$ for all $j \in \hat{n} \setminus \{i_1, i_2, \ldots, i_m\}$. Then

$$(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{-\{\ell\}} = \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n,$$

where

$$\begin{aligned}
\kappa_j &= \pi_j, \qquad \forall j \in \hat{n} \setminus \{i_1, i_2, \ldots, i_m\},\\
\kappa_{i_1} &= \pi_{i_1}^{-\{\ell\}},\\
\kappa_{i_2} &= (\pi_{i_1} \circledtriangleright_{L_{i_2-1}} \pi_{i_2})^{-\{\ell\}},\\
\kappa_{i_3} &= (\pi_{i_1} \circledtriangleright_{L_{i_2-1}} \pi_{i_2} \circledtriangleright_{L_{i_3-1}} \pi_{i_3})^{-\{\ell\}},\\
&\vdots\\
\kappa_{i_m} &= (\pi_{i_1} \circledtriangleright_{L_{i_2-1}} \pi_{i_2} \circledtriangleright_{L_{i_3-1}} \cdots \circledtriangleright_{L_{i_m-1}} \pi_{i_m})^{-\{\ell\}},
\end{aligned}$$

and $L_{i_k-1} = (K_1 \cup K_2 \cup \ldots \cup K_{i_k-1}) \setminus \{\ell\}$.

Iterative application of this Theorem always leads to the desired marginal distribution and fully corresponds to the Shachter's marginalization procedure. In fact, application of the aniticipating operator somehow corresponds to the *inheritance of parents* in his *edge reversal rule*. So one cannot be surprised that the computational complexity of this process strongly depends on the number of occurences 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 deleted variables). Beginning from the second occurence of this variable we should replace distribution π_{i_k} by an expression containing one or more anticipating operators, and, in addition to it, this expression still has to be marginalized. Thus it may easily happen that iterative application of this theorem becomes computationally intractable due to its enormous time and memory consumption.

Most effective marginalizing procedures are based on the following (unfortunately also rather complex) assertion, which is a generalization of Theorem 11 from [3]. It describes conditions, under which a number of variables may be deleted in one, computationally simple step.

Theorem 3 *Let $\pi_1, \pi_2, \ldots, \pi_n$ be a generating sequence and $(j_1, j_2, \ldots, j_m)$ be a subsequence of $\hat{n}$ such that there exists $s \in Z = \{j_1, \ldots, j_m\}$, for which*

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

Then, denoting $L = \bigcup_{j \in Z} K_j$, $\mu = (\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{\downarrow K_s}$, and for all $j \notin Z$

$$\bar{L}_j = \bigcup_{i \in \hat{j} \setminus Z} K_i,$$

marginal distribution $(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{\downarrow L}$ can be expressed as a compositional model

$$(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{\downarrow L} = \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n,$$

where

$$\begin{array}{lll} \kappa_j = \pi_j & \text{for} & j \in Z, \\ \kappa_j = \mu^{\downarrow L \cap \bar{L}_j} & \text{for} & j \notin Z. \end{array}$$

Proof Let $\{\ell_1, \ell_2, \ldots, \ell_m\} = (K_1 \cup \ldots \cup K_n) \setminus L$ be any ordering of indices to be eliminated. Let $\nu_1^1, \nu_2^1, \ldots, \nu_n^1$ be a generating sequence received by application of Theorem 2 to the sequence $\pi_1, \pi_2, \ldots, \pi_n$ and the index ℓ_1. What can be said about the generating sequence $\nu_1^1, \nu_2^1, \ldots, \nu_n^1$?

1. $(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{-\{\ell_1\}} = (\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{\downarrow L \cup \{\ell_2, \ldots, \ell_m\}} = \nu_1^1 \triangleright \nu_2^1 \triangleright \ldots \triangleright \nu_n^1$;
2. For all $j \in Z$, $\nu_j^1 = \pi_j$;

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 $\bar{L}_j$ but not ℓ_1. Therefore, the respective set of indices contains $K_j \setminus \{\ell_1\}$ and is contained in $\bar{L}_j \setminus \{\ell_1\}$.

