

Graphical Models for Conditional Independence Structures

B. Vantaggi

Dip. Metodi e Modelli Matematici
Univ. “La Sapienza”, Rome, Italy
vantaggi@dmmm.uniroma1.it

Abstract

In this paper we study conditional independence structures arising from conditional probabilities and lower conditional probabilities. Such models are based on notions of stochastic independence apt to manage also those situations where zero evaluations on *possible* events are present: this is particularly crucial for lower probability. The “graphoid” properties of such models are investigated, and the representation problem of conditional independence structures is dealt with by generalizing classical separation criteria for undirected and directed acyclic graphs.

Keywords. Graphical models, conditional independence, lower probability, separation criteria.

1 Introduction

Graphs are widely used in probability, statistics and artificial intelligence to represent conditional independence structures induced by uncertainty measures ([9], [10], [16], [17], [18], [22]). Therefore graphical models are based on conditional independence, but there is generally no agreement on the choice of the formal relevant definition, because sometimes the intuitive meaning is not caught. Almost all definitions are generalizations of the classical definition of stochastic independence, and so their framework is the Kolmogorovian approach to conditional probability.

It is well known that such formalization of stochastic independence presents some counterintuitive aspects: for example, an event A of zero probability is stochastically independent of itself, while it is natural to require that *every* event is dependent on itself. Essentially, critical situations arise when 0 (or 1) probabilities on *possible* events are involved. To overcome such controversial situations other definitions of stochastic independence have been introduced [4], [5] in the most general framework of conditional probability as

given by de Finetti [12], Krauss [15], Dubins [13] (see Section 2). This approach to conditional probability allows to manage zero probability (while in the usual approach the classic definition of probability of a conditional event $E|H$ does not make sense if $P(H)$ is zero) and *direct* partial assessments through the concept of coherence, and it gives the possibility of extending any coherent assessment (for an overview, see for example [6]). The formalization of *stochastic independence* between two possible events A and B given in [5] implies *logical independence* of A and B .

In Section 3 the main properties of this definition are recalled also in the case of random variables (see [20]). In particular, the attention is focused on checking graphoid properties (see [17]).

In [7] the authors extend stochastic independence definition of two events to more general uncertainty function: lower (upper) conditional probabilities, where the problem of 0 (or 1) values is even more crucial.

In this paper the definition given in [7] is extended to conditional independence and it is studied the case of random variables. Moreover, the closure of such independence models with respect to graphoid properties is checked. It is proved that the two kinds of conditional independence models – induced by lower probability and probability – are closed with respect to the following properties: decomposition and its reverse, weak union, contraction and its reverse, intersection and its reverse. Notice that such structures are not closed with respect to symmetry.

One of the aims of this paper is to represent graphically these conditional independence structures. The classical separation criteria for undirected graphs [16] and the d-separation [17] for directed graphs are not completely apt to represent our conditional independence models (as we have shown in [19] by some examples). So in Section 6 a generalization of these criteria is presented. However, both these “generalized criteria” induce graphoid structures, so they are

not apt to describe models not (necessarily) closed with respect to symmetry (called *a-graphoid*): so we introduce another criterion able to represent not symmetric statements.

2 Conditional probability

Let \mathcal{A} be a finite Boolean algebra and denote $\mathcal{A}^0 = \mathcal{A} \setminus \{\emptyset\}$, where \emptyset is the *impossible* event.

Definition 1 *Given a Boolean algebra \mathcal{A} , a conditional probability on $\mathcal{A} \times \mathcal{A}^0$ is a function $P(\cdot|\cdot)$ into $[0, 1]$, which satisfies the following conditions:*

- (i) $P(\cdot|H)$ is a finitely additive probability on \mathcal{A} for any $H \in \mathcal{A}^0$
- (ii) $P(H|H) = 1$ for every $H \in \mathcal{A}^0$
- (iii) $P(E \wedge A|H) = P(E|H)P(A|E \wedge H)$, whenever $E, A \in \mathcal{A}$ and $H, E \wedge H \in \mathcal{A}^0$

Note that (iii) reduces, when $H = \Omega$ (where Ω is the *certain* event), to classic “chain rule” for probability

$$P(E \wedge A) = P(E)P(A|E).$$

In the case $P_0(\cdot) = P(\cdot|\Omega)$ is strictly positive on \mathcal{A}^0 , any conditional probability can be derived as a ratio by this unique “unconditional” probability P_0 . As proved in [15] in all other cases to set a similar representation we need to resort to a finite family $\mathcal{P} = \{P_0, \dots, P_k\}$ of unconditional probabilities:

- every P_α is defined on a proper subalgebra (taking $\mathcal{A}_0 = \mathcal{A}$) $\mathcal{A}_\alpha = \{E \in \mathcal{A}_{\alpha-1} : P_{\alpha-1}(E) = 0\}$;
- for each event $B \in \mathcal{A}^0$ there exists an unique α such that $P_\alpha(B) > 0$ and for every conditional event $E|H$ it holds $P(E|H) = \frac{P_\alpha(E \wedge H)}{P_\alpha(H)}$ with $P_\alpha(H) > 0$.

The class of probabilities $\mathcal{P} = \{P_0, \dots, P_k\}$ is said to *agree* with the conditional probability $P(\cdot|\cdot)$.

Such theory of conditional probability allows to handle also *partial* probability assessment on arbitrary set of conditional events $\mathcal{F} = \{E_1|H_1, \dots, E_n|H_n\}$ through the coherence: an assessment is coherent if it is the restriction of a conditional probability defined on $\mathcal{A} \times \mathcal{A}^0$, where \mathcal{A} is the algebra generated by $\{E_1, H_1, \dots, E_n, H_n\}$. A characterization of coherence has given in [2]:

Theorem 1 *Let \mathcal{F} be an arbitrary finite family of conditional events and \mathcal{A}_0 denotes the set of atoms C_r generated by the events $E_1, H_1, \dots, E_n, H_n$. For a real function P on \mathcal{F} the following two statements are equivalent:*

- (i) P is a coherent conditional probability on \mathcal{F} ;
- (ii) there exists (at least) a class of unconditional probabilities $\{P_0, P_1, \dots, P_k\}$, with P_0 on \mathcal{A}_0 and P_α

($\alpha > 0$) being defined on $\mathcal{A}_\alpha = \{E \in \mathcal{A}_{\alpha-1} : P_{\alpha-1}(E) = 0\}$, such that for any $E_i|H_i \in \mathcal{F}$ there is a unique P_α , with $P_\alpha(H_i) > 0$, and

$$P(E_i|H_i) = \frac{\sum_{C_r \subseteq E_i \wedge H_i} P_\alpha(C_r)}{\sum_{C_r \subseteq H_i} P_\alpha(C_r)}.$$

The class of probabilities $\mathcal{P} = \{P_0, \dots, P_k\}$ agreeing with the given coherent assessment P is not unique. But, fixed one class $\mathcal{P} = \{P_0, \dots, P_k\}$ for each event H there is a unique α such that $P_\alpha(H) > 0$ and α is said *zero-layer* of H according to \mathcal{P} , and it is denoted by the symbol $\circ(H)$. In particular, for every probability we have $\circ(\Omega) = 0$, while we define $\circ(\emptyset) = \infty$. The *zero-layer* of a conditional event $E|H$ is defined (see [5]) as

$$\circ(E|H) = \circ(E \wedge H) - \circ(H).$$

The crucial role of zero-layers is recalled in Section 3.

