A Sufficient Condition for Belief Function Construction From Conditional Belief Functions

Mieczyslaw A. Klopotek and Sławomir T. Wierzchoń
Institute of Computer Science, Polish Academy of Sciences
PL 01-237 Warsaw, 21 Ordona St., Fax: (48-22) 37-65-64
Phone: (48-22) 36-28-85 ext. 45,
e-mail: klopotek@ipipan.waw.pl, http://www.ipipan.waw.pl/~klopotek/

Key words: evidence theory, graphoidal structures, conditional belief functions
for **V Workshop sobre Aspectos Teoricos de la Inteligencia Artificial ATIA'98**

SUMMARY

It is commonly acknowledged that we need to accept and handle uncertainty when reasoning with real world data. The most profoundly studied measure of uncertainty is the probability. However, the general feeling is that probability cannot express all types of uncertainty, including vagueness and incompleteness of knowledge. The Mathematical Theory of Evidence or the Dempster-Shafer Theory (DST) [1, 12] has been intensely investigated in the past as a means of expressing incomplete knowledge. The interesting property in this context is that DST formally fits into the framework of graphoidal structures [13] which implies possibilities of efficient reasoning by local computations in large multivariate belief distributions given a factorization of the belief distribution into low dimensional component conditional belief functions. But the concept of conditional belief functions is generally not usable because composition of conditional belief functions is not granted to yield joint multivariate belief distribution, as some values of the belief distribution may turn out to be negative [4, 13, 15].

To overcome this problem creation of an adequate frequency model is needed. In this paper we suggest that a Dempster-Shafer distribution results from "clustering" (merging) of objects sharing common features. Upon "clustering" two (or more) objects become indistinguishable (will be counted as one) but some attributes will behave as if they have more than one value at once. The next elements of the model needed are the concept of conditional independence and that of merger conditions. It is assumed that before merger the objects move closer in such a way that conditional distributions of features for the objects to merge are identical. The traditional conditional independence of feature variables is assumed before merger (thereafter only the DST conditional independence holds). Furthermore it is necessary that the objects get "closer" before the merger independly for each feature variable and only those areas merge where the conditional distributions get identical in each variable.

The paper demonstrates that within this model, the graphoidal properties hold and a sufficient condition for non-negativity of the graphoidally represented belief function is presented and its validity demonstrated.

A Sufficient Condition for Belief Function Construction From Conditional Belief Functions

1 Introduction

It is commonly acknowledged that we need to accept and handle uncertainty when reasoning with real world data. The most profoundly studied measure of uncertainty is the probability. There exist methods of so-called graphoidal representation of joint probability distribution - called Bayesian networks [11] - allowing for expression of qualitative independence, causality, efficient reasoning, explanation, learning from data and sample generation. However, the general feeling is that probability cannot express all types of uncertainty, including vagueness and incompleteness of knowledge. The Mathematical Theory of Evidence or the Dempster-Shafer Theory (DST) [1, 12] has been intensely investigated in the past as a means of expressing incomplete knowledge. The interesting property in this context is that DST formally fits into the framework of graphoidal structures [13] which implies possibilities of efficient reasoning by local computations in large multivariate belief distributions given a factorization of the belief distribution into low dimensional component conditional belief functions [16]. This in turn could qualify DST for usage in expert systems dealing with uncertainty. But the concept of conditional belief functions is generally hardly usable for representation of belief functions in learning and sample generation because composition of conditional belief functions is not granted to yield joint multivariate belief distribution, as some values of the belief distribution may turn out to be negative [4, 13, 15]. We call this the well-formedness problem of decomposition into conditional belief functions.

