*Int. J. General Systems*, Vol. 23, pp. 323–341
Reprints available directly from the publisher
Photocopying permitted by license only



# DESCRIPTION OF STRUCTURES OF STOCHASTIC CONDITIONAL INDEPENDENCE BY MEANS OF FACES AND IMSETS
## 3rd part: examples of use and appendices[1]

MILAN STUDENÝ

*Institute of Information Theory and Automation, Czech Academy of Sciences, Pod vodárenskou věží 4, 182 08 Prague 8, Czech Republic*

*Local Abstract (3rd part)* This part contains several examples that illustrate the implementation of the facial deductive mechanism and show how to transform information about conditional independence structure given in the form of dependency models, Bayesian or Markov networks into imsets. Another example indicates that the facial deductive mechanism is indeed more powerful than the semigraphoid one. A simple method of proving the probabilistic soundness of an inference rule is presented. Two appendices contain some supplementary results about the structural ordering of imsets. The advantages and disadvantages of and prospects for the presented approach are discussed in the Conclusions.

INDEX TERMS:[2] Conditional independence, imset, Bayesian network, Markov network, facial implication, portrait.

## PREFACE

This paper is the last installment of the three-part work *Description of structures of stochastic conditional independence by means of faces and imsets*. Recall that the first part of the work describes the motivation for the presented theory and defines basic concepts while the second part contains its mathematical fundamentals. This part gives several examples illustrating the implementation of the corresponding deductive mechanism for conditional independence (CI) and the relation to the other existent methods of description of CI-structures. All these examples are gathered in one section. Moreover, this part is supplemented with two appendices. Appendix A contains the proof of Assertion 2.2 from the second installment of the series and Appendix B studies the operation of *contraction* of imsets which may be used in a probabilistic expert system as "restriction of the set of symptoms". The paper also contains a Conclusions section (for the entire series) which treats the prospects of the presented theory.

Every definition, result, example or section throughout all series is denominated by two numbers: the first number indicates the paper where it can be found and the second number is its location within that paper.

---

[1]This research was supported by the internal grants of Czech Academy of Sciences n. 27510 "Explanatory power of probabilistic expert systems: theoretical background" and n. 27564 "Knowledge derivation for probabilistic expert systems".

[2]AMSclassification: 68T 30, 62B 10.

Of course, the reader should be familiar with preceding two parts of the work in order to understand adequately all the examples. However, some examples are easily understandable without thorough study of those papers. Namely, the last two examples illustrating the combining of Bayesian networks and the deriving of inference rules can be read almost immediately (only a few definitions from sections 1.1, 1.2 and Def 2.10 from §2.3.1 are needed) and considered as simple instructions.

## NOTATION

Throughout the paper we will deal with the following situation: A *finite* set $N$ having at least *two elements* called the *basic set* is given, i.e. $2 \leq \text{card}\, N < \infty$. The class of all its subsets will be denoted by $\exp N$. The class of *nontrivial* subsets of $N$, i.e. subsets having at least two elements will be denoted by $\mathcal{U}$:

$$\mathcal{U} = \{S \subset N;\ \text{card}\, S \geq 2\}.$$

For a set $T \subset N$, its *indicator* i.e. the zero-one function on $\exp N$ (possibly restricted to $\mathcal{U}$), is defined as follows:

$$\delta_T(S) = \begin{cases} 1 & \text{in case } S = T \\ 0 & \text{in case } S \neq T. \end{cases}$$

For disjoint sets $A, B \subset N$ the juxtaposition $AB$ will stand for their union $A \cup B$. The symbol $A \perp B \mid C$ is used for the CI-statement saying that $A$ is conditionally independent of $B$ given $C$ (see Def 1.2, §1.1.1), $\langle \cdot, \cdot \rangle$ denotes the scalar product (see Def 1.13b, §1.3.1).

The structural ordering (see Def 2.8, §2.2.1) will be denoted by $\prec$. The letter $E$ usually denotes its base (i.e. the set of *elementary imsets*, see Def 2.7, §2.2.1), while the letter $A$ its skeleton (Def 1.17, §1.3.4).

The symbol $\vdash_{sem}$ stands for semigraphoid derivability (Def 1.7, §1.1.3), $\models$ is used for probabilistic implication (Def 1.8, §1.1.3) and $\mapsto$ for facial implication (Def 2.3, §2.1.2). Facial equivalence (see Def 2.4, §2.1.3) will be denoted by $\approx$, the set of nonnegative integers (including zero) by $\mathbb{Z}^+$ and the set of positive integers (natural numbers) by $\mathbb{N}$. Finally, the following symbols for classes of functions on $\mathcal{U}$ will be used:

| | |
|---|---|
| $R(\mathcal{U})$ | the class of real functions on $\mathcal{U}$ |
| $Z(\mathcal{U})$ | the class of integer-valued functions on $\mathcal{U}$ |
| $Z_{norm}(\mathcal{U})$ | the class of normalized integer-valued functions on $\mathcal{U}$, see Def 1.10, sect. 1.2 |
| $Z(\prec)$ | the class of structural imsets, see Def 2.8, §2.2.1 |

## 3.1 EXAMPLES OF USE

This section contains several examples showing how to make facial implication implementable on computers. First, all skeletal imsets (for the structural ordering) in

**Table 3.1** Values $\langle a, u\rangle$ for $a \in A$ and $u \in E$

| | $\delta_N - \delta_S - \delta_T$ | $\delta_N - \delta_S - \delta_R$ | $\delta_N - \delta_T - \delta_R$ | $\sigma_S$ | $\delta_T$ | $\delta_R$ |
|---|---|---|---|---|---|---|
| $a_1 = \delta_N$ | 1 | 1 | 1 | 0 | 0 | 0 |
| $a_2 = \delta_N + \delta_S$ | 0 | 0 | 1 | 1 | 0 | 0 |
| $a_3 = \delta_N + \delta_T$ | 0 | 1 | 0 | 0 | 1 | 0 |
| $a_4 = \delta_N + \delta_R$ | 1 | 0 | 0 | 0 | 0 | 1 |
| $a_5 = 2\delta_N + \delta_S + \delta_T + \delta_R$ | 0 | 0 | 0 | 1 | 1 | 1 |

the case of 3 symptoms are found and the testing of facial implication by means of the skeleton is illustrated. Then two types of internal computer representation of faces are discussed (namely by means of *generating imsets and portraits*). Further examples illustrate three ways of inputing information about CI-structure (individual CI-statements, Markov and Bayesian networks). The section is concluded by two examples of use illustrating the advantages of the proposed method over previous methods. The first one shows how a whole CI-structure can be inferred using the new approach. The second one says how "to prove easily" probabilistic soundness of "prospective" inference rules.

### *3.1.1 Testing Facial Implication*

Before an example illustrating the testing of facial implication we find the skeleton in case of 3 variables.

EXAMPLE 3.1 (skeleton in case of 3 variables)
Suppose that card $N = 3$; let $\{S, T, R\}$ be the list of its subsets of cardinality 2. In this case there exist 6 elementary imsets (see Def 2.7, §2.2.1), namely:

$$E = \{\delta_N - \delta_S - \delta_T, \delta_N - \delta_S - \delta_R, \delta_N - \delta_T - \delta_R, \delta_S, \delta_T, \delta_R\}.$$

By Consequence 2.5, §2.2.4 the structural ordering has the skeleton i.e. the least finite subset $A \subset Z_{norm}(\mathcal{U})$ satisfying:

$$[0 \prec u] \Leftrightarrow [\forall a \in A \quad \langle a, u\rangle \geq 0] \quad \text{whenever } u \in Z(\mathcal{U}). \tag{3.1}$$

Thus, to show that the set

$$A = \{\delta_N, \delta_N + \delta_S, \delta_N + \delta_T, \delta_N + \delta_R, 2\delta_N + \delta_S + \delta_T + \delta_R\}$$

is indeed the structural skeleton it suffices to verify that it is a minimal subset of $Z_{norm}(\mathcal{U})$ satisfying (3.1). It is done in three steps below.