Now, iterative application of Theorem 2 to the generating sequences ν_1^i, $\nu_2^i, \ldots \nu_n^i$ and the indices ℓ_{i+1} yields sequences $\nu_1^{i+1}, \nu_2^{i+1}, \ldots, \nu_n^{i+1}$ for $i = 1, \ldots, m-1$. Analogous to the first step, we can see that for all of these sequences:

1. $(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{-\{\ell_1,\ldots,\ell_i\}} = (\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{\downarrow L \cup \{\ell_{i+1},\ldots,\ell_m\}} = \nu_1^i \triangleright \nu_2^i \triangleright \ldots \triangleright \nu_n^i$;

2. For all $j \in Z$, $\nu_j^i = \pi_j$;

3. For each $j \notin Z$, ν_j^i is a distribution of variables $X_{K_j \setminus \{\ell_1,\ldots,\ell_i\}}$ and possibly some other variables from $X_{\bar{L}_j \setminus \{\ell_1,\ldots,\ell_i\}}$.

Therefore, $(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{\downarrow L} = \nu_1^m \triangleright \nu_2^m \triangleright \ldots \triangleright \nu_n^m$ and to finish the proof we have to show that we can transform the sequence $\nu_1^m, \ldots, \nu_n^m$ into the required sequence $\kappa_1, \ldots, \kappa_n$ without changing the generated multidimensional distribution.

The elements with indices $j_1, \ldots, j_n$ need no change, as

$$\pi_j = \nu_j^m = \kappa_j,$$

for all $j \in Z$. Therefore, what has remained to be shown is that substituting ν_j^m with κ_j (for $j \notin Z$) does not change the generated distribution.

Denote by L_j (for all $j = 1, 2, \ldots, n$) the sets of indices of variables for which the distributions ν_j^m are defined. Clearly, for $j \in Z$, $L_j = K_j$. For $j \notin Z$, we have shown above that

$$K_j \setminus \{\ell_1, \ldots, \ell_m\} \subseteq L_j \subseteq \bar{L}_j \setminus \{\ell_1, \ldots, \ell_m\} \subseteq \bar{L}_j \cap L \subseteq K_s,$$

(the last inclusion follows from the theorem assumptions) and therefore (using $K_j \setminus \{\ell_1, \ldots, \ell_m\} = K_j \cap L$)

$$(L_1 \cup L_2 \cup \ldots \cup L_{j-1}) \cup L_j \supseteq \bar{L}_j \cap L \cap K_s \supseteq L_j.$$

This enables us to apply Lemma 2, getting

$$(\nu_1^m \triangleright \nu_2^m \triangleright \ldots \triangleright \nu_{j-1}^m) \triangleright \nu_j^m = \\ = (\nu_1^m \triangleright \nu_2^m \triangleright \ldots \triangleright \nu_{j-1}^m) \triangleright (\nu_1^m \triangleright \ldots \triangleright \nu_j^m)^{\downarrow L \cap K_s \cap \bar{L}_j}.$$

Since both $(\nu_1^m \triangleright \ldots \triangleright \nu_j^m)$ and μ are marginal distributions of $\pi_1 \triangleright \ldots \triangleright \pi_n$, their common marginals must equal each other:

$$(\nu_1^m \triangleright \ldots \triangleright \nu_j^m)^{\downarrow L \cap K_s \cap \bar{L}_j} = \mu^{\downarrow L \cap \bar{L}_j} = \kappa_j,$$

and therefore

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

Repeating this considerations for all $j \notin Z$, one can substitute ν_j^m by κ_j for all $j \notin Z$, which finishes the proof. ∎

