ELSEVIER

Fuzzy Sets and Systems 103 (1999) 487–505

FUZZY sets and systems

Independence concepts in possibility theory: Part II[1]

Luis M. de Campos*, Juan F. Huete

Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, E-18071, Granada, Spain

Received May 1996; received in revised form April 1997

Abstract

From both a theoretical and a practical point of view, the study of the concept of independence has great importance in any formalism that manages uncertainty. In *Independence Concepts in Possibility Theory: Part I* (de Campos and Huete, Fuzzy Sets and Systems 103 (1999) 127–152) several independence relationships were proposed, using different comparison criteria between conditional possibility measures, and using Hisdal conditioning as the conditioning operator. In this paper, we follow the same approach, but considering possibility measures as particular cases of consonant plausibility measures and, therefore, using Dempster conditioning instead of Hisdal's. We formalize several intuitive ideas to define independence relationships, namely 'not to modify', 'not to gain' and 'to obtain similar' information after conditioning, and study their properties. We also compare the results with the previous ones obtained in Part I using Hisdal conditioning. Finally, the marginal problem, i.e., how to obtain a joint possibility distribution from a set of marginals, and the problem of factorizing large possibility distributions, in terms of its conditionally independent components, are considered.

Keywords: Independence; Possibility theory; Dempster conditioning; Marginal problem

1. Introduction

The concept of independence allows us to represent that our belief about certain events or variables does not change when additional information is known. Therefore, we can use independence in reasoning models in order to perform inferences considering only relevant information, which implies that more efficient algorithms can be developed. Therefore, in addition to the theoretical interest, the study of the concept of independence has great practical significance. When we consider different frameworks for representing uncertain information, we can find several papers about the concept of independence [4, 5, 10, 22, 26, 27]. In this paper, we focus our attention on the concept of conditional independence within the possibilistic framework [29, 12]. Some recent works studying this topic are also to be found in the literature [1–3, 7, 13, 16, 9, 17, 19].

In order to establish a conditional independence relationship between variables, our approach is to compare the previous ('a priori') information with the information that we obtain after a new piece of information has been known ('a posteriori'). We present different comparison criteria ('not to modify', 'not to gain' and 'to obtain similar' information after conditioning), which give rise to different approaches to the concept of independence. Moreover, in order to

* Corresponding author. E-mail: lci@decsai.ugr.es.

[1] This work has been supported by the DGICYT under Project n. PB92-0939.

PII: S0165-0114(97)00161-9

evaluate the different definitions of independence, a set of properties or axioms has been selected that would seem reasonable to demand from any relationship that capture the notion of independence (the well-known *graphoid* axioms [24]), and our definitions are tested against this set. These axioms would also permit us to compare the given definitions with the definitions of independence obtained for other formalisms.

In our study, we consider possibility measures as a special case of Dempster/Shafer evidence measures [11, 25], namely consonant plausibility measures. An in-depth study of the same concept, but considering possibility measures in a closer way to fuzzy sets, can be found in *Independence Concepts in Possibility Theory: Part I* [6]. In this paper, we present a parallel study to the one developed in [6], but using Dempster conditioning instead of Hisdal's. However, in order to make the paper self-contained, some of the concepts and methods used in [6] will be briefly reconsidered.

The paper is organized according to the following scheme: first, in Section 2 we briefly study possibility measures as particular cases of Dempster/Shafer measures of evidence. Section 3 starts by introducing several intuitive definitions of conditional independence, as well as the abstract properties that these definitions should verify. Next, we formalize the previous definitions in the framework of possibility theory, using Dempster conditioning, and study their properties. In Section 4, we study the marginal problem, that is to say, how to construct a joint possibility distribution from a set of marginals, and as an application, the problem of storage of large possibility distributions is considered in Section 5. Finally, Section 6 contains the concluding remarks and some proposals for future research.

2. Evidence measures and possibility measures

Perhaps one of the most general formalisms for dealing with numerical uncertainty is that of fuzzy measures [28]. A fuzzy measure is a mapping, g, from the power set of a given finite reference set D_X (from which a variable X takes its values) to the interval $[0,1]$; for any $A \subseteq D_X$, $g(A)$ represents our degree of belief in the occurrence of the event A (i.e., the value of the variable X belongs to A), and g must satisfy the following properties: $g(\emptyset)=0$, $g(D_X)=1$ (limit conditions) and for all $A,B \subseteq D_X$, if $A \subseteq B$, then $g(A) \leqslant g(B)$ (monotonicity condition). However, fuzzy measures are usually too general for practical purposes, and we often have to restrict ourselves to considering appropriate subclasses, having a richer set of properties that make a more efficient computation possible (for an interesting classification of fuzzy measures, see [21]). Two of the most interesting subclasses of fuzzy measures are probability and possibility measures [29, 15], which in turn belong to another wider and well-known class of fuzzy measures: evidence measures [11,25]. In order to study the concept of independence for possibility measures, and compare it with that of probabilities, we will use evidence measures as the common reference class. So, we briefly recall some basic concepts relating to these measures.

Evidence measures (belief and plausibility) are particular cases of fuzzy measures, based on the concept of basic probability assignment, m:

Definition 1. A basic probability assignment (b.p.a.), m, is a mapping from $\mathcal{P}(D_X)$ (the power set of D_X) to the unit interval

$$m : \mathcal{P}(D_X) \to [0,1]$$

that satisfies the following conditions:
1. $m(\emptyset) = 0$,
2. $\sum_{A \subseteq D_X} m(A) = 1$.

A b.p.a. m can be interpreted as follows: 'There exists an unknown element u belonging to the set D_X, and $m(A)$ represents that portion of the total belief exactly committed to hypothesis A (u belongs to A) given a piece of evidence'; in other words, $m(A)$ represents the direct support of evidence for A, without considering the evidence for any proper subset of A.

Using the definition above, the concepts of belief and plausibility measures can be introduced:

Definition 2. Given a b.p.a. m, the belief measure associated with m is defined by means of

$$\mathrm{Bel} : \mathcal{P}(D_X) \to [0,1]$$

where, for each $A \subseteq D_X$

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

Bel(A) gives the total belief about the fact 'the unknown element u belongs to A', and obviously satisfies the limit and monotonicity conditions, so Bel is a fuzzy measure. Every subset A in D_X such that $m(A)>0$ is called a focal element of m. Given a belief measure, the plausibility measure may be defined as its dual measure:

Definition 3. Let Bel be a belief measure; its associated plausibility measure, Pl, is given by

$\mathrm{Pl}:\mathscr{P}(D_X)\to[0,1]$

such that for each $A\subseteq D_X$

$\mathrm{Pl}(A)=1-\mathrm{Bel}(\bar{A})$,

where $\bar{A}$ represents the complement of A in D_X.

Note that Pl is also a fuzzy measure. As Bel($\bar{A}$) measures the doubt on A, Pl(A) represents the extent to which the evidence does not rule out A. Plausibility measures can be directly obtained from the values associated to the b.p.a. m by means of the following expression:

$$\mathrm{Pl}(A)=\sum_{B\cap A\neq\emptyset} m(B).$$

Let us review how probability and possibility measures can be considered as particular cases of evidence measures:

2.1. Bayesian evidence measures: probability measures

A Bayesian evidence measure is any evidence measure satisfying

$\mathrm{Pl}(A)=\mathrm{Bel}(A)$ for all $A\subseteq D_X$.

This is equivalent to
1. $\mathrm{Bel}(\emptyset)=0$;
2. $\mathrm{Bel}(D_X)=1$;
3. $\mathrm{Bel}(A\cup B)=\mathrm{Bel}(A)+\mathrm{Bel}(B)$ whenever $A\cap B=\emptyset$.

The bayesian structure implies that the evidence assignment has no degree of freedom, i.e., the singletons of D_X are the only possible focal elements ($m(A)=0\ \forall A\subseteq D_X$ such that $|A|>1$). In this case, the b.p.a., m, is equivalent to a probability distribution p, or, in other words, every probability distribution can be associated with a Bayesian evidence measure where $m(\{x\})=p(x)$.

2.2. Consonant evidence measures: possibility measures

Any evidence measure is said to be consonant if it satisfies
1. $\mathrm{Bel}(\emptyset)=0$;
2. $\mathrm{Bel}(D_X)=1$;
3. $\mathrm{Bel}(A\cap B)=\min\{\mathrm{Bel}(A),\mathrm{Bel}(B)\}$, for all A, $B\subseteq D_X$.

The following well-known theorem represents a characterization of consonant evidence measures:

Theorem 1. *An evidence measure is consonant if and only if the focal elements of the b.p.a. m are nested, i.e., there exists a family of subsets of* D_X, A_i, $i=1,2,\dots,r$, *such that* $A_i\subset A_j$ *whenever* $i<j$ *and* $\sum_{i=1}^{r} m(A_i)=1$.

Consonant evidence measures are the prototypes for possibility measures, where the plausibility measure (Pl) in Dempster–Shafer theory plays the role of the possibility measure, Π, and the belief measure (Bel) plays the role of the necessity measure, N. Another point of view is considering possibility measures as an extreme case for the monotonicity condition (note that, as $A,B\subseteq A\cup B$, then we have $\mathrm{Pl}(A\cup B)\geqslant\max\{\mathrm{Pl}(A),\mathrm{Pl}(B)\}$):

$$\forall A,B\subseteq D_X,\quad \Pi(A\cup B)=\max\{\Pi(A),\Pi(B)\},\qquad (1)$$

where $\Pi(A)$ represents the possibility of the event A. So, $\Pi(A)=1$ means that the event A is completely possible and $\Pi(A)=0$ means that the event A is impossible (particularly $\Pi(X)=1$ and $\Pi(\emptyset)=0$). Any possibility measure verifies that, given two events A and $\bar{A}$,

$$\max\{\Pi(A),\Pi(\bar{A})\}=1$$

which expresses that if we consider two contradictory and exhaustive events, at least one of them must be completely possible.

