Marginalization Algorithm for Compositional Models

Radim Jiroušek
Inst. of Inf. Th. and Automation
Academy of Sciences
Prague
radim@utia.cas.cz

Václav Kratochvíl
Inst. of Inf. Th. and Automation
Academy of Sciences
Prague
Vaclav@Kratochvil.biz

Abstract

The paper deals with a problem of marginalization of multidimensional probability distributions represented by compositional models, more precisely by perfect sequence models. It appears that from the computational point of view, the solution is more efficient than any known marginalization process for Bayesian networks. This is because the process, which is in the paper described in a form of an algorithm, takes advantage of the fact that perfect sequence models have some information explicitly encoded, which can be got from Bayesian networks only by application of rather computationally expensive procedures.

Keywords: Compositional model, Multidimensional distribution, Bayesian network, Marginalization, Algorithm.

1 Introduction

Ability to represent and process multidimensional probability distributions is a necessary condition for application of probabilistic methods in Artificial Intelligence. Most popular approaches to solve this task are based on Graphical Markov Models (GMM - [8]), from which Bayesian Networks (BN - [3]) are most often used. In the present paper we will deal with an alternative to GMM, so called Compositional Models (CM - [4, 5, 6]) and present an algorithm enabling us to compute marginal distributions from really multidimensional ones.

A possible solution of this task for BNs is given in papers by R. Shachter [9, 10][1]. 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 in case we did not employ *marginalization by reduction* theoretically supported by Theorem 3 of our paper. This theorem, namely, takes advantage of the main difference between Bayesian networks and compositional models revealed in [7]. This advantage consists in the fact that compositional models express explicitly some marginals, whose computation in Bayesian network may be computationally expensive.

2 Notation and Basic Properties

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 $\{X_i\}_{i \in K}$.

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 used by Glenn Shafer [11] and Prakash Shenoy [12], $\pi^{\downarrow L}$. These symbols are used when we want to highlight the variables, for which the mar-

[1] Another solution can be found in [12].

ginal distribution is defined. If we want to specify variables which are to be deleted in the process of marginalization, we will use the symbol π^{-M}, where M is the set of indices of the variables, which do not appear among the arguments of the resulting marginal distribution. Thus, for $\pi(x_K)$ and $M = K \setminus L$, $\pi^{-M} = \pi^{\downarrow L}$.

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

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

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)

$$\forall x_M \in \times_{i\in M}\mathbf{X}_i \; (\kappa(x_M) = 0 \Longrightarrow \pi(x_M) = 0).$$

Since the outcome of the composition (if it is defined) is a new distribution, we can iteratively repeat the application of this operator composing 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* parameterized by the index set M we understand computation of the following distribution

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

The most basic properties of the introduced operators are expressed in the following assertions (for their proofs see [4] and [5]).

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 *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*

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

3 Perfect Sequence Models

Now we will focus our attention on marginalization of distributions given by a special subclass of generating sequences. ¿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 $\{X_i\}_{i\in K_j}$.

Definition 3 We call a generating sequence $\pi_1, \pi_2, \ldots, \pi_n$ *perfect* if for all $j = 2, \ldots, n$

$$(\pi_1 \triangleright \ldots \triangleright \pi_{j-1}) \triangleright \pi_j = \pi_j \triangleright (\pi_1 \triangleright \ldots \triangleright \pi_{j-1})$$

hold true.

Perfect sequences have a lot of pleasant properties, which are advantageous for multidimensional distributions representation. The most important one is expressed in the following assertion.

Theorem 1 *A generating sequence $\pi_1, \pi_2, \ldots, \pi_n$ is perfect* iff *all the distributions π_i are marginal to the represented distribution, i.e. for all $i = 1, 2, \ldots, n$*

$$(\pi_1 \triangleright \ldots \triangleright \pi_n)^{\downarrow K_i} = \pi_i.$$

Now, let us 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. The proof of the following assertion, which holds not only for perfect but for all generating sequences, can be found in [5].