The last Theorem offers us a possibility to substatially reduce a dimension of a considered compositional model in one step. Unfortunately, it gives us no instructions how to find a set of indices Z (along with the index s) meeting the necessary assumptions required for application of the Theorem. For this, the following two simple lemmata will be useful. To formulate them in a transparent way we will use the following auxiliary symbol. Having a set $Z \subset \hat{n}$ and $j \notin Z$ the symbol $W(Z,j)$ denotes the following subset of indices:

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

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

Lemma 4 *If for $Z \subset \hat{n}$ ($\emptyset \neq Z \neq \hat{n}$) there exists $s \in Z$, for which $s \in \bigcap_{j \notin Z} W(Z,j)$, then s and Z meet all the assuptions of Theorem 3.*

Proof. For s meeting the assuption of this Lemma

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

for all $j \notin Z$, and therefore

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

∎

Lemma 5 *Let nonempty $Z \subset \hat{n}$ be different from $\hat{n}$. If for some $j \notin Z$, $W(Z,j) = \{j\}$ than there does not exists $s \in Z$, such that s and Z meet the assuptions of Theorem 3.*

Proof. $W(Z,j) = \{j\}$ means that $W(Z,j) \cap Z = \emptyset$. So it also means that

$$\left(\bigcup_{i \in Z} K_i \right) \cap K_j$$

is not contained in any K_s for $s \in Z$, and therefore there cannot exist $s \in Z$ containing

$$(\bigcup_{i \in Z} K_i) \cap (\bigcup_{i \notin Z} K_i),$$

because $j \notin Z$. ∎

5 Marginalization algorithm

In this section we will briefly formulate the main ideas of an effective algorithm for marginalization of compositional models. The algorithm is based on application of Lemma 1 and Theorems 1 – 3. Our goal is to minimize use of Theorem 2. 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 Theorem 1. Therefore, we use it 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, due to Lemma 1, some of the distributions may be deleted after application of Theorem 1, and therefore their commutative repetition is reasonable.

2. When the idea of step 1 is not applicable, we will try to apply Theorem 3 (this possibility is discussed below in more details). In case of a success we will continue again with step 1.

3. If neither step 1 nor step 2 is applicable we will marginalize the resulting generating sequence using iteratively Theorem 2.

From the point of view of effectivity of the marginalizing procedure, the most influencial is a sofisticated realization of step 2. It is important to realize that Lemmata 4 and 5 offer us a basis for a much more efficient procedure than testing all possible subsets $Z \subset \hat{n}$. If such Z exists (along with the corresponding $s \in Z$) , it can always be found with the help of the process we shall now briefly describe.

Consider a situation when we are to compute

$$(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{\downarrow M},$$

and let

$$Z = \{j \in \hat{n} : K_j \cap M \neq \emptyset\}\,.$$

We start with computing $W(Z, j)$ for all $j \notin Z$. As a rule, we cannot expect that there would be

$$s \in Z \cap \left(\bigcap_{j \notin Z} W(Z, j) \right)$$

(in such a case we would have got, due to Lemma 4, a required solution). First we have to add to Z (due to Lemma 5) all the indices $j \notin Z$, for which $W(Z, j) = \{j\}$. With the new Z we should proceed as before: compute $W(Z, j)$ for all $j \notin Z$ and add those $j \notin Z$ to Z, for which $W(Z, j) = \{j\}$.

When there does not exists $j \notin Z$, for which $W(Z, j) = \{j\}$ we start looking for $s \in Z$, for which Theorem 3 could be applied. Now, we can again ask whether there exists

$$s \in Z \cap \left(\bigcap_{j \notin Z} W(Z,j) \right).$$

In positive case we found a way, how to apply Theorem 3. In opposite case Z will be be increased. We take $s \notin Z$ (preferably such that[1] $K_s \cap L$ is the largest possible), add it to Z and find new $W(Z,j)$ sets. Then add to Z all $j \notin Z$, for which $s \notin W(Z,j)$. Repeating incremental enlargening of Z (not changing s) will finish either with a couple s and Z, for which Theorem 3 is applicable, or, geting $Z = \hat{n}$ we learn that s must be added to Z. This step may be repeated with the original Z increased by the previous s and a new $s \in Z$.