I. $[0 \prec u] \Rightarrow [\forall a \in A\ \langle a, u\rangle \geq 0]$ whenever $u \in Z(\mathcal{U})$.
Table 3.1 gives the values of $\langle a, u\rangle$ for $a \in A$ and $u \in E$.
Whenever $n \cdot u = \Sigma_{v\in E} k_v \cdot v$ for $n \in \mathbb{N}$, $k_v \in \mathbb{Z}^+$ (see Def 1.15, §1.3.3) then for all $a \in A \quad n \cdot \langle a, u\rangle = \Sigma_{v\in E} k_v \cdot \langle a, v\rangle \geq 0$.

II. $[\forall a \in A\ \langle a, u\rangle \geq 0] \Rightarrow [u = \Sigma_{v\in E} k_v \cdot v$ for some $k_v \in \mathbb{Z}^+]$ whenever $u \in Z(\mathcal{U})$.
Notice that the equality $a_1 + a_5 = a_2 + a_3 + a_4$ (see the table) has an important role in the sequel. The required implication can be proved by induction on $\langle a_1 + a_5, u\rangle$

$= \langle a_2 + a_3 + a_4, u\rangle \geq 0$. In case that $\langle a_1 + a_5, u\rangle = 0$ we easily derive $[\langle a_i, u\rangle = 0$ for all $i]$, hence $u \equiv 0$ and simply put $k_v = 0$ for every $v \in E$.

In case $\langle a_1 + a_5, u\rangle > 0$ at least one of 6 cases listed below occurs and one can find $w \in E$ having $\langle a_i, w\rangle = 1$ just for the selected $a_i \in A$:

$$[\langle a_1, u\rangle, \langle a_2, u\rangle > 0] \rightarrow w = \delta_N - \delta_T - \delta_R$$

$$[\langle a_1, u\rangle, \langle a_3, u\rangle > 0] \rightarrow w = \delta_N - \delta_S - \delta_R$$

$$[\langle a_1, u\rangle, \langle a_4, u\rangle > 0] \rightarrow w = \delta_N - \delta_S - \delta_T$$

$$[\langle a_2, u\rangle, \langle a_5, u\rangle > 0] \rightarrow w = \delta_S$$

$$[\langle a_3, u\rangle, \langle a_5, u\rangle > 0] \rightarrow w = \delta_T$$

$$[\langle a_4, u\rangle, \langle a_5, u\rangle > 0] \rightarrow w = \delta_R.$$

Clearly, $u - w$ satisfies the induction assumption.

III. $\forall \tilde{a} \in A\ \exists u \in Z(\mathcal{U})\quad [\langle \tilde{a}, u\rangle < 0\ \&\ \forall a \in A \setminus \{\tilde{a}\}\ \langle a, u\rangle \geq 0]$.

This can be shown as follows:

$$\tilde{a} = a_1 = \delta_N \rightarrow u = -\delta_N + \delta_S + \delta_T + \delta_R$$

$$\tilde{a} = a_2 = \delta_N + \delta_S \rightarrow u = -\delta_S + \delta_T + \delta_R$$

$$\tilde{a} = a_3 = \delta_N + \delta_T \rightarrow u = \delta_S - \delta_T + \delta_R$$

$$\tilde{a} = a_4 = \delta_N + \delta_R \rightarrow u = \delta_S + \delta_T - \delta_R$$

$$\tilde{a} = a_5 = 2\delta_N + \delta_S + \delta_T + \delta_R \rightarrow u = \delta_N - \delta_S - \delta_T - \delta_R.$$

EXAMPLE 3.2 (testing of facial implication)
Consider the situation from the preceding example and suppose that the following structural imsets $u$, $v$ are given (the "decomposition" is in braces):

$$u = \delta_N - \delta_R \ (= \{\delta_N - \delta_S - \delta_R\} + \delta_S)$$
$$v = 2\delta_N + 3\delta_S - 2\delta_T - 2\delta_R \ (= 2\,\{\delta_N - \delta_T - \delta_R\} + 3\delta_S).$$

The question is whether $u \mapsto v$. It is possible to recognize it without the skeleton, as it follows from Lemma 2.2(b), §2.1.2:

$$[u \mapsto v] \Leftrightarrow [\exists k \in \mathbb{Z}^+\ k \cdot u - v = \sum_{w \in E} k_w \cdot w \text{ for some } k_w \in \mathbb{Z}^+].$$

Indeed, one can write: $5u - v = 3\{\delta_N - \delta_S - \delta_R\} + 2\delta_T$ and therefore $u \mapsto v$. In this case $u$ has to be multiplied 5 times to find a concrete "decomposition". But how to estimate the needed "factor" in the general case? One can avoid this problem in

**Table 3.2** Scalar products with skeletal imsets

| | $\delta_N - \delta_R$ | $2\delta_N + 3\delta_S - 2\delta_T - 2\delta_R$ | $\delta_N$ | $2\delta_N - \delta_S - \delta_T - \delta_R$ |
|---|---|---|---|---|
| $a_1 = \delta_N$ | 1 | 2 | 1 | 2 |
| $a_2 = \delta_N + \delta_S$ | 1 | 5 | 1 | 1 |
| $a_3 = \delta_N + \delta_T$ | 1 | 0 | 1 | 1 |
| $a_4 = \delta_N + \delta_R$ | 0 | 0 | 1 | 1 |
| $a_5 = 2\delta_N + \delta_S + \delta_T + \delta_R$ | 1 | 3 | 2 | 1 |

case when the skeleton is known (or at least a finite class inducing the structural ordering is known) by means of Assertion 2.1(§2.1.7). Then

$$[u \mapsto v] \Leftrightarrow [\forall a \in A \ \langle a, v\rangle > 0 \Rightarrow \langle a, u\rangle > 0].$$

Thus, Table 3.2 of scalar products implies the conclusion $u \mapsto v$ immediately. The converse question i.e. whether $v \mapsto u$ clearly has a negative answer. You can see that this conclusion can be hardly achieved using the "decomposition" approach.

### *3.1.2 Representation of Faces in Computer*

Now, we are going to deal with the question of how to represent faces in a computer. In fact, there are two possibilities discussed below.

The first way uses the above mentioned identification of faces with classes of $\approx$-equivalence, i.e. it describes a face by its generating imset (§2.1.6). Nevertheless, it has a minor disadvantage: the representation of CI-structure is ambiguous in the sense that the same CI-structures can be represented by different imsets. In fact, two imsets represent the same CI-structure if they are facially equivalent (Summary 2.3, §2.3.2). Certainly, some of the imsets within a class of facial equivalence are exceptional. For example, the structural imsets $u_1 = \delta_N - \delta_S$ and $u_2 = 2\delta_N - 2\delta_S - \delta_T$ are $\approx$-equivalent and $0 \lessdot u_2 - u_1$ but not conversely (i.e. not $0 \lessdot u_1 - u_2$). One can suggest representing $\approx$-classes by imsets "minimal" in this sense. Although these imsets are usually unique (in case card $N = 3$) two such imsets may exist; for example $\delta_N$ and $2\delta_N - \delta_S - \delta_T - \delta_R$ form such a pair of "minimal" $\approx$-equivalent normalized imsets (see Table 3.2 in §3.1.1).

The second way, which avoids the mentioned ambiguity, is to represent faces by means of portraits (§2.1.8). This approach removes the above mentioned problem (see Summary 2.2b). The reader can object: *Portraits are evidently better for representation of faces in a computer. Why does the author deal with generating imsets at all?* I have two reasons for keeping in mind both types of representation.

