Foundations of Compositional Models: Structural Properties

R. Jiroušek[a,b] *and *V. Kratochvíl*[a,b] †

[a]Institute of Information Theory and Automation, Prague, CZ

[b]Faculty of Management, University of Economics, Jindřichův Hradec, CZ

1 Introduction

This paper is a follow-up of [Jiroušek, 2011], where basic properties of compositional models, as one of the approaches to multidimensional probability distributions representation and processing, were introduced. Similar to other methods, such as Bayesian networks, the compositional models take advantage of the properties of conditional independence to decrease the number of parameters necessary for their representation. In fact, each of the approaches for multidimensional model representation has a way to specify the system of conditional independence relationships valid for the considered probability distribution.

One of the first applications of graphs for this purpose appears in papers from the field of genetics by [Wright, 1921]. However, it is known that not all systems of conditional independence statements induced by a probability distribution can be described by a single graph. It was shown by [Verma, 1987] that *the spectrum of probabilistic dependencies is in fact so rich that it cannot be cast into any representation scheme that uses a polynomial amount of storage.* Being unable to provide a perfect mapping at a reasonable cost, one compromises the requirement that a respective tool such as a graph represents each and every dependency of a probability distribution, and allows some independencies to escape the representation [Geiger and Pearl, 1988].

The compositional models that were introduced as an algebraic alternative to graphical models do not use graphs to represent conditional independence statements. Here, these statements are encoded in a sequence of distributions to which an operator of composition – the key element of this theory – is applied in order to assemble a multidimensional model from its low-dimensional parts. More precisely, a sequence of sets of variables – which will be called a *model structure* in this paper – plays the same role as a graph in the case of graphical

*1st author: radim @utia.cas.cz
†2nd author: velorex @utia.cas.cz

modeling.

Recall that several ways to read conditional independence statements have been designed in the area of graphical models. In the case of undirected graphs, conditional independence relations can be uncovered using a graph separation criterion. In the case of acyclic directed graphs, the so-called *d-separation* criterion designed by Pearl is often used [Pearl, 1988]. An alternative test for d-separation was devised in [Lauritzen, Dawid, Larsen, and Leimer, 1990]. It is based on the notion of moralized ancestral graphs.

In Section 3.2 of this paper we show a way to read conditional independence relations for compositional models. A related topic is the so-called equivalence problem, i.e., the problem of recognizing whether two different structures induce the same system of conditional independence statements. For Bayesian networks the problem was solved by [Verma, Pearl, 1991]; two acyclic directed graphs induce the same independence structure if they have the same *adjacencies* and *immoralities* (the latter are special induced subgraphs). Later, a unique representation of a class of equivalent graphs was found, the so-called *essential graph* [Andersson, Madigan, and Perlman, 1997]. In the present paper both these problems are solved for the compositional models in Section 3.3.

Having two different structures inducing the same system of conditional independence statements, it may be of special importance to have an easy way to transform one structure into the other using some elementary operations. This issue was treated in a case of directed graphs in [Chickering, 1995] by *legal arrow reversal*. To solve this problem for compositional models, we introduce a special system of operations in Section 4.

Section 5, incorporated into the text at the suggestion of the anonymous reviewer, describes the relationship between the compositional and graphical approaches to multidimensional probability distribution representation.

In the last two sections of this paper we show how the structural properties are manifested in the properties of the multidimensional probability distributions represented in the form of compositional models.

2 Basic notions and notation

In this paper, we deal with a finite system of finite-valued variables $\{u, v, x, \ldots\}$, sets of which will be denoted by upper-case Roman characters such as K, U, V, W, Z, with possible indices. Ordered sequences of variable sets will be denoted by calligraphic characters like $\mathcal{P} = (U, W, Z, V)$, $\mathcal{P}' = (K_1, K_2, K_3, K_4, K_5)$, or, $\mathcal{P}'' = (K_1, K_3, K_5, K_4, K_2)$. Notice that here $\mathcal{P}' \neq \mathcal{P}''$ because $\mathcal{P}''$ is a reordering of $\mathcal{P}'$. Symbol $|\mathcal{P}|$ denotes the number of sets in the sequence, i.e., for the previously introduced sequences $|\mathcal{P}| = 4$ and $|\mathcal{P}'| = |\mathcal{P}''| = 5$.

Lower-case Greek characters will denote probability distributions, e.g., $\pi(K)$ will denote a probability distribution defined for variables from K. Its marginal distribution for variables from $U \subset K$ will be denoted by either simply $\pi(U)$, or $\pi^{\downarrow U}$. For $U = \emptyset$, $\pi^{\downarrow \emptyset} = 1$.