3 Logical independence

In the sequel, a *possible* event denotes an event different from \emptyset and Ω . Two distinct non-trivial partitions \mathcal{E}_1 and \mathcal{E}_2 of Ω are logically independent if the “finer” partition \mathcal{E} (called also *set of atoms*) generated by them, coincides with the set of all possible logical products between the events of \mathcal{E}_1 and \mathcal{E}_2 , i.e.

$$\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2 = \{C = C_1 \wedge C_2 \neq \emptyset : C_1 \in \mathcal{E}_1 ; C_2 \in \mathcal{E}_2\}.$$

Hence, in such case the cardinality $|\mathcal{E}|$ of \mathcal{E} is equal to $|\mathcal{E}_1| \cdot |\mathcal{E}_2|$. A logical constraint exists between two partitions if they are not logical independent, i.e. some logical products of the kind $C_1 \wedge C_2$ is not possible. In particular, the events A and B are logical independent if the partitions $\mathcal{E}_1 = \{A, A^c\}$ and $\mathcal{E}_2 = \{B, B^c\}$ are logical independent, so the set of atoms is $\{A \wedge B, A \wedge B^c, A^c \wedge B, A^c \wedge B^c\}$.

Analogously, the partitions $\mathcal{E}_1, \dots, \mathcal{E}_n$ are logically independent if the set of atoms \mathcal{E} generated by \mathcal{E}_i 's is obtained as logical product $\mathcal{E} = \mathcal{E}_1 \times \dots \times \mathcal{E}_n$, i.e. $C_1 \wedge \dots \wedge C_n \neq \emptyset$ with $C_i \in \mathcal{E}_i$ ($i = 1, \dots, n$).

Obviously, if n partitions are logically independent, then arbitrary subsets of these partitions are logically independent too. However, n partitions $\mathcal{E}_1, \dots, \mathcal{E}_n$ are not necessarily logically independent, even if every subset of $n - 1$ partitions are logically independent; it follows that there is some logical constraint of the kind $C_1 \wedge \dots \wedge C_n = \emptyset$, with $C_i \in \mathcal{E}_i$. For example, suppose $\mathcal{E}_1 = \{A, A^c\}$, $\mathcal{E}_2 = \{B, B^c\}$ and $\mathcal{E}_3 = \{C, C^c\}$ are three distinct partitions of Ω with $A \wedge B \wedge C = \emptyset$. All

the couples of these partitions are logically independent, but the partition \mathcal{E}_1 is not logically independent from the partition $\mathcal{E}_2 \times \mathcal{E}_3$ (i.e. generated by $\{\mathcal{E}_2, \mathcal{E}_3\}$). The same conclusion is reached replacing \mathcal{E}_1 by \mathcal{E}_2 or \mathcal{E}_3 .

Given n partitions and a logical constraint among them, it is possible to find the minimal subset $\{\mathcal{E}_1, \dots, \mathcal{E}_k\}$ of partitions generating such constraint: it means that $\mathcal{E}_1, \dots, \mathcal{E}_k$ are not logical independent, i.e. there exists at least a combination of atoms, with $C_i \in \mathcal{E}_i$, such that $C_1 \wedge \dots \wedge C_k = \emptyset$, while for all $j = 1, \dots, k$ we have $C_1 \wedge \dots \wedge C_{j-1} \wedge C_{j+1} \wedge \dots \wedge C_k \neq \emptyset$. This does not imply that all the subsets of the k partitions are logically independent, because there could be another different logical constraint involving some subset.

We will say that such set of partitions $\{\mathcal{E}_1, \dots, \mathcal{E}_k\}$ is the *minimal set* generating the given logical constraint.

4 Conditional independence for probabilities

It is well known that the classical definition of stochastic independence of two events

$$P(A \wedge B) = P(A)P(B) \quad (1)$$

gives rise to counter-intuitive situations when one of the events has probability 0 or 1. For instance an event A with $P(A) = 0$ is stochastically independent of itself, while it is natural (due to the intuitive meaning of independence) to require for any event to be dependent on itself. Other classical formulations are

$$P(A|B) = P(A) \quad (2)$$

and

$$P(A|B) = P(A|B^c), \quad (3)$$

that are equivalent to (1) for events such that the probability of B is different from 0 and 1, but in that “extreme” cases (without positivity assumption) they may even lack meaning in the Kolmogorovian approach to conditional probability.

Anyway, also considering the stronger formulation (3) in the more general framework of de Finetti [12] some critical situations continue to exist.

Example 1 Consider two incompatible possible events A and B , i.e. $A \wedge B = \emptyset$. Let P be a conditional probability such that $P(A|B^c) = 0$. Obviously condition (3) holds, since $P(A|B)$ must be 0 because of the given logical relation. Note that the previous consideration is valid for $P(B) \in [0, 1]$. Therefore, according to all the classical formulations

A and B are stochastically independent, although they are logically dependent.

For this reason other different definitions of stochastic independence have been proposed, we consider that given in [5] in the most general framework of coherent probability assessment and extended to conditional independence in [20]. In the sequel, to avoid cumbersome notation, the conjunction symbol \wedge among events is omitted.

Definition 2 Given a coherent conditional probability P , defined on a family \mathcal{F} containing $\mathcal{D} = \{A|BC, A|B^cC, A^c|BC, A^c|B^cC\}$, A is conditionally independent of B given C with respect to P if both the following conditions hold:

$$(i) P(A|BC) = P(A|B^cC);$$

$$(ii) \text{ there exists a class } \{P_\alpha\} \text{ of probabilities agreeing with the restriction of } P \text{ to the family } \mathcal{D}, \text{ such that } \circ(A|BC) = \circ(A|B^cC) \text{ and } \circ(A^c|BC) = \circ(A^c|B^cC).$$

Remark 1 Even if condition (ii) may give rise to different zero-layers (corresponding to different agreeing classes), nevertheless what is essential is that they satisfy the two equalities corresponding to the conditioning event of \mathcal{D} .

To distinguish this notion of conditional independence from the classic one we call it cs-independence and denote by the symbol \perp_{cs} .

The definition of cs-independence coincides with the formulations (1), (2) and (3), when $P(A|BC)$ and $P(B|C)$ are values on $(0, 1)$, while $P(C)$ must be greater than 0.