If D_X is a finite set (as we use it), then the possibility values for the singletons in D_X completely define the possibility measure:

$$\forall A \subseteq D_X, \quad \Pi(A) = \max\{\pi(x) \mid x \in A\},$$

where $\pi(x) = \Pi(\{x\})$ and π is a mapping from D_X to $[0,1]$ called *possibility distribution*. This mapping is normalized, i.e., there exists $x_0 \in D_X$, such that $\pi(x_0) = 1$. $\pi(x)$ represents the degree in which $x \in D_X$ is the possible value that the variable X takes. Therefore, $\Pi(A)$ is the possibility that the value of the variable X is in A.

2.3. Marginalization and conditioning operators

In order to illustrate the tools we shall need to use, let us first consider the case of probabilities, where the concept of conditional independence has been extensively studied. Consider two variables, X and Y, taking their values in the sets $D_X = \{x_1, x_2, \ldots, x_n\}$ and $D_Y = \{y_1, y_2, \ldots, y_m\}$, respectively, and let p be a bidimensional probability distribution defined on $D_X \times D_Y$; then X and Y are said to be independent if

$$p(x \mid y) = p(x),$$
$$\forall x \in D_X,\ y \in D_Y \text{ such that } p(y) > 0.$$

This definition asserts the independence of X and Y when all the conditional probabilities on X given any value for Y are equal to the marginal probability on X. Therefore, these concepts, marginalization and conditioning, should also be considered within the possibilistic framework. As we can consider both possibility and probability measures as particular cases of evidence measures, we shall study the concept of marginal evidence measure and conditional evidence measure and then, these concepts will be particularized for the narrower class of possibility measures. We focus our attention on plausibility measures.

Definition 4. Let Pl be a bidimensional plausibility measure defined on $D_X \times D_Y$. The marginal plausibility measure Pl_X on X (analogously for Pl_Y on Y) is defined as

$$\mathrm{Pl}_X(A) = \mathrm{Pl}(A \times D_Y) = \sum_{C \cap A \times D_Y \neq \emptyset} m(C) = \sum_{C_X \cap A \neq \emptyset} m(C), \quad \forall A \subseteq D_X, \tag{2}$$

where C_X is the projection of C on X, i.e., $C_X = \{x \in D_X \mid (x, y) \in C \text{ for some } y \in D_Y\}$.

Considering consonant evidence measures, i.e., possibility measures, the marginal possibility measure is defined in the same way:

Definition 5. Given a bidimensional possibility measure Π defined on $D_X \times D_Y$, the marginal possibility measure on X, Π_X, (analogously on Y) is defined as:

$$\Pi_X(A) = \Pi(A \times D_Y), \quad \forall A \subseteq D_X.$$

As the projections of nested focal sets are also nested, it is obvious that the marginal possibility measure is indeed a possibility measure. Thus, the marginal possibility distribution on X (analogously on Y) can be defined by means of:

$$\pi_X(x) = \Pi_X(\{x\}) = \Pi(x \times D_Y) = \max_{y \in D_Y} \pi(x, y) \quad \forall x \in D_X. \tag{3}$$

Now, we shall consider conditional evidence measures. In this case, there are several ways of defining the conditioning (see [23] for a review). Here, we shall use the concept of conditional evidence measure given by Dempster [11] and Shafer [25].

Definition 6. Let Pl be a bidimensional plausibility measure defined on $D_X \times D_Y$. The conditional plausibility measure given $[Y = B]$, $\mathrm{Pl}_{X \mid Y=B}$ on X (analogously for $\mathrm{Pl}_{Y \mid X=A}$ on Y) is defined as

$$\mathrm{Pl}_{X \mid Y=B}(A \mid B) = \frac{\mathrm{Pl}(A \times B)}{\mathrm{Pl}_Y(B)} = \frac{\mathrm{Pl}(A \times B)}{\mathrm{Pl}(D_X \times B)}. \tag{4}$$

If we have a consonant evidence measure, then the conditional possibility measure can be defined in the same way:

Definition 7. Let Π be a bidimensional possibility measure defined on $D_X \times D_Y$. The conditional possibility measure given $[Y = B]$, $\Pi_{X \mid Y=B}$ on X (analogously for $\Pi_{Y \mid X=A}$ on Y) is defined by means of

$$\Pi_{X \mid Y=B}(A \mid B) = \frac{\Pi(A \times B)}{\Pi_Y(B)}.$$

We must note that conditional possibility measures are also possibility measures. So, we may limit our attention to conditional possibility distributions. More precisely, the possibility distribution on X, conditioned to the event $[Y = y]$, denoted by $\pi_d(. \mid y)$ is defined as

$$\pi_d(x \mid y) = \frac{\pi(x, y)}{\pi_Y(y)} = \frac{\pi(x, y)}{\max_{x' \in D_X} \pi(x', y)}. \quad (5)$$

From now on, to simplify the notation of a marginal possibility distribution, we shall drop the subindex, thus writing $\pi(x)$ and $\pi(y)$ instead of $\pi_X(x)$ and $\pi_Y(y)$, respectively.

3. Possibilistic conditional independence

In this section we follow the same methodology developed in [6], where an intuitive approach to the concept of independence was proposed. If we denote by $I(X \mid Z \mid Y)$ the assertion 'X is independent of Y given Z', in order to establish the concept of possibilistic conditional independence, a natural approach consists in comparing the previous knowledge about X with the knowledge that we obtain after knowing a new piece of information about Y. Therefore, a comparison between the conditional possibility distributions, $\pi_d(x \mid z)$ and $\pi_d(x \mid yz)$, must be carried out. The same general idea underlies in the probabilistic framework [10], in different formalisms to represent the uncertainty [27] and also in the framework of possibility measures [17]. With this comparison we try to detect a change in our current belief when a new piece of information is known. Bearing in mind that we have uncertain knowledge, different comparison criteria can be considered. The most obvious way (and also the strictest one) to define conditional independence is the following:

Definition 8 (*Not modifying the information*). Given any value of the variable Z, knowing the value that the variable Y takes does not modify our information about the values that variable X can take.

A softer definition of conditional independence can be obtained if we relax the notion of not modifying the information. In this case, an increase in our uncertainty after conditioning is allowed.

Definition 9 (*Not gaining information*). Given any value of the variable Z, if we know the value that the variable Y takes, we do not gain additional information about the values of X.

Finally, a generic similarity relationship between conditional possibility distributions can also be used to establish the independence:

Definition 10 (*Obtaining similar information*). Given any value of the variable Z, if we know the value that the variable Y takes, we obtain a piece of information about X similar to the one prior to learning the value of Y.

Given these intuitive notions of independence, the next step is to formalize them within the possibilistic framework and then study the set of properties that each definition verifies. Pearl [24] identified the following set of axioms or properties that seems reasonable to demand from any relationship that attempts to capture the intuitive idea of independence (a semantic interpretation of the axioms is to be found in [6, 24]).

A1. *Trivial Independence*:
$I(X \mid Z \mid \emptyset)$.
A2. *Symmetry*:
$I(X \mid Z \mid Y) \Rightarrow I(Y \mid Z \mid X)$.
A3. *Decomposition*:
$I(X \mid Z \mid Y \cup W) \Rightarrow I(X \mid Z \mid Y)$.
A4. *Weak Union*:
$I(X \mid Z \mid Y \cup W) \Rightarrow I(X \mid Z \cup Y \mid W)$.
A5. *Contraction*:
$I(X \mid Z \mid Y)$ and $I(X \mid Z \cup Y \mid W)$
$\Rightarrow I(X \mid Z \mid Y \cup W)$.
A6. *Intersection*:
$I(X \mid Z \cup W \mid Y)$ and $I(X \mid Z \cup Y \mid W)$
$\Rightarrow I(X \mid Z \mid Y \cup W)$.

In order to formalize the previous definitions in the framework of possibility measures, we shall consider that X, Y and Z are disjoint variables or subsets of variables in a finite set of n variables, and π is an n-dimensional possibility distribution on these variables. Any generic value that these variables can take on will be denoted by x, y, z, and particular instances for these variables will be denoted by subscripted or Greek letters.

The above possibilistic independence criteria can be formalized using different comparison operators $\odot$ between conditional possibilities, i.e., $\pi_d(x \mid z) \odot \pi_d(x \mid yz)$. The comparison operators will represent the notions of not modifying, not gaining, and obtaining similar information after conditioning. In the following subsections, different comparison operators will be considered, and thereafter, for each one of them, we shall study the set of axioms that the corresponding definition of independence verifies.

3.1. The equality operator

The first idea was to define independence when we *do not modify* the original information at all after conditioning. An obvious way to capture this idea is to use an equality relationship as the comparison operator. Formally:

Definition 11. (D1) *Not modifying the information.*

$$I(X \mid Z \mid Y) \Leftrightarrow \pi_d(x \mid yz) = \pi_d(x \mid z), \quad \forall x, y, z \text{ such that } \pi(yz) > 0. \tag{6}$$

Observe that we must impose that the two conditional measures involved are defined, and this is the case if $\pi(yz) > 0$. This definition has also been proposed by Studený [27] and, in a slighty different form, by Fonck [17].

The next proposition shows a simple characterization of the previous definition:

Proposition 1. *The definition of independence* D1 *is equivalent to*

$$I(X \mid Z \mid Y) \Leftrightarrow \pi_d(xy \mid z) = \pi_d(x \mid z)\pi_d(y \mid z), \quad \forall x, y, z \text{ such that } \pi(z) > 0.$$

Proof. The proof is very simple, so we omit it. □

Let us study what properties definition D1 satisfies:

Proposition 2. *The independence relationship* D1 *satisfies the axioms* A1–A5, *and if the possibility distribution is strictly positive, it also satisfies* A6.

Proof. The axioms A1, A2 and A5 have a direct proof. Axiom A4 may be deduced directly from A3. So, we only prove the axioms A3 and A6.

Decomposition: $I(X \mid Z \mid Y \cup W) \Rightarrow I(X \mid Z \mid Y)$.

We find that $\forall xyzw$ such that $\pi(yzw) > 0$, $\pi_d(x \mid yzw) = \pi_d(x \mid z)$. Then

$$\frac{\pi(xyzw)}{\pi(yzw)} = \frac{\pi(xz)}{\pi(z)}, \text{ i.e., } \pi(xyzw) = \pi(yzw)\frac{\pi(xz)}{\pi(z)}, \quad \forall xyzw \text{ such that } \pi(yzw) > 0.$$

But the last equality is also true if $\pi(yzw) = 0$, i.e., it is always true. Then, by taking maximum in w on both sides we obtain $\pi(xyz) = \pi(yz)(\pi(xz)/\pi(z))$ $\forall xyz$, and therefore $\pi_d(x \mid yz) = \pi_d(x \mid z)$ $\forall xyz$ such that $\pi(yz) > 0$.