6 Example

Let us consider distributions $\pi_1, \pi_2, \ldots, \pi_{13}$ with corresponding sets of variables (as shown in Figure 6)

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

They define a generating sequence

$$\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_{13}.$$

Our goal is to compute

$$(\pi_1 \triangleright \ldots \triangleright \pi_{13})^{\downarrow\{2,3,5\}}.$$

First, deletion of distribution π_{13} is enabled by Lemma 1. Now, all the variables appearing only in one distribution may be marginalized out using Theorem 1. So,

$$\begin{aligned} &(\pi_1 \triangleright \ldots \triangleright \pi_{13})^{-\{1,7,9,14,17,20\}} \\ &\quad = \pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4^{-\{9\}} \triangleright \pi_5 \triangleright \pi_6^{-\{1\}} \triangleright \pi_7^{-\{14\}} \triangleright \pi_8 \triangleright \pi_9^{-\{17\}} \triangleright \pi_{10} \triangleright \pi_{11} \triangleright \pi_{12}^{-\{7\}}. \end{aligned}$$

Looking at this distribution we immediately see that distribution $\pi_9^{-\{17\}} = \pi_9^{\downarrow\{19\}}$ may be ommitted because of Lemma 1. In fact, we actually do not need to calculate marginal $\pi_9^{-\{17\}}$ and may just leave π_9 out.

After this simplification we can see that also variable X_{16} appears among the arguments of only one distribution and Lemma 1 may be used once more

$$\begin{aligned} &(\pi_1 \triangleright \ldots \triangleright \pi_{13})^{-\{1,7,9,14,16,17,20\}} \\ &\quad = \pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4^{-\{9\}} \triangleright \pi_5 \triangleright \pi_6^{-\{1\}} \triangleright \pi_7^{-\{14\}} \triangleright \pi_8^{-\{16\}} \triangleright \pi_{10} \triangleright \pi_{11} \triangleright \pi_{12}^{-\{7\}}. \end{aligned}$$

