On stochastic conditional independence: the problems of characterization and description

Milan Studený *

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, Prague, Czech Republic
E-mail: studeny@utia.cas.cz

The topic of this survey are structures of stochastic conditional independence. Two basic questions are dealt with: the problem of characterization of conditional independence models and the problem of their mathematical description and computer representation. Basic formal properties of conditional independence are recapitulated and the problem of axiomatic characterization of stochastic conditional independence models is mentioned. Classic graphical methods of description of these structures are recalled, in particular the method which uses chain graphs. Limitation of graphical approaches motivated an attempt at a non-graphical approach. A certain method of description of stochastic conditional independence models which uses non-graphical tools called 'structural imsets' is outlined.

Keywords: conditional independence, chain graphs, structural imsets.

AMS Subject classification: 68T30, 62H05.

1. Introduction

Motivation of this paper stems from general significance of the concept of *conditional independence* (CI). Classic concept of stochastic independence of two random vectors $[\xi_i]_{i\in A}$ and $[\xi_i]_{i\in B}$ has interpretation of their mutual *irrelevance* which means that the values $[\xi_i]_{i\in B}$ have no relevance to the task of estimation of the values of $[\xi_i]_{i\in A}$ and conversely.

The concept of conditional independence has similar interpretation. Consider a random vector $[\xi_i]_{i\in N}$ and suppose that $A, B, C \subseteq N$ are pairwise disjoint sets. Then the situation when $[\xi_i]_{i\in A}$ is conditionally independent of $[\xi_i]_{i\in B}$ given $[\xi_i]_{i\in C}$ can be interpreted in terms of relevance as follows. Knowing the values of $[\xi_i]_{i\in C}$ the values of $[\xi_i]_{i\in A}$ and the values of $[\xi_i]_{i\in B}$ have no relevance each other. Alternative interpretation is in terms of 'decomposition' which means that the system of variables $[\xi_i]_{i\in A\cup B\cup C}$ can be decomposed without loss of information

* This research has been supported by the grants GAČR n. 201/01/1482 and GAAVČR n. A1075104.

into two subsystems, namely $[\xi_i]_{i\in A\cup C}$ and $[\xi_i]_{i\in B\cup C}$. More exactly, the joint distribution of $[\xi_i]_{i\in A\cup B\cup C}$ is recoverable from its marginals for $A\cup C$ and $B\cup C$ by means of a certain product formula. Thus, CI describes certain *qualitative* relation among variables and every CI statement represents a piece of general 'structural' information about a stochastic system.

Formal properties of stochastic CI were widely studied in *probabilistic reasoning* which is an area of artificial intelligence where uncertainty processing is based on probability theory. Moreover, CI appears in two branches of mathematical statistics. For example, in analysis of *contingency tables*, well-known graphical log-linear models can be interpreted in terms of CI (see Chapter 4 of [29]). Similarly, in *multivariate analysis*, certain graphs describing CI structures are associated with well-known structural equation models (see Section 5 of [3]).

But the significance of the concept of CI is not limited to probabilistic (and statistical) framework. This concept has been introduced in several non-probabilistic frameworks, namely in various *calculi for dealing with uncertainty* in artificial intelligence (for more detailed overview see [51,45,13]). Note for illustration that several other contributions in [63] dealt with CI in alternative frameworks, namely with CI in the framework of

- possibility theory [4,14,60],
- coherence theory [7],
- theory of imprecise probabilities [9],
- theory of evidence [5],
- hypergraphs [31].

To conclude this motivation part let me formulate two basic questions of interest of this paper.

1. *The problem of characterization* is the question what are the structures of stochastic CI among a finite number of discrete random variables.

2. *The problem of description* is the question how to describe these structures mathematically to have both the possibility of suitable interpretation of the mathematical tool and the possibility of relevant computer implementation.

Note that the phrase 'CI structure' is used throughout the paper in general sense, mainly in motivation parts. It corresponds to exact mathematical concept of 'CI model' introduced in Definition 4.

The content of the following sections is as follows. Basic concepts are recalled in Section 2, in particular the concept of CI. After preliminaries classic formal properties of stochastic CI known as semi-graphoid properties are recapitulated (Section 3). This is a starting point for a more general question of axiomatic characterization of formal properties of CI mentioned afterwards (Section 4). Then the problem of description of CI structures is treated. Section 5 briefly summarizes classic graphical methods. The need for non-graphical methods is justified in Section 6. Finally, Section 7 describes roughly a certain non-graphical approach which tries to overcome limitation of graphical methods.

2. Basic concepts

Throughout the paper N denotes a non-empty finite set of *factors*. These factors correspond to random variables, say ξ_i, $i \in N$. Every non-empty subset A of N corresponds to a random vector $[\xi_i]_{i \in A}$. Every random variable ξ_i has a certain measurable space $(X_i, \mathcal{X}_i)$ as a frame, that is the set X_i of possible values (= sample space) endowed with σ-algebra $\mathcal{X}_i$. Under this situation $(X_A, \mathcal{X}_A)$ will be used as a shorthand for $(\prod_{i \in A} X_i, \prod_{i \in A} \mathcal{X}_i)$ which serves as a frame for ξ_A. Moreover, let us accept the convention that $(X_\emptyset, \mathcal{X}_\emptyset)$ is (an appended) measurable space endowed with trivial σ-algebra $\mathcal{X}_\emptyset \equiv \{\emptyset, X_\emptyset\}$.

Note that those readers who are not familiar with basic measure-theoretical concepts may consider finite non-empty sets X_i, $i \in N$ instead of measurable spaces $(X_i, \mathcal{X}_i)$, $i \in N$. Respective definitions in this special *discrete case* will be systematically recalled in the sequel.

Given $x = [x_i]_{i \in N} \in X_N$ and $\emptyset \neq A \subseteq N$ the projection of x to X_A is denoted by $x_A \equiv [x_i]_{i \in A}$.

Definition 1. By *probability distribution* over N will be understood any probability measure with domain $(X_N, \mathcal{X}_N)$. Its *marginal* for $\emptyset \neq A \subset N$ is defined by the formula

$$P^A(Y) = P(Y \times X_{N \setminus A}) \quad \text{for} \quad Y \in \mathcal{X}_A .$$

Moreover, $P^N \equiv P$ and $P^\emptyset$ is the unique probability measure on $(X_\emptyset, \mathcal{X}_\emptyset)$ by convention. A probability distribution over N is *marginally continuous* if there exists a collection of σ-finite measures μ_i on $(X_i, \mathcal{X}_i)$, $i \in N$ such that P is absolutely continuous with respect to the product measure $\prod_{i \in N} \mu_i$. Recall that P is absolutely continuous with respect to a σ-finite measure μ on $(X_N, \mathcal{X}_N)$, written by $P \ll \mu$, if $\mu(Y) = 0 \;\Rightarrow\; P(Y) = 0$ for every $Y \in \mathcal{X}_N$. Well-known Radon-Nikodym theorem [44] then implies the existence of density of P with respect to μ, called also Radon-Nikodym derivative.