2.1 Conditional independence

One of the most important notions of this paper, a concept of *conditional independence*, generalizes the well-known independence of variables.

Definition 2.1. *Consider a probability distribution $\pi(K)$ and three disjoint subsets $U, V, Z \subseteq K$ such that both $U, V \neq \emptyset$. We say that groups of variables U and V are* conditionally independent *given Z for probability distribution π (in symbols $U \perp\!\!\!\perp V | Z[\pi]$) if*

$$\pi^{\downarrow U \cup V \cup Z} \pi^{\downarrow Z} = \pi^{\downarrow U \cup Z} \pi^{\downarrow V \cup Z}.$$

In many basic books on probabilistic multidimensional models (e.g., [Cowell, Dawid, Lauritzen, and Spiegelhalter, 1999, Pearl, 1988, Studený, 2005]), one can find the following important properties of conditional independence that are universal and common for different formalization:

symmetry $\quad U \perp\!\!\!\perp V | Z[\pi] \Leftrightarrow V \perp\!\!\!\perp U | Z[\pi] \quad$ (2.1)

decomposition $\quad U \perp\!\!\!\perp (V \cup W) | Z[\pi] \Rightarrow U \perp\!\!\!\perp V | Z[\pi] \quad$ (2.2)

weak union $\quad U \perp\!\!\!\perp (V \cup W) | Z[\pi] \Rightarrow U \perp\!\!\!\perp W | (V \cup Z)[\pi] \quad$ (2.3)

contraction $\quad U \perp\!\!\!\perp W | (V \cup Z)[\pi]\ \&\ U \perp\!\!\!\perp V | Z[\pi] \Rightarrow U \perp\!\!\!\perp (V \cup W) | Z[\pi] \quad$ (2.4)

where U, V, W, and Z denote disjoint subsets of variables. Ternary relations that obey the four properties listed above are often called *semigraphoids* [Pearl, Paz, 1987].

2.2 Compositional models

In [Jiroušek, 2011] we summarized results on probabilistic compositional models whose systems of independence relations we are going to study in this paper. To be able to introduce these models we have to recall the operator of composition and a couple of its most important properties that were proved in [Jiroušek, 2011].

Definition 2.2. *For two arbitrary distributions $\pi_1(K_1)$ and $\pi_2(K_2)$, for which*[1] *$\pi_1^{\downarrow K_1 \cap K_2} \ll \pi_2^{\downarrow K_1 \cap K_2}$, their* composition *is given by the following formula*[2]

$$(\pi_1 \triangleright \pi_2) = \frac{\pi_1^{\downarrow K_1} \pi_2^{\downarrow K_2}}{\pi_2^{\downarrow K_1 \cap K_2}}.$$

In case $\pi_1^{\downarrow K_1 \cap K_2} \not\ll \pi_2^{\downarrow K_1 \cap K_2}$ the composition remains undefined.

[1] $\pi_1^{\downarrow K_1 \cap K_2} \ll \pi_2^{\downarrow K_1 \cap K_2}$ denotes that the distribution $\pi_1^{\downarrow K_1 \cap K_2}$ is absolutely continuous with respect to distribution $\pi_2^{\downarrow K_1 \cap K_2}$, which, in our finite setting, means that whenever $\pi_1^{\downarrow K_1 \cap K_2}$ is positive then $\pi_2^{\downarrow K_1 \cap K_2}$ must also be positive.

[2] In this paper $\frac{0 \cdot 0}{0} = 0$.

Lemma 2.3. *Consider distributions $\pi_1(K_1)$ and $\pi_2(K_2)$. If $\pi_1 \triangleright \pi_2$ is defined, it is a distribution for variables $K_1 \cup K_2$ and*

$$(\pi_1 \triangleright \pi_2)^{\downarrow K_1} = \pi_1.$$

Moreover, for any U such that $K_1 \cap K_2 \subseteq U \subseteq K_1 \cup K_2$

$$(\pi_1 \triangleright \pi_2)^{\downarrow U} = \pi_1^{\downarrow K_1 \cap U} \triangleright \pi_2^{\downarrow K_2 \cap U}.$$

Lemma 2.4. *Let $\kappa(K_1 \cup K_2) = \pi_1(K_1) \triangleright \pi_2(K_2)$ be defined. Then*

$$(K_1 \setminus K_2) \perp\!\!\!\perp (K_2 \setminus K_1) | (K_1 \cap K_2)[\kappa].$$

