Is It Possible to Define Graphical Models in Dempster-Shafer Theory of Evidence?

Radim Jiroušek
Institute of Information Theory and Automation, Academy of Sciences
and
Faculty of Management, University of Economics
Jarošovská 1117/ II, 377 01 Jindřichův Hradec
Czech Republic
radim@utia.cas.cz

Abstract

The paper answers the question raised in the title only partially. We show that one can quite easily introduce graphical models corresponding to triangulated (decomposable) graphs, i.e. decomposable models. Analyzing all graphical 4-dimensional models we show that representation of decomposable models is very efficient and, simultaneously, that cycles in graphs can bring problems to which we do not bring a solution.

Introduction

When applying AI models to problems of practice one has always to cope with what is often called "curse of multidimensionality". Only exceptionally one can manage with tens of variables, usually one needs hundreds or even thousands of them. This is why a number of models have been developed in probability theory enabling efficient representation and processing multidimensional probability distributions; let us mention above all Graphical Markov Models. There are several special classes coming under this heading: the class of *graphical models* (in the sense the term was used by the authors of pioneer papers as e.g. (Daroch, Lauritzen and Speed 1980, Edwards and Havránek 1985)) is the oldest, *Bayesian networks* are perhaps the most popular, and *decomposable models* are the most efficient.

Considering only the binary case (i.e. all the considered variables achieve only two values) one needs 2^n numbers to represent a general probability distribution. The problem of multidimensionality becomes even more alarming when considering models within Dempster-Shafer theory of evidence, where one needs (in the binary case) $2^{(2^n)}$ numbers to represent a general n-dimensional model. So it is quite evident that one has to focus to models that can be represented by a much smaller number of parameters.

In this paper we are going to introduce *graphical models*, and especially their proper subfamily *decomposable models*, within Dempster-Shafer theory.

Basic Notions

Set notation

In the whole paper we will deal with a finite number of variables $X_1, X_2, \ldots, X_n$ each of which is specified by a finite set $\mathbf{X}_i$ of its values. So, we will consider multidimensional space of discernment

$$\mathbf{X}_N = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_n,$$

and its *subspaces*. For $K \subset N = \{1, 2, \ldots, n\}$, $\mathbf{X}_K$ denotes a Cartesian product of those $\mathbf{X}_i$, for which $i \in K$:

$$\mathbf{X}_K = \times_{i \in K} \mathbf{X}_i.$$

A *projection* of $x = (x_1, x_2, \ldots, x_n) \in \mathbf{X}_N$ into $\mathbf{X}_K$ will be denoted $x^{\downarrow K}$, i.e. for $K = \{i_1, i_2, \ldots, i_\ell\}$

$$x^{\downarrow K} = (x_{i_1}, x_{i_2}, \ldots, x_{i_\ell}) \in \mathbf{X}_K.$$

Analogously, for $K \subset L \subseteq N$ and $A \subset \mathbf{X}_L$, $A^{\downarrow K}$ will denote a *projection* of A into $\mathbf{X}_K$:

$$A^{\downarrow K} = \{y \in \mathbf{X}_K : \exists x \in A \ \ (y = x^{\downarrow K})\}.$$

Let us remark that we do not exclude situations when $K = \emptyset$. In this case $A^{\downarrow \emptyset} = \emptyset$.

In addition to the projection, in this text we will need also the opposite operation which will be called a join. By a *join* of two sets $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_L$ we will 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 the join of the corresponding sets is just their Cartesian product

$$A \otimes B = A \times B.$$

For $K = L$, $A \otimes B = A \cap B$. If $K \cap L \neq \emptyset$ and $A^{\downarrow K \cap L} \cap B^{\downarrow K \cap L} = \emptyset$ then also $A \otimes B = \emptyset$.

