Epistemic Irrelevance on Sets of Desirable Gambles

Serafín Moral
Dpto. de Ciencias de la Computación e IA
Universidad de Granada
18071 - Granada (Spain)
smc@decsai.ugr.es

Abstract

This paper studies graphoid properties for epistemic irrelevance in sets of desirable gambles. For that aim, the basic operations of conditioning and marginalization are expressed in terms of variables. Then, it is shown that epistemic irrelevance is an asymmetric graphoid. The intersection property is verified in probability theory when the global probability distribution is positive in all the values. Here it is always verified due to the handling of zero probabilities in sets of gambles. It is also presented an asymmetrical D-separation principle, by which this type of independence relationships can be represented in directed acyclic graphs.

Keywords. Desirables gambles, imprecise probabilities, conditioning, epistemic independence, epistemic irrelevance.

1 Introduction

Coherent sets of desirable gambles [13, 14] are a very general model for imprecise probability. They are more informative than convex sets of probability measures as they can provide the behavior conditioned to events of probability zero. Though its generality, this model has a simple mathematical formulation, allowing to express concepts as conditioning, combination and marginalization in a very simple way. Perhaps, its main drawback may be the difficulty of elicitation from an expert due to the implications of considering as desirable or not the gambles in the frontier.

Independence is one of the key concepts for every theory of uncertainty. In imprecise probability this concepts is richer than in the classical theory of probability allowing a number of different interpretations [1, 3]. In the model of desirable sets of gambles the most natural definition is epistemic irrelevance: adding information about one event, then the information about the other does not change. As in credal sets [12, 3], this concept is not symmetrical. If we impose irrelevance in both directions we obtain epistemic independence.

This paper tries to investigate the graphoid properties for epistemic irrelevance in sets of desirable gambles, extending the work by Cozman and Walley [3] in which this concept was studied for credal sets of probabilities. As in that setting, the symmetry and redundancy properties are not verified. However, if we consider the symmetrical version of independence, then contraction is not verified, and this is a much more important property from our point of view. We also investigate how these models of independence can be represented by directed acyclic graphs, by introducing the AD-criterion (Asymmetrical D-separation). This representation will allow in some cases to discriminate between the two possible orientations of the arcs in cases that are indistinguishable when probabilistic independences are considered.

The paper is organized as follows. Section 2 considers the basics concepts of desirable sets of gambles, but expressed in the language of variables instead of the language of events. This is more usual for the study of graphoid axioms. Section 3 presents the graphoid axioms proving all of them except symmetry and redundancy, and including intersection, a property that can be proved only with strictly positive distributions in the probabilistic case. Section 4 presents the AD-separation principle and Section 5 is devoted to the conclusions.

2 Sets of Desirable Gambles

This section introduces the notation and the fundamental concepts of sets of desirable gambles.

We shall consider variables $X, Y, Z, W, \ldots$ taking values on finite sets $U_X, U_Y, U_Z, U_W, \ldots$, respectively. If X, Y are variables, (X, Y) will be the joint variable taking values on $U_X \times U_Y$.

A gamble about X is a real function, f, defined on U_X. Each value $f(x)$ represents the gain associated to the result $X = x$. Let $\mathcal{L}(X)$ be the set of all the possible gambles about X. A piece of information about X is a set of gambles, $\mathcal{D}(X) \subseteq L(X)$, satisfying the following properties (a coherent set of gambles) [14]:

D1. $0 \notin \mathcal{D}(X)$.

D2. if $f \geq 0, f \neq 0$, then $f \in \mathcal{D}(X)$

D3. if $f \in \mathcal{D}(X)$ and $\lambda > 0$, then $\lambda.f \in \mathcal{D}(X)$

D4. if $f, g \in \mathcal{D}(X)$, then $f + g \in \mathcal{D}(X)$

$\mathcal{D}(X)$ represents the set of strictly desirable gambles, and it is different to the desirable gambles considered in [15, 6] where the 0 gamble was considered as desirable. Considering only strict desirability has some advantages as allowing to express how to calculate conditional probabilities to events of probability 0 [14].

If $\mathcal{D}_1(X)$ and $\mathcal{D}_2(X)$ are coherent sets then, $\mathcal{D}_1(X)$ is said to be less informative than $\mathcal{D}_2(X)$ if and only if $\mathcal{D}_1(X) \subseteq \mathcal{D}_2(X)$.

If $\mathcal{R}(X)$ is a set of gambles, then the set of gambles generated by application of properties D2, D3, and D4 (the intersection of all the sets verifying these properties and containing $\mathcal{R}(X)$) will be called the *natural extension* of $\mathcal{R}(X)$ and denoted by $\overline{\mathcal{R}(X)}$. If the natural extension verifies D1, then it will a coherent set of gambles, more precisely it will be the least informative coherent set of gambles containing $\mathcal{R}(X)$.

This idea will be used to extend the concept of natural extension to other situations in which we do not have an initial set $\mathcal{R}(X)$, but an specific condition that should be verified by $\mathcal{D}(X)$. Natural extension will determine the least informative coherent set verifying that condition. For this to make sense it is necessary to prove that there is at least a coherent set verifying the property.

Sets of desirable gambles are a more general approach to represent uncertainty than convex credal sets in the following sense. A set of desirable gambles $\mathcal{D}(X)$ defines a convex credal set by,

$$\mathcal{M}(X) = \{P : E_P[f] \geq 0, \forall f \in \mathcal{D}(X)\} \qquad (1)$$

However, the same convex set can be defined by different sets of desirable gambles [14], as shown by the following example [5].