Example 1 (continued) – It is easy to check, directly by definition, that the event A is not cs-independent of B , in fact

$$\circ(A|B) = \circ(AB) - \circ(B) = \circ(\emptyset) - \circ(B) = \infty,$$

while being $AB^c = A$ and B^c possible events, the zero-layer $\circ(A|B^c) = \circ(A) - \circ(B^c)$ is finite.

In general, the following result has been proved in [5] and for the conditional case in [19]:

Theorem 2 Let AC, BC, B^cC be possible events. Let P be a coherent conditional probability such that $A \perp_{cs} B|C [P]$, then A and B are logically independent (i.e., none of the events $ABC, AB^cC, A^cBC, A^cB^cC$ is impossible).

In the aforementioned papers there is also a theorem characterizing stochastic independence of two logically independent events A and B in terms of probabilities $P(B|C), P(B|AC)$ and $P(B|A^cC)$, giving up any direct reference to the zero-layers.

Indeed, in [20] the definition of cs-independence has been extended to the case of finite sets of events.

Definition 3 Let $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ be three different partitions of Ω such that \mathcal{E}_2 is not trivial. The partition \mathcal{E}_1 is stochastically independent of \mathcal{E}_2 given \mathcal{E}_3 with respect to a coherent conditional probability P (in symbols $\mathcal{E}_1 \perp_{cs} \mathcal{E}_2 | \mathcal{E}_3 [P]$) iff $C_{i_1} \perp_{cs} C_{i_2} | C_{i_3} [P]$ for every $C_{i_1} \in \mathcal{E}_1, C_{i_2} \in \mathcal{E}_2, C_{i_3} \in \mathcal{E}_3$ such that $C_{i_2} \wedge C_{i_3} \neq \emptyset$.

In the quoted paper the following result has been proved

Theorem 3 Let $\mathcal{E}_1, \mathcal{E}_2$ be two partitions of Ω and P be a coherent conditional probability. Then

$$\mathcal{E}_1 \perp_{cs} \mathcal{E}_2 | D [P] \implies A \perp_{cs} B | D [P]$$

for every A logically dependent on \mathcal{E}_1 , and B logically dependent on \mathcal{E}_2 .

By the previous result also conditional independence for finite algebras $\mathcal{A}_1, \mathcal{A}_2$ have been characterized: \mathcal{A}_1 is cs-independent of \mathcal{A}_2 given a partition \mathcal{E} if the atoms \mathcal{E}_1 of \mathcal{A}_1 are cs-independent of \mathcal{E}_2 , the atoms of \mathcal{A}_2 , conditionally to every event in \mathcal{E} , i.e.

$$\mathcal{A}_1 \perp_{cs} \mathcal{A}_2 | \mathcal{E} [P] \iff \mathcal{E}_1 \perp_{cs} \mathcal{E}_2 | \mathcal{E} [P].$$

Starting from the previous considerations the case of random variables can be dealt. Let $X = (X_1, \dots, X_n)$ be a random vector with values in $R_X \subseteq \mathbb{R}^n$. The partition \mathcal{E} of the sure event Ω generated by X is denoted by $\mathcal{E}_X = \{X = x : x \in R_X\}$.

Definition 4 Let (X, Y, Z) be a finite discrete random vector with values in $R \subseteq R_X \times R_Y \times R_Z$ and $\mathcal{E}_X, \mathcal{E}_Y, \mathcal{E}_Z$ the partitions generated by X, Y and Z , respectively. Let P be a coherent conditional probability on \mathcal{F} containing $\{A|BC : A \in \mathcal{E}_X, B \in \mathcal{E}_Y, C \in \mathcal{E}_Z\}$: then X is stochastically cs-independent of Y given Z with respect to P (in symbol $X \perp_{cs} Y | Z [P]$) iff

$$\mathcal{E}_X \perp_{cs} \mathcal{E}_Y | \mathcal{E}_Z [P].$$

Note that Definition 4 does not require the classical assumptions: $R = R_X \times R_Y \times R_Z$ (logical constraints among the variables cannot be considered); the probability assessments are complete and positive.

The set \mathcal{M}_P of stochastic cs-independence statements induced by P of the form $X_I \perp_{cs} X_J | X_K$, where I, J and K are three disjoint subsets, is called *cs-independence model*. Every stochastic cs-independence model induced by P is closed with respect to the properties listed below ([20]). Every property can be interpreted as the requirement that \mathcal{M}_P is closed under respective ‘‘inference rules’’.

Decomposition property

$$X_I \perp_{cs} [X_J, X_K] | X_W [P] \implies X_I \perp_{cs} X_J | X_W [P];$$

Reverse decomposition property

$$[X_I, X_J] \perp_{cs} X_W | X_K [P] \implies X_I \perp_{cs} X_W | X_K [P];$$

Weak union property

$$X_I \perp_{cs} [X_J, X_K] | X_W [P] \implies X_I \perp_{cs} X_J | [X_W, X_K] [P];$$

Contraction property

$$X_I \perp_{cs} X_W | [X_J, X_K] [P] \& X_I \perp_{cs} X_J | X_K [P] \implies$$

$$X_I \perp_{cs} [X_J, X_W] | [X_K] [P];$$

Reverse contraction property

$$X_I \perp_{cs} X_W | [X_J, X_K] [P] \& X_J \perp_{cs} X_W | X_K [P] \implies$$

$$[X_I, X_J] \perp_{cs} X_W | [X_K] [P];$$

Intersection property

$$X_I \perp_{cs} X_J | [X_W, X_K] [P] \& X_I \perp_{cs} X_W | [X_J, X_K] [P] \implies$$

$$X_I \perp_{cs} [X_J, X_W] | [X_K] [P];$$

Reverse intersection property

$$X_I \perp_{cs} X_W | [X_J, X_K] [P] \& X_J \perp_{cs} X_W | [X_I, X_K] [P] \implies$$

$$[X_I, X_J] \perp_{cs} X_W | [X_K] [P].$$

Hence, these models satisfy all graphoid properties (see [17], [18]) except symmetry property

$$X_I \perp_{cs} X_J | X_K [P] \implies X_J \perp_{cs} X_I | X_K [P]$$

and reverse weak union property

$$[X_J, X_W] \perp_{cs} X_I | [X_K] [P] \implies X_J \perp_{cs} X_I | [X_W, X_K] [P].$$

In [19] also models (called *a-graphoid*) closed with respect to reverse weak union (but not necessarily with respect to symmetry) have been classified. The possible lack of symmetry is not counterintuitive, as explained in [5]: for example the validity of $A \perp_{cs} B$ means intuitively that the occurrence of B with positive probability does not ‘‘influence’’ the probability of A ; but it does not necessarily entail, conversely, that the occurrence of the ‘‘unexpected’’ (zero probability) event A should ‘‘influence’’ B . On the other hand, when the probabilities of A and B are in $(0, 1)$, if $A \perp_{cs} B$, then the symmetric statement $B \perp_{cs} A$ holds. So when the probability P is strictly positive, the cs-independence model induced by P is closed with respect to graphoid properties: symmetry, decomposition, weak union, contraction and intersection.

