Kybernetika

Jiřina Vejnarová

A comparison of evidential networks and compositional models

Kybernetika, Vol. 50 (2014), No. 2, 246–267

Persistent URL: http://dml.cz/dmlcz/143792

KYBERNETIKA — VOLUME *50* (2014), NUMBER 2, PAGES 246–267

A COMPARISON OF EVIDENTIAL NETWORKS AND COMPOSITIONAL MODELS

Jiřina Vejnarová

Several counterparts of Bayesian networks based on different paradigms have been proposed in evidence theory. Nevertheless, none of them is completely satisfactory. In this paper we will present a new one, based on a recently introduced concept of conditional independence. We define a conditioning rule for variables, and the relationship between conditional independence and irrelevance is studied with the aim of constructing a Bayesian-network-like model. Then, through a simple example, we will show a problem appearing in this model caused by the use of a conditioning rule. We will also show that this problem can be avoided if undirected or compositional models are used instead.

Keywords: evidence theory, conditioning, independence, directed graphs

Classification: 68T37

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 factorisation 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 factorisation 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 representatives 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 in, for example, possibility theory [5], evidence theory [4] or in the more general frameworks of valuation-based systems [15] and credal sets [7], while counterparts of decomposable models are, more or less, omitted.

DOI: 10.14736/kyb-2014-2-0246

In this paper, summarizing results published in several conference proceedings [21, 22, 23] we confine ourselves to evidence theory. Previously proposed counterpart of Bayesian networks, so-called directed evidential networks [4], does not seem to us satisfactory because of the conditional independence concept (or its misinterpretation), as we will demonstrate on a simple example. We will present a new concept of evidential networks based on the independence concept introduced in [18] (see also [11]). For this purpose a new conditioning rule for variables is proposed to enable a reasonable relationship between conditional independence and irrelevance based on this rule. Nevertheless, problems exist also in this type of models, caused by the use of a conditioning rule. Again, this problem will be demonstrated through a simple example, which also indicates that this problem can be avoided if undirected or compositional models are used instead.

The paper is organised as follows. After a brief overview of basic concepts necessary for understanding the paper in Section 2, Section 3 is devoted to conditioning in evidence theory, especially from the viewpoint of evidential networks (including the definition of a new conditioning rule for variables and proof of its correctness). In Section 4 conditional independence and irrelevance are recalled and their relationship as it applies to the above-mentioned conditioning rule is studied. Finally, in Section 5, after demonstrating the problems connected with directed evidential networks, we will recall compositional models and we will compare evidential networks with compositional models (and decomposable models).

2. BASIC CONCEPTS

In this section we will recall, as briefly as possible, basic concepts from evidence theory [14] concerning sets and set functions.

2.1. Set projections and extensions

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 a *multidimensional frame of discernment*[1]

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

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.

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 $K = \{i_1, i_2, \ldots, i_k\}$

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

Analogously, for $M \subset K \subseteq N$ and $A \subset \mathbf{X}_K$, $A^{\downarrow M}$ will denote a *projection* of A into $\mathbf{X}_M$. In this case

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

[1] Let us note that already Kong [13] dealt with multidimensional belief functions.

In addition to the projection, in this text we will also need an inverse operation that is usually called a cylindrical extension. The *cylindrical extension* of $A \subset \mathbf{X}_K$ to $\mathbf{X}_L$ $(K \subset L)$ is the set

$$A^{\uparrow L} = \{x \in \mathbf{X}_L : x^{\downarrow K} \in A\}.$$

Clearly

$$A^{\uparrow L} = A \times \mathbf{X}_{L \setminus K}.$$

A more complex case is to make a common extension of two sets, which will be called a join [1]. By a *join* 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 $C \neq C^{\downarrow K} \bowtie C^{\downarrow L}$.

Let us also note that if K and L are disjoint, then the join of A and B is just their Cartesian product, $A \bowtie B = A \times B$, and if $K = L$ then $A \bowtie B = A \cap B$. If $K \cap L \neq \emptyset$ and $A^{\downarrow K \cap L} \cap B^{\downarrow K \cap L} = \emptyset$ then $A \bowtie B = \emptyset$ as well. Generally,

$$A \bowtie B = (A \times \mathbf{X}_{L \setminus K}) \cap (B \times \mathbf{X}_{K \setminus L}),$$

i. e., a join of two sets is the intersection of their cylindrical extensions.

2.2. Set functions

In evidence theory [14], two dual measures are used to model the uncertainty: belief and plausibility measures. Each 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]

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 $A \in \mathcal{P}(\mathbf{X}_N)$

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

[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.

where A^C is the set complement of $A \in \mathcal{P}(\mathbf{X}_N)$. Furthermore, basic assignment can be computed from the 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 determine values of the remaining two.

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

$$m^{\downarrow M}(A) = \sum_{\substack{B \subseteq \mathbf{X}_K \\ B^{\downarrow M} = A}} m(B). \tag{4}$$

Having two basic assignments m_1 and m_2 on $\mathbf{X}_K$ and $\mathbf{X}_L$, respectively ($K, L \subseteq N$), we say that these assignments are *projective* if

$$m_1^{\downarrow K \cap L} = m_2^{\downarrow K \cap L},$$

which occurs if and only if there exists a basic assignment m on $\mathbf{X}_{K \cup L}$ such that both m_1 and m_2 are marginal assignments of m.

Dempster's rule of combination [14] has, from the time it first appeared, been frequently criticised by many authors. Therefore, many alternatives to it were suggested by various authors. From the viewpoint of this paper, the most important is the *conjunctive combination rule* [3], which is, in fact, an unnormalised Dempster's rule defined for m_1 and m_2 on the same space $\mathbf{X}_K$ by the formula

$$(m_1 \circledcirc m_2)(C) = \sum_{A, B \subseteq \mathbf{X}_K A \cap B = C} m_1(A) m_2(B).$$

The result of this rule is one of the examples of unnormalised basic assignment.