From this point forward we will consider distributions $\pi_1(K_1), \dots, \pi_n(K_n)$. To avoid too many parentheses, whenever we speak about $\pi_1 \triangleright \pi_2 \triangleright \dots \triangleright \pi_n$, we assume that the operators are realized from left to right, i.e.,

$$\pi_1 \triangleright \pi_2 \triangleright \dots \triangleright \pi_n = (((\pi_1 \triangleright \pi_2) \triangleright \pi_3) \triangleright \dots \triangleright \pi_{n-1}) \triangleright \pi_n, \qquad (2.5)$$

and that this expression is defined. In this way, Formula (2.5) represents a multidimensional distribution of variables $K_1 \cup K_2 \cup \dots \cup K_n$, and we call the sequence $\pi_1(K_1), \pi_2(K_2), \dots, \pi_n(K_n)$ a *generating sequence* for distribution $\kappa = \pi_1 \triangleright \pi_2 \triangleright \dots \triangleright \pi_n$.

Based on the above convention we get the following assertion as a direct corollary of Lemma 2.3.

Corollary 2.5. *Consider a compositional model κ with a generating sequence $\pi_1(K_1)$, $\pi_2(K_2)$, $\dots, \pi_n(K_n)$ (i.e., $\kappa = \pi_1 \triangleright \pi_2 \triangleright \dots \triangleright \pi_n$). Then, for all $i = 1, \dots, n$,*

$$\kappa^{\downarrow K_1 \cup \dots \cup K_i} = \pi_1 \triangleright \dots \triangleright \pi_i.$$

3 Structures

Consider a compositional model defined by a generating sequence $\pi_1(K_1), \pi_2(K_2), \dots, \pi_n(K_n)$. Then the sequence of sets $\mathcal{P} = (K_1, K_2, \dots, K_n)$ is said to be its *structure*, and we use the symbol $K_i \in \mathcal{P}$ to express the fact that K_i is a member of this sequence. In what follows, the symbol $\widehat{K}(\mathcal{P})$ will always denote the set of all the variables appearing in structure $\mathcal{P}$, i.e., for the considered structure $\mathcal{P} = (K_1, K_2, \dots, K_n)$, $\widehat{K}(\mathcal{P}) = K_1 \cup K_2 \cup \dots \cup K_n$.

In the following text, we are going to study what happens when we change the ordering of the distributions in a generating sequence. Specifically, we want to characterize the changes that have no impact on the generated multidimensional distribution. It appears that it is sufficient to study elementary changes when a set from the structure is moved to a new position (and all the other sets in the structure are respectively shifted). Such a move when the k-th set is placed into the l-th position will be denoted by $(k \curvearrowright l)$, and if this move is applied to structure $\mathcal{P}$ the resulting structure will be denoted by $\mathcal{P}(k \curvearrowright l)$. I.e., for $\mathcal{P} = (U, W, Z, V)$ the reordering $\mathcal{P}(1 \curvearrowright 3) = (W, Z, U, V)$, and for $\mathcal{P}' = (K_1, K_2, K_3, K_4, K_5)$, its reordering $\mathcal{P}'(5 \curvearrowright 4) = (K_1, K_2, K_3, K_5, K_4)$, which can be further reordered as, e.g., $\mathcal{P}'(5 \curvearrowright 4)(2 \curvearrowright 5) = (K_1, K_3, K_5, K_4, K_2)$.

With respect to its position in a structure, each set $U \in \mathcal{P}$ can be split into two disjoint parts. We denote them $R(\mathcal{P}, U)$ and $S(\mathcal{P}, U)$, where $R(\mathcal{P}, U)$ denotes the subset of those variables from $U \in \mathcal{P}$, which are not in any set preceding U in the sequence $\mathcal{P}$. Conversely, $S(\mathcal{P}, U)$ denotes the subset of the remaining variables from U which appear in at least one of the sets preceding U in $\mathcal{P}$.