5 Conditional independence for lower probabilities

Extending a given coherent conditional probability assessment to a new event we do not necessarily get a unique value, but a bounded interval (for more details see [3] and for a relevant discussion see [8]), and the

bounds are called lower and upper conditional probabilities, respectively (other authors prefer to speak of *imprecise* probabilities, see [14]).

Given an arbitrary finite family \mathcal{F} of conditional events, a function \underline{P} on \mathcal{F} is a coherent lower conditional probability if there exists a non-empty family of conditional probabilities $\mathbf{P} = \{P(\cdot|\cdot)\}$ on \mathcal{F} (*dominating family*) whose lower envelope is \underline{P} , i.e., for any $E|H \in \mathcal{F}$,

$$\underline{P}(E|H) = \min_{\mathbf{P}} P(E|H).$$

The element P of the dominating family \mathbf{P} such that $P(E_i|H_i) = \underline{P}(E_i|H_i)$ for $E_i|H_i \in \mathcal{F}$ is called *i-minimal probability* and its agreeing class \mathcal{P} (see Section 2) is called *i-minimal class* [7].

It is well known that upper conditional probabilities $\overline{P}(E|H) = \max_{\mathbf{P}} P(E|H)$ (upper envelope of a class of conditional probabilities) are the dual functions of lower probabilities, i.e. $\overline{P}(E|H) = 1 - \underline{P}(E^c|H)$. By the previous transformation we can deal with both uncertainty functions in an unified way, so in the sequel we refer only to lower conditional probabilities.

The important role in inferential processes of imprecise probabilities leads to a generalization of the notion of stochastic independence. This topic is controversial, in fact there are several different formulations [10], [11], which essentially aim at being generalizations of formulations (1), (2), (3) of stochastic independence. Nevertheless, in the context of lower probabilities the role of zero values is even more delicate and some authors, as for example Cozman [10], require the positivity condition to avoid the related problems (even if he claims “future research must investigate the consequences of abandoning the positivity condition”).

In [7] a definition of independence for lower probabilities, which is able to handle zero lower probability values, is given.

Definition 5 *Given a coherent lower conditional probability \underline{P} , defined on a family \mathcal{F} containing $\mathcal{D} = \{A|B, A|B^c, A^c|B, A^c|B^c\}$, A is independent of B with respect to \underline{P} (in symbols $A \perp_{cs}^* B [P]$) if there exists a dominating class \mathbf{P} such that, for every $P \in \mathbf{P}$, it holds $A \perp_{cs} B [P]$.*

The latter requirement of Definition 5 can be limited only to i-minimal probabilities, for any event in \mathcal{F} .

Remark 2 *Definition 5 requires only the existence of a class satisfying independence condition. In other words, it is not required that different dominating classes induce the same independence statements.*

More details on the suitability of referring the inde-

pendence to a class of probabilities generating a given lower conditional probabilities are in [7].

Moreover, in the aforementioned paper the main properties of this definition have been deepened and, in particular, the validity of the following implication is proved

$$A \perp_{cs}^* B [P] \implies \underline{P}(A|B) = \underline{P}(A|B^c) = \underline{P}(A). \quad (4)$$

Essentially, the statement $A \perp_{cs}^* B [P]$ means that $A \perp_{cs} B [P]$, for every P belonging to (at least one) dominating class. Also in the case of lower probabilities, cs-independence implies *logical independence*.

Definition 5 can be extended easily to conditional independence.

Definition 6 *Given a coherent lower conditional probability \underline{P} , defined on a family \mathcal{F} containing $\mathcal{D} = \{A|BC, A|B^cC, A^c|BC, A^c|B^cC\}$, A is conditionally independent of B given C with respect to \underline{P} (in symbols $A \perp_{cs}^* B|C [P]$) if there exists a dominating class \mathbf{P} such that, for every $P \in \mathbf{P}$, it holds $A \perp_{cs} B|C [P]$.*

Following the line of the proof in [7], we get

$$A \perp_{cs}^* B|C \implies \underline{P}(A|BC) = \underline{P}(A|B^cC) = \underline{P}(A|C) \quad (5)$$

Definition 6 can be extended to finite sets of events, as done in Section 4 for probabilities:

Definition 7 *Let $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}$ be three finite partitions of Ω . Given a coherent lower conditional probability \underline{P} , we say that \mathcal{E}_1 is stochastically independent of \mathcal{E}_2 given \mathcal{E} with respect to \underline{P} (in symbols $\mathcal{E}_1 \perp_{cs}^* \mathcal{E}_2|\mathcal{E} [P]$) iff there exists a dominating class \mathbf{P} such that, for every $P \in \mathbf{P}$, it holds $C_{i_1} \perp_{cs}^* C_{i_2}|C_{i_3} [P]$ for every $C_{i_1} \in \mathcal{E}_1$, $C_{i_2} \in \mathcal{E}_2$, $C_{i_3} \in \mathcal{E}$ such that $C_{i_2} \wedge C_{i_3} \neq \emptyset$.*

From Theorem 3 and Definition 6 it follows that given two different non trivial partitions $\mathcal{E}_1, \mathcal{E}_2$ of Ω and a lower probability \underline{P} , then

$$\mathcal{E}_1 \perp_{cs}^* \mathcal{E}_2|D [P] \implies A \perp_{cs}^* B|D [P]$$

for every A and B logically dependent on $\mathcal{E}_1, \mathcal{E}_2$, respectively. Therefore, a finite algebra \mathcal{A}_1 is cs-independent of another finite algebra \mathcal{A}_2 if the same independence relation is valid between their partitions, i.e.

$$\mathcal{A}_1 \perp_{cs}^* \mathcal{A}_2|\mathcal{E} \iff \mathcal{E}_1 \perp_{cs}^* \mathcal{E}_2|H \text{ for each } H \in \mathcal{E}.$$

Starting from the previous considerations, the case of random variables can be faced also in this framework.

Definition 8 *Let (X, Y, Z) be a finite discrete random vector with values in $R \subseteq R_X \times R_Y \times R_Z$ and $\mathcal{E}_X, \mathcal{E}_Y, \mathcal{E}_Z$ the partitions generated by X, Y and Z ,*

respectively. Let \underline{P} be a coherent lower conditional probability, X is stochastically cs-independent of Y given Z with respect to \underline{P} (in symbol $X \perp_{cs} Y | Z [\underline{P}]$) iff

$$\mathcal{E}_X \perp_{cs}^* \mathcal{E}_Y | \mathcal{E}_Z [\underline{P}].$$

Since conditional independence for lower probabilities \underline{P} is related to a dominating class \mathbf{P} , we refer to \mathbf{P} a set of conditional cs-independence statements induced by \underline{P} .