- The first one is that structural imsets lead to further possible description of CI-structures, namely by means of validity of so-called product formulas. This approach, explained in [Studený, 1993], can make interpretation of facial CI-structures more natural.
- The second reason is connected with the intended input of information about CI-structure. In probabilistic expert systems, one can expect that incomplete information about CI-structure will come from various sources: as a result of a statistical test or as an expert's estimate. The task is usually to gather this information and then either to deduce some "new" information or to "build up" the model. When entering individual CI-statements, imsets are a more natural (di-

rect) "translation" than portraits. To form the corresponding portrait one has to compute a lot of scalar products (specifically, one must be computed for each skeletal imset; note that the cardinality of the skeleton rapidly increases with card $N$). Moreover, the representation of portraits seems to require more memory in a computer. It is not reasonable to make this "translation procedure" after each input. A better way is to gather the information in the form of imsets (simple summing of imsets!) and make the "translation" to portraits only before deducing some "required" information about CI-structure.

### *3.1.3 CI-statements as Input*

The next topic of our discussion is the input of information about CI-structure. The theory was constructed with the intention of making possible various types of input and their combination. Therefore dependency models are taken to be the basic input. They enable experts to enter undirected graphs (Markov networks) or directed acyclic graphs (Bayesian networks) but also of individual CI-statements as a result of an affirmative answer of a statistical test.

Note that the so-called *"principle of positive information about CI-structure"* is accepted here, i.e. we enter only sure positive information about conditional independence. This should be emphasized especially when the source of information is a graph (where just a missing link or arrow means positive information about independence) in order to avoid the mistake "ignorance means independence".

All three above mentioned possibilities for original sources of information will be illustrated by examples. They will show that sometimes one can find a short cut.

EXAMPLE 3.3 (individual CI-statements as input)
Let $N = \{1, 2, 3, 4\}$. Suppose that we have obtained positive information about CI-structure in the form of following four CI-statements:

$$I: \quad 1 \perp 4 \mid 3 \quad 1 \perp 3 \mid 2 \quad 2 \perp 3 \mid 4 \quad 2 \perp 4 \mid 13.$$

The task is to recognize whether CI-statements $t_1 : 1 \perp 3 \mid 4$ resp. $t_2 : 3 \perp 4 \mid 1$ are implied by the given list. The input of information in the form of individual CI-statements is very simple: one can translate each CI-statement by the primary mapping $i$ (Def 2.10, §2.3.1) into a structural imset. In this case we obtain the following set of structural imsets $L = \{u_1, \ldots, u_4\}$:

$$u_1 = \delta_{\{1,3,4\}} - \delta_{\{1,3\}} - \delta_{\{3,4\}} \quad u_2 = \delta_{\{1,2,3\}} - \delta_{\{1,2\}} - \delta_{\{2,3\}}$$

$$u_3 = \delta_{\{2,3,4\}} - \delta_{\{2,4\}} - \delta_{\{3,4\}} \quad u_4 = \delta_N - \delta_{\{1,2,3\}} - \delta_{\{1,3,4\}} + \delta_{\{1,3\}}.$$

The information can be then gathered by summing these imsets: indeed, it follows from Lemma 2.1, §2.1.1 and Lemma 2.2 (a), §2.1.2 that for each $v \in Z(<)$ $[L \mapsto v]$ iff $(\Sigma_{u \in L}\, u) \mapsto v$. Thus our task is to decide whether $u = \Sigma_{i=1}^{4} u_i = \delta_N + \delta_{\{2,3,4\}} - \delta_{\{1,2\}} - \delta_{\{2,3\}} - \delta_{\{2,4\}} - 2\delta_{\{3,4\}}$ facially implies either $v_1 = \delta_{\{1,3,4\}} - \delta_{\{1,4\}} - \delta_{\{3,4\}}$ or $v_2 = \delta_{\{1,3,4\}} - \delta_{\{1,3\}} - \delta_{\{1,4\}}$. The equality

$$u - v_1 + \{\delta_N - \delta_{\{1,2,4\}} - \delta_{\{1,3,4\}} + \delta_{\{1,4\}}\} + \{\delta_{\{1,2,4\}} - \delta_{\{1,2\}} - \delta_{\{2,4\}}\}$$
$$+ \{\delta_{\{2,3,4\}} - \delta_{\{2,3\}} - \delta_{\{3,4\}}\}$$

says that $u \mapsto v_1$. Note that $I \vdash_{sem} t_1$. To show that $u \not\mapsto v_2$ it is necessary to find a

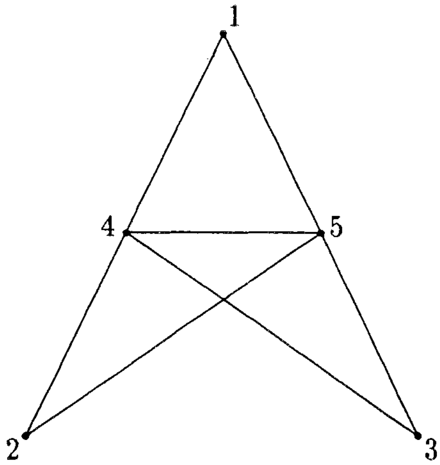


**Figure 3.1** Markov network.

skeletal imset $a$ such that $\langle a, v_2\rangle > 0$ and $\langle a, u\rangle = 0$ (proceed as in Example 3.2). The list of skeletal imsets is in 'code' form in Table 2.1, §2.2.4. As a result of the check procedure one should find two such skeletal imsets, one of which is $a = \delta_N + \delta_{\{1,3,4\}} + \delta_{\{2,3,4\}} + \delta_{\{3,4\}}$ (namely $m_{\mathcal{T}}$ for $\mathcal{T}_{min} = \{\{3, 4\}\}$). Note that the co-atomic face corresponding to $a$ is probabilistically representable; it also implies $I \not\Vdash t_2$.

### *3.1.4 Undirected Graphs as Input*

As mentioned in the Introduction, undirected graphs (UGs) are widely used in the literature to represent CI-structures (the concept on *Markov network* from [Pearl, 1988]). A triplet $\langle A, B, C\rangle$ is represented in a UG iff every path from $A$ to $B$ meets $C$. Every dependency model defined by this condition is probabilistically representable (see [Geiger and Pearl, 1990b]). The elementary method of input for a UG is to form the corresponding dependency model and enter it as described in the preceding example.[3]

Nevertheless, for special graphs, called *triangulated* or *chordal* (specified by the condition that every cycle of the length 4 or more possesses a chord), which correspond to well-known *decomposable models* [Lauritzen, *et al.*, 1984], a short cut is possible. The method giving directly a pertinent structural imset will be described in the following example.

EXAMPLE 3.4 (UG input)
Put $N = \{1, 2, 3, 4, 5\}$ and consider the graph in Figure 3.1. To show the elementary method first find the corresponding dependency model. One can easily verify that the following 6 triplets are the only 'pairwise' triplets represented by the given UG (symmetric triplets are omitted):

$$1 \perp 2 \mid 45 \quad 1 \perp 2 \mid 345 \quad 1 \perp 3 \mid 45 \quad 1 \perp 3 \mid 245 \quad 2 \perp 3 \mid 45 \quad 2 \perp 3 \mid 145.$$

By translation using the primary mapping and summing one can get the imset:

$$u_1 = 3\delta_N - \delta_{\{1,2,4,5\}} - \delta_{\{1,3,4,5\}} - \delta_{\{2,3,4,5\}} - \delta_{\{1,4,5\}} - \delta_{\{2,4,5\}} - \delta_{\{3,4,5\}} + 3\delta_{\{4,5\}}.$$

[3]Note that it suffices to limit input to triplets of the form $\langle\{a\}, \{b\}, C\rangle$ (the so-called pairwise form of dependency model) which correspond to elementary structural imsets (this argumentis based on Proposition I from [Matúš, 1992]; it can be also derived using Theorem 2.1 §2.1.4).

