What is the Difference between Bayesian Networks and Compositional Models?

Radim Jiroušek

Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Praha 8, CZECH REPUBLIC
Tel: +420 26605 2046, Fax: +420 286 581 419, E-mail: radim@utia.cas.cz

Abstract

In this contribution we shall discuss a relation of two types of multidimensional models introduced within the framework of probability theory, which appeared to be in a sense equivalent: Bayesian networks and compositional models. Based on a simple example we shall analyse algorithms transforming one type of the model into the other. In this way we shall demonstrate a principal difference, which explains why the compositional models are more efficient for computations.

Keywords:

Probability, Multidimensional distribution, Graphical Markov model, Operator of composition.

Introduction

In the previous lectures presented in a series of the Czech-Japan Seminars ([1] - [4]) we showed that all distributions representable by Bayesian networks can also be represented by an alternative way: in the form of *compositional models*[1]. These models abandon the necessity to describe the dependence structure of modelled distributions by graphs. In contrast to this, it describes directly how such a multidimensional distribution is computed – *composed* – from a system of low-dimensional distributions. However, as we shall see from this paper, it is not the only advantage of compositional models. We shall show that compositional models (especially those defined by perfect sequences) are more suitable for computations.

Notation

In this paper we deal with a system of finite-valued random variables, whose indices are from a finite set N. Their probability distributions are denoted by Greek letters. Thus, $\pi(x_N)$ will denote an $|N|$-dimensional distribution of variables $\{X_i\}_{i\in N}$, $\pi(x_J)$ (for $J \subset N$) is its marginal distribution for variables $\{X_i\}_{i\in J}$. This marginal distribution will also be denoted simply $\pi^{(J)}$. When considering disjoint subsets $L_1, L_2 \subset N$, the symbol $\pi(x_{L_1}|x_{L_2})$ denotes the respective conditional distribution of variables $\{X_i\}_{i\in L_1}$ given $\{X_i\}_{i\in L_2}$ (we do not exclude situations when either L_1 or L_2 is empty; $\pi(\emptyset) = 1, \pi(x_{L_1}|\emptyset) = \pi(x_{L_1})$).

By a Bayesian network representing a distribution $\pi(x_N)$, we understand a couple consisting of an acyclic directed graph $G = (N, E)$ and a system of conditional probability distributions

$$\{\nu_i(x_i|x_{pa(i)})\}_{i\in N},$$

such that

$$\pi(x_N) = \prod_{i\in N} \nu_i(x_i|x_{pa(i)}).$$

$pa(i)$ thus denotes the set of all parents of node i in the considered graph G.

For compositional models, the most important notion is the following one introducing a possibility to compose two low-dimensional distributions.

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

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

where $\pi(x_M) \ll \kappa(x_M)$ denotes that $\pi(x_M)$ is *dominated* by $\kappa(x_M)$:

$$\kappa(x_M) = 0 \Longrightarrow \pi(x_M) = 0$$

for all combinations of values x_M.

Compositional models and their basic properties

The following assertion summarize those properties of the operator of composition proven in previous papers, which will be necessary in this text.

[1] Here we speak only about probabilistic models. Possibilistic compositional models were introduced by J. Vejnarová; for more references see [10].

Lemma 1 *Let for probability distributions* $\pi(x_K)$ *and* $\kappa(x_L)$

$$\pi(x_{L\cap K}) \ll \kappa(x_{L\cap K})$$

(i.e., $\pi \triangleright \kappa$ *is defined). Then*

1. $\pi \triangleright \kappa$ *is a probability distribution of variables* $\{X_i\}_{i\in L\cup K}$ *and its marginal distribution for variables* $\{X_i\}_{i\in K}$ *equals* π*:*

$$(\pi \triangleright \kappa)(x_K) = \pi(x_K).$$

2. *If* $\pi(x_K)$ *and* $\kappa(x_L)$ *are consistent (i.e.,* $\pi(x_{K\cap L}) = \kappa(x_{K\cap L})$*) then*

$$\pi \triangleright \kappa = \kappa \triangleright \pi.$$

From this assertion one can immediately see that if $L \subseteq K$ then

$$\pi(x_K) \triangleright \kappa(x_L) = \pi(x_K).$$

As said already above, the main significance of the operator of composition is in the fact that it can form multidimensional distributions from systems of low-dimensional ones. First, let us stress that if not specified otherwise by brackets, operators $\triangleright$ are always applied from left to right. It means that