It can easily be generalised [3] to the case when m_1 is defined on $\mathbf{X}_K$ and m_2 is defined on $\mathbf{X}_L$ ($K \neq L$) in the following way (for any $C \in \mathbf{X}_{K \cup L}$):

$$(m_1 \circledcirc m_2)(C) = \sum_{\substack{A \subseteq \mathbf{X}_K, B \subseteq \mathbf{X}_L \\ A^{\uparrow L \cup K} \cap B^{\uparrow L \cup K} = C}} m_1(A) m_2(B). \tag{5}$$

3. CONDITIONING

Conditioning is one of the most important topics of any theory dealing with uncertainty. When studying Bayesian-network-like multidimensional models one can hardly avoid it.

3.1. Conditioning of events

In evidence theory the "classical" conditioning rule is the so-called *Dempster's rule of conditioning* [14] defined for any $\emptyset \neq A \subseteq \mathbf{X}_N$ and $B \subseteq \mathbf{X}_N$ such that $Pl(B) > 0$ by the formula

$$m(A|_D B) = \frac{\sum_{C \subseteq \mathbf{X}_N : C \cap B = A} m(C)}{Pl(B)}$$

and $m(\emptyset|_D B) = 0$. This conditioning rule can be viewed as a special case of Dempster's rule of combination.

From this formula one can immediately obtain:

$$\begin{aligned} Bel(A|_D B) &= \frac{Bel(A \cup B^C) - Bel(B^C)}{1 - Bel(B^C)}, \\ Pl(A|_D B) &= \frac{Pl(A \cap B)}{Pl(B)}. \end{aligned} \tag{6}$$

This is not the only possible way to perform conditioning. Another – in a way symmetric – conditioning rule is the following one, called *focusing* [10], defined for any $\emptyset \neq A \subseteq \mathbf{X}_N$ and $B \subseteq \mathbf{X}_N$ such that $Bel(B) > 0$ by the formula

$$m(A|_F B) = \begin{cases} \dfrac{m(A)}{Bel(B)} & \text{if} \quad A \subseteq B, \\ 0 & \text{otherwise.} \end{cases}$$

From the following two equalities one can see in which sense these two conditioning rules are symmetric:

$$\begin{aligned} Bel(A|_F B) &= \frac{Bel(A \cap B)}{Bel(B)}, \\ Pl(A|_F B) &= \frac{Pl(A \cup B^C) - Pl(B^C)}{1 - Pl(B^C)}. \end{aligned} \tag{7}$$

These are, of course, only examples of conditioning rules, there exist a great number of them, see e. g. [8].

3.2. Conditional variables

In [20] we presented the following two definitions of conditioning by variables, based on the Dempster conditioning rule and focusing, we proved that these definitions are correct, i. e. , these rules define (generally different) basic assignments (for more details see [20]). As we recently realized, Xu and Smets [24] dealt with similar problem, however they avoided normalization.

Nevertheless, the usefulness of the above-mentioned conditioning rules for multidimensional models is rather questionable, as we shall see in Section 4.3. This fact led us to the following proposal of a new conditioning rule.

Definition 3.1. 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:

$$m_{X_K|X_L}(A|B) = \frac{\displaystyle\sum_{\substack{C \subseteq \mathbf{X}_{K \cup L}: \\ C^{\downarrow K} = A \,\&\, C^{\downarrow L} = B}} m(C)}{m^{\downarrow L}(B)} \tag{8}$$

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

Although this conditioning rule does not have its event preimage, it is sensible in conditioning of variables (in other words, it is correctly defined) as expressed by Theorem 3.2.

Theorem 3.2. The set function $m_{X_K|X_L}$ defined for any fixed $B \subseteq \mathbf{X}_L$ such that $m^{\downarrow L}(B) > 0$ by Definition 3.1 is a basic assignment on $\mathbf{X}_K$.

Proof. Let $B \subseteq \mathbf{X}_L$ be such that $m^{\downarrow L}(B) > 0$. As nonnegativity of $m_{X_K|X_L}(A|B)$ for any $A \subseteq \mathbf{X}_K$ and the fact that $m_{X_K|X_L}(\emptyset|_p B) = 0$ follow directly from the definition, to prove that $m_{X_K|X_L}$ is a basic assignment it is enough to show that

$$\sum_{A \subseteq \mathbf{X}_K} m_{X_K|X_L}(A|B) = 1.$$

To check it, let us sum up the values in the numerator in (8)

$$\begin{aligned}\sum_{A \subseteq \mathbf{X}_K} \sum_{\substack{C \subseteq \mathbf{X}_{K \cup L}: \\ C^{\downarrow K} = A \& C^{\downarrow L} = B}} m(C) &= \sum_{\substack{C \subseteq \mathbf{X}_{K \cup L} \\ C^{\downarrow L} = B}} m(C) \\ &= m^{\downarrow L}(B),\end{aligned}$$

where the last equality follows directly from (4). □

The difference among Dempster conditioning rule, focusing and conditioning rule introduced in Definition 3.1 is illustrated by the following simple example.