Nevertheless, as the graph is chordal the direct method also can be used. Simply identify the class of maximal cliques $\mathcal{C}$ of the graph and then take the the imset $u$ with the following natural extension:

$$\bar{u} = \delta_{\cup \mathcal{C}} + \sum_{\emptyset \neq \mathcal{B} \subset \mathcal{C}} (-1)^{\operatorname{card} \mathcal{B}} \cdot \delta_{\cap \mathcal{B}}$$

($\cap \mathcal{B}$ is the intersection of sets from $\mathcal{B}$, $\cup \mathcal{C}$ is the union of sets from $\mathcal{C}$).

The reasoning of this formula can be found in [Studený, 1992b]. In our example $\mathcal{C} = \{\{1, 4, 5\}, \{2, 4, 5\}, \{3, 4, 5\}\}$ and the corresponding imset is

$$u_2 = \delta_N - \delta_{\{1,4,5\}} - \delta_{\{2,4,5\}} - \delta_{\{3,4,5\}} + 2\delta_{\{4,5\}}.$$

The reader can easily verify that indeed $u_1 \approx u_2$.

REMARK 3.1 (multiple joint conditional independence)
The preceding example shows how imsets can be used to describe multiple joint conditional independence i.e. conditional independence of more than 2 random variables under condition of another set. Namely, the imset $u_2$ describes independence of 1,2,3 given $\{4, 5\}$. Note that it can be also formulated in terms of dependency models, for example by couple of CI-statements: $1 \perp 2 \mid 45$ & $12 \perp 3 \mid 45$. Nevertheless, there is no simultaneously symmetric and non-redundant expression by means of triplets.

### *3.1.5 Directed Acyclic Graphs as Input*

Directed acyclic graphs (DAGs) represent another classical way to describe CI-structures. The terminology differs: Pearl [1988] calls them *Bayesian networks*, Shachter [1990] *probabilistic influence diagrams*, Smith [1989] simply *influence diagrams*, Lauritzen et al. [1990] *directed Markov fields*. A triplet $\langle A, B, C\rangle$ is represented in a DAG iff $C$ *d*-separates[4] $A$ and $B$ (see [Pearl, *et al.*, 1990]). There are other equivalent criteria: the method of 'moral graphs' proposed by Lauritzen et al. [1990] or the procedure from [Smith, 1989] mentioned below. Certainly, one possible way to translate the CI-structure given by a DAG into an imset is to enter the dependency model made of the triplets represented by the DAG. Nevertheless, an elegant short cut is derived in the following example.

EXAMPLE 3.5 (DAG input)
A simple formula "translating" DAGs into imsets is derived here. Namely, for a DAG the formula:

$$\bar{u} = \delta_N - \delta_\emptyset + \sum_{k \in N} \{\delta_{\pi(k)} - \delta_{(k)\cup\pi(k)}\},$$

where $\pi(k)$ denotes the set of parents of the node $k$ i.e. origins of arcs[5] to $k$, gives

[4]An undirected path from $A$ to $B$ is $d$-separated by $C$ iff either some tail-to-tail or tail-to-head node on it meets $C$ or, on the other hand, the set of descendants of some head-to-head node on it (including that node) has empty intersection with $C$.

[5]Most authors use term 'arc' instead of 'arrow'.

the natural extension of a structural imset describing exactly the CI-structure represented by the given DAG.

Indeed, the originally proposed way to obtain triplets represented by a DAG is to form a sequence of its nodes compatible with the orientation of the arcs, then define the following sequence of triplets (so-called *input list*):

$$I: \quad \langle k, \{1, \ldots, k-1\} \setminus \pi(k), \pi(k) \rangle \quad k = 1, \ldots, n$$

(whenever $\{1, \ldots, k-1\} \setminus \pi(k) = \emptyset$ the corresponding triplet is omitted) and then form their semigraphoid closure (see [Smith, 1989], the equivalence of this criterion to $d$-separation is shown in [Pearl *et al.*, 1990]). As the resulting dependency model is probabilistically representable (see [Geiger and Pearl, 1990a]) it follows that (Consequence 2.6, §2.3.1 and Consequence 2.8, §2.3.4):

$$[I \vdash_{sem} t] \Leftrightarrow [i(I) \mapsto i(t)] \Leftrightarrow [I \models t] \quad \text{whenever } t \in T_*(N).$$

Thus by Lemma 2.1 and Lemma 2.2(a) the imset $u = \Sigma_{t \in I}\, i(t)$ describes the same CI-structure. Nevertheless $\bar{u} = \Sigma_{k=1}^{n} \{\delta_{\{1,\ldots,k\}} - \delta_{\{1,\ldots,k-1\}} - \delta_{\{k\} \cup \pi(k)} + \delta_{\pi(k)}\}$ gives the formula above. Note that the formula does not depend on the choice of a "compatible" sequence!

Remark 3.2 Theoretically, probability measures could also serve as input information about CI-structure. In this case its multiinformation function $M$ can give directly the portrait as follows:

$$\{a \in A; \exists u \in E \quad \langle M, u \rangle = 0 \;\&\; \langle a, u \rangle > 0\}.$$

### *3.1.6 Combining Information Sources*

The following example shows that facial implication is indeed more powerful than the semigraphoid derivability. The example combines two inputs, namely two DAGs. Another important feature is that it illustrates how a whole CI-structure can be derived using the new approach. In fact, another example of this operation can be found in the literature, namely reversing arcs from [Shachter, 1990] can be understood as derivation of a new DAG from an original one. I think that the presented method can be also used for these purposes.

Example 3.6 (combining Bayesian networks)
Let $N = \{1, 2, 3, 4\}$. Suppose that positive information about CI-structure was given by two experts in the form of the DAGs in Figure 3.2. One can find the corresponding imsets $u_1$ and $u_2$ by the formula from Example 3.5:

$$u_1 = \delta_N - \delta_{\{1,2,4\}} - \delta_{\{1,2,3\}} + \delta_{\{1,2\}} + \delta_{\{1,3\}}$$
$$u_2 = \delta_N - \delta_{\{1,3,4\}} - \delta_{\{2,3,4\}} + \delta_{\{2,4\}} + \delta_{\{3,4\}}.$$

To get the imset gathering both pieces of information simply sum these imsets:

$$u = 2\delta_N - \delta_{\{1,2,3\}} - \delta_{\{1,2,4\}} - \delta_{\{1,3,4\}} - \delta_{\{2,3,4\}} + \delta_{\{1,2\}} + \delta_{\{1,3\}} + \delta_{\{2,4\}} + \delta_{\{3,4\}}.$$

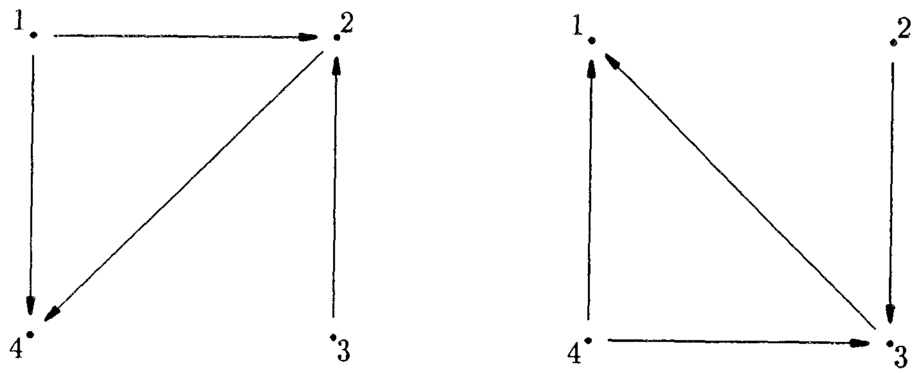


**Figure 3.2** Given DAGs.

The question is whether the influence diagram in Figure 3.3a is implied.