If the *dominating* measures μ_i, $i \in N$ are fixed, by *marginal density* for $\emptyset \neq A \subseteq N$, denoted by f_A, is understood the Radon-Nikodym derivative of P^A with respect to $\mu_A \equiv \prod_{i \in A} \mu_i$. Moreover, $f_\emptyset(x) = 1$ for any $x \in X_\emptyset$. $\triangle$

It can be shown (see [49], Proposition 1) that P is marginally continuous if it is absolutely continuous with respect to the product of its one-dimensional marginals $\prod_{i \in N} P^{\{i\}}$. An important special case of a marginally continuous distribution is any non-degenerate multidimensional Gaussian distribution where $(X_i, \mathcal{X}_i)$ is the set of real numbers with Borel σ-algebra for every $i \in N$. Moreover, every discrete probability measure over N is automatically marginally continuous since one can always consider the counting measure on X_i in place of μ_i, $i \in N$.

Thus, the reader can accept the following simplified version of the above definition. *Probability density* over N is any non-negative real function p on X_N with $\sum\{p(x)\,;\, x \in X_N\} = 1$. It is called *strictly positive* if $p(x) > 0$ for all $x \in X_N$. The corresponding probability measure P on $(X_N, \mathcal{X}_N)$ where $\mathcal{X}_N$ is the power set of X_N is then

$$P(Y) = \sum_{y \in Y} p(y) \quad \text{for} \quad Y \subseteq X_N\,.$$

The *marginal density* p^A for $\emptyset \neq A \subset N$ is defined as follows

$$p^A(y) = \sum \{\, p(y,z)\,;\, z \in X_{N\setminus A}\,\} \quad \text{for every} \quad y \in X_A\,, \tag{1}$$

and $p^{\emptyset}(*) \equiv 1$ by convention.

Remark 2. Note that the situation described in Definition 1 is not completely general. In fact, 'logical independence' [7] or 'valuation independence' [13] of considered random variables is implicitly assumed which means that fixing on a value of ξ_i, $i \in N$ does not impose any restriction of the range of ξ_j, $i \neq j \in N$. Formally, the joint sample space has the form of a product space. On the other hand, fixing on a value of ξ_i may influence the probability of occurence of values of ξ_j for $j \neq i$ which is modelled by a probability distribution P. Therefore, no restriction concerning 'stochastic independence' of considered random variables is involved in Definition 1 as explained above (arbitrary discrete distribution is marginally continuous).

Note that 'logical contraints', i.e. the situations when fixing on a value of ξ_i excludes the occurence of some values of ξ_j for $j \neq i$ (differently from the case of zero probability of the value ξ_j which still lets the occurence of this value possible - even if improbable) can be modelled by taking a general joint sample space for ξ_i, $i \in N$. However, this necessitates much more general definition of the concept of CI - see Remark 5.

On the other hand, 'classic' Kolgomorovian probabilistic approach which is treated in this paper does not allow to distinguish between different levels of zero probability events. There are other approaches which allow to do so, for example the coherence theory [8]. This approach takes the concept of conditional probability as a primitive concept; the corresponding definition of CI then involves additional requirements in comparison with the definition given below (Definition 4). Nevertheless these two definitions definitely coincide in discrete case with strictly positive probability distributions.

Thus, the concept of marginally continuous distribution is general enough for our purposes. It includes both discrete distributions and typical continuous distributions. Respective CI statements are described by special triplets of sets.

Definition 3. By a *disjoint triplet* over N will be understood any triplet $\langle A, B|C\rangle$ of pairwise disjoint subsets of N. The class of disjoint triplets over N will be denoted by $\mathcal{T}(N)$. $\triangle$

Conditional independence of random vectors ξ_A and ξ_B given ξ_C, where $\langle A, B|C\rangle \in \mathcal{T}(N)$, is formally defined in terms of their joint distribution over $A \cup B \cup C$. The following definition of CI is limited to the case of marginally continuous distributions (for explanation see Remark 5).

Definition 4. Suppose that $\langle A, B|C\rangle \in \mathcal{T}(N)$ and P is a marginally continuous distribution over N. Let $\mu_N \equiv \prod_{i\in N} \mu_i$ be the product of fixed dominating measures. One says that A is *conditionally independent of B given C with respect to P* and writes $A \perp\!\!\!\perp B \,|\, C\ [P]$ if the equality

$$f_{A\cup B\cup C}(x_{A\cup B\cup C}) \cdot f_C(x_C) = f_{A\cup C}(x_{A\cup C}) \cdot f_{B\cup C}(x_{B\cup C}) \tag{2}$$

holds for μ_N-almost every $x \in X_N$ (here f_A, $A \subseteq N$ are marginal densities from Definition 1). Recall that the phrase 'for μ_N-almost every $x \in X_N$' means that $\mu_N\,\{x \in X_N\,;\ (2) \text{ does not hold}\,\} = 0$. This slight technicality cannot be avoided since the densities (Radon-Nikodym derivatives) are determined uniquely only within equivalence 'equality almost everywhere' - see [44].
The *conditional independence model* induced by P is then defined as follows

$$\{\,\langle A, B|C\rangle \in \mathcal{T}(N)\,;\ A \perp\!\!\!\perp B \,|\, C\ [P]\,\}\,.$$

Abbreviation CI model is used in the sequel. $\triangle$

Note that the definition above actually does not depend on the choice of dominating measures (see Section 2.3.1 in [59]). In discrete case, that is in case when X_i, $i \in N$ are finite non-empty sets, the above definition takes the following simple form: one has $A \perp\!\!\!\perp B \,|\, C\ [P]$ iff

$$p^{A\cup B\cup C}(a,b,c) \cdot p^C(c) = p^{A\cup C}(a,c) \cdot p^{B\cup C}(b,c)$$

holds for every $a \in X_A$, $b \in X_B$, $c \in X_C$ (here p^A, $A \subseteq N$ are given by (1) and equality is required for <u>every</u> a, b, c since the equality 'almost everywhere' with respect to the counting measure on X_N is usual equality). Some of the readers may be familiar with alternative equivalent definitions in the discrete case. They are usually formulated in terms of *conditional density* which is defined by

$$p_{A|C}(a,c) = \frac{p^{A\cup C}(a,c)}{p^C(c)} \quad \text{whenever} \quad p^C(c) > 0\,,$$

for disjoint $A, C \subseteq N$. Here is an overview of common equivalent definitions of $A \perp\!\!\!\perp B \,|\, C\ [P]$ in discrete case.