Example 1 *Assume* $U_X = \{x_1, x_2, x_3, x_4\}$ *and a credal set given by a convex set with two probability distributions as extreme points:* $p_1 = (0, 0, 0.25, 0.75), p_2 = (0, 0, 0.5, 0.5)$. *This credal set can be defined by two sets of desirable gambles,* $\mathcal{D}_i(X) = \overline{\mathcal{R}_i(X)}, i = 1, 2$, *where*

$$\mathcal{R}_1(X) = \{f : f(x_3) + 3f(x_4) > 0, f(x_3) + f(x_4) > 0\}$$

$$\mathcal{R}_2(X) = \mathcal{R}_1(X) \cup \{f : f(x_3) = f(x_4) = 0, f(x_1) + f(x_2) > 0\}$$

It is immediate to show that $\mathcal{R}_1(X)$ *imposes some restrictions on the probabilities, which imply* $p(x_1) = p(x_2) = 0$. *So, the restrictions in* $\mathcal{R}_2(X)$ *are also trivially verified. Where is the difference between both sets of desirable gambles? As, we will see later they express different ways of calculating conditional information in the case of events of a lower prevision equal to 0. Desirable gambles contain information about how to calculate conditional information even for events of upper prevision equal to zero.*

Given a credal convex set $\mathcal{M}(X)$ we always can consider the least informative set of accepted gambles for which $\mathcal{M}(X)$ is obtained by equation (1). This is the set $\mathcal{D}(X)$ generated by gambles $\mathcal{R}(X)$, where

$$\mathcal{R}(X) = \{f : E_P(f) > 0, \forall P \in \mathcal{M}(X)\} \qquad (2)$$

A coherent set of desirable gambles, $\mathcal{R}(X)$, assigns a lower and an upper prevision for each real function $f : U_X \to I\!R$ defined by,

$$\begin{aligned} \underline{E}[f] &= \sup \{\alpha : f - \alpha \in \mathcal{R}(X)\} \\ \overline{E}[f] &= \inf \{\alpha : -f + \alpha \in \mathcal{R}(X)\} \end{aligned} \qquad (3)$$

If $B \subseteq U_X$, then the lower (upper) probability of B, $\underline{p}(B)(\overline{p}(B))$, is the lower (upper) prevision of the indicator function I_B of B.

2.1 Marginalization and Extension

If we have a coherent set of gambles $\mathcal{D}(X, Y)$ about the values of two variables, we define the marginalization of this set to one of its variables, for example X. First, a gamble f in U_X is identified with the gamble f' in $U_X \times U_Y$ with $f'(x, y) = f(x), \forall y \in U_Y$. In these conditions, the marginal of a set of $\mathcal{D}(X, Y)$ on X are the gambles $\mathcal{D}(X, Y)^{\downarrow X} = \mathcal{D}(X, Y) \cap \mathcal{L}(X)$ (i.e. the accepted gambles that only depends on the value of variable X).

Example 2 *Assume that* $U_X = \{0, 1\}, U_Y = \{a, b\}$ *and that* $\mathcal{D}(X, Y) = \overline{\{f\}}$ *where* $f(0, a) = 2, f(0, b) = 1, f(1, a) = -2, f(1, b) = -3$, *then* $\mathcal{D}(X, Y)^{\downarrow X} = \overline{\{f'\}}$ *where* $f'(0) = 2, f'(1) = -2$.

Extension is the reverse operator to marginalization. It transforms a coherent set of gambles defined for one variable $\mathcal{D}(X)$ into a coherent set relative to one more variable (X,Y). First, we are going to consider the concept of weak extension. The definition is simple. After identifying a gamble f about X with the gamble f' about (X,Y) as above, then its weak extension, $\mathcal{D}(X)^{\uparrow X,Y}$, is simply the natural extension on (X,Y) of $\mathcal{D}(X)$: all gambles generated by application of properties D2, D3, and D4 to gambles f' where $f \in \mathcal{D}(X)$. In the following, we will make an extensive use of this identification of a gamble defined for a variable with the corresponding gamble in a higher dimension. We even will denote both gambles with the same symbol.

If $\mathcal{D}(X,Y)$ and $\mathcal{D}(Y,Z)$ are coherent sets, then the natural extension of $\mathcal{D}(X,Y)^{\uparrow X,Y,Z} \cup \mathcal{D}(Y,Z)^{\uparrow X,Y,Z}$ will be denoted as $\mathcal{D}(X,Y) \oplus \mathcal{D}(Y,Z)$. Gambles in $\mathcal{D}(X,Y) \oplus \mathcal{D}(Y,Z)$ are defined on $U_X \times U_Y \times U_Z$. This operator first extend both coherent sets to a common frame, and then it takes the natural extension of the union. This natural extension is equal to $\mathcal{D}(X,Y) \cup \mathcal{D}(Y,Z)$ and all gambles $f = g + h$ where $g \in \mathcal{D}(X,Y)$ and $h \in \mathcal{D}(Y,Z)$.

It is immediate to show that $(\mathcal{D}(X,Y) \oplus \mathcal{D}(Y,Z))^{\downarrow X,Y} = \mathcal{D}(X,Y) \oplus \mathcal{D}(Y,Z)^{\downarrow Y}$. This property is fundamental for local computation in valuation based systems [9] and it will be used in the proofs of the rest of the paper.

The strong extension of a set of gambles $\mathcal{D}(X)$ to (X,Y) will be denoted as $\mathcal{D}(X)^{\Uparrow X,Y}$ and it is the natural extension of $\mathcal{R} = \{f_{y_0} : f \in \mathcal{D}_X, y_0 \in U_Y\}$ where f_{y_0} is the gamble defined on $U_X \times U_Y$ by $f_{y_0}(x,y_0) = f(x)$, and $f_{y_0}(x,y) = 0.0$, when $y \neq y_0$. This type of extension will be used to add variables in conditional sets. It verifies that $\left(\mathcal{D}(X)^{\Uparrow X,Y}\right)^{\downarrow X} = \mathcal{D}(X)$ and it is more informative than weak extension $\mathcal{D}(X)^{\uparrow X,Y}$.