For $\mathcal{P} = (K_1, \ldots, K_n)$ it means that

$$R(\mathcal{P}, K_1) = K_1 \quad \textit{and} \quad R(\mathcal{P}, K_i) = K_i \setminus (K_1 \cup \ldots \cup K_{i-1}) \; \forall i = 2, \ldots, n,$$

and

$$S(\mathcal{P}, K_1) = \emptyset \quad \textit{and} \quad S(\mathcal{P}, K_i) = K_i \cap (K_1 \cup \ldots \cup K_{i-1}) \; \forall i = 2, \ldots, n.$$

We say that a set $K_i \in \mathcal{P}$ is *reducible* in $\mathcal{P}$ if $K_i = S(\mathcal{P}, K_i)$ and *irreducible* otherwise. In other words, the reducible set does not introduce any "new" variable to the sequence, i.e., $K_i \subseteq K_1 \cup \ldots \cup K_{i-1}$.

Consider the structure $\mathcal{P} = (K_1, \ldots, K_5) = (\{u\}, \{u, v, w\}, \{u, v, x\}, \{w, y\}, \{u, y, z\})$ and its reordering $\mathcal{P}' = \mathcal{P}(3 \curvearrowright 1)$. For the respective R and

	$R(\mathcal{P}, \cdot)$	$S(\mathcal{P}, \cdot)$	$R(\mathcal{P}', \cdot)$	$S(\mathcal{P}', \cdot)$
$K_1 = \{u\}$	$\{u\}$	$\emptyset$	$\emptyset$	$\{u\}$
$K_2 = \{u, v, w\}$	$\{v, w\}$	$\{u\}$	$\{w\}$	$\{u, v\}$
$K_3 = \{u, v, x\}$	$\{x\}$	$\{u, v\}$	$\{u, v, x\}$	$\emptyset$
$K_4 = \{w, y\}$	$\{y\}$	$\{w\}$	$\{y\}$	$\{w\}$
$K_5 = \{u, y, z\}$	$\{z\}$	$\{u, y\}$	$\{z\}$	$\{u, y\}$

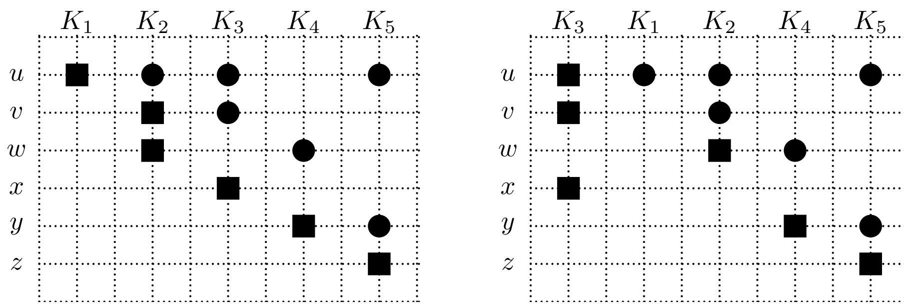
Tab. 1: R and S-parts of the structures $\mathcal{P}$ and $\mathcal{P}' = \mathcal{P}(3 \curvearrowright 1)$.

S-parts, see Table 1. Notice that K_1 is reducible in $\mathcal{P}'$ but not in $\mathcal{P}$.

3.1 Persegrams

To visualize the structure of a compositional model (and its generating sequence) we use a tool called a *persegram*. This visualization tool was originally designed in [Jiroušek, 2008] in a slightly different way.

Definition 3.1. *The* persegram *of a structure $\mathcal{P} = (K_1, \ldots, K_n)$ is a table in which rows correspond to variables from $\widehat{K}(\mathcal{P}) = K_1 \cup \ldots \cup K_n$ (in an arbitrary order) and columns to sets $K_1, \ldots, K_n$ in the respective ordering. A position in the table is marked if the respective set contains the corresponding variable. Markers for the first occurrence of each variable (i.e., the leftmost markers in rows) are* box-markers*, and for other occurrences there are* bullets.



(a) $\mathcal{P} = (K_1, K_2, K_3, K_4, K_5)$

(b) $\mathcal{P}(3 \curvearrowright 1) = (K_3, K_1, K_2, K_4, K_5)$

Fig. 1: Persegram of structure $(K_1, \ldots, K_5)$ and its reordering

In Figure 1 we can see persegrams of structures $\mathcal{P}$ and $\mathcal{P}' = \mathcal{P}(3 \curvearrowright 1)$ defined in Example 3.

Notice the difference between these persegrams. By reordering the columns – sets from the structure – several markers change their shapes. For example, marker $[K_1, u]$ is a box-marker in $\mathcal{P}$ but a bullet in $\mathcal{P}'$. Conversely, $[K_3, u]$ is a bullet in $\mathcal{P}$ but a box-marker in $\mathcal{P}'$.

