# Belief Networks and Local Computations

Radim Jiroušek[1]

**Abstract** This paper is one of many attempts to introduce graphical Markov models within Dempster-Shafer theory of evidence. Here we take full advantage of the notion of factorization, which in probability theory (almost) coincides with the notion of conditional independence. In Dempster-Shafer theory this notion can be quite easily introduced with the help of the operator of composition.

Nevertheless, the main goal of this paper goes even further. We show that if a belief network (a D-S counterpart of a Bayesian network) is to be used to support decision, one can apply all the ideas of Lauritzen and Spiegelhalter's local computations.

**Key words:** Operator of composition, Factorization, Decomposable models, Conditioning.

## 1 Introduction

Graphical Markov models (GMM) [9], a technique which made computations with multidimensional probability distributions possible, opened doors for application of probabilistic methods to problem of practice. Here we have in mind especially application of the technique of local computations for which theoretical background was laid by Lauritzen and Spiegelhalter [10]. The basic idea can be expressed in a few words: a multidimensional distribution represented by a Bayesian network is first converted into a decomposable model which allows for efficient computation of conditional probabilities.

The goal of this paper is to show that the same ideas can be employed also within Dempster-Shafer theory of evidence [11].

Faculty of Management of University of Economics, and Institute of Information Theory and Automation, Academy of Sciences Czech Republic radim@utia.cas.cz

In this paper we consider a finite setting: space $\mathbf{X}_N = \mathbf{X}_1 \times \ldots \times \mathbf{X}_n$, and its subspaces (for $K \subseteq N$) $\mathbf{X}_K = \bigtimes_{i \in K} \mathbf{X}_i$. For a point $x = (x_1, \ldots, x_n) \in \mathbf{X}_N$ its projection into subspace $\mathbf{X}_K$ is denoted $x^{\downarrow K} = (x_{i, i \in K})$. Analogously, for $A \subseteq \mathbf{X}_N$, $A^{\downarrow K} = \{y \in \mathbf{X}_K : \exists x \in A, x^{\downarrow K} = y\}$. By a *join* of two sets $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_L$ we understand a set

$$A \otimes B = \{x \in \mathbf{X}_{K \cup L} : x^{\downarrow K} \in A \;\; \& \;\; x^{\downarrow L} \in B\}.$$

Notice that if $K$ and $L$ are disjoint, then $A \otimes B = A \times B$, if $K = L$ then $A \otimes B = A \cap B$.

In view of this paper it is important to realize that for $C \subseteq \mathbf{X}_{K \cup L}$, $C \subseteq C^{\downarrow K} \otimes C^{\downarrow L}$, and that the equality $C = C^{\downarrow K} \otimes C^{\downarrow L}$ holds only for some of them.

## 2 Basic Assignments

A *basic assignment* (ba) $m$ on $\mathbf{X}_K$ ($K \subseteq N$) is a function $m : \mathcal{P}(\mathbf{X}_K) \to [0, 1]$, for which

$$\sum_{\emptyset \neq A \subseteq \mathbf{X}_K} m(A) = 1.$$

If $m(A) > 0$, then $A$ is said to be a *focal element* of $m$. Recall that

$$Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B), \quad \text{and} \quad Pl(A) = \sum_{B \subseteq \mathbf{X}_K : B \cap A \neq \emptyset} m(B).$$

Having a ba $m$ on $\mathbf{X}_K$ one can consider its *marginal assignment* on $\mathbf{X}_L$ (for $L \subseteq K$), which is defined (for each $\emptyset \neq B \subseteq \mathbf{X}_L$):

$$m^{\downarrow L}(B) = \sum_{A \subseteq \mathbf{X}_K : A^{\downarrow L} = B} m(A).$$

**Definition 1 (Operator of composition).** For two arbitrary ba's $m_1$ on $\mathbf{X}_K$ and $m_2$ on $\mathbf{X}_L$ ($K \neq \emptyset \neq L$) a *composition* $m_1 \triangleright m_2$ is defined for each $C \subseteq \mathbf{X}_{K \cup L}$ by one of the following expressions:

[a] if $m_2^{\downarrow K \cap L}(C^{\downarrow K \cap L}) > 0$ and $C = C^{\downarrow K} \otimes C^{\downarrow L}$ then

$$(m_1 \triangleright m_2)(C) = \frac{m_1(C^{\downarrow K}) \cdot m_2(C^{\downarrow L})}{m_2^{\downarrow K \cap L}(C^{\downarrow K \cap L})};$$

[b] if $m_2^{\downarrow K \cap L}(C^{\downarrow K \cap L}) = 0$ and $C = C^{\downarrow K} \times \mathbf{X}_{L \setminus K}$ then

$$(m_1 \triangleright m_2)(C) = m_1(C^{\downarrow K});$$

[c] in all other cases $(m_1 \triangleright m_2)(C) = 0$.

Let us stress that the operator of composition is something other than the famous Dempster's rule of combination [2]. While Dempster's rule was designed to combine different (independent) sources of information (it realizes fusion of sources), the operator of composition was designed to assemble (compose) factorizing basic assignments from their pieces. Notice that, e.g., for computation of $(m_1 \triangleright m_2)(C)$ it suffices to know only the values of $m_1$ and $m_2$ for the respective projections of set $C$, whereas computing Dempster's combination of $m_1$ and $m_2$ for set $C$ requires knowledge of, roughly speaking, the entire basic assignments $m_1$ and $m_2$. This is an indisputable (computational) advantage of the factorization considered in this paper. Unfortunately, the operator of composition is neither commutative nor associative. In [8, 7] we proved a number of properties concerning the operator of composition; the following ones are the most important for the purpose of this paper.

**Proposition 1.** *Let $m_1$ and $m_2$ be ba's defined on $\mathbf{X}_K$, $\mathbf{X}_L$, respectively. Then:*

1. *$m_1 \triangleright m_2$ is a ba on $\mathbf{X}_{K\cup L}$;*
2. *$(m_1 \triangleright m_2)^{\downarrow K} = m_1$;*
3. *$m_1 \triangleright m_2 = m_2 \triangleright m_1 \quad \Longleftrightarrow \quad m_1^{\downarrow K\cap L} = m_2^{\downarrow K\cap L}$.*

From Property 1 one immediately gets that for basic assignments $m_1, m_2, \ldots, m_r$ defined on $\mathbf{X}_{K_1}, \mathbf{X}_{K_2}, \ldots, \mathbf{X}_{K_r}$, respectively, the formula $m_1 \triangleright m_2 \triangleright \ldots \triangleright m_r$ defines a (possibly multidimensional) basic assignment defined on $\mathbf{X}_{K_1\cup\ldots\cup K_r}$. However, to avoid ambiguity (recall that the operator is not associative) we have to say that, if not specified otherwise by parentheses, the operators will always be applied from left to right, i.e.,

$$m_1 \triangleright m_2 \triangleright \ldots \triangleright m_r = (\ldots(m_1 \triangleright m_2) \triangleright \ldots \triangleright m_{r-1}) \triangleright m_r.$$

Nevertheless, when designing the process of local computations for compositional models in D-S theory, which is intended to be an analogy to the process proposed by Lauritzen and Spiegelhalter in [10], one needs a type of associativity (see also [12]) expressed in the following assertion proved in [6].

**Proposition 2.** *Let $m_1, m_2$ and $m_3$ be ba's on $\mathbf{X}_{K_1}, \mathbf{X}_{K_2}$ and $\mathbf{X}_{K_3}$, respectively, such that $K_2 \supseteq K_1 \cap K_3$, and*

$$m_1^{\downarrow K_1\cap K_2}(C^{\downarrow K_1\cap K_2}) > 0 \Longrightarrow m_2^{\downarrow K_1\cap K_2}(C^{\downarrow K_1\cap K_2}) > 0.$$

*Then $(m_1 \triangleright m_2) \triangleright m_3 = m_1 \triangleright (m_2 \triangleright m_3)$.*

### *Belief Networks and Decomposable Models*

In this subsection we introduce a Dempster-Shafer counterpart to GMM's. Studying properly probabilistic GMM's one can realize that it is the notion of *factorization* that makes it possible to represent multidimensional probability distributions efficiently. Focusing only on Bayesian networks one can see that they can be defined in probability theory in several different ways. Here we will proceed according to a rather theoretical approach which defines a Bayesain network as a probability distribution factorizing with respect to a given *acyclic directed graph* (DAG). The factorization guarantees that the independence structure of a probability distribution represented by a Bayesian network is in harmony with the well-known *d-separation criterion* [4, 9].