It can be described by the imset $u_3 = \delta_N - \delta_{\{1,2,4\}} - \delta_{\{2,3,4\}} + \delta_{\{2,4\}} + \delta_{\{3,4\}}$. The equality $u - u_3 = \{\delta_N - \delta_{\{1,2,3\}} - \delta_{\{1,3,4\}} + \delta_{\{1,3\}}\} + \delta_{\{1,2\}}$ says $u \mapsto u_3$ and therefore by Consequence 2.8, §2.3.4 the DAG in Figure 3.3a is probabilistically implied by those from Figure 3.2. The Bayesian network in Figure 3.3b can be derived similarly.

Note that the conclusion from Example 3.6 cannot be shown using semigraphoid derivation. For example, the triplet $\langle 3, 4, \emptyset \rangle$ represented by the DAG in Figure 3.3a cannot be derived by semigraphoid axioms from the triplets represented by the DAGs

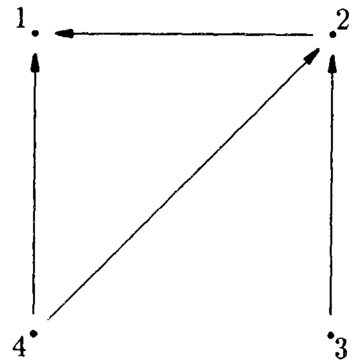


**Figure 3.3a** The first implied DAG.

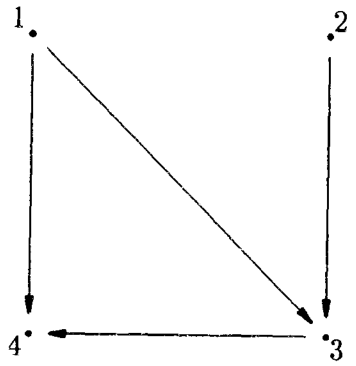


**Figure 3.3b** The second implied DAG.

in Figure 3.2 (as they form a semigraphoid).[6] Consequently, the DAG in Figure 3.3a (and similarly the one in Figure 3.3b) cannot be derived using the method of reversing arcs [Shachter, 1990] which is based on semigraphoid inference rules (see [Smith, 1989]).

### *3.1.7 Deriving Inference Rules*

As already mentioned, the probabilistic implication cannot be characterized by means of a finite number of probabilistically sound inference rules [Studený, 1992a]. Therefore, it would be nice to be able to derive 'automatically' such inference rules. The following example describes a simple method how to verify soundness of inference rules suspected to be sound; note that it can be used to derive all inference rules (A.3)–(A.7) from [Studený, 1994].

EXAMPLE 3.7 (deriving inference rule)
Suppose that $A$, $B$, $C$, $D$, $E$ are nonempty and pairwise disjoint subsets of the basic set $N$. Our task is to verify the probabilistic soundness of the following "inference rule":

$$[A \perp B \mid CD \;\&\; A \perp C \mid DE \;\&\; A \perp D \mid BE \;\&\; A \perp E \mid BC]$$
$$\Leftrightarrow [A \perp E \mid CD \;\&\; A \perp D \mid BC \;\&\; A \perp C \mid BE \;\&\; A \perp B \mid DE]$$

For this purpose "translate" every triplet of the first line to imsets by the primary mapping (see Def 2.10, §2.3.1):

$$u_1 = \delta_{ABCD} - \delta_{ACD} - \delta_{BCD} + \delta_{CD}$$
$$u_2 = \delta_{ACDE} - \delta_{ADE} - \delta_{CDE} + \delta_{DE}$$
$$u_3 = \delta_{ABDE} - \delta_{ABE} - \delta_{BDE} + \delta_{BE}$$
$$u_4 = \delta_{ABCE} - \delta_{ABC} - \delta_{BCE} + \delta_{BC}.$$

Similarly "translate" triplets from the second line:

$$u_5 = \delta_{ACDE} - \delta_{ACD} - \delta_{CDE} + \delta_{CD}$$
$$u_6 = \delta_{ABCD} - \delta_{ABC} - \delta_{BCD} + \delta_{BC}$$
$$u_7 = \delta_{ABCE} - \delta_{ABE} - \delta_{BCE} + \delta_{BE}$$
$$u_8 = \delta_{ABDE} - \delta_{ADE} - \delta_{BDE} + \delta_{DE}.$$

It is easy to see $u_1 + u_2 + u_3 + u_4 = u_5 + u_6 + u_7 + u_8$ and this implies the probabilistic soundness of the "inference rule" above. Indeed, by Lemma 2.2(b),

[6] In fact, it can be done by the inference rule (A.7) from [Studený, 1994] which is independent of the semigraphoid inference rules.

§2.1.2 $\{u_1, u_2, u_3, u_4\} \mapsto u_5, u_6, u_7, u_8$ and conversely $\{u_5, u_6, u_7, u_8\} \mapsto u_1, u_2, u_3, u_4$. Hence Consequence 2.8, §2.3.4 gives the required probabilistic implication.

## 3.2 APPENDIX A (proof of Assertion 2.2)

First, we derive some necessary conditions for structural imsets.

CONSEQUENCE A.1 Introduce a multiset $r^B$ for each $B \in \mathcal{U}$ as follows:

$$r^B(K) = \begin{cases} 1 & \text{in case } B \subset K \\ 0 & \text{otherwise.} \end{cases}$$

Whenever $u \in Z(<)$ then $\forall\, B \in \mathcal{U}\ \langle r^B, u\rangle = \Sigma\{u(K); B \subset K\} \geq 0$.

*Proof* Put $\mathcal{T} = \{K \in \mathcal{U}; B \subset K\}$, clearly $m_{\mathcal{T}} = r^B$, use Lemma 2.8 and Theorem 2.4b from §2.2.3. ■

FACT A Whenever $\mathcal{W} \subset \mathcal{U}$ is ascending (see Lemma 2.8) and $u \in Z(\mathcal{U})$ then the condition $[\forall B \in \mathcal{W}\ \langle r^B, u\rangle = 0]$ implies $[\forall K \in \mathcal{W}\ u(K) = 0]$.

Hint: take maximal $B \in \mathcal{W}$ satisfying $u(B) \neq 0$.

CONSEQUENCE A.2 For any $Z \subset N$, card $Z \geq 3$ and $\mathcal{S} \subset \{K \in \mathcal{U}; K \subsetneq Z\}$ with card $\mathcal{S} \geq 2$,

$$[C, D \in \mathcal{S}\quad C \neq D] \Rightarrow Z = C \cup D. \tag{3.2}$$

Suppose that $u \in Z(<)$ satisfies the condition:

$$u(K) \leq 0 \quad \text{whenever } S \subsetneq K \neq Z \text{ for some } S \in \mathcal{S}. \tag{3.3}$$

Then $0 \leq 2u(Z) + \Sigma_{S\in\mathcal{S}}\, u(S)$.

*Proof* Let $\mathcal{T} = \{K \in \mathcal{U}; \exists S \in \mathcal{S}\ S \subset K\}$, try to compute $m_{\mathcal{T}}$ (see Lemma 2.8, §2.2.3), by (3.2) $\mathcal{T}_{min} = \mathcal{S}$. Hence $m_{\mathcal{T}}(S) = 1$ for $S \in \mathcal{S}$. Now, the aim is to show:

$$\underline{m_{\mathcal{T}}}(A) \leq 1 \quad \text{whenever } A \subsetneq Z. \tag{3.4}$$

*Indeed,* it is evident when card $A \leq 2$; proceed by induction on card $A$. Supposing card $A \geq 3$ the nontrivial case is $A \in \mathcal{T} \setminus \mathcal{S}$. For any relevant couple $S, T \subset N$ with $A = S \cup T$ verify $m_{\mathcal{T}}(\mathcal{S}) + m_{\mathcal{T}}(T) - m_{\mathcal{T}}(S \cap T) \leq 1$. By the induction assumption only the case $m_{\mathcal{T}}(S) = m_{\mathcal{T}}(T) = 1$ is interesting. Thus, $C \subset S$ and $D \subset T$ for some $C, D \in \mathcal{S}$. As $A \subsetneq Z$ by (3.2) $C = D$. Hence $C \subset S \cap T$ gives $m_{\mathcal{T}}(S \cap T) = 1$, i.e. $m_{\mathcal{T}}(S) + m_{\mathcal{T}}(T) - m_{\mathcal{T}}(S \cap T) = 1$.