From the definition of conditional independence for lower probabilities and the results presented in Section 3, we get the following result

Theorem 4 *Let $\mathcal{M}_{\underline{P}}$ be a cs-independence model induced by the lower conditional probability \underline{P} , then $\mathcal{M}_{\underline{P}}$ is closed with respect to decomposition property and its reverse, weak union property, contraction property and its reverse, intersection property and its reverse.*

6 Graphical representation of independence structures

Graphical models have their origin in several areas: artificial intelligence, probability and statistics. Their applicability is due to two main factors: graphical visualization of independence statements, which facilitates communication between field experts and statistician; secondly, computational complexity reduction. In this paper we focus our attention just on the first aspect: representation of the conditional cs-independence structures.

Since the usual graphical models are not apt to describe cs-independence models, we present some notions of graphs slightly different from the usual ones.

An *l-graph* is a triplet $G = (V, E, \mathcal{B})$, where V is a finite set of *vertices*, E is a set of *edges* defined as a subset of $V \times V$ (i.e. set of all ordered pairs of distinct vertices), and the family $\mathcal{B} = \{B : B \subseteq V\}$ of subsets of vertices.

The vertices are represented by circles, and each $B \in \mathcal{B}$ by a box enclosing those circles corresponding to vertices in B .

Definition of l-graph differs from that of graph (see [16], [17]), since our interest for \mathcal{B} is to gather the sets of variables linked by some logical constraint. More precisely, every vertex $v \in V$ or subset $I \subseteq V$ is associated to a variable X_v or to a random vector X_I , respectively, and a box $B = \{v : v \in J\}$ visualizes the *minimal set* of random variables $\{X_v : v \in J\}$ whose partitions generate the given logical relation (see Section 3).

Consider the l-graph in Figure 1, which has no edges

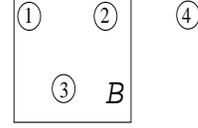


Figure 1: Graphical representation of variables linked by a logical constraint

The box $B = \{1, 2, 3\}$ denotes a logical component and represents a logical relation among the variables X_1, X_2, X_3 : it could be, for example, that the event $\{X_1 = 1, X_2 = 0, X_3 = 0\}$ is not possible. The box B is used to indicate where the logical constraint is localized.

An edge $(u, v) \in E$ is called *undirected* if also $(v, u) \in E$; whereas an edge (u, v) such that its “opposite” (v, u) is not in E is called *directed*. An undirected edge (u, v) is represented by $u - v$, whereas $u \rightarrow v$ is used for a directed edge. If an l-graph has only undirected edges, it is called *undirected* l-graph, while if all edges are directed, the graph is said to be *directed*. In directed l-graphs, if $u \rightarrow v$, then u is said to be a *parent* of v and v a *child* of u .

A *path* of length n from u to v is a sequence $u = u_1, \dots, u_n = v$ of vertices such that either $(u_i, u_{i+1}) \in E$ and $(u_{i+1}, u_i) \notin E$ or $(u_{i+1}, u_i) \in E$ and $(u_i, u_{i+1}) \notin E$ for $i = 1, \dots, n - 1$.

A *directed path* of length n from u to v is a sequence $u = u_1, \dots, u_n = v$ of vertices such that $(u_i, u_{i+1}) \in E$ and $(u_{i+1}, u_i) \notin E$ for $i = 1, \dots, n - 1$.

If there is a path from u to v , we say that u leads to v , and we denote it as $u \mapsto v$. An *n-cycle* is a path of length n that begins and ends in the same vertex $u_1 = u_n$. A directed l-graph is *acyclic* if it contains no cycles.

The vertices u such that $u \mapsto v$ and there is no path from v to u , are the *ancestors* $an(v)$ of v ; the *descendants* $ds(u)$ of u are the vertices v such that $u \mapsto v$ and there is no path from v to u .

6.1 A separation criterion for undirected l-graph

To represent conditional cs-independence relations, we need to introduce a “rule”, called separation criterion, that allows to read the relations directly from l-graphs.

Definition 9 *Let V_1, V_2, S be three disjoint subsets of V . The set V_1 is *u-separated* from V_2 by S in the l-graph $G = (V, E, \mathcal{B})$ if the following conditions hold: (u_1) every path from $u \in V_1$ to $v \in V_2$ goes through a*

vertex in S

(u_2) there is no $B_i \in \mathcal{B}$ such that $B_i \subseteq V_1 \cup V_2 \cup S$, and both sets $B_i \cap V_1$ and $B_i \cap V_2$ are not empty.

The condition (u_2) can be rewritten as follows:

- $\forall B \in \mathcal{B}$ s.t. $B \subseteq V_1 \cup V_2 \cup S$ one has either $B \cap V_1 = \emptyset$ or $B \cap V_2 = \emptyset$;
- $\forall B \in \mathcal{B}$ s.t. $B \subseteq V_1 \cup V_2 \cup S$ one has either $B \subseteq V_1 \cup S$ or $B \subseteq V_2 \cup S$.

The statement “the set of vertices V_1 is u-separated by V_2 given S ” in an l-graph $G = (V, E, \mathcal{B})$ is denoted as $(V_1, V_2|S)_G^u$. The difference between u-separation criterion and the classical one [16] is established by condition (u_2), where logical components are considered. Therefore, to detect the properties of u-separation, we must check the graphoid properties verified by the relation $X \perp_{\mathcal{B}} Y|Z$:

$$\forall B \in \mathcal{B} \quad B \subseteq X \cup Y \cup Z \implies B \subseteq X \cup Z \text{ or } B \subseteq Y \cup Z.$$

Theorem 5 *The relation $X \perp_{\mathcal{B}} Y|Z$ is a graphoid.*

Proof. We must prove that the relation verifies symmetry, decomposition, weak union, contraction and intersection properties. Symmetry is trivial.

Suppose now $X \perp_{\mathcal{B}} Y \cup W|Z$. Any $B \subseteq X \cup Y \cup Z \in \mathcal{B}$ is such that $B \subseteq X \cup W \cup Y \cup Z$ and by assumption $B \subseteq X \cup Z$ or $B \subseteq W \cup Y \cup Z$, then being W and $Y \cup Z$ disjoint we get $B \subseteq X \cup Z$ or $B \subseteq Y \cup Z$. Therefore, the statement $X \perp_{\mathcal{B}} Y \cup W|Z$ implies $X \perp_{\mathcal{B}} Y|Z$, i.e. decomposition property is satisfied.

