Graphical Representation of Asymmetric Graphoid Structures

B. VANTAGGI
Università "La Sapienza," Roma, Italy

Abstract

Independence models induced by some uncertainty measures (e.g. conditional probability, possibility) do not obey the usual graphoid properties, since they do not satisfy the symmetry property. They are efficiently representable through directed acyclic l-graphs by using L-separation criterion.

In this paper, we show that in general there is not a l-graph which describes completely all the independence statements of a given model; hence we introduce in this context the notion of minimal I-map and we show how to build it, given an ordering on the variables. In addition, we prove that, for any ordering, there exists an I-map for any asymmetric graphoid structure.

Keywords

conditional independence models, directed acyclic graph, L-separation criterion, I-map

1 Introduction

The use of graphs to describe conditional independence structures (the set of conditional independence statements "X is independent of Y given Z") induced by probability distributions has a long and rich tradition; one can distinguish three main classic approaches based on *undirected graphs* [12], *directed acyclic graphs* [14], or *chain graphs* [15]. These graphical structure obey graphoid properties (symmetry, decomposition, weak union, contraction, intersection). On the other hand, the independence models based on the classic definition of stochastic independence in the usual probabilistic setting, have semi-graphoid structure (they satisfy all graphoid properties except intersection). However, if the probability distribution is strictly positive, the independence model has a graphoid structure. Hence, the lack of intersection property is due to zero probability on some of the possible events. Actually, it is well-known (see, for example, [4, 6]) that the classic definition of stochastic independence presents counter-intuitive situations when zero or one probability events are involved: for example, a possible event with zero or one probability is independent of itself.

We stress that zero probability values are interesting not only from a merely theoretical point of view, but they are met in many real problems, for example in medical diagnosis [7] , statistical mechanics, physics, etc. [11].

The counter-intuitive situations cannot be avoided within the usual framework of conditional probability. In the more general framework (de Finetti [8], Dubins [9]), a definition of stochastic independence (called cs-independence), which avoids these critical situations, has been introduced in [4] and the main properties have been studied. We recall that the aforementioned definition agrees with the classic one when the probabilities of the relevant events are different from 0 and 1.

The main properties connected with graphoid structures were proved in [16]: these independence models generally are not closed with respect to the symmetry property. Hence, the classic separation criterion are not apt to represent asymmetric independence statements, so in [17] a new separation criterion (called L-separation) for directed acyclic l-graphs has been introduced. It has been shown also that L-separation criterion satisfy *asymmetric graphoid* properties (graphoid properties except symmetry).

In this paper we deepen the problem of representing such cs-independence model, together with the logical constraints, using L-separation criterion in directed acyclic l-graphs. In particular, Example 1 shows that cs-independence structures are richer than the graphical ones, i.e. for some independence model there is no graph able to describe all the independence statements. Hence, in Section 5 we define in this context (analogously to [14, 10]) the notion of minimal I-map for a given independence model $\mathcal{M}$: a directed acyclic l-graph such that every statement represented by it is in $\mathcal{M}$, while the graph obtained by removing any arrow from it would represent an independence statement not in $\mathcal{M}$.

Moreover, in Section 5 we show how to build such minimal I-maps underling the differences arising from the lack of symmetry property, and, in addition, we prove that any ordering on the variables gives rise to an I-map for any independence model $\mathcal{M}$ obeying to asymmetric graphoid properties.

On the other hand, the ordering has a crucial role: in fact, if a perfect I-map (able to describe all the independence statements) exists, it can be built using only some specific ordering on the variables.

2 Independence in a coherent probability setting

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

$$P(A \wedge B) = P(A)P(B) \tag{1}$$

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 classic formulations are $P(A|B) = P(A)$ and

$P(A|B) = P(A|B^c)$, 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.

Anyway, some critical situations related to logical dependence continue to exist (see [16]) also considering the last stronger formulation in the more general framework of de Finetti [8]:

Definition 1 *Given a Boolean algebra $\mathcal{A}$, a conditional probability on $\mathcal{A} \times \mathcal{A}^0$ (with $\mathcal{A}^0 = \mathcal{A} \setminus \{\emptyset\}$) 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 the 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 (Kolmogorov's definition) by this unique "unconditional" probability P_0.

As proved in [6], in all other cases to get a similar representation we need to resort to a finite family $\mathcal{P} = \{P_0, \ldots, P_k\}$ of unconditional probabilities:
- every P_α is defined on a proper set of events (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 $E|H$ one has $P(E|H) = \frac{P_\alpha(E \wedge H)}{P_\alpha(H)}$ with $P_\alpha(H) > 0$.

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

Such theory of conditional probability allows to handle also *partial* probability assessment on an arbitrary set of conditional events $\mathcal{F} = \{E_1|H_1, \ldots, E_n|H_n\}$ through the concept of 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 $\{E_1, H_1, \ldots, E_n, H_n\}$. A characterization of coherence was proven in [3]:

Theorem 1 *Let $\mathcal{F}$ be an arbitrary finite family of conditional events and $\mathcal{C}$ denote the set of atoms C_r generated by the events $E_1, H_1, \ldots, E_n, H_n$. 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 a class of unconditional probabilities $\{P_0, \ldots P_k\}$, with P_0 defined 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, given 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 [4]) as

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

In the sequel, to avoid cumbersome notation, the conjunction symbol $\wedge$ among events is omitted.