Intersection: $I(X \mid Z \cup Y \mid W) \,\&\, I(X \mid Z \cup W \mid Y) \Rightarrow I(X \mid Z \mid Y \cup W)$.

Considering a strictly positive distribution, and given the independence relationships on the left-hand side of the implication, we find that

$$\pi_d(x \mid yzw) = \pi_d(x \mid yz) = \pi_d(x \mid zw), \quad \forall xzyw.$$

In particular, we have $\pi_d(x \mid yz) = \pi_d(x \mid zw)$, $\forall xzyw$, and therefore $\pi(xzw) = \pi(zw)\pi_d(x \mid yz)$ $\forall xyzw$. So, by taking the maximum in w we obtain $\pi(xz) = \pi(z)\pi_d(x \mid yz)$ $\forall xyz$, and then $\pi_d(x \mid z) = \pi_d(x \mid yz) = \pi_d(x \mid yzw)$, i.e., $I(X \mid Z \mid Y \cup W)$. □

As we have already commented, in the first part of this paper [6], we developed a similar study of different definitions of independence using Hisdal conditioning instead of Dempster's. When we used Hisdal conditioning together with the equality operator, we obtained the properties A1 and A3–A6. So, there are differences between the two forms of conditioning: A1, A3–A6 hold in both cases, but A6 is only true for strictly positive distributions in the case of Dempster conditioning. Moreover, A2 holds for Dempster conditioning but it does not hold for Hisdal's.

3.2. The inclusion operator

The concept of independence as a non gain of information after conditioning can be adequately formalized (see [6]) by using an inclusion relationship [14] between possibility distributions. This relationship establishes when a possibility distribution is more or less informative than another one.

Definition 12. Given two possibility distributions, π and π', defined on the same reference set D_X, then π' is included in (or is less informative than) π if and only if $\pi(x) \leqslant \pi'(x)$, $\forall x \in D_X$.

Using this relationship, the independence in the sense of not gaining information can be defined as follows:

Definition 13. (D2) *Not gaining information.*

$$I(X \mid Z \mid Y) \Leftrightarrow \pi_{\mathrm{d}}(x \mid yz) \geqslant \pi_{\mathrm{d}}(x \mid z),$$
$$\forall x, y, z \text{ such that } \pi(yz) > 0. \quad (7)$$

Proposition 3. *The independence relationship* D2 *satisfies the axioms* A1–A3 *and* A5.

Proof. The axioms A1, A2 and A5 can be immediately proven. We only show the proof for the axiom A3.

Decomposition: $I(X \mid Z \mid Y \cup W) \Rightarrow I(X \mid Z \mid Y)$.

From $I(X \mid Z \mid Y \cup W)$ we obtain

$$\frac{\pi(xz)}{\pi(z)} \leqslant \frac{\pi(xyzw)}{\pi(yzw)} \quad \forall xyzw \text{ such that } \pi(yzw) > 0$$

and therefore

$$\pi(yzw)\frac{\pi(xz)}{\pi(z)} \leqslant \pi(xyzw)$$
$$\forall xyzw \text{ such that } \pi(yzw) > 0.$$

This inequality also holds if $\pi(yzw) = 0$, i.e., it is always true. Then, by taking the maximum in w we obtain $\pi(yz)[\pi(xz)/\pi(z)] \leqslant \pi(xyz)$ $\forall xyz$, and thus $\pi_{\mathrm{d}}(x \mid z) \leqslant \pi_{\mathrm{d}}(x \mid yz)$ $\forall xyz$ such that $\pi(yz) > 0$. □

Unfortunately, the important property of Weak Union (A4) does not hold in general; in the next counterexample, we consider four bivaluated variables X, Y, Z, W having the following joint possibility distribution:

x_1yzw	$\pi(x_1yzw)$		x_2yzw	$\pi(x_2yzw)$
$x_1y_1z_1w_1$	0.3		$x_2y_1z_1w_1$	0.4
$x_1y_1z_1w_2$	0.4		$x_2y_1z_1w_2$	0.4
$x_1y_1z_2w_1$	1		$x_2y_1z_2w_1$	1
$x_1y_1z_2w_2$	1		$x_2y_1z_2w_2$	1
$x_1y_2z_1w_1$	0.5		$x_2y_2z_1w_1$	0.7
$x_1y_2z_1w_2$	0.5		$x_2y_2z_1w_2$	0.7
$x_1y_2z_2w_1$	1		$x_2y_2z_2w_1$	1
$x_1y_2z_2w_2$	1		$x_2y_2z_2w_2$	1

In that case, we find that $\pi(x \mid yzw) \geqslant \pi(x \mid z)$, $\forall xyzw$, i.e., $I(X \mid Z \mid Y \cup W)$ holds; however, for example, $\pi(x_1 \mid y_1z_1w_1) = 0.75 < 0.4/0.4 = 1.0 = \pi(x_1 \mid y_1z_1)$, so that the inequalities $\pi(x \mid yzw) \geqslant \pi(x \mid yz)$ are not always true. Therefore, $I(X \mid Y \cup Z \mid W)$ does not hold.

When we used the inclusion operator together with Hisdal conditioning in [6], we obtained the properties A1–A5. So, in this case, Hisdal conditioning performs better than Dempster's, because the latter fails to satisfy A4.

3.3. Default conditioning

We think that the problem with the concept of independence studied in the previous subsection is due to the fact that the idea of independence as non-gain of information has not been carried out till the finish: if after conditioning we lose information, it would be more convenient to keep the initial information. That is, if in a very specific context we do not have much information, then we can use the information available in a less specific context. This idea implies a change in the definition of conditioning. Therefore, we use a new conditioning operator, called *Dempster default conditioning*, denoted by $\pi_{\mathrm{d_c}}(.\,|\,.)$ (which is analogous to the one defined in [6] on the basis of Hisdal conditioning):

$$\pi_{\mathrm{d_c}}(x \mid y) = \begin{cases} \pi(x) & \text{if } \pi(xy) \geqslant \pi(x)\pi(y)\ \forall x, \\ \pi_{\mathrm{d}}(x \mid y) & \text{if } \exists x' \text{ such that} \\ & \quad \pi(x'y) < \pi(x')\pi(y). \end{cases} \quad (8)$$

The idea is the following: if after conditioning we obtain a less informative distribution, then we preserve the previous more precise information; otherwise, we use the usual (Dempster) conditional possibility distribution.

Using this conditioning, a new definition of independence can be stated as follows:

Definition 14. (D3) *Default conditioning.*

$$I(X \mid Z \mid Y) \Leftrightarrow \pi_{d_c}(x \mid yz) = \pi_{d_c}(x \mid z), \quad \forall xyzw. \tag{9}$$

Note that, as the default conditioning is always defined for every value of the variables involved, it is not necessary to impose any restriction for the equality of the two default conditional possibility distributions $\pi_{d_c}(x \mid yz)$ and $\pi_{d_c}(x \mid z)$.

Proposition 4. *The independence relationship* D3 *verifies the axioms* A1 *and* A3–A6 (A6 *even for non-strictly positive distributions*).

Proof. Once again, the axioms A1 and A5 are immediately deduced, and the axiom A4 can be obtained directly once the axiom A3 is proven.

Decomposition: $I(X \mid Z \mid Y \cup W) \Rightarrow I(X \mid Z \mid Y)$.

$I(X \mid Z \mid Y \cup W)$ means that $\pi_{d_c}(x \mid yzw) = \pi_{d_c}(x \mid z)$ $\forall xyzw$. Our aim is to prove that $\pi_{d_c}(x \mid yz) = \pi_{d_c}(x \mid z)$ $\forall xyz$. For any given z, we shall study the two different cases that may appear:

(1) Suppose that $\pi_{d_c}(x \mid z) = \pi(x)$, i.e., $\pi(xz) \geqslant \pi(x)\pi(z)$ $\forall x$.

Then, we find that $\pi_{d_c}(x \mid yzw) = \pi_{d_c}(x \mid z) = \pi(x)$, $\forall xyw$. Therefore, using the definition of default conditioning, we have $\pi(xyzw) \geqslant \pi(x)\pi(yzw)$, $\forall xyw$, and taking the maximum in w, we obtain $\pi(xyz) \geqslant \pi(x)\pi(yz)$ $\forall xy$, that is $\pi_{d_c}(x \mid yz) = \pi(x) = \pi_{d_c}(x \mid z)$ $\forall xy$.

(2) Suppose that $\pi_{d_c}(x \mid z) = \pi(xz)/\pi(z) \neq \pi(x)$, i.e., exists $\delta \in D_X$ such that $\pi(\delta z) < \pi(\delta)\pi(z)$. In that case we find that

$$\pi_{d_c}(x \mid z) = \frac{\pi(xz)}{\pi(z)} = \pi_{d_c}(x \mid yzw), \quad \forall xyw$$

and therefore

$$\frac{\pi(xz)}{\pi(z)} = \frac{\pi(xyzw)}{\pi(yzw)}, \quad \forall xyw.$$

So, we obtain $\max_w\{\pi(xyzw)\pi(z)\} = \max_w\{\pi(xz)\pi(yzw)\}$ and then $\pi(xyz)\pi(z) = \pi(xz)\pi(yz)$ $\forall xy$. Now, we must prove that $\pi_{d_c}(x \mid yz) \neq \pi(x)$. Considering that $\pi_{d_c}(. \mid z) \neq \pi(.)$, that is to say, there is $\delta \in D_X$ such that $\pi(\delta z) < \pi(\delta)\pi(z)$, then we have, for all x, y

$$\frac{\pi(xyz)}{\pi(yz)} = \frac{\pi(xz)}{\pi(z)},$$

particularly, for δ we have $\dfrac{\pi(\delta yz)}{\pi(yz)} = \dfrac{\pi(\delta z)}{\pi(z)} < \pi(\delta)$,

so that

$$\forall xy, \ \pi_{d_c}(x \mid yz) = \frac{\pi(xyz)}{\pi(yz)} = \frac{\pi(xz)}{\pi(z)} = \pi_{d_c}(x \mid z).$$

Therefore, we deduce $I(X \mid Z \mid Y)$.