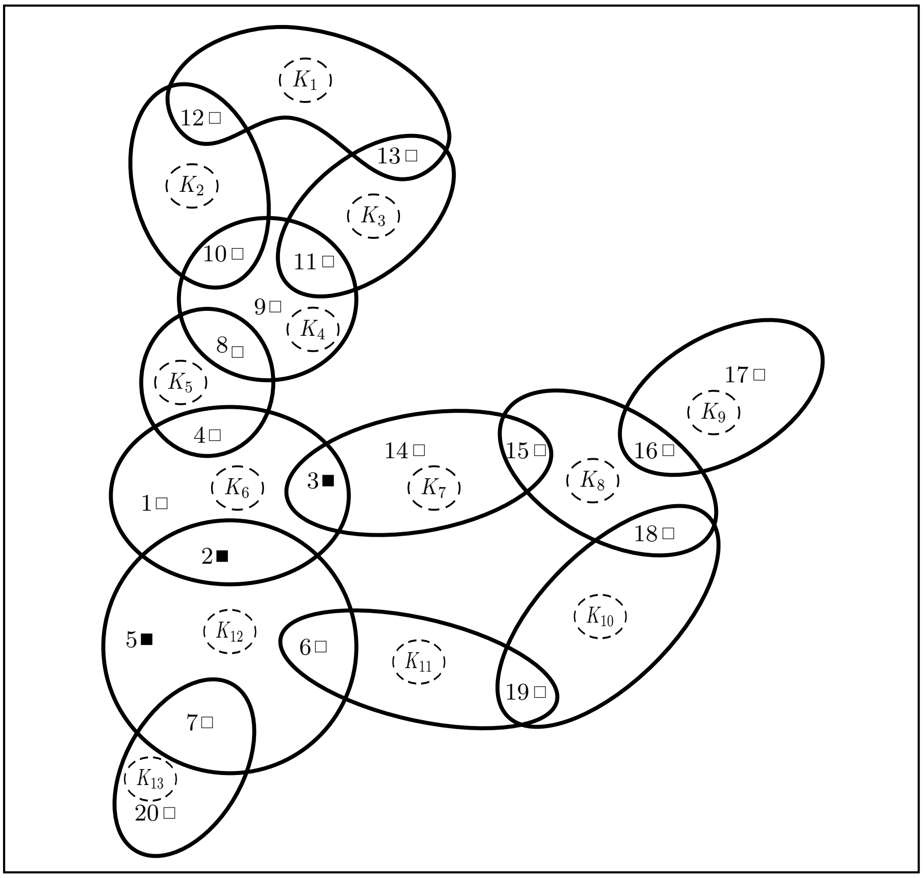

Figure 1: Sets of variables, for which distributions $\pi_1, \pi_2, \ldots, \pi_{13}$ are defined
.

Now, we start applying ideas from step 2 to find out whether Theorem 3 may be applied. This process is summarized in Table 1. We start with $Z = \{6, 12\}$. All $W(Z, j) \neq \{j\}$, and therefore we do not apply Lemma 5. Since $\bigcap_{j \notin Z} W(Z, j) = \emptyset$, Theorem 3 cannot be applied to Z. It means that we have to start enlarging this set, i.e. we start considering $s \notin Z$. For the first choise 3 indices come into consideration: $5, 7, 11$. Let us choose 7. Therefore, we start considering $Z = \{6, 7, 12\}$. Then we have to add to Z also all $j \notin Z$, for which $s \notin W(Z, j)$. In the first step it means that we have to add $\{5, 11\}$ to Z, in the second step $\{4, 10\}$ and so on. After 4 steps Z contains indices of all the distributions, which means that 7 does not come into consideration for application of Theorem 3. It results in necessity to add 7 to Z and we have

[1]For meaning of L see Theorem 3.

Table 1: Finding whether Theorem 3 may be applied

Z	s	$j \notin Z : s \notin W(Z,j)$
$6,12$	7	$5,11$ $4,10$ $2,3,8$ 1
$6,7,12$	5	$8,11$ 10
$6,7,12$	8	⋮

to choose another s. Also at this moment 3 indices come into consideration: $5,8,11$. Choosing 5 this time we get after two steps $Z = \{5,6,7,8,10,11,12\}$, which with $s = 7$ meet the assumptions of Theorem 3. According to this Theorem we get

$$\begin{aligned}(\pi_1 \triangleright \ldots \triangleright \pi_{13})^{-\{1,7,9,10,11,12,13,14,16,17,20\}} &= (\pi_1 \triangleright \ldots \triangleright \pi_{13})^{\downarrow\{2,3,4,5,6,8,15,18,19\}} \\ &= \mu^{\downarrow\emptyset} \triangleright \mu^{\downarrow\emptyset} \triangleright \mu^{\downarrow\emptyset} \triangleright \mu^{\downarrow\{8\}} \triangleright \pi_5 \triangleright \pi_6^{-\{1\}} \triangleright \pi_7^{-\{14\}} \triangleright \pi_8^{-\{16\}} \triangleright \pi_{10} \triangleright \pi_{11} \triangleright \pi_{12}^{-\{7\}} \\ &= \mu^{\downarrow\{8\}} \triangleright \pi_5 \triangleright \pi_6^{-\{1\}} \triangleright \pi_7^{-\{14\}} \triangleright \pi_8^{-\{16\}} \triangleright \pi_{10} \triangleright \pi_{11} \triangleright \pi_{12}^{-\{7\}} \\ &= \mu^{\downarrow\{8\}} \triangleright \pi_5 \triangleright \pi_6^{\downarrow\{2,3,4\}} \triangleright \pi_7^{\downarrow\{3,15\}} \triangleright \pi_8^{\downarrow\{15,18\}} \triangleright \pi_{10} \triangleright \pi_{11} \triangleright \pi_{12}^{\downarrow\{2,5,6\}},\end{aligned}$$