In this framework the following definition of stochastic independence has been proposed in [4] and extended to conditional independence in [16]:

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

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

(ii) there exists a class $\{P_\alpha\}$ *of probabilities agreeing with the restriction of P to the family* $\mathcal{D}$*, such that*

$$\circ(A|BC) = \circ(A|B^cC) \quad \text{and} \quad \circ(A^c|BC) = \circ(A^c|B^cC).$$

Note that if $0 < P(A|BC) = P(A|B^cC) < 1$ (so $0 < P(A^c|BC) = P(A^c|B^cC) < 1$), then both equalities in condition (ii) are trivially satisfied

$$\circ(A|BC) = 0 = \circ(A|B^cC) \quad \text{and} \quad \circ(A^c|BC) = 0 = \circ(A^c|B^cC).$$

Hence, in this case condition (i) completely characterizes conditional cs-independence, and, in addition, this definition coincides with the classic formulations when also $P(B|C)$ and $P(C)$ are in $(0,1)$. However, in the other cases (when $P(A|BC)$ is 0 or 1) condition (i) needs to be "reinforced" by the requirement that also their zero-layers must be equal, otherwise we can meet critical situations (see, e.g. [6]).

Observation 1 *Even if different agreeing classes generated by the restriction of P on* $\mathcal{D}$ *may give rise to different zero-layers, it has been proved in [5, 6] that condition (ii) of Definition 2 either holds for all the agreeing classes of P or for none of them.*

Notice that for every event A this notion of stochastic independence is always irreflexive (also when the probability of A is 0 or 1) because $\circ(A|A) = 0$, while $\circ(A|A^c) = \infty$. Moreover, conditional independence of two possible events A and B imply the *logical independence* of A and B, i.e. all the events of the kind $A^* \wedge B^*$ is possible, with A^* - analogously B^* - is either A or A^c. (see [4]).

In [4, 16] theorems characterizing stochastic and conditional independence of two logically independent events A and B in terms of probabilities $P(B|C), P(B|AC)$ and $P(B|A^cC)$ is given, giving up any direct reference to the zero-layers.

Theorem 2 *Let A, B be two events logically independent with respect to the event C. If P is a coherent conditional probability such that $P(A|BC) = P(A|B^cC)$, then $A \perp\!\!\!\perp_{cs} B \mid C$ if and only if one of the following conditions holds:*

(a) $0 < P(A|BC) < 1$*;*

(b) $P(A|BC) = 0$ *and the extension of P to $B|C$ and $B|AC$ satisfies one of the following conditions*

1. $P(B|C) = 0$, $P(B|AC) = 0$,
2. $P(B|C) = 1$, $P(B|AC) = 1$,
3. $0 < P(B|C) < 1$, $0 < P(B|AC) < 1$;

(c) $P(A|BC) = 1$ *and the extension of P to $B|C$ and $B|A^cC$ satisfies one of the following conditions*

1. $P(B|C) = 0$, $P(B|A^cC) = 0$,
2. $P(B|C) = 1$, $P(B|A^cC) = 1$,
3. $0 < P(B|C) < 1$, $0 < P(B|A^cC) < 1$.

Indeed, in [16] the definition of cs-independence has been extended to the case of finite sets of events and to finite random variables.

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 $\mathcal{E}_1 \perp\!\!\!\perp_{cs} \mathcal{E}_2 | \mathcal{E}_3\ [P]$) iff $C_{i_1} \perp\!\!\!\perp_{cs} C_{i_2} | C_{i_3}\ [P]$ 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 $C_{i_2} \wedge C_{i_3} \neq \emptyset$.*

Let $X = (X_1, \ldots, X_n)$ 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 $\mathcal{E}_X$, $\mathcal{E}_Y$, $\mathcal{E}_Z$ be 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 $X \perp\!\!\!\perp_{cs} Y | Z\ [P]$) iff $\mathcal{E}_X \perp\!\!\!\perp_{cs} \mathcal{E}_Y | \mathcal{E}_Z\ [P]$.*

Note that in Definition 4 it is not required that the domain of the random vector (X, Y, Z) must be $R = R_X \times R_Y \times R_Z$, so logical constraints among the variables can be considered.

The set $\mathcal{M}_P$ of cs-independence statements induced by a coherent conditional probability P of the form $X_I \perp\!\!\!\perp_{cs} X_J | X_K$, where I, J and K are three disjoint subsets, is called *cs-independence model*.

Every cs-independence model induced by P is closed with respect to the following properties (for the proof see [16]):

Decomposition property

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

Reverse decomposition property

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

Weak union property

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

Contraction property

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

Reverse contraction property

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

Intersection property

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

Reverse intersection property

$X_I \perp\!\!\!\perp_{cs} X_W | [X_J, X_K] \, [P] \,\&\, X_J \perp\!\!\!\perp_{cs} X_W | [X_I, X_K] \; [P] \Rightarrow \; [X_I, X_J] \perp\!\!\!\perp_{cs} X_W | [X_K] \, [P]$.