Theorem 2 *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 $(1, 2, \ldots, n)$ such that $\ell \notin K_j$ for all $j \notin \{i_1, i_2, \ldots, i_m\}$. Then

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

where

$$\begin{aligned}
\kappa_j &= \pi_j, \qquad \forall j \notin \{i_1, i_2, \ldots, i_m\}, \\
\kappa_{i_1} &= \pi_{i_1}^{-\{\ell\}}, \\
\kappa_{i_2} &= (\pi_{i_1} \textcircled{\triangleright}_{L_{i_2-1}} \pi_{i_2})^{-\{\ell\}}, \\
\kappa_{i_3} &= (\pi_{i_1} \textcircled{\triangleright}_{L_{i_2-1}} \pi_{i_2} \textcircled{\triangleright}_{L_{i_3-1}} \pi_{i_3})^{-\{\ell\}}, \\
&\vdots \\
\kappa_{i_m} &= (\pi_{i_1} \textcircled{\triangleright}_{L_{i_2-1}} \pi_{i_2} \textcircled{\triangleright}_{L_{i_3-1}} \cdots \\
&\qquad\qquad \textcircled{\triangleright}_{L_{i_m-1}} \pi_{i_m})^{-\{\ell\}},
\end{aligned}$$

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 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*. However, its application can be computationally rather expensive. It is simple only in case when the variable to be delete is *boundary*, i.e. when this variable is contained among the arguments of only one of the distributions from a generating sequence in question. Then Theorem 2 simplifies into the following form.

Corollary *Let $\pi_1, \pi_2, \ldots, \pi_n$ be a generating sequence. If $\ell \in K_i$ for some $i \in \{1, \ldots, n\}$ and $\ell \notin K_j$ for all $j \neq i$ then*

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

More effective marginalizing procedures are however based on the following assertion, which was proven in [1]. But first let us define an auxiliary notion of a *reduction of a generating sequence*, which will simplify formulations in the following text.

Definition 4 *Let $\pi_1, \pi_2, \ldots, \pi_n$ be a generating sequence, $(j_1, j_2, \ldots, j_m)$ a subsequence of $\{1, 2, \ldots, n\}$ and $s \in Z = \{j_1, \ldots, j_m\}$ be such that*

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

Then we say that s and Z determine a reduction *of generating sequence $\pi_1, \ldots, \pi_n$ (or simply that (s, Z) is a reduction).*

Theorem 3 *Let $s \in Z$ and $Z = \{j_1, \ldots, j_m\}$ determine a reduction of a perfect sequence $\pi_1, \pi_2, \ldots, \pi_n$. Then, denoting*

$$\bar{L}_j = \bigcup_{i \in \{1, \ldots, j\} \setminus Z} K_i,$$

for all $j \notin Z$, marginal distribution $(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)^{\downarrow L}$ can be expressed

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

where

$$\begin{aligned}
\kappa_j &= \pi_j & \text{for} \quad & j \in Z, \\
\kappa_j &= \pi_s^{\downarrow K_s \cap \bar{L}_j} & \text{for} \quad & j \notin Z.
\end{aligned}$$

4 Marginalization Algorithm

In the current section, which is unfortunately rather technical, we present the main results of this contribution. Considering a perfect generating sequence $\pi_1(x_{K_1}), \pi_2(x_{K_2}), \ldots, \pi_n(x_{K_n})$ and a set of indices $L \subset (K_1 \cup K_2 \cup \ldots \cup K_n)$ we are going to describe an algorithm performing computation of a generating sequence $\kappa_1(x_{L_1}), \ldots, \kappa_m(x_{L_m})$, representing the required marginal distribution:

$$\begin{aligned}
&(\pi_1(x_{K_1}) \triangleright \pi_2(x_{K_2}) \triangleright \ldots \triangleright \pi_n(x_{K_n}))^{\downarrow L} \\
&\qquad = \kappa_1(x_{L_1}) \triangleright \kappa_2(x_{L_2}) \triangleright \ldots \triangleright \kappa_m(x_{L_m}).
\end{aligned}$$

For this we have to keep in mind that in practical situations distribution $\pi_1 \triangleright \ldots \triangleright \pi_n$ is defined for hundreds (if not thousands) of variables while the required marginal distribution

$\kappa_1 \triangleright \ldots \triangleright \kappa_m$ should have only small number of arguments (usually tens at maximum).

The complete algorithm is depicted in Figure 2 and will be described in Section 4.3. As the reader can see, the algorithm consists in (cyclical) employment of five procedures, four of which are quite simple (these are described in Section 4.1) and only one, to which whole Section 4.2 is devoted, is more complicated.