In view of this paper it is important to realize that if $x \in C \subseteq \mathbf{X}_{K \cup L}$, then $x^{\downarrow K} \in C^{\downarrow K}$ and $x^{\downarrow L} \in C^{\downarrow L}$, which means that always

$$C \subseteq C^{\downarrow K} \otimes C^{\downarrow L}.$$

However, it does not mean that $C = C^{\downarrow K} \otimes C^{\downarrow L}$. For example, considering 3-dimensional frame of discernment $\mathbf{X}_{\{1,2,3\}}$ with $\mathbf{X}_i = \{a_i, \bar{a}_i\}$ for all three $i = 1, 2, 3$, and $C = \{a_1a_2a_3, \bar{a}_1a_2a_3, a_1a_2\bar{a}_3\}$ one gets

$$\begin{aligned} C^{\downarrow\{1,2\}} \otimes C^{\downarrow\{2,3\}} &= \{a_1a_2, \bar{a}_1a_2\} \otimes \{a_2a_3, a_2\bar{a}_3\} \\ &= \{a_1a_2a_3, \bar{a}_1a_2a_3, a_1a_2\bar{a}_3, \bar{a}_1a_2\bar{a}_3\} \neq C. \end{aligned}$$

Assignment notation

The role of a probability distribution from a probability theory is in Dempster-Shafer theory played by any of the set functions: belief function, plausibility function or basic (*probability or belief*) assignment. Knowing one of them, one can deduce the two remaining. In this paper we will use exclusively basic assignments.

A *basic assignment* m on $\mathbf{X}_K$ ($K \subseteq N$) is a function

$$m : \mathcal{P}(\mathbf{X}_K) \longrightarrow [0,1],$$

for which

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

For the sake of this paper it is reasonable to consider only normalized basic assignments, for which $m(\emptyset)$ equals always 0. If $m(A) > 0$, then A is said to be a *focal element* of m.

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

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

Basic assignment m is said to be *Bayesian* if all its focal elements are *singletons* (i.e. $m(A) > 0 \implies |A| = 1$). In this case it is easy to show that belief and plausibility functions coincide, and correspond to a probability measure on $\mathbf{X}_N$.

Operator of composition

Compositional models were introduced for probability theory in (Jiroušek 1997) as an alternative to Bayesian networks for efficient representation of multidimensional measures. They were based on a recurrent application of the operator of composition. Analogous operator within the framework of Dempster-Shafer theory was introduced in (Jiroušek, Vejnarová and Daniel 2007).

Definition 1 ***Operator of Composition.*** For two arbitrary basic assignments 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 $(m_1 \triangleright m_2)(C) = 0$.

First of all we want to stress that the operator of composition is something else than the famous Dempster's rule of combination (Dempster 1967). For example it is (in contrary to Dempster's rule) neither commutative nor associative. In (Jiroušek, Vejnarová and Daniel 2007) we proved a number of properties concerning the operator of composition. In view of the forthcoming text those presented in the following assertion are the most important.

Theorem 1 ***Basic Properties.*** *Let m_1 and m_2 be basic assignments defined on $\mathbf{X}_K, \mathbf{X}_L$, respectively. Then:*

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

Though it is not a topic of this paper, let us mention that analogously to probability theory, one can easily introduce models equivalent to Bayesian networks also in Dempster-Shafer theory with the help of the operator of composition.

Conditional Independence

Let us stress at the very beginning of this section that for the purpose of this paper we cannot make do with the notion of conditional independence that is used by most of the other authors. Because of this, we will start discussing a generally accepted notion of unconditional independence (Studený 1993) (some authors call it *marginal* independence (Shenoy 1994), non-interactivity (Klir, G. J. 2006), marginal non-interactivity (Ben Yaghlane, Smets and Mellouli 2002a), independence in random sets (Couso, Moral and Walley 1999), etc.). In the cited papers the term was introduced either with the help of the conjunctive combination rule, or with the help of the commonality function. It was showed in (Jiroušek and Vejnarová 2010) that the following definition is fully equivalent.

Definition 2 ***Unconditional Independence.*** Let m be a basic assignment on $\mathbf{X}_N$ and $K, L \subset N$ be nonempty disjoint. We say that groups of variables X_K and X_L are *independent* with respect to basic assignment m (in notation $K \perp\!\!\!\perp L\ [m]$) if for all $A \subseteq \mathbf{X}_{K\cup L}$

$$m^{\downarrow K\cup L}(A) = \begin{cases} m^{\downarrow K}(A^{\downarrow K}) \cdot m^{\downarrow L}(A^{\downarrow L}) & \text{if } A = A^{\downarrow K} \times A^{\downarrow L}, \\ 0 & \text{otherwise.} \end{cases}$$

Notice that this notion possesses the most important property of independence from probability theory:

If $K \perp\!\!\!\perp L\ [m]$ then m is uniquely given by $m^{\downarrow K}$ and $m^{\downarrow L}$.