Hence, these models satisfy all graphoid properties (see [14],[15]) except the symmetry property

$$X_I \perp\!\!\!\perp_{cs} X_J | X_K \; [P] \Rightarrow \; X_J \perp\!\!\!\perp_{cs} X_I | X_K \; [P]$$

and reverse weak union property

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

In [16] the models closed with respect to reverse weak union property, but not necessarily with respect to symmetry, (called *a-graphoid*) were classified. The possible lack of symmetry is not counterintuitive (see [4, 6]). Obviously, when the probability P is strictly positive on possible events, the cs-independence model induced by P is closed with respect to graphoid properties.

3 Basic graphical concepts

A *l-graph* G is a triplet $(V, E, \mathcal{B})$, where V is a finite set of *vertices*, E is a set of *edges* (i.e. a subset of ordered pairs of distinct vertices of $V \times V \setminus \{(v,v) \; : v \in V\}$) and $\mathcal{B}$ is a family (possibly empty) of subsets of vertices. The elements of the family $\mathcal{B} = \{B \;, \; B \subseteq V\}$ are represented graphically by boxes enclosing the vertices in B. If $\mathcal{B}$ is empty, then the l-graph is a graph.

The attention in the sequel will be focused on *directed acyclic* l-graphs, and to introduce this kind of l-graphs we need to recall some basic notion from graph theory. A directed l-graph is a l-graph whose set of vertices E satisfies the following property: $(u,v) \in E \Rightarrow (v,u) \notin E$. A directed edge $(u,v) \in E$ is represented by an arrow pointing from u to v, $u \to v$. We say that u is a *parent* of v and v a *child* of u. The set of parents of v is denoted by $pa(v)$ and the set of children of u by $ch(u)$.

A *path* from u to v is a sequence of distinct vertices $u = u_1, \ldots, u_n = v$, $n \geq 1$ such that either $u_i \to u_{i+1}$ or $u_{i+1} \to u_i$ for $i = 1, \ldots, n-1$. A *directed path* from u to v is a sequence $u = u_1, \ldots, u_n = v$ of distinct vertices such that $u_i \to u_{i+1}$ for all $i = 1, \ldots, n-1$. If there is a directed path from u to v, we say that u is an ancestor of v or v a descendant of u and we write $u \mapsto v$. The symbols $an(v)$ and $ds(u)$ denote the set of *ancestors* of v and the set of *descendants* of u (vertices that $u \in an(v)$ and $v \in ds(u)$), respectively. Note that, according to our definition, a sequence consisting of one vertex is a directed path of length 0, and therefore every vertex is its own descendent and ancestor, i.e. $u \in an(u), u \in ds(u)$.

A *reverse directed path* from u to v is a sequence $u = u_1, \ldots, u_n = v$ of distinct vertices such that $u_i \leftarrow u_{i+1}$ for all $i = 1, \ldots, n-1$.

A *n-cycle* is a sequence of $u_1, \ldots, u_n$, with $n > 3$, such that $u_n \to u_1$ and $u_1, \ldots, u_n$ is a directed path. A directed graph is *acyclic* if it contains no cycles.

Given an acyclic directed graph G, the relation $\mapsto$ defines a *partial ordering* $\prec_G$ on the set of vertices, in particular for any $u, v \in V$ we have that if $u \in an(v)$, then $u \prec_G v$, while if $u \in ds(v)$, then $v \prec_G u$.

3.1 L-graphs and logical constraints

In Section 2 the relationship between logical independence and stochastic cs-independence has been shown, so we need to visualize which variables are linked by a logical constraint, and for this purpose we refer to the family $\mathcal{B}$ of subsets of vertices. Since, given a random vector $X = (X_1, \ldots, X_n)$, a vertex i is associated with each random variable X_i, by means of the boxes $B \in \mathcal{B}$, we visualize the sets of random variables linked by a logical constraint (more precisely, a logical constraint involves the events of the partitions generated by the random variables). Recall that the partitions $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are *logically independent* if for every choice $C_i \in \mathcal{E}_i$, with $i = 1, ..., n$, the conjunction $C_1 \wedge \ldots \wedge C_n \neq \emptyset$.

Obviously, if n partitions are logically independent, then arbitrary subsets of these partitions are logically independent.

However, n partitions $\mathcal{E}_1, \ldots, \mathcal{E}_n$ need not be logically independent, even if every $n-1$ partitions can be logically independent; it follows that there is a *logical constraint* such that an event of the kind $C_1 \wedge \ldots \wedge C_n$ is impossible, 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 that partitions are logically independent, but they are not logically independent. Actually, the partition $\mathcal{E}_1$ is not logically independent of the partition 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 some logical constraints among such partitions, it is possible, for each constraint, to find the *minimal subset* $\{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$ of partitions generating it. Actually, $\mathcal{E}_1, \ldots, \mathcal{E}_k$ are such that $C_1 \wedge \ldots \wedge C_k = \emptyset$, with $C_i \in \mathcal{E}_i$, and, in addition, for all $j = 1, \ldots, k$, $C_1 \wedge \ldots \wedge C_{j-1} \wedge C_{j+1} \wedge \ldots \wedge C_k \neq \emptyset$. Such set of partitions $\{\mathcal{E}_1, \ldots, \mathcal{E}_k\}$ is said the *minimal set* generating the given logical constraint, and it is singled-out graphically by the box $B = \{1, \ldots, k\}$, which includes exactly the vertices associated to the corresponding random variables $X_1, \ldots, X_k$. Then, in the sequel we call the boxes $B \in \mathcal{B}$ *logical components*.

