On Weakness of Evidential Networks*

Jiřina Vejnarová
Institute Information Theory and Automation
Academy of Sciences of the Czech Republic
vejnar@utia.cas.cz

Abstract

In evidence theory several counterparts of Bayesian networks based on different paradigms have been proposed. We will present, through simple examples, problems appearing in two kinds of these models caused either by the conditional independence concept (or its misinterpretation) or by the use of a conditioning rule. The latter kind of problems can be avoided if undirected models are used instead.

1 Introduction

When applying models of artificial intelligence to any practical problem one must cope with two basic problems: uncertainty and multidimensionality. The most widely used models managing these issues are, at present, so-called *probabilistic graphical Markov models*.

The problem of multidimensionality is solved in these models with the help of the notion of conditional independence, which enables factorization of a multidimensional probability distribution into small parts, usually marginal or conditional low-dimensional distributions (e.g. in *Bayesian networks*), or generally into low-dimensional factors (e.g. in *decomposable models*). Such a factorization not only decreases the storage requirements for representation of a multidimensional distribution but it usually also induces efficient computational procedures allowing inference from these models.

Probably the most popular representative of these models are *Bayesian networks*, while from the computational point of view so-called *decomposable models* are the most advantageous. Naturally, several attempts to construct an analogy of Bayesian networks have also been made in other frameworks as e.g. in possibility theory [5], evidence theory [4] or in the more general frameworks of valuation-based systems [13] and credal sets [7], while counterparts of decomposable models are, more or less, omitted.

In this contribution we will confine ourselves to evidence theory, where several counterparts of Bayesian networks based on different paradigms have been proposed

*The support of Grant GAČR 201/08/0539 is gratefully acknowledged.

[4, 13, 19]. We will present, through two simple examples, problems appearing in these models caused either by the conditional independence concept (or its misinterpretation) or by the use of different conditioning rules. The latter kind of problems can be avoided if undirected models are used instead.

2 Basic Concepts

In this section we will briefly recall basic concepts from evidence theory [12] concerning sets and set functions.

2.1 Set Projections and Joins

For an index set $N = \{1, 2, \ldots, n\}$ let $\{X_i\}_{i \in N}$ be a system of variables, each X_i having its values in a finite set $\mathbf{X}_i$. In this paper we will deal with *multidimensional frame of discernment* $\mathbf{X}_N = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_n$, and its *subframes* (for $K \subseteq N$) $\mathbf{X}_K = \bigtimes_{i \in K} \mathbf{X}_i$. When dealing with groups of variables on these subframes, X_K will denote a group of variables $\{X_i\}_{i \in K}$ throughout the paper.

For $M \subset K \subseteq N$ and $A \subset \mathbf{X}_K$, $A^{\downarrow M}$ will denote a *projection* of A into $\mathbf{X}_M$:

$$A^{\downarrow M} = \{y \in \mathbf{X}_M \mid \exists x \in A : y = x^{\downarrow M}\},$$

where, for $M = \{i_1, i_2, \ldots, i_m\}$,

$$x^{\downarrow M} = (x_{i_1}, x_{i_2}, \ldots, x_{i_m}) \in \mathbf{X}_M.$$

In addition to the projection, in this text we will also need an opposite operation, which will be called a join. By a *join*[1] of two sets $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_L$ $(K, L \subseteq N)$ we will understand a set

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

Let us note that for any $C \subseteq \mathbf{X}_{K \cup L}$ naturally $C \subseteq C^{\downarrow K} \bowtie C^{\downarrow L}$, but generally $C \neq C^{\downarrow K} \bowtie C^{\downarrow L}$.

2.2 Set Functions

In evidence theory [12] two dual measures are used to model the uncertainty: belief and plausibility measures. Both of them can be defined with the help of another set function called a *basic (probability* or *belief) assignment* m on $\mathbf{X}_N$, i.e.,

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

where $\mathcal{P}(\mathbf{X}_N)$ is the power set of $\mathbf{X}_N$, and $\sum_{A \subseteq \mathbf{X}_N} m(A) = 1$. Furthermore, we assume that $m(\emptyset) = 0$.[2]

[1] This term and notation are taken from the theory of relational databases [1].

[2] This assumption is not generally accepted, e.g. in [2] it is omitted. The consequences of this omission will be mentioned several times throughout this paper.