Moreover, for any $B \subseteq X \cup Y \cup W \cup Z$, by assumption it follows $B \subseteq Y \cup W \cup Z$ or $B \subseteq X \cup Z \subseteq X \cup W \cup Z$, so weak union property $X \perp_{\mathcal{B}} Y \cup W|Z \implies X \perp_{\mathcal{B}} Y|W \cup Z$ is satisfied.

Now suppose that $X \perp_{\mathcal{B}} W|Z$ and $X \perp_{\mathcal{B}} Y|W \cup Z$, so every $B \subseteq X \cup Y \cup W \cup Z$ is such that $B \subseteq Y \cup W \cup Z$ or $B \subseteq X \cup W \cup Z$, the latter implies $B \subseteq X \cup Z$ or $B \subseteq W \cup Z$ by second assumption. Therefore, $X \perp_{\mathcal{B}} W|Z$ and $X \perp_{\mathcal{B}} Y|W \cup Z$ imply $X \perp_{\mathcal{B}} Y \cup W|Z$, i.e. contraction property holds.

Suppose now that $X \perp_{\mathcal{B}} W|Y \cup Z$ and $X \perp_{\mathcal{B}} Y|W \cup Z$, if $B \subseteq X \cup Y \cup W \cup Z$ then $B \subseteq X \cup Y \cup Z$ or $B \subseteq W \cup Y \cup Z$ for the first assumption, $B \subseteq X \cup W \cup Z$ or $B \subseteq Y \cup W \cup Z$ for the second assumption. To prove the intersection property ($X \perp_{\mathcal{B}} Y|W \cup Z$ and $X \perp_{\mathcal{B}} W|Y \cup Z$ imply $X \perp_{\mathcal{B}} Y \cup W|Z$), we must verify that $B \subseteq X \cup Z$ or $B \subseteq Y \cup W \cup Z$. The two assumptions imply that $B \subseteq Y \cup W \cup Z$ or the validity of both the relations $B \subseteq X \cup W \cup Z$ and $B \subseteq X \cup Y \cup Z$; in the second case $B \subseteq X \cup Z$.

It is well known that classical vertex separation criterion for undirected graphs satisfies graphoid properties [16], so by the previous result we can conclude

that also vertex u-separation criterion verify the same properties.

Corollary 1 *The vertex u-separation verifies graphoid properties.*

Remark 3 *Composition property*

$$(X, Y|Z)_G^u \& (X, W|Z)_G^u \implies (X, Y \cup W|Z)_G^u$$

need not hold. Obviously, the hypotheses imply that every path going from X to $Y \cup W$ is blocked by Z , in fact the classical vertex separation verifies such property. But, the second condition (u_2) characterizing u-separation criterion can fail.

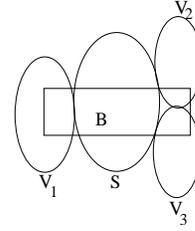


Figure 2: Counter-example for composition property

In fact, it is enough to consider $B \in \mathcal{B}$ such that $B \cap V_i \neq \emptyset$ for every $i = 1, 2, 3$ and $B \subseteq V_1 \cup V_2 \cup V_3 \cup S$. Hence, $B \cap (V_2 \cup V_3) \neq \emptyset$ and $\neg((V_1, V_2 \cup V_3|S)_G^u)$. On the other hand, $B \not\subseteq V_1 \cup V_2 \cup S$ and $B \not\subseteq V_1 \cup V_3 \cup S$. The mentioned situation is illustrated in Figure 2, where the ovals represent the sets of vertices. Given a cs-independence models \mathcal{M} , we say that the l-graph G represents \mathcal{M} if $(V_1, V_2|S)_G^u$ implies $X_{V_1} \perp_{cs}^ X_{V_2} | X_S \in \mathcal{M}$. Actually, stochastic independence models can violate composition property: we may have $X \perp_{cs} X_j [P]$ for $j \in J$ while the statement $X \perp_{cs} X_J [P]$ is not valid. It is well known that separation criteria in [16], [17] do not reflect such feature, in fact they satisfy composition property.*

Now we just show a simple example to better understand the proposed criterion.

Example 2 Consider the random vector (X_1, X_2, X_3) , where the variables X_i ($i = 1, 2, 3$) are binary and X_1 and X_2 are logically linked by the constraint: the event $\{X_1 = 1\} \wedge \{X_2 = 1\}$ (denoted also as $\{X_1 = 1, X_2 = 1\}$) is impossible. Let \mathcal{M} be the independence model formed by the statements $X_2 \perp_{cs}^* X_3$, $X_3 \perp_{cs}^* X_2$, $X_2 \perp_{cs}^* X_3 | [X_1]$, $X_3 \perp_{cs}^* X_2 | [X_1]$.

\mathcal{M} can be represented “completely” thanks to u-separation criterion (i.e. all the statements and only those in \mathcal{M} are represented) by the undirected l-graph drawn in Figure 3.

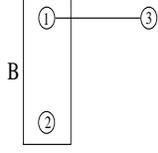


Figure 3: Graphical representation of \mathcal{M} .

The logical constraint between X_1 and X_2 is visualized in the l-graph by the logical component B . Otherwise the graph would represent inexistent independence statement. Note that \mathcal{M} cannot be represented completely neither by directed acyclic graphs using d-separation (or its generalization called D-separation [17]).

The cs-independence model \mathcal{M} can be induced, for example, by the following probability P

$$\begin{aligned} P(X_2 = 1, X_3 = 1) &= P(X_2 = 1, X_3 = 0) = \frac{1}{6}, \\ P(X_1 = 0, X_2 = 0, X_3 = 1) &= \frac{1}{3} \\ P(X_1 = 0, X_2 = 0, X_3 = 0) &= \frac{1}{3} \end{aligned}$$

The computations to check that P actually induces the statements in \mathcal{M} are trivial and there is no need to specify the zero-layers of $\{X_1 = 1, X_3 = 1\}$ and $\{X_1 = 1, X_3 = 0\}$.

6.2 Separation criteria for directed acyclic l-graph

In [19] two different separation criteria for directed acyclic l-graph have been given to represent cs-independence models. The first called *dl-separation* is a generalization of d-separation criterion [17] obtained introducing the notion of logical components as done for undirected l-graphs. Before introducing the new separation criterion, we recall the classical definition of blocked path given in [17].

Definition 10 *Let G be an acyclic directed graph. A path u_1, \dots, u_m in G is blocked by a set of vertices $S \subseteq V$, whenever there exists a triplet of connected vertices u, v, w such that of the following condition holds:*

1. *either $u \rightarrow v, v \rightarrow w$ or $w \rightarrow v, v \rightarrow u$, and $v \in S$*
2. *$v \rightarrow u, v \rightarrow w$ and $v \in S$*
3. *$u \rightarrow v, w \rightarrow v$ and $v \cup ds(v) \notin S$*

The conditions can be illustrated by Figure 4 where the grey vertices are those belonging to S .