We will denote as $\mathcal{D}(X,Y) \otimes \mathcal{D}(Y,Z)$ the natural extension of $\mathcal{D}(X,Y)^{\uparrow X,Y,Z} \cup \mathcal{D}(Y,Z)^{\Uparrow X,Y,Z}$. This operation is not symmetrical. The only difference with above one is that now we use strong extension in the second element. It is also verified that $(\mathcal{D}(X,Y) \otimes \mathcal{D}(Y,Z))^{\downarrow X,Y} = \mathcal{D}(X,Y) \otimes \mathcal{D}(Y,Z)^{\downarrow Y}$ and $(\mathcal{D}(X,Y) \otimes \mathcal{D}(Y,Z))^{\downarrow Y,Z} = \mathcal{D}(X,Y)^{\downarrow Y} \otimes \mathcal{D}(Y,Z)$.

We also have that $\mathcal{D}(X) \otimes \mathcal{D}(X,Y) = \mathcal{D}(X) \oplus \mathcal{D}(X,Y)$, as in this case we do not have an extension operator in the second element.

2.2 Conditioning

Conditioning has a very simple definition in terms of desirable gambles. The set of desirable conditional gambles conditioned to set B is $\mathcal{D}(X|B) = \{f \in \mathcal{L}(X) : f.I_B \in \mathcal{D}(X)\}$, where I_B is the indicator function of B.

Example 3 *Assume the situation of Example 1. Though, $\mathcal{D}_1(X)$ and $\mathcal{D}_2(X)$ have associated the same set of probability distributions, they represent different behavior under conditioning: $\mathcal{D}_1(X|\{x_1,x_2\})$ is the set of gambles f, with $f(x_1) \geq 0, f(x_2) \geq 0$, and $f(x_1) > 0$ or $f(x_2) > 0$, which has as associated convex set the set of all the probabilities with $P(\{x_1,x_2\}) = 1.0$; whereas $\mathcal{D}_2(X|\{x_1,x_2\})$ is the set of desirable gambles generated by $\{f : f(x_1) + f(x_2) > 0\}$, for which the associated convex set has a single probability distribution: $P(x_1) = P(x_2) = 0.5, P(x_3) = P(x_4) = 0.0$.*

If we know a conditional set of gambles $\mathcal{D}(X|B)$, and not the original set $\mathcal{D}(X)$, we can apply natural extension to $\mathcal{D}(X|B)$ to determine an unconditional set of accepted gambles associated to this conditional set: the least informative set $\mathcal{D}(X)$ such that its conditional set is $\mathcal{D}(X|B)$. It is not difficult to show that if $\mathcal{R} = \{f \in \mathcal{D}(X|B) : f(x) = 0 \text{ if } x \notin B\}$, then this natural extension is $\mathcal{D}(X) = \overline{\mathcal{R}}$.

If $\mathcal{B}$ is a finite partition of U_X, and we have a coherent set of gambles $\mathcal{D}(X|B_i)$ for every $B_i \in \mathcal{B}$, then the natural extension of this partition is the least informative coherent set $\mathcal{D}_{\mathcal{B}}(X)$ such that for every $B_i \in \mathcal{B}$ we have that $\mathcal{D}_{\mathcal{B}}(X|B_i) = \mathcal{D}(X|B_i)$. It is not difficult to show that if all $\mathcal{D}(X|B_i)$ are coherent, then this natural extension is given by the gambles $f \in \mathcal{L}(X)$ such that $f.I_{B_i} \in \mathcal{D}(X|B_i) \cup \{0\}, \forall B_i \in \mathcal{B}, f \neq 0$.

If $\mathcal{R}_i = \{f \in \mathcal{D}(X|B_i) : f(x) = 0 \text{ if } x \notin B_i\}$, then we have that $\mathcal{D}_{\mathcal{B}}(X) = \overline{\cup_{B_i \in B} \mathcal{R}_i}$.

2.3 Conditional Credal Sets

Now, we consider the specification of conditional information of one variable with respect to the values of the other one. We have two variables, X and Y, which are specified as marginal information about X and conditional information about Y for each one of the possible values of X.

A set of desirable gambles conditioned to the elements of X, $\mathcal{D}_v(Y|X)$, is a coherent set of gambles about Y, $\mathcal{D}_v(Y|X=x)$, for each possible value, x, of X. It represents our current attitude to accept gambles assuming that the value of X is equal to x. We make an assumption, $X = x$, and $\mathcal{D}_v(Y|X=x)$ contains the desired gambles for Y under it. A global $\mathcal{D}(X,Y)$ and a conditional $\mathcal{D}_v(Y|X)$ are compatible if and only if for all x, $\mathcal{D}_v(Y|X=x)$ is the conditioning of $\mathcal{D}(X,Y)$ given $X = x$ and marginalized to Y afterward; i.e. $\mathcal{D}_v(Y|X=x) = \mathcal{D}(X,Y|X=x)^{\downarrow Y}$.

If we have a family of coherent sets conditioned to the elements of X, $\mathcal{D}_v(Y|X)$, its natural extension $\mathcal{D}_v(X,Y)$ is the least informative coherent set compatible with this family of conditional sets. $\mathcal{D}_v(X,Y)$ is very easy to obtain: it is the set of gambles $f \in \mathcal{L}(X,Y)$ such that $f \neq 0$ and for every $x_0 \in U_X$, the gamble f_{x_0} in U_Y, given by $f_{x_0}(y) = f(x_0,y)$, is in $\mathcal{D}_v(Y|X=x_0) \cup 0$. The proof is based in showing that this set $\mathcal{D}_v(X,Y)$ gives rise to these conditional sets and that every other set $\mathcal{D}'(X,Y)$ with these conditional sets contains $\mathcal{D}_v(X,Y)$. The first part is based in applying the definitions. For the second part, assume f such that for every $x_0 \in U_X$, $f_{x_0} \in \mathcal{D}'(X,Y|X=x_0)^{\downarrow Y} \cup \{0\} = \mathcal{D}_v(Y|X=x_0) \cup \{0\}$. This means that $r_{x_0} \in \mathcal{D}'(X,Y|X=x_0) \cup \{0\}$ where r_{x_0} is defined on $U_X \times U_Y$ by $r_{x_0}(x,y) = f_{x_0}(y) = f(x_0,y)$. If $r_{x_0} \in \mathcal{D}'(X,Y|X=x_0)$ and $r_{x_0} \neq 0$, then r'_{x_0} given by $r'_{x_0}(x,y) = r_{x_0}(x,y) = f(x_0,y)$ if $x = x_0$ and 0.0, otherwise is in $\mathcal{D}'(X,Y)$. The result is obtained by applying property D4 and taking into account that $f = \sum_{r_{x_0} \neq 0} r'_{x_0}$.