There are several generalizations of the notion of independence from Definition 2 to *conditional independence*. The most frequently used is the following one, which is called *factorization* by Shenoy (1994) and *conditional non-interactivity* by Ben Yaghlane, Smets and Mellouli (2002b).

Definition 3 ***Conditional Non-interactivity.*** Let m be a basic assignment on $\mathbf{X}_N$ and $K, L, M \subset N$ be disjoint (K and L nonempty). We say that groups of variables X_K and X_L are *conditionally non-interactive given* X_M with respect to basic assignment m (denoted $K \perp_m L|M$) if for all $A \subseteq \mathbf{X}^{\downarrow K\cup L\cup M}$

$$m^{\downarrow K\cup L\cup M}(A) \circledcap m^{\downarrow M}(A^{\downarrow M}) = m^{\downarrow K\cup M}(A^{\downarrow K\cup M}) \circledcap m^{\downarrow L\cup M}(A^{\downarrow L\cup M}),$$

where $\circledcap$ denotes the *conjunctive combination rule*, i.e. *non-normalized Depmpster's rule of combination.*

Nevertheless, this notion does not help us to define graphical models because nothing like *factorization lemma* (see Theorem 2 below) can be proved for conditional non-interactivity. Moreover, as it was showed by Studený, when the notion from Definition 2 is used, then it can happen that for two consistent overlapping basic assignments there does not exist their common extension with the required conditional non-interactivity (for the Studený's example see (Ben Yaghlane, Smets and Mellouli 2002b) or (Jiroušek and Vejnarová 2010)). It means that the notion of conditional non-interactivity does not support multidimensional model construction. Therefore, in this paper we will use the notion of conditional independence, which was introduced in (Jiroušek 2007, Jiroušek and Vejnarová 2010), and which differs from the notion of conditional non-interactivity.

Definition 4 ***Conditional Independence.*** Let m be a basic assignment on $\mathbf{X}_N$ and $K, L, M \subset N$ be disjoint, $K \neq \emptyset \neq L$. We say that groups of variables X_K and X_L are *conditionally independent given* X_M *with respect to* m (and denote it by $K \perp\!\!\!\perp L|M\,[m]$), if for any $A \subseteq \mathbf{X}_{K\cup L\cup M}$ such that $A = A^{\downarrow K\cup M} \otimes A^{\downarrow L\cup M}$ the equality

$$m^{\downarrow K\cup L\cup M}(A) \cdot m^{\downarrow M}(A^{\downarrow M}) \\ = m^{\downarrow K\cup M}(A^{\downarrow K\cup M}) \cdot m^{\downarrow L\cup M}(A^{\downarrow L\cup M})$$

holds true, and $m^{\downarrow K\cup L\cup M}(A) = 0$ for all the remaining $A \subseteq \mathbf{X}_{K\cup L\cup M}$, for which $A \neq A^{\downarrow K\cup M} \otimes A^{\downarrow L\cup M}$.

It is not difficult to show that for $M = \emptyset$ both conditional non-interactivity and conditional independence coincide with the notion of unconditional independence from Definition 2. It was also showed by Ben Yaghlane, Smets and Mellouli (Theorem 5.3 in (2002b)) that if $K \perp_m L|M$ then for all focal elements A of m

$$A = A^{\downarrow K\cup M} \otimes A^{\downarrow L\cup M}.$$

Moreover, both conditional non-interactivity and conditional independence meet all the following *semigraphoid axioms* (presented here for conditional independence (Pearl 1988, Studený 1993, Lauritzen 1996, Studený 2002)):

(A1) $K \perp\!\!\!\perp L \,|\, M\,[m] \implies L \perp\!\!\!\perp K \,|\, M\,[m]$

(A2) $K \perp\!\!\!\perp L\cup M \,|\, J\,[m] \implies K \perp\!\!\!\perp M \,|\, J\,[m]$

(A3) $K \perp\!\!\!\perp L\cup M \,|\, J\,[m] \implies K \perp\!\!\!\perp L \,|\, M\cup J\,[m]$