Example 3.3. 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

$$m(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5, \tag{9}$$

$$m(\{(a_1, a_2, \bar{a}_3), (\bar{a}_1, \bar{a}_2, a_3)\}) = .5. \tag{10}$$

Let us compute conditional basic assignments of X_1 and X_2 given X_3 for the above mentioned conditioning rules. Marginal basic assignment of X_3 is

$$m(\{\bar{a}_3\}) = .5, \qquad m(\mathbf{X}_3) = .5.$$

From these values one gets

$$Bel(\{\bar{a}_3\}) = .5, \qquad Bel(\mathbf{X}_3) = 1$$

and

$$Pl(\{a_3\}) = .5, \quad Pl(\{\bar{a}_3\}) = 1, \quad Pl(\mathbf{X}_3) = 1.$$

Therefore, one can perform conditioning by $\{\bar{a}_3\}$ and $\mathbf{X}_3$ for any conditioning rule (and also by $\{a_3\}$ by Dempster's rule). Doing that one obtains:

$$m(\mathbf{X}_1 \times \mathbf{X}_2 | \{\bar{a}_3\}) = 1,$$

$$m(\{(a_1, a_2), (\bar{a}_1, \bar{a}_2)\} | \mathbf{X}_3) = 1,$$

while

$$m(\mathbf{X}_1 \times \mathbf{X}_2|_F\{\bar{a}_3\}) = 1,$$
$$m(\{(a_1, a_2), (\bar{a}_1, \bar{a}_2)\}|_F\mathbf{X}_3) = .5,$$
$$m(\mathbf{X}_1 \times \mathbf{X}_2|_F\mathbf{X}_3) = .5,$$

and

$$m(\mathbf{X}_1 \times \mathbf{X}_2|_D\{\bar{a}_3\}) = 1,$$
$$m(\{(a_1, a_2), (\bar{a}_1, \bar{a}_2)\}|_D\{\bar{a}_3\}) = 1,$$
$$m(\{(a_1, a_2), (\bar{a}_1, \bar{a}_2)\}|_D\mathbf{X}_3) = .5,$$
$$m(\mathbf{X}_1 \times \mathbf{X}_2|_D\mathbf{X}_3) = .5,$$

from which the difference can be seen. ♢

The main disadvantage of the new conditioning rule is, that conditional by the whole subframework need not be identical with the marginals (as e. g. , in the above example). In this respect it differs from both Dempster conditioning rule and focusing. It is caused by the fact, that conditioning via (8) means that the value of X_L belongs exactly to B (and not to any other set), while in focusing it may belong to any subset of B and using Dempster rule even to any subset intersecting it.

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 [17]).

4.1. Independence

In evidence theory the most common notion of independence is that of a random set independence [6]. It has already been proven [18] that it is also the only sensible one, as, e. g. , application of strong independence to two bodies of evidence may generally lead to a model which is beyond the framework of evidence theory.

Definition 4.1. 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*[3] *with respect to a basic assignment* m (in notation $K \perp\!\!\!\perp L\ [m]$) if

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

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

[3]This independence concept is a generalization of the concept of evidential independence on different frameworks, i. e. joint basic assignment can be obtained from the marginals using Dempster rule.

This notion can be generalised in various ways [3, 15, 18]; the concept of conditional non-interactivity from [3], based on conjunctive combination rule (5), is used for construction of directed evidential networks in [4] (cf. also Section 5.1). In this paper we will use the concept of conditional independence introduced in [11, 18], utilizing the concept of a join as a generalization of Cartesian product (cf. also end of Section 2.1).

Definition 4.2. 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

$$m^{\downarrow K\cup L\cup M}(C) \cdot m^{\downarrow M}(C^{\downarrow M}) \quad = \quad m^{\downarrow K\cup M}(C^{\downarrow K\cup M}) \cdot m^{\downarrow L\cup M}(C^{\downarrow L\cup M}) \tag{11}$$

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

It has been proven in [18] that this conditional independence concept satisfies the so-called 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, 15]. Contrary to these conditional independence concepts, it is *consistent with marginalisation* [16]; in other words, the multidimensional model of conditionally independent variables keeps the original marginals (for more details see [18]).

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 — in other words, is irrelevant to — the evidence concerning the other variable.

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

$$m_{X_K|X_L}(A|B) = m(A) \tag{12}$$

for any $A \subseteq \mathbf{X}_K$.[4]

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 the Dempster conditioning rule nor that based on focusing, imply independence even in cases when the relation is symmetric, as can be seen from examples in [20].

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

$$m_{X_K|X_L X_M}(A|B) = m_{X_K|X_M}(A|B^{\downarrow M}) \tag{13}$$

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 (12) and (13) stands for an abstract conditioning rule (any of those mentioned in the previous section or elsewhere [8]).

[4] Let us note that a somewhat weaker definition of irrelevance can be found in [2], where equality is substituted by proportionality. The reason for this weakening is the simple fact that the authors do not require normality. This notion has later been generalised using the conjunctive combination rule [3].

However, the validity of (12) and (13) may depend on the choice of the conditioning rule, as we showed in [20] — more precisely, irrelevance with respect to one conditioning rule need not imply irrelevance with respect to the other.

4.3. Relationship between independence and irrelevance

As mentioned at the end of the preceding section, different conditioning rules lead to different irrelevance concepts. Nevertheless, when studying the relationship between (conditional) independence and irrelevance based on the Dempster conditioning rule and focusing, we realised that they do not differ too much from each other, as suggested by the following summary of results presented in [20].

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 [20].

Theorem 4.3. 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

$$m_{X_K|_F X_L X_M}(A|_F B) = m_{X_K|_F X_M}(A|_F B^{\downarrow M}) \tag{14}$$

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

From this point of view, focusing seems to be slightly superior to the Dempster conditioning rule, but it is still not satisfactory. However, the new conditioning rule introduced by Definition 3.1 is more promising, as we can see in the following theorem.

Theorem 4.4. 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 3.1.