A set $A \in \mathcal{P}(\mathbf{X}_N)$ is a *focal element* if $m(A) > 0$. Let $\mathcal{F}$ denote the set of all focal elements, a focal element $A \in \mathcal{F}$ is called an $m-atom$ if for any $B \subseteq A$ either $B = A$ or $B \notin \mathcal{F}$. In other words, $m-atom$ is a setwise-minimal focal element.

Belief and *plausibility measures* are defined for any $A \subseteq \mathbf{X}_N$ by the equalities

$$Bel(A) = \sum_{B \subseteq A} m(B), \qquad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \tag{1}$$

respectively. It is well-known (and evident from these formulae) that for any $A \in \mathcal{P}(\mathbf{X}_N)$

$$Bel(A) \leq Pl(A), \qquad Pl(A) = 1 - Bel(A^C), \tag{2}$$

where A^C is the set complement of $A \in \mathcal{P}(\mathbf{X}_N)$. Furthermore, basic assignment can be computed from belief function via Möbius inverse:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B), \tag{3}$$

i.e. any of these three functions is sufficient to define values of the remaining two.

For a basic assignment m on $\mathbf{X}_K$ and $M \subset K$ a *marginal basic assignment* of m is defined (for each $A \subseteq \mathbf{X}_M$):

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

3 Conditioning

Conditioning belongs to the most important topics of any theory dealing with uncertainty. From the viewpoint of the construction of Bayesian-network-like multidimensional models it seems to be inevitable.

3.1 Conditioning of Events

In evidence theory the "classical" conditioning rule is the so-called *Dempster's rule of conditioning* defined for any $\emptyset \neq A \subseteq \mathbf{X}_N$ and $B \subseteq \mathbf{X}_N$ such that $Pl(B) > 0$ by the formulae

$$\begin{aligned} Bel(A|_D B) &= \frac{Bel(A \cup B^C) - Bel(B^C)}{1 - Bel(B^C)}, \\ Pl(A|_D B) &= \frac{Pl(A \cap B)}{Pl(B)}. \end{aligned} \tag{4}$$

Let us note that in [2] a bit different formulae are used: conditional beliefs and plausibilities are not normalized. It corresponds to the omission of the assumption $m(\emptyset) = 0$.

This is not the only possibility how to condition, another — in a way symmetric — conditioning rule is the following one, called *focusing* defined for any $\emptyset \neq A \subseteq \mathbf{X}_N$ and $B \subseteq \mathbf{X}_N$ such that $Bel(B) > 0$ by the formulae

$$\begin{aligned} Bel(A|_F B) &= \frac{Bel(A \cap B)}{Bel(B)}, \\ Pl(A|_F B) &= \frac{Pl(A \cup B^C) - Pl(B^C)}{1 - Pl(B^C)}. \end{aligned} \tag{5}$$

Formulae (4) and (5) are, in a way, evidential counterparts of conditioning in probabilistic framework. Let us note that the seemingly "natural" way of conditioning

$$m(A|_P B) = \frac{m(A \cap B)}{m(B)} \tag{6}$$

is not possible, since $m(A|_P B)$ need not be a basic assignment. It is caused by a simple fact that m, in contrary to Bel and Pl, is not monotonous with respect to set inclusion.

3.2 Conditional Variables

However, from the viewpoint of evidential networks conditioning of variables is of primary interest. In [18] we presented two definitions of conditioning by variables, based on Dempster conditioning rule and focusing, we proved that these definitions are correct, nevertheless, their usefulness for multidimensional models is rather questionable, as thoroughly discussed in the above-mentioned paper.

Therefore, in [19] we proposed a new conditioning rule which is, in a way, a generalization of (6).

Definition 1 *Let X_K and X_L ($K \cap L = \emptyset$) be two groups of variables with values in $\mathbf{X}_K$ and $\mathbf{X}_L$, respectively. Then the conditional basic assignment of X_K given $X_L \in B \subseteq \mathbf{X}_L$ (for B such that $m^{\downarrow L}(B) > 0$) is defined as follows:*

$$m_{X_K|_P X_L}(A|_P B) = \frac{\displaystyle\sum_{\substack{C \subseteq \mathbf{X}_{K \cup L}: \\ C^{\downarrow K} = A \& C^{\downarrow L} = B}} m(C)}{m^{\downarrow L}(B)} \tag{7}$$

for any $A \subseteq \mathbf{X}_K$.