If we have a marginal information $\mathcal{D}_m(X)$ about X and a conditional family of sets $\mathcal{D}_v(Y|X)$ of Y given X, then their natural extension will be the natural extension of $\mathcal{D}_m(X)$ and $D_v(X,Y)$, $\mathcal{D}_m(X) \oplus \mathcal{D}_v(X,Y)$, and it will be also denoted as $\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X)$. This natural extension is also equal to $\mathcal{D}_m(X) \otimes \mathcal{D}_v(X,Y)$ and will be denoted as $\mathcal{D}_m(X) \otimes \mathcal{D}_v(Y|X)$ too. As these sets are closed for properties D2, D3 and D4, then $\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X)$ will contain $\mathcal{D}_m(X) \cup D_v(X,Y)$ and all the gambles $f+g$ where $f \in \mathcal{D}_m(X)$, $g \in D_v(X,Y)$. If $\mathcal{D}_m(X)$ and $\mathcal{D}_v(Y|X)$ are coherent (they verify D1) then $\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X)$ is always coherent.

Theorem 1 *If $D_m(X)$ and $\mathcal{D}_v(Y|X)$ are coherent, then $(\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X))^{\downarrow X} = \mathcal{D}_m(X)$ and the conditional information $(\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X))(Y|X = x_0) = ((\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X))|X = x_0)^{\downarrow Y}$ is equal to $\mathcal{D}_v(Y|X=x_0)$.*

Proof.- First we have to prove that $(\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X))^{\downarrow X} = \mathcal{D}_m(X)$. Assume $f \in \mathcal{D}_m(X)$, then $f \in (\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X))$ and as f is constant on Y (it is defined only for variable X) we have that $f \in (\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X))^{\downarrow X}$. Reciprocally, assume that $f \in (\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X))^{\downarrow X}$. Then $f \in \mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X) = \mathcal{D}_m(X) \oplus \mathcal{D}_v(X,Y)$ and it is constant on the values of Y. We have three possibilities: $f \in D_m(X), f \in \mathcal{D}_v(X,Y)$, or $f = h + r$, with $h \in D_m(X), r \in \mathcal{D}_v(X,Y)$. In all of them we have to prove that $f \in D_m(X)$. In the first case is trivial. Assume that $f \in D_v(X,Y)$. We have that for every x_0, the gamble f_{x_0} defined as above is in $D_v(Y|X=x_0)$, but f_{x_0} is constant on Y and $D_v(Y|X=x_0)$ is coherent. Therefore $f_{x_0}(y) = f(x_0,y) > 0$. As f is constant on Y, then $f \in \mathcal{L}(X)$ and as $f > 0$, by property D2, it belongs to coherent set $D_m(X)$. Assume now that $f = h + r$, where $h \in \mathcal{D}_m(X)$, and $r \in \mathcal{D}_v(X,Y)$. As h and f are constant on Y, r will be constant on Y too. We can repeat the argument that we used for f in the case of $f \in D_v(X,Y)$, but now applied to r, showing that $r \in D_m(X)$. As $h, r \in D_m(X)$, we have that $f = h + r \in D_m(X)$.

Now, we prove $((\mathcal{D}_m(X) \oplus \mathcal{D}_v(X,Y))|X = x_0)^{\downarrow Y}$ is equal to $\mathcal{D}_v(Y|X=x_0)$. If $f \in \mathcal{D}_v(Y|X=x_0)$, then it is defined on the values of Y. The function $f'_{x_0}(x,y) = f(y)$, if $x = x_0$, and 0.0 otherwise, is in $\mathcal{D}_v(X,Y)$. Therefore $f \in \mathcal{D}_v(X,Y|X=x_0)$, and as conditioning is monotonic (if we add more gambles to a coherent set, then all conditional sets will be at least as informative as before adding them) we have that $f \in (\mathcal{D}_m(X) \oplus \mathcal{D}_v(X,Y))|X = x_0)$. As f is constant on X, we have that $f \in (\mathcal{D}_m(X) \oplus \mathcal{D}_v(X,Y))|X = x_0)^{\downarrow Y}$. Reciprocally, assume $f \in ((\mathcal{D}_m(X) \oplus \mathcal{D}_v(X,Y))|X = x_0)^{\downarrow Y}$, then f is constant on Y and $f.I_{X=x_0} \in \mathcal{D}_m(X) \oplus \mathcal{D}_v(X,Y)$. We are going to prove that $f.I_{X=x_0} \in \mathcal{D}_v(X,Y)$. We have three options $f.I_{X=x_0} \in D_m(X), f.I_{X=x_0} \in \mathcal{D}_v(X,Y)$, or $f.I_{X=x_0} = h + r$, with $h \in D_m(X), r \in \mathcal{D}_v(X,Y)$. In the third case is trivial. In the first case, if $f.I_{X=x_0} \in D_m(X)$, then it is constant on Y. As $f.I_{X=x_0}(x,y) = 0.0$, for $x \neq x_0$, and $D_m(X)$ is coherent, we have that $f.I_{X=x_0}(x_0,y) > 0$, and therefore $f.I_{X=x_0}(x_0,y) \in D_m(X)$. If $f.I_{X=x_0} = h + r$, with $h \in D_m(X), r \in \mathcal{D}_v(X,Y)$. If we fix a value $x \neq x_0$, we have that $f.I_{X=x_0}$ and h are constant on Y. Then $r = f.I_{X=x_0} - h$, will be constant too for $x \neq x_0$. By the definition of $\mathcal{D}_v(X,Y)$, this implies that $r(x,y) \geq 0$, for $x \neq x_0$. Therefore $h(x) \leq 0$, for $x \neq x_0$ and $h(x_0) > 0$ and $r(x_0,y) \neq 0, \forall y$. Let us modify h and r to h' and r' which are equal to h and r for $X = x_0$ and equal to 0.0 otherwise. We continue having $f.I_{X=x_0} = h' + r'$ with $r' \in \mathcal{D}_v(X,Y)$. As $h' \geq 0$ and $h' \neq 0$, we have $h' \in \mathcal{D}_v(X,Y)$, and as this set is coherent we have $f.I_{X=x_0} = h' + r' \in \mathcal{D}_v(X,Y)$. By the definition of conditioning, this implies that $f \in \mathcal{D}_v(X,Y|X=x_0)$ and as f is constant on Y, we have that $f \in \mathcal{D}_v(X,Y|X=x_0)^{\downarrow Y} = \mathcal{D}_v(Y|X=x_0)$. ■