P r o o f. Let X_K and X_L be conditionally independent given X_M; then for any $A \subseteq \mathbf{X}_{K\cup L\cup M}$ such that $A = A^{\downarrow K\cup M} \bowtie A^{\downarrow L\cup M}$

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

and $m(A) = 0$ otherwise. From this equality we immediately obtain that for all A such that $m^{\downarrow L\cup M}(A^{\downarrow L\cup M}) > 0$ (it also implies that $m^{\downarrow M}(A^{\downarrow M}) > 0$) equality

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

is satisfied. Let us note that the left-hand side of this equality is equal to $m_{X_K|X_{L\cup M}}(A^{\downarrow K}|A^{\downarrow L\cup M})$, because of the fact that A is the only focal element whose projections on $\mathbf{X}_K$ and $\mathbf{X}_{L\cup M}$ are equal $A^{\downarrow K}$ and $A^{\downarrow L\cup M}$, respectively. Under similar considerations one arrives to the fact that the right-hand side equals $m_{X_K|X_M}(A^{\downarrow K}|A^{\downarrow M})$, which means that X_L is irrelevant to X_K given X_M. □

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

Example 4.5. Let X_1 and X_2 be two binary variables (with values in $\mathbf{X}_i = \{a_i, \bar{a}_i\}$) with joint basic assignment m defined as follows:

$$\begin{aligned} &m(\{a_1a_2\}) = .25,\\ &m(\{a_1\}\times\mathbf{X}_2) = .25,\\ &m(\mathbf{X}_1\times\{a_2\}) = .25,\\ &m(\mathbf{X}_1\times\mathbf{X}_2\setminus\{\bar{a}_1\bar{a}_2\}) = .25. \end{aligned}$$

From these values one can obtain

$$m^{\downarrow 2}(\{a_2\}) = m^{\downarrow 2}(\mathbf{X}_2) = .5.$$

Evidently, it is not possible to condition by $\{\bar{a}_2\}$; so we have to confine ourselves to conditioning by $\{a_2\}$ and $\mathbf{X}_2$:

$$\begin{aligned} m_{X_1|X_2}(\{a_1\}|\{a_2\}) &= m_{X_1|X_2}(\{a_1\}|\mathbf{X}_2) = .5 = m^{\downarrow 1}(\{a_1\}),\\ m_{X_1|X_2}(\{\bar{a}_1\}|\{a_2\}) &= m_{X_1|X_2}(\{\bar{a}_1\}|\mathbf{X}_2) = \ 0 = m^{\downarrow 1}(\{\bar{a}_1\}),\\ m_{X_1|X_2}(\mathbf{X}_1|\{a_2\}) &= \ m_{X_1|X_2}(\mathbf{X}_1|\mathbf{X}_2) \ = .5 = m^{\downarrow 1}(\mathbf{X}_1), \end{aligned}$$

i. e., X_1 and X_2 are irrelevant. But they are not independent, as the focal elements are not rectangles, which contradicts Definition 4.1. ◇

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

Theorem 4.6. Let K, L, M be disjoint subsets of N such that $K, L \neq \emptyset$; $m_{X_K|X_M}$ be a (given) conditional basic assignment of X_K given X_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 3.1, then X_K and X_L are independent given X_M (with respect to a joint basic assignment $m = m_{X_K|X_M}\cdot m_{X_{L\cup M}}$[5] defined on $\mathbf{X}_{K\cup L\cup M}$).

[5] Let us note that due to Theorem 3.2 $m_{X_{L\cup M}}$ is marginal to m and $m_{X_K|X_M}$ can be re-obtained from m via Definition 3.1.

P r o o f. Irrelevance of X_L to X_K given X_M means that for any $A \subseteq \mathbf{X}_K$ and any $B \subseteq \mathbf{X}_{L\cup M}$ such that $m^{\downarrow L\cup M}(B) > 0$

$$m_{X_K|X_{L\cup M}}(A|B) = m_{X_K|X_M}(A|B^{\downarrow M}).$$

Multiplying both sides of this equality by $m^{\downarrow L\cup M}(B) \cdot m^{\downarrow M}(B^{\downarrow M})$ one obtains

$$\begin{aligned} m_{X_K|X_{L\cup M}}(A|B) \cdot m^{\downarrow L\cup M}(B) \cdot m^{\downarrow M}(B^{\downarrow M}) \\ = m_{X_K|X_M}(A|B^{\downarrow M}) \cdot m^{\downarrow L\cup M}(B) \cdot m^{\downarrow M}(B^{\downarrow M}), \end{aligned}$$

which is equivalent to

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

Therefore, the equality (11) is satisfied for $C \subseteq \mathbf{X}_{K\cup L\cup M}$ such that $C = A \times B$, where $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_{L\cup M}$. Due to Theorem 3.2 it is evident that

$$\sum_{A\subseteq \mathbf{X}_K, B\subseteq \mathbf{X}_{L\cup M}} m(A \times B) = 1,$$

and therefore equality (11) is also trivially satisfied for any other $C = C^{\downarrow K\cup M} \bowtie C^{\downarrow L\cup M}$, and $m(C) = 0$ otherwise as well. Therefore, X_K and X_L are independent given X_M with respect to a joint basic assignment $m = m_{X_K|X_M} \cdot m_{X_{L\cup M}}$. □

This theorem makes it possible to define evidential networks in a way analogous to Bayesian networks, but simultaneously a question arises: are these networks advantageous in comparison with other multidimensional models in this framework? The following section (more precisely, its last subsection) brings at least a partial answer to this question.

5. EVIDENTIAL NETWORKS AND COMPOSITIONAL MODELS

In this section we will deal with directed evidential networks [4] and evidential networks studied in Section 5.3. These two models differ not only in the conditioning rule, but also, and it seems to be more important, by the interpretation of the graph structure of the model, as we shall see later.

5.1. Directed evidential networks

In this subsection we will deal with directed evidential networks which differ, as already mentioned at the beginning of this section, by the conditioning rule and by the interpretation of the graph structure of the model.