For Bayesian networks, this factorization principle can be formulated in the following way (here $pa(i)$ denotes the set of parents of a node $i$ of the considered DAG, and $fam(i) = pa(i) \cup \{i\}$): measure $\pi$ is a Bayesian network with a DAG $G = (N, E)$ if for each $i = 2, \ldots, |N|$ (assuming that this ordering of nodes is such that $k \in pa(j) \Longrightarrow k < j$) marginal distribution $\pi^{\downarrow\{1,2,\ldots,i\}}$ factorizes with respect to couple $(\{1, 2, \ldots, i-1\}, fam(i))$. And this is the definition which can be directly taken over into Dempster-Shafer theory.

**Definition 2 (Belief network).** We say that a ba $m$ is a *belief network* (BN) with a DAG $G = (N, E)$ if for each $i = 2, \ldots, |N|$ (assuming the enumeration meets the property that $k \in pa(j) \Longrightarrow k < j$) marginal ba $m^{\downarrow\{1,2,\ldots,i\}}$ factorizes in the following sense: $m^{\downarrow\{1,2,\ldots,i\}} = m^{\downarrow\{1,2,\ldots,i-1\}} \triangleright m^{\downarrow fam(i)}$.

From this definition, which differs from those used in [3, 12], we immediately get the following description of a BN.

**Proposition 3 (Closed form for BN).** Let $G = (N, E)$ be a DAG, and $1, 2, \ldots, |N|$ be its nodes ordered in the way that parents are before their children. Ba $m$ is a BN with graph $G$ if and only if

$$m = m^{\downarrow fam(1)} \triangleright m^{\downarrow fam(2)} \triangleright \ldots \triangleright m^{\downarrow fam(|N|)}.$$

Taking advantage of the notion of factorization which is based on the operator of composition, we can also introduce decomposable ba's. In harmony with decomposable probability distribution, decomposable ba's are defined as those factorizing with respect to *decomposable graphs*, i.e. undirected graphs whose *cliques* (maximal sets of nodes inducing complete subgraphs) $C_1, C_2, \ldots, C_r$ can be ordered to meet the so-called *running intersection property* (RIP): for all $i = 2, \ldots, r$ there exists $j, 1 \le j < i$, such that $K_i \cap (K_1 \cup \ldots \cup K_{i-1}) \subseteq K_j$.

**Definition 3 (Decomposable ba).** Consider a decomposable graph $G = (N, F)$ with cliques $C_1, C_2, \ldots, C_r$ and assume the cliques are ordered to meet RIP. We say that a ba $m$ is *decomposable* (Dba) with respect to $G = (N, F)$ if for each $i = 2, \ldots, r$ marginal ba $m^{\downarrow C_1 \cup \ldots \cup C_i}$ factorizes in the following sense:

$$m^{\downarrow C_1 \cup \ldots \cup C_i} = m^{\downarrow C_1 \cup \ldots \cup C_{i-1}} \triangleright m^{\downarrow C_i}.$$

Analogously to the closed form for a BN we get also closed form for Dba, which is again an immediate consequence of the definition.

**Proposition 4 (Closed form for Dba).** Let $G = (N, F)$ be decomposable with cliques $C_1, C_2, \ldots, C_r$ and assume the cliques are ordered to meet RIP. Ba $m$ is decomposable with respect to $G$ if and only if

$$m = m^{\downarrow C_1} \triangleright m^{\downarrow C_2} \triangleright \ldots \triangleright m^{\downarrow C_r}.$$

### *Conditioning*

Unfortunately, there is no generally accepted way of conditioning in D-S theory. Though we do not have an ambition to fill in this gap, we need a tool which will enable us to answer questions like: *What is a belief for values of variable $X_j$ if we know that variable $X_i$ has a value $a$?* In probability theory the answer is given by conditional probability distribution $\pi(X_j|X_i = a)$. Let us study a possibility to obtain this conditional distribution with the help of the probabilistic operator of composition[1].