4.1 Simple Procedures

All the following four simple procedures, which are based on the theoretical properties introduced in the previous sections, somehow modify a generating sequence. Therefore, the considered generating sequence must be a part of an input, and a modified sequence must form an output of each of these procedures. So, when describing the procedures we assume that they have the sequence $\pi_1(x_{K_1}), \pi_2(x_{K_2}), \ldots, \pi_n(x_{K_n})$ available.

The first procedure realizes the simplest possible marginalization based on application of Lemma 1. Therefore, the procedure must be informed, which variables may be deleted. It is done through the index set L, which contains indices of all the variables that must be retained in the required marginal distribution.

Truncation of an unavailing tail $[L]$**.**

Find the smallest m, for which
$K_1 \cup \ldots \cup K_m \supseteq L$;
Delete distributions $\pi_{m+1}, \ldots, \pi_n$
from the input sequence.

The second procedure does not realize any marginalization but simplifies the respective generating sequence. The fact that this procedures does not change the respective multidimensional distribution follows immediately from Lemma 1.

Deletion of redundant elements.

Delete all π_j from the input
sequence, for which
$K_j \subseteq K_1 \cup \ldots \cup K_{j-1}$.

The following two procedures decrease dimension of the represented distribution by one. They marginalize one variable, namely X_j, out. Therefore, index j is an input parameter of these procedures. The correctness of these procedures follows from Theorem 2. Whilst General marginalization procedure may always be applied, Simple marginalization may be applied only when j is an element of only one K_j, i.e. when X_j is a boundary variable.

Simple marginalization $[j]$**.**

Find $\ell \in \{1, 2, \ldots, n\} : j \in K_\ell$;
Substitute π_ℓ with $\pi_\ell^{-\{j\}}$.

General marginalization $[j]$**.**

```
k ← 0;
FOR ℓ ∈ {1,...,n} DO {
   IF j ∈ K_ℓ THEN {
      Substitute π_ℓ with
      (κ_1 ⊳_{L_2} κ_2 ⊳_{L_3} ... ⊳_{L_k} κ_k ⊳_L π_ℓ)^{-{j}};
      k ← k+1;  κ_k ← π_ℓ;  L_k ← L
   }
   L ← L ∪ (K_j \ {j})
}.
```

4.2 Marginalization by means of Reduction

The procedures described in the previous subsection could decrease dimensionality of the multidimensional distribution either by one, or by more than one but only when the variables to be deleted appeared "in the tail" of the generating sequence. In this subsection we describe another possibility, which proves to be very efficient in many situations, especially when the number of the variables to be deleted is really high. This procedure realizes the situation described by Theorem 3. For this, one has to find a reduction (s, Z) such that Z contains all indices of the variables, for which the computed marginal distribution should be defined ($L \subseteq Z$). And it is this very search for reduction what makes the process rather complicated.

To find a reduction, the process employs sets $W(Z, j)$ (defined below) and their properties, which were proven in [1].

Having a set $Z \subset \{1, \ldots, n\}$ and $j \notin Z$ the symbol $W(Z, j)$ denotes the following subset of indices:

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

(notice that sets $W(Z,j)$ depend not only on Z and j but naturally also on the considered generating sequence).

Lemma 3 *If for nonempty $Z \subsetneq \{1,\ldots,n\}$ there exists $s \in Z$, for which $s \in \bigcap_{j \notin Z} W(Z,j)$, then s and Z determine a reduction (of the considered generating sequence).*

Lemma 4 *If for $j \notin Z \subsetneq \{1,\ldots,n\}$ ($Z \neq \emptyset$) $W(Z,j) \cap Z = \emptyset$ then for any reduction determining couple s and Z'*

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

The whole procedure of marginalization by means of reduction is described in Figure 1. It employs four simple steps we are now going to describe and which are theoretically supported by Theorem 3 and Lemmata 3 and 4. Quite naturally, all these procedures work with the generating sequence in question $\pi_1(x_{K_1}), \pi_2(x_{K_2}), \ldots, \pi_n(x_{K_n})$ (but only one of them - Marginalization $[s, Z']$ - modifies it).

The first procedure extends Z and computes all sets $W(Z,j)$. This means that Z must be in both input and output and $W(Z,j)$ in output of the procedure. The extension is based on Lemma 4, which, in case that $W(Z,j) = \{j\}$, says that all reductions coming into consideration are such, that the reduction set contains not only Z but also j, and therefore j is automatically added to Z[2].