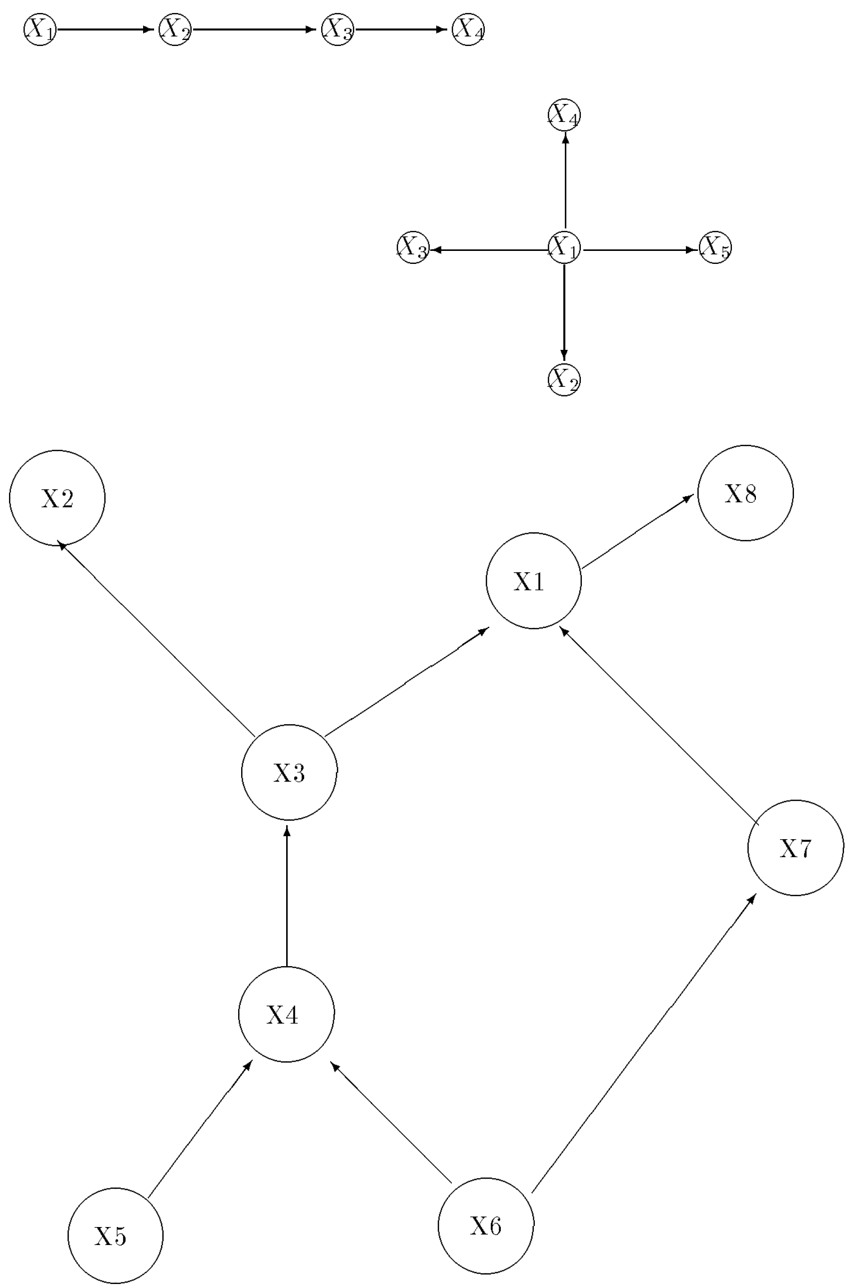
To overcome this problem creation of an adequate frequency model of DST is needed. In this paper we suggest that a Dempster-Shafer distribution results from "clustering" (merging) of objects sharing common features. Upon "clustering" two (or more) objects become indistinguishable (will be counted as one) but some attributes will behave as if they have more than one value at once. E.g. a paper written by two (co-)authors may be considered as a merger of two ideas of the two authors (two "papers") that only at some points (for some attributes) may be clearly separated into the contributions of each of them.

This of course does not suffice. The next elements of the model needed are the concept of conditional independence and that of merger conditions. It is assumed that before merger the objects move closer in such a way that conditional distributions of features for the objects to merge are identical. The traditional conditional independence of feature variables is assumed before merger (thereafter only the DST conditional independence holds). Furthermore it is necessary that the objects get "closer" before the merger independly for each feature variable and only those areas merge where the conditional distributions get identical in each variable.

The paper demonstrates that within this model, the graphoidal properties hold and a sufficient condition for non-negativity of the graphoidally represented belief function is presented and its validity proven.

2 Formal Definition of the Dempster-Shafer Theory of Evidence

Let us make the remark that if an object is described by a set of discrete attributes $X_1, X_2, ..., X_n$ taking values from their respective domains $\Xi_1, \Xi_2, ..., \Xi_n$ then we can think of it as being described by a complex attribute X having vector values, that is the domain



(a)

(b)

(c)

Figure 1: a) A chain-like bayesian network. b) A star-like bayesian network c) A general bayesian network, with generated data available at *http* : *//www.ipipan.waw.pl/˜klopotek/ds/szampony.zip*

Ξ of X is equal:

$$\Xi = \{(x_1, x_2, ..., x_n) | x_i \in \Xi_i \forall i = 1, ..., n\}$$

.

So let X below be a complex attribute in the above-mentioned sense. We say that Ξ, the domain of X is our space of discourse spanned by the attribute X. We shall also briefly say that X is our space of discourse instead.

Let us define the basic functions of the DST.