Vertex d-separation criterion requires that every path going from one set to the other is blocked.

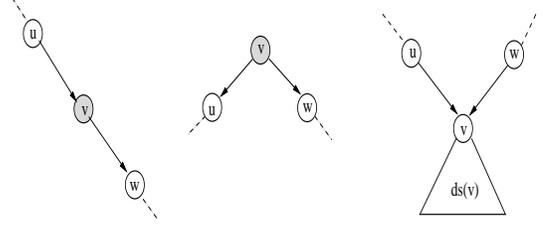


Figure 4: Blocked paths

Our generalization is the following

Definition 11 *Let $G = (V, E, \mathcal{B})$ be an acyclic directed l-graph and let V_1, V_2 and S be three disjoint set of vertices of V . The set of vertices S dl-separates V_1 from V_2 in G (in symbol $(V_1, V_2 | S)_G^{dl}$), whenever every path in G from V_1 to V_2 is blocked by S and the following condition holds:*

(u_2) *there is no $B_i \in \mathcal{B}$ such that $B_i \subseteq V_1 \cup V_2 \cup S$, and both sets $B_i \cap V_1$ and $B_i \cap V_2$ are not empty.*

It is immediate to check that the possible boxes B_i in the three situations of Figure 4 can be formed only by $\{u, v\}$ and $\{v, w\}$, but cannot be $\{u, w\}$.

Example 3 Let X_i ($i = 1, 2, 3$) be binary variables such that $\{X_1 = 0, X_2 = 0, X_3 = 0\}$ is impossible. Consider the lower conditional probability \underline{P} defined as follows

$$\begin{aligned} \underline{P}(X_1 = 1 | X_2 = 1) &= 0.2 = \underline{P}(X_1 = 1 | X_2 = 0) \\ \underline{P}(X_2 = 1 | X_1 = 1) &= 0.3 = \underline{P}(X_2 = 1 | X_1 = 0) \\ \underline{P}(X_3 = 1 | X_2 = 1) &= 0 = \underline{P}(X_3 = 1 | X_1 = 1) \\ \underline{P}(X_3 = 1 | X_2 = 0) &= \frac{4}{5}, \quad \underline{P}(X_3 = 1 | X_1 = 0) = \frac{5}{8} \\ \underline{P}(X_1 = 1 | X_3 = 1) &= 0 = \underline{P}(X_2 = 1 | X_3 = 1) \\ \underline{P}(X_1 = 1 | X_3 = 0) &= \frac{1}{3} \quad \underline{P}(X_2 = 1 | X_3 = 0) = \frac{15}{22} \end{aligned}$$

Because of the logical constraint the only possible independence statements induced by \underline{P} are between the couples of variables. The given assessment cannot induce the statements $X_3 \perp_{cs}^* X_1, X_3 \perp_{cs}^* X_2$ and their symmetric because (for $i = 1, 2$)

$$\begin{aligned} \underline{P}(X_3 = 1 | X_i = 1) &\neq \underline{P}(X_3 = 1 | X_i = 0) \\ \underline{P}(X_i = 1 | X_3 = 1) &\neq \underline{P}(X_i = 1 | X_3 = 0). \end{aligned}$$

On the other hand, we can show that

$$\mathcal{M}_{\underline{P}} = \{X_1 \perp_{cs}^* X_2, X_2 \perp_{cs}^* X_1\}.$$

is induced by the i-minimal class

$$\begin{aligned} P_0^1(X_1 = 1, X_2 = 1, X_3 = 0) &= 0.06 \\ P_0^1(X_1 = 1, X_2 = 0, X_3 = 0) &= 0.14 \\ P_0^1(X_1 = 0, X_2 = 1, X_3 = 0) &= 0.24 \\ P_0^1(X_1 = 0, X_2 = 0, X_3 = 1) &= 0.56 \end{aligned}$$

and P_0^1 is 0 elsewhere, while

$$\begin{aligned}
P_1^1(X_1 = 1, X_2 = 0, X_3 = 1) &= 0.3 \\
P_1^1(X_1 = 0, X_2 = 1, X_3 = 1) &= 0.5 \\
P_1^1(X_1 = 1, X_2 = 1, X_3 = 1) &= 0.2;
\end{aligned}$$

and

$$\begin{aligned}
P_0^2(X_1 = 1, X_2 = 1, X_3 = 1) &= 0.05 \\
P_0^2(X_1 = 1, X_2 = 1, X_3 = 1) &= 0.05 \\
P_0^2(X_1 = 1, X_2 = 0, X_3 = 1) &= 0 \\
P_0^2(X_1 = 1, X_2 = 0, X_3 = 0) &= 0.1 \\
P_0^2(X_1 = 0, X_2 = 1, X_3 = 1) &= 0.1 \\
P_0^2(X_1 = 0, X_2 = 1, X_3 = 0) &= 0.3 \\
P_0^2(X_1 = 0, X_2 = 0, X_3 = 1) &= 0.4.
\end{aligned}$$

It is easy to check that these two classes are minimal for \underline{P} and moreover, both imply that $X_1 \perp_{cs} X_2$ and $X_2 \perp_{cs} X_1$.

$\mathcal{M}_{\underline{P}}$ can be represented by the l-graph in Figure 5.

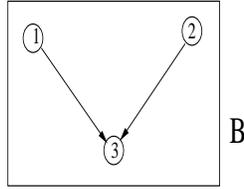


Figure 5: Acyclic directed l-graphs describing $\mathcal{M}_{\underline{P}}$.

From Theorem 5 and from the properties of blocked path it follows

Corollary 2 *Vertex dl-separation satisfies graphoid properties.*

However, dl-separation does not verify composition property for the same reasons presented for u-separation. Since dl-separation does not allow to represent *not symmetric* statements, we introduce another criterion apt for this aim.

Definition 12 *Let G be an acyclic directed l-graph. A path u_1, \dots, u_n , $n \geq 1$ in G is blocked by a set of vertices $S \subseteq V$, whenever there exists a triplet of consecutive vertices w, v, u in the path such that one of the following three condition holds:*

1. $u \rightarrow v \rightarrow w$ and $v \in S$ (i.e. w, v, u is a reverse directed path)
2. $u \leftarrow v \rightarrow w$ and $v \in S$
3. $u \rightarrow v \leftarrow w$ and $ds(v) \notin S$

Note that Definition 12 strictly depends on the direction of the path (see the first condition).

Figure 6 shows a case where the directed path w, v, u is not blocked ($v \in S$).