where

$$\mu^{\downarrow\{8\}} = (\pi_1 \triangleright \ldots \triangleright \pi_4)^{\downarrow\{8\}}.$$

After this step, all the other attempts to find Z and s, for which Theorem 3 could be applied, fail. So we have to start applying Theorem 2.

Now, we have a 9-dimensional distribution and our goal is to get 3-dimensional one – distribution of variables X_2, X_3, X_5. So, we have to marginalize 6 variables out with indices $4,6,8,15,18,19$. This situation is demonstrated in Figure 6. Let us apply Theorem 2 to delete variable X_8:

$$\begin{aligned}&(\pi_1 \triangleright \ldots \triangleright \pi_{13})^{\downarrow\{2,3,4,5,6,15,18,19\}} \\ &\quad = \mu^{\downarrow\{\emptyset\}} \triangleright (\mu^{\downarrow\{8\}} \textcircled{\triangleright}_{\emptyset} \pi_5)^{-\{8\}} \triangleright \pi_6^{\downarrow\{2,3,4\}} \triangleright \pi_7^{\downarrow\{3,15\}} \triangleright \pi_8^{\downarrow\{15,18\}} \triangleright \pi_{10} \triangleright \pi_{11} \\ &\qquad \triangleright \pi_{12}^{\downarrow\{2,5,6\}}.\end{aligned}$$

Let us denote

$$\kappa_1(x_4) = (\mu^{\downarrow\{8\}} \textcircled{\triangleright}_{\emptyset} \pi_5)^{-\{8\}} = \left(\frac{\mu^{\downarrow\{8\}}\pi_5}{\pi_5^{\downarrow\{8\}}}\right)^{\downarrow\{4\}} = \sum_{x_8 \in \mathbf{X}_8} \frac{\mu(x_8)\pi_5(x_4,x_8)}{\pi_5(x_8)}.$$

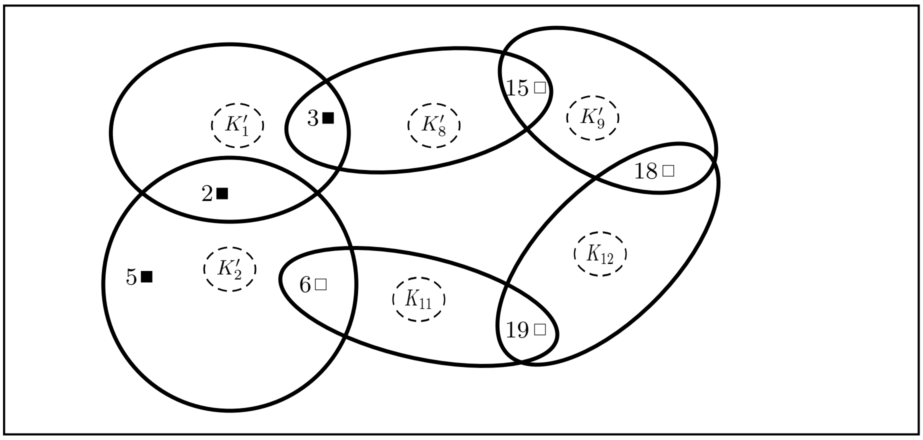

Figure 2: Modified sets of variables corresponding to 9-dimensional model.

