Relationship between properties characterizing independence equivalence in Bayesian networks and compositional models

Václav Kratochvíl*

Institute of Inf. Theory and Automation
The Academy of Sciences of the Czech Rep.
velorex@utia.cas.cz

Abstract: We present a simple characterizations of equivalent compositional model structures based on (A) invariant properties and (B) local transformations. It has been shown in [1] that one can transform any compositional model into Bayesian network representing the same joint probability distribution and vice versa.

Moreover, every assertion of independence induced by a Bayesian network structure is also induced by the structure of the respective compositional model that is created from the Bayesian network and vice versa.

That is why wa can simply compare characterization of equivalent compositional model structures together with the known characteristics of equivalent Bayesian network structures. We show which (A) invariant properties and (B)local transformations correspond each other. In opposite case we show what is its (invariant property, transformation) meaning in the other model structure.

Keywords: Compositional model, Bayesian network, equivalence problem

1 Introduction

The ability to represent and process multidimensional probability distributions is a necessary condition for application of probabilistic methods in Artificial Intelligence. Among the most popular approaches are the methods based on Graphical Markov Models, e.g., Bayesian Networks. An alternative approach to Graphical Markov Models are the so-called compositional models.

A Bayesian network for a set of variables $N = \{u_1, \ldots, u_{|N|}\}$ represents a joint probability distribution over those variables. It consists of (1) a network structure that encodes assertions of conditional independence in the distribution and (2) a set of conditional probability distribution corresponding to that structure. The network structure is an acyclic directed graph (DAG for short) such that each variable $u \in U$ has a corresponding node u in the structure.

Consider a system of variables N. Let $G = (N, E)$ is a dag. Denote for each $i \in N$: $pa(i) = \{j \in N : (j \to i) \in E\}$. We say that a probability distribution $\kappa(N)$ is a Bayesian network with graph G if

$$\kappa(N) = \prod_{i \in N} \kappa(i|pa(i)) \tag{1}$$

Similarly, compositional model for a set of variables $N = \{u_1, \ldots, u_{|N|}\}$ represents a joint probability distribution over those variables as well. It is represented by a sequence of low-dimensional probability distributions that are composed together by *operator of composition* $\triangleright$

*This work was supported by National Science Foundation of the Czech Republic under Grants No. ICC/08/E010, and 201/09/1891, and by Ministry of Education under Grants No. 1M0572 and 2C06019.

(which is defined in Definition 2). Hence, by a compositional model we understand a sequence $\pi_1(K_1), \ldots, \pi_n(K_n)$ representing a joint distribution $\pi_1(K_1) \triangleright \ldots \triangleright \pi_n(K_n)$ where the operator $\triangleright$ is applied from left to right. Note that unlike the Bayesian networks, structure of a compositional model is not defined anywhere sideways: *We say that the sequence of sets of variables $K_1, \ldots, K_n$ is the structure of compositional model $\pi_1(K_1), \ldots, \pi_n(K_n)$.*

For both compositional model and Bayesian network and for any given structure, there is a corresponding set of probability distributions that can be represented using a model with that structure/network. Two model structures are *equivalent* if the set of distributions that can be represented using one of the structure/dag is identical to the set of distributions that can be represented using the other. Or in another words: *Two model structures are* equivalent *if they induce the same set of conditional independence assertions - the same independence model.* Because equivalence is reflexive, symmetric, and transitive, the relation defines a set of equivalence classes over structure/dag. Note that the notion of equivalence is of particular importance for learning these models from data.

2 Notation

In this section we introduce our notation and properties necessary to discuss equivalence in both Bayesian networks and compositional models.

2.1 Bayesian network

We write $u \leftrightarrow_G v$ to denote that there is an *edge* or an adjacency between nodes u and v in G which means that either $u \to v$ or $u \leftarrow v$ in G. The set of edges in directed graph G is the collection of two-element subsets of N:

$$\mathcal{E}(G) = \{\{u, v\}; u \leftrightarrow_G v\}.$$

We say that distinct nodes u, v, w form an *immorality* in G and write $(u, v) \rightsquigarrow w[G]$ if $u \to w$ in $G, v \to w$ in G, and $u \not\leftrightarrow_G v$. In fact an immorality in dag G is nothing but a special induced subgraph of G: $G[u, v, w]$.