The function m is defined as

Definition 1 *The* Pseudo-Mass Function m *is defined as* m*:*$2^\Xi \to [-1, 1]$ *with*

$$\sum_{A \in 2^\Xi} m(A) = 1$$

$$m(\emptyset) = 0$$

$$\forall_{A \in 2^\Xi} \quad 0 \le \sum_{A \subseteq B} m(B)$$

If additionally

$$\forall_{A \in 2^\Xi} \quad m(A) \ge 0$$

holds, then we shall call m the (proper) mass function.

Definition 2 *Whenever* $m(A) \ne 0$*, we say that* A *is the focal point of the Bel-Function.*

We define the Bel-function as follows.

Definition 3 *The* Pseudo-Belief Function *is defined as* Bel*:*$2^\Xi \to [-1, 1]$ *with* $\Xi = \Xi_1 \times Xi_2 \times ... \times \Xi_n$ *being the space spanned by the attribute* $X = X_1 \times X_2 \times \ldots \times X_n$ *with*

$$Bel(A) = \sum_{B \subseteq A} m(B)$$

for any set $A \in 2^\Xi$ *where* $m(A)$ *is a Pseudo-Mass Function (see def.1 above). If additionally* m *is a (proper) mass function, then* Bel *above is called (proper) belief function.*

Generally, only proper mass, belief, commonality and plausibility functions are studied. The corresponding pseudo-functions are result of conditioning in DST and - as they take negative values - they have no direct interpretation.

Let us make the remark, that the above def.3 is easily derived from the usual axiomatic system for DST (compare [2]).

Let us also introduce the Pl-Function:

Definition 4 *The* Pseudo-Plausibility Function *is defined as*

$$\forall_{A \in 2^\Xi} \ Pl(A) = 1 - Bel(\Xi - A)$$

where Bel *is a pseudo-belief function. If* Bel *above is a (proper) belief function, then* Pl *is a (proper) plausibility function.*

For completeness let us recall also the Q-Function of the DS-Theory.

Table 1: a) Marginal distribution of X_1 in Fig.1a,b), b) - conditional distributions in Fig.1a), c) - conditionals in Fig.1b).

(a)

X_1	m_{X_1}
{a}	0.4
{b}	0.4
{a,b}	0.2

(b)

X_i	X_{i+1}	$m_{X_{i+1}\|X_i}$
{a} ×	{a}	0.293333
{a} ×	{b}	-0.126667
{a} ×	{a,b}	-0.166667
{b} ×	{a}	-0.126667
{b} ×	{b}	0.293333
{b} ×	{a,b}	-0.166667
{a,b} ×	{a}	0.3
{a,b} ×	{b}	0.3
{a,b} ×	{a,b}	0.4

(c)

X_1	X_i	$m_{X_i\|X_1}$
{a} ×	{a}	0.166667
{a} ×	{b}	-0.0833333
{a} ×	{a,b}	-0.0833333
{b} ×	{a}	-0.0833333
{b} ×	{b}	0.166667
{b} ×	{a,b}	-0.0833333
{a,b} ×	{a}	0.35
{a,b} ×	{b}	0.35
{a,b} ×	{a,b}	0.3

Definition 5 *The (proper)* Commonality Function Q *is defined as* $Q{:}2^{\Xi} \to [0,1]$ *with* $\Xi = \Xi_1 \times Xi_2 \times ... \times \Xi_n$ *being the space spanned by the attribute* $X = X_1 \times X_2 \times \ldots \times X_n$ *with*

$$\forall_{A \in 2^{\Xi}}\ Q(A) = \sum_{A \subseteq B} m(B)$$

where m(A) is a Pseudo-Mass Function in the sense of the DS-Theory (see def.1 above and notice the difference to def.3 in that the sum is taken over supersets, not subsets of A).

Notice that there exists no Pseudo-Commonmality function.

Beside the above definition a characteristic feature of the DS-Theory is the so-called DS-rule of combination of independent evidence:

Definition 6 *Let* Bel_{E_1} *and* Bel_{E_2} *represent independent information over the same space of discourse. Then:*

$$Bel_{E_1,E_2} = Bel_{E_1} \oplus Bel_{E_2}$$

calculated as:

$$m_{E_1,E_2}(A) = c \cdot \sum_{B,C;A=B\cap C} m_{E_1}(B) \cdot m_{E_2}(C)$$

(c - normalizing constant) represents the Combined Belief-Function of Two Independent Beliefs

Let us also introduce the marginalization and extension operations: first for sets.

Definition 7 *Let* $X = X_1 \times X_2 \times \ldots \times X_n$ *and* $\Xi = \Xi_1 \times \Xi_2 \times \ldots \times \Xi_n$*. Let A be a set* $A \in 2^{\Xi}$*. Let* $Y = X_{i_1} \times X_{i_2} \times \ldots \times X_{i_k}$*, where indices* $\{i_1, \ldots, i_k\}$ *are all distinct and are subset of {1,...,n}. The set B is the projection (marginalization) of the set A onto the (sub)space Y (denoted* $B = A^{\downarrow Y}$*) iff for every element* $(v_1, \ldots, v_n) \in A$ *the element* $(v_{i_1}, v_{i_2}, \ldots, v_{i_k})$ *belongs to B.*
We shall say also say that A is an extension of B.