In directed evidential networks [4] conditional beliefs are assigned to arcs, i. e., as many conditionals are assigned to every node as is the number of its parents. These conditionals are subsequently combined by the conjunctive combination rule (5).

In this section we will, using a simple example, demonstrate problems caused by this approach.

$A \subseteq \mathbf{C}_i$	$m_i(A)$
$\{h_i\}$	.49
$\{t_i\}$	.49
$\{h_i, t_i\}$	.02

	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$
$\{h_1\}$	.24	.24	.01
$\{t_1\}$	.24	.24	.01
$\{h_1, t_1\}$	.01	.01	~ 0

Tab. 1. Basic assignments m_i and joint basic assignments $m_{12} = m_1 \cdot m_2$.

m	$\{b\}$			$\{\bar{b}\}$			$\{b, \bar{b}\}$		
	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$
$\{h_1\}$	.24	0	0	0	.24	0	0	0	.01
$\{t_1\}$	0	.24	0	.24	0	0	0	0	.01
$\{h_1, t_1\}$	0	0	0	0	0	0	.01	.01	~ 0

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

Example 5.1. Let us consider two fair coin tosses 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 toss 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 a random set independence at the beginning of Section 4 and the right part of Table 1).

Now, let us consider one more variable, B, expressing the fact that 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 1 and (due to deterministic dependence of the values of B on the values of C_1 and C_2) the joint basic assignment of the three variables is in Table 2.

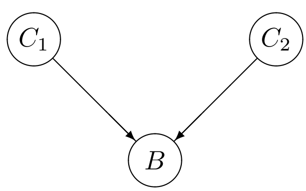

Fig. 1. Directed graph from Example 2.

The approach suggested by Ben Yaghlane et al [4] starts 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

$A \subseteq \mathbf{C}_i$	$m_i(A)$
$\{h_i\}$	.49
$\{t_1\}$	.49
$\{h_1, t_1\}$	.02

$D \subseteq \mathbf{B}$	$m_{.\|i}(D)$
$\{b\}$	.49
$\{\bar{b}\}$	.49
$\{b, \bar{b}\}$	.02

Tab. 3. Basic assignments m_i and conditional basic assignments $m_{.|i}$.

m	$\{b^*\}$			$\{b, \bar{b}\}$		
	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$	$\{h_2\}$	$\{t_2\}$	$\{h_2, t_2\}$
$\{h_1\}$	.0624	.0624	.0025	.0001	.0001	~ 0
$\{t_1\}$	.0624	.0624	.0025	.0001	.0001	~ 0
$\{h_1, t_1\}$	.0025	.0025	.0001	~ 0	~ 0	~ 0

Tab. 4. Joint basic assignment of variables C_1, C_2 and B based on the conjunctive combination rule; b^* stands for either b or $\bar{b}$.

problem caused by this approach 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 $C_i, i = 1, 2$ can be found in the right part of Table 3. 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 the event that the bell rings also does not depend on the result of one coin.

The values of joint basic assignments are computed from Table 3 using the (non-normalised) conjunctive combination rule. Results of these computations can be found in Table 4.

It is evident that the independence (non-interactivity) between coins C_1 and C_2 has been substituted by conditional non-interactivity, which does not make sense, as C_1 is strongly dependent on C_2 whenever B is known. $\diamondsuit$

In the last subsection of this section we will show how to model this problem using our approach to evidential networks. Before doing that we have to recall the concept of compositional models.

5.2. Compositional models

Compositional models are based on the concept of the operator of composition of basic assignments, introduced in [12] in the following way.

Definition 5.2. 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 [11, 12]) it follows that the operator of composition is not commutative in general, but it preserves the first marginal (in case of projective basic assignments both of them). In both of these aspects it differs from the conjunctive combination rule. Let us illustrate this fact by the following simple example.

Example 5.3. Let X_1, X_2 and X_3 be three binary variables with values in $\mathbf{X}_1 = \{a_1, \bar{a}_1\}, \mathbf{X}_2 = \{a_2, \bar{a}_2\}$, $\mathbf{X}_3 = \{a_3, \bar{a}_3\}$ and m_1 and m_2 be two basic assignments on $\mathbf{X}_1 \times \mathbf{X}_3$ and $\mathbf{X}_2 \times \mathbf{X}_3$ respectively, both of them having only two focal elements:

$$\begin{aligned} m_1(\{(a_1, \bar{a}_3), (\bar{a}_1, \bar{a}_3)\}) &= .5, \\ m_1(\{(a_1, \bar{a}_3), (\bar{a}_1, a_3)\}) &= .5, \\ m_2(\{(a_2, \bar{a}_3), (\bar{a}_2, \bar{a}_3)\}) &= .5, \\ m_2(\{(a_2, \bar{a}_3), (\bar{a}_2, a_3)\}) &= .5. \end{aligned} \tag{15}$$

Since their marginals are projective

$$\begin{aligned} m_1^{\downarrow 3}(\{\bar{a}_3\}) &= m_2^{\downarrow 3}(\{\bar{a}_3\}) = .5, \\ m_1^{\downarrow 3}(\{a_3, \bar{a}_3\}) &= m_2^{\downarrow 3}(\{a_3, \bar{a}_3\}) = .5, \end{aligned}$$

there exists (at least one) common extension of both of them.