4 Separation criterion for directed acyclic graphs

To represent conditional cs-independence models we need to recall L-separation criterion . In fact, the classic separation criterion for directed acyclic graphs (see [14]), known as d-separation (where d stands for directional), is not suitable for our purposes, because it induces a graphoid structure, and so it is not useful to describe a model where symmetry property may not hold (see Example 1).

Definition 5 *Let G be an acyclic directed graph. A path $u_1, \ldots, u_n$, $n \geq 1$ in G is blocked by a set of vertices $S \subset V$, whenever there exists $1 < i < n$ such that one of the following three condition holds:*

1. $u_{i+1} \rightarrow u_i \rightarrow u_{i-1}$ *(i.e. u_{i-1}, u_i, u_{i+1} is the reverse directed path) and* $u_i \in S$
2. $u_{i-1} \leftarrow u_i \rightarrow u_{i+1}$ *and* $u_i \in S$
3. $u_{i-1} \rightarrow u_i \leftarrow u_{i+1}$ *and* $ds(u_i) \notin S$

The three conditions of Definition 5 are illustrated by Figure 1 (the grey vertices belong to S).

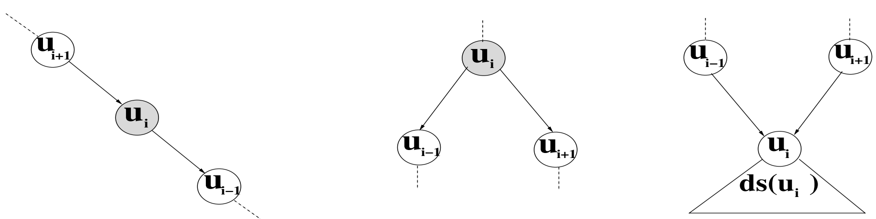


Figure 1: Blocked paths

Note that the definition of blocked path strictly depends on the direction of the path, in fact the main difference between our notion and that used in d-separation criterion [14] consists essentially in condition 1. of Definition 5. The path u_{i-1}, u_i, u_{i+1} drawn in the left-side of Figure 1 is blocked by u_i, while its reverse is not blocked by u_i because of the direction. Hence, the reverse path of a blocked one is not necessarily blocked according to our definition, so the blocking path notion does not satisfy the symmetry property.

The second and third cases of Definition 5 are like in d-separation criterion.

Definition 6 *Let G be a directed acyclic l-graph and let U, W and S be three pairwise disjoint sets of vertices of V. We say that U is L-separated from W by S in G and write symbol $(U,W|S)^l_G$, whenever every path in G from U to W is blocked by S and moreover, the following "logical separation" condition holds*

$$\forall B \in \mathcal{B} \text{ s.t. } B \subseteq U \cup W \cup S \text{ one has either } B \cap U = \emptyset \text{ or } B \cap W = \emptyset. \tag{2}$$

Figure 2 clarifies when condition (2) holds (the set of vertices V_i and S are represented as ovals).

Figure 2: Representation of logical components: in the left-side V_1 and V_2 are not connected, in the right-side they are connected by B

Since the notion of blocked path is not necessarily symmetric, it follows that $(U,W|S)^l_G \not\Rightarrow (W,U|S)^l_G$. Actually, the lack of symmetry property depends on the notion of blocked path and not on the condition of logical separation (2).

Theorem 3 *[17] Let $G = (V,E,\mathcal{B})$ be a graph. The following properties hold*

1. *(Decomposition property)*
$$(U,W \cup Z|S)^l_G \Longrightarrow (U,W|S)^l_G$$
2. *(Reverse decomposition property)*
$$(U \cup Z,W|S)^l_G \Longrightarrow (U,W|S)^l_G$$
3. *(Weak union property)*
$$(U,W \cup Z|S)^l_G \Longrightarrow (U,W|Z \cup S)^l_G$$
4. *(Reverse weak union property)*
$$(U \cup Z,W|S)^l_G \Longrightarrow (U,W|Z \cup S)^l_G.$$
5. *(Contraction property)*
$$(U,W|S)^l_G \;\&\; (U,Z|W \cup S)^l_G \Longrightarrow (U,W \cup Z|S)^l_G$$
6. *(Reverse contraction property)*
$$(U,W|S)^l_G \;\&\; (Z,W|U \cup S)^l_G \Longrightarrow (U \cup Z,W|S)^l_G$$
7. *(Intersection property)*
$$(U,W|Z \cup S)^l_G \;\&\; (U,Z|W \cup S)^l_G \Longrightarrow (U,W \cup Z|S)^l_G$$
8. *(Reverse intersection property)*
$$(U,W|Z \cup S)^l_G \;\&\; (Z,W|U \cup S)^l_G \Longrightarrow (U \cup Z,W|S)^l_G$$

5 Minimal I-map

Given an independence model $\mathcal{M}$ over a set of variables (possibly) linked by a set of logical constraints, we look for a directed acyclic l-graph G describing all the statements T in $\mathcal{M}$ and localizing the set of variables involved in some logical constraint. But, generally, it is not always feasible to have such graph G (i.e. describing all the independence statements) for a given $\mathcal{M}$ as shown by the following example.