Intersection: $I(X \mid Z \cup W \mid Y)$ and $I(X \mid Z \cup Y \mid W)$ $\Rightarrow I(X \mid Z \mid Y \cup W)$.

Once again, given any z, let us consider two different cases:

(1) Suppose there exist y_0 and w_0 such that $\forall x$, $\pi(xy_0zw_0) \geqslant \pi(x)\pi(y_0zw_0)$, i.e., $\forall x$, $\pi_{d_c}(x \mid y_0zw_0) = \pi(x)$.

Then, from $I(X \mid Z \cup W \mid Y)$ and $I(X \mid Z \cup Y \mid W)$, we have $\pi_{d_c}(x \mid zw) = \pi_{d_c}(x \mid yz) = \pi_{d_c}(x \mid yzw) \forall xyw$, and therefore we can easily deduce that $\pi_{d_c}(x \mid yzw) = \pi(x)$, $\forall xyw$. So, we obtain $\pi(xyzw) \geqslant \pi(x)\pi(yzw)$ $\forall xyw$, and then we have, for all x

$$\max_{yw} \pi(xyzw) \geqslant \max_{yw} \{\pi(x)\pi(yzw)\},$$

which implies $\pi(xz) \geqslant \pi(x)\pi(z)$ $\forall x$, and then $\pi_{d_c}(x \mid z) = \pi(x) = \pi_{d_c}(x \mid yzw)$.

(2) Supppose that for all y and w, $\pi_{d_c}(x \mid yzw) = [\pi(xyzw)/\pi(yzw)]$ $\forall x$. Then from the hypothesis we obtain that

$$\pi_{d_c}(x \mid zw) = \frac{\pi(xzw)}{\pi(zw)} = \frac{\pi(xyz)}{\pi(yz)} = \pi_{d_c}(x \mid yz) = \pi_{d_c}(x \mid yzw), \quad \forall xyw.$$

Therefore, we find that $\max_w\{\pi(xzw)\pi(yz)\} = \max_w\{\pi(xyz)\pi(zw)\}$ and then $\pi(xz)\pi(yz) = \pi(xyz)\pi(z)$, i.e.,

$$\frac{\pi(xyz)}{\pi(yz)} = \frac{\pi(xz)}{\pi(z)}.$$

Finally,

$$\pi_{d_c}(x \mid yzw) = \frac{\pi(xyzw)}{\pi(yzw)} = \frac{\pi(xyz)}{\pi(yz)} = \frac{\pi(xz)}{\pi(z)} = \pi_{d_c}(x \mid z),$$

and therefore, $I(X \mid Z \mid Y \cup W)$ holds. □

In order to see that D3 does not verify Symmetry in general, consider the following counterexample, where X, Y, Z are bivaluated variables and π is the following possibility distribution:

xyz	$\pi(xyz)$
$x_1y_1z_1$	1.0
$x_1y_1z_2$	0.3
$x_1y_2z_1$	0.6
$x_1y_2z_2$	0.1
$x_2y_1z_1$	0.6
$x_2y_1z_2$	0.2
$x_2y_2z_1$	0.4
$x_2y_2z_2$	0.1

In this case, we find that $\pi_{d_c}(x \mid yz) = \pi_{d_c}(x \mid z)$ ($= \pi(x)$), i.e., $I(X \mid Y \mid Z)$. However, $\pi_{d_c}(y_2 \mid x_2z_2) = 0.5 \neq 0.33 = \pi_{d_c}(y_2 \mid z_2)$, and therefore $\neg I(Y \mid Z \mid X)$.

So, by using the default conditioning we have recovered the important property of Weak Union, but we have lost Symmetry. This property can easily be recovered by using a symmetrized definition of independence, such as $I^s(X \mid Z \mid Y) \Leftrightarrow I(X \mid Z \mid Y)$ and $I(Y \mid Z \mid X)$, but in this case it is not clear so far that all the other axioms will still be satisfied.

So, in the case of using the default conditioning to define independence, the two underlying conditioning operators, Dempster's and Hisdal's [6], perform equally well with respect to the graphoid axioms: both verify A1, A3–A6 and fail A2.

With respect to the relationships existing among the three definitions of independence considered so far, it is obvious that D1 → D2, but (contrary to what happened when using Hisdal conditioning [6]), D3 is not implied by D1 and D2 is not implied by D3. The following examples show these facts:

In this first example, we have three bivaluated variables X, Y and Z, whose joint possibility distribution is given in the table below.

xyz	$\pi(xyz)$
$x_1y_1z_1$	1.0
$x_1y_1z_2$	0.4
$x_1y_2z_1$	0.6
$x_1y_2z_2$	0.0
$x_2y_1z_1$	0.5
$x_2y_1z_2$	0.1
$x_2y_2z_1$	0.3
$x_2y_2z_2$	0.0

It may be seen that $\pi_d(x \mid yz) = \pi_d(x \mid z)$ $\forall xyz$ such that $\pi(yz) > 0$ (i.e., for all yz except y_2z_2). So, $I(X \mid Z \mid Y)$ holds when using definition D1. However, $\pi_{d_c}(x_2 \mid z_2) = 0.25$, whereas $\pi_{d_c}(x_2 \mid y_2z_2) = 0.5$, so that $\neg I(X \mid Z \mid Y)$ if we use definition D3.

In the second example, once again we have three bivaluated variables X, Y and Z, and the following joint possibility distribution:

xyz	$\pi(xyz)$
$x_1y_1z_1$	1.0
$x_1y_1z_2$	0.3
$x_1y_2z_1$	0.5
$x_1y_2z_2$	0.7
$x_2y_1z_1$	0.8
$x_2y_1z_2$	0.25
$x_2y_2z_1$	0.4
$x_2y_2z_2$	0.6

It is easy to check that $\pi(xyz) \geqslant \pi(x)\pi(yz)$ $\forall xyz$, and $\pi(xz) \geqslant \pi(x)\pi(z)$ $\forall xz$, so that $\pi_{d_c}(x \mid yz) = \pi(x) = \pi_{d_c}(x \mid z)$ $\forall xyz$, i.e., $I(X \mid Z \mid Y)$ is true if we use definition D3. However, $\pi_d(x_2 \mid y_1z_2) = 0.833 < 0.857 = \pi_d(x_2 \mid z_2)$, hence $I(X \mid Z \mid Y)$ is not true when we use definition D2.

3.4. Similarity operators

In order to define independence relationships, the third intuitive idea was to use a similarity criterion between conditional distributions. Thus, given a similarity relationship $\simeq$, defined on the set of possibility distributions for the variable X, the independence may be defined in the following way:

Definition 15. (D4) *Obtaining similar information.*

$$I(X \mid Z \mid Y) \Leftrightarrow \pi_d(. \mid yz) \simeq \pi_d(. \mid z),$$
$$\forall y,z \text{ such that } \pi(yz) > 0. \qquad (10)$$

First, we shall study what kind of properties of $\simeq$ are sufficient to guarantee that some of the axioms hold. Using the same criterion as in [6], we consider that an equivalent relationship is a good candidate to define independence in the sense of similarity. In that case, we find that A1 (Trivial Independence) and A5 (Contraction) are immediately obtained, and we can guarantee that A3 (Decomposition) is verified if and only if A4 (Weak Union) is verified. Therefore, we must look for the additional conditions which assure the fulfilment of A3. Particularly, we shall impose that the similarity relationship verifies the following property.

Definition 16 (*Maximum Property*). We say that a similarity relationship $\simeq$ satisfies the maximum property if, for any family $\{\pi_s\}$ of possibility distributions verifying

$$\pi_s(x) = \frac{f_s(x)}{\lambda_s}, \quad \forall x$$

where λ_s are positive real numbers less than or equal to 1 (therefore $\max_x f_s(x) = \lambda_s$), and with $\pi'(x)$ being the possibility distribution obtained by means of

$$\pi'(x) = \frac{\max_s f_s(x)}{\max_s \lambda_s}, \quad \forall x,$$

and π any other possibility distribution, then

$$\pi_s \simeq \pi \ \forall s \ \Rightarrow \ \pi' \simeq \pi.$$

The next proposition says that by using this property we can guarantee the fulfilment of the Decomposition and Intersection axioms.

Proposition 5. *Given an equivalence relationship between possibility distributions, $\simeq$, a sufficient condition for the fulfilment of Decomposition is that $\simeq$ verifies the maximum property. Moreover, for the case of strictly positive distributions, the fulfilment of the properties above also guarantees the fulfilment of Intersection.*

Proof.

Decomposition: $I(X \mid Z \mid Y \cup W) \Rightarrow I(X \mid Z \mid Y)$.

We find that $\pi_d(. \mid yzw) \simeq \pi_d(. \mid z)$, $\forall yzw$ such that $\pi(yzw) > 0$. For any y,z such that $\pi(yz) > 0$, let us define, $f_w^{yz}(x) = \pi(xyzw)$, $\lambda_w^{yz} = \pi(yzw)$ and $P_W^{yz} = \{w \in D_W \mid \pi(yzw) > 0\}$. In that case, $\pi_d(x \mid yzw) = f_w^{yz}(x)/\lambda_w^{yz}$ $\forall w \in P_W^{yz}$ and

$$\pi'^{yz}(x) = \frac{\max_{w \in P_W^{yz}} f_w^{yz}(x)}{\max_{w \in P_W^{yz}} \lambda_w^{yz}} = \frac{\max_{w \in D_W} f_w^{yz}(x)}{\max_{w \in D_W} \lambda_w^{yz}}$$
$$= \frac{\pi(xyz)}{\pi(yz)} = \pi_d(x \mid yz).$$

Then, using the maximum property we obtain $\pi_d(. \mid yz) \simeq \pi_d(. \mid z)$ $\forall yz$ such that $\pi(yz) > 0$, thus concluding that $I(X \mid Z \mid Y)$ holds.