- In terms of conditional density over A given $B \cup C$ and C:

$$p_{A|B\cup C}(a|b,c) = p_{A|C}(a|c)$$

  for every $a \in X_A$, $b \in X_B$, $c \in X_C$ with $p^{B\cup C}(b,c) > 0$.
- In terms of conditional density over $A \cup B$ given C (a descriptive version):

$$p_{A\cup B|C}(a,b|c) = p_{A|C}(a|c) \cdot p_{B|C}(b|c)$$

  for every $a \in X_A$, $b \in X_B$, $c \in X_C$ with $p^C(c) > 0$.
- in terms of conditional density over $A \cup B$ given C (a non-descriptive version): for every $c \in X_C$ with $p^C(c) > 0$ the conditional density $p_{A\cup B|C}(*|c)$ is the product of densities over A and B.
- Factorization property: there exist real functions f on $X_{A\cup C}$ and g on $X_{B\cup C}$ such that

$$p^{A\cup B\cup C}(a,b,c) = f(a,c) \cdot g(b,c)$$

  for every $a \in X_A$, $b \in X_B$, $c \in X_C$.
- Cross-interchangeability condition:

$$p^{A\cup B\cup C}(a,b,c) \cdot p^{A\cup B\cup C}(\overline{a},\overline{b},c) = p^{A\cup B\cup C}(\overline{a},b,c) \cdot p^{A\cup B\cup C}(a,\overline{b},c)$$

  for every $a, \overline{a} \in X_A$, $b, \overline{b} \in X_B$, $c \in X_C$.

It is left to the reader to verify that the conditions above are equivalent in the above mentioned discrete case. The most common definition is the first one which corresponds to interpretation of CI in terms of irrelevance. The factorization property is eloquent, it leads directly to interpretation of CI in terms of decomposition. Special cross-interchangeability condition has specific interpretation from the point of view of statistical physics [39]. Various attempts to introduce the concept of CI in non-probabilistic frameworks are based on analogy with previous forms of definition.

Remark 5. For sake of brevity I omitted general definition of CI (for arbitrary probability distributions over N). Definition of this type is too technical. One has to come back to measure-theoretical groundings in order to recall general concept of (say non-regular) conditional probability with respect to a σ-algebra, which is a special function defined uniquely only within certain equivalence. This enables one to introduce the concept of CI of two σ-algebras given the third one as the equality of certain conditional probabilities within respective equivalence (for details see [15]). CI for random variables is then introduced as CI of respective σ-algebras.

3. Basic formal properties of conditional independence

Several authors independently accentuated some simple basic formal properties of stochastic CI. In modern statistics, the first author was probably Dawid [12] who characterized certain statistical concepts, for example the concept of sufficient statistics, in terms of generalized stochastic CI and used such properties to derive his results in an elegant way, without tedious calculation. Analogous properties, formulated in terms of σ-algebras were treated by other statisticians, for example by Mouchart and Rolin [38]. Spohn [48] studied the concept of CI from the point of view of philosophical logic and formulated the same properties like Dawid. Note that implicit remarks on properties of CI (in alternative frameworks) also appeared in literature. One of reviewers of this paper pointed out that Nambiar [40] interpreted multivariate dependencies treated in theory of relational databases as CI statements.

Finally, the importance of CI in probabilistic reasoning was explicitly discerned and highlighted by Pearl and Paz [42]. They introduced the concept of *semi-graphoid* to name 'formal independence models' (i.e. subsets of $\mathcal{T}(N)$) satisfying these formal properties. These well-known properties are formulated in the next lemma (the proof in discrete case is left to the reader, for general case see Lemma 2.1 in [59]).

Lemma 6. Let P be a probability distribution over N. Supposing $A, B, C, D \subseteq N$ are pairwise disjoint sets the following properties hold.

1. $A \perp\!\!\!\perp \emptyset \,|\, D\ [P]$ triviality,
2. $A \perp\!\!\!\perp B \,|\, D\ [P] \;\Rightarrow\; B \perp\!\!\!\perp A \,|\, D\ [P]$ symmetry,
3. $A \perp\!\!\!\perp B \cup C \,|\, D\ [P] \;\Rightarrow\; A \perp\!\!\!\perp C \,|\, D\ [P]$ decomposition,
4. $A \perp\!\!\!\perp B \cup C \,|\, D\ [P] \;\Rightarrow\; A \perp\!\!\!\perp B \,|\, C \cup D\ [P]$ weak union,
5. $\{\, A \perp\!\!\!\perp B \,|\, C \cup D\ [P]\ \&\ A \perp\!\!\!\perp C \,|\, D\ [P] \,\} \;\Rightarrow\; A \perp\!\!\!\perp B \cup C \,|\, D\ [P]$ contraction.

Note that some authors [43,46] regarded the above properties as suitable axioms for abstract *conditional irrelevance*. Indeed, the most of CI models arising in miscellaneous alternative uncertianty calculi [51] exhibit the same formal properties. Semi-graphoids can be also viewed from purely algebraic point of view which leads to interesting related inference tasks [36]. Their further generalization leads to the concept of *separoid* which is a specific algebraic structure on a join semi-lattice [13].

4. Problem of axiomatic characterization

In general, by the problem of axiomatic characterization is meant the task to characterize CI models over N (see Definition 4) without a reference to 'underlying' probability distributions, namely in terms of properties of formal independence models. Of course, one is interested in formal properties of semi-graphoid type applicable to any set of factors N which can also be interpreted as inference rules of a certain 'formal axiomatic theory' [37] in sense of mathematical logic. For more detailed explanation of preceding sketchy sentence see [50].

Pearl and Paz [42] raised a natural conjecture that semi-graphoids coincide with CI models (induced by discrete distributions). Unfortunately, their conjecture was refuted [49] by finding another property of CI models:

$$\{ A \perp\!\!\!\perp B \,|\, C \cup D \;[P] \;\&\; C \perp\!\!\!\perp D \,|\, A \;[P] \;\&\; C \perp\!\!\!\perp D \,|\, B \;[P] \;\&\; A \perp\!\!\!\perp B \,|\, \emptyset \;[P] \} \quad \Leftrightarrow$$
$$\Leftrightarrow \;\{ C \perp\!\!\!\perp D \,|\, A \cup B \;[P] \;\&\; A \perp\!\!\!\perp B \,|\, C \;[P] \;\&\; A \perp\!\!\!\perp B \,|\, D \;[P] \;\&\; C \perp\!\!\!\perp D \,|\, \emptyset \;[P] \}\,.$$

Note that this result was achieved using some tools of information theory, namely by application of the properties of multiinformation function respectively of entropy function. This information-theoretical point of view (which is also somehow behind the method from Section 7) made it possible to derive later even a stronger result [50]:

Theorem 7. CI models induced by discrete probability distributions cannot be characterized in terms of a finite number of formal properties of semi-graphoid type.