Although we said above, that it makes little sense for conditioning of events, it is sensible in conditioning of variables, as expressed by Theorem 1 proven in [19]. The above-mentioned problem of non-monotonicity is avoided, because a marginal basic assignment is always greater than (or equal to) the joint one.

Theorem 1 *The set function $m_{X_K|_P X_L}$ defined for any fixed $B \subseteq \mathbf{X}_L$, such that $m^{\downarrow L}(B) > 0$ by Definition 1 is a basic assignment on $\mathbf{X}_K$.*

4 Conditional Independence and Irrelevance

Independence and irrelevance need not be (and usually are not) distinguished in the probabilistic framework, as they are almost equivalent to each other. Similarly, in possibilistic framework adopting De Cooman's measure-theoretical approach [9] (particularly his notion of almost everywhere equality) we proved that the analogous concepts are equivalent (for more details see [15]).

4.1 Independence

In evidence theory the most common notion of independence is that of random set independence [6]. It has already been proven [16] that it is also the only sensible one.

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

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

for all $A \subseteq \mathbf{X}_{K\cup L}$ for which $A = A^{\downarrow K} \times A^{\downarrow L}$, and $m(A) = 0$ otherwise.

This notion can be generalized in various ways [3, 13, 16]; the concept of conditional non-interactivity from [3], based on conjunctive combination rule, is used for construction of directed evidential networks in [4] (cf. also Section 5.3). In this paper we will use the concept introduced in [10, 16], as we consider it more suitable: in contrary to other conditional independence concepts [3, 13] it is *consistent with marginalization* [14], in other words, the multidimensional model of conditionally independent variables keeps the original marginals (for more details see [16]).

Definition 3 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 the equality

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

holds for any $A \subseteq \mathbf{X}_{K\cup L\cup M}$ such that $A = A^{\downarrow K\cup M} \bowtie A^{\downarrow L\cup M}$, and $m(A) = 0$ otherwise.

It has been proven in [16] that this conditional independence concept satisfies so-called the semi-graphoid properties taken as reasonable to be valid for any conditional independence concept and it has been shown in which sense this conditional independence concept is superior to previously introduced ones [3, 13].

4.2 Irrelevance

Irrelevance is usually considered to be a weaker notion than independence (see e.g. [6]). It expresses the fact that a new piece of evidence concerning one variable cannot influence the evidence concerning the other variable, in other words is irrelevant to it.

More formally: group of variables X_L is *irrelevant* to X_K ($K \cap L = \emptyset$) if for any $B \subseteq \mathbf{X}_L$ such that the left-hand side of the equality is defined

$$m_{X_K|X_L}(A|B) = m(A) \tag{8}$$

for any $A \subseteq \mathbf{X}_K$.[3]

It follows from the definition of irrelevance that it need not be a symmetric relation. Let us note, that in the framework of evidence theory neither irrelevance based on Dempster conditioning rule nor that based on focusing even in cases when the relation is symmetric, imply independence, as can be seen from examples in [18].

Generalization of this notion to conditional irrelevance may be done as follows. Group of variables X_L is *conditionally irrelevant* to X_K given X_M (K, L, M disjoint, $K \neq \emptyset \neq L$) if

$$m_{X_K|X_L X_M}(A|B) = m_{X_K|X_M}(A|B^{\downarrow M}) \tag{9}$$

is satisfied for any $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_{L \cup M}$, such that both sides are defined.

Let us note that the conditioning in equalities (8) and (9) stands for an abstract conditioning rule (any of those mentioned in the previous section or some other [8]). However, the validity of (8) and (9) may depend on the choice of the conditioning rule.

4.3 Relationship Between Independence and Irrelevance

As mentioned at the end of preceding section, different conditioning rules lead to different irrelevance concepts. Nevertheless, when studying the relationship between (conditional) independence and irrelevance based on Dempster conditioning rule and focusing we realized that they do not differ too much from each other, as suggested by the following summary.

For both conditioning rules:

- Irrelevance is implied by independence.
- Irrelevance does not imply independence.
- Irrelevance is not symmetric, in general.
- Even in case of symmetry it does not imply independence.
- Conditional independence does not imply conditional irrelevance.