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

Therefore, in order to construct a multidimensional distribution it is sufficient to determine a sequence – we will call it a *generating sequence* – of low-dimensional distributions $\pi_1, \pi_2, \ldots, \pi_n$. To simplify all the expressions in the sequel let us make the following convention: whenever we will speak about a distribution π_k in this and the following sections, if not specified explicitly otherwise, the distribution π_k will always be assumed to be a distribution of variables $\{X_i\}_{i\in K_k}$, which means it is a distribution $\pi_k(x_{K_k})$.

With respect to application of operators of composition, it is important to realize that the operator is non-commutative and also non-associative. So, generally

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

Nevertheless, under special conditions these properties hold true. For example, the following assertion was proven in [5].

Lemma 2 *If* $K_1 \supseteq (K_2 \cap K_3)$ *then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleright \pi_2.$$

The next assertion expresses a property, which is very important from the computational point of view (for proof see also [5]).

Lemma 3 *If* π_1, π_2 *and* π_3 *are such that* $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ *is defined then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright (\pi_2 \circledcirc_{K_1} \pi_3),$$

where

$$\pi_2 \circledcirc_{K_1} \pi_3 = (\pi_3^{((K_1\setminus K_2)\cap K_3)} \pi_2) \triangleright \pi_3.$$

Perfect sequence models

Not all generating sequences are equally efficient in their representations of multidimensional distributions. Among them, so-called perfect sequences hold an important position.

Definition 2 A generating sequence of probability distributions $\pi_1, \pi_2, \ldots, \pi_n$ is called *perfect* if $\pi_1 \triangleright \ldots \triangleright \pi_n$ is defined and for all $k = 2, \ldots n$

$$(\pi_1 \triangleright \ldots \triangleright \pi_{k-1}) \triangleright \pi_k = \pi_k \triangleright (\pi_1 \triangleright \ldots \triangleright \pi_{k-1}).$$

From this definition one can hardly see the importance of perfect sequences. This importance becomes clearer from the following characterization theorem (for its proof see [6]).

Theorem 1 *A sequence of distributions* $\pi_1, \pi_2, \ldots, \pi_n$ *is perfect* iff *all the distributions from this sequence are marginals of the distribution* $(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)$.

In other words, the theorem claims that when considering that low-dimensional distributions π_k are carriers of local information, the constructed multidimensional distribution represents global information, faithfully reflecting all of the local input[2]. This is why we will be so much interested in perfect sequence models.

The following assertion (whose proof is in [5]) shows that we may restrict our attention only to perfect sequences, because each generating sequence, for which $\pi_1 \triangleright \ldots \triangleright \pi_n$ is defined, can easily be transformed into a perfect one.

Theorem 2 *If* $\pi_1 \triangleright \ldots \triangleright \pi_n$ *is defined then the sequence* $\kappa_1, \ldots, \kappa_n$ *computed by the following process*

$$\begin{aligned}
\kappa_1 &= \pi_1,\\
\kappa_2 &= \kappa_1^{(K_2\cap K_1)} \triangleright \pi_2,\\
\kappa_3 &= (\kappa_1 \triangleright \kappa_2)^{(K_3\cap(K_1\cup K_2))} \triangleright \pi_3,\\
&\vdots\\
\kappa_n &= (\kappa_1 \triangleright \ldots \triangleright \kappa_{n-1})^{(K_n\cap(K_1\cup\ldots K_{n-1}))} \triangleright \pi_n
\end{aligned}$$

is perfect and

$$\pi_1 \triangleright \ldots \triangleright \pi_n = \kappa_1 \triangleright \ldots \triangleright \kappa_n.$$

[2]For futher comments regarding application of probabilistic models in expert systems see [9].