More specifically, for every natural number $n \geq 3$ there exists a formal property of CI (of semi-graphoid type) which applies on a set of factors of cardinality n but which cannot be revealed on a set of less cardinality, namely

$$\{ A \perp\!\!\!\perp B_1 \,|\, B_2 \;\&\; A \perp\!\!\!\perp B_2 \,|\, B_3 \;\&\; \ldots \;\&\; A \perp\!\!\!\perp B_{n-1} \,|\, B_n \;\&\; A \perp\!\!\!\perp B_n \,|\, B_1 \} \quad \Leftrightarrow$$
$$\Leftrightarrow \;\{ A \perp\!\!\!\perp B_2 \,|\, B_1 \;\&\; A \perp\!\!\!\perp B_3 \,|\, B_2 \;\&\; \ldots \;\&\; A \perp\!\!\!\perp B_n \,|\, B_{n-1} \;\&\; A \perp\!\!\!\perp B_1 \,|\, B_n \}\,.$$

Note that, moreover, the properties cannot be derived as a consequence of other formal properties of CI valid on sets of factors of less cardinality.

Remark 8. Although semi-graphoid properties do not characterize stochastic CI models, several results on their relative completeness were achieved. Geiger, Paz, Pearl [21] and independently Matúš [32] showed that classic (unconditional) independence models can be characterized in terms of properties derivable from semi-graphoid properties. The same conclusion holds in full-context case, when disjoint triplets involving all variables in N are treated (see independently achieved results by Geiger and Pearl in [22] and by Malvestuto [30]). Specific relative completeness results concern graphical models represented by acyclic directed graphs [20,46] and models generates by pairs of CI statements [56].

The problem of axiomatic characterization has close connection to the *implication problem* for a distribution framework. By a distribution framework is meant 'consistently' defined class $\mathcal{P}$ of probability measures over N for every set of factors N. Three examples of distribution frameworks are the class of discrete probability distributions, the class of strictly positive discrete probability distributions and the class of non-degenerate Gaussian distributions. Given a set of disjoint triplets $\Sigma \subseteq \mathcal{T}(N)$ its *logical closure* relative to $\mathcal{P}$ is the set of disjoint triplets over N representing CI statements with respect to $\underline{\text{every}}$ $P \in \mathcal{P}$ such that all triplets from Σ are CI statements with respect to $\overline{P}$. The respective implication problem is the task to characterize logical closure with help of formal properties of CI models induced by $P \in \mathcal{P}$, called *axioms* of CI relative to $\mathcal{P}$. The above mentioned results on relative completeness of semi-graphoid properties can be viewed as results solving the implication problem for special distribution frameworks and specific restrictions of the class of CI statements.

5. Graphical methods

The second basic question of this paper is the problem of description of stochastic CI structures. In general, by the problem of description is meant the task to describe (some of) CI models over N (see Definition 4) by means of simpler mathematical objects of discrete mathematics, namely those which offer acceptable interpretation for humans. Traditional methods which comply with requirements for easy interpretability and computer feasibility use graphs whose nodes correspond to factors (= variables). Classic types of graphs used for this purpose admit either directed or undirected edges betwen nodes.

Definition 9. Let H be an undirected graph having N as the set of nodes and $B_1, \ldots, B_n$, $n \geq 1$ is an ordered partition of N into non-empty blocks, called a *chain*. The corresponding *chain graph* (CG) G is made from H by directing every edge in H between nodes from different blocks in direction from the former block to the latter block. Graph obtained in this way is called a chain graph over N. Its directed edges are called *arrows*, undirected edges are called *lines*. $\triangle$

Note that chain graphs can be equivalently introduced as graphs without loops, multiple edges and directed cycles (that is, sequences of nodes $u_1, \ldots, u_n$, $n \geq 4$ with $u_n = u_1$ such that $u_i \rightarrow u_{i+1}$ or $u_i - u_{i+1}$ for $i = 1, \ldots, n$ and where $u_j \rightarrow u_{j+1}$ for at least one $j \in \{1, \ldots, n\}$) - see Lemma 2.1 in [57].

CGs were introduced by Lauritzen and Wermuth in mid-eighties [25], but they had some predecessors [24]. One of the aims of introducing them was to give an unifying point of view on two traditionally separate but analogous graphical approaches. The first was using *undirected graphs* (UGs) named also 'Markov networks' [43] that is graphs allowing lines only. The second was using *acyclic directed graphs* called 'Bayesian networks' [43], that is graphs allowing arrows only and with forbidden directed cycles. The majority of authors in this field became accustomed to the phrase 'directed acyclic graph' and abbreviation DAG which is not completely accurate (since adjectives do not commute). Comparison between classic classes of graphs used for description of CI structures is illustrated by Figure 1.

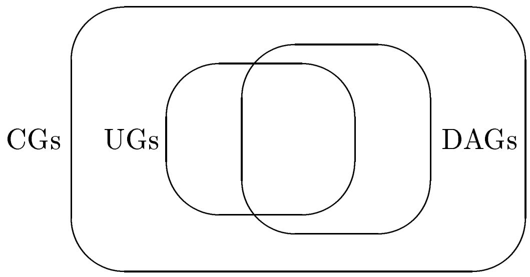

Figure 1. Relationships among classic graphical models.

Remark 10. The aim of this 'historical' remark is to explain the reader how miscellaneous graphical criteria (e.g. Definition 11 below) used to relate (identify) graphs with formal independence models (= subsets of $\mathcal{T}(N)$) and consequently with probability distributions were invented. It has close connection to the implication problem mentioned in Remark 8.

The way of identification of classic graphs with CI structures was a result of certain development. For example, in case of UGs the original interpretation of a line $u — v$ was a certain 'conditional dependence statement'. More exactly, an UG over N was intended to represent a class of strictly positive (discrete) probability distributions over N, called *Markovian distributions* (with respect to the graph). Distribution P of this kind was defined by the requirement that if a pair of nodes $\{u, v\}$ is not a line in the graph, then $\{u\} \perp\!\!\!\perp \{v\} \,|\, N \setminus \{u, v\}$ $[P]$. However, other CI statements appeared to be valid in every Markovian distribtion. Thus, the original 'pairwise Markov condition' was often replaced by stonger 'local Markov condition' (for details see Section 3.2.1 in [29]). Nevertheless, this condition was not strong enough to identify all CI statements valid with respect to every (strictly positive) Markovian distribution. Thus, the development ended with 'global Markov condition' which identifies the whole collection of CI statements shared by (positive) Markovian distributions. These shared CI statements can be identified directly on basis of the graph G by means of *separation criterion for undirected graphs* as follows. The CI statements correspond to those triplets $\langle A, B|C\rangle \in \mathcal{T}(N)$ for which every path in G between A and B has a node in C. The graph G represents in this way the CI structure which is shared by Markovian distributions (with respect to G). Thus, one can regard the graph as a representative of the whole (common) CI structure which is the point of view taken in this paper.