Example 1. Let (X_1, X_2, X_3, X_4) be a random vector such that the range of X_i is $\{0, 1\}$, let us denote $A_i = (X_i = 1)$ (so $A_i^c = (X_i = 0)$), and suppose that $A_1 \subset A_2$.

Consider the following coherent conditional probability

$P(A_1A_2) = \frac{1}{5}$, $P(A_1^cA_2) = \frac{3}{10}$, $P(A_1^cA_2^c) = \frac{1}{2}$,

$P(A_3A_4|A_1A_2) = P(A_3A_4|A_1^cA_2) = P(A_3A_4^c|A_1A_2) = P(A_3A_4^c|A_1^cA_2) = 0$,

$P(A_3^cA_4|A_1A_2) = \frac{2}{5} = P(A_3^cA_4|A_1^cA_2)$,

$P(A_3^cA_4^c|A_1A_2) = \frac{3}{5} = P(A_3^cA_4^c|A_1^cA_2)$,

$P(A_4|A_2A_3) = \frac{2}{5}$, $P(A_4|A_2^cA_3) = \frac{3}{20}$, $P(A_2|A_3) = \frac{1}{5}$,

$P(A_1|A_2A_3A_4) = \frac{1}{2}$, $P(A_1|A_2A_3A_4^c) = \frac{2}{5}$.

Since $P(A_1|A_2) = \frac{2}{5}$, it follows from condition (b) 3. of Theorem 2 the validity of the statements $A_3A_4 \perp\!\!\!\perp_{cs} A_1|A_2$ and $A_3A_4^c \perp\!\!\!\perp_{cs} A_1|A_2$; moreover from condition (a) of the same theorem it follows that also $A_3^cA_4 \perp\!\!\!\perp_{cs} A_1|A_2$ and $A_3^cA_4^c \perp\!\!\!\perp_{cs} A_1|A_2$ hold, so we have (by Definition 3 and Definition 4) that $(X_3, X_4) \perp\!\!\!\perp_{cs} X_1|X_2$.

While, the statement $X_1 \perp\!\!\!\perp_{cs} (X_3, X_4)|X_2$ does not hold under P, in fact we have $P(A_1|A_2A_3A_4) = \frac{1}{2} \neq P(A_1|A_2)$.

The validity of the two conditional independence statements $X_3 \perp\!\!\!\perp_{cs} X_4|X_2$ and $X_4 \perp\!\!\!\perp_{cs} X_3|X_2$ follows from these equalities $P(A_3|A_2A_4) = 0 = P(A_3|A_2A_4^c)$ and $P(A_4|A_2) = 0.4 = P(A_4|A_2A_3) = P(A_4|A_2A_3^c)$.

Note that $P(A_3|A_4) = 0 = P(A_3|A_4^c)$ and $P(A_4) = 0.2 = P(A_4|A_3) = P(A_4|A_3^c)$, so $X_4 \perp\!\!\!\perp_{cs} X_3$ and its symmetric statement hold under P.

Therefore, the independence model $\mathcal{M}_P$ (which has a-graphoid structure) contains the statements $(X_3, X_4) \perp\!\!\!\perp_{cs} X_1|X_2$, $X_3 \perp\!\!\!\perp_{cs} X_4|X_2$, $X_3 \perp\!\!\!\perp_{cs} X_4|(X_1, X_2)$; $X_4 \perp\!\!\!\perp_{cs} X_3|X_2$, $X_4 \perp\!\!\!\perp_{cs} X_3|(X_1, X_2)$, $X_3 \perp\!\!\!\perp_{cs} X_4$, $X_4 \perp\!\!\!\perp_{cs} X_3$.

Note that $\mathcal{M}_P$ is not completely representable by a directed acyclic l-graph.

Hence, we need to introduce, analogously as in [14], the notion of I-map.

Definition 7 *A directed acyclic l-graph G is an I-map for a given independence model $\mathcal{M}$ iff every independence statement represented by means of L-separation criterion in G is also in $\mathcal{M}$.*

Thus an I-map G for $\mathcal{M}$ may not represent every statement of $\mathcal{M}$, but the ones it represents are actually in $\mathcal{M}$, it means that the set $\mathcal{M}_G$ of statements described by G is contained in $\mathcal{M}$.

An I-map G for $\mathcal{M}$ is said *minimal* if removing any arrow from the l-graph G the obtained l-graph will no longer be an I-map for $\mathcal{M}$.

Given an independence model $\mathcal{M}$ over a random vector $(X_1,...,X_n)$, let $\pi = (\pi_1,...,\pi_n)$ be any ordering of the given variables, and, in addition, for any j, let $U_{\pi_j} = \{\pi_1,...,\pi_{j-1}\}$ be the set of indexes before π_j, and D_{π_j} the minimal subset of U_{π_j} such that $X_{\pi_j} \perp\!\!\!\perp_{cs} X_{R_{\pi_j}} | X_{D_{\pi_j}}$ where $R_{\pi_j} = U_{\pi_j} \setminus D_{\pi_j}$; moreover, let $W_{\pi_j} = \{v \in U_{\pi_j} : v \in D_{\pi_k} \cap D_{\pi_i}, i \neq k, i \leq j, k \leq j\}$ and S_{π_j} the maximal subset of U_{π_j} such that $X_{S_{\pi_j}} \perp\!\!\!\perp_{cs} X_{\pi_j} | X_{W_{\pi_j}}$.