Marginalization for perfect sequence models

Unfortunately, we cannot get a marginal distribution $(\pi_1 \triangleright \pi_2)^{(L)}$ (for $L \subseteq K_1 \cup K_2$) as a composition of some marginal distributions. Namely, it is easy to show that generally

$$(\pi_1 \triangleright \pi_2)^{(L)} \neq \pi_1^{(K_1 \cap L)} \triangleright \pi_2^{(K_2 \cap L)}.$$

The following assertion (proven in [6]) gives an instruction how to compute a marginal distribution, whose dimensionality is by one smaller than the dimensionality of the original distribution. Since we do not impose any condition on the deleted variable, by iterative application of this assertion we can (at least theoretically) compute an arbitrary marginal distribution.

For the sake of simplicity, we will use the following simplified notation:

$$\begin{aligned}&\pi_k^{[\ell]} = \pi_k(x_{K_k \setminus \{\ell\}}),\\ &(\pi_1 \triangleright \ldots \triangleright \pi_n)^{[\ell]}\\ &\qquad = (\pi_1 \triangleright \ldots \triangleright \pi_n)(x_{(K_1 \cup \ldots \cup K_n) \setminus \{\ell\}})\end{aligned}$$

for $\ell \in K_k$.

Theorem 3 *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 some

$$\{i_1, i_2, \ldots, i_m\} \subseteq \{1, 2, \ldots, n\}$$

(assuming $(i_1 < i_2 < \ldots < i_m)$) such that $\ell \notin K_j$ for all

$$j \in \{1, 2, \ldots, n\} \setminus \{i_1, i_2, \ldots, i_m\}.$$

Then

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

where

$$\begin{aligned}&\kappa_j = \pi_j \text{ for all } j \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_m\},\\ &\kappa_{i_1} = \pi_{i_1}^{[\ell]},\\ &\kappa_{i_2} = (\pi_{i_1} \circledcirc_{L_{i_2-1}} \pi_{i_2})^{[\ell]},\\ &\kappa_{i_3} = (\pi_{i_1} \circledcirc_{L_{i_2-1}} \pi_{i_2} \circledcirc_{L_{i_3-1}} \pi_{i_3})^{[\ell]},\\ &\qquad\vdots\\ &\kappa_{i_m} = (\pi_{i_1} \circledcirc_{L_{i_2-1}} \pi_{i_2} \circledcirc_{L_{i_3-1}} \ldots \circledcirc_{L_{i_m-1}} \pi_{i_m})^{[\ell]},\end{aligned}$$

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

As a special case of this assertion one can immediately conclude that if a variable to be deleted is contained only in one distribution from a generating sequence then marginalization of a multidimensional model can simply be done just by marginalizing the respective low-dimensional distribution. Recall that it, in a way, corresponds to the node deletion step of a Shachter's procedure for marginalization of Bayesian networks ([8]), which allows deletion of *childless* nodes. We shall, discuss this fact in more detail later in Example.