Note that analogous development was observed in case of DAGs (see Section 3.2.2 in [29]). The difference is that the development ended with two different but equivalent criteria to determine the maximal collection of CI statements valid with respect to the corresponding class of Markovian distributions. The group around Lauritzen [28] generalized an incomplete criterion from [24] to a step-wise *moralization criterion*. Here, testing whether a disjoint triplet is represented in the graph consists in transformation of the directed graph into a certain UG, called the moral graph, and then using the separation criterion for undirected graphs. Geiger, Verma and Pearl [18] proposed a direct *d-separation criterion* (d stand for 'directional') which examines paths in the original directed graph whether they are blocked. The definition of blocking at a node of a path depends on direction of arrows 'entering' the node. The criteria were shown to be equivalent in [28].

Interpretation of classic CGs is in accordance with traditional interpretation of UGs and DAGs. Lauritzen [27] introduced *moralization criterion for chain graphs* which simultaneously generalizes the separation criterion for undirected graphs and the moralization criterion for acyclic directed graphs. Note that theory of classic CGs was later developed and deepened by results of Frydenberg [17] (e.g. characterization of equivalent CGs).

Definition 11. Let G be a CG over N and $\langle A, B|C\rangle \in \mathcal{T}(N)$ is a disjoint triplet over N. Testing whether $\langle A, B|C\rangle$ is represented in G according to the moralization criterion for CGs consists of three steps.

1. Restriction of G to the set of ancestors T of $A \cup B \cup C$ in G, that is the set of $u \in N$ such that there exists a sequence $u = u_1, \ldots, u_n$, $n \geq 1$ such that $u_n \in A \cup B \cup C$ and $u_i \to u_{i+1}$ or $u_i \;—\; u_{i+1}$ in G for $i = 1, \ldots, n$. The resulting induced graph is denoted by G_T.

2. Moralization of G_T means that an edge between different nodes u, v of G_T is added if there exists a path $u = w_1 \to w_2 \;—\; \ldots \;—\; w_{m-1} \leftarrow w_m = v$, $m \geq 3$. The middle part of the path is composed of lines but it may consist of a single node. After that direction of possible arrows is forgotten so that the resulting *moral graph* of G_T is an undirected graph over T.

3. If every path between a node in A and a node in B in the moral graph contains a node in C then $\langle A, B|C\rangle$ is represented in G according to the moralization criterion.

Then one writes $A \perp\!\!\!\perp B \,|\, C\ [G]$. $\triangle$

The moralization criterion is simply illustrated by Figure 2. The tested disjoint triplet $\langle a, d \,|\, \{b, e, g\}\rangle$ is not represented in the given CG because the path $a \;—\; f \;—\; c \;—\; d$ in the moral graph is outside the conditioning set.

The next step is to introduce the class of Markovian distributions with respect to a CG.

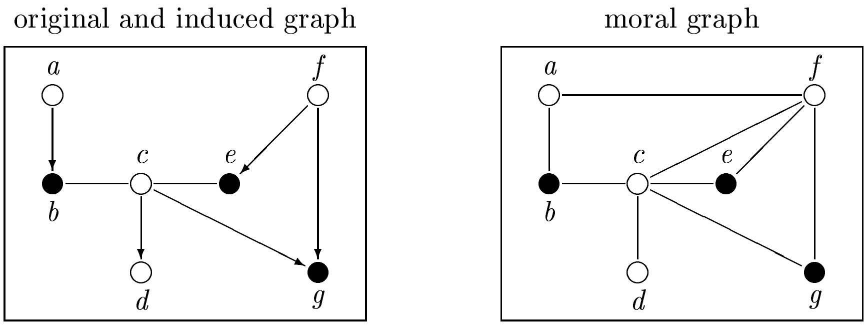

Figure 2. Testing $\langle a, d \,|\, \{b, e, g\} \rangle$ according to the moralization criterion for chain graphs.

Definition 12. Supposing G is a CG over N one says that a probability distribution P over N is *Markovian* with respect to G if

$$A \perp\!\!\!\perp B \,|\, C\ [G] \ \text{ implies } \ A \perp\!\!\!\perp B \,|\, C\ [P] \quad \text{for every } \langle A, B|C\rangle \in \mathcal{T}(N)\,.$$

Say that P is *perfectly Markovian* if the converse implication holds as well. △

Justification for use of chain graphs as mathematical tools for description of stochastic CI models follows from the next result proved in [58].

Theorem 13. For every chain graph G over N there exists a strictly positive discrete probability distribution P over N which is perfectly Markovian with respect to G.

Thus, every CG indeed represents a stochastic CI model, not only the collection of CI statements shared by all Markovian distributions (see Remark 10). Note that analogous results were earlier achieved for UGs [16,22] and for DAGs [19]. Analogously to the case of DAGs the substantial step of the proof of Theorem 13 is direct c-separation criterion (c stands for 'chain') for identification of CI statements represented in a CG. This criterion introduced in [6] generalizes d-separation criterion for DAGs [43] and is equivalent to the above mentioned moralization criterion for CGs (see Consequence 4.1 in [58]).

Remark 14. Even more general types of graphs were recently proposed for description of CI structures. For example, general directed graphs which allow multiple edges and directed cycles [47], reciprocal graphs [23], joint-response chain graphs [11], chain graphs with alternative interpretation [2], partial ancestral graphs [3] and annotated graphs [41].

6. Why non-graphical methods?

However, the main problem of graphical approaches is that they can hardly describe all stochastic CI structures (at least in discrete case). Let me illustrate it in case of four factors. In recently finished series of papers [33–35] all CI models induced by discrete distributions over four factors were found (for an overview see [55]). The resulting number makes it possible to make comparison in this case. One has 64 UG models, 185 DAG models, 200 CG-models and 18300 stochastic CI models over four factors (the number of semi-graphoids is 26424). I hope that the numbers are eloquent.

However, they are theoretical reasons why even above mentioned advanced graphical approaches do not overcome this limitation of graphical methods. Namely, the number of graphs is exponential in the number of factors while the number of stochastic CI models is superexponential (see Section 3.6 in [59]).

Note that limitation to a certain 'nice' class of graphical models is possible but it may lead to serious methodological errors in learning CI models from statistical data. More exactly, limited graphical framework may force acceptance of CI statements which are not supported by data on basis of CI statements which are supported by data (see Section 1.1 in [59] or [62]).