[2]By application of an heuristic rule, j is added to Z also in some other situations, namely, when $W(Z,j) \cap Z = \emptyset$

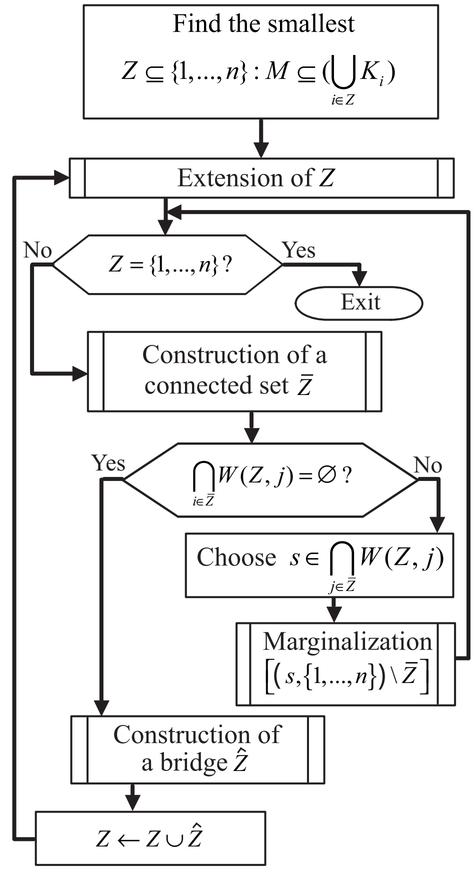

Figure 1: Marginalization by means of reduction

Extension of Z.

$Z^c \leftarrow \{1,\ldots,n\} \setminus Z$;
WHILE $Z^c \neq \emptyset$ DO {
  Choose any $j \in Z^c$;
  $W(Z,j) \leftarrow \left\{ s \in Z : \left(\bigcup_{i \in Z} K_i \right) \cap K_j \subseteq K_s \right\}$;
  IF $W(Z,j) = \emptyset$ THEN {
    $Z \leftarrow Z \cup \{j\}$;
    $Z^c \leftarrow \{1,\ldots,n\} \setminus Z$
  }
  ELSE { $Z^c \leftarrow Z^c \setminus \{j\}$
  }
}.

In [1] we have shown how sets $W(Z,j)$ can be used to find a required reduction. Lemma 3 says that we have already succeeded if there exists $s \in \bigcap_{j \notin Z} W(Z,j)$. We could do this test immediately when Z is received (by computing $W(Z,j)$). However, it is rather improbable that we would find such s for all $j \notin Z$

Therefore we place the test after a construction of a connected set $\bar{Z}$, which is a smallest potential set of indices of distributions, which has a chance to be deleted with the help of Theorem 3 (for more details, the reader is referred to [1]).

Construction of a connected set $\bar{Z}$.

```
Choose $j \notin Z$;
$\bar{L} \leftarrow \bigcup_{i\in\hat{n}} K_i \setminus \bigcup_{i\in Z} K_i$;  $\bar{Z} \leftarrow \{j\}$;
WHILE
$Z' \leftarrow \{k \notin Z \cup \bar{Z}: \exists \ell \in \bar{Z}: K_k \cap K_\ell \cap \bar{L} \neq \emptyset\} \neq \emptyset$
   DO {
     $\bar{Z} \leftarrow \bar{Z} \cup Z'$;
}.
```

If the distributions with indices from $\bar{Z}$ can be deleted from the generating sequence (i.e. if there exists $s \in \bigcap_{j \notin Z} W(Z,j)$) we perform the marginalization guaranteed by Theorem 3.