Example 1. *Consider a graph G from Figure 1b. Then there are two immoralities $(u, v) \rightsquigarrow x[G], (u, w) \rightsquigarrow x[G]$ in G. $\mathcal{E}(G) = \{\{v, w\},\{u, x\}, \{v, x\},\{v, w\},\{v, z\},\{w, z\}\}$.*

2.2 Compositional models

Definition 2. *For two arbitrary distributions $\pi_1(U)$ and $\pi_2(V)$ their* composition *is given by the formula*

$$\pi_1(U) \triangleright \pi_2(V) = \frac{\pi_1(U)\pi_2(V)}{\pi_2(U \cap V)}$$

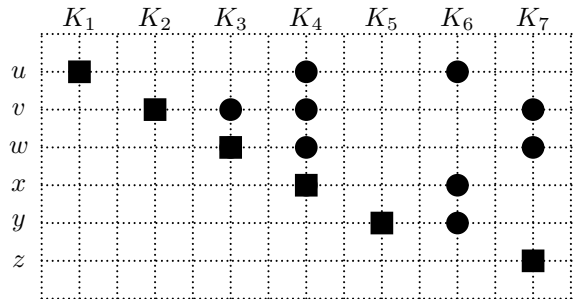
if $\pi_1(U \cap V) \ll \pi_2(U \cap V)$, otherwise the composition remains undefined.

The symbol $\pi_1(K) \ll \pi_2(K)$ means that $\pi_1(K)$ is dominated *by $\pi_2(K)$, which in its turn means (in the considered finite setting) $\forall x \in \times_{j \in K} \mathbf{X}_j; (\pi_2(x) = 0 \Longrightarrow \pi_1(x) = 0)$. Moreover, if there is a product of two zeros in the numerator and we take $\frac{0 \cdot 0}{0} = 0$*

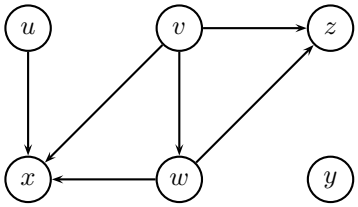
Considering a compositional model with structure $\mathcal{P} = K_1, \ldots, K_n$, each set K_i can be divided into two disjoint parts with respect to the structure $\mathcal{P}$. We denote them $R(K_i^{\mathcal{P}})$ and $S(K_i^{\mathcal{P}})$, where $R(K_i^{\mathcal{P}}) = K_i \backslash (K_1 \cup \ldots \cup K_{i-1})$ and $S(K_i^{\mathcal{P}}) = K_i \cap (K_1 \cup \ldots \cup K_{i-1})$.

It has the following meaning: $R(K_i^{\mathcal{P}})$ denotes the variables first occurring in the structure (as well as in model itself). $S(K_i^{\mathcal{P}})$ denotes those variables that have been already used. Observe that $K_i^{\mathcal{P}} = R(K_i^{\mathcal{P}}) \cup S(K_i^{\mathcal{P}})$.

Example 3. *Let $\pi_1(K_1) \rhd \ldots \rhd \pi_7(K_7)$ be a compositional model such that. $K_1 = \{u\}, K_2 = \{v\}, K_3 = \{v, w\}, K_4 = \{u, v, w, x\}, K_5 = \{y\}, K_6 = \{u, x, y\}$, and $K_7 = \{v, w, z\}$. Then one can see its structure $\mathcal{P} = K_1, \ldots, K_6$ depicted on Figure 1a where columns correspond to $\{K_i\}_{i=1,\ldots,7}$. Rows correspond to variables. Sets $\{R(K_i)\}_{i=1,\ldots,7}$ are denoted by box-markers while $\{S(K_i)\}_{i=1,\ldots,7}$ by bullets. Observe that box-marker is always the first marker in each row.*



(a) Visualization of a compositional model structure



(b) Dag - Bayesian network structure

Figure 1: Two different structures that induce the same independence model

We write $u \leftrightarrow_{\mathcal{P}} v$ to denote that there is a set $K_i \in \mathcal{P}$ such that $u, v \in K_i$ and $u \in R(K_i^{\mathcal{P}})$ or $v \in R(K_i^{\mathcal{P}})$. The *connection set* of $\mathcal{P}$ is the collection of two-element subsets of N:

$$\mathcal{E}(\mathcal{P}) = \{\{u, v\} : u, v \in N, u \leftrightarrow_{\mathcal{P}} v\}$$