The subset $\Theta_\pi = \{X_{\pi_j} \perp\!\!\!\perp_{cs} X_{R_{\pi_j}} | X_{D_{\pi_j}}, X_{S_{\pi_j}} \perp\!\!\!\perp_{cs} X_{\pi_j} | X_{W_{\pi_j}} : j = 1,...n\}$ is said the *basic list* of $\mathcal{M}$ relative to π. From the basic list Θ_π and the set of logical components $\mathcal{B}$, a directed acyclic l-graph G (related to π) is obtained by drawing the boxes $B \in \mathcal{B}$ and designating D_{π_j} as parents of vertex π_j (for any vertex $v \in D_{\pi_j}$, an arrow goes from v to π_j), moreover, for any vertex $\pi_i \in U_{\pi_j} \setminus S_{\pi_j}$ such that $\pi_i \in ds(w)$, with $w \in W_{\pi_j}$, but $\pi_i \notin an(\pi_j)$ draw an arrow from π_i to π_j.

This construction of G from the basic list differs from the classic construction given for directed acyclic graphs with d-separation [14] essentially for the second part, which is useful to avoid the introduction of symmetric statements not in the given independence model. For example, consider the independence model $\mathcal{M} = \{X_1 \perp\!\!\!\perp_{cs} X_3 | X_2\}$ and considering the ordering $\pi = (2,3,1)$, the related directed acyclic l-graph is obtained following these steps: draw an arrow from 2 to 3, then consider the vertex 1 and draw an arrow from 2 to 1; now since $3 \in ds(2)$ (i.e. $D_3 = \{2\}$), but $3 \notin an(1)$ and, since the statement $X_3 \perp\!\!\!\perp_{cs} X_1 | X_2$ is not in $\mathcal{M}$, we must draw an arrow from 3 to 1.

Now, we must prove that such directed acyclic l-graph obtained from the basic list Θ_π is an I-map for $\mathcal{M}$.

Theorem 4 *Let $\mathcal{M}$ be an independence model over a set of random variables linked by a set of logical constraints. Given an ordering π on the random variables, if $\mathcal{M}$ is an a-graphoid, then the directed acyclic l-graph G generated by the basic list Θ_π is an I-map for $\mathcal{M}$.*

Proof: For an a-graphoid of one variable it is obvious that the directed acyclic l-graph is an I-map. Suppose for a-graphoid structure with less than k variables that the directed acyclic l-graph is an I-map.

Let $\mathcal{M}$ be an independence model under k variables. Given an ordering π on the variables, let X_n be the last variable according to π (n denotes the vertex in G associated to X_n), $\mathcal{M}'$ the a-graphoid formed by removing all the independence statements involving X_n from $\mathcal{M}$ and G' the directed acyclic l-graph formed by removing n and all the arrows going to n (they cannot depart from n because is the last vertex) in G.

Since X_n is the last variable in the ordering π, it cannot appear in any set of parents D_{π_j} (with $j < k$), and the basic list $\Theta' = \Theta \setminus \{X_n \perp\!\!\!\perp_{cs} X_{R_n} | X_{D_n}\}$ generates $\mathcal{G}'$. Since $\mathcal{M}'$ has $k-1$ variables, $\mathcal{G}'$ is an I-map of it.

G is an I-map of $\mathcal{M}$ iff the set $\mathcal{M}_G$ of the independence statements represented in G by L-separation criterion is also in $\mathcal{M}$.

If X_n does not appear in T, then, being $T = (X_I \perp\!\!\!\perp_{cs} X_J | X_K) \in \mathcal{M}_G$, T must be represented also in $\mathcal{G}'$, if it were not, then there would be a path in $\mathcal{G}'$ from I to J that is not blocked (according to L-separation) by K. But then it must be not blocked also in G, since the addition of a vertex and some arrows going to the new vertex cannot block a path. Since $\mathcal{G}'$ is an I-map for of $\mathcal{M}'$, T must be an element of it, but $\mathcal{M}' \subset \mathcal{M}$, so $T \in \mathcal{M}$.

Otherwise (if X_n appears in T), T falls into one of the following three situations:

1. suppose that $T = ((X_I, X_n) \perp\!\!\!\perp_{cs} X_J | X_K) \in \mathcal{M}_G$, let $X_n \perp\!\!\!\perp_{cs} X_{R_n} | X_{D_n} \in \mathcal{M}$ (by construction). Obviously J and D_n have no vertices in common, otherwise we would have a path from a vertex in $j \in J \cap D_n$ pointing to n, so by L-separation n would not be separated from J given K in G.