This motivated an attempt to develop a non-graphical method of description of CI structures which hopefully keeps some assets of graphical methods. The method proposed below is partially inspired by an information-theoretical approach. It can be used both in case of discrete distributions, in case of non-degenerate Gaussian distributions or even in mixed case of increasingly popular of conditional Gaussian distributions [29].

7. Non-graphical approach

The above mentioned non-graphical method was presented in a series of papers [53,52] but a more up-to-date exposition is to appear in later chapters of [59]. Basic mathematical tools for description of stochastic CI models over N are certain integer-valued functions on the power set of N.

Definition 15. By an *imset* over N is understood an integer-valued function on the class $\{A\,;\, A \subseteq N\}$. Every disjoint triplet $\langle A, B|C\rangle \in \mathcal{T}(N)$ corresponds to a *semi-elementary imset* $u_{\langle A,B|C\rangle}$ defined as follows:

$$u_{\langle A,B|C\rangle}\colon \quad \begin{array}{rcll} A\cup B\cup C,\, C & \longrightarrow & +1, & \\ A\cup C,\, B\cup C & \longrightarrow & -1, & \\ Z & \longrightarrow & 0 & \qquad \text{otherwise.} \end{array}$$

It is called *elementary* if A and B are singletons. An imset u over N is called *structural* if there exists a natural number l and a sequence of (possibly repeated) semi-elementary imsets $u_1, \ldots, u_r$, $r \geq 0$ such that $l \cdot u = \sum_{i=1}^{r} u_i$. △

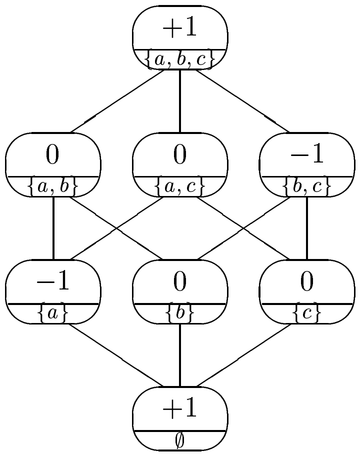

Figure 3. Hasse diagram of a semi-elementary imset.

Note that the word **imset** is an abbreviation for **i**nteger-valued **m**ulti**set**. Multiset is a common concept from combinatorial theory describing systems of sets where some sets can be taken repeatedly. So, multisets generalize hypergraphs in a certain sense.

For a small number of factors (= variables) imsets can be visualized by special pictures. The power set of N is a well-known distributive lattice and can be represented by means of its *Hasse diagram*. Nodes of this diagram correspond to subsets of N and a link between two nodes is made if the symmetric difference of represented sets is a singleton. A function on the power set can be visualized in such a way that one writes the corresponding values into respective nodes. An example of such a diagram is in Figure 3.

However, for description of stochastic CI structures only structural imsets are used. In fact, they were introduced as imsets which can be obtained as a conical combination of semi-elementary imsets (with rational coefficients). However, they can be equivalently introduced as imsets obtained as a conical combination of elementary imsets. There are theoretical reasons for this apparently superfluous terminological distinction: elementary imsets correspond (in sense of the next definition) to atomic (= minimal non-trivial) stochastic CI models.

Definition 16. Let u be a structural imset over N and $\langle A,B|C\rangle \in \mathcal{T}(N)$ a disjoint triplet over N. One writes $A \perp\!\!\!\perp B \,|\, C\; [u]$ and says that $\langle A,B|C\rangle$ is represented in u if there exists a natural number k and a structural imset w such that

$$k \cdot u \;=\; u_{\langle A,B|C\rangle} + w\,.$$

A probability measure P over N is *Markovian* with respect to u if

$$A \perp\!\!\!\perp B \,|\, C \; [u] \quad \text{implies} \quad A \perp\!\!\!\perp B \,|\, C \; [P] \qquad \text{for every } \langle A, B|C \rangle \in \mathcal{T}(N)\,.$$

It is called *perfectly Markovian* if the converse implication is true as well. △

Thus, structural imsets can be used for description of certain formal independence models. Note that they can be utilized for description of all stochastic CI models induced by probability distributions with finite multiinformation. Recall that the multiinformation of a probability distribution P over N [49] is the relative entropy of P with respect to the product of its one-dimensional marginals $\prod_{i \in N} P^{\{i\}}$. The class of distributions with finite multiinformation is a subclass of the class of marginally continuous distributions (see Section 2.3.4 of [59]).

Those readers who are not familiar with basic measure-theoretical concepts should know at least the fact that every discrete distribution has finite multiinformation. The same claim is true for non-degenerate Gaussian distributions. The main result is the following one (see Section 5.3 of [59], discrete case [53]).

Theorem 17. Let P be a probability distribution over N with finite multiinformation. Then there exists a structural imset u over N such that P is perfectly Markovian with respect to u.

Thus, every discrete stochastic CI structure can be described by a structural imset. On the other hand, there exists a structural imset describing a formal independence model which is not a stochastic CI model for any discrete probability measure.

Some readers may object that structural imsets are far from reasonable interpretation. Perhaps they appreciate the following equivalent definition of Markovian distribution in terms of a certain product formula (see Section 4.5 of [59], discrete case [52]).

Theorem 18. Let u be a structural imset over N. Then a probability distribution P over N with finite multiinformation is Markovian with respect to u iff the following formula holds

$$\prod_{S \subseteq N} f_S(x_S)^{u^+(S)} = \prod_{S \subseteq N} f_S(x_S)^{u^-(S)} \quad \text{for } \mu_N\text{-almost every } x \in X_N\,.$$

Here $\mu_N \equiv \prod_{i \in N} \mu_i$ denotes suitable produt of dominating measures and u^+ are u^- are positive and negative parts of u (i.e. $u^+(S) = \max\{u(S), 0\}$ and $u^-(S) = \max\{-u(S), 0\}$).

Note that in discrete case the formula above reduces to the form

$$\prod_{S \subseteq N} p^S(x_S)^{u^+(S)} = \prod_{S \subseteq N} p^S(x_S)^{u^-(S)} \quad \text{for every } x \in X_N\,.$$

Thus, every discrete stochastic CI models has (through the concept of structural imset) interpretation which is analogous to the interpretation of well-known log-linear models from mathematical statistics (see e.g. Chapter 4 in [29]). Further consequence of the formula in Theorem 18 is that the class of marginals

$$\{\, P^T\,;\, T \subseteq S \ \text{ for some } S \text{ with } u(S) < 0 \,\}$$

determines uniquely (always wider) class of marginals

$$\{\, P^T\,;\, T \subseteq S \ \text{ for some } S \text{ with } u(S) > 0 \,\}\,.$$

This often means (for 'nice' structural imsets) that P is determined uniquely by some of its proper marginals.