Applying (5) to the marginals (15) one obtains

$$\begin{aligned} (m_1 \circledcirc m_2)(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) &= .25, \\ (m_1 \circledcirc m_2)(\mathbf{X}_1 \times \{a_2\} \times \{\bar{a}_3\}) &= .25, \\ (m_1 \circledcirc m_2)(\{a_1\} \times \mathbf{X}_2 \times \{\bar{a}_3\}) &= .25, \\ (m_1 \circledcirc m_2)(\{(a_1, a_2, \bar{a}_3), (\bar{a}_1, \bar{a}_2, a_3)\}) &= .25 \end{aligned} \tag{16}$$

with marginal basic assignments

$$\begin{aligned} (m_1 \circledcirc m_2)^{\downarrow 13}(\{(a_1, \bar{a}_3), (\bar{a}_1, \bar{a}_3)\}) &= .5, \\ (m_1 \circledcirc m_2)^{\downarrow 13}(\{(a_1, \bar{a}_3)\}) &= .25, \\ (m_1 \circledcirc m_2)^{\downarrow 13}(\{(a_1, \bar{a}_3), (\bar{a}_1, a_3)\}) &= .25, \\ (m_1 \circledcirc m_2)^{\downarrow 23}(\{(a_2, \bar{a}_3), (\bar{a}_2, \bar{a}_3)\}) &= .5, \\ (m_1 \circledcirc m_2)^{\downarrow 23}(\{(a_2, \bar{a}_3)\}) &= .25, \\ (m_1 \circledcirc m_2)^{\downarrow 23}(\{(a_2, \bar{a}_3), (\bar{a}_2, a_3)\}) &= .25. \end{aligned}$$

On the other hand, composition of basic assignments m_1 and m_2 defined by (15) is by Definition 5.2

$$(m_1 \triangleright m_2)(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5, \tag{17}$$
$$(m_1 \triangleright m_2)(\{(a_1, a_2, \bar{a}_3), (\bar{a}_1, \bar{a}_2, a_3)\}) = .5, \tag{18}$$

which keeps, due to projectivity, both marginals. ♢

Furthermore, the 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 [11, 12]; here we will confine ourselves to the following theorem (proven in [11]) expressing the relationship between conditional independence and the operator of composition.

Theorem 5.4. 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 applications of the operator of composition, one gets a multidimensional 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 adopt 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:[6]

$$m_1 \triangleright m_2 \triangleright \ldots \triangleright m_n = (\ldots((m_1 \triangleright m_2) \triangleright m_3) \triangleright \ldots) \triangleright m_n. \tag{19}$$

Therefore, multidimensional model (19) is specified by an ordered sequence of low-dimensional basic assignments – a *generating sequence* $m_1, m_2, \ldots, m_n$.

5.3. 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

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

[6]Naturally, we will use parentheses to change the ordering in which the operators are to be performed.

The property explaining why we call these sequences "perfect" is expressed by the following assertion proven in [11].

Theorem 5.5. 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 present a simple algorithm for the construction of an evidential network from a perfect sequence of basic assignments.

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.[7] 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 $(X_i \rightarrow X_j)$ 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 (8).

Example 5.6. Let us illustrate this simple procedure for reconstruction of the evidential network from the perfect sequence of distributions

$$m_1(T), m_2(G, B), m_3(D, T, G), m_4(R, B), m_5(W, R, D).$$

Regarding the described process we have to start by ordering the variables. The first variable must be T, as m_1 contains only this variable. The second one must be B or G, as both are contained in m_2 and none of them in m_1, and the third one the other. Let us choose first G and then B. The fourth must be D, as it is the only variable contained in m_3 and neither in m_1 nor in m_2. Analogously, fifth one must be R and the last one W. Hence, we have the ordering

$$T, G, B, D, R, W.$$

Node T first appears among the variables of m_1. As this distribution is defined only for this single variable there is no edge leading to this node. The second variable G

[7]Let us note that variables $X_1, X_2, \ldots, X_k$ may be ordered arbitrarily, nevertheless, for the chosen ordering proof of Theorem 5.7 is much simpler than in the general case.

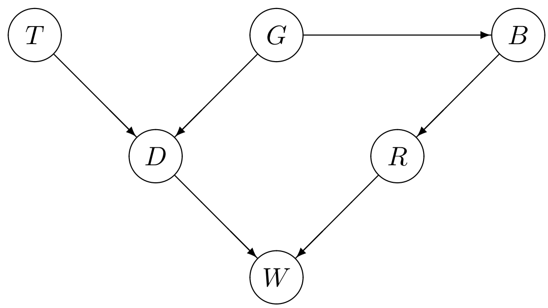

Fig. 2. Acyclic directed graph of an evidential network reconstructed from a perfect sequence m_1, m_2, m_3, m_4, m_5.

appears first in distribution $m_2(G, B)$. The arrow $(B \to G)$ would be inserted into the graph if either B would be before variable G in the selected ordering, or, if it were among the variables for which some of the preceding distributions would be defined. None of these conditions is met and therefore the edge $(B \to G)$ is not included into the graph. The third variable B appears again among the variables for which m_2 is defined. This time, however, G precedes B in the chosen ordering of the variables, and therefore edge $(G \to B)$ is included into the graph.

Variable D appears first among the arguments of m_3. Both the remaining variables T, G are among the arguments of the preceding distributions (m_1 and m_2, respectively) and therefore we have to define both the edges $(T \to D)$ and $(G \to D)$. Similarly, we have to include edges $(B \to R)$ (B is among the variables for which m_2 is defined) and $(D \to W)$ and $(R \to W)$ (D and R are among the arguments of m_3 and m_4, respectively). The resulting graph is in Figure 2.

The only arbitrariness in the ordering of variables is that of ordering of those from m_2, as none of them appeared in preceding basic assignment. If we use the ordering T, B, G, D, R, W, then the arrow $(B \to G)$ will be included instead of $(G \to B)$. $\diamondsuit$

Generally, 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 sets forth that any of them is induced by the perfect sequence.