We shall distinguish one special extension: the empty extension.

Definition 8 *Let $X = p \cup q$ where p, q are disjoint sets of variables. Ξ, Ξ_p, Ξ_q be spaces spanned by X, p, q. Let $B \subseteq \Xi_p$. Let $A \subseteq \Xi_p \times \Xi_q$ such that $A = B \times \Xi_q$. Then we say that A is the empty extension of B, denoted $A = B^{\uparrow X}$.*

Now let us define the marginalization and extension for Bel-Functions:

Definition 9 *Let p, q be disjoint sets of variables. Let $X = p \cup q$ be our space of discourse for which the m, and its Bel, Pl, Q functions are defined. The m function marginalized (projected) onto the subspace p, denoted as $m^{\downarrow p}$ is defined as:*

$$\forall_{B;B\subseteq \Xi_p} m^{\downarrow p}(B) = \sum_{A;B=A^{\downarrow p}} m(A)$$

.The functions $Bel^{\downarrow p}$, $Pl^{\downarrow p}$, $Q^{\downarrow p}$ are defined accordingly to Bel and Pl definitions above with respect to $m^{\downarrow p}$ as their mass function.

Definition 10 *Let p be our space of discourse for which the m, and its Bel, Pl, Q functions are defined. The m function empty-extended onto the superspace $X = p \cup q$, denoted as $m^{\uparrow X}$ is defined as:*

$$\forall_{A;A\subseteq \Xi, A=(A^{\downarrow p})^{\uparrow X}} \quad m^{\uparrow X}(A) = m(A^{\downarrow p})$$

and

$$\forall_{A;A\subseteq \Xi, A\neq (A^{\downarrow p})^{\uparrow X}} \quad m^{\uparrow X}(A) = 0$$

otherwise. The functions $Bel^{\uparrow X}$, $Pl^{\uparrow X}$ $Q^{\uparrow X}$ are defined accordingly to Bel and Pl definitions above with respect to $m^{\uparrow X}$ as their mass function.

Please notice that the operator $\oplus$ is defined for combination of Bel's only for the same space of discourse. Should it happen, however that Bel_1 is defined over the space $p \cup r$, and Bel_2 over $q \cup r$, then instead of writing:

$$Bel_{1,2} = Bel_1^{\uparrow p\cup q\cup r} \oplus B_2^{\uparrow\uparrow p\cup q\cup r}$$

we will simply write

$$Bel_{1,2} = Bel_1 \oplus Bel_1$$

whenever no misunderstandings may occur.

We shall define also the conditional belief function $Bel^{|q}$.

Definition 11 *Let Bel be a belief function defined for the set of variables X and let q be a subset of this set. Then $Bel^{|q}$ is any pseudo-belief function fitting the equation*

$$Bel = Bel^{|q} \oplus Bel^{\downarrow q}$$

Let us also say what we mean by variable independence.

Definition 12 *Let Bel be a pseudo-belief function defined for the set of variables X and let p, q be non-intersecting subsets of this set. We say that p, q are independent, $p \perp q$, iff*

$$Bel^{\downarrow p\cup q} = Bel^{\downarrow p} \oplus Bel^{\downarrow q}$$

Definition 13 *Let Bel be a pseudo-belief function defined for the set of variables X and let p, q, r be non-intersecting subsets of this set. We say that p, q are independent given r $p \perp q | r$ iff*

$$Bel^{\downarrow p\cup q\cup r|r} \oplus Bel^{\downarrow r} = Bel^{\downarrow p\cup r|r} \oplus Bel^{\downarrow q\cup r|r} \oplus Bel^{\downarrow r}$$

Remark: Instead of writing $Bel^{\downarrow p\cup r|r}$ we shall write for brevity $Bel_{p|r}$.

Table 2: The function $m = m_{X_1} \oplus m_{X_2|X_1} \oplus m_{X_3|X_2} \oplus m_{X_4|X_3}$ is not a proper belief function

X1	X2	X3	X4	m
		....		...
{a} ×	{a} ×	{a} ×	{a}	0.131974
{a} ×	{a} ×	{a} ×	{b}	0.0260108
{a} ×	{a} ×	{a} ×	{a,b}	0.0352418
{a} ×	{a} ×	{b} ×	{a}	9.40444e-05
{a} ×	{a} ×	{b} ×	{b}	0.0432588
{a} ×	{a} ×	{b} ×	{a,b}	0.000353778
{a} ×	{a} ×	{a,b} ×	{a}	0.01772
{a} ×	{a} ×	{a,b} ×	{b}	0.01772
{a} ×	{a} ×	{a,b} ×	{a,b}	0.0236267
{a} ×	{b} ×	{a} ×	{a}	9.40444e-05
{a} ×	{b} ×	{a} ×	{b}	-2.91556e-05
{a} ×	{b} ×	{a} ×	{a,b}	-3.82222e-05
		....		...