Then we get

$$(\pi_1 \triangleright \ldots \triangleright \pi_{13})^{\downarrow\{2,3,4,5,6,15,18,19\}}$$
$$= \kappa_1 \triangleright \pi_6^{\downarrow\{2,3,4\}} \triangleright \pi_7^{\downarrow\{3,15\}} \triangleright \pi_8^{\downarrow\{15,18\}} \triangleright \pi_{10} \triangleright \pi_{11} \triangleright \pi_{12}^{\downarrow\{2,5,6\}},$$

and can start marginalizing variable X_4 out. Analogously to the preceding step we get

$$(\pi_1 \triangleright \ldots \triangleright \pi_{13})^{\downarrow\{2,3,5,6,15,18,19\}}$$
$$= \kappa_2 \triangleright \pi_7^{\downarrow\{3,15\}} \triangleright \pi_8^{\downarrow\{15,18\}} \triangleright \pi_{10} \triangleright \pi_{11} \triangleright \pi_{12}^{\downarrow\{2,5,6\}},$$

where

$$\kappa_2(x_2, x_3) = (\kappa_1 \textcircled{\triangleright}_{\emptyset} \triangleright \pi_6^{\downarrow\{2,3,4\}})^{-\{4\}} = \sum_{x_4 \in \mathbf{X}_4} \frac{\kappa_1(x_4)\pi_6(x_2, x_3, x_4)}{\pi_6(x_4)}.$$

Let us show how to marginalize, for example, X_{18}. The rest will be left to the reader.

$$(\pi_1 \triangleright \ldots \triangleright \pi_{13})^{\downarrow\{2,3,5,6,15,19\}}$$
$$= \kappa_2 \triangleright \pi_7^{\downarrow\{3,15\}} \triangleright \pi_8^{\downarrow\{15\}} \triangleright \left(\pi_8^{\downarrow\{15,18\}} \textcircled{\triangleright}_{\{2,3,15,18\}} \pi_{10}\right)^{-\{18\}} \triangleright \pi_{11} \triangleright \pi_{12}^{\downarrow\{2,5,6\}}$$
$$= \kappa_2 \triangleright \pi_7^{\downarrow\{3,15\}} \triangleright \kappa_3 \triangleright \pi_{11} \triangleright \pi_{12}^{\downarrow\{2,5,6\}},$$

where

$$\kappa_3(x_{15,19}) = \left(\pi_8^{\downarrow\{15,18\}} \textcircled{\triangleright}_{\{2,3,15,18\}} \pi_{10}\right)^{-\{18\}} = \left(\pi_8^{\downarrow\{15,18\}} \triangleright \pi_{10}\right)^{-\{18\}}.$$

Let us still mention that we could delete $\pi_8^{\downarrow\{15\}}$ from the generating sequence because of Lemma 1.

Acknowledgements

The research was partially supported by VŠE, under grant IGA 611074 and AV ČR under grant A2075302.

References

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

[2] R. Jiroušek, Marginalization in composed probabilistic models. In: *Proc. of the 16th Conf. Uncertainty in Artificial Intelligence UAI'00* (C. Boutilier and M. Goldszmidt eds.), Morgan Kaufmann Publ., San Francisco, California, 2000, pp. 301-308.

[3] R. Jiroušek, Decomposition of multidimensional distributions represented by perfect sequences. *Annals of Mathematics and Artificial Intelligence*, **35**(2002) pp. 215-226.

[4] R. Jiroušek, What is the Difference between Bayesian Networks and Compositional Models? In: H. Noguchi, H. Ishii, M. Inuiguchi (eds.) Proceedings of the 7th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty. Awaji Yumebutai ICC, Japan, 2004, pp. 191-196.

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

[6] R.D. Shachter, Evaluating Influence Diagrams, Operations Res., **34**(1986), pp. 871-890.

[7] R.D. Shachter, Probabilistic inference and influence diagrams. *Oper. Res.*, **36**(1988), pp. 589-604.

[8] G. Shafer, *Probabilistic Expert Systems*. SIAM, Philadelphia, 1996.