(A4) $(K \perp\!\!\!\perp L \,|\, M\cup J\,[m])$ & $(K \perp\!\!\!\perp M \,|\, J\,[m]) \implies K \perp\!\!\!\perp L\cup M \,|\, J\,[m]$

The main difference between our notion of conditional independence and that of conditional non-interactivity is in the fact that

- our concept does not suffer from the *inconsistency with marginalization*: for two consistent basic assignments m_1 and m_2 (over $\mathbf{X}_K$ and $\mathbf{X}_L$, respectively) one can always take their composition $m_1 \triangleright m_2$, which, due to Theorem 1, is an extension of both m_1 and m_2 and
$$K\setminus L \perp\!\!\!\perp L\setminus K| \, K\cap L\,[m_1 \triangleright m_2];$$
- for our notion, the Dempster-Shafer's counterpart of the probabilistic factorization lemma has been proved in (Jiroušek 2010); see Theorem 2 below.

Both these properties are important from the point of view of multidimensional model representation. They express the fact that validity of a conditional independence relation means that the respective basic assignment may be defined with a smaller number of parameters (either just by its marginals, or with the "factor functions" as expressed by Theorem 2).

The reader interested in a more thorough comparison od the notions of conditional independence and conditional non-interactivity (illustrated with examples) is referred to (Vejnarová 2010).

Definition 5 ***Simple Factorization.*** Consider two nonempty sets $K \cup L = N$. We say that basic assignment m *factorizes with respect to* (K, L) if there exist two nonnegative set functions

$$\phi : \mathcal{P}(\mathbf{X}_K) \longrightarrow [0, +\infty), \quad \psi : \mathcal{P}(\mathbf{X}_L) \longrightarrow [0, +\infty),$$

such that for all $A \subseteq \mathbf{X}_{K\cup L}$

$$m(A) = \begin{cases} \phi(A^{\downarrow K}) \cdot \psi(A^{\downarrow L}) & \text{if } A = A^{\downarrow K} \otimes A^{\downarrow L} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2 ***Factorization Lemma.*** *Let $K, L \subseteq N$ be nonempty, $K \cup L = N$. m factorizes with respect to (K, L) if and only if*

$$K\setminus L \perp\!\!\!\perp L\setminus K \,|\, K\cap L\,[m].$$

Theorem 3 ***Factorization of Composition.*** *Let $K, L \subseteq N$ be nonempty, $K \cup L = N$. m factorizes with respect to (K, L) if and only if*

$$m = m^{\downarrow K} \triangleright m^{\downarrow L}.$$

Example

Let us illustrate the space-saving power of the introduced simple factorization on the simplest possible examples. Consider for a moment only binary case: we assume $|\mathbf{X}_i| = 2$ for all i.

Considering 2-dimensional situation, we can take into account only independent variables: $\{1\} \perp\!\!\!\perp \{2\}[m]$. In this case, according to Definition 2 all focal elements of m are Cartesian products of subsets of $\mathbf{X}_1$ with subsets of $\mathbf{X}_2$. Since each $\mathbf{X}_i$ has three nonempty subsets we get that m has maximally 9 focal elements and can be uniquely described just with 6 numbers: with basic assignments $m^{\downarrow\{1\}}$ and $m^{\downarrow\{2\}}$ (see the left column of the Binary part in Table 1).

A little bit more complex situation occurs when considering 3-dimensional space of discernment (see the right column of the Binary part in Table 1). Now we can consider factorization $(K, L) = (\{1, 2\}, \{2, 3\})$ corresponding to the conditional independence $\{1\} \perp\!\!\!\perp \{3\}|\{2\}[m]$.