The only difference between these conditioning rules is expressed by the following theorem proven in [18].

Theorem 2 *Let X_K and X_L be conditionally independent groups of variables given X_M under joint basic assignment m on $\mathbf{X}_{K \cup L \cup M}$ (K, L, M disjoint, $K \neq \emptyset \neq L$). Then*

$$m_{X_K|_F X_L X_M}(A|_F B) = m_{X_K|_F X_M}(A|_F B^{\downarrow M}) \tag{10}$$

for any $m^{\downarrow L \cup M}$-atom $B \subseteq \mathbf{X}_{L \cup M}$ such that $B^{\downarrow M}$ is $m^{\downarrow M}$-atom and $A \subseteq \mathbf{X}_K$.

---

[3] Let us note that somewhat weaker definition of irrelevance one can found in [2], where equality is substituted by proportionality. This notion has been later generalized using conjunctive combination rule [3].

From this point of view focusing seems to be slightly superior to Dempster conditioning rule, but still it is not satisfactory. However, the new conditioning rule introduced by Definition 1 is more promising, as suggested by the following theorem, proven in [19].

Theorem 3 *Let K, L, M be disjoint subsets of N such that $K, L \neq \emptyset$. If X_K and X_L are independent given X_M (with respect to a joint basic assignment m defined on $X_{K \cup L \cup M}$), then X_L is irrelevant to X_K given X_M under the conditioning rule given by Definition 1.*

The reverse implication is not valid in general, which expresses the expected property: conditional independence is stronger than conditional irrelevance.

However, in Bayesian networks also the reverse implication plays an important role, as for the inference, the network is usually transformed into a decomposable model. Nevertheless, the following assertion proven in [20] holds true.

Theorem 4 *Let K, L, M be disjoint subsets of N such that $K, L \neq \emptyset$ and $m_{X_K|_P X_{L \cup M}}$ be a (given) conditional basic assignment of X_K given $X_{L \cup M}$ and $m_{X_{L \cup M}}$ be a basic assignment of $X_{L \cup M}$. If X_L is irrelevant to X_K given X_M under the conditioning rule given by Definition 1, then X_K and X_L are independent given X_M (with respect to a joint basic assignment $m = m_{X_K|_P X_{L \cup M}} \cdot m_{X_{L \cup M}}$ defined on $\mathbf{X}_{K \cup L \cup M}$).*

5 (Directed) Evidential Networks and Compositional Models

In this section we will deal with directed evidential networks [4] and evidential networks [20]. These two models differ not only by the conditioning rule, but also, and it seems to be more important, by the interpretation of graph structure of the model.

While in evidential networks conditional basic assignment is assigned to every node given its parents (analogously to Bayesian networks), in directed evidential networks conditional beliefs are assigned to arcs, i.e. to every node as many conditionals are assigned as is the number of its parents. These conditionals are subsequently combined by the conjunctive combination rule.

The difference between directed evidential networks and compositional models will be described in Section 5.3 by a simple example, while the lost of information in evidential networks (in comparison with compositional models) in Section 5.4. Before doing that we need to recall the concept of compositional models.

5.1 Compositional models

Compositional models are based on the concept of the operator of composition of basic assignments, introduced in [11] in the following way.

Definition 4 *For two arbitrary basic assignments m_1 on $\mathbf{X}_K$ and m_2 on $\mathbf{X}_L$ a composition $m_1 \triangleright m_2$ is defined for all $C \subseteq \mathbf{X}_{K \cup L}$ by one of the following expressions:*

(a) *if* $m_2^{\downarrow K\cap L}(C^{\downarrow K\cap L}) > 0$ *and* $C = C^{\downarrow K} \bowtie C^{\downarrow L}$ *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*

$$(m_1 \triangleright m_2)(C) = 0.$$

From the basic properties of this operator (proven in [10, 11]) it follows that operator of composition is not commutative in general, but it preserves first marginal (in case of projective basic assignments both of them). In both these aspects it differs from conjunctive combination rule. Furthermore, operator of composition is not associative and therefore its iterative applications must be made carefully, as we will see later.

A lot of other properties possessed by the operator of composition can be found in [10, 11], nevertheless here we will confine ourselves to the following theorem (proven in [10]) expressing the relationship between conditional independence and operator of composition.