Marginalization $[s, Z']$.

```
$L' \leftarrow \emptyset$;
FOR $\ell \in \{1, \ldots, n\}$ DO {
    IF $\ell \notin Z'$ THEN {
      $L' \leftarrow L' \cup (K_s \cap K_\ell)$;
      Substitute $\pi_\ell$ with $\pi_s^{\downarrow L'}$
    }
}.
```

If the distributions with indices from $\bar{Z}$ cannot be deleted from the generating sequence (i.e. if there does not exists $s \in \bigcap_{j \notin Z} W(Z,j)$) we have to increase the set Z and continue with looking for a reduction with a greater Z. In case that there exist a couple $j, k \in \bar{Z}$ for which $W(Z,j) \cap W(Z,k) \cap Z = \emptyset$ then it is quite clear that both these indices must be added to Z (not to loose a chance of finding a reduction). If $K_j \cap K_k \cap \bar{L} = \emptyset$ then it can be shown that in the next step some other sets from $\bar{Z}$ would be added to Z. Therefore we add to Z not only j and k but also the smallest group of indices, which are found with the help of a shortest path in an auxiliary graphs.

In [1] we have shown that in rather special situations it can happen that there does not exist a couple $j, k \in \bar{Z}$ with

$$W(Z,j) \cap W(Z,k) \cap Z = \emptyset.$$

In this case we cannot use the above mentioned sophisticated rule and therefore we add to Z only the smallest group of indices, which cannot be avoided.

Construction of a bridge $\hat{Z}$.

```
$L \leftarrow \bigcup_{i\in Z} K_i$;
Search for a couple
$j, k \in \bar{Z} : W(Z,j) \cap W(Z,k) \cap Z = \emptyset$;
IF such a couple exists THEN {
    Construct a graph $G = (V, E)$:
    $V = \bar{Z}$
    $E = \{\ell_1, \ell_2 \in \bar{Z} : K_{\ell_1} \cap K_{\ell_2} \cap \bar{L} \neq \emptyset\}$;
    Assign to $\hat{Z}$ the nodes on the
    shortest path from $j$ to $k$;
}
ELSE {
    $\hat{Z} \leftarrow \{i \in \bar{Z} : K_i \cap L \neq \emptyset\}$;
}.
```

4.3 Algorithm

Up to now, we have commented all the procedures appearing in the scheme in Figure 2 describing the marginalization algorithm:

- Truncation of an unavailing tail,
- Deletion of redundant elements,
- Simple marginalization $[j]$,
- General marginalization $[j]$,
- Marginalization by means of reduction.

So, we could stop describing the marginalization algorithm if it were not for the decision block "Is the condition for reduction fulfilled". What is the purpose of this block? One of the purposes of this decision block is to prevent unsuccessful repetition of Marginalization by means of reduction. As we could see from the last subsection, it can happen that no reduction pair is found in this process. If this happens, there is no other way to get the required marginal distribution than to start application of the General marginalization process, which can always be applied.

And there is one additional purpose of the discussed decision block, which is, regarding