Given a global set $\mathcal{D}(X,Y)$, we can calculate the marginal set $\mathcal{D}_m(X) = \mathcal{D}(X,Y)^{\downarrow X}$ and the family of conditional gambles $\mathcal{D}_v(Y|X)$, but it we compute $\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X)$ we will obtain a set of desirable gambles, which in general is less informative (it contains less gambles) than the original one $\mathcal{D}(X,Y)$: $\mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X) \subseteq \mathcal{D}(X,Y)$. A completely analogous situation occurs with credal sets [5].

Another important property of conditioning is that it commutes with marginalization.

Theorem 2 *If we have three variables X, Y, Z, and a coherent set $\mathcal{D}(X,Y,Z)$, then we have that for all $y \in U_Y$, $(\mathcal{D}_v(X,Y,Z)^{\downarrow X,Y}|Y = y) = (\mathcal{D}_v(X,Z|Y = y))^{\downarrow X}$*

Proof.-

Assume $f \in (\mathcal{D}_v(X,Y,Z)^{\downarrow X,Y}|Y = y)$. Then we have that $f \in (\mathcal{D}(X,Y,Z)^{\downarrow X,Y}|Y = y)^{\downarrow X}$. This implies that f is constant on Y and Z and $f.I_{Y=y} \in \mathcal{D}(X,Y,Z)^{\downarrow X,Y}$. As marginalization is an intersection, $f.I_{Y=y} \in \mathcal{D}(X,Y,Z)$. By definition, $f \in \mathcal{D}(X,Y,Z|Y = y)$. Now, as f is constant on Y, we have that $f \in \mathcal{D}(X,Y,Z|Y = y)^{\downarrow X,Z} = \mathcal{D}_v(X,Z|Y = y)$. As f is constant on Z we have that $f \in \mathcal{D}_v(X,Z|Y = y)^{\downarrow X}$.

Reciprocally, if $f \in (\mathcal{D}_v(X,Z|Y = y))^{\downarrow X}$, then this function is defined only for variable X (it is constant for the other variables) and it belongs to $\mathcal{D}_v(X,Z|Y = y)$. This implies that $f.I_{Y=y} \in \mathcal{D}(X,Y,Z)$. As this function is constant on Z we have that $f.I_{Y=y} \in \mathcal{D}(X,Y,Z)^{\downarrow X,Y}$. By definition, we have that $f \in (\mathcal{D}(X,Y,Z)^{\downarrow X,Y}|Y = y)$, and as f is constant on Y $f \in (\mathcal{D}(X,Y,Z)^{\downarrow X,Y}|Y = y)^{\downarrow Y} = (\mathcal{D}_v(X,Y,Z)^{\downarrow X,Y}|Y = y)$. ∎

As a consequence, if we have a global set $\mathcal{D}(X,Y,Z)$, it makes sense to talk about $D_v(X|Y)$ and it is equal to the family of coherent sets gambles on X indexed by the values of Y, and that is obtained either by marginalizing $\mathcal{D}(X,Y,Z)$ on (X,Y) and then conditioning on the values of Y or by conditioning first and marginalizing afterward.

2.4 Epistemic Irrelevance

The definition given here of epistemic irrelevance is a bit stronger than the one considered in [12, 3]. In the case of unconditional independence it corresponds to the irrelevant natural extension as was defined in [1]. In the case of conditional independence to the values of a variable, we add a requirement that the global information is given as a marginal information on this variable and a conditional information given the values of this variable.

Given three variables, X, Y, and Z, we say that X is irrelevant to Y given Z if and only if the global set of gambles $\mathcal{D}(X,Y,Z)$ can be expressed as $\mathcal{D}_m(X,Z) \otimes \mathcal{D}_v(Y|Z)$, where $\mathcal{D}_m(X,Z)$ is the marginal set on variables (X,Z) and $\mathcal{D}_v(Y|Z)$ is a family of gambles about Y for each value of Z.

This definition implies that for every $x \in U_X, z \in U_Z$, the set $\mathcal{D}_v(Y|Z = z, X = x)$ is equal to the set $\mathcal{D}_v(Y|Z = z)$. But it also implies that $\mathcal{D}(X,Y,Z) = \mathcal{D}_m(X,Z) \oplus \mathcal{D}_v(Y|Z,X)$. This is easy to prove taking into account that the equality $\mathcal{D}_v(Y|Z = z, X = x) = \mathcal{D}_v(Y|Z = z), \forall x, z$ is equivalent to $\mathcal{D}_v(X,Y,Z) = \mathcal{D}_v(Y,Z)^{\Uparrow X,Y,Z}$, where $\mathcal{D}_v(X,Y,Z)$ is the natural extension of the family of coherent sets $\mathcal{D}_v(Y|Z,X)$ and $\mathcal{D}_v(Y,Z)$ the natural extension of $D_v(Y|Z)$.