We say that distinct variables u, v, w form a *F-condition* in $\mathcal{P}$ and write $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$ if $\exists K_i \in \mathcal{P}$ such that $w \in R(K_i^{\mathcal{P}}), \{u, v\} \subseteq S(K_i^{\mathcal{P}})$, and $u \nleftrightarrow_{\mathcal{P}} v$. The set of all F-conditions induced by $\mathcal{P}$ is denoted by $\mathcal{F}(\mathcal{P})$. Observe that if $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$ then $u \leftrightarrow_{\mathcal{P}} w$ and $v \leftrightarrow_{\mathcal{P}} w$.

Example 4. *One can find the following connections and F-conditions in structure $\mathcal{P}$ from Figure 1a: $\mathcal{F}(\mathcal{P}) = \{\langle u, v|x\rangle, \langle u, w|x\rangle\}$, $\mathcal{E}(\mathcal{P}) = \{\{v, w\}$,$\{u, x\}$, $\{v, x\}$,$\{v, w\}$,$\{v, z\}$,$\{w, z\}\}$. Notice that $\mathcal{E}(\mathcal{P}) = \mathcal{E}(G)$ for G from Figure 1b and $\mathcal{F}(\mathcal{P})$ corresponds to its immoralities.*

For another characteristics, we need a definition of *structure core*, which is closely related to the special shape of a structure - the so-called *reduced structure*.

Definition 5. *For a structure $\mathcal{P}$, its* core *$\mathcal{C}(\mathcal{P})$ is a collection of $K_i \in \mathcal{P}$ such that $K_i \neq S(K_i)$ and $K_i \neq S(K_j)$ for all $K_j \in \mathcal{P}$ such that $R(K_j^{\mathcal{P}}) \neq \emptyset$.*

Structure where set of columns coincide with its core is called reduced structure.

Example 6. *Observe that for $\mathcal{P} = K_1, \ldots, K_7$ from Figure1a holds that $\mathcal{C}(\mathcal{P}) = \{K_1, K_4, K_5, K_7\}$. Indeed, notice $K_2 = S(K_3^{\mathcal{P}}), K_3 = S(K_7^{\mathcal{P}})$ and $K_6 = S(K_6^{\mathcal{P}})$.*

The following lemma is stated without proof. The proof is based on the characteristics of local transformations and avid reader may find it in [3]. We denote the reduced form of structure $\mathcal{P}$ by symbol $\mathcal{P}^{\mathcal{C}}$.

Lemma 7. *Every structure can be transformed into its equivalent reduced form in the following way:*

1: *$\mathcal{P}' = \mathcal{P}$ where all columns K_i such that $K_i = S(K_i^{\mathcal{P}})$ are removed;*
2: ***while** $\mathcal{P}'$ is not reduced **do***
3: $(i, j) = \{i, j \in \{1, \ldots, |\mathcal{P}'|\}$ *such that* $i < j; K_i^{\mathcal{P}'} = S(K_j^{\mathcal{P}'})\}$; {*such pair exists since $\mathcal{P}'$ is not reduced*}

4: $\mathcal{P}' = K_1^{\mathcal{P}'}, \ldots, K_{i-1}^{\mathcal{P}'}, K_j^{\mathcal{P}'}, K_{i+1}^{\mathcal{P}'}, \ldots, K_{j-1}^{\mathcal{P}'}, K_{j+1}^{\mathcal{P}'}, \ldots, K_{|\mathcal{P}'|}^{\mathcal{P}'}$ {*remove column* K_i *and move column* K_j *on* $i-th$ *position*}

5: ***end while***

6: $\mathcal{P}^{\mathcal{C}} = \mathcal{P}'$;

7: ***return*** $\mathcal{P}^{\mathcal{C}}$;

Example 8. *Following the algorithm presented in the previous lemma and structure* $\mathcal{P}$ *from Figure 1a, one can create the following sequence of equivalent structures (See Figure 2) where the last one is reduced:* $K_1, \ldots, K_7 \Rightarrow K_1, K_2$, K_3,$K_4, K_5, K_7 \Rightarrow K_1, K_3, K_4, K_5, K_7 \Rightarrow K_1, K_7, K_4, K_5 = \mathcal{P}^{\mathcal{C}}$. *See Example 6 and observe that* $\mathcal{P}^{\mathcal{C}}$ *consists of columns that are contained in* $\mathcal{C}(\mathcal{P})$ *only. Moreover, notice that* $\mathcal{E}(\mathcal{P}) = \mathcal{E}(\mathcal{P}^{\mathcal{C}})$ *and* $\mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}^{\mathcal{C}})$.