How one can find all sets $A \subseteq \mathbf{X}_{\{1,2,3\}}$ for which $A = A^{\downarrow\{1,2\}} \otimes A^{\downarrow\{2,3\}}$? It is obvious that this property is met by all singletons. However, there are also other sets having this property like $\{a_1a_2a_3, \bar{a}_1\bar{a}_2\bar{a}_3\}$ or

Table 1: Simple factorization space requirements

	Binary: $\mathbf{X}_i = \{a_i, \bar{a}_i\}$		Ternary: $\mathbf{X}_i = \{a_i, \bar{a}_i, \hat{a}_i\}$	
frame of discernment $\mathbf{X}$	$\mathbf{X}_1 \times \mathbf{X}_2$	$\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3$	$\mathbf{X}_1 \times \mathbf{X}_2$	$\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3$
$\lvert\mathbf{X}\rvert$	4	8	9	27
$\lvert\{A \subseteq \mathbf{X} : A \neq \emptyset\}\rvert$	15	255	511	134 217 727
factorization (K, L)	$(\{1\}, \{2\})$	$(\{1,2\}, \{2,3\})$	$(\{1\}, \{2\})$	$(\{1,2\}, \{2,3\})$
$\lvert\{A \neq \emptyset : A = A^{\downarrow K} \otimes A^{\downarrow L}\}\rvert$	9	99	49	124 999
number of factors	$2 \times 3 = 6$	$2 \times 15 = 30$	$2 \times 7 = 14$	$2 \times 511 = 1\,022$

$\{a_1 a_2 \bar{a}_3, \bar{a}_1 \bar{a}_2 a_3, \bar{a}_1 a_2 \bar{a}_3\}$. The answer to the question suggests itself when one realizes that it is equivalent to the question: for which $B \subseteq \mathbf{X}_{\{1,2\}}$ and $C \subseteq \mathbf{X}_{\{2,3\}}$ it holds that

$$(B \otimes C)^{\downarrow\{1,2\}} = B \quad \text{and} \quad (B \otimes C)^{\downarrow\{2,3\}} = C,$$

because these equivalences hold true if and only if

$$B^{\downarrow\{2\}} = C^{\downarrow\{2\}}.$$

Realizing that for 9 subsets $B \subseteq \mathbf{X}_{\{1,2\}}$ the respective projection $B^{\downarrow\{2\}}$ is $\mathbf{X}_{\{2\}}$, for 3 subsets it is $\{a_2\}$ and for another three it is $\{\bar{a}_2\}$, and that the same holds also for the projection of subsets of $\mathbf{X}_{\{2,3\}}$, one gets that there are

$$9 \times 9 + 3 \times 3 + 3 \times 3 = 99$$

nonempty subsets $A \subseteq \mathbf{X}_{\{1,2,3\}}$ for which $A = A^{\downarrow\{1,2\}} \otimes A^{\downarrow\{2,3\}}$.

For more details about this example and also how the respective numbers look like also for a ternary case see Table 1.

Graphical Models

In probability theory, graphical models were defined as probability distributions (measures) factorizing with respect to a system of subsets forming cliques of a graph (Daroch, Lauritzen and Speed 1980, Edwards and Havránek 1985). In this paper we will need just a couple of terms from graph theory; a graph consists of *nodes* and *edges*, if all pairs of nodes are connected in a graph then the graph is *complete*. By a *clique* we understand maximal subset of nodes inducing a complete subgraph. For example, graph in Figure 1(a) has three cliques: $\{1,2,3,4\}$, $\{3,4,5\}$, $\{6\}$; graph (b) in this Figure has five cliques: $\{1,2,3\}$, $\{1,4\}$, $\{3,6\}$, $\{4,5\}$, $\{5,6\}$); the last graph (c) has only two cliques: $\{1,2,3,4\}$ and $\{3,4,5,6\}$. In all these graphs we can see several *cycles*: for example $1-4-5-3$ in graph (a) and $1-4-5-6-3$ in graph (b). The difference between these two cycles is that while the latter cycle is *chordless*, in the former cycle there is an edge $4-3$ forming a *chord*.

For introducing a Dempster-Shafer counterpart of graphical models one needs a notion of a more complex factorization than simple factorization introduced in Definition 5. The above introduced factorization enables us to consider only graphs with two cliques; for example see graph in Figure 1(c). In this specific case factorization with respect to this graph means factorization with respect to $(\{1,2,3,4\}, \{3,4,5,6\})$.