3 Graphoidal properties

We say that a conditional independence relation has graphoidal properties if the following holds:

1. (symmetry) Suppose r,s and v are disjoint subsets of w. Then $r \perp s | v$ iff $s \perp r | v$
2. (decomposition) Suppose r,s,t and v are disjoint subsets of w. If $r \perp s \cup t | v$ then $r \perp s | v$
3. (weak union) Suppose r,s,t and v are disjoint subsets of w. If $r \perp s \cup t | v$ then $r \perp s | v \cup t$
4. (contraction) Suppose r,s,t and v are disjoint subsets of w. If $r \perp s | v$ and $r \perp t | v \cup s$ then $r \perp s \cup t | v$
5. (intersection) Suppose r,s,t and v are disjoint subsets of w. If $r \perp s | v \cup t$ and $r \perp t | v \cup s$ then $r \perp s \cup t | v$

The proof of these properties for the DST is due to [13, p.215-219] and [6] the last property (intersection) has been proven for a wider class of belief functions than in [13].

The graphidal properties allow to define a DST bayesian network as

Definition 14 *[7] A belief network is a pair (D, Bel) where D is a dag (directed acyclic graph) and Bel is a belief function called the* underlying distribution*. Each node i in D corresponds to a variable X_i in Bel, a set of nodes I corresponds to a set of variables X_I and x_i, x_I denote values drawn from the domain of X_i and from the (cross product) domain of X_I respectively. Each node in the network is regarded as a storage cell for any distribution $Bel^{\downarrow\{X_i\}\cup X_{\pi(i)}|X_{\pi(i)}}$ where $X_{\pi(i)}$ is a set of nodes corresponding to the parent nodes $\pi(i)$ of i. The underlying distribution represented by a belief network is computed via:*

$$Bel = \bigoplus_{i=1}^{n} (Bel^{\downarrow\{X_i\}\cup X_{\pi(i)}})^{|X_{\pi(i)}}$$

Please notice the local character of valuation of a node: to valuate the node i corresponding to variable X_i only the marginal $Bel^{\downarrow\{X_i\}\cup X_{\pi(i)}}$ needs to be known (e.g. from data) and not the entire belief function.

Table 3: the function $m = m_{X_1} \oplus m_{X_2|X_1} \oplus m_{X_3|X_1} \oplus m_{X_4|X_1} \oplus m_{X_5|X_1}$ is not a proper belief function

X1	X2	X3	X4	X5	m
		....			...
{a} ×	{b} ×	{b} ×	{b} ×	{a,b}	-0.000107315
{a} ×	{b} ×	{b} ×	{a,b} ×	{a}	0.0022038
{a} ×	{b} ×	{b} ×	{a,b} ×	{b}	-0.000107315
		....			...

4 The problem

Let us illustrate the problem of well-formedness of graphoidal decomposition in DST with Bayesian networks in Fig.1a) and b). Table 1a) gives marginal distribution of X_1 in Fig.1a,b), table 1b) - conditional distributions in Fig.1a), table 1c) - conditionals in Fig.1b).

In Fig.1a) $m_{X_1} \oplus m_{X_2|X_1} \oplus m_{X_3|X_2}$ and in Fig.1b) $m_{X_1} \oplus m_{X_2|X_1} \oplus m_{X_3|X_1} \oplus m_{X_4|X_1}$ are proper belief functions (with non-negative values of m). But in Fig.1a) the function $m = m_{X_1} \oplus m_{X_2|X_1} \oplus m_{X_3|X_2} \oplus m_{X_4|X_3}$ is not a proper belief function, as visible in the table 2. It is easily checked that all marginals over each individual variable of $m = m_{X_1} \oplus m_{X_2|X_1} \oplus m_{X_3|X_2} \oplus m_{X_4|X_3}$ are identical.

Also in the Fig.1b) the function $m = m_{X_1} \oplus m_{X_2|X_1} \oplus m_{X_3|X_1} \oplus m_{X_4|X_1} \oplus m_{X_5|X_1}$ is not a proper belief function as visible in the table 3

Hence, in DST, sample generation from a network and therefore the development of learning algorithms identifying graphoidal structure from data, reasoning from sample data, understanding of causality and of mechanisms giving rise to belief distributions is hampered. E.g. beside [5], the known sample generation algorithms [3, 8, 9, 10, 17] do not use conditional belief functions and therefore (1) conditional independence between variables cannot be pre-specified for the sample and (2) a single generator pass may fail to generate a single sample element.