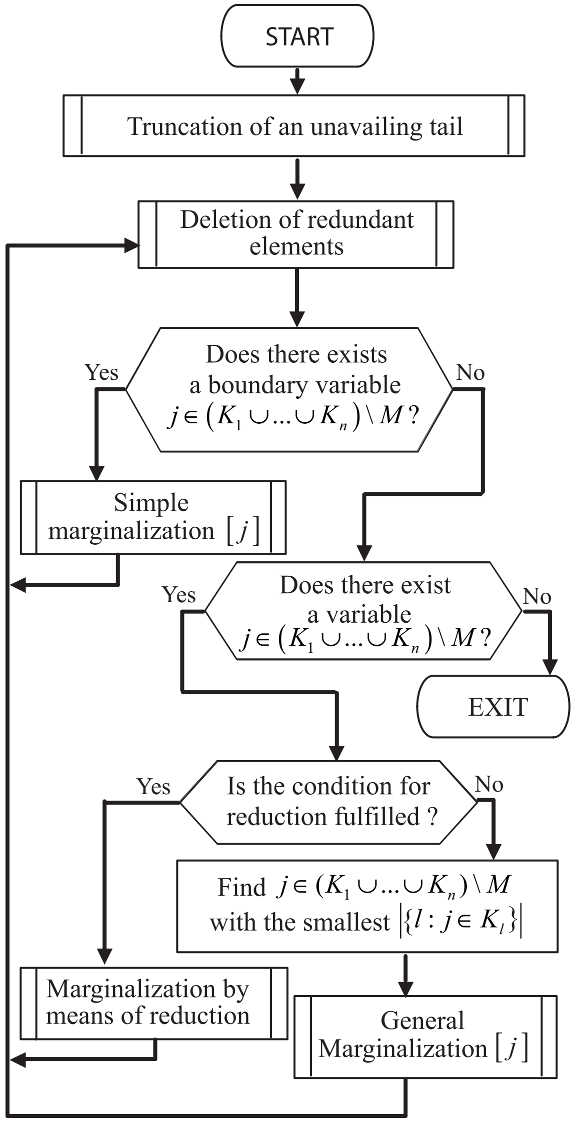

Figure 2: Marginalization algorithm

problems of computational optimality, even more important. The process of marginalization by reduction is so complicated, that it is quite clear that it does not pay off to apply it when only several variables are to be marginalized away. In such a case multiple application of "General marginalization" may be faster. Not speaking about the fact that chances to find a reduction, when only several superfluous variables have left, are rather small. The simplest realization of the discussed decision step is to set up a boundary for the number of variables to be deleted, for which the process of marginalization by reduction is employed. To find a more sophisticated realization of this step is a task for the experiments, which are currently performed. The results of them will be presented at the conference.

5 Conclusions

In this paper we have described an algorithm for marginalization in compositional models, more precisely, models represented by perfect sequences. The algorithm is based on theoretical properties proven in several assertions published in [1]. The algorithm is currently realized in the system MUDIM[3] and its efficiency is being tested on artificially generated data. The first results show that a proper realization of the decision step "Is the condition for reduction fulfilled?" is even more important than we expected.

As we do not have any theoretical support to answer a question whether there exists a suitable reduction in the considered compositional model, it may happen that we apply the process of marginalization by means of reduction even in situations when a reduction does not exist. Moreover, this algorithm is computationally expensive. So, we expected that this type of marginalization should be switched off for small models, and generally also when only a small number of variables are to be deleted. Therefore we were quite pleasantly surprised that this approach can increase computational efficiency even for small "laboratory" situations (see first two rows of Table 1).

In Table 1 we refer to computations with two models constructed for artificially generated data. Let us stress that it would be easy to construct a model, for which the reduction substantially decreases the computational time. Nevertheless, on purpose we are presenting examples which are, in a way, from this point of view inconvenient. They represent distributions of 24 and 100 variables. The other difference between the models is that in the first one there exists a reduction whilst no reduction can be found for the sec-

[3]The system is realized in **R** language and all the computations commented below were performed with AMD Athlon64 3000+, 1024MB RAM computer.

Table 1: Computational time in seconds

	reduction switched on	reduction switched off
Model 1 marginalization from 24 to 3 variables	0.12 s	0.69 s
Model 1 marginalization from 24 to 3 variables	0.53 s	0.61 s
Model 2 marginalization from 100 to 2 variables	12.45 s	1.67 s
Model 2 marginalization from 100 to 4 variables	29.41 s	3.06 s

ond model. The difference between the first two rows of Table 1 shows that the efficiency of the process does not depend only on the model – probability distribution, but also on which variables are to be marginalized out. This also explains the difference between the 3rd and 4th rows of the Table. The last one refers to the situation when we deliberately selected the most “inconvenient” variables for marginalization.

More experiments with both real and artificial data will be commented at the conference. By this time we also hope to introduce a heuristic rule realizing the critical decision step “Is the condition for reduction fulfilled?”

Acknowledgements

The research was supported by grants: GA AV ČR no. A2075302 and MŠMT no. 1M0572.

References

[1] Vl. Bína, R. Jiroušek: Marginalization in Multidimensional Compositional Models. Submitted for publication to Kybernetika.

[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: Composition of Probability Measures on Finite Spaces. In: Proc. of the 13th Conf. Uncertainty in Artificial Intelligence UAI'97 (D. Geiger, P.P. Shenoy eds.), Morgan Kaufmann Publ., 1997, pp. 274-281.

[5] 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.

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

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

[8] S. L. Lauritzen: Graphical Models. Clarendon Press, Oxford 1996.

[9] R. D. Shachter: Evaluating Influence Diagrams. Oper. Res. **34** (1986), 871–890.

[10] R. D. Shachter: Probabilistic inference and influence diagrams. Oper. Res. **36** (1988), 589–604.

[11] G. Shafer: Probabilistic Expert Systems. SIAM, Philadelphia 1996.

[12] T. Schnidt, P. P. Shenoy: Some improvements to the Shenoy-Shafer and Hugin architectures for computing marginals. Artificial Intelligence **102** (1998), 323-333.