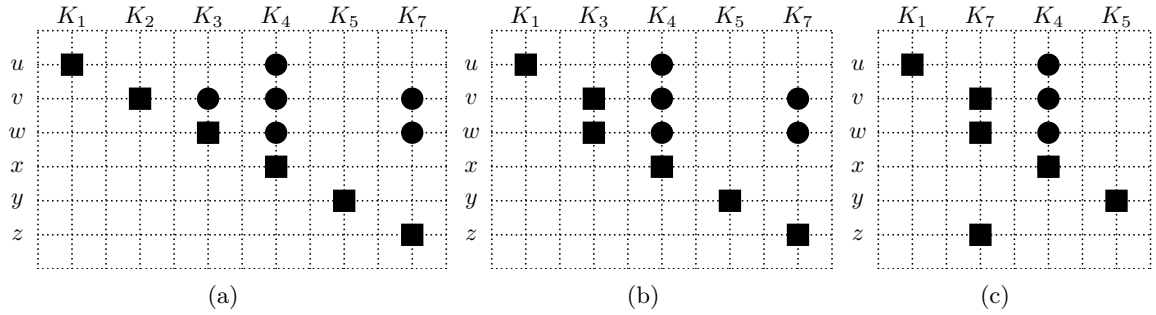


Figure 2: The process of reduction

For the purpose of characterization of equivalent structures, we introduce the following additional notation: By a symbol $S(\mathcal{P}^{\mathcal{C}})$ we understand a collection of sets $\{S(K_i^{\mathcal{P}^{\mathcal{C}}})\}$ for all $K_i^{\mathcal{P}^{\mathcal{C}}}$.

Example 9. *For a structure* $\mathcal{P}$ *from Figure 1a holds that* $S(\mathcal{P}^{\mathcal{C}}) = \{\{u, v, w\}\}$. *Indeed, see its reduced form* $\mathcal{P}^{\mathcal{C}}$ *depicted on Figure 2c.*

3 Equivalence

From the definition of Bayesian network (1) one can immediately see that a Bayesian network $\kappa(N)$ is uniquely defined by a system of conditional distributions $\{i|pa(i)\}_{i\in N}$. Define $\pi(\{i\} \cup pa(i))$ such that $\pi(i|pa(i)) = \kappa(i|pa(i))$. To do this, one can take, for example, a uniform distribution $\mu(pa(i))$ and set $\pi(\{i\} \cup pa(i)) = \kappa(i|pa(i))\mu(pa(i))$.

Now we immediately see that the considered Bayesian network can be expressed in the form $\kappa = \pi_{j_1} \rhd \ldots \rhd \pi_{j_N}$, if the permutation $j_1, \ldots, j_N$ is such that all parents of a node are always before it: $j_k \in pa(j_l) \Rightarrow k < l$. Therefore, each Bayesian network can be represented as a compositional model. Similarly, we can find a Bayesian network representation for a distribution defined as a compositional model. Since this article deals with the independence relations induced by structure of the corresponding model only, we show how to convert structure of a compositional model into equivalent dag:

Algorithm 10. *Having a structure* $\mathcal{P} = K_1, \ldots, K_n$ *we first order (in an arbitrary way) all the variables from* $(K_1 \cup \ldots \cup K_n) = N$. *Then a graph* $G(N, E)$ *of the constructed Bayesian network is defined in the following way:* $u, v \in N$, $(u \to v) \in E$ *if there exists* $K_i \in \mathcal{P}$ *such that (A)* $u \in S(K_i^{\mathcal{P}})$, $v \in R(K_i^{\mathcal{P}})$ *or (B)* $u, v \in R(K_i^{\mathcal{P}})$ *and* u *is in the ordering defined in the first step before* v.

Observe that for a structure $\mathcal{P}$ and dag G obtained from $\mathcal{P}$ by application of Algorithm 10 holds that $\mathcal{E}(\mathcal{P}) = \mathcal{E}(G)$ and $\mathcal{F}(\mathcal{P})$ corresponds to the set of immoralities induced by G. (Notice that if one apply this algorithm on structure $\mathcal{P}$ from Figure 1a then the dag from Figure 1b will be obtained.)