Intersection: $I(X \mid Y \cup Z \mid W) \,\&\, I(X \mid Z \cup W \mid Y)$
$\Rightarrow I(X \mid Z \mid Y \cup W)$.

Considering strictly positive distributions, we find that $\pi_d(. \mid yzw) \simeq \pi_d(. \mid yz)$ and $\pi_d(. \mid yzw) \simeq \pi_d(. \mid zw)$ for all yzw. Using the fact that $\simeq$ is an equivalence relationship, particularly the properties of symmetry and transitivity, we have that $\pi_d(. \mid yz) \simeq \pi_d(. \mid wz)$ $\forall yzw$. Let $f_w^z(x) = \pi(xzw)$ and $\lambda_w^z = \pi(zw)$, then $\pi_d(x \mid zw) = f_w^z(x)/\lambda_w^z$, and

$$\pi'^z(x) = \frac{\max_w f_w^z(x)}{\max_w \lambda_w^z} = \frac{\pi(xz)}{\pi(z)} = \pi_d(x \mid z).$$

Therefore, using the maximum property, we obtain $\pi_d(. \mid z) \simeq \pi_d(. \mid yz)$. Now, taking into account that $\pi_d(. \mid yzw) \simeq \pi_d(. \mid yz)$, then once again using transitivity and symmetry we have $\pi_d(. \mid yzw) \simeq \pi_d(. \mid z)$, $\forall yzw$. □

The following corollary can be immediately deduced from the previous proposition:

Corollary 1. *The independence relationship* D4, *where $\simeq$ is any equivalence relationship verifying the maximum property, satisfies the axioms* A1, *and* A3–A5. *Moreover, if the possibility distribution is strictly positive, then* A6 *is also verified.*

Once again, the only property excluded from this context is A2 (Symmetry). In this case, and if we require to the similarity relationship the appropriate properties for each case, both Hisdal and Dempster conditioning perform almost identically [6]: they do not verify A2 and do verify A1, A3–A6, although A6 only for strictly positive distributions in the case of Dempster conditioning.

Different examples of relationships $\simeq$, which are appropriate to define the independence D4 follow (they were also successfully used in [6]):

Isoordering: In that case, we are considering that a possibility distribution essentially establishes an ordering among the values that the variable can take on, and the possibility degrees are of a secondary importance. So, the similarity can be asserted when the two possibility distributions establish the same ordering among their possibility values, i.e.,

$$\pi \simeq \pi' \Leftrightarrow \forall x, x' \; [\pi(x) < \pi(x') \Leftrightarrow \pi'(x) < \pi'(x')].$$

Resemblance: In that case, we consider that two possibility distributions are similar when the possibility degrees for each distribution, for each value, are alike. Formally, let m be any positive integer and let $\{\alpha_k\}_{k=0,\dots,m}$ be real numbers such that $\alpha_0 < \alpha_1 < \cdots < \alpha_m$, with $\alpha_0 = 0$ and $\alpha_m = 1$. We denote $I_k = [\alpha_{k-1}, \alpha_k)$, $k = 1, \dots, m-1$, and $I_m = [\alpha_{m-1}, \alpha_m]$. Then, the similarity relationship $\simeq$ is defined by means of

$$\pi \simeq \pi' \Leftrightarrow \forall x$$
$$\exists k \in \{1, \dots, m\} \text{ such that } \pi(x), \pi'(x) \in I_k.$$

An equivalent version, using α-cuts, is the following:

$$\pi \simeq \pi' \Leftrightarrow C(\pi, \alpha_k) = C(\pi', \alpha_k) \quad \forall k = 1, \dots, m-1,$$

where $C(\pi, \alpha) = \{x \mid \pi(x) \geqslant \alpha\}$.

α_0-Equality: In this case, we are considering a threshold α_0, and supposing that only for values greater than α_0 it is considered interesting to differentiate between the possibility degrees of two distributions. In terms of α-cuts, this relationship may be expressed as follows:

$$\pi \simeq \pi' \Leftrightarrow C(\pi, \alpha) = C(\pi', \alpha) \quad \forall \alpha \geqslant \alpha_0,$$

which is equivalent to

$$\pi \simeq \pi'$$
$$\Leftrightarrow C(\pi, \alpha_0) = C(\pi', \alpha_0) \text{ and } \pi(x) = \pi'(x)$$
$$\forall x \in C(\pi, \alpha_0).$$

We should note that the above similarity relationships are equivalence relationships and verify the maximum property. Therefore, using Proposition 5, we may conclude that:

Corollary 2. *The definition of independence* D4, *where the similarity relationship* $\simeq$ *is either Isoordering, Resemblance or* α_0*-Equality, verifies the properties* A1 *and* A3–A5. *Moreover, if the possibility distribution is strictly positive, then* D4 *also verifies* A6.

Finally, we shall see that the Symmetry axiom does not hold in general using the following examples, which show cases where $I(X \mid \emptyset \mid Y)$ but $\neg I(Y \mid \emptyset \mid X)$.

Isoordering: Let X, Y be bivaluated variables, with the following joint possibility distribution

xy	$\pi(xy)$
x_1y_1	1
x_1y_2	0.8
x_2y_1	0.7
x_2y_2	0.7

In that case, considering the marginal distribution on X, we obtain the ordering $x_2 \prec x_1$, and after conditioning to y ($\pi_d(.\mid y_1)$ and $\pi_d(.\mid y_2)$), the same ordering results. So, we obtain $I(X \mid \emptyset \mid Y)$. On the other hand, if we marginalize on Y, we have $y_2 \prec y_1$, but when conditioning to x_2 we obtain $y_2 \not\prec y_1$, so $\pi(y)$ and $\pi_d(y \mid x_2)$ are not similar, hence $\neg I(Y \mid \emptyset \mid X)$.

Resemblance: We use the same distribution as above, and consider the following discretization for the unit interval: $I_1 = [0, 0.7) I_2 = [0.7, 0.9); I_3 = [0.9, 1]$. In that case, we find that $\pi(x_1), \pi_d(x_1 \mid y) \in I_3$ $\forall y$ and $\pi(x_2), \pi_d(x_2 \mid y) \in I_2$ $\forall y$, and therefore $I(X \mid \emptyset \mid Y)$. However, $\pi(y_2) \in I_2$ but $\pi_d(y_2 \mid x_2) \in I_1$, hence $I(Y \mid \emptyset \mid X)$ does not hold.

α_0-Equality: Consider two variables X, Y, with X taking values in $D_X = \{x_1, x_2\}$ and Y taking values

in $D_Y = \{y_1, y_2, y_3\}$. Let us use any threshold $\alpha_0 > 0.5$, and suppose the following joint possibility distribution:

xy	$\pi(xy)$
x_1y_1	1.0
x_1y_2	0.4
x_1y_3	1.0
x_2y_1	0.5
x_2y_2	0.2
x_2y_3	0.4

In that case, we find that $\pi(x_1) = \pi_d(x_1 \mid y) = 1\ \forall y$, and $\pi(x_2), \pi_d(x_2 \mid y) < \alpha_0\ \forall y$. Then, as our interest is in the equality only for values greater than the threshold, we can conclude that $I(X \mid \emptyset \mid Y)$ holds. On the other hand, we may find that $\pi_d(y_3 \mid x_1) = \pi(y_3) = 1 \neq 0.8 = \pi_d(y_3 \mid x_2)$ and then the independence relationship $I(Y \mid \emptyset \mid X)$ does not hold.

The following table summarizes the different properties that the different definitions of independence we are considering verify. The symbol 'X' means that the corresponding property holds, and 'P' means that the property holds only for strictly positive distributions.

	A1	A2	A3	A4	A5	A6
D1 (Eq. (6))	X	X	X	X	X	P
D2 (Eq. (7))	X	X	X		X	
D3 (Eq. (8))	X		X	X	X	X
D4 (Eq. (10))	X		X	X	X	P

4. The marginal problem

Following the scheme given in [6], we shall study the *marginal problem*: suppose that X, Y, Z are three disjoint subsets of variables, and that π_1 and π_2 are two possibility distributions defined over XZ and YZ respectively. The problem is how to construct, from π_1 and π_2, a joint possibility distribution, π, over XYZ. In that case, it is reasonable to assume that there exists some kind of conditional independence relationship among the variables X, Y and Z, particularly $I(X \mid Z \mid Y)$. Another natural requirement is that the marginal distributions of π over the domains XZ and YZ should coincide with the original distributions π_1 and π_2. Therefore, starting out from π_1 and π_2, we want to construct a joint possibility distribution π satisfying the following conditions:

1. X and Y must be independent given Z, i.e. $I(X \mid Z \mid Y)$.
2. The marginal measure on XZ must be preserved, i.e.: $\pi(xz) = \max_y \pi(xyz) = \pi_1(xz)$.
3. The marginal measure on YZ must be preserved, i.e.: $\pi(yz) = \max_x \pi(xyz) = \pi_2(yz)$.

If we want to meet these requirements, we must impose a compatibility condition on the original distributions, π_1 and π_2, which ensures that both distributions represent the same information about the common variable Z. This compatibility relationship is defined in the following way:

Definition 17. Let π_1 and π_2 be two possibility distributions on XZ and YZ, respectively. We say that π_1 and π_2 are compatible on Z if and only if

$$\forall z \in D_Z, \quad \pi_1(z) = \pi_2(z).$$

If the two original distributions are not compatible on Z, then obviously we cannot build π, in fact in this case it makes no sense to try it, because we are mixing incoherent information. In the case of compatibility, we must fix an independence criterion that allows us to determine the joint distribution $\pi(xyz)$. We shall impose that X and Y are conditionally independent given Z, using definition D1. The next proposition shows that the use of D1 guarantees the fulfilment of the previous requirements.

Proposition 6. *Let π_1 and π_2 be two possibility distributions, defined on XZ and YZ, respectively, and compatible on Z. Then, the joint possibility distribution, π, when defined by means of*

$$\pi(xyz) = \begin{cases} \pi_1(x \mid z)\pi_2(yz) \\ \quad = \pi_1(xz)\pi_2(y \mid z) \\ \quad = \dfrac{\pi_1(xz)\pi_2(yz)}{\pi_1(z)} & \text{if } \pi_1(z) > 0, \\ 0 & \text{if } \pi_1(z) = 0 \end{cases}$$

satisfies:

1. $I_{D1}(X \mid Z \mid Y)$, *i.e.*, $\pi_d(x \mid yz) = \pi_d(x \mid z)$, $\forall xyz$ *such that* $\pi(yz) > 0$.
2. $\pi(xz) = \pi_1(xz)$, $\forall xz$.
3. $\pi(yz) = \pi_2(yz)$, $\forall yz$.