Define a *degenerated* one-dimensional probability distribution $\kappa_{|i;a}$ as a distribution of variable $X_i$ achieving probability 1 for value $X_i = a$, i.e.,

$$\kappa_{|i;a}(X_i = x) = \begin{cases} 1 & \text{if } x = a, \\ 0 & \text{otherwise.} \end{cases}$$

Now, compute $(\kappa_{|i;a} \triangleright \pi)^{\downarrow\{j\}}$ for a probability distribution $\pi$ of variables $X_K$ with $i, j \in K$:

$$(\kappa_{|i;a} \triangleright \pi)^{\downarrow\{j\}}(y) = ((\kappa_{|i;a} \triangleright \pi)^{\downarrow\{j,i\}})^{\downarrow\{j\}}(y) = (\kappa_{|i;a} \triangleright \pi^{\downarrow\{j,i\}})^{\downarrow\{j\}}(y)$$
$$= \sum_{x \in \mathbf{X}_i} \frac{\kappa_{|i;a}(x) \cdot \pi^{\downarrow\{j,i\}}(y, x)}{\pi^{\downarrow\{i\}}(x)} = \frac{\pi^{\downarrow\{j,i\}}(y, a)}{\pi^{\downarrow\{i\}}(a)} = \pi^{\downarrow\{j,i\}}(y|a).$$

Using an analogy, we consider in this paper that a proper answer to the above-raised question, in a situation when ba $m$ is taken into consideration, is given by $(m_{|i;a} \triangleright m)^{\downarrow\{j\}}$ (or rather by the corresponding $Bel$ function), where $m_{|i;a}$ is a ba on $\mathbf{X}_i$ with only one focal element $m(\{a\}) = 1$. This idea is moreover supported by the semantics of $m_{|i;a}$; this ba expresses the fact that we are *sure* that variable $X_i$ takes the value $a$. Therefore $m_{|i;a} \triangleright m$ is a ba arising from $m$ by enforcing it to have a marginal for variable $X_i$ that is equal to $m_{|i;a}$ (see Property 2 of Proposition 1). In other words it describes the relationships among all variables from $X_N$ which is encoded in $m$, when we know that $X_i$ takes value $a$.

---

[1] In probability theory the operator of composition is defined for distributions $\pi(X_K)$ and $\kappa(X_L)$, for which $\pi^{\downarrow K\cap L}$ is absolutely continuous with respect to $\kappa^{\downarrow K\cap L}$, for each $x \in \mathbf{X}_{L\cup K}$ by the formula

$$(\pi \triangleright \kappa)(x) = \frac{\pi(x^{\downarrow K})\kappa(x^{\downarrow L})}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})}.$$

For the precise definition and its properties see [5]

## 3 Local Computations

As said in Introduction, by *local computations* we understand a realization of the ideas published by Lauritzen and Spiegelhalter [10]. They proposed to compute a conditional probability (as for example $\pi(X_d|X_i = a, X_j = b, X_k = c))$ for a distribution $\pi$ represented in a form of a Bayesian network in the following two steps.

1. Bayesian network is transformed into a decomposable model representing the same probability distribution $\pi$;
2. the required conditional distribution is computed by a process consisting of computations with the marginal distributions corresponding to the cliques of the respective decomposable graph.

This means that to get the desired conditional distribution one needs to know only the structure of the decomposable models (e.g. the respective decomposable graph) and the respective system of marginal distributions.

And it is the goal of this section to show that practically the same computational process can be realized also in D-S theory.

#### *Conversion of a BN into Dba*

The process realizing this step can be directly taken over from probability theory [4]. If $G = (N, E)$ is a DAG of some belief network, then undirected graph $G = (N, \bar{E})$, where

$$\bar{E} = \left\{ \{i, j\} \in \binom{N}{2} : \exists k \in N \ \{i, j\} \subseteq fam(k) \right\},$$

is a so-called *moral graph* from which one can get the necessary decomposable graph $G = (V, F)$ (which will be uniquely specified by a system of its cliques $C_1, C_2, \ldots, C_r$) by any heuristic approach used for moral graph triangulation [1] (it is known that the process of looking for an optimal triangulated graph is a NP hard problem). Then it is an easy task to compute the necessary marginal ba's $m^{\downarrow C_1}, \ldots, m^{\downarrow C_r}$ when one realizes that there must exist an ordering (let it be the ordering $C_1, C_2, \ldots, C_r$) of the cliques meeting RIP and simultaneously

$$i \in pa(j) \implies f(i) \leq f(j),$$

where $f(k) = \min(\ell : k \in C_\ell)$.