Theorem 5.7. For a belief network defined by the above-described procedure the following independence statements are satisfied for any $j = 2, \ldots k$:

$$\{j\} \perp\!\!\!\perp (\{i : i < j\} \setminus pa(j)) \mid pa(j), \tag{20}$$

where $pa(j)$ is the set of parents of the node j.

P r o o f. Let $j \in K_\ell$, $j \notin K_1 \cup \ldots \cup K_{\ell-1}$. Due to the fact that

$$m_1 \triangleright m_2 \triangleright \ldots \triangleright m_{\ell-1} \triangleright m_\ell = (\cdots (m_1 \triangleright m_2) \triangleright \cdots \triangleright m_{\ell-1}) \triangleright m_\ell$$

and Theorem 5.4 we have that

$$K_\ell \setminus (K_1 \cup \ldots \cup K_{\ell-1}) \perp\!\!\!\perp (K_1 \cup \ldots \cup K_{\ell-1}) \setminus K_\ell \,|\, K_\ell \cap (K_1 \cup \ldots \cup K_{\ell-1}) \,. \tag{21}$$

It is evident that $(K_1 \cup \ldots \cup K_{\ell-1}) \setminus K_\ell = \{i : i < j\} \setminus pa(j)$, let us denote this set by L. Now, there are two possibilities: either $K_\ell \cap (K_1 \cup \ldots \cup K_{\ell-1}) = pa(j)$ (if j does not have any parents appearing first in K_ℓ) or $K_\ell \cap (K_1 \cup \ldots \cup K_{\ell-1}) \subset pa(j)$ (otherwise).

In the first case either $K_\ell \setminus (K_1 \cup \ldots \cup K_{\ell-1}) = \{j\}$ and we immediately obtain (20), or $K_\ell \setminus (K_1 \cup \ldots \cup K_{\ell-1}) \supset \{j\}$ and (20) follows from (21) due to $K \cup M \perp\!\!\!\perp L|I\ [m] \Rightarrow K \perp\!\!\!\perp L|I\ [m]$ (following for any mutually disjoint sets I, K, L, M from semi-graphoid properties), where $K = \{j\}$, $M = K_\ell \setminus (K_1 \cup \ldots \cup K_{\ell-1}) \setminus \{j\}$ and $I = K_\ell \cap (K_1 \cup \ldots \cup K_{\ell-1}) = pa(j)$.

In the latter case, we start by application of the implication $K \cup M \perp\!\!\!\perp L|I\ [m] \Rightarrow K \perp\!\!\!\perp L|M \cup I\ [m]$, whose validity for any mutually disjoint sets I, K, L, M again follows from the semi-graphoid properties, to $K = K_\ell \setminus (K_1 \cup \ldots \cup K_{\ell-1}) \setminus pa(j), M = (K_\ell \setminus (K_1 \cup \ldots \cup K_{\ell-1})) \cap pa(j)$ and $I = K_\ell \cap (K_1 \cup \ldots \cup K_{\ell-1})$. As $M \cup I = pa(j)$, we can then proceed analogous to the previous paragraph to obtain (20). □

Example 5.6. (Continued) Let us note that basic assignments m_1, m_2 and m_3 form a perfect sequence. The graph in Figure 1 can easily be obtained from this perfect sequence via the algorithm presented above. Together with the system of (conditional) basic assignments m_1, m_2 and $m_{B|C_1C_2}$ contained in Table 5 (due to a lack of space, x_i stands for either h_i or t_i, $i = 1, 2$, the remaining values of conditional basic assignment are undefined) it forms an evidential network.

	$\{h_1h_2\}$	$\{h_1t_2\}$	$\{t_1h_2\}$	$\{t_2t_2\}$	$\{x_1h_2, x_1t_2\}$	$\{h_1x_2, t_1x_2\}$	$\mathbf{C_1} \times \mathbf{C_2}$
$\{b\}$	1	0	0	1	0	0	0
$\{\bar{b}\}$	0	1	1	0	0	0	0
$\{b, \bar{b}\}$	0	0	0	0	1	1	1

Tab. 5. Conditional basic assignment $m_{B|C_1C_2}$ of variable B given C_1 and C_2.

From Table 5 the difference between the approach of Ben Yaghlane et al. [4] and our models is evident. Here the conditional dependence between the results on the two coins given the ringing of the bell (or not) is kept. The resulting model clearly coincides with values contained in Table 2. ♢

5.4. Evidential network vs. compositional model

At the end of Section 4 we asked the question whether evidential networks are advantageous in comparison with other multidimensional models in this framework.

In the preceding subsection we showed that they are more powerful than directed evidential networks, as they are able to distinguish between conditional and unconditional independence. Nevertheless, the following example demonstrates that although any compositional model can be transformed into an evidential network, it may happen that the latter is more imprecise than the former.

Example 5.8. 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

$$m(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5, \tag{22}$$
$$m(\{(a_1, a_2, \bar{a}_3), (\bar{a}_1, \bar{a}_2, a_3)\}) = .5. \tag{23}$$

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

$$m_{X_1|X_{23}}(A|B) = m_{X_1|X_3}(A|B^{\downarrow\{3\}}), \tag{24}$$

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

$$m_{X_1|X_{23}}(\mathbf{X}_1|\mathbf{X}_2 \times \{\bar{a}_3\}) = m_{X_1|X_3}(\mathbf{X}_1|\{\bar{a}_3\}) = 1, \tag{25}$$
$$m_{X_1|X_{23}}(\mathbf{X}_1|\{(a_2, \bar{a}_3), (\bar{a}_2, a_3)\}) = m_{X_1|X_3}(\mathbf{X}_1|\mathbf{X}_3) = 1. \tag{26}$$

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

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

The conditional independence of X_1 and X_2 given X_3 and relation (24) correspond to a directed graph in Figure 3, which leads to the following system of (conditional) basic assignments:

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

and $m_{X_1|X_3}$ as suggested in the right-hand side of (25) and (26).