Theorem 5 *Let* m *be a joint basic assignment on* $\mathbf{X}_M$, $K, L \subseteq M$. *Then* $(K \setminus L) \perp\!\!\!\perp (L \setminus K)|(K \cap L)\ [m]$ *if and only if*

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

for any $A \subseteq \mathbf{X}_{K\cup L}$.

Now, let us consider a system of low-dimensional basic assignments $m_1, m_2, \ldots, m_n$ defined on $\mathbf{X}_{K_1}, \mathbf{X}_{K_2}, \ldots, \mathbf{X}_{K_n}$, respectively. Composing them together by multiple application of the operator of composition, one gets multidimensional a basic assignment on $\mathbf{X}_{K_1\cup K_2\cup\ldots\cup K_n}$. However, since we know that the operator of composition is neither commutative nor associative, we have to properly specify what "composing them together" means.

To avoid using too many parentheses let us make the following convention. Whenever we write the expression $m_1 \triangleright m_2 \triangleright \ldots \triangleright m_n$ we will understand that the operator of composition is performed successively from left to right:[4]

$$m_1 \triangleright m_2 \triangleright \ldots \triangleright m_n = (\ldots((m_1 \triangleright m_2) \triangleright m_3) \triangleright \ldots) \triangleright m_n. \tag{11}$$

Therefore, multidimensional model (11) is specified by an ordered sequence of low-dimensional basic assignments — a *generating sequence* $m_1, m_2, \ldots, m_n$.

[4]Naturally, if we want to change the ordering in which the operators are to be performed we will do so using parentheses.

5.2 Evidential network generated by a perfect sequence

From the point of view of artificial intelligence models used to represent knowledge in a specific area of interest, a special role is played by the so-called *perfect sequences*, i.e., generating sequences $m_1, m_2, \ldots, m_n$, for which

$$\begin{aligned} m_1 \triangleright m_2 &= m_2 \triangleright m_1, \\ m_1 \triangleright m_2 \triangleright m_3 &= m_3 \triangleright (m_1 \triangleright m_2), \\ &\vdots \\ m_1 \triangleright m_2 \triangleright \ldots \triangleright m_n &= m_n \triangleright (m_1 \triangleright \ldots \triangleright m_{n-1}). \end{aligned}$$

The property explaining why we call these sequences "perfect" is expressed by the following assertion proven in [10].

Theorem 6 *A generating sequence $m_1, m_2, \ldots, m_n$ is perfect if and only if all assignments $m_1, m_2, \ldots, m_n$ are marginal assignments of the multidimensional assignment $m_1 \triangleright m_2 \triangleright \ldots \triangleright m_n$:*

$$(m_1 \triangleright m_2 \triangleright \ldots \triangleright m_n)^{\downarrow K_j} = m_j,$$

for all $j = 1, \ldots, n$.

Now, let us recall a simple algorithm for the construction of an evidential network from a perfect sequence of basic assignments [17].

Having a perfect sequence $m_1, m_2, \ldots, m_n$ (m_ℓ being the basic assignment of X_{K_ℓ}), we first order all the variables for which at least one of the basic assignments m_ℓ is defined in such a way that first we order (in an arbitrary way) variables for which m_1 is defined, then variables from m_2 which are not contained in m_1, etc.[5] Finally we have

$$\{X_1, X_2, X_3, \ldots, X_k\} = \{X_i\}_{i \in K_1 \cup \ldots \cup K_n}.$$

Then we get a graph of the constructed evidential network in the following way:

1. the nodes are all the variables $X_1, X_2, X_3, \ldots, X_k$;

2. there is an edge $(X_i \to X_j)$ if there exists a basic assignment m_ℓ such that both $i, j \in K_\ell$, $j \notin K_1 \cup \ldots \cup K_{\ell-1}$ and either $i \in K_1 \cup \ldots \cup K_{\ell-1}$ or $i < j$.

Evidently, for each j the requirement $j \in K_\ell$, $j \notin K_1 \cup \ldots \cup K_{\ell-1}$ is met exactly for one $\ell \in \{1, \ldots, n\}$. It means that all the parents of node X_j must be from the respective set $\{X_i\}_{i \in K_\ell}$ and therefore the necessary conditional basic assignments $m_{j|pa(j)}$ can easily be computed from basic assignment m_ℓ via (7).