#### *Computation of Conditional ba*

In comparison with the previous step, this computational process is much more complex. We have to show that having a decomposable ba $m = m^{\downarrow C_1} \triangleright \ldots \triangleright m^{\downarrow C_r}$ one can compute $(m_{|i;a} \triangleright m)^{\downarrow \{j\}}$ locally.

For this, we take advantage of the famous fact that if $C_1, C_2, \ldots, C_r$ can be ordered to meet RIP, then for each $k \in \{1, 2, \ldots, r\}$ there exists an ordering meeting RIP for which $C_k$ is the first one. So consider any $C_k$ for which

$i \in C_k$, and find the ordering meeting RIP which starts with $C_k$. Without loss of generality let it be $C_1, C_2, \ldots, C_r$ (so, $i \in C_1$).

Considering ba $m$ decomposable with respect to a graph with cliques $C_1, C_2, \ldots, C_r$, our goal is to compute

$$(m_{|i;a} \triangleright m)^{\downarrow\{j\}} = \left(m_{|i;a} \triangleright (m^{\downarrow C_1} \triangleright m^{\downarrow C_2} \triangleright \ldots \triangleright m^{\downarrow C_r})\right)^{\downarrow\{j\}}.$$

However, at this moment we have to assume that $m^{\downarrow\{i\}}(\{a\})$ is positive. Under this assumption we can apply Proposition 2 $r-1$ times getting

$$\begin{aligned} m_{|i;a} \triangleright (m^{\downarrow C_1} \triangleright m^{\downarrow C_2} \triangleright \ldots \triangleright m^{\downarrow C_r}) \\ &= m_{|i;a} \triangleright (m^{\downarrow C_1} \triangleright m^{\downarrow C_2} \triangleright \ldots \triangleright m^{\downarrow C_{r-1}}) \triangleright m^{\downarrow C_r} \\ &= \ldots = m_{|i;a} \triangleright m^{\downarrow C_1} \triangleright m^{\downarrow C_2} \triangleright \ldots \triangleright m^{\downarrow C_r}, \end{aligned}$$

from which computationally local process[2]

$$\begin{aligned} \bar{m}_1 &= m_{|i;a} \triangleright m^{\downarrow C_1}, \\ \bar{m}_2 &= \bar{m}_1^{\downarrow C_2 \cap C_1} \triangleright m^{\downarrow C_2}, \\ \bar{m}_3 &= (\bar{m}_1 \triangleright \bar{m}_2)^{\downarrow C_3 \cap (C_1 \cup C_2)} \triangleright m^{\downarrow C_3}, \\ &\vdots \\ \bar{m}_r &= (\bar{m}_1 \triangleright \ldots \triangleright \bar{m}_{r-1})^{\downarrow C_r \cap (C_1 \cup \ldots C_{r-1})} \triangleright m^{\downarrow C_r}, \end{aligned}$$

yields a sequence $\bar{m}_1, \ldots, \bar{m}_r$, for which $m_{|i;a} \triangleright m = \bar{m}_1 \triangleright \ldots \triangleright \bar{m}_r$, and each $\bar{m}_k = (m_{|i;a} \triangleright m)^{\downarrow C_k}$. Therefore, to compute $(m_{|i;a} \triangleright m)^{\downarrow\{j\}}$ it is enough to find any $k$ such that $j \in C_k$ because in this case $(m_{|i;a} \triangleright m)^{\downarrow\{j\}} = \bar{m}_k^{\downarrow\{j\}}$.