An important question is the question of theoretical possibility of computer implementation of the method. Structural imsets provide a certain numerical inference mechanism which is based on arithmetic operations with integers. Further results of [59,53] concern this question. One of the results (see Sections 5.2 and 6.2.2 in [59]) says that for every N there exists a finite class of non-negative imsets (the number of these imsets depends on cardinality of N) which allows one to reduce the question of testing inference between structural imsets over N to a standard task of linear programming (namely checking whether scalar product of some integral vectors are non-negative).

Characterization of Markov equivalent structural imsets (that is, imsets describing the same CI structure) is a consequence of this result. Moreover, a direct formula which makes it possible to 'translate' some classic graphs into structural imsets describing the same CI structure is proposed (see Section 7.2 of [59]).

8. Conclusion

The method of structural imsets makes it possible to describe all stochastic CI structures arising in discrete framework and in continuous Gaussian framework. However, it is not a finished theory. There are lots of open problems motivated mainly by practical requirements (see Chapter 8 of [59]).

A substantial step towards wider acceptance of this approach can be suitable *visualization* for humans. Related question is relevant *interpretation* of these models of CI structure (which is however more likely a philosohical question than a mathematical question). Hasse diagrams are good visualization tools for a small number of factors but impractical for more than 5 factors. Perhaps some of 'sparse' structural imsets (that is imsets with a small number of non-zero values) can be visualized in the form of graphs whose nodes are sets (with non-zero imset value). I have in mind an analogue of joint trees [10] or valuation networks [45]. I guess that it is feasible for structural imsets which correspond to some classic graphical models (e.g. decomposable models).

Another bunch of open questions concerns *computer implementation*. The way of implementation is clear from theoretical point of view but related problems with memory demands (when the number of variables increases) appear to be very serious. The task which seems to be the most important one from statistical point of view is the task to develop good methods for *learning* CI structures induced by structural imsets. This remains to be a topic of research New Millenium.

Acknowledgements

I would like to thank to both reviewers for their valuable comments. This is also a suitable place to express my thanks to the organizing committee of the conference *Partial Knowledge and Uncertainty: Independence, Conditioning, Inference*, in particular to Prof. Romano Scozzafava and Dr. Barbara Vantaggi, for organizing such a pleasant event which held in Rome, May 2000.

References

[1] S. A. Andersson, D. Madigan and M. D. Perlman, A characterization of Markov equivalence for acyclic digraphs, Annals of Statistics 25 (1997) 505-541.

[2] S. A. Andersson, D. Madigan and M. D. Perlman, Alternative Markov properties for chain graphs, Scandinavian Journal of Statistics 28 (2001) 33-85.

[3] S. A. Andersson, D. Madigan, M. D. Perlman and T. S. Richardson, Graphical Markov models in multivariate analysis, in: *Multivariate Analysis, Design of Experiments, and Survey Sampling, Statistical Textbooks Monographs 159*, Dekker, 1999, pp. 187-229.

[4] S. Benferhat, D. Dubois and H. Prade, Possibilistic independence and plausible reasoning, in: *Advances in Fuzzy Systems - Applications and Theory 8*, World Scientific, 1995, pp. 47-63.

[5] B. Ben Yaghlane, P. Smets and K. Mellouli, Independence concepts for belief functions, in Proceedings of 8th International Conference on Information Processing and Management of Uncertainty in Knowledge Based Systems IPMU 2000 vol. I., pp. 357-364.

[6] R. R. Bouckaert and M. Studený, Chain graphs: semantics and expressiveness, in: *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, eds. Ch. Froidevaux and J. Kohlas, Lecture Notes in Artificial Intelligence 946, Springer, 1995, pp. 67-76.

[7] G. Coletti and R. Scozzafava, Null events and stochastic independence, Kybernetika 34 (1998) 69-78.

[8] G. Coletti and R. Scozzafava, Zero probabilities in stochastic independence, in: *Information, Uncertainty, Fusion*, eds. B. Bouchon-Meunier, R. R. Yager and L. A. Zadeh, Kluwer, 2000, pp. 185-196.

[9] I. Couso, S. Moral and P. Walley, A survey of concepts of independence for imprecise probabilities, Risk, Decision and Policy 5 (2000) 1-17.

[10] R. G. Cowell, A. P. Dawid, S. L. Lauritzen and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems*, Springer-Verlag, 1999.

[11] D. R. Cox and N. Wermuth, *Multivariate Dependencies - Models, Analysis, and Interpretation*, Chapman and Hall, 1996.

[12] A. P. Dawid, Conditional independence in statistical theory, Journal of Royal Statistical Society series B 41 (1979) 1-31.

[13] A. P. Dawid, Separoids: a general framework for conditional independence and irrelevance, to appear in Annals of Mathematics and Artificial Intelligence.
[14] L. M. de Campos and J. F. Huete, Independence concepts in possibility theory, parts I. and II., Fuzzy Sets and Systems 103 (1999) 127-152 487-505.
[15] J.-P. Florens, M. Mouchart and J.-M. Rolin, *Elements of Bayesian Statistics*, Marcel Dekker, 1990.
[16] M. Frydenberg, Marginalization and collapsability in graphical interaction models, Annals of Statistics 18 (1990) 790-805.
[17] M. Frydenberg, The chain graph Markov property, Scandinavian Journal of Statistics 17 (1990) 333-353.
[18] D. Geiger, T. Verma and J. Pearl, Identifying independence in Bayesian networks, Networks 20 (1990) 507-534.
[19] D. Geiger and J. Pearl, On the logic of causal models, in: *Uncertainty in Artificial Intelligence 4*, eds. R. D. Shachter, T. S. Lewitt, L. N. Kanal and J. F. Lemmer, North-Holland, 1990, pp. 3-14.
[20] D. Geiger and J. Pearl, Logical and algorithmic properties of conditional independence and their application to Bayesian networks, Annals of Mathematics and Artificial Intelligence 2 (1990) 165-178.
[21] D. Geiger, A. Paz and J. Pearl, Axioms and algorithms for inferences involving probabilistic independence, Information and Computation 91 (1991) 128-141.
[22] D. Geiger and J. Pearl, Logical and algorithmic properties of conditional independence and graphical models, Annals of Statistics 21 (1993) 2001-2021.
[23] J. T. A. Koster, Markov properties of nonrecursive causal models, Annals of Statistics 24 (1996) 2148-2177.
[24] H. Kiiveri, T. P. Speed and J. B. Carlin, Recursive causal models, Journal of Australian Mathematical Society series A 36 (1984) 30-52.
[25] S. L. Lauritzen and N. Wermuth, Mixed interaction models, Research report R-84-8, Institute of Electrical Systems, University of Aalborg 1984 (the research report was later modified and became a basis of the journal paper [26]).
[26] S. L. Lauritzen and N. Wermuth, Graphical models for associations between variables, some of which are qualitative and some quantitative, Annals of Statistics 17 (1989) 31-57.
[27] S. L. Lauritzen, Mixed graphical association models, Scandinavian Journal of Statistics 16 (1989) 273-306.
[28] S. L. Lauritzen, A. P. Dawid, B. N. Larsen and H.-G. Leimer, Independence properties of directed Markov fields, Networks 20 (1990) 491-505.
[29] S. L. Lauritzen, *Graphical Models*, Oxford Statistical Science Series 17, Clarendon Press, 1996.
[30] F. M. Malvestuto, A unique formal system for binary decomposition of database relations, probability distributions and graphs, Information Sciences 59 (1992) 21-52.
+ F. M. Malvestuto, M. Studený, Comment on "A unique formal ... graphs", Information Sciences 63 (1992) 1-2.
[31] F. M. Malvestuto, A hypergraph-theoretic analysis of collapsibility and decomposability for extended log-linear models, Statistics and Computing 11 (2001) 155-169.
[32] F. Matúš, Stochastic independence, algebraic independence and abstract connectedness, Theoretical Computer Science 134 (1994) 455-471.
[33] F. Matúš and M. Studený, Conditional independencies among four random variables I., Combinatorics, Probability and Computing 4 (1995) 269-278.
[34] F. Matúš, Conditional independencies among four random variables II., Combinatorics, Probability and Computing 4 (1995) 407-417.