3 Semigraphoid Axioms

A relation of independence $I(X,Y|Z)$ between variables verifies the semigraphoid axioms if and only if it fulfills the following properties [7, 4, 10]

Symmetry.- $I(X,Y|Z) \Rightarrow I(Y,X|Z)$

Redundancy.- $I(X,Y|X)$

Decomposition.- $I(X,(W,Y)|Z) \Rightarrow I(X,Y|Z)$

Weak Union.- $I(X,(W,Y)|Z) \Rightarrow I(X,Y|(W,Z))$

Contraction.- $I(X,Y|Z)$ and $I(X,W|(Y,Z)) \Rightarrow I(X,(W,Y)|Z)$

The relation $I(X,Y|Z)$ is to be read as X is independent (or irrelevant) to Y given Z.

When the following axiom is verified:

Intersection.- $I(X,Y|(Z,W))$ and $I(W,Y|(X,Z)) \Rightarrow I((X,W),Y|Z)$

then the set of independences is said to be a graphoid.

As some of the definitions of independence are non symmetrical, we have to introduce the reverse versions of these properties, following Cozman and Walley [3]. The axioms as they are written will be called the direct versions. If the variables are in reverse order, we call them the reverse versions. Then, the reverse contraction is

Reverse Contraction.- $I(Y,X|Z)$ and $I(W,X|(Y,Z)) \Rightarrow I((W,Y),X|Z)$

Epistemic irrelevance does not verify the symmetry property. The examples in [3] are also valid here. They are given for credal sets. To translate them to sets of gambles we can use the transformation in equation (2). Though redundancy can look trivial, it is not verified in the direct version. This property implies that $\mathcal{D}(X,Y)$ can be represented as $\mathcal{D}_m(X) \otimes \mathcal{D}_v(Y|X) = \mathcal{D}_m(X) \oplus \mathcal{D}_v(Y|X)$, and, as it was said in Section 2.3, this is not always possible. All the graphoid properties except symmetry are verified (at least in one of its versions, direct or reverse) as the following theorem proves. The most unusual property is the intersection axiom for general sets of desirable gambles, without any additional condition. But this is not surprising due to the handling of zero probabilities in this theory.

Theorem 3 *Epistemic conditional irrelevance verifies reverse redundancy, reverse and direct decomposition, reverse weak union, direct and reverse contraction and direct intersection.*

Proof.- Reverse redundancy says that $I(X,Y|Y)$ and this is equivalent to represent the global $\mathcal{D}(X,Y) = \mathcal{D}_m(X,Y) \otimes \mathcal{D}_v(Y|Y)$, and this is always possible if $\mathcal{D}_m(X,Y) = \mathcal{D}(X,Y)$ and $D_v(Y|Y)$ is the trivial set of gambles (it only contains gambles $f \neq 0$, $f \geq 0$.

Direct decomposition says that $I(X,(W,Y)|Z) \Rightarrow I(X,Y|Z)$. $I(X,(W,Y)|Z)$ is equivalent to $\mathcal{D}(X,W,Y,Z) = \mathcal{D}_m(X,Z) \otimes \mathcal{D}_v(W,Y|Z)$. Then we can prove that $\mathcal{D}(X,Y,Z) = \mathcal{D}(X,W,Y,Z)^{\downarrow X,Y,Z} = \mathcal{D}_m(X,Z) \otimes \mathcal{D}_v(W,Y|Z)^{\downarrow Y,Z} = \mathcal{D}_m(X,Z) \otimes \mathcal{D}_v(Y|Z)$, where $\mathcal{D}_v(W,Y|Z)^{\downarrow Y,Z}$ is calculated by taking the marginal on Y of each conditional set of gambles on $\mathcal{D}_v(W,Y|Z=z)$ for the different values $z \in U_Z$.

For reverse decomposition, we have to prove $I((W,Y),X|Z) \Rightarrow I(Y,X|Z)$. If $I((W,Y),X|Z)$, then we have that $\mathcal{D}(X,W,Y,Z) = \mathcal{D}(W,Y,Z) \otimes \mathcal{D}_v(X|Z)$. From here, and marginalizing on (X,Y,Z) we can obtain $\mathcal{D}(X,Y,Z) = \mathcal{D}(X,W,Y,Z)^{\downarrow X,Y,Z} = \mathcal{D}(W,Y,Z)^{\downarrow Y,Z} \otimes \mathcal{D}_v(X|Z) = \mathcal{D}(Y,Z) \otimes \mathcal{D}_v(X|Z)$, which is equivalent to the desired property $I(Y,X|Z)$.

For reverse weak union, we have to prove $I((W,Y),X|Z) \Rightarrow I(Y,X|(W,Z))$. If $I((W,Y),X|Z)$, then we have that $\mathcal{D}(X,W,Y,Z) = \mathcal{D}(W,Y,Z) \otimes \mathcal{D}_v(X|Z)$. Let us consider $\mathcal{D}_v(X|Z,W=w) = \mathcal{D}_v(X|Z), \forall w \in U_W$, then we have $\mathcal{D}(X,W,Y,Z) = \mathcal{D}(W,Y,Z) \otimes \mathcal{D}_v(X|Z,W)$ and therefore, the desired independence $I(Y,X|(W,Z))$.

For contraction we have to prove $I(X,Y|Z)$ and $I(X,W|(Y,Z)) \Rightarrow I(X,(W,Y)|Z)$. If $I(X,W|(Y,Z))$, then $\mathcal{D}(X,Y,W,Z) = \mathcal{D}(X,Y,Z) \otimes \mathcal{D}_v(W|Y,Z)$. As $I(X,Y|Z)$ we have that $\mathcal{D}(X,Y,Z) = \mathcal{D}(X,Z) \otimes \mathcal{D}_v(Y|Z)$. The desired independence is obtained if we take into account that for every z $\mathcal{D}_v(Y|Z=z) \otimes \mathcal{D}_v(W|Y,Z=z)$ is a set of desirable gambles about (W,Y) which is equal to $D_v(W,Y|Z=z)$.