Observe that there is one-to-one correspondence between bullets in the column of K_i and variables from $S(\cdot, K_i)$. Similarly, box-markers of K_i correspond to $R(\cdot, K_i)$.

3.2 Structural independence

Considering a compositional model, i.e., a multidimensional distribution generated by a generating sequence, one can see that it is possible, using Lemma 2.4, to deduce a number of conditional independence relations that must hold for this distribution. It is not surprising. The same property holds for a Bayesian network where one can read a system of necessary conditional independence relations from the corresponding acyclic directed graph. To determine all the independence relations induced by a structure of a generating sequence (we will call them *structural independencies*) we use the above-defined persegram. Structural independencies are indicated by the absence of a *trail connecting relevant markers and avoiding others* – see the following definition.

Definition 3.2. *A sequence of markers $m_0, \ldots, m_t$ in a persegram of a structure $\mathcal{P}$ is called a Z-*avoiding trail ($Z \subseteq \widehat{K}(\mathcal{P})$) that connects m_0 and m_t *if it meets the following five conditions:*

1. *neither m_0 nor m_t corresponds to a variable from Z*
2. *for each $s = 1, \ldots, t$, the couple (m_{s-1}, m_s) is either in the same row (i.e., a horizontal connection) or in the same column (a vertical connection);*

3. *each vertical connection must be adjacent to a box-marker (i.e., at least one of the markers in the vertical connection is a box-marker) - the so-called regular vertical connection;*

4. *no horizontal connection corresponds to a variable from Z;*

5. *vertical and horizontal connections regularly alternate with the following possible exception:*

   at most, two vertical connections may be in direct succession if their common adjacent marker is a box-marker of a variable from Z

If a Z-avoiding trail connects two markers corresponding to variables u and v, we say that these variables are connected by a Z-avoiding trail. *This situation is denoted by $u \leftrightsquigarrow_Z v[\mathcal{P}]$.*

Similarly to a probability distribution π which induces a ternary relation on disjoint sets of variables $U \perp\!\!\!\perp V|Z[\pi]$, a persegram, or more precisely a structure $\mathcal{P} = (K_1, \ldots, K_n)$, also introduces a ternary relation on triples of variable sets.

Definition 3.3. *Consider a structure $\mathcal{P} = (K_1, \ldots, K_n)$ and three disjoint sets $U, V, Z \subset \widehat{K}(\mathcal{P})$ such that $U, V \neq \emptyset$. We say that sets of variables U and V are* conditionally independent given Z in $\mathcal{P}$ *(in symbol $U \perp\!\!\!\perp V|Z[\mathcal{P}]$), if for each $u \in U$ it holds that there does not exist $v \in V$ such that $u \leftrightsquigarrow_Z v[\mathcal{P}]$.*

So, we have defined two types of (conditional) independence for groups of variables: the independence induced by probability distributions and that induced by structures. As already mentioned above, we will call the latter case *structural independence* to distinguish between these two types.

When the independent sets are singletons, we speak about *elementary relations* and denote them in a simplified form $u \perp\!\!\!\perp v|Z$ (instead of the more precise notation $\{u\} \perp\!\!\!\perp \{v\}|Z$). It is important to realize that, for any structure $\mathcal{P}$, all of its structural independencies are uniquely determined by the system of elementary relations $u \not\perp\!\!\!\perp v|Z[\mathcal{P}]$ in the following sense:

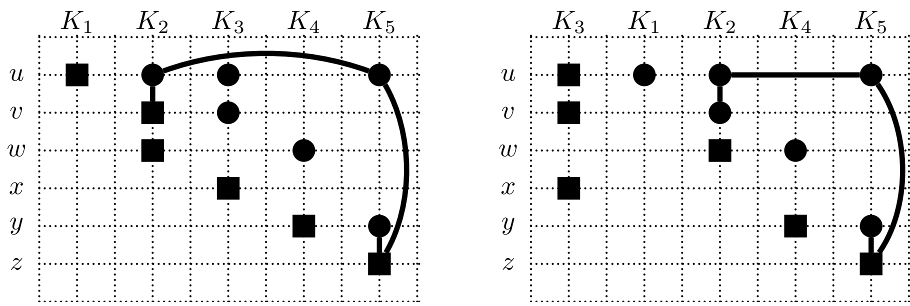
$$u \perp\!\!\!\perp v|Z[\mathcal{P}] \ \& \ u \perp\!\!\!\perp w|Z[\mathcal{P}] \implies u \perp\!\!\!\perp \{v, w\}|Z[\mathcal{P}], \tag{3.1}$$

because structural elementary relations are deduced from the nonexistence of a sequence of markers $u \leftrightsquigarrow_Z v[\mathcal{P}]$. This property, naturally, does not hold for probabilistic independence, and therefore it is quite natural that rule (3.1), which we will call an *extension* in the following text, cannot be deduced from the semigraphoid rules (2.1) - (2.4).