Both theories are able to model the same set of probability distributions. Let us compare tools to determine the equivalence of structures.

3.1 Direct characterization

In this section we present several invariant properties of equivalent compositional model structures as well as corresponding properties of equivalent dags. The first direct characterization of equivalent compositional model structures corresponds to the well known direct characterization of equivalent dags which is closely related to the so called *essential graph*. (See conversion Algorithm 10.)

- Compositional model: *Two structures $\mathcal{P}, \mathcal{P}'$ over N are independence equivalent if and only if $\mathcal{E}(\mathcal{P}) = \mathcal{E}(\mathcal{P}')$ and $\mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}')$.*
- Bayesian network: *Two dags G, G' over N are equivalent if and only if $\mathcal{E}(G) = \mathcal{E}(G')$ and the graphs G, G' have the same immoralities*

Unlike Bayesian networks, this characterization is not very appropriate for compositional models. For example, try to compare *connection sets* corresponding to structures from Figures 1a and 2c. It is quite difficult even for such simple structures. For this reason, further characterization based on columns was derived.

Lemma 11. *Two structures $\mathcal{P}, \mathcal{P}'$ over N are independence equivalent if and only if $\mathcal{C}(\mathcal{P}) = \mathcal{C}(\mathcal{P}')$ and $S(\mathcal{C}^{\mathcal{P}}) = S(\mathcal{C}^{\mathcal{P}})$.*

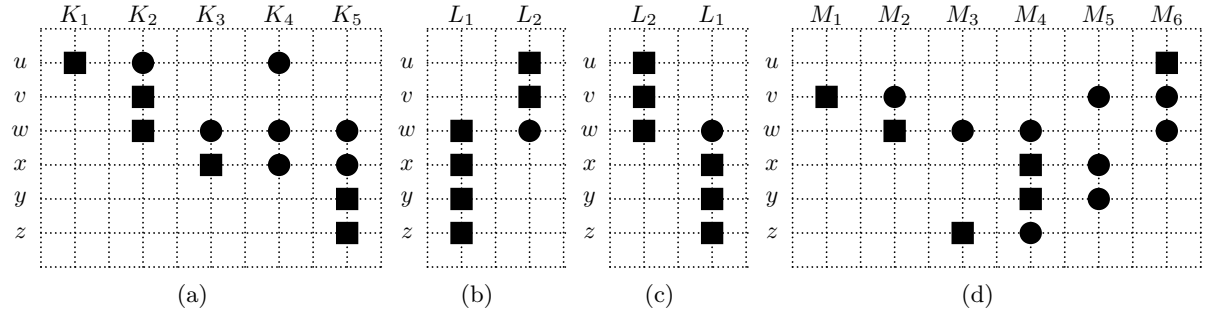


Figure 3: Independence equivalent structures

What is the corresponding property of *structure core* in the area of equivalent dags? Note that core correspond to a collection of sets $\{pa(u) \cup \{u\}\}$ for some $u \in N$. Nevertheless, a very similar method for testing equivalence of dags over N is partially based on the approach from [2] where one writes a 'formal ratio' for every dag G over N as follows: in the nominator one lists sets $pa_G(u) \cup \{u\}$ for $u \in N$ while in the denominator one lists the sets $pa_G(u)$ for $u \in N$. Then the cancelation is performed: one occurrence of a set $U \subseteq N$ in the denominator is canceled against one occurrence of U in the nominator. For example the dag G from Figure 1b induces the following 'ratio':

$$\frac{u \cdot v \cdot zvw \cdot xuvw \cdot wv \cdot y}{\emptyset \cdot \emptyset \cdot vw \cdot uvw \cdot v \cdot \emptyset} = \frac{u \cdot zvw \cdot xuvw \cdot y}{uvw}$$