It is also evident, that if both i and j are in the same basic assignment and not in previous ones, then the direction of the arc depends only on the ordering of the variables. This might lead to different independences, nevertheless, the following theorem proven in [17] sets forth that any of them is induced by the perfect sequence.

[5] Let us note that variables $X_1, X_2, \ldots, X_k$ may be ordered arbitrarily, nevertheless, for the above ordering proof of Theorem 7 is simpler than in the general case.

Table 1: Basic assignments m_i and conditional basic assignments $m_{.|i}$.

$A \subseteq \mathbf{C}_i$	$m_i(A)$
$\{h_i\}$	0.49
$\{t_1\}$	0.49
$\{h_1, t_1\}$	0.02

$D \subseteq \mathbf{B}$	$m_{.\|i}(D)$
$\{b\}$	0.49
$\{\bar{b}\}$	0.49
$\{b, \bar{b}\}$	0.02

Table 2: Joint basic assignment m of variables C_1, C_2 and B.

m	$\{b\}$			$\{\bar{b}\}$			$\{b, \bar{b}\}$		
	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$
$\{h_1\}$	0.24	0	0	0	0.24	0	0	0	0.01
$\{t_1\}$	0	0.24	0	0.24	0	0	0	0	0.01
$\{h_1, t_1\}$	0	0	0	0	0	0	0.01	0.01	~ 0

Theorem 7 *For a belief network defined by the above procedure the following independence statements are satisfied for any $j = 2, \ldots k$:*

$$\{j\} \perp\!\!\!\perp (\{i < j\} \setminus pa(j)) \mid pa(j). \tag{12}$$

5.3 Example: two coins toss

Let us consider two fair coins toss expressed by variables C_1 and C_2 with values in $\mathbf{C}_1$ and $\mathbf{C}_2$, respectively ($\mathbf{C}_i = \{h_i, t_i\}$), and the basic assignments m_1 and m_2 (contained in the left part of Table 1) expressing the fact that the result of any of the coins may from time to time be unknown. The results of tossing two coins are usually considered to be independent, therefore the joint basic assignment m_{12} is just a product of these m_1 and m_2 (cf. definition of random set independence at the beginning of Section 4).

Now, let us consider one more variable B expressing the fact the bell is ringing, i.e. $\mathbf{B} = \{b, \bar{b}\}$. It happens only if the result on both coins is the same (two heads or two tails). It is evident, that B depends on both C_1 and C_2, which corresponds to the graph in Figure 5.3 and (due to deterministic dependence of the values of B

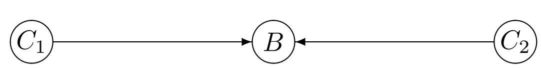

Figure 1: Graph G from Example: two coin toss.

on the values of C_1 and C_2) the joint basic assignment of the three variables is in Table 2. The above-mentioned graph can easily be obtained from perfect sequence of basic assignments m_1, m_2 and $m_3 \equiv m$ (contained in Tables 1 and 2) via the algorithm presented in the preceding section.

Table 3: Joint basic assignment of variables C_1, C_2 and B based conjunctive combination rule; b^* stands for either b or $\bar{b}$.

m	$\{b^*\}$			$\{b, \bar{b}\}$		
	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$
$\{h_1\}$	0.0624	0.0624	0.0025	0.0001	0.0001	~ 0
$\{t_1\}$	0.0624	0.0624	0.0025	0.0001	0.0001	~ 0
$\{h_1, t_1\}$	0.0025	0.0025	0.0001	~ 0	~ 0	~ 0

The approach suggested by Ben Yaghlane et al. [4] is completely different. The authors start from belief functions of C_1 and C_2 and conditional belief functions of B given C_1 and C_2, respectively. To make the difference between these two approaches more apparent we will use basic assignments instead of belief functions (belief functions, nevertheless, can be easily obtained from them by (1)). The conditional basic assignments of B given C_1 and C_2, respectively, can be found in the right part of Table 1. Let us note that these conditional basic assignments do not depend on the condition, as the results of tossing two coins are independent and therefore also the event that the bell rings does not depend on the result at one coin.