To illustrate the notion of a Z-avoiding trail, consider structure $\mathcal{P} = (K_1, K_2, K_3, K_4, K_5)$ and its reordering $\mathcal{P}(3 \curvearrowright 1)$ – for the respective persegrams see Figure 2. In both of these persegrams the same sequence of markers is traced out:

$$[K_2, v], [K_2, u], [K_5, u], [K_5, z], [K_5, y].$$

Notice that in Figure 2a, this sequence of markers is a Z-avoiding trail for each Z for which $\{z\} \subseteq Z \subseteq \{w, x, z\}$. For each such Z thus $v \leftrightsquigarrow_Z y[\mathcal{P}]$.



(a) $\mathcal{P}$: Z-avoiding trail,$\{z\} \subseteq Z \subseteq \{w, x, z\}$

(b) $\mathcal{P}'$: seq. of markers

Fig. 2: Sequence of markers in structure $(K_1, \ldots, K_5)$ and its reordering

If we want to stress which markers are box-markers and empasize a type of connection between the consecutive markers, we may also write this sequence in the following form:

$$[K_2, v]^{\blacksquare} \updownarrow [K_2, u]^{\bullet} \longleftrightarrow [K_5, u]^{\bullet} \updownarrow [K_5, z]^{\blacksquare} \updownarrow [K_5, y]^{\bullet};$$

$\longleftrightarrow$ denotes a horizontal connection and $\updownarrow$ vertical connection.

On the other hand, it is evident that the sequence of the same markers does not represent a Z-avoiding trail in the persegram of $\mathcal{P}(3 \curvearrowright 1)$ (see Figure 2b). This is because in this case the vertical connection $[K_2, v]^{\bullet} \updownarrow [K_2, u]^{\bullet}$ is not regular in $\mathcal{P}(3 \curvearrowright 1)$. Nevertheless, one can see that $v \leftrightsquigarrow_Z y[\mathcal{P}(3 \curvearrowright 1)]$, too, because for all Z for which $\{z\} \subseteq Z \subseteq \{w, x, z\}$,

$$[K_3, v]^{\blacksquare} \updownarrow [K_3, u]^{\blacksquare} \longleftrightarrow [K_5, u]^{\bullet} \updownarrow [K_5, z]^{\blacksquare} \updownarrow [K_5, y]^{\bullet}$$

is a Z-avoiding trail in the persegram in Figure 2b.

The following two theorems justify the concept of Z-avoiding trails as well as the related notion of structural independence, and the notation used. The first one says, among others, that the relation of structural independence meets all the semigraphoid axioms. Note that the *extension* property has already been mentioned above in formula (3.1).

Theorem 3.4. *For any structure $\mathcal{P} = (K_1, \ldots, K_n)$ the corresponding ternary* *relation of structural independence has the following properties:*

symmetry	$U \perp\!\!\!\perp V \mid Z[\mathcal{P}] \Leftrightarrow V \perp\!\!\!\perp U \mid Z[\mathcal{P}]$	(3.2)
decomposition	$U \perp\!\!\!\perp (V \cup W) \mid Z[\mathcal{P}] \Rightarrow U \perp\!\!\!\perp V \mid Z[\mathcal{P}]$	(3.3)
weak union	$U \perp\!\!\!\perp (V \cup W) \mid Z[\mathcal{P}] \Rightarrow U \perp\!\!\!\perp W \mid (V \cup Z)[\mathcal{P}]$	(3.4)
contraction	$U \perp\!\!\!\perp W \mid (V \cup Z)[\mathcal{P}]$ & $U \perp\!\!\!\perp V \mid Z[\mathcal{P}] \Rightarrow U \perp\!\!\!\perp (V \cup W) \mid Z[\mathcal{P}]$	(3.5)
extension	$U \perp\!\!\!\perp V \mid Z[\mathcal{P}]$ & $U \perp\!\!\!\perp W \mid Z[\mathcal{P}] \Rightarrow U \perp\!\!\!\perp (V \cup W) \mid Z[\mathcal{P}]$	(3.6)
intersection	$U \perp\!\!\!\perp W \mid (V \cup Z)[\mathcal{P}]$ & $U \perp\!\!\!\perp V \mid (W \cup Z)[\mathcal{P}]$ $\Rightarrow U \perp\!\!\!\perp (V \cup W) \mid Z[\mathcal{P}]$	(3.7)

where U, V, W, and Z are disjoint subsets of $\widehat{K}(\mathcal{P})$; U, V, W are nonempty.