Proof. Requirements 2 and 3 are immediate and, using them, the independence property is immediate too. □

Example 1. Let π_1 and π_2 be the possibility distributions displayed in the tables below. Assuming $I(X \mid Z \mid Y)$, we can construct the following joint possibility distribution π:

x	z	$\pi_1(xz)$
x_1	z_1	0.25
x_1	z_2	0
x_2	z_1	0.5
x_2	z_2	1

y	z	$\pi_2(yz)$
y_1	z_1	0.5
y_1	z_2	0.75
y_2	z_1	0
y_2	z_2	1

x	z	y	$\pi(xyz)$
x_1	z_1	y_1	0.25
x_1	z_1	y_2	0
x_1	z_2	y_1	0
x_1	z_2	y_2	0
x_2	z_1	y_1	0.5
x_2	z_1	y_2	0
x_2	z_2	y_1	0.75
x_2	z_2	y_2	1

□

5. Storage of large possibility distributions

Regardless of the way that we interpret a possibility distribution, one important question we have to deal with is how to store the values of a joint distribution. Obviously, we can simply use a table of values, but then we need a size which is exponential (in the number of variables) to store all the information. In this section we present two approaches, where the use of the conditional independence relationship D1 permits us to reduce the memory requirements. The independence relationships are used to factorize the joint possibility distribution in terms of its conditionally independent components. Then, using the method presented in the previous section, we can construct the original distribution without losing information.

Both methods use a tree structure to represent the joint distribution, the first one will be called Possibilistic Tree Structure (PT) (and is similar to the one proposed in [6], but now using Dempster conditioning instead of Hisdal's) and the second one will be called Dependence Tree Structure (DT). The main differences between them are the way in which the possibility values are stored, the different meaning of the nodes, and how the set of independence relationships is used to construct each tree.

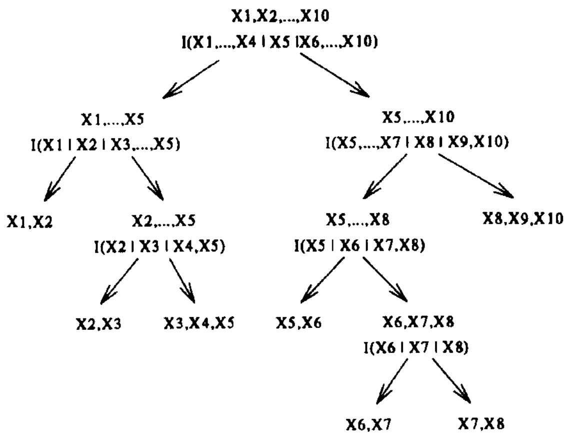

Fig. 1. Possibilistic tree.

Possibilistic tree structures. Let X, Y, Z be disjoint subsets of variables and π a joint possibility distribution. If $I(X \mid Z \mid Y)$ holds using definition D1, i.e., $\pi_d(x \mid yz) = \pi_d(x \mid z)$, then we can recover the joint distribution on XYZ by means of its marginals on XZ and YZ, using the equality $\pi(xyz) = \pi(xz)\pi(yz)/\pi(z)$. In order to generalize this idea, consider a set U with n variables, $U = \{X_1, X_2, \ldots, X_n\}$, and let X_i be the variable (equivalently for a set of variables) such that $I(\{X_1, \ldots, X_{i-1}\} \mid X_i \mid \{X_{i+1}, \ldots, X_n\})$. In that case, we can split the initial distribution into two components defined on $\{X_1, \ldots, X_{i-1}, X_i\}$ and $\{X_i, X_{i+1}, \ldots, X_n\}$. The same idea can be recursively applied to both subsets, thus forming the PT structure. We only have to store the marginal possibility distributions on the leaves of the tree. Let us look at an example:

Example 2. Let $U = \{X_1, \ldots, X_{10}\}$ be a set of bivaluated variables, and suppose that the independence relationships indicated in the nodes of the tree in Fig. 1 are true. We only have to store the following set of marginal distributions:

$$\pi(x_1, x_2), \pi(x_2, x_3), \pi(x_3, x_4, x_5), \pi(x_5, x_6), \pi(x_6, x_7),$$
$$\pi(x_7, x_8), \pi(x_8, x_9, x_{10})$$

and therefore we considerably reduce the memory requirements (we store only 36 values instead of the $2^{10} = 1024$ values that completely define the joint distribution).

The joint distribution can be obtained by combining the marginal distributions in a bottom-up approach, using the method proposed in Proposition 6; for example, in Fig. 1, $\pi(x_2,x_3,x_4,x_5) = \pi(x_2x_3)\pi(x_3x_4x_5)/\pi(x_3)$, and also $\pi(x_1,x_2,x_3,x_4,x_5) = \pi(x_1,x_2)\pi(x_2,x_3,x_4,x_5)/\pi(x_2)$. This process can be continued until the root node is reached.

The general method for building a Possibilistic Tree is based on the recursive procedure *SPLIT*, which is described below. The parameters F and *node* in *SPLIT(F,node)* represent a subset of variables in U and a node in the tree, respectively. Initially, the algorithm takes as the input a node, *root*, labelled with U, and gives as the output the tree structure.

- The recursive procedure *SPLIT(F,node)* is defined as follows:
  - Find disjoint subsets L, R, S in F such that $L \cup S \cup R = F$ and $I(L \mid S \mid R)$.
  - If this is possible, then attach the set S to *node*, create two new child nodes of *node*, *leftchild* and *rightchild* with respective labels $L \cup S$ and $R \cup S$, and call *SPLIT*($L \cup S$,*leftchild*) and *SPLIT*($R \cup S$,*rightchild*).
  - Otherwise, attach the marginal possibility distribution π_F to *node*.

The next proposition shows how to obtain the joint distribution.

Proposition 7. *Let T be a Possibilistic Tree for the set of variables $U = \{X_1,X_2,\ldots,X_n\}$, L_j, $j = 1,\ldots,m$ the leaves in T and I_k, $k = 1,\ldots,r$ the internal nodes in T. Then the joint possibility distribution on $X_1,X_2,\ldots,X_n$ can be obtained by means of*

$$\pi = \prod_{j=1}^{m} \pi_{L_j} \left(\prod_{k=1}^{r} \pi_{I_k} \right)^{-1},$$

where π_{L_j} is the marginal possibility distribution stored in the leaf L_j, and π_{I_k} represents the marginal possibility distribution over the set of variables attached to node I_k (which splits the set of variables that constitute the label of I_k into two conditionally independent subsets).

Proof. The proof is immediate, using the independence relationships represented in the possibilistic tree and the definition of independence D1. □

If we are interested in obtaining some marginal possibility distributions (instead of the joint distribution), then we can take advantage of the independence relationships represented in the PT, and we do not use all the variables in the tree, i.e., it is not necessary to construct the joint distribution first (using the previous proposition) and after marginalizing (see the algorithm proposed in [6], which can be adapted to this case by simply changing the combination operator).

Dependence tree structures. In order to better explain what Dependence Trees are, it is interesting to remark that a Possibilistic Tree has the following characteristics:

- There are two kind of nodes: leaf nodes, which store possibility distributions for subsets of variables, and internal nodes that store conditional independence assertions.
- The leaf nodes store marginal possibility distributions.
- The possibility distributions for different leaf nodes may share common variables, i.e., the leaf nodes do not represent mutually exclusive subsets of variables.
- The tree does not explicitly represent the marginal distributions for the subsets of variables attached to the internal nodes, although these distributions are needed to construct the joint distribution. Anyway, these distributions can be easily computed from the distributions stored in the leaves (note that the Possibilistic Trees constructed in [6] using Hisdal conditioning did not need to use these 'internal' distributions).

Now, we look for a different tree representation, which will use only one kind of node instead of two, will store conditional distributions instead of marginals, the nodes will represent mutually exclusive subsets of variables, and the computation of the joint distribution will be done directly using the information stored in the nodes.

In a Dependence Tree we use a similar scheme to the one employed in probabilistic dependence graphs [24]. In these structures, nodes represent variables, links represent direct dependence relationships among variables, which are quantified by means of conditional probability distributions, and the absence of a link between two variables represents a conditional independence relationship. In our case, and in order to always obtain a tree structure, we will allow the nodes to represent subsets of variables.