Clearly (3.4) implies $\overline{m_{\mathcal{T}}}(Z) \leq 2$. On the other hand take $C, D \in \mathcal{S}$, $C \neq D$ and by (3.2) and the convexity inequality get: $m_{\mathcal{T}}(Z) \geq m_{\mathcal{T}}(C) + m_{\mathcal{T}}(D) - \underline{m_{\mathcal{T}}}(C \cap D) \geq 2$. Then using Lemma 2.8 and Theorem 2.4b (§2.2.3) write:

$$0 \leq \langle m_{\mathcal{T}}, u\rangle = \sum_{K\in\mathcal{T}} m_{\mathcal{T}}(K)\cdot u(K) = 2u(Z) + \sum_{S\in\mathcal{S}} u(S) + \sum_{K\in\mathcal{T}\setminus\mathcal{S}\cup\{Z\}} m_{\mathcal{T}}(K)\cdot u(K).$$

By (3.3) $\Sigma_{K\in\mathcal{T}\setminus\mathcal{S}\cup\{Z\}}\, m_{\mathcal{T}}(K)\cdot u(K) \leq 0$, hence the desired inequality. ■

LEMMA A.1 Suppose that $u \in E$ ($E$ denotes the set of elementary imsets, see Def 2.7, §2.2.1) and $n\cdot u = v + w$ where $n \in \mathbb{N}$ $v, w \in Z(<)$. Then $v = k\cdot u$ for some $k \in \mathbb{Z}^+$.

*Proof* Divide $E$ into disjoint subsets $E_i$ $i = 1, \ldots, \text{card } N - 1$:

$$E_i = \{u \in E;\ \bar{u} = \delta_{S\cup T} - \delta_S - \delta_T + \delta_{S\cap T} \quad \text{where card } S = \text{card } T = i\}.$$

Three cases will be distinguished.

I. $u \in E_1$ i.e. $u = \delta_A$ with card $A = 2$.

In this case put $\mathcal{W} = \mathcal{U}\setminus\{A\}$ and for each $B \in \mathcal{W}$ write $0 = \langle r^B, n\cdot u\rangle = \langle r^B, v\rangle + \langle r^B, w\rangle$ (notation $r^B$ is from Consequence A.1). By Consequence A.1 both $\langle r^B, v\rangle$ and $\langle r^B, w\rangle$ are nonnegative and therefore vanish. By Fact $A$, $[\forall K \in \mathcal{W}\ v(K) = 0]$ i.e. $v = k\cdot\delta_A$ where $k = \langle r^A, v\rangle \geq 0$ by Consequence A.1.

II. $u \in E_2$ i.e. $u = \delta_A - \delta_B - \delta_C$ where card $A = 3$ and $B, C, D$ are subsets of $A$ of cardinality 2.

Put $\mathcal{W} = \mathcal{U}\setminus\{A, B, C, D\}$ and as in I. derive $[\forall K \in \mathcal{W}\ v(K) = w(K) = 0]$. Then $0 = \langle r^B, n\cdot u\rangle = \langle r^B, v\rangle + \langle r^B, w\rangle$ gives $\langle r^B, v\rangle = \langle r^B, w\rangle = 0$ by Consequence A.1 and hence $v(A) + v(B) = 0 = w(A) + w(B)$. Similarly $v(A) + v(C) = 0 = w(A) + w(C)$. Finally, let $Z = A$ $\mathcal{S} = \{B, C, D\}$ and by Consequence A.2 derive $0 \leq 2v(A) + v(B) + v(C) + v(D)$, $0 \leq 2w(A) + w(B) + w(C) + w(D)$ i.e. $0 \leq v(D), w(D)$ and $0 = n\cdot u(D) = v(D) + w(D)$ implies $v(D) = 0$. Hence $v = k\cdot(\delta_A - \delta_B - \delta_C)$ where $k = \langle r^A, v\rangle \geq 0$ by Consequence A.1.

III. $u \in E_i$ where $i \geq 3$ i.e. $u = \delta_{S\cup T} - \delta_S - \delta_T + \delta_{S\cap T}$ where card $S$ = card $T$ $= i \geq 3$ and card $S\setminus T$ = card $T \setminus S = 1$.

First introduce $\mathcal{A} = \{K \in \mathcal{U};\ K \subset S \cup T\}$ and divide it into four disjoint subclasses:

$$\mathcal{A}_{S\cup T} = \{K \in \mathcal{A};\ K\setminus S \neq \emptyset\ \&\ K\setminus T \neq \emptyset\}$$

$$\mathcal{A}_S = \{K \in \mathcal{A};\ K \subset S\ \&\ K\setminus T \neq \emptyset\}$$

$$\mathcal{A}_T = \{K \in \mathcal{A};\ K\setminus S \neq \emptyset\ \&\ K \subset T\}$$

$$\mathcal{A}_{S\cap T} = \{K \in \mathcal{A};\ K \subset S\ \&\ K \subset T\}$$

The conclusion will be derived in 6 steps:

(i) $\forall K \in \mathcal{U}\setminus\mathcal{A}$ $v(K) = 0 = w(K)$.
Take $\mathcal{W} = \mathcal{U}\setminus\mathcal{A}$ and repeat the procedure from I.

(ii) $v(S \cup T) = -v(S) = -v(T) = v(S \cap T)$ & $w(S \cup T) = -w(S) = -w(T) = w(S \cap T)$.

By consequence A.1 and (i) $0 \leq \langle r_s, v \rangle = v(S) + v(S \cup T)$ and $0 \leq w(S) + w(S \cup T)$. Since $0 = n \cdot (u(S) + u(S \cup T)) = \{v(S) + v(S \cup T)\} + \{w(S) + w(S \cup T)\}$ it follows that $v(S) + v(S \cup T) = 0 = w(S) + w(S \cup T)$. Similarly, using $r^T$ derive $v(T) + v(S \cup T) = 0 = w(T) + w(S \cup T)$. An analogous procedure for $r^{S \cap T}$ yields $v(S \cap T) + v(S) + v(T) + v(S \cup T) = 0 = w(S \cap T) + w(S) + w(T) + w(S \cup T)$.

(iii) $\forall K \in \mathcal{A}_{S \cup T} \setminus \{S \cup T\} \quad v(K) = 0 = w(K)$.

To verify this by reverse induction on card $K$ it suffices to show:

$$[\forall L \in \mathcal{A}_{S \cup T} \setminus \{S \cup T\} \quad K \subsetneq L \quad v(L) = 0 = w(L)] \Rightarrow v(K) = 0 = w(K).$$

Let $Z = S \cup T$ and $\mathcal{S} = \{S, T, K\}$. According the induction assumption, (i) and Consequence A.2 get:

$$0 \leq 2v(S \cup T) + v(S) + v(T) + v(K) \quad 0 \leq 2w(S \cup T) + w(S) + w(T) + w(K)$$

i.e. by (ii) $0 \leq v(K)$, $w(K)$ and $0 = n \cdot u(K) = v(K) + w(K)$ implies $v(K) = 0 = w(K)$.

(iv) $\forall K \in \mathcal{A}_S \setminus \{S\} \quad v(K) = 0 = w(K)$.