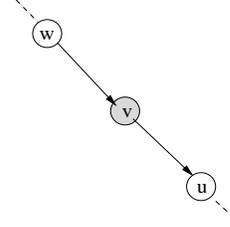


Figure 6: Not blocked directed path

The difference between the introduced notion of blocked path and that used in d-separation criterion [17] is illustrated by means of Figure 4 and Figure 6. The path w, v, u on the left of Figure 4 is blocked by v , while its reverse is represented on Figure 6 is not blocked by v because of the direction. Hence, the reverse path of one blocked is not necessarily blocked according to our definition, so the blocking path notion does not satisfy symmetric property. The second and third cases of Definition 12 are like in d-separation criterion.

Definition 13 *Let G be an directed acyclic l-graph and let U, W and S be three pairwise disjoint set of vertices of V . We say that U is l-separated from W by S in G and write symbol $(U, W|S)_G^l$, whenever every path in G from U to W is blocked by S and moreover, the following “logical separation” condition holds $(u_2) \forall B \in \mathcal{B}$ s.t. $B \subseteq U \cup W \cup S$ one has either $B \cap U = \emptyset$ or $B \cap W = \emptyset$.*

Corollary 3 *Vertex l-separation verifies a-graphoid properties.*

Therefore, l-separation helps to represent structures not necessarily closed with respect to symmetric property (as cs-independence models or other models known in literature see for example [1], [10], [11], [21]).

Example 4 Let X_1, X_2, X_3 be three binary variables. Consider the lower conditional probability P

$$\begin{aligned}
\underline{P}(X_1 = 1|X_2 = 1, X_3 = 1) &= 0 \\
\underline{P}(X_1 = 1|X_2 = 0, X_3 = 1) &= 0 \\
\underline{P}(X_2 = 1, X_3 = 1) &= 0.1 \\
\underline{P}(X_1 = 0, X_2 = 1, X_3 = 1) &= 0.1 \\
\underline{P}(X_1 = 0, X_2 = 0, X_3 = 1) &= 0.1 \\
\underline{P}(X_2 = 1|X_1 = 1, X_3 = 1) &= 0.25 \\
\underline{P}(X_1 = 1|X_2 = 1, X_3 = 0) &= 0.6 \\
\underline{P}(X_1 = 1|X_2 = 0, X_3 = 0) &= 0.6
\end{aligned}$$

It is easy to find a dominating class such that the induced cs-independence model is

$$\mathcal{M} = \{X_1 \perp_{cs}^* X_2 \mid X_3 [P]\}.$$

On the other hand, the statement $X_2 \perp_{cs}^* X_1 \mid X_3$ cannot be induced by \underline{P} , since

$$\begin{aligned} \underline{P}(X_2 = 1 \mid X_1 = 1, X_3 = 1) &= 0.25 \neq \\ &\neq \frac{0.1}{0.1 + 0.1} = 0.5 = \underline{P}(X_2 = 1 \mid X_1 = 0, X_3 = 1). \end{aligned}$$

The l-graph representing completely \mathcal{M} according to the l-separation criterion is drawn in Figure 7.

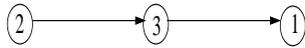


Figure 7: Graphical representation for \mathcal{M} .

In fact, the path 1, 3, 2 is blocked by 3, while the path 2, 3, 1 is not blocked by 3 because of the direction.

References

- [1] B. Bouchon-Meunier, G. Coletti and C. Marsala. Independence and Possibilistic Conditioning. To appear in *Annals of Mathematics and Artificial Intelligence*.
- [2] G. Coletti. Coherent Numerical and Ordinal Probabilistic Assessments. *IEEE Trans. on Systems, Man, and Cybernetics*, 24(12):1747–1754, 1994.
- [3] G. Coletti and R. Scozzafava. Characterization of Coherent Conditional Probabilities as a Tool for their Assessment and Extension. *International Journal of Uncertainty, Fuzziness and Knowledge-Based System*, 4(2):103–127, 1996.
- [4] G. Coletti and R. Scozzafava. Null Events and Stochastic Independence. *Kibernetika*, 34(1):69–78, 1998.
- [5] G. Coletti and R. Scozzafava. Zero Probabilities in Stochastic Independence. *Information, Uncertainty, Fusion* (Eds. B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh), Kluwer, Dordrecht, 185–196, 2000. Selected papers from IPMU '98.
- [6] G. Coletti and R. Scozzafava. Conditioning and Inference in Intelligent Systems. *Soft Computing*, 3:118–130, 1999.
- [7] G. Coletti and R. Scozzafava. Stochastic Independence for Upper and Lower Probabilities in a Coherent Setting. In *Proc. of IPMU 2000*, Madrid (Spain), 1:341–348, 2000.
- [8] G. Coletti and R. Scozzafava. The Role of Coherence in Eliciting and Handling Imprecise Probabilities and its Application to Medical Diagnosis. *Information Sciences*, 130:41–65, 2000.
- [9] R.G. Cowell, A.P. Dawid, S.L. Lauritzen and D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems*, Springer-Verlag, 1999.
- [10] F.G. Cozman. Credal Networks. *Artificial Intelligence*, 120:199–233, 2000.
- [11] L. de Campos and S. Moral. Independence Concepts for Convex Sets of Probabilities. In P.Besnard, S. Hanks (Eds.), *Proc. XI Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, CA, 1995, pp.108–115.
- [12] B. de Finetti. Sull'Impostazione Assiomatica del Calcolo delle Probabilità. *Annali dell'Università di Trieste*, 19:3–55, 1949. (Eng. transl.: Ch. 5 in *Probability, Induction, Statistics*, London: Wiley, 1972).
- [13] L.E. Dubins. Finitely Additive Conditional Probabilities, Conglomerability and Disintegration. *Annals of Probability*, 3:89–99, 1975.
- [14] ISIPTA'99. *Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications*, de Cooman, G., Cozman, F.G., Moral, S., Walley, P. (eds.), Imprecise Probabilities Project, Universiteit Gent, Belgium, 1999.
- [15] P.H. Krauss. Representation of Conditional Probability Measures on Boolean Algebras. *Acta Math. Acad. Sci. Hungar.* 19:229–241, 1968.
- [16] S.L. Lauritzen. *Graphical Models*. Clarendon Press, Oxford, 1996.
- [17] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- [18] M. Studeny and R.R. Bouckaert. On Chain Graph Models for Description of Conditional Independence Structures. *The Annals of Statistics*, 26(4):1434–1495, 1998.
- [19] B. Vantaggi. Conditional Independence and Graphical Models in a Coherent Setting. *PhD Thesis*, Università di Perugia, Italy, 1999.
- [20] B. Vantaggi. Conditional Independence in a Coherent Finite Setting. *Annals of Mathematics and Artificial Intelligence*, to appear.
- [21] J. Vejnarova. Conditional Independence Relations in Possibility Theory, *Int. J. of Uncertainty, Fuzziness and Knowledge-based Systems*, 8(3), 253–269, 2000.
- [22] J. Witthaker. *Graphical Models in Applied Multivariate Statistics*. Wiley & Sons, New York, 1990.