Considering compositional model structure, the *structure core* is a set of columns that are not removed during 'cancelation' with sets $\mathcal{S}(K_i)$. Hence if one defines a 'ratio' where in the nominator one lists sets $\{K_i^{\mathcal{P}}\}_{i=1,\ldots,n}$ while in the denominator one lists the sets $\{\mathcal{S}(K_i^{\mathcal{P}})\}_{i=2,\ldots,n}$, then after cancelation the nominator correspond to $\mathcal{C}(\mathcal{P})$ while denominator to $\mathcal{S}(\mathcal{C}^{\mathcal{P}})$ - see Algorithm 10. These two characterization coincide. For example the structure from Figure 1a induces the following 'ratio':

$$\frac{u \cdot v \cdot vw \cdot uvwx \cdot y \cdot uxy \cdot vwz}{\emptyset \cdot v \cdot uvw \cdot \emptyset \cdot uxy \cdot vw} = \frac{u \cdot zvw \cdot xuvw \cdot y}{uvw}$$

3.2 Indirect characterization

In the case of Bayesian networks, there is a local transformation - the so called *legal arrow reversal* - which enables to generate all dags that are equivalent to a given one. By a *legal arrow reversal* we understand the change of dag K into dag L by replacement of an arrow $u \to v$ (in K) by $u \leftarrow v$ (in L) under the condition that $pa_K(u) \cup \{u\} = pa_K(v)$ (here $u, v \in N$ are some distinct nodes). Two dags, if one is obtained from the other by legal arrow reversal, are equivalent. For example: the change of direction of the arrow $v \to x$ in graph on Figure 1b is *legal arrow reversal*.

In case of compositional model, there are three local transformation - the so called *IE operations* - which similarly enables to generate all structures that are equivalent to a given one. Restriction to a set of reduced structures suggests that two equivalent structures consist of the same columns (possibly in different order). I.e. one is a permutation of the other. It is no surprise that two of the elementary operations are transposition of adjacent columns: *constant transposition and box transposition*. The third transformation that removes/adds columns composed from bullets only ($K_i^{\mathcal{P}} = \mathcal{S}(K_i^{\mathcal{P}})$) is called *simple reduction/extension*.

Let $\mathcal{P} = K_1, \ldots, K_n$ and $\sigma = (k-1, k)$ such that $k \leq n$. $\sigma = (k-1, k)$ is a transposition and $\mathcal{P}\sigma = K_1, \ldots, K_{k-2}, K_k, K_{k-1}, K_{k+2}, \ldots, K_n$. ($K_{k-1}$ and K_k changed their positions).

For $\mathcal{P} = K_1, \ldots, K_n$ and $k \in \{2, \ldots, n\}$ a transposition $\sigma = (k-1, k)$ is *constant* if $R(K_{k-1}^{\mathcal{P}}) \cap K_k^{\mathcal{P}} = \emptyset$. We say that $\mathcal{P}\sigma$ is *constant transposition* of $\mathcal{P}$. Observe that $\mathcal{P}' = K_1, K_2, K_3, K_5, K_4$ is *constant transposition* of $\mathcal{P} = K_1, \ldots, K_5$ from Figure 3a.

We call transposition $\sigma = (k-1, k)$ a *box transposition* in $\mathcal{P} = K_1, \ldots, K_n$ if $k \in \{2, \ldots, n\}$ and $\mathcal{S}(K_{k-1}^{\mathcal{P}}) \subseteq \mathcal{S}(K_k^{\mathcal{P}}) \subseteq K_{k-1}^{\mathcal{P}}$. We say that $\mathcal{P}\sigma$ is *box transposition* of $\mathcal{P}$. Observe that structures from Figures 3b and 3c are mutually its own *box transpositions*. Similarly, $\mathcal{P}'' = K_2, K_1, K_3, K_4, K_5$ is *box transposition* of $\mathcal{P} = K_1, \ldots, K_5$ from Figure 3a.

In the case of indirect characterization, the individual operations do not meet each other in a 1:1 ratio. *Legal arrow reversal* can reflect the application of several *IE operations* in the corresponding structure and vice versa. Rather, it is closer to realize parallels between *IE operations* and *formal ratio* that was mentioned at the end of the previous chapter. Specifically, *simple reduction* corresponds precisely to the classical cancelation of the common factor.

References

[1] R. Jiroušek. Multidimensional Compositional Models. *Preprint DAR - ÚTIA 2006/4, ÚTIA AV ČR, Prague* (2006).

[2] M. Studený. On mathematical description of probabilistic conditional independence structures *thesis for DrSc. degree, ÚTIA AV ČR, Prague* (2001).

[3] V. Kratochvíl. Equivalence Problem in Compositional Models *WUPES'09 , Eds: Kroupa T., Vejnarová J., Liblice, CZ* (2009).