Relation of compositional models and Bayesian networks

The reader certainly noticed that any distribution $\kappa(x_N)$ represented by a Bayesian network (consisting of an acyclic graph $G = (N, E)$ and a system of conditional probability distributions $\{\nu_i(x_i|x_{pa(i)})\}_{i \in N}$) can also be represented as a compositional model. In this case, namely,

$$\kappa(x_N) = \prod_{i \in N} \nu_i(x_i|x_{pa(i)}),$$

and distributions $\pi_i(x_{fam(i)})$ $(fam(i) = \{i\} \cup pa(i))$ can easily be found, for which

$$\pi_i(x_i|x_{pa(i)}) = \nu_i(x_i|x_{pa(i)}).$$

To do it, one can take, for example, a uniform distribution $\mu(x_{pa(i)})$ and set

$$\pi_i(x_{fam(i)}) = \nu_i(x_i|x_{pa(i)})\mu(x_{pa(i)}).$$

Now we immediately see that the distribution $\kappa(x_N)$ represented by the considered Bayesian network can be expressed in the form

$$\kappa(x_N) = \pi_{i_1} \triangleright \pi_{i_2} \triangleright \ldots \triangleright \pi_{i_{|N|}},$$

if the permutation $i_1, i_2, \ldots, i_{|N|}$ is such that all parents of a node are always before it[3]:

$$i_k \in pa(i_\ell) \implies k < \ell.$$

Therefore, each Bayesian network can be represented by a generating sequence and, due to Theorem 2, also by a perfect sequence. In the next paragraph we shall formalize this procedure in a slightly more efficient way and show how to find a Bayesian network representation for a distribution defined by a perfect sequence. Thus we will see that the class of distributions represented by Bayesian networks is equivalent to the class of distributions represented by perfect sequences. However, when studying these transformation processes in detail, we will see the advantage of perfect sequence representation.

Transformation of a perfect sequence into a Bayesian network

The following simple procedure transforming an arbitrary perfect sequence $\pi_1, \ldots, \pi_n$ into a Bayesian network consists of a definition of a graph and of a computation of conditional distributions $\kappa(x_i|x_{pa(i)})$ defining the constructed Bayesian network.

[3] An existence of such an ordering is guaranteed by acyclicity of the considered graph G.

Algorithm PSM $\longrightarrow$ BN

(a) *Having a perfect sequence $\pi_1(x_{K_1}), \ldots, \pi_n(x_{K_n})$ we first order (in an arbitrary way) all the indices of the considered variables, i.e.*

$$\{1, 2, 3, \ldots, |N|\} = K_1 \cup \ldots \cup K_n = N.$$

(b) *A graph of the constructed belief network is defined in the following way:*

1. *the set of nodes is N;*
2. *there is an edge $(i \rightarrow j)$ if there exists a distribution π_k such that all the following three conditions hold:*
   (i) $i, j \in K_k$,
   (ii) $j \notin K_1 \cup \ldots \cup K_{k-1}$
   (iii) either $i \in K_1 \cup \ldots \cup K_{k-1}$ *or i is in the ordering defined in step* **(a)** *before j.*

(c) *For each j the requirement $j \in K_k$, $j \notin K_1 \cup \ldots \cup K_{k-1}$ is met exactly for one $k \in \{1, \ldots, s\}$. It means that all the parents of node X_j must be in the respective set X_{K_k} and therefore the necessary conditional distribution $\nu_j(x_j|x_{pa(j)})$ can be easily computed from distribution $\pi_k(x_{K_k})$.*

Regarding this algorithm, it is important to realize that different orderings of indices defined in the first step of the algorithm may lead to different graphs. It corresponds to the fact that usually several acyclic directed graphs are equivalent in the sense that they define equivalent Bayesian networks (they have the same underlying graphs and the same list of *immoralities*).

Transformation of a Bayesian network into a perfect sequence

At the beginning of this section we have presented an idea how to find a perfect sequence model representing the distribution defined by a Bayesian network. Let us now formalize it in the way that the length of the resulting perfect sequence may be smaller than the number of variables of the considered distribution.