Before studying the general case, let us see an illustrative example:

Example 3. Let $U = \{X_1, X_2, \ldots, X_8\}$ be a set of variables and suppose that $I(X_1 \mid X_2X_3 \mid X_4X_5X_6X_7X_8)$ holds. In that case, the set of variables U can be split as in Fig. 2(a), where in the root node, (X_2X_3), we store the marginal distribution $\pi_{X_2X_3}$ and for each leaf we store the conditional possibility distributions, $\pi_{X_1 \mid X_2X_3}$ and $\pi_{X_4..X_8 \mid X_2X_3}$. Thus, using the above independence relationship, we can obtain the joint distribution by means of

$$\pi = \pi_{X_2X_3} * \pi_{X_1 \mid X_2X_3} * \pi_{X_4..X_8 \mid X_2X_3}.$$

Now, let us suppose that the independence relationship $I(X_8 \mid X_2X_3 \mid X_4X_5X_6X_7)$ is verified. Again, we can split the leaf $(X_4X_5X_6X_7X_8)$ as in Fig. 2(b), where the conditional distributions $\pi_{X_8 \mid X_2X_3}$ and $\pi_{X_4..X_7 \mid X_2X_3}$ must be stored at nodes (X_8) and $(X_4X_5X_6X_7)$, respectively. Note that

$$\pi_{X_4..X_8 \mid X_2X_3} = \pi_{X_8 \mid X_2X_3} * \pi_{X_4..X_7 \mid X_2X_3}$$

and therefore the joint distribution can be written as

$$\pi = \pi_{X_2X_3} * \pi_{X_1 \mid X_2X_3} * \pi_{X_8 \mid X_2X_3} * \pi_{X_4..X_7 \mid X_2X_3}.$$

Suppose now that $I(X_6X_7 \mid X_4X_5 \mid X_2X_3)$ also holds. In this case, we can create a new level in the tree (see Fig. 2(c)). Again, at each new node, we store the conditional possibility distribution given its parent. Note that $\pi_{X_4X_5 \mid X_2X_3} * \pi_{X_6X_7 \mid X_4X_5} = \pi_{X_4X_5 \mid X_2X_3} * \pi_{X_6X_7 \mid X_2..X_5} = \pi_{X_4..X_7 \mid X_2X_3}$, so that the joint distribution can be decomposed as

$$\pi = \pi_{X_2X_3} * \pi_{X_1 \mid X_2X_3} * \pi_{X_8 \mid X_2X_3} * \pi_{X_4X_5 \mid X_2X_3} * \pi_{X_6X_7 \mid X_4X_5}.$$

Finally, if the independence relationship $I(X_6 \mid X_4X_5 \mid X_7)$ holds, the tree structure becomes as in Fig. 2(d), with $\pi_{X_6 \mid X_4X_5}$ and $\pi_{X_7 \mid X_4X_5}$ being the conditional distributions stored at nodes (X_6) and (X_7) respectively, and verifying $\pi_{X_6X_7 \mid X_4X_5} = \pi_{X_6 \mid X_4X_5} * \pi_{X_7 \mid X_4X_5}$. Therefore, using all the independence relationships considered, the joint possibility distribution can be obtained by means of

$$\pi = \pi_{X_2X_3} * \pi_{X_1 \mid X_2X_3} * \pi_{X_8 \mid X_2X_3} * \pi_{X_4X_5 \mid X_2X_3} * \pi_{X_6 \mid X_4X_5} * \pi_{X_7 \mid X_4X_5}.$$

So, we have to store a marginal possibility distribution in the root node, and for any other node in the structure we store the conditional possibility distributions given its parent node in the tree. Note that the joint possibility distribution can be written (as it happens for probabilistic dependence graphs [24]), as a product of the distributions stored in the nodes. □

Given a joint possibility distribution, the following procedures, *DIV* and *FACT*, permit us to build a Dependence Tree structure. *DIV* has as input a set of variables, U, and the root node for the tree structure. Therefore, initially it must be called as *DIV*(U,*root*). This procedure produces the Dependence Tree as output. On the other hand, *FACT* is a recursive procedure, used by *DIV*, that takes as input the list of nodes at level i in the tree, denoted by $NL(i)$, and then, using independence relationships, the tree grows by creating a new level $i+1$.

DIV(U,*root*):

- Find disjoint subsets L, R, P in U such that $U = L \cup R \cup P$ and $I(L \mid P \mid R)$ holds.
- If this is possible and $P \neq \emptyset$, then label the root node with the subset P, attach to it the marginal possibility distribution π_P, create two child nodes of *root*, with labels L and R, and insert in the list of nodes at level 1, $NL(1)$, the subsets L and R. Then call the procedure *FACT*($NL(1)$) and return the tree obtained.

  If $P = \emptyset$, i.e., $I(L \mid \emptyset \mid R)$, then create two new tree structures with roots $root_L$ and $root_R$, and call to *DIV*(L,$root_L$) and *DIV*(R,$root_R$).
- Otherwise, attach the joint distribution π_U to *root*.

FACT($NL(i)$): *We denote by n any node in* $NL(i)$, *by N the subset of variables in n, by* p_n *the parent of node n in the structure, and by* P_n *the subset of variables in* p_n.

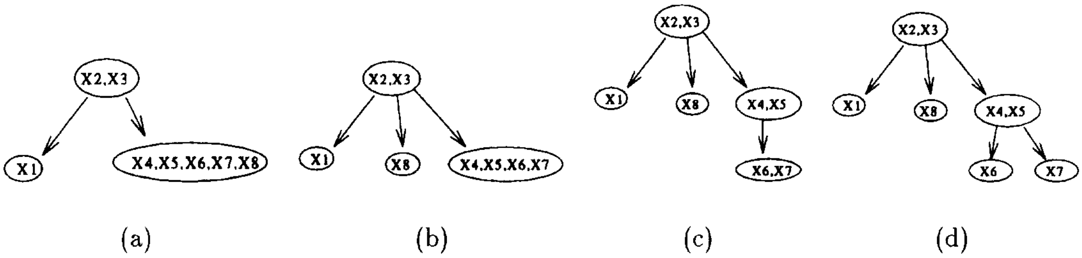

Fig. 2. Construction of a Dependence Tree.

- Whenever there is a node $n \in NL(i)$ verifying that exist disjoint subsets $L, R \subseteq N$ such that $L \cup R = N$ and $I(L \mid P_n \mid R)$, then remove node n from the structure and N from $NL(i)$, create two new child nodes of p_n with labels L and R, and insert in the list of nodes $NL(i)$ the subsets L and R.
- For each node $n \in NL(i)$
  - If there are disjoint subsets $N_i, N_{i+1} \subseteq N$ such that $N_i \cup N_{i+1} = N$ and $I(P_n \mid N_i \mid N_{i+1})$, then remove N from $NL(i)$, change the label of node n from N to N_i and attach to n the conditional distribution $\pi_{N_i \mid P_n}$; create a new child node of n with the label N_{i+1}, and insert N_{i+1} in the list of nodes at level $i+1$, $NL(i+1)$.
  - Otherwise, attach the conditional distribution $\pi_{N \mid P_n}$ to node n.
- If $NL(i+1) \neq \emptyset$ then call $FACT(NL(i+1))$.

Using these procedures, a tree structure representing the original joint distribution may be obtained. We need to store a marginal possibility distribution in the root node, and for any other node n in the tree, we store the conditional possibility distributions of the label of n given the label of its parent p_n in the tree, $\pi_{N \mid P_n}$. Then, the joint possibility distribution can be obtained, in a similar way to the case of the probabilistic dependence graph, as follows:

Proposition 8. *Let T be a Dependence Tree for a set of variables U, with nodes $n_1, \ldots, n_m$. Then, the joint possibility distribution on the variables in U can be obtained by means of*

$$\pi = \prod_{i=1}^{m} \pi_{N_i \mid P_{n_i}},$$

where N_i is the set of variables that form the label of node n_i, and P_{n_i} is the set of variables that form the label of the parent node of n_i in T (if n_1 is the root node, it is assumed that $P_{n_1} = \emptyset$).

Proof. The proof is immediate, taking into account the independence relationships used to build the dependence tree. □

Bearing in mind the way in which the Dependence Tree has been built, we can obtain the following graphical independence criterion for these structures:

Proposition 9. *Let T be a Dependence Tree for a set of variables U, and let X, Y, Z be disjoint subsets in U. If all the paths connecting nodes that include a variable in X to nodes that include a variable in Y contain some node whose associated subset of variables is included in Z, then $I(X \mid Z \mid Y)$.*

Proof. The result can be easily deduced using the independence relationships represented in the structure and the properties of the definition of independence D1. □

The previous proposition permits us to deduce new independence relationships from the ones used to build the tree. For example, for the dependence tree in Fig. 2(d), we can deduce that $I(X_6 \mid X_2X_4X_5 \mid X_1X_8)$ is a true conditional independence assertion. Let us see how this assertion can be derived from the axioms verified by D1: we start out from three of the independence statements used to construct the tree, namely $I(X_1|X_2X_3|X_4..X_8)$ (I1), $I(X_8|X_2X_3|X_4..X_7)$ (I2) and $I(X_6X_7|X_4X_5|X_2X_3)$ (I3) From (I1) we obtain, using weak union and symmetry $I(X_4..X_7|X_2X_3X_8|X_1)$ (I4);

from (I2) we obtain $I(X_4..X_7|X_2X_3|X_8)$ (I5) by using symmetry; then, from (I4) and (I5) we deduce $I(X_4..X_7|X_2X_3|X_1X_8)$ (I6) by applying contraction; now, from (I6), and using symmetry, decomposition, weak union and symmetry (in this order), we obtain $I(X_6|X_2X_3X_4X_5|X_1X_8)$ (I7); if we apply symmetry, decomposition and once again symmetry to (I3), we get $I(X_6\,|\,X_4X_5\,|\,X_2X_3)$ (I8); now, contraction applied to (I7) and (I8) produces $I(X_6\,|\,X_4X_5\,|\,X_1X_2X_3X_8)$ (I9); finally, decomposition and weak union, when applied to (I9), give $I(X_6\,|\,X_2X_4X_5\,|\,X_1X_8)$. This derivation gives us an idea of the power of Proposition 9.

It should be noted that the algorithms proposed to obtain PT and DT structures representing joint possibility distributions are not deterministic, i.e., we have some freedom to decide what independence assertions we should test for, and in the case that more than one independence assertion were true, we have also freedom to select which of them to use to build the tree. In other words, we could obtain different PT and different DT representations of the same joint distribution, depending on the search strategy we use. This leaves open the important topic of studying heuristic methods to find optimal PT and DT representations of joint distributions (with the word *optimal* meaning, for example, minimum storage requirements, or maximum efficiency when using these structures for inference tasks). It should also be pointed out that the two tree representations proposed do not rely on the uncertainty formalism being used, possibility theory in this case, but they can be also used for other uncertainty theories, provided that we have the appropriate concept of conditional independence within each theory.

Although the two tree representations of joint distributions discussed, Possibilistic Trees and Dependence Trees are quite different, there are some connections between them; for example, from a Dependence Tree we can always obtain an equivalent Possibilistic Tree (i.e., representing the same set of conditional independence assertions), but the converse is not necessarily true. In Fig. 3, a PT equivalent to the Dependence Tree of Fig. 2(d) is displayed. The transformation process is quite obvious.