5 The Solution

In our solution to the problem of well-formedness of belief functions composed from conditional belief functions below we impose the restriction that in the bayesian network no two parents of a node are directly connected.

The fundamental idea behind the approach is to replace the conditional belief function with a specially defined conditional probability function while splitting some values of variables into subvalues. These subvalues take care of differences between belief function values between subsets and supersets of elementary values of variables. The proper generation of samples is run with these special conditional probability functions in a very traditional way, and after completion of sample generation the split values are again joined.

The main difficulties we encounter with handling conditional belief functions is that the conditional independence in DST is radically different from probabilistic independence and that the conditional mass functions m take negative values.

To overcome negativeness, we assume that the conditional belief functions are represented in terms of so-called K functions as introduced in [5]. Given that X is the set of all variables in the conditional belief function and q the set of conditioning variables, we have:

$$K_{|q}(A) = \sum_{B;A^{\downarrow q} \subseteq B^{\downarrow q}, A^{\downarrow X-q} = B^{\downarrow X-q},} m(B)$$

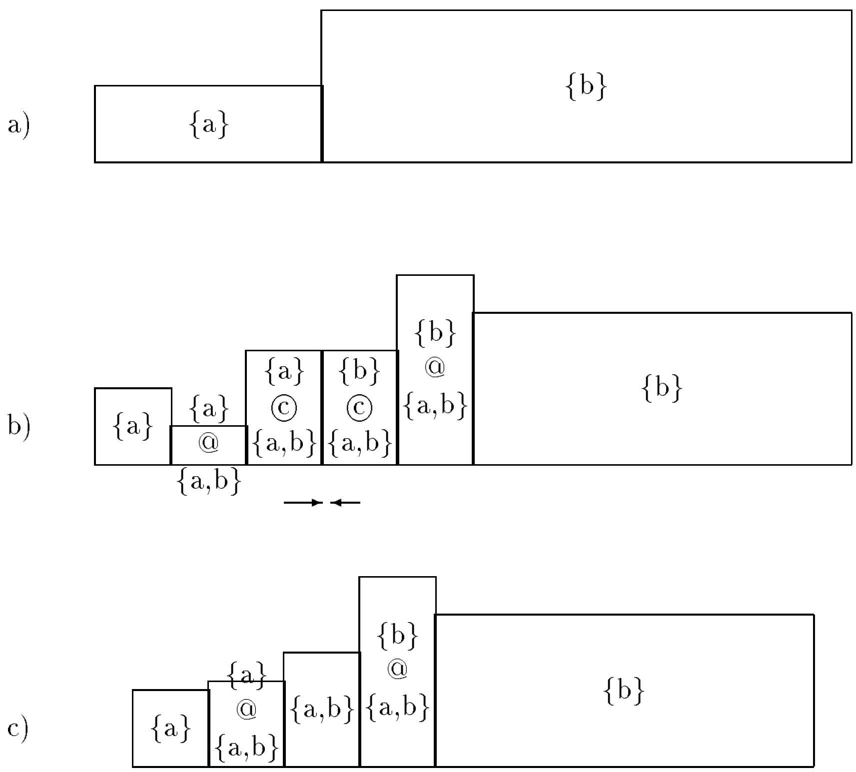


Figure 2: An illustration of the merging process of the values of variable X1 with the domain {a,b}. a) A symbolic representation of the original conditional distribution of $K(X2|X1 = \{a\})$ (to the left) and $K(X2|X1 = \{b\})$ (to the right). b) In the neighboring region of X1={a} and X1={b} the distributions of $K(X2|X1 = .)$ tend to get to a balance (equal to the future $K(X2|X1 = \{a,b\})$), hence are denoted $K(X2|X1 = \{a\}©\{a,b\})$ and $K(X2|X1 = \{b\}©\{a,b\})$. As a "side-effect" the distributions of areas further away change so as to compensate so that proper distribution of the region is reestablished: $K(X2|X1 = \{a\}) = 0.5 \cdot (K(X2|X1 = \{a\}©\{a,b\}) + K(X2|X1 = \{a\}@\{a,b\}))$ and $K(X2|X1 = \{b\}) = 0.5 \cdot (K(X2|X1 = \{b\}©\{a,b\}) + K(X2|X1 = \{b\}@\{a,b\}))$. c) Subpopulations in neighboring regions with identical distributions of X2 X1={a}©{a,b} and X1={b}©{a,b} are merging: The conditional distribution is kept ($K(X2|X1 = \{a\}©\{a,b\}) = K(X2|X1 = \{b\}©\{a,b\}) = K(X2|X1 = \{a,b\})$), but the objects acquire "double identity" with respect to X1: their X1 value is equal to the set {a,b}.