Algorithm BN $\longrightarrow$ PSM

(a) *Having a Bayesian network with graph $G = (N, E)$ and system of conditional distributions $\{\nu_i(x_i|x_{pa(i)})\}_{i \in N}$ construct an enumeration of indices from N in the following way:*

1. *Assign 1 to any source (parentless node) of G.*
2. *assign the next number to a node whose all parents have already been enumerated. If there are more such nodes choose any with the greatest number of parents.*

(b) *Let $1, 2, \ldots, |N|$ be the elements of N ordered according to the enumeration constructed in the step (a). Auxiliary sets $R_1, \ldots, R_n$ and $S_1, \ldots, S_n$ are constructed by the following process:*

1. *set $r \leftarrow 0, n \leftarrow 1$;*
2. *while $r < |N|$ perform:*
   (i) find the maximal integer m such that for all $k = 1, \ldots, m$
   $pa(r+k) = pa(r+1) \cup \{r + i : 0 < i < k\}$.
   (ii) $R_n \leftarrow pa(r+1)$;
   (iii) $S_n \leftarrow \{r+1, r+2, \ldots, r+m\}$;
   (iv) $r \leftarrow r + m$;
   (v) $n \leftarrow n + 1$;

(c) *For each $i = 1, 2, \ldots, n-1$ compute:*
$K_i = R_i \cup S_i$,
$\kappa_i = (\kappa_1 \triangleright \ldots \triangleright \kappa_{i-1})^{(R_i)} \prod_{j \in S_i} \nu_j(x_j|x_{pa(j)})$.

Example

Let $N = \{1, 2, 3, \ldots, 8\}$ and assume that the generating sequence

$$\pi_1(x_1, x_2), \pi_2(x_2, x_3), \pi_3(x_2, x_4), \pi_4(x_3, x_5), \pi_5(x_4, x_6), \pi_6(x_5, x_6, x_7, x_8)$$

is perfect. Let us have a look how the algorithm PSM $\longrightarrow$ BN proceed when applied to this sequence.

In step (a) consider the (auxiliary) ordering $\{1, 2, 3, \ldots, 8\}$. Then, the acyclic directed graph is defined in step (b). In the following text we shall show how the sets of parents $pa(j)$ are constructed for this graph.

From the conditions 2.(i) and 2.(ii) we see that only those nodes, which are indices of the variables appearing among the argument of that distribution, where X_j appears the first time, may be in $pa(j)$. Let us apply this rule to $j = 1$. X_1 appears first time among the arguments of π_1. The other argument of this distribution is X_2 and therefore only the node 2 should be considered as a potential parent of the node 1. However, the condition 2.(iii) is not fulfilled for $i = 2$ and $j = 1$ and therefore the node 1 does not have any parent.

Analogously, for $j = 2$ we get that there is again only one potential parent: $i = 1$. For this couple the condition 2.(iii) holds true because 1 is (in the considered ordering $j = 1, 2, \ldots, 8$) before 2. Therefore $pa(2) = \{1\}$.

For $j = 3$ there is again only one potential parent; node 2 (both X_2 and X_3 are among the arguments of π_2 and it is the first appearance of X_3). In this

case the condition 3.(iii) is fulfilled because X_2 is an argument of π_1. Thus we get $pa(3) = \{2\}$.

Continuing in this way for $j = 4, 5, 6, 7, 8$ we get a graph, which is in Figure 1.

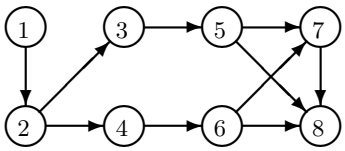

Figure 1: Acyclic graph of a Bayesian network

Operation of the algorithm is finished by computation of the necessary system of conditional distributions performed in step (c):

$$\begin{aligned}
\nu_1(x_1) &= \pi_1(x_1), & \nu_2(x_2|x_1) &= \pi_1(x_2|x_1),\\
\nu_3(x_3|x_2) &= \pi_2(x_3|x_2), & \nu_4(x_4|x_2) &= \pi_3(x_4|x_2),\\
\nu_5(x_5|x_3) &= \pi_4(x_5|x_3), & \nu_6(x_6|x_4) &= \pi_5(x_6|x_4),\\
\nu_7(x_7|x_5,x_6) &= \pi_6(x_7|x_5,x_6),\\
\nu_8(x_8|x_5,x_6,x_7) &= \pi_6(x_8|x_5,x_6,x_7).
\end{aligned}$$