Decomposable models

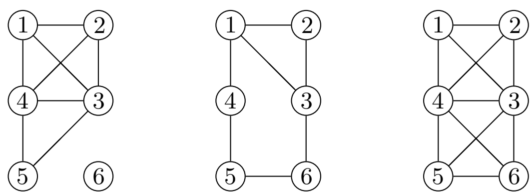
The above mentioned 2-clique graphs belong to a greater class of graphs factorization with respect to which does not bring great problems. Here we have in mind a family of *decomposable graphs*. These are the graphs which do not contain a chordless cycle of length greater then three. In Figure 1 we have two decomposable graphs, namely (a) and (c). Graph in Figure 1(b) contains a chordless cycle of length 5: $1-3-6-5-4$ (graph (c) indeed contains the same cycle, but this cycle has not only one but three chords).

It is well known that cliques $K_1, K_2, \ldots, K_r$ of a decomposable graph can be ordered to meet so called *Running Intersection Property* (RIP): for all $i = 2, \ldots, r$ there exists $j, 1 \le j < i$, such that

$$K_i \cap (K_1 \cup \ldots \cup K_{i-1}) \subseteq K_j.$$

This offers us a possibility to define decomposable models using Definition 5 recursively.

Definition 6 ***Decomposable Basic Assignments.*** We say that a basic assignment m is *decomposable* if it *factorizes with respect to a decomposable graph* in the following sense (let $K_1, K_2, \ldots, K_r$ be cliques of the considered decomposable graph ordered so that they meet RIP): for all $i = 2, \ldots, r$ the marginal $m^{\downarrow K_1 \cup \ldots \cup K_i}$ factorizes (in the sense of Definition 5) with respect to $(K_1 \cup \ldots \cup K_{i-1}, K_i)$.



(a) **(b)** **(c)**

Figure 1: Graphs with 6 nodes

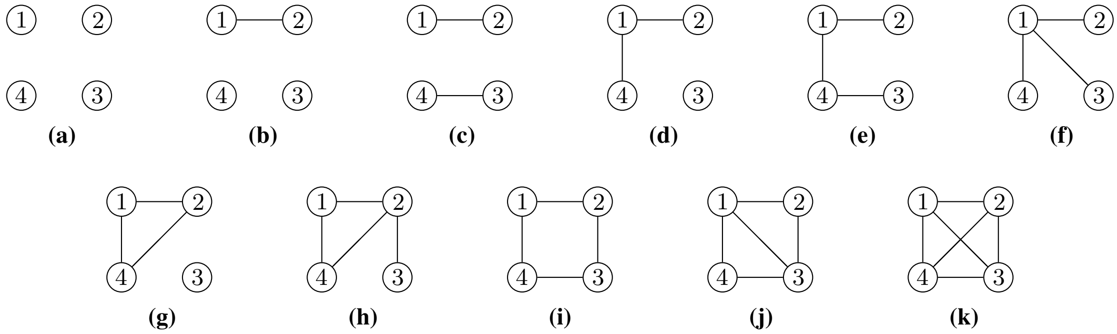


Figure 2: Graphical models for 4 variables

By a repeated application of Theorem 3 one can see that decomposable models can easily be represented by a system of its marginals.

Theorem 4 ***Composition of Decomposable Models.*** *Consider a decomposable graph with cliques $K_1, K_2, \ldots, K_r$. If this ordering meets RIP then m is decomposable with respect to the graph in question if and only if*

$$m = (\ldots((m^{\downarrow K_1} \triangleright m^{\downarrow K_2}) \triangleright m^{\downarrow K_3}) \triangleright \ldots \triangleright m^{\downarrow K_{r-1}}) \triangleright m^{\downarrow K_r}.$$

Notice that thanks to Theorem 2 one can deduce that for a decomposable basic assignment m it is possible to read the system of conditional independence relations valid for m exactly in the same way as it is done for decomposable probabilistic measures: If $G = (N, E)$ is a decomposable graph with respect to which decomposable basic assignment m factorizes, and if nodes i and j are separated in G by set M then

$$i \perp\!\!\!\perp j \mid M \; [m].$$

However, let us stress once more: This possibility holds only for conditional independence (Definition 4) and not for conditional non-interactivity.

4-dimensional models

In the previous subsection we have showed that introducing decomposable models in Dempster-Shafer theory does not bring great difficulties. But when considering the whole class of graphical models we are getting into problems, which have not been solved, yet. Let us illustrate these problems on the simplest possible case, on 4-dimensional models.