The values of joint basic assignments is computed from Tables 1 using (non-normalized) conjunctive combination rule. Results of these computations can be found in Table 3.

It is evident that the independence (non-interactivity) between coins C_1 and C_2 is not valid any more — it has been substituted by conditional non-interactivity, which does not make a sense, as C_1 is strongly dependent on C_2 whenever B is known.

5.4 Evidential Network vs Compositional Model

Theorem 3 makes it possible to define evidential networks in a way analogous to Bayesian networks, but simultaneously brings a question: are these networks advantageous in comparison with other multidimensional models in this framework? The following example brings, at least partial, answer to this question.

Example 1 Let X_1, X_2 and X_3 be three binary variables with values in $\mathbf{X}_i = \{a_i, \bar{a}_i\}, i = 1, 2, 3$, and m be a basic assignment on $\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3$ defined as follows

$$m(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5,$$
$$m(\{(a_1, a_2, \bar{a}_3), (\bar{a}_1, \bar{a}_2, a_3)\}) = .5.$$

Variables X_1 and X_2 are conditionally independent given X_3 with respect to m. Therefore also X_2 is irrelevant to X_1 given X_3, i.e.

$$m_{X_1|X_{23}}(A|B) = m_{X_1|X_3}(A|B^{\downarrow\{3\}}), \tag{13}$$

for any focal element B of $m^{\downarrow\{23\}}$. As both $m^{\downarrow\{23\}}$ and $m^{\downarrow\{3\}}$ have only two focal elements, namely $\mathbf{X}_2 \times \{\bar{a}_3\}$ and $\{(a_2, \bar{a}_3), (\bar{a}_2, a_3)\}$ and $\{\bar{a}_3\}$ and $\mathbf{X}_3$, respectively, we have

$$m_{X_1|_P X_{23}}(\mathbf{X}_1|\mathbf{X}_2 \times \{\bar{a}_3\}) = m_{X_1|_P X_3}(\mathbf{X}_1|\{\bar{a}_3\}) = 1, \quad (14)$$
$$m_{X_1|_P X_{23}}(\mathbf{X}_1|\{(a_2, \bar{a}_3), (\bar{a}_2, a_3)\}) = m_{X_1|_P X_3}(\mathbf{X}_1|\mathbf{X}_3) = 1. \quad (15)$$

Using these conditionals and the marginal basic assignment $m^{\downarrow\{23\}}$ we get a basic assignment $\tilde{m}$ different from the original one, namely

$$\tilde{m}(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5,$$
$$\tilde{m}(\mathbf{X}_1 \times \{(a_2, \bar{a}_3), (\bar{a}_2, a_3)\}) = .5.$$

Furthermore, if we interchange X_1 and X_2 we get yet another model, namely

$$\hat{m}(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5,$$
$$\hat{m}(\mathbf{X}_2 \times \{(a_1, \bar{a}_3), (\bar{a}_1, a_3)\}) = .5. \qquad \diamondsuit$$

The conditional independence of X_1 and X_2 given X_3 and relation (13) correspond to a directed graph in Figure 5.4, which leads to the following system of (conditional)

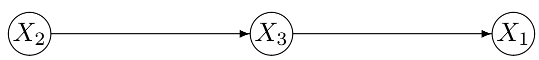

Figure 2: Graph G from Example 1.

basic assignments:

$$m^{\downarrow 2}(\mathbf{X}_2) = 1,$$
$$m_{X_3|_P X_2}(\{\bar{a}_3\}|\mathbf{X}_2) = m_{X_3|_P X_2}(\mathbf{X}_2|\mathbf{X}_2) = 1,$$

and $m_{X_1|_P X_3}$ as suggested in right-hand side of (14) and (15).

The final model

$$\check{m}(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5,$$
$$\check{m}(\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3) = .5.$$

is again different, as instead of basic assignment $m^{\downarrow 23}$ (as in Example 1) we used its marginal and conditional.

Therefore it is evident, that evidential networks are less powerful than e.g. compositional models [10], as any of these threedimensional basic assignments can be obtained from two twodimensional ones using the operator of composition.

6 Conclusions

This contribution was devoted to two kinds of multidimensional models with directed graph structure, namely directed evidential networks and evidential networks.

In directed evidential networks the graph structure is used in different sense than in Bayesian networks (it resembles rather so-called pseudobayesian networks), which may lead to senseless results, as we presented by a simple example.