What should be stressed at this moment is that all these computations are *local*[4], they do not require additional auxiliary memory and their time requirements are liner with the number of probabilities defining the compositional model

Let us, now, apply the BN $\longrightarrow$ PSM algorithm to the resulting Bayesian network consisting of a directed acyclic graph in Figure 1 and the respective conditional distributions $\nu_1, \ldots, \nu_8$. First, realizing the step (a) we have to construct an enumeration (ordering) of nodes of the graph. Notice, that result of this step is not unique. A possible ordering can be $1, 2, 3, 4, 5, 6, 7, 8$ but also $1, 2, 4, 6, 3, 5, 7, 8$ (and several others). Considering the latter one, step (b) of the algorithm yields the system of auxiliary sets R_i and S_i as shown in Table 1.

Table 1: Auxiliary sets R_i, S_i

i	R_i	S_i
1	$\emptyset$	$\{1,2\}$
2	$\{2\}$	$\{4\}$
3	$\{4\}$	$\{6\}$
4	$\{2\}$	$\{3\}$
5	$\{3\}$	$\{5\}$
6	$\{5,6\}$	$\{7,8\}$

Afterwards, in the step (c) the following 6 distributions are computed

$$\begin{aligned}
&\kappa_1(x_1,x_2) = \nu_1(x_1)\cdot\nu_2(x_2|x_1),\\
&\kappa_2(x_2,x_4) = \kappa_1(x_2)\cdot\nu_4(x_4|x_2),\\
&\kappa_3(x_4,x_6) = (\kappa_1\triangleright\kappa_2)(x_4)\cdot\nu_6(x_6|x_4),\\
&\kappa_4(x_2,x_3) = (\kappa_1\triangleright\kappa_2\triangleright\kappa_3)(x_2)\cdot\nu_3(x_3|x_2),\\
&\kappa_5(x_3,x_5) = (\kappa_1\triangleright\kappa_2\triangleright\kappa_3\triangleright\kappa_4)(x_3)\cdot\nu_5(x_5|x_3),\\
&\kappa_6(x_5,x_6,x_7,x_8) = (\kappa_1\triangleright\kappa_2\triangleright\kappa_3\triangleright\kappa_4\triangleright\kappa_5)(x_5,x_6)\\
&\qquad\qquad\cdot\nu_7(x_7|x_5,x_6)\cdot\nu_8(x_8|x_5,x_6,x_7),
\end{aligned}$$

which form a perfect sequence defining the same distribution as the considered Bayesian network, and therefore also

$$\pi_1\triangleright\pi_2\triangleright\pi_3\triangleright\pi_4\triangleright\pi_5\triangleright\pi_6 = \kappa_1\triangleright\kappa_2\triangleright\kappa_3\triangleright\kappa_4\triangleright\kappa_5\triangleright\kappa_6.$$

Let us, however, have a closer look at the computation of the distributions κ_i. Computations of κ_1 and κ_2 are obviously local. Also computations of κ_3, κ_4 and κ_5 are local, because the sequence $\kappa_1, \kappa_2, \ldots, \kappa_6$ is perfect and therefore

$$\begin{aligned}
&(\kappa_1\triangleright\kappa_2)(x_4) = \kappa_2(x_4),\\
&(\kappa_1\triangleright\kappa_2\triangleright\kappa_3)(x_2) = \kappa_1(x_2),\\
&(\kappa_1\triangleright\kappa_2\triangleright\kappa_3\triangleright\kappa_4)(x_3) = \kappa_4(x_3).
\end{aligned}$$