Also use reverse induction and prove:

$$[\forall L \in \mathcal{A}_S \setminus \{S\} \quad K \subsetneq L \quad v(L) = 0 = w(L)] \Rightarrow v(K) = 0 = w(K).$$

From (i), (iii), the induction assumption and again (ii) it follows that:

$$\langle r^K, v \rangle = v(K) + v(S) + v(S \cup T) = v(K)$$

$$\langle r^K, w \rangle = w(K) + w(S) + w(S \cup T) = w(K).$$

By Consequence A.1, $0 \leq v(K)$, $w(K)$ and $0 = n \cdot u(K) = v(K) + w(K)$.

(v) $\forall K \in \mathcal{A}_T \setminus \{T\} \quad v(K) = 0 = w(K)$.

The method is the same as in (iv).

(vi) $\forall K \in \mathcal{A}_{S \cap T} \setminus \{S \cap T\} \quad v(K) = 0 = w(K)$.

Repeat the procedure from (iv), but to verify $\langle r^k, v \rangle = v(K)$ and $\langle r^K, w \rangle = w(K)$ use (i), (iii), (iv), (v), the induction assumption and then (ii).

(vii) The steps (i)–(vi) together imply $v = k \cdot (\delta_{S \cup T} - \delta_S - \delta_T + \delta_{S \cap T})$ where $k = \langle r^{S \cup T}, v \rangle \geq 0$ by Consequence A.1. ■

ASSERTION 2.2 The set of elementary imsets $E$ is the base of $<$.

*Proof* Owing to Assertion 1.3 (§1.3.3) it suffices to show that $E$ is a minimal set establishing $<$. Thus, we prove by contradiction:

$$\forall u \in E \quad n \cdot u \neq \sum_{n \in E \setminus \{u\}} k_v \cdot v \quad \text{for any } n \in \mathbb{N},\ k_v \in \mathbb{Z}^+.$$

Suppose not. Then choose $v \in E \setminus \{u\}$ with $k_v > 0$ (it is possible owing to $u \neq 0$) and put $w = n \cdot u - v$. As $0 < v, w$ by Lemma A.1, $v = k \cdot u$ for some $k \in \mathbb{Z}^+$. Since $u, v \in Z_{norm}(\mathcal{U})$ by Lemma 1.1, section 1.2 $v = u$ and this contradicts the choice of $v$.

## 3.3. APPENDIX B (contraction of imsets)

An operation of contraction can be defined for structural imsets. It has significance of restriction of the model of CI-structure to a subset of the basic set. Some results concerning this operation are mentioned below.

DEFINITION B.1 (contraction, intersection)
For an imset $u \in Z(\mathcal{U})$ and a subset $V$ of the basic set introduce the *contraction of u to V* as follows:

$$(u \,\Delta\, V)(S) = \Sigma \{u(K);\, K \in \mathcal{U} \quad S = K \cap V\} \quad \text{whenever } S \in \mathcal{U}.$$

Evidently, $u \,\Delta\, V$ is vanishing outside $V$.
Given $r \in R(\mathcal{U})$ and $V \subset N$ define the *intersection of r by V*, denoted $r \wedge V$, as:

$$(r \wedge V)(S) = \underline{r}(S \cap V) \quad \text{whenever } S \in \mathcal{U}. \qquad (\text{for } \underline{r} \text{ see Def 1.12, section 1.2})$$

Further properties of scalar product quasiorderings can be derived under additional assumptions regarding the inducing class.

LEMMA B.1 Suppose that $\preceq$ is the scalar product ordering induced $C \subset R(\mathcal{U})$.
a) If $C$ is closed under intersection $\preceq$ then

$$v \preceq u \Rightarrow (v \,\Delta\, V) \preceq (u \,\Delta\, V) \quad \text{whenever } u, v \in Z(\mathcal{U}),\ V \subset N. \qquad (V.6)$$

b) If every $r \in C$ is nonnegative and nondecreasing (with respect to inclusion), then

$$0 \preceq u \Rightarrow (u \,\Delta\, V) \preceq u \quad \text{whenever } u \in Z(\mathcal{U}),\ V \subset N. \qquad (V.7)$$

In particular, $0 \leq u$ implies $0 \preceq u$ (let $V = \emptyset$).

*Proof* It is easy to see that:

$$\langle r, u \,\Delta\, V\rangle = \langle r \wedge V, u\rangle \quad \text{whenever } r \in R(\mathcal{U}),\ u \in Z(\mathcal{U}),\ V \subset N.$$

a) Therefore $[(v \,\Delta\, V) \preceq (u \,\Delta\, V)] \Leftrightarrow [\forall r \in C \quad \langle r \wedge V, v\rangle \leq \langle r \wedge V, u\rangle]$. Nevertheless, $r \in C$ implies $(r \wedge V) \in C$ according to the assumption.
b) Clearly $r \wedge V \leq r$ for any $r \in C$ and $V \subset N$. Write for arbitrary $u \in Z^+(\mathcal{U})$:

$$0 \leq \langle r - (r \wedge V), u\rangle = \langle r, u\rangle - \langle r \wedge V, u\rangle = \langle r, u\rangle - \langle r, u \,\Delta\, V\rangle. \quad \blacksquare$$

LEMMA B.2 The class of completely convex set functions (see Def 2.9, §2.2.2) is closed under intersection.

*Proof* Whenever $m$ is a completely convex function and $A, B \subset N$ write: $\underline{(m \wedge V)}(A \cup B) + \underline{(m \wedge V)}(A \cap B) - \underline{(m \wedge V)}(A) - \underline{(m \wedge V)}(B) = \underline{m}((A \cap V) \cup (B \cap V)) + \underline{m}((A \cap V) \cap (B \cap V)) - \underline{m}(A \cap V) - \underline{m}(B \cap V) \geq 0$. ■

CONSEQUENCE B.1 The structural ordering satisfies in addition to (V.1)–(V.5) the properties:

$$v \prec v \Rightarrow (u \,\Delta\, V) \prec (u \,\Delta\, V) \quad \text{whenever } u, v \in Z(\mathcal{U}),\ V \subset N \qquad \text{(V.6)}$$

$$0 \leq u \Rightarrow (u \,\Delta\, V) \prec u \quad \text{whenever } u \in Z(\mathcal{U}),\ V \subset N. \qquad \text{(V.7)}$$

*Proof* By Theorem 2.4b, §2.2.3 $\prec$ is induced by the class of completely convex set functions. Combine Lemma B.1, Lemma B.2 and Lemma 2.7b, §2.2.2. ■

Note that the same method can be used to show that the standard multiinformation ordering (see Def 2.18, §2.3.5) satisfies (V.6) and (V.7).

## CONCLUSIONS

The aim of entire series of papers was to introduce a new approach to the description of CI-structures. It provides the concept of face as an alternative to the concept of dependency model for description of these structures. It was indicated how to represent faces in a computer and the connection between faces and dependency models was established. Here I try to discuss advantages and disadvantages of and prospects for the presented approach.

### *Contribution*

First, I would like to highlight what I consider to be the contribution made by this series of papers.

1. The presented theory gives a deductive mechanism to infer probabilistically valid consequences of positive information about CI-structure. This mechanism is more powerful than the well-known semigraphoid mechanism (treated more or less explicitly for example in [Dawid, 1979], [Spohn, 1980], [Pearl, 1988], [Smith, 1989]). The approach makes it possible to prove further independent qualitative properties of stochastic CI i.e. to verify probabilistic soundness of conjectured inference rules[7]—the simple method is described in Example 3.7.
2. Although the mentioned deductive mechanism involves an infinite number of independent inference rules (see Remark 2.3, §2.3.2) it is "finitely implementable" from a theoretical point of view. Namely, whenever the skeleton is at your disposal (its existence was proved) you can provide a simple efficient implication algorithm—see Example 3.2.