Consider all graphs with 4 nodes (see Figure 2). The complete graph (k) is uninteresting because it has only one clique and thus any distribution factorizes with respect to this graph. As the reader can easily check from the following list, with the exception of graph (i) all the remaining graphs are decomposable. Here is the list of the respective cliques; each sequence is ordered so that it meets RIP:

(a) : $\{1\}, \{2\}, \{3\}, \{4\}$
(b) : $\{1, 2\}, \{3\}, \{4\}$
(c) : $\{1, 2\}, \{3, 4\}$
(d) : $\{1, 2\}, \{1, 4\}, \{3\}$
(e) : $\{1, 2\}, \{1, 4\}, \{3, 4\}$
(f) : $\{1, 2\}, \{1, 3\}, \{1, 4\}$
(g) : $\{1, 2, 4\}, \{3\}$
(h) : $\{1, 2, 4\}, \{2, 3\}$
(j) : $\{1, 2, 3\}, \{1, 3, 4\}$

Table 2: Graphical 4-dimensional models

graph	MNFE	NP
(a)	81	12
(b)	135	21
(c)	225	30
(d)	297	33
(e)	711	45
(f)	783	45
(g)	765	258
(h)	2115	270
(i)	2961	60
(j)	9 999	510
(k)	65 535	65 535

MNFE - maximum number of focal elements
NP - number of parameters defining model

So, the only non-decomposable graph is graph (i) representing a cycle of length 4: $1 - 2 - 3 - 4$. For this specific situation a reasonable solution can be proposed as follows:

Factorization with respect to 4-cycle. A 4-dimensional basic assignment m (over $\mathbf{X}_{\{1,2,3,4\}}$) factorizes with respect to graph in Figure 2(i) if there exist four nonnegative set functions

$$\phi : \mathcal{P}(\mathbf{X}_{\{1,2\}}) \to [0, +\infty), \quad \psi : \mathcal{P}(\mathbf{X}_{\{2,3\}}) \to [0, +\infty),$$
$$\rho : \mathcal{P}(\mathbf{X}_{\{3,4\}}) \to [0, +\infty), \quad \mu : \mathcal{P}(\mathbf{X}_{\{1,4\}}) \to [0, +\infty),$$

such that for all $A \subseteq \mathbf{X}_{\{1,2,3,4\}}$: If

$$A = A^{\downarrow\{1,2,3\}} \otimes A^{\downarrow\{1,3,4\}} = A^{\downarrow\{1,2,4\}} \otimes A^{\downarrow\{2,3,4\}}$$

then

$$m(A) = \phi(A^{\downarrow\{1,2\}}) \cdot \psi(A^{\downarrow\{2,3\}}) \cdot \rho(A^{\downarrow\{3,4\}}) \cdot \mu(A^{\downarrow\{1,4\}}),$$

and $m(A) = 0$ if either $A \neq A^{\downarrow\{1,2,3\}} \otimes A^{\downarrow\{1,3,4\}}$ or $A \neq A^{\downarrow\{1,2,4\}} \otimes A^{\downarrow\{2,3,4\}}$.

This definition guarantees that m has the same dependence structure as in probability theory a distribution factorizing with respect to graph (i):

4-cycle Factorization Lemma. *If a basic assignment m over $\mathbf{X}_{\{1,2,3,4\}}$ factorizes with respect to graph in Figure 2(i) then*

$$\{1\} \perp\!\!\!\perp \{3\}|\{2,4\}[m] \ \textit{and}\ \{2\} \perp\!\!\!\perp \{4\}|\{1,3\}[m].$$

The ideas of a factorization over a 4-cycle can be generalized to factorization with respect to a general cycle graph. In such a graph $G = (N, E)$ any couple of non-adjacent nodes i, j divides the set of all the remaining nodes $N \setminus \{i, j\}$ into two (nonempty) groups K, L such that K and L are separated by $\{i, j\}$. Each such a couple i, j poses one requirement concerning a form of a focal element A:

$$A = A^{\downarrow K\cup\{i,j\}} \otimes A^{\downarrow L\cup\{i,j\}}. \qquad (*)$$

Since there are $|N|\cdot(|N|-3)/2$ non-adjacent pairs of nodes in G it means that the same number of equalities of the form of $(*)$ have to hold for any focal element of a basic assignment m factorizing with respect to G. Naturally, values $m(A)$ have to be expressible in a form of a product of $|N|$ factor functions.