For reverse contraction we have to prove $I(Y,X|Z)$ and $I(W,X|(Y,Z)) \Rightarrow I((W,Y),X|Z)$. $I(W,X|(Y,Z))$ implies that $\mathcal{D}(X,Y,W,Z) = \mathcal{D}(Y,W,Z) \otimes \mathcal{D}_v(X|Y,Z)$. As $I(Y,X|Z)$, $\mathcal{D}_v(X|Y,Z)$ does not depend of the value of Y, with which we obtain the desired equality $\mathcal{D}(X,Y,W,Z) = \mathcal{D}(Y,W,Z) \otimes \mathcal{D}_v(X|Z)$.

For intersection, we have to prove $I(X,Y|(Z,W))$ and $I(W,Y|(X,Z)) \Rightarrow I((X,W),Y|Z)$. If we assume $I(X,Y|(Z,W))$ then $\mathcal{D}(X,Y,W,Z) = \mathcal{D}(X,Z,W) \otimes D_v(Y|Z,W)$. As $I(W,Y|(X,Z))$, we also have $\mathcal{D}(X,Y,W,Z) = \mathcal{D}(X,Z,W) \otimes D_v(Y|Z,X)$. As $\mathcal{D}(X,Z,W)$ does not depend on Y, we have that for every $z,x,w, D_v(Y|Z=z,W=w) = D_v(Y|Z=z,X=x)$. From here, we have that $D_v(Y|Z=z,W=w)$ does not depend of w, and therefore, the desired decomposition: $\mathcal{D}(X,Y,W,Z) = \mathcal{D}(X,Z,W) \otimes D_v(Y|Z)$.

■

4 The Asymmetrical D-Separation Criterion

Assume a set of variables $(X_i)_{i \in I}$ and a directed acyclic graph [7] G such that there is a node of the graph associated to each variable X_i.

Here we modify the classical D-separation criterion to represent independences [8, 7]. As usual, we shall consider paths between pair of variables in which arcs can be used in any direction (direct or reverse). If $(X_k)_{k \in K}$ is a vector of observed variables and we have a path between variables X_i and X_j, we say that this path is blocked under this set of observations if and only if at least one of the two following conditions happens:

1. There is a variable in the path X_l through which the path passes with converging arrows and this variable and none of its descendants is in $(X_k)_{k \in K}$.

2. There is a variable in the path X_l with $l \in K$, such that the path does not leave it with an incoming arc. This variable can be X_i or X_j.

In the second condition, if observed variable X_l is the last variable in the path, i.e. X_j, then the path is blocked, as the path does not leave variable X_l. If X_l is the first variable or an intermediate variable, then we have to consider the following variable in the path, X_n. It there is a link from X_l to X_n the path is blocked in X_l. If the link is from X_n to X_l, then the path is not blocked.

This criterion will be called AD-separation. In classical D-separation, the path can leave the blocking variable through an arc which is used in the reverse direction. Now, this case is not included and makes the new criterion non-symmetrical.

Given a graph, and $K, L, M \subseteq I$, we say that the graph induces the independence

$I_G((X_k)_{k\in K},(X_l)_{l\in L}|(X_m)_{m\in M})$ if and only if every path from a variable $X_k, k \in K$, to a variable $X_l, l \in L$, is blocked by a variable under observations $(X_m)_{m\in M}$.

To simplify the notation we shall write $I_G((X_k)_{k\in K},(X_l)_{l\in L}|X_{m\in M})$, as $I_G(K,L|M)$. All the properties verified by epistemic irrelevance are verified by the models of independence represented by graphs and asymmetrical D-separation.

Theorem 4 *If G is a directed acyclic graph, then I_G verifies reverse redundancy, reverse and direct decomposition, reverse weak union, direct and reverse contraction, and direct intersection.*

Proof.-

Reverse redundancy says that $I_G(J,K|K)$. This is immediate, as every path to a variable $X_k, k \in K$ is blocked in the node X_k.

Direct and reverse decomposition are immediate from the definition of I_G.

For reverse weak union, we have to prove $I_G((J \cup K),L|M) \Rightarrow I_G(J,L|K \cup M)$. Now consider a path from a node X_i to a node $X_l, l \in L$. The only way that this path is blocked by $(X_m)_{m\in M}$ and not by $(X_m)_{m\in K\cup M}$ is that it is blocked in a node X_k with a descendant $X_{k'}$ with $k' \in K$. As $I_G((J\cup K),L|M)$ the path going from $X_{k'}$ to X_k using reverse arcs till X_k, and then using the path going from X_k to X_l should be blocked by $(X_m)_{m\in M}$. Then, there should be another node blocking the original path different from X_k and closer to X_l. As this argument can not be repeated indefinitely, the path is blocked by $(X_m)_{m\in K\cup M}$. Now, the condition for $I_G(J,L|K\cup M)$ is proved.

For contraction we have to prove $I(J,K|L)$ and $I(J,M|K\cup L) \Rightarrow I(J,K\cup M|L)$. Paths from a variable $X_j, j \in J$, to a variable $X_k, k \in K$, are blocked by $(X_l)_{l\in L}$. We only have to prove the blocking for paths from $X_j, j \in J$, to a variable $X_m, m \in M$. This path is blocked by $(X_l)_{l\in L\cup K}$. If it is blocked by a variable $X_l, l \in L$, then there is no problem: the same variable blocks the path under $(X_l)_{l\in L}$. If it is variable $X_k, k \in K$. Then the part of the path going from X_j to X_k should be blocked by $(X_l)_{l\in L}$. With which the complete path is also blocked.

For reverse contraction we have to prove $I(K,J|L)$ and $I(M,J|K\cup L) \Rightarrow I(K\cup M,J|L)$. The proof is completely analogous to the one for direct contraction.