This simple idea can be quite naturally generalized in the following sense. Considering a model with basic assignment $m$ and having a prior information about values of variables $X_{i_1} = a_1, \ldots, X_{i_t} = a_t$, the goal may be to compute

$$(m_{|i_1,\ldots,i_t;a_1,\ldots,a_t} \triangleright m)^{\downarrow\{j\}} = (m_{|i_1;a_1} \triangleright \ldots \triangleright m_{|i_t;a_t} \triangleright m)^{\downarrow\{j\}}.$$

It can be done easily just by repeating the described computational process as many times as the number of given values (in our case $t$). This is possible because ba $m_{|i_1;a_1} \triangleright m = \bar{m}_1 \triangleright \ldots \triangleright \bar{m}_r$ is again decomposable and therefore ba's $\bar{m}_1, \ldots, \bar{m}_r$ can be again reordered so that the respective sequence of index sets meets RIP and index $i_2$ belongs to the first index set, and so on. However, and it is important to stress it, in this case we have to assume that the combination of given values, which specifies the condition, is a focal element of ba $m$, i.e., regarding the condition specified above, we have to assume that $m^{\downarrow\{i_1,\ldots,i_t\}}(\{a_1, \ldots, a_t\}) > 0$.

[2] Notice that due to the assumption that $C_1, \ldots, C_r$ meets RIP, for each $k$ there exists $\ell$ such that $(\bar{m}_1 \triangleright \ldots \triangleright \bar{m}_{k-1})^{\downarrow C_k \cap (C_1 \cup \ldots C_{k-1})} = \bar{m}_\ell^{\downarrow C_k \cap (C_1 \cup \ldots C_{k-1})}$, which ensures locality of the described computations.

# 4 Conclusions

In the paper we have shown that with the help of the operator of composition it is possible to define BN's as a D-S counterpart of Bayesian networks. Moreover, we have shown that under the assumption that a given condition is a focal element of a ba represented by a BN, one can realize a process yielding a basic assignment representing a *conditional belief*. This computational process can be performed locally, i.e., all the computations involves only marginal distributions of the respective ba. The only weak point of the presented approach is that it can be applied only under an additional assumption requiring that the prior information specifying the condition is a focal element of the ba represented by the given BN.

**Acknowledgements** This work was supported by GAČR under the grants ICC/08/E010, and 201/09/1891, and by MŠMT ČR under grants 1M0572 and 2C06019.

# References

1. Cano, A., Moral, S.: Heuristic algorithms for the triangulation of graphs. In: Bouchon-Meunier B, Yager RR, Zadeh LA (eds) Advances in Intelligent Computing IPMU '94. Paris, France (1995)
2. Dempster, A.: Upper and lower probabilities induced by a multi-valued mapping. Ann. of Math. Stat. 38, 325–339 (1967)
3. Dempster, A. P., Kong, A.: Uncertain evidence and artificial analysis. J. Stat. Planning and Inference 20, 355-368 (1988)
These are very powerful models for representing and manipulating belief functions in multidimensional spaces.
4. Jensen, F.V.,: Bayesian Networks and Decision Graphs. IEEE Computer Society Press, New York (2001)
5. Jiroušek, R.: Composition of probability measures on finite spaces. In: Geiger D, Shenoy PP (eds) Proc. of the 13th Conf. Uncertainty in Artificial Intelligence UAI'97. Morgan Kaufmann Publ., San Francisco, California (1997)
6. Jiroušek, R.: A note on local computations in Dempster-Shafer theory of evidence. Accepted to: 7th Symposium on Imprecise Probabilities and Their Applications, Innsbruck (2011)
7. Jiroušek, R., Vejnarová, J., Compositional Models and Conditional Independence in Evidence Theory. Int. J. Approx. Reason. 52, 316-334 (2011)
8. Jiroušek, R., Vejnarová, J., Daniel, M.: Compositional models of belief functions. In: de Cooman G, Vejnarová J, Zaffalon M (eds) Proc. of the 5th Symposium on Imprecise Probabilities and Their Applications, Praha, (2007)
9. Lauritzen, S.L.: Graphical models. Oxford University Press (1996)
10. Lauritzen, S.L., Spiegelhalter, D.J.: Local computation with probabilities on graphical structures and their application to expert systems. J. of Royal Stat. Soc. series B 50, 157–224 (1988)
11. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton, New Jersey (1976)
12. Shenoy, P.P., Shafer, G.: Axioms for probability and belief-function propagation. In: Shachter RD, Levitt T, Lemmer JF, Kanal LN (eds) Uncertainty in Artificial Intelligence 4, North-Holland, (1990)