   Since there is an arrow from every vertex in D_n to n and every path from n to J is blocked by K in G, then every path from D_n to J must be blocked by K in G. Therefore, every path from both D_n and I to J are blocked by K in G. Now, if there is a logical component $B \in \mathcal{B}$ such that $B \subseteq D_n \cup I \cup J \cup K$ and both $B \cap (D_n \cup I)$ and $B \cap J$ are not empty, then remove a suitable vertex in B from D_n, w.l.g. Hence, the statement $(X_I, X_{D_n}) \perp\!\!\!\perp_{cs} X_J | X_K$ belongs to $\mathcal{M}_G$. This statement does not contain the variable X_n, hence, being G' an I-map for $\mathcal{M}' \subset \mathcal{M}$, then $(X_I, X_{D_n}) \perp\!\!\!\perp_{cs} X_J | X_K \in \mathcal{M}$.

   Since $\mathcal{M}$ is closed under a-graphoid properties, (by weak union property) $X_n \perp\!\!\!\perp_{cs} X_J | (X_I, X_{D_n}, X_K) \in \mathcal{M}$ and it follows $(X_I, X_{D_n}, X_n) \perp\!\!\!\perp_{cs} X_J | X_K \in \mathcal{M}$ (using reverse contraction property), so $(X_I, X_n) \perp\!\!\!\perp_{cs} X_J | X_K \in \mathcal{M}$ by decomposition property.

2. suppose that $T = (X_I \perp\!\!\!\perp_{cs} (X_J, X_n) | X_K) \in \mathcal{M}_G$, it means, by definition of L-separation and from the assumption that n is the last vertex in the ordering, that every path going from I to $J \cup n$ is L-separated by K. Therefore, if there is no path as in condition 1. of Definition 5, then in the remaining two cases, also the statement $T_1 = ((X_J, X_n) \perp\!\!\!\perp_{cs} X_I | X_K) \in \mathcal{M}_G$, so the proof goes in the same line of that in step 1.

   Otherwise, (if there is a path as in condition 1 of Definition 5), then $I \not\subseteq an(n)$. Therefore, there is a subset $W_n \subseteq U_n$ such that every path between n and $I \cup K$ is blocked by W_n. Note that, $W_n = W^1 \cup W^2$ (W^1 or W^2 can be empty) with $W^2 \subseteq D_n$ and $W^1 \subseteq an(D_n)$. Moreover, let $J = J^1 \cup J^2 \cup J^3$ (J^1 or J^2 or J^3 can be empty) with $J^1 \subseteq ds(W) \cap an(K)$, while $J^2 \subseteq W$ and $J^3 = J \setminus (J^1 \cup J^2)$, so for any $j \in J^3$ one has that either $j \in an(W)$ or $j \in ds(W) \cap an(n)$.
   By construction, one has that every path between $n \cup J^3$ and $I \cup K \cup J^1$ is

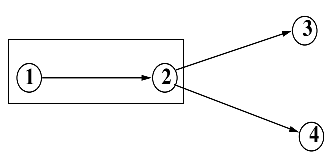
blocked by W_n. Hence, one has that $(X_I, X_K, X_{J^1}) \perp\!\!\!\perp_{cs} (X_n, X_{J^3}) | X_{W_n}$ and its symmetric statement belong to $\mathcal{M}$.

Therefore, one has $X_I \perp\!\!\!\perp_{cs} (X_n, X_{J^3}) | (X_{W_n}, X_K, X_{J^1}) \in \mathcal{M}$ by weak union property. Since also $T_2 = (X_I \perp\!\!\!\perp_{cs} (X_{W_n}, X_{J^1}) | X_K) \in \mathcal{M}_G$ and since that statement T_2 does not involve n, $T_2 \in \mathcal{M}$, so the statement $X_I \perp\!\!\!\perp_{cs} (X_n, X_{W_n}, X_{J_1}, X_{J_3}) | X_K)$ belong to $\mathcal{M}$ (by contraction property), and it follows that $X_I \perp\!\!\!\perp_{cs} (X_n, X_J) | X_K$ belongs to $\mathcal{M}$ (by reverse decomposition).

3. suppose that $T = (X_I \perp\!\!\!\perp_{cs} X_J | (X_K, X_n)) \in \mathcal{M}_G$. It must be the case that I is L-separated by J given K in G for if it were not, then there would be a path from some vertex in I to some vertex in J not passing trough K. But I is separated by J given n and K, so this path would pass through n; but n is the last vertex in the ordering, so all arrows go on it. Hence, it cannot block any unblocked path, and so $T_1 = (X_I \perp\!\!\!\perp_{cs} X_J | X_K) \in \mathcal{M}_G$.

   The statements T_1 and T imply that either $(X_I, X_n) \perp\!\!\!\perp_{cs} X_J | X_K$ or $X_I \perp\!\!\!\perp_{cs} (X_J, X_n) | X_K$ holds in G: in fact, if both I and J are connected to n, since n is the last vertex (from n an arrow cannot leave), then there is a directed path from I to n and another from J to n, so that one would get $X_I \perp\!\!\!\perp_{cs} X_J | (X_K, X_n) \notin \mathcal{M}_G$. So, the conclusion follows by step 1 and 2. □

Example 1 (continued) – The following pictures show the minimal I-map obtained by means of the proposed procedure for two possible orderings: $(1,2,3,4)$ on the left-side and $(3,4,1,2)$ on the right-side