For intersection, we have to prove $I_G(J,K|L\cup M)$ and $I_G(L,K|J\cup M) \Rightarrow I_G(J\cup L,K|M)$. Assume a path between a variable X_l and a variable $X_k, k \in K$. If it was blocked by $(X_m)_{m\in L\cup M}$ or by $(X_m)_{m\in J\cup M}$ and not by $(X_m)_{m\in M}$, then it is blocked in a variable X_i, with $i \in L$ or $i \in J$. Assume without loss of generality that $i \in L$. As $I_G(L,K|J\cup M)$, then the path from X_i to X_k, should be again blocked by $(X_m)_{m\in J\cup M}$ in a node closer to X_k. We can then repeat the same reasoning. As this process should be finite, we can conclude the path must be blocked by observations $(X_m)_{m\in M}$. And this finally proves $I_G(J\cup L,K|M)$.

■

A minimal I-map of an epistemic irrelevance model is a graph such that

1. If $I_G(J,K|L)$, then

$$\mathcal{D}_v((X_k)_{k\in K}|(X_j)_{j\in J\cup L}) = \mathcal{D}_v((X_k)_{k\in K}|(X_j)_{j\in L})$$

2. $ND(X_i) - \Pi(X_i)$ is epistemic irrelevant to X_i given $\Pi(X_i)$, where $ND(X_i)$ is the set of non descendant variables of X_i.

3. If we remove an arc, then some of the above conditions is not verified.

Observe as AD-separation only implies the weaker version of epistemic irrelevance considered in [3]. If we consider only the strong version of irrelevance, then AD-separation should be made much weaker.

A first difference with probabilistic I-maps by using the classical D-separation criterion is that, in that case, any ordering of the variables gives rise to a minimal I-map. Now, this is not always the case. Given a total order in the set of variables, a directed acyclic graph is compatible with that order when for every pair of variables X_i and X_j such that X_i is a parent of X_j, then X_i comes before than X_j in the order ($X_i \leq X_j$). We can have two dependent variables X and Y and a directed graph with a link from X to Y which is a minimal I-map, but not the graph with a link from Y to X, i.e. we can have $\mathcal{D}(X,Y) = \mathcal{D}_m(X)\otimes\mathcal{D}_v(Y|X)$, but not the decomposition $\mathcal{D}(X,Y) = \mathcal{D}_m(Y)\otimes\mathcal{D}_v(X|Y)$. If we consider the order Y, X then there is not a minimal I-map in which X is not an ascendant of Y. This can help to determine the correct direction of arcs. With two variables and probabilistic independences, this is never possible.

Another difference with probabilistic I-maps is that due to the verification of intersection, if we have an order of the variables and a minimal I-map compatible with it, then this is the only minimal I-map compatible with the order. This was guaranteed in the probabilistic case only with positive probabilities.

5 Conclusions

In this paper we have studied the main properties of epistemic irrelevance for sets of desirable gambles and the representation of a family of irrelevance relationships by means of a directed graph.

In the future, we plan to study more powerful representations of conditional independence relationships as chain graphs [11]. In these graphs, we have directed and undirected links. Undirected links could be used to represent dependences that are not expressed in a conditional way. If we have a situation with three variables X, Y, Z and we are given the independence, X is irrelevant to Z given Y, then there is no directed graph representing this independence. However, we could represent it with a chain graph with an undirected link between X and Y and a directed link from Y to Z.

Another topic of future study is the computational implications of this definition, following Cozman [2]. In particular, whether the introduction of decomposition of global information in the definition of epistemic irrelevance and the new criterion for representing independences can change some of the developments given in [2].

Acknowledgments

I want to thank two anonymous referees for useful and encouraging comments that have made possible to produce this paper. The initial draft contained a lot of imprecisions and some errors. Reviewers pointed out them or raised the right questions that allowed me to discover the problems in the initial draft. Part of the content of this paper is due to them.

This paper has been written with support of CICYT under project TIC97-1135-C04-01.

References

[1] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5:165–181, 2000.

[2] F.G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[3] F.G. Cozman and P. Walley. Graphoid properties of epistemic irrelevance and independence. Unpublished manuscript, 2000.

[4] A.P. Dawid. Conditional independence. In S. Kotz and C.B. Read amd D.L. Banks, editors, *Encyclopedia of Statistical Sciences, Update Volume 2*, pages 146–153. Wiley, New York, 1999.

[5] S. Moral and A. Cano. Strong conditional independence for credal sets. Submitted to Annals of Mathematics and Artificial Intelligence, 2000.

[6] S. Moral and N. Wilson. Revision rules for convex sets of probabilities. In G. Coletti, D. Dubois, and R. Scozzafava, editors, *Mathematical Models for Handling Partial Knowledge in Artificial Intelligence*, pages 113–128. Plenum Press, 1995.

[7] J. Pearl. *Probabilistic Reasoning with Intelligent Systems.* Morgan & Kaufman, San Mateo, 1988.

[8] J. Pearl and T. Verma. The logic of representing dependencies by directed graphs. In *Proceedings of the AAAI'87 Conference*, pages 347–379. AAAI/MIT Press, 1987.

[9] G. Shafer and P.P. Shenoy. Local computation in hypertrees. Working Paper N. 201, School of Business, University of Kansas, Lawrence, 1988.

[10] M. Studeny. Semigraphoids and structures of probabilistic conditional independence. *Annals of Mathematics and Artificial Intelligence*, 21:71–98, 1997.

[11] Milan Studeny. On separation criterion and recovery algorithm for chain graphs. In *Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence (UAI–96)*, pages 509–516, Portland, Oregon, 1996.

[12] P. Vicig. Epistemic independence for imprecise probabilities. *International Journal of Approximate Reasoning*, 24:235–250, 2000.

[13] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

[14] P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.

[15] N. Wilson and S. Moral. A logical view of probability. In A. Cohn, editor, *Proceedings of the Eleventh European Conference on Artificial Intelligence (ECAI'94)*, pages 386–390, London, 1994. Wiley.