But no analogous simple rule can be applied to computation of $(\kappa_1\triangleright\kappa_2\triangleright\kappa_3\triangleright\kappa_4\triangleright\kappa_5)(x_5,x_6)$. For this we have to apply Theorem 3:

$$\begin{aligned}
&(\kappa_1\triangleright\kappa_2\triangleright\kappa_3\triangleright\kappa_4\triangleright\kappa_5)(x_5,x_6)\\
&\quad = ((((\kappa_1\triangleright\kappa_2\triangleright\kappa_3\triangleright\kappa_4\triangleright\kappa_5)^{[1]})^{[2]})^{[3]})^{[4]}.
\end{aligned}$$

Computing $(\kappa_1\triangleright\kappa_2\triangleright\kappa_3\triangleright\kappa_4\triangleright\kappa_5)^{[1]}$ is simple, because X_1 appears only in π_1 and therefore (cf. Theorem 3)

$$(\kappa_1\triangleright\kappa_2\triangleright\kappa_3\triangleright\kappa_4\triangleright\kappa_5)^{[1]} = \kappa_1^{[1]}\triangleright\kappa_2\triangleright\kappa_3\triangleright\kappa_4\triangleright\kappa_5,$$

which equals $\kappa_2\triangleright\kappa_3\triangleright\kappa_4\triangleright\kappa_5$, because κ_1 and κ_2 are consistent and therefore $\kappa_1(x_2)\triangleright\kappa_2 = \kappa_2\triangleright\kappa_1(x_2) = \kappa_2$ (see the comment after Lemma 1).

Applying Theorem 3 to eliminate variable x_2 we get

$$(\kappa_2\triangleright\kappa_3\triangleright\kappa_4\triangleright\kappa_5)^{[2]} = \kappa_2^{[2]}\triangleright\kappa_3\triangleright(\kappa_2\circledcirc_{\{4,6\}}\kappa_4)^{[2]}\triangleright\kappa_5.$$

Let us denote

$$\begin{aligned}
\hat\kappa(x_3,x_4) &= (\kappa_2\circledcirc_{\{4,6\}}\kappa_4)^{[2]}\\
&= (\kappa_4(\emptyset)\kappa_2(x_2,x_4)\triangleright\kappa_4(x_2,x_3))^{[2]}\\
&= (\kappa_2(x_2,x_4)\triangleright\kappa_4(x_2,x_3))^{[2]}.
\end{aligned}$$

Since analogously to the previous case $\kappa_2^{[2]}\triangleright\kappa_3 = \kappa_3$, deletion of the third variable (X_3) is performed as follows

$$(\kappa_3\triangleright\hat\kappa\triangleright\kappa_5)^{[3]} = \kappa_3\triangleright\hat\kappa^{[3]}\triangleright(\hat\kappa\circledcirc_{\{4,6\}}\kappa_5)^{[3]}.$$

Denoting

$$\begin{aligned}
\hat{\hat\kappa}(x_4,x_5) &= (\hat\kappa\circledcirc_{\{4,6\}}\kappa_5)^{[3]}\\
&= (\kappa_5(\emptyset)\hat\kappa(x_3,x_4)\triangleright\kappa_5(x_3,x_5))^{[3]}\\
&= (\hat\kappa(x_3,x_4)\triangleright\kappa_5(x_3,x_5))^{[3]}
\end{aligned}$$

[4]Notice that we understand this notion more restrictively in comarison with the understanding used in [7]

we see that

$$(\kappa_3 \triangleright \hat{\kappa} \triangleright \kappa_5)^{[3]} = \kappa_3 \triangleright \hat{\kappa}^{[3]} \triangleright \hat{\hat{\kappa}}(x_4, x_5) = \kappa_3 \triangleright \hat{\hat{\kappa}}(x_4, x_5),$$

where the last modification is allowed because of Lemma 2 and the comment after Lemma 1. In this way we eventually received the possibility to compute the distribution κ_6 from the required perfect sequence

$$\kappa_6 = (\kappa_3 \triangleright \hat{\hat{\kappa}}(x_4, x_5))^{[4]} \nu_7(x_7|x_5, x_6) \cdot \nu_8(x_8|x_5, x_6, x_7).$$