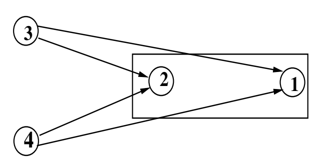


Figure 3: Two possible I-Maps for the independence model $\mathcal{M}_P$ of Example 1

Actually, the picture in the left-side represents the independence statements $(X_3, X_4) \perp\!\!\!\perp_{cs} X_1 | X_2$, $X_3 \perp\!\!\!\perp_{cs} X_4 | X_2$, $X_4 \perp\!\!\!\perp_{cs} X_3 | X_2$ and those implied by a-graphoid properties; while that one on the right-side describes the statement $X_3 \perp\!\!\!\perp_{cs} X_4$ and its symmetric one. Note that these two graphs actually are minimal I-maps; in fact removing any arrow from them, we may read independence statements not in $\mathcal{M}_P$. The block $B = \{1,2\}$ localizes the logical constraint $A_1 \subset A_2$.

If for a given independence model over n variables there exists a perfect map G, then (at least) one of $n!$ orderings among the variables will generate the l-graph G. More precisely, such orderings, which give rise to G, are all the orderings compatible with the partial order induced by G.

6 Conclusions

The L-separation criterion for directed acyclic graphs has been recalled together with its main properties. This is very useful for effective description of independence models induced by different uncertainty measures [1, 2, 4, 5, 6, 13, 16, 18, 19]. In fact, these models cannot be represented efficiently by the well-known graphical models [12, 14], because the related separation criteria satisfy the symmetry property.

In this paper, we have considered the L-separation criterion introduced in [16], which satisfies asymmetric graphoid properties. We have shown that for some independence models there is not a perfect map even using L-separation criterion.

Therefore, the notion of minimal I-map has been redefined in this context and we have shown how to build it given an ordering on the variables. In addition, we have proved that for any ordering on the variables there is a minimal I-map for a given independence model obeying to asymmetric graphoid properties.

References

[1] B. Bouchon-Meunier, G. Coletti, C. Marsala. Independence and Possibilistic Conditioning. *Annals of Mathematics and Artificial Intelligence*, 35:107–124, 2002.

[2] L. de Campos, S. Moral. Independence Concepts for Convex Sets of Probabilities. In *Proc. XI Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, CA, 108-115, 1995.

[3] G. Coletti. Coherent numerical and ordinal probabilistic assessments. *IEEE Trans. on Systems, Man, and Cybernetics*, 24(12):1747-1754, 1994.

[4] G. Coletti, R. Scozzafava. Zero probabilities in stochastic independence. In *Information, Uncertainty, Fusion* (Eds. B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh), Kluwer, Dordrecht 185-196, 2000. (Selected papers from IPMU '98).

[5] G. Coletti, R. Scozzafava. Stochastic Independence in a Coherent Setting. *Annals of Mathematics and Artificial Intelligence*, 35:151–176, 2002.

[6] G. Coletti, R. Scozzafava. *Probabilistic logic in a coherent setting* (Trends in logic n.15, Kluwer, Dordrecht/Boston/London) 2002.

[7] F.T. de Dombal, F. Gremy. *Decision Making and Medical Care.* North Holland, 1976.

[8] B. de Finetti. Sull'impostazione assiomatica del calcolo delle probabilitá. *Annali dell'Universitá di Trieste*, 19:3-55, 1949. (Eng. transl.: Ch. 5 in Probability, Induction, Statistics, London: Wiley, 1972).

[9] L.E. Dubins. Finitely additive conditional probabilities, conglomerability and disintegration. *Annals of Probability*, 3:89-99, 1975.

[10] D. Geiger, T. Verma, J. Pearl. Identifying Independence in Bayesian Networks. *Networks*, 20:507-534, 1990.

[11] J.R. Hill. Comment on "Graphical models". *Statistical Science*, 8:258-261, 1993.

[12] S.L. Lauritzen. *Graphical models*, Clarendon Press, Oxford, 1996.

[13] S.Moral. Epistemic irrelevance on sets of desiderable gambles. In *Proc. of the Fird Int. Symposium on Imprecise Probabilities and Their Applications* (Eds. G. de Cooman, T.L. Fine, T. Seidenfeld), 247-254, 2001.

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

[15] M. Studeny, R.R. Bouckaert. On chain graph models for description of conditional independence structures. *The Annals of Statistics*, 26(4):1434-1495, 1998.

[16] B. Vantaggi. Conditional Independence in a Coherent Finite Setting. *Annals of Mathematic and Artificial Intelligence*, 32:287-314, 2001.

[17] B. Vantaggi. The L-separation criterion for description of cs-independence models. *International Journal of Approximate Reasoning*, 29:291-316, 2002.

[18] B. Vantaggi. Graphical models for conditional independence structures. In *Proc. of the Fird Int. Symposium on Imprecise Probabilities and Their Applications* (Eds. G. de Cooman, T.L. Fine, T. Seidenfeld), 332-341, 2001.

[19] J. Vejnarova. Conditional Independence Relations in Possibility Theory. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 8(3):253-269, 2000.

Barbara Vantaggi is with the Department of Metodi e Modelli Matematici per le Scienze Applicate at Universitá "La Sapienza" di Roma, Italy. E-mail: vantaggi@dmmm.uniroma1.it