Table 4: A conditional mass function m (a) and its corresponding cumulative mass function K (b).

(a)

X1	X2	m
{a} ×	{a}	0.166667
{a} ×	{b}	-0.0833333
{a} ×	{a,b}	-0.0833333
{b} ×	{a}	-0.0833333
{b} ×	{b}	0.166667
{b} ×	{a,b}	-0.0833333
{a,b} ×	{a}	0.35
{a,b} ×	{b}	0.35
{a,b} ×	{a,b}	0.3

(b)

X1	X2	K
{a} ×	{a}	0.516667
{a} ×	{b}	0.266667
{a} ×	{a,b}	0.216667
{b} ×	{a}	0.266667
{b} ×	{b}	0.516667
{b} ×	{a,b}	0.216667
{a,b} ×	{a}	0.35
{a,b} ×	{b}	0.35
{a,b} ×	{a,b}	0.3

For example, given m in table 4 (a) we get K in table 4 (b): K-function is nonnegative. For any level of conditioning variables the conditioned variables form a probability distribution.

Now we extend the set of values of every variable. If the set S is a set of values of an attribute, then we define the function $MY()$ as $MY(S) = S$ and $SU()$ as $SU(S) = \emptyset$. S is a V-expression. For any V-expression V for any proper non-empty subset $s \subset MY(V)$ we define V-expressions $s©V$ and $s@V$ and define functions $MY(s©V) = MY(s@V) = s$, $SU(s©V) = SU(s@V) = V$. The only element of the set $\{S\}^n$ is a V(n)-expression. $MY(S^n) = S$ and $SU(S^n) = \emptyset$. For any V-expression V for any proper non-empty subset $s \subset MY(V)$, V(n)-expressions are elements of the set: V_n=$\{s©V,\ s@V\}^n$-$\{s©V\}^n$ and for every $v_n \in V_n$ $MY(v_n) = s$, $SU(v_n) = V$. Thus each V(n)-expression is a vector of n V-expressions.

Let X_j be a node in the belief network with n successors and let $\pi(X_j)$ be the set of its predecessors in the network. Let $K_{X_j|\pi(X_j)}$ be the K-function associated with this node. We transform it into a conditional probability function by replacing X_j with X'_j taking its values from the set of V(n)-expressions over the set of values of X_j, and every variable $X_i \in \pi(X_j)$ is replaced with X_i" taking its values from the set of V-expressions over the set of values of X_j. $P(x'_j|x_{i1}", \ldots, x_{ik})"$ is calculated as follows:

1. If $SU(x_{i1}") = \ldots = SU(x_{ik}") = \emptyset$ then for any subset of values s from the domain of X_j $\sum_{x'_j;MY(x'_j)=s} P(x'_j|x_{i1}", \ldots, x_{ik}") = K_{X_j|\pi(X_j)}(x'_j|x_{i1}" \ldots, x_{ik}")$.
2. If $SU(x'_j) \neq \emptyset$ then $P(x'_j|x_{i1}", \ldots, x_{ik}") = P(SU(x'_j)^n|x_{i1}", \ldots, x_{ik}")$.
3. If $x_{il}" = MY(x_{il}")©SU(x_{il}")$ then
$P(x'_j|x_{i1}", \ldots, x_{il}", \ldots, x_{ik}") = P(x'_j|x_{i1}", \ldots, SU(x_{il}"), \ldots, x_{ik}")$
4. if $x_{il}" = MY(x_{il}")@SU(x_{il}")$ let x^*_{il} denote either $x_{il}"$ or $SU(x_{il}")$ and otherwise let x^*_{il} denote only $x_{il}"$. Then
$P(x'_j|x_{i1}", \ldots, MY(x_{il}"), \ldots, x_{ik}") = average_{x^*_{i1},\ldots,x^*_{ik}}(P(x'_j|x^*_{i1}, \ldots, x^*_{ik}))$

To meet the well-formedness criterion, $P(x'_j|x_{i1}", \ldots, x_{ik})"$ i has to be non-negative everywhere.
If X_j is a parent of another node in the network on the $h-th$ outgoing edge, then the respective $x_j"$ acts as the $h-th$ element of the vector x'_j.

With such a transformed probability distribution we generate the sample and then replace all the V- and V(n) expressions V with $MY(V)$.

If X_j is the conditioning variable in the nth branch then the composite (vector) values act as if they were the nth component of the vector.

After random generation the variable values collapse back to what they were before.