However, to get this distribution we had to compute $\hat{\kappa}$, $\hat{\hat{\kappa}}$ and $(\kappa_3 \triangleright \hat{\hat{\kappa}}(x_4, x_5))^{[4]}$, and all these three computations were not local; we had to compute distributions for the groups of variables, which did not appear in the input Bayesian network (in a sense it corresponds to *inheriting parents* in the Shachter's procedure - see [8].)

1 Conclusions

Analyzing the PSM $\longrightarrow$ BN algorithm we showed that all the computations are always local – computation of each of the necessary conditional distributions can always be done from only one of the distributions from which the perfect sequence consists of. On the other hand, on an example we showed that the opposite algorithm BN $\longrightarrow$ PSM is algorithmically more complex; for some (in fact for most of) Bayesian networks the computation of a corresponding perfect sequence model requires computation of some marginal distributions, which is generally known to be a difficult task. From this we see that, though both the models represent the same probability distribution, perfect sequence model has some of the marginal distributions "pre-computed" and therefore it is not surprising that some computational procedures are less demanding than the analogous procedures for Bayesian networks.

Let us conclude the paper by another simple example supporting the above presented statement. As mentioned in the paragraph dealing with marginalization for compositional model we formulated a simple rule for deletion of variables appearing among the arguments of only one distribution. It somehow corresponds to the Shachter's deletion rule ([8]) concerning childless nodes in Bayesian networks. Considering the network with acyclic directed graph in Figure 1 this rule can directly be applied only to node 8, whereas in the equivalent compositional model the corresponding rule can be applied to three variables: X_1, X_7 and X_8.

Acknowledgment

The research was supported by Grant Agency of the Czech Academy of Sciences by grant no. A2075302 and Austrian-Czech grant AKTION KONTAKT 2004/19.

References

[1] Jiroušek R.: Construction of Multidimensional Models. Part 1: Probabilistic Approach. In: Ramík, Jaroslav, Novák, Vilém (eds.) Methods for Decision Support in Environment with Uncertainty - Applications in Economics, Business and Engineering. Ostrava: Ostravská universita, 2000, pp 111-118.

[2] Jiroušek R.: An anticipating operator and its role for computation of marginal distributions. In: Ishii, H., Inuiguchi, M. (eds.). Proceedings of the 3rd Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty. Osaka: Osaka University, 2001, pp 69-75.

[3] Jiroušek R.: Detection of conditional independence relations from persegrams. In: Jiroušek, R., Vejnarová J. (eds.). Proceedings of the 4th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty. Jindřichův Hradec, VŠE, 2001, pp 52-62.

[4] Jiroušek R.: On computational procedures for probabilistic compositional models. In: Junzo Watada (ed) Proceedings of the 5th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty. Osaka University, Japan, 2002, pp 3-10.

[5] Jiroušek R., Composition of probability measures on finite spaces. In: *Proc. of the 13th Conf. Uncertainty in Artificial Intelligence UAI'97*, eds. D. Geiger and P. P. Shenoy, Morgan Kaufmann Publ., San Francisco, California, 1997, pp. 274-281.

[6] Jiroušek R., 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.

[7] Lauritzen S. L. and Spiegelhalter D. J., Local computation with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society series B* **50** (1988), pp. 157–224.

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

[9] Shafer G., *Probabilistic Expert System*. Siam, 1996.

[10] Vejnarová J., On construction and interpretation of multidimensional possibility distributions. In: Jiroušek, R., Vejnarová J. (eds.). Proceedings of the 4th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty. Jindřichův Hradec, VŠE, 2001, pp 152-165.