Proof. ***Symmetry***, ***decomposition*** and ***extension*** are trivial consequences of the definition of structural independence. The remaining three properties will be proved by contradiction.

Weak union. Assume $U \not\perp\!\!\!\perp W \mid (V \cup Z)[\mathcal{P}]$. Due to the assumption $U \perp\!\!\!\perp (V \cup W) \mid Z[\mathcal{P}]$ and from the decomposition rule we know that $U \perp\!\!\!\perp W \mid Z[\mathcal{P}]$. It means that there must exist a $(V \cup Z)$-avoiding trail $u \leftrightsquigarrow_{V \cup Z} w$ for some $u \in U$ and $w \in W$ such that it contains two vertical connections in direct succession with a common adjacent box-marker for a variable from V. Considering the first appearance of a marker corresponding to variable from V, denote it v, we in fact constructed $u \leftrightsquigarrow_Z v$, which contradicts $U \perp\!\!\!\perp V \mid Z[\mathcal{P}]$ that is a consequence of the assumption.

Contraction. Assuming $U \perp\!\!\!\perp V \mid Z[\mathcal{P}]$ and $U \not\perp\!\!\!\perp (V \cup W) \mid Z[\mathcal{P}]$ means that there exists a Z-avoiding trail $u \leftrightsquigarrow_V w$ for some $u \in U$ and $w \in W$. However, similar to the previous step, since we assume that $U \perp\!\!\!\perp W \mid (V \cup Z)[\mathcal{P}]$, the considered trail must contain two vertical connections in direct succession with a common adjacent box-marker for a variable from V. This, again, suggests the existence of a Z-avoiding trail connecting variable u with a variable from V, which contradicts to $U \perp\!\!\!\perp V \mid Z[\mathcal{P}]$.

Intersection. Let us assume the opposite. If $U \not\perp\!\!\!\perp (V \cup W) \mid Z[\mathcal{P}]$, then for a certain $u \in U$ there must exist $v \in (V \cup W)$ such that $u \leftrightsquigarrow_Z v$. Consider the shortest $\tau = u \leftrightsquigarrow_Z v$ connecting u with a node from $(V \cup W)$. Without loss of generality, we can assume that $v \in V$. Assumption of $U \perp\!\!\!\perp V \mid (W \cup Z)[\mathcal{P}]$ induces that τ has a horizontal connection in a $w \in W$. Therefore by cutting the rest of τ off, one can create $\tau' = u \leftrightsquigarrow_Z w$ that is shorter than τ and connects u with a node from $(V \cup W)$, which is impossible, because we chose τ to be the shortest connecting trail. □

The next important theorem was originally proven in [Jiroušek, 2008]. Here we present a new and more elegant (and hopefully also more transparent) proof based on the ideas of [Verma, Pearl, 1990]. This theorem reveals the relation between both types of conditional independence in sets of variables: structural independence and probabilistic independence.

Theorem 3.5. *Consider a generating sequence $\pi_1, \ldots, \pi_n$ with structure $\mathcal{P} = (K_1, \ldots, K_n)$. Then for arbitrary three disjoint subsets $U, V, Z \subset \widehat{K}(\mathcal{P})$ such that $U \neq \emptyset$ and $V \neq \emptyset$ holds that*

$$U \perp\!\!\!\perp V|Z[\mathcal{P}] \Longrightarrow U \perp\!\!\!\perp V|Z[\pi_1 \rhd \ldots \rhd \pi_n]. \tag{3.8}$$

Proof. The proof will be led by an induction on the number of sets in the structure. The assertion is obvious for $|\mathcal{P}| = 1$ (there is no structural independence). Suppose it holds for all structures of a length less than n and let us prove it for $\mathcal{P} = (K_1, \ldots, K_n)$.