Figure 2 provides with an insight into the background idea of the process of transformation of K distributions into $P(|)$ distributions in case of a single successor. The idea is that of splitting a conditional distribution into parts that compensate one another to achieve the original distribution. With more (say n) successors one shall imagine a multi-dimentional ($n+1$ dim.) picture with processes of splitting running independenmtly. Out of the 2^n parts resulting from splitting, only one will merge with the neighboring value of the conditioning variable.

6 Concluding Remarks

The proposed process of generation of samples from conditional belief functions has several significant advantages over previously known algorithms:

- a single sample object is generated in a single generator cycle (Previous generators required 1 to n passes because of contradictions in component belief functions, which are now eliminated by usage of conditional belief functions)
- The conditional independence structure can be pre-specified for the generated sample
- As a result a sufficient well-formedness criterion for conditional belief functions is developed ensuring that the joint belief distribution represented by a set of conditional belief functions really exists.

To verify the above sample generation algorithm, and hence the well-formedness criterion, a program has been implemented allowing to generate the sample from conditional belief functions and to test DST conditional independence properties of the sample. The independence test is based on a previously elaborated layered independence test [4]. The PC algorithm of Spirtes/Glymour/Scheines [14] has been successfully tested for multivariate belief distributions for samples generated by our approach. Fig.1c) represents one of the networks recovered.

References

[1] Dempster A.P.: *Upper and lower probabilities induced by a multi-valued mapping*, Ann. Math. Stat. 38 (1967), 325-339

[2] Halpern J.Y., Fagin R.: Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence* 54(1992), 275-317

[3] Kaempke T.: About assessing and evaluating uncertain information within the theory of evidence. Decision Support Systems 4:433-439, 1988

[4] Kłopotek M.A., Matuszewski A., Wierzchoń S.T.: Overcoming negative-valued conditional belief functions when adapting traditional knowledge acquisition tools to Dempster-Shafer Theory, Proc. CESA'96 IMACS Multiconference (Conference on Expert System Applications), Lille-France, 9-12 July 1996, Vol.2, pp.948-953.

[5] Kłopotek M.A.: *Methods of Identification and Interpretations of Belief Distributions in the Dempster-Shafer Theory* (in Polish). Publisher: Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland, 1998, ISBN 83-900820-8-x.

[6] Kłopotek M.A.: On (Anti)Conditional Independence in Dempster-Shafer Theory. *Journal Mathware and Softcomputing* **5**(1998)1, 69-89.

[7] Kłopotek M.A.: Beliefs in Markov Trees - From Local Computations to Local Valuation. [in:]R. Trappl, Ed.: *Cybernetics and Systems Research*, Proc. 12th European Meeting on Cybernetics and System Research, Vienna 5-8 April 1994, World Scientific Publishers, Vol.1. 351-358.

[8] Kreinovich V. et al.: Monte-Carlo methods make Dempster-Shafer formalism feasible. In: Advances in the Dempster-Shafer Theory of Evidence R.Yager, M. Fedrizzi, J. Kacprzyk: (eds) John Willey, New York, 175-191, 1994.

[9] Moral S, Wilson N. Markov-chain Monte-Carlo algorithm for the calculation of the Dempster-Shafer belief Proc. 12th Nt. Conf. On AI (AAAI-94) 269-274, 1994.

[10] Moral S, Wilson N. Importance sampling Monte-Carlo algorithm for the calculation of Dempster-Shafer belief, Proc. 6th Int.Conf. IPMU'96, Granada 1-5.7.1996, Vol. III, pp. 1337-1344

[11] Pearl J.: *Probabilistic Reasoning in Intelligent Systems:Networks of Plausible Inference*, Morgan Kaufmann, San Mateo CA, 1988.

[12] Shafer G.: *A Mathematical Theory of Evidence* , Princeton University Press, Princeton, 1976

[13] Shenoy P.P.: Conditional Independence in Valuation-based Systems. *International Journal of Approximate Reasoning*, 10, 203-234, 1994.

[14] Spirtes P., Glymour C., Scheines R.: *Causation, Prediction and Search*, Lecture Notes in Statistics 81, Springer-Verlag, 1993.

[15] Studeny M.: Formal properties of conditional independence in different calculi of AI, Sonderforschungsbericht 343, Diskrete Strukturen in der Mathematik, Universität Bielefeld, 1994.

[16] Wierzchoń S.T., Kłopotek M.A.: Modified Component Valuations in Valuation Based Systems As A Way to Optimize Query Processing. *Journal of Intelligent Information Systems* **9**(1997), 157-180.

[17] Wilson N.: A Monte Carlo algorithm for Dempster-Shafer belief Proc. 7th Conf. On Uncertainty in AI, B.D'Ambrossio, P. Smets, AP. Bonisome (Eds.), Morgen-Kaufmann, 414-417, 1991