Evidential networks, in contrary, keep the sense of the graphical structure known from Bayesian networks, nevertheless their weakness consists in conditioning, which may destroy the structure of the original focal elements.

From this point of view compositional models seem to be more appropriate multidimensional models in the framework of evidence theory than these two kinds of networks.

References

[1] C. Beeri, R. Fagin, D. Maier, M. Yannakakis, On the desirability of acyclic database schemes, *J. of the Association for Computing Machinery,* **30** (1983), 479–513.

[2] B. Ben Yaghlane, Ph. Smets, K. Mellouli, Belief functions independence: I. the marginal case. *Int. J. Approx. Reasoning*, **29** (2002), 47–70.

[3] B. Ben Yaghlane, Ph. Smets, K. Mellouli, Belief functions independence: II. the conditional case. *Int. J. Approx. Reasoning*, **31** (2002), 31–75.

[4] B. Ben Yaghlane, Ph. Smets, K. Mellouli, Directed evidential networks with conditional belief functions. In: Nielsen TD, Zhang NL (eds.) *Proceedings of ECSQARU 2003,* 291–305.

[5] S. Benferhat, D. Dubois, L. Gracia, H. Prade, Directed possibilistic graphs and possibilistic logic. In: B. Bouchon-Meunier, R.R. Yager, (eds.) *Proc. of IPMU'98*, Editions E.D.K. Paris, 1470–1477.

[6] I. Couso, S. Moral, P. Walley, Examples of independence for imprecise probabilities, *Proceedings of ISIPTA'99,* eds. G. de Cooman, F. G. Cozman, S. Moral, P. Walley, 121–130.

[7] F. G. Cozman, Credal networks, *Artificial Intelligence Journal,* **120** (2000), 199-233.

[8] M. Daniel, Belief conditioning rules for classic belief functions, *Proceedings of WUPES'09,* eds. T. Kroupa, J. Vejnarová, 46–56.

[9] G. De Cooman, Possibility theory I – III. *Int. J. General Systems* **25** (1997), 291–371.

[10] R. Jiroušek, J. Vejnarová, Compositional models and conditional independence in Evidence Theory, *Int. J. Approx. Reasoning,* **52** (2011), 316-334.

[11] R. Jiroušek, J. Vejnarová, M. Daniel, Compositional models for belief functions. In: De Cooman G, Vejnarová J, Zaffalon M (eds.) *Proceedings of ISIPTA'07.* Praha, 243–252.

[12] G. Shafer, *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, New Jersey (1976).

[13] P. P. Shenoy, Conditional independence in valuation-based systems. *Int. J. Approx. reasoning,* **10** (1994), 203–234.

[14] M. Studený, Formal properties of conditional independence in different calculi of artificial intelligence. *Proceedings of ECSQARUÃ93*, eds. K. Clarke, R. Kruse, S. Moral, Springer-Verlag, 1993, 341–348.

[15] J. Vejnarová, Conditional independence relations in possibility theory. *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems* **8** (2000), 253–269.

[16] J. Vejnarová, On conditional independence in evidence theory, *Proceedings of ISIPTA'09,* eds. T. Augustin, F. P. A. Coolen, S. Moral, M. C. M. Troffaes, Durham, UK, 2009, 431–440.

[17] J. Vejnarová, An alternative approach to evidential network construction, *Combining Soft Computing and Statistical Methods in Data Analysis.* Eds: Borgelt Ch., Gonzales-Rodriguez G., Trutschnig W., Lubiano M.A., Gil M.A., Grzegorzewski P., Hryniewicz O., Oviedo 2010, 619–626.

[18] J. Vejnarová, Conditioning, conditional independence and irrelevance in evidence theory, *Proceedings of ISIPTA'11,* eds. F. Coolen, G. de Cooman, T. Fetz, M. Oberguggenberger, Innsbruck, Austria, 2011, 381–390.

[19] J. Vejnarová, Conditioning in evidence theory from the perspective of multidimensional models, *Proceedings of IPMU'12,* eds. S. Greco et al., Part III, CCIS 299, 2012, 450459.

[20] J. Vejnarová, Evidential Networks from a Different Perspective, *Proeedings of SMPS'12,* to apperar.