To simplify the readings, put $R = R(\mathcal{P}, K_n)$ and $S = S(\mathcal{P}, K_n)$. Note that each set of U, V, Z can be expressed as a union of two disjoint parts $\bar{U}, \bar{V}, \bar{Z} \subseteq (K_1 \cup \ldots \cup K_{n-1})$ and $R_U, R_V, R_Z \subseteq R$, respectively, i.e., $\bar{U}, \bar{V}, \bar{Z}, R$ are disjoint. Then $U \perp\!\!\!\perp V|Z[\mathcal{P}]$ can be equivalently written as

$$(\bar{U} \cup R_U) \perp\!\!\!\perp (\bar{V} \cup R_V)|(\bar{Z} \cup R_Z)[\mathcal{P}]. \tag{3.9}$$

The proof will be performed for several special cases characterized by which sets from R_U, R_V, R_Z are empty. Theoretically, we can distinguish eight situations. However, using the symmetry (3.2) and the fact that R_U and R_V meeting (3.9) cannot be nonempty simultaneously (all couples of variables from R can be connected by a regular vertical connection) it is enough to investigate the following four cases:

(i) $R_U, R_V, R_Z = \emptyset$,

(ii) $R_U \neq \emptyset$; $R_V, R_Z = \emptyset$,

(iii) $R_Z \neq \emptyset$; $R_U, R_V = \emptyset$,

(iv) $R_U, R_Z \neq \emptyset$; $R_V = \emptyset$.

Ad (i) So, we assume $\bar{U} \perp\!\!\!\perp \bar{V}|\bar{Z}[\mathcal{P}]$ where $\bar{U}, \bar{V}, \bar{Z}$ are three disjoint subsets of $\widehat{K}(\mathcal{P})$ disjoint with R (i.e. $\bar{U}, \bar{V}, \bar{Z} \subset (K_1 \cup \ldots \cup K_{n-1})$). Since each $\bar{Z}$-avoiding trail in a persegram of $(K_1, \ldots, K_{n-1})$ is also a $\bar{Z}$-avoiding trail in a persegram of $\mathcal{P} = (K_1, \ldots, K_{n-1}, K_n)$, it is obvious that

$$\bar{U} \perp\!\!\!\perp \bar{V}|\bar{Z}[\mathcal{P}] \implies \bar{U} \perp\!\!\!\perp \bar{V}|\bar{Z}[(K_1, \ldots, K_{n-1})].$$

Therefore, using the induction hypothesis we get $\bar{U} \perp\!\!\!\perp \bar{V}|\bar{Z}[\pi_1 \rhd \ldots \rhd \pi_{n-1}]$, which implies $\bar{U} \perp\!\!\!\perp \bar{V}|\bar{Z}[\pi_1 \rhd \ldots \rhd \pi_n]$ since $\pi_1 \rhd \ldots \rhd \pi_{n-1}$ is, due to Corollary 2.5, a marginal of $\pi_1 \rhd \ldots \rhd \pi_n$.

Ad (ii) In this case we assume that $(\bar{U} \cup R_U) \perp\!\!\!\perp \bar{V}|\bar{Z}[\mathcal{P}]$, and $\bar{U}, \bar{V}, \bar{Z}, R$ are disjoint. Based on this, we will also show that

$$(\bar{U} \cup R_U \cup (S \setminus \bar{Z})) \perp\!\!\!\perp \bar{V}|\bar{Z}[\mathcal{P}]. \tag{3.10}$$

Namely, the negation of (3.10) corresponds to the existence of a trail $v \leftrightsquigarrow_{\bar{Z}} s$ for some $v \in \bar{V}$ and $s \in (S \setminus \bar{Z})$. Considering the shortest such trail (this means, among other things, that the connection to the marker corresponding to variable s is vertical) we can see that it can be extended by two markers (connections)

$$\longleftrightarrow [K_n, s]^{\bullet} \updownarrow [K_n, r]^{\blacksquare}$$

for any $r \in R_U$, which contradicts the assumption $(\bar{U} \cup R_U) \perp\!\!\!\perp \bar{V} | \bar{Z}[\mathcal{P}]$. This means that relation (3.10) holds.

Applying decomposition property (3.3) to relation (3.10) we get

$$(\bar{U} \cup (S \setminus \bar{Z})) \perp\!\!\!\perp \bar{V} | \bar{Z}[\mathcal{P}],$$

which satisfies the conditions of the previously proven case (i), namely, $(\bar{U} \cup (S \setminus \bar{Z}))$, $\bar{V}$, $\bar{Z}$ and R are disjoint, and therefore

$$(\bar{U} \cup (S \setminus \bar{Z})) \perp\!\!\!\perp \bar{V} | \bar{Z}[\pi_1 \triangleright \ldots \triangleright \pi_n]. \tag{3