From a Possibilistic Tree we can also obtain a Dependence Tree, but in this case we cannot guarantee that both trees represent the same conditional independences. For example, from the Possibilistic Tree in Fig. 1 we can construct the DT displayed in

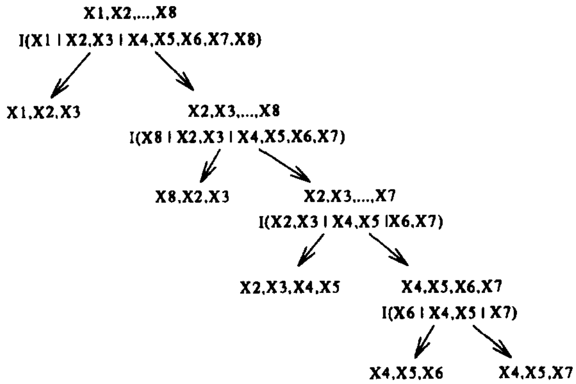

Fig. 3. A Possibilistic Tree equivalent to the Dependence Tree in Fig. 2(d).

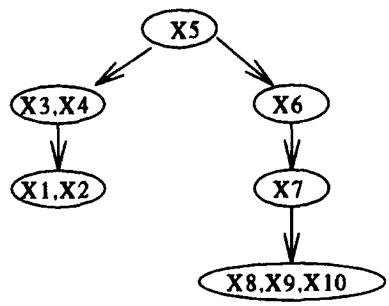

Fig. 4. Dependence tree obtained from the Possibilistic Tree in Fig. 1.

Fig. 4 (using the independence relationships represented explicitly in the PT, and some others which can be deduced from them through the axioms, concretely $I(X_5\,|\,X_3X_4\,|\,X_1X_2)$, $I(X_5\,|\,X_6\,|\,X_7..X_{10})$ and $I(X_6\,|\,X_7\,|\,X_8X_9X_{10})$). Note that the independence statement $I(X_1\,|\,X_2\,|\,X_3X_4X_5)$, which appears in the PT, cannot be deduced from the independence statements used to build the DT.

Although this brief analysis of the relationships between Possibilistic and Dependence Trees should be further investigated, it seems to point out that Possibilistic Trees are more expressive than Dependence Trees, from the point of view of the independence relationships that can be represented. However, DT structures could be extended to more general structures, using graphs instead of trees, which would be expected to have a representational power comparable to or even greater than possibilistic trees.

6. Concluding remarks

We have proposed and analyzed several concepts of possibilistic conditional independence. In order to define it, our approach has been based on using several criteria to compare conditional possibility distributions (before and after obtaining new pieces of information, i.e., 'a priori' and 'a posteriori' possibility distributions). Particularly, in this paper we have approached possibility measures as consonant plausibility measures, and therefore Dempster conditioning has been used (in the first part of this paper [6] we developed a similar study using Hisdal conditioning instead of Dempster's).

Three different comparison criteria have been proposed: the first establishes the independence when the 'a priori' information is not modified at all after conditioning (D1); then we relaxed this criterion and formalized the independence when we obtain less precise information (D2, D3) or obtain similar information (D4) after conditioning. In order to compare these independence criteria, we used a well-known set of axioms (the graphoid axioms) that capture the intuitive notion of independence. We saw that D1 satisfied all the axioms, whereas for D3 and D4, the only axiom which was not verified is Symmetry, and D2 did not verify Weak Union and Intersection.

Moreover, we have studied the marginal problem, i.e., how to construct a joint possibility distribution from marginal distributions, assuming a conditional independence relationship. As a direct application, we found that it is possible to factorize a joint possibility distribution using the independence criterion D1, and then recover the original distribution. This amounts to a considerable saving in the storage requirements of large joint possibility distributions and should lead to efficient inference algorithms, using local computations.

It is obvious that a lot remains to be done. For example, considering other points of view to define independence which are not based on conditioning; studying the relationships of Possibilistic Trees with other structures to perform inferences within this framework, such as Hypergraphs [15, 20], and also to tackle the problem of how to construct efficiently a Possibilistic Tree from a joint possibility distribution; it is also interesting to consider how to use these structures with respect to the problem of extracting or estimating possibility distributions (from expert judgments or from raw data); in the latter case, we could use the expert knowledge to construct a Possibilistic Tree and then estimate the possibility distribution for each leaf, or we could design procedures to build the tree directly from databases.

Another interesting problem is the use of more complex structures to store possibility distributions, such as Dependence Graphs [24] (and not only Dependence Trees). In these structures we are representing dependence and independence relationships among variables. Therefore, an important problem is that of propagating (i.e., updating using local computation) the information using the independence relationships represented in the Dependence Graph. Finally, the study of the consequences of a non-symmetrical definition of independence with respect to its graphical representation (by means of Possibilistic Trees or Dependence Graphs) is also an interesting task.

References

[1] S. Benferhat, D. Dubois, H. Prade, Expressing independence in a possibilistic framework and its application to default reasoning, in: A. Cohn (Ed.), 11th European Conf. on Artificial Intelligence, Wiley, New York, 1994, pp. 150–154.

[2] L.M. de Campos, Independence relationships in possibility theory and their application to learning belief networks, in: G. Della Riccia, R. Kruse, R. Viertl (Eds.), Mathematical and Statistical Methods in Artificial Intelligence, CTSM Courses and Lectures, vol. 363, Springer, Wien, 1995, pp. 119–130.

[3] L.M. de Campos, J. Gebhardt, R. Kruse, Axiomatic treatment of possibilistic independence, in: C. Froidevaux, J. Kohlas (Eds.), Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Lecture Notes in Artificial Intelligence, vol. 946, Springer, Berlin, 1995, pp. 77–88.

[4] L.M. de Campos, J.F. Huete, Independence concepts in upper and lower probabilities, in: B. Bouchon-Meunier, L. Valverde, R.R. Yager (Eds.), Uncertainty in Intelligent Systems, North-Holland, Amsterdam, 1993, pp. 49–59.

[5] L.M. de Campos, J.F. Huete, Learning non probabilistic belief networks, in: M. Clarke, R. Kruse, S. Moral (Eds.), Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Lecture Notes in Computer Science, vol. 747, Springer, Berlin, 1993, pp. 57–64.

[6] L.M. de Campos, J.F. Huete, Independence concepts in possibility theory: Part I, Fuzzy Sets and Systems 103 (1999) 127–152.

[7] L.M. de Campos, J.F. Huete, S. Moral, Possibilistic independence, Proc. 3rd European Congr. on Intelligent Techniques and Soft Computing (EUFIT), vol. 1, 1995, pp. 69–74.

[8] L.M. de Campos, S. Moral, Independence concepts for convex sets of probabilities, in: P. Besnard, S. Hanks (Eds.), Proc. 11th Conf. on Uncertainty in Artificial Intelligence, Morgan and Kaufmann, San Mateo, 1995, pp. 108–115.
[9] G. de Cooman, E.E. Kerre, A new approach to possibilistic independence, in: Proc. 3rd IEEE Internat. Conf. on Fuzzy Systems, 1994, pp. 1446–1451.
[10] A.D. Dawid, Conditional independence in statistical theory, J. Roy. Statist. Soc. Ser. B 41 (1979) 1–31.
[11] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Statist. 38 (1967) 325–339.
[12] D. Dubois, Belief structures, possibility theory, decomposable confidence measures on finite sets, Comput. Artificial Intelligence 5 (1986) 403–417.
[13] D. Dubois, L. Farinas del Cerro, A. Herzig, H. Prade, An ordinal view of independence with applications to plausible reasoning, in: R. López de Mántaras, D. Poole (Eds.), Proc. 10th Conf. on Uncertainty in Artificial Intelligence, Morgan and Kaufmann, 1994, pp. 195–203.
[14] D. Dubois, H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, New York, 1988.
[15] D. Dubois, H. Prade, Inference in possibilistic hypergraphs, in: B. Bouchon-Meunier, R.R. Yager, L.A. Zadeh (Eds.), Uncertainty in Knowledge Bases, Lecture Notes in Computer Science, vol. 521, Springer, Berlin, 1990, pp. 250–259.
[16] L. Farinas del Cerro, A. Herzig, Possibility theory and independence, Proc. 5th Internat. Conf. on Information Processing and Management of Uncertainty in Knowledge Based Systems, IPMU, 1994, pp. 820–825.
[17] P. Fonck, Conditional independence in posibility theory, in: R. López de Mántaras, D. Poole (Eds.), Proc. 10th Conf. on Uncertainty in Artificial Intelligence, Morgan and Kaufmann, San Mateo, 1994, pp. 221–226.
[18] E. Hisdal, Conditional possibilities, independence and non-interaction, Fuzzy Sets and Systems 1 (1978) 283–297.
[19] J.F. Huete, Aprendizaje de Redes de Creencia mediante la Detección de Independencias: modelos no Probabilisticos, Thesis, Universidad de Granada, Spain, 1995 (in Spanish).
[20] R. Kruse, J. Gebhardt, F. Klawonn, Foundations of Fuzzy Systems, Wiley, New York, 1994.
[21] M.T. Lamata, S. Moral, Classification of fuzzy measures, Fuzzy Sets and Systems 33 (1989) 243–253.
[22] S.L. Lauritzen, A.P. Dawid, B.N. Larsen, H.G. Leimer, Independence properties of directed Markov fields, Network 20 (1990) 491–505.
[23] S. Moral, L.M. de Campos, Updating uncertain information, in: B. Bouchon-Meunier, R.R. Yager, L.A. Zadeh (Eds.), Uncertainty in Knowlede Bases, Lecture Notes in Computer Science, vol. 521, Springer, Berlin, 1991, pp. 58–67.
[24] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan and Kaufmann, San Mateo, 1988.
[25] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[26] P.P. Shenoy, Conditional independence in uncertainty theories, in: D. Dubois, M.P. Wellman, B.D'Ambrosio, P. Smets (Eds.), Proc. 8th Conf. on Uncertainty in Artificial Intelligence, Morgan and Kauffmann, San Mateo, 1992, pp. 284–291.
[27] M. Studený, Formal properties of conditional independence in different calculi of A.I., in: M. Clarke, R. Kruse, S. Moral (Eds.), Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Lecture Notes in Computer Science, vol. 747, Springer, Berlin, 1993, pp. 341–348.
[28] M. Sugeno, Theory of fuzzy integrals and its applications, Thesis, Tokyo Institute of Technology, Japan (1974).
[29] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28.