[35] F. Matúš, Conditional independencies among four random variables III. final conclusion, Combinatorics, Probability and Computing 8 (1999) 269-276.

[36] F. Matúš, On the length of the semigraphoid inference, in: *Proceedings of WUPES'2000, 5th Workshop on Uncertainty Processing, June 21-24, 2000, Jindřichův Hradec, Czech Republic*, pp. 176-180.

[37] E. Mendelson, *Introduction to Mathematical Logic*, D. van Nostrand, 1979.

[38] M. Mouchart and J.-M. Rolin, A note on conditional independence with statistical application, Statistica 44 (1984) 557-584.

[39] J. Moussouris, Gibbs and Markov properties over undirected graphs, Journal of Statistical Physics 10 (1974) 11-31.

[40] K. K. Nambiar, Some analytic tools for the design of relational database systems, in Proceedings of the 6th Conference on Very Large Data Bases 1980, pp. 417-428.

[41] A. Paz, R. Y. Geva and M. Studený, Representation of irrelevance relations by annotated graphs, Fundamenta Informaticae 42 (2000) 149-199.

[42] J. Pearl and A. Paz, Graphoids: a graph-based logic for reasoning about relevance relations, in: *Advances in Artificial Intelligence II.*, eds. B. Du Boulay, D. Hogg and L. Steels, North-Holland, 1987, pp. 357-363.

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

[44] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, 1974.

[45] P. P. Shenoy, Conditional independence in valuation-based systems, International Journal of Approximate Reasoning 10 (1994) 203-234.

[46] J. Q. Smith, Influence diagrams for statistical modelling, Annals of Statistics 17 (1989) 654-672.

[47] P. Spirtes, Directed cyclic graphical representations of feedback models, in: *Uncertainty in Artificial Intelligence 11*, eds. P. Besnard and S. Hanks, Morgan Kaufmann, 1995, pp. 491-498.

[48] W. Spohn, Stochastic independence, causal independence and shieldability, Journal of Philosophical Logic 9 (1980) 73-99.

[49] M. Studený, Multiinformation and the problem of characterization of conditional independence relations, Problems of Control and Information Theory 18 (1989) 3-16.

[50] M. Studený, Conditional independence relations have no finite complete characterization, in: *Information Theory, Statistical Decision Functions and Random Processes volume B*, eds. S. Kubík and J.Á. Víšek, Kluwer (also Academia), 1992, pp. 377-396.

[51] M. Studený, Formal properties of conditional independence in different calculi of AI, in: *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, eds. M. Clarke, R. Kruse and S. Moral, Lecture Notes in Computer Science 747, Springer, 1993, pp. 341-348.

[52] M. Studený, Description of conditional independence structures by means of imsets: a connection with product formula validity, in: *Uncertainty in Intelligent Systems*, eds. B. Bouchon-Meunier, L. L. Valverde and R. R. Yager, Elsevier, 1993, pp. 179-194.

[53] M. Studený, Description of structures of stochastic conditional independence by means of faces and imsets (a series of 3 papers), 1st part: introduction and basic concepts, 2nd part: basic theory, 3rd part: examples of use and appendices, International Journal of General Systems 23 (1994/1995) 123-137 201-219 323-341.

[54] M. Studený, Semigraphoids are two-antecedental approximations of stochastic conditional independence models, in: *Uncertainty in Artificial Intelligence 10*, eds. R. L. de Mantaras and D. Poole, Morgan Kaufmann, 1994, pp. 546-552.

[55] M. Studený and P. Boček, CI-models arising among 4 random variables, in: *Proceedings of WUPES'94 (3rd workshop on Uncertainty Processing in Expert Systems), September 11-15, 1994, Třešť, Czech Republic*, pp. 268-282.

[56] M. Studený, Semigraphoids and structures of probabilistic conditional independence, Annals of Mathematics and Artificial Intelligence 21 (1997) 71-98.

[57] M. Studený, A recovery algorithm for chain graphs, International Journal of Approximate Reasoning 17 (1997) 265-293.

[58] M. Studený and R. R. Bouckaert, On chain graph models for description of conditional independence structures, Annals of Statistics 26 (1998) 1434-1495.

[59] M. Studený, On mathematical description of probabilistic conditional independence structures, a survey monograph (thesis for DrSc degree), Institute of Information Theory and Automation, Prague (May 2001).

[60] J. Vejnarová, Conditional independence relations in possibility theory, International Journal of Uncertainty Fuzziness and Knowledge-Based Systems 8 (2000) 253-269.

[61] T. Verma and J. Pearl, Causal netwoks: semantics and expressiveness, in: *Uncertainty in Artificial Intelligence 4*, eds. R. D. Shachter, T. S. Lewitt, L. N. Kanal and J. F. Lemmer, North-Holland, 1990, pp. 69-76.

[62] Y. Xiang, S. K. M. Wong, N. Cercone, Critical remarks on single link search in learning belief networks, in: *Uncertainty in Artificial Intelligence 12*, eds. E. Horvitz and F. Jensen, Morgan Kaufmann, 1996, pp. 564-571.

[63] Book of extended abstracts of the conference "Partial Knowledge and Uncertainty: Independence, Conditioning, Inference", May 4-6, Rome, Italy, eds. R. Scozzafava and B. Vantaggi.