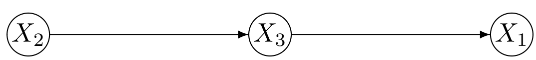

Fig. 3. Graph G from Example 5.8.

The final model

$$\check{m}(\mathbf{X}_1 \times \mathbf{X}_2 \times \{\bar{a}_3\}) = .5, \tag{27}$$

$$\check{m}(\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3) = .5. \tag{28}$$

is again different, since instead of basic assignment $m^{\downarrow 23}$ we used its marginal and conditional.

Therefore it is clearly seen that evidential networks are less powerful than, e. g., compositional models [11], as any of these three-dimensional basic assignments can be obtained from two two-dimensional ones using the operator of composition. $\diamondsuit$

Let us also note that both the original model (22) and (23), as well as the last one (27) and (28), can be expressed with the help of an undirected graph. Both of these models factorise with respect to the graph in Figure 4.

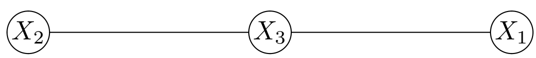

Fig. 4. Undirected graph G.

The difference between compositional models and undirected models is due to the fact that compositional models are defined also for non-projective basic assignments, in contrast to decomposable models.

6. CONCLUSIONS

This paper is devoted to two kinds of multidimensional models with directed graph structure, namely, directed evidential networks [4] and evidential networks. While the former models are mentioned only marginally, as a motivation for our research, the latter are studied in more detail.

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

We presented a new conditioning rule for variables, which appeared to be suitable for the definition of (conditional) irrelevance having a sensible relationship with conditional independence. This allows us to define evidential networks in a way analogous to Bayesian networks.

Despite the fact that these models are able to distinguish between unconditional and conditional independence, they are still not an optimal model. Their weakness consists in conditioning, which may destroy the structure of the original focal elements and may lead to more imprecise models.

From this point of view, compositional models or those based on undirected graphs seem to be more appropriate multidimensional models in the framework of evidence theory than these two kinds of networks.

7. ACKNOWLEDGEMENT

The support by Grant GA ČR 13-20012S is gratefully acknowledged.

(Received March 20, 2013)

REFERENCES

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

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

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

[4] B. Ben Yaghlane, Ph. Smets, and K. Mellouli: Directed evidential networks with conditional belief functions. In: Proc. ECSQARU 2003 (T. D. Nielsen and N. L. Zhang, eds.), pp. 291–305.

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

[6] I. Couso, S. Moral, P. Walley: Examples of independence for imprecise probabilities. In: Proc. ISIPTA'99 (G. de Cooman, F. G. Cozman, S. Moral, and P. Walley, eds.), pp. 121–130.

[7] F. G. Cozman: Credal networks. Artificial Intelligence J. *120* (2000), 199–233.

[8] M. Daniel: Belief conditioning rules for classic belief functions. In: Proc. WUPES'09 (T. Kroupa and J. Vejnarová, eds.), pp. 46–56.

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

[10] J. W. Guan and D. A. Bell: Evidence Theory and its Applications. Vol. 1. North-Holland, 1991.

[11] R. Jiroušek and J. Vejnarová: Compositional models and conditional independence in evidence theory. Internat. J. Approx. Reasoning *52* (2011), 316–334.

[12] R. Jiroušek, J. Vejnarová, and M. Daniel: Compositional models for belief functions. In: Proc. ISIPTA'07 (G. De Cooman, J. Vejnarová, and M. Zaffalon, eds.) Praha, pp. 243–252.

[13] A. Kong: Multivariate Belief Functions and Graphical Models. Doctoral disertation, Department of Statistics, Harvard University, 1986.

[14] G. Shafer: A Mathematical Theory of Evidence. Princeton University Press, Princeton, New Jersey 1976.

[15] P. P. Shenoy: Conditional independence in valuation-based systems. Internat. J. Approx. Reasoning *10* (1994), 203–234.

[16] M. Studený: Formal properties of conditional independence in different calculi of artificial intelligence. In: Proc. ECSQARU'93 (K. Clarke, R. Kruse, and S. Moral, eds.), Springer-Verlag, 1993, pp. 341–348.

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

[18] J. Vejnarová: On conditional independence in evidence theory. In: Proc. ISIPTA'09 (T. Augustin, F. P A. Coolen, S. Moral, M. C. M. Troffaes, eds.), Durham 2009, pp. 431–440.

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

[20] J. Vejnarová: Conditioning, conditional independence and irrelevance in evidence theory. In: Proc. ISIPTA'11 (F. Coolen, G. de Cooman, T. Fetz, and M. Oberguggenberger, eds.), Innsbruck 2011, pp. 381–390.

[21] J. Vejnarová: Conditioning in evidence theory from the perspective of multidimensional models. In: Proc. IPMU'12 (S. Greco et al., eds.), Part III, CCIS 299, 2012, pp. 450–459.

[22] J. Vejnarová: Evidential networks from a different perspective. In: Synergies of Soft Computing and Statistics for Intelligent Data Analysis, Soft Methods In Probability and Statistics, Heidelberg 2012, pp. 429–436.

[23] J. Vejnarová: On weakness of evidential networks. In: Proc. 9th Workshop on Uncertainty Processing, pp. 190–203.

[24] H. Xu and Ph. Smets: Evidential reasoning with conditional belief functions. In: Proc. Tenth Conference on Uncertainty in Artificial Intelligence (UAI'94), pp. 598–605.

Jiřina Vejnarová, Institute of Information Theory and Automation — Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: vejnar@utia.cas.cz