[7]Note that many authors, influenced by results of Armstrong [1974] in the theory of relational databases, called these inference rules 'axioms'.

3. The presented approach allows the input of information about CI-structure in various forms. The basic connection with dependency models enables you to enter individual CI-statements (as a result of a statistical test or expert's testimony)—see Example 3.3. Moreover, both Bayesian networks and decomposable models (i.e. chordal undirected graphs) have a simple, natural form of input—see Example 3.4, 3.5. Note that information from different sources can be easily combined (see Example 3.6).
4. In contrast to the classical approach where only individual CI-statements were inferred this deductive mechanism can also infer the validity of a whole CI-structure (for example the validity of a Bayesian network)—this was shown in Example 3.6.
5. In contrast to both types of graphical representation of information about CI-structure (undirected and directed acyclic graphs) this approach involves all possible probabilistic CI- structures—see Consequence 2.9, §2.3.4. Of course, semigraphoids also involve all CI-structures, but as already mentioned, the structural faces approach says more.
6. The result about an equivalent description of validity of a facial model by means of the so-called product formula from [Studený, 1993] is the first step to a justified interpretation of all these models of CI-structure. Thus, it seems to me that the facial approach to description of CI-structures has at least as sound an interpretation as hierarchical log-linear models in statistics.
7. The internal computer representation of faces by means of generating imsets enables you to join and save information without its loss (in contrast to reversing arcs in the influence diagram method [Shachter, 1990]). As concerns representation in a computer, an imset requires card $\mathcal{U} = 2^{\text{card}N} - \text{card}\, N - 1$ integers in a memory of a computer while a dependency model $4^{\text{card}N} + 2^{\text{card}N} - 2 \cdot 3^{\text{card}N}$ bits (although in the case of a semigraphoid it can be reduced to $\text{card}\, N \cdot (\text{card}\, N - 1) \cdot 2^{\text{card}N-3}$).
8. Multiple joint conditional independence has a natural internal representation by means of imsets—see Remark 3.1, §3.1.4.

*Tasks to be Solved*

There remain certain problems:

1. So, far I have no sufficiently convenient characterization of the skeleton allowing an algorithm to construct it for an arbitrary given number of variables or attributes. This may be a serious problem as the verification of completeness of the class of skeletal imsets was very tedious in the case of 4 attributes—see [Studený, 1991].
2. Unfortunately, structural faces do not completely correspond to probabilistic CI-structures. There exists a structural face which is not probabilistically representable (even in the case of 4 attributes, namely all 6 co-atomic faces corresponding to the skeletal imsets from the class IX mentioned in Table 2.1, §2.2.4). I must emphasize that the essential step to this discovery was made by my colleague F. Matúš [1994] who found that there exists a matroid which is not probabilistically representable. Knowing this I succeeded in finding fur-

ther independent properties of probabilistically representable dependency models which do not hold for general structural semigraphoids:

$$A \perp B \mid C \;\&\; A \perp B \mid D \;\&\; B \perp C \mid A \;\&\; C \perp D \mid \emptyset \Rightarrow B \perp AC \mid \emptyset \qquad \text{(B.1)}$$

$$A \perp B \mid D \;\&\; A \perp C \mid B \;\&\; B \perp C \mid A \;\&\; C \perp D \mid \emptyset \Rightarrow AB \perp C \mid \emptyset \qquad \text{(B.2)}$$

The proofs of these properties are still in manuscript; Matúš and I plan to publish it later after deriving more results.

Finally, I would like to mention some prospects for the presented approach:

1. Although the described theory did not fulfill the original aim to completely correspond to probabilistic CI-structures it seems that it can be successfully modified in order to be able to describe precisely CI-structures within a special "nice" class of probability measures $\Phi$. In fact, this is the very reason for the development of the theory of faces for the general scalar product ordering on imsets (see section 2.1). Consider the multiinformation ordering (see Def 2.18, §2.3.5) corresponding to $\Phi$; in case that it is finitely established there is a fair hope that faces with respect to this multiinformation ordering both fit CI-structures within $\Phi$ and have an efficient deductive mechanism. This may also be the case for the standard multiinformation ordering! So far, I don't know how the standard multiinformation ordering (in case of 4 attributes) looks.
2. Perhaps a natural way of input of graphical models (i.e. undirected graphs) could be found. A proper set of triplets giving the dependency model represented by such a graph as its semigraphoid closure is needed.

ACKNOWLEDGMENT

I am indebted to František Matúš who helped me to improve the presented theory and to an anonymous referee who helped me to improve substantially its presentation. I also thank Pavel Boček for writing a computer program clarifying the situation in the case of four attributes and Václav Kelar for help with figures.

REFERENCES

Armstrong, W. W. (1974), [1974], "Dependency structures of database relationships." In: *Proceedings of IFIP 1974*, North-Holland, pp. 580–583.

Dawid, A. P. [1979], "Conditional independence in statistical theory." *Journal of the Royal Statistical Society B*, **41**(1), pp. 1–31.

Geiger, D. and J. Pearl [1990a], "On the logic of causal models." In: *Uncertainty in Artificial Intelligence 4*, edited by R. D. Shachter, T. S. Lewitt, L. N. Kanal and J. F. Lemmer, North Holland, Amsterdam, pp. 3–14.

Geiger, D. and J. Pearl [1990b], "Logical and algorithmic properties of conditional independence and their application to Bayesian networks." *Annals of Mathematics and Artificial Intelligence*, **2**(1/4), pp. 165–178.

Lauritzen, S. L., T. P. Speed and K. Vijayan [1984], "Decomposable graphs and hypergraphs." Journal of Australian Mathematical Society A, **36**(1),pp. 12–29.

Lauritzen, S. L., A. P. Dawid, B. N. Larsen and H.-G. Leimer [1990], "Independence properties of directed Markov fields." *Networks*, **20**(5), pp. 491–505.

Matúš, F. [1992], "Ascending and descending conditional independence relations." In: *Information Theory Statistical Decision Functions Random Processes: transactions of the 11th Praque Conference* *vol. B*, edited by S. Kubík and J. A. Víšek, Kluwer, Dordrecht (also Academia, Prague), pp. 189–200.

Matúš, F. [1994], "Probabilistic conditional independence structures and matroid theory: backgrounds." *International Journal of General Systems*, **22**(2), pp. 185–196.

Pearl, J. [1988], *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufman, San Mateo.

Pearl, J., D. Geiger and T. Verma [1990], "The logic of influence diagrams." Chapter 4 in *Influence Diagrams, Belief Nets and Decision Analysis*, edited by R. M. Oliver and J. Q. Smith, John Wiley, Chichester.

Shachter, R. D. [1990], "An ordered examination of influence diagrams." *Networks*, **20**(5), pp. 535–563.

Smith, J. Q. [1989], "Influence diagrams for statistical modelling." *Annals of Statistics*, **17**(2), pp. 654–672.

Spohn, W. [1980], "Stochastic independence, causal independence and shieldability." *Journal of Philosophical Logic*, **9**(1), pp. 73–99.

Studený, M. [1991], "Convex set functions I and II." research reports n. 1733 and n. 1734, Institute of Information Theory and Automation, Prague, November 1991.

Studený, M. [1992a], "Conditional independence relations have no finite complete characterization." In: *Information Theory Statistical Decision Functions Random Processes: transactions of the 11th Prague Conference vol. B*, edited by S. Kubík and J. A. Víšek, Kluwer, Dordrecht (also Academia, Prague), pp. 377–396.

Studený, M. [1992b], "Multiinformation and conditional independence II." research report n. 1751, Institute of Information Theory and Automation, Prague, September 1992.

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

Studený, M. [1994], "Structural semigraphoids." *International Journal of General Systems*, **22**(2), pp. 207–217.

*For a photo and biography of the author, please see 22(2), page 217, of this journal. The new author's e-mail address is studeny@utia.cas.cz*