Conclusion

The main message of this contribution was to show that the new version of the notion of conditional independence introduced in Definition 4 is superior to the notion of conditional irrelevance (used e.g. in (Shenoy 1994, Couso, Moral and Walley 1999, Yaghlane, Smets and Mellouli 2002b)), because it supports introduction of factorization, especially development of decomposable models. Using a simple example of factorization with respect to a cycle graph we gave a hint how it will be possible to define a notion of factorization with respect to general graphs.

Acknowledgments

This work was supported by GAČR under the grant no. ICC/08/E010, and 201/09/1891, and by MŠMT ČR under grants 1M0572 and 2C06019.

References

Ben Yaghlane, B.; Smets, Ph.; and Mellouli, K. 2002a. Belief Function Independence: I. The Marginal Case. *Int. J. of Approximate Reasoning* 29(1): 47-70.

Ben Yaghlane, B.; Smets, Ph.; and Mellouli, K. 2002b. Belief Function Independence: II. The Conditional Case. *Int. J. of Approximate Reasoning* 31(1-2): 31-75.

Couso, I.; Moral, S.; and Walley, P. 1999. Examples of Independence for Imprecise Probabilities. In *Proceedings of ISIPTA'99,* 121-130.

Daroch, J.N.; Lauritzen, S.; and Speed, T.P. 1980. Markov Fields and Log Linear Interaction Models for Contingency Tables. *Ann. Stat.* 8: 522-539.

Dempster, A. 1967. Upper and Lower Probabilities Induced by a Multi-Valued Mapping. *Annals of Mathematical Statistics* 38: 325-339.

Edwards, D.E.; and Havránek, T. 1985. A Fast Procedure for Model Search in Multidimensional Contingency Tables. *Biometrika* 72(2): 339-351.

Jiroušek, R. 1997. Composition of Probability Measures on Finite Spaces. In *Proc. of the 13th Conf. Uncertainty in Artificial Intelligence UAI'97*, 274–281. Morgan Kaufmann Publ., San Francisco, California.

Jiroušek, R. 2007. On a Conditional Irrelevance Relation for Belief Functions Based on the Operator of Composition. In *Dynamics of Knowledge and Belief, Proceedings of the Workshop at the 30th Annual German Conference on Artificial Intelligence,* 28-41. Fern Universität in Hagen, Osnabrück.

Jiroušek, R. 2010. Factorization and Decomposable Models in Dempster-Shafer Theory of Evidence. Accepted to Workshop on the Theory of Belief Functions, April 2010, Brest, France.

Jiroušek, R.; and Vejnarová, J. 2010. Compositional Models and Conditional Independence in Evidence Theory. Accepted to *Int. J. Approx. Reasoning*.

Jiroušek, R.; Vejnarová, J.; and Daniel, M. 2007. Compositional Models of Belief Functions. In *Proc. of the 5th Symposium on Imprecise Probabilities and Their Applications, ISIPTA'07,* 243–252.

Klir, G. J. 2006. *Uncertainty and Information. Foundations of Generalized Information Theory.* Wiley, Hoboken.

Lauritzen, S. L. 1996. *Graphical Models*. Oxford University Press.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press, New Jersey.

Shenoy, P. P. 1994. Conditional Independence in Valuation-Based Systems, *Int. J. of Approximate Reasoning*, 10(3): 203-234.

Shenoy, P. P. 1997. Binary Join Trees for Computing Marginals in the Shenoy-Shafer Architecture, *Int. J. of Approximate Reasoning* 17(2-3): 239-263.

Studený, M. 1993. Formal Properties of Conditional Independence in Different Calculi of AI. In *Proceedings of European Conference on Symbolic and quantitative Approaches to Reasoning and Uncertainty ECSQARU'93*, 341–351, Springer-Verlag.

Studený, M. 2002. On Stochastic Conditional Independence: the Problems of Characterization and Description. *Annals of Mathematics and Artificial Intelligence* 35: 323-341.

Vejnarová, J. 2010. A Thorough Comparison of Two Conditional Independence Concepts for Belief Functions. Accepted to Workshop on the Theory of Belief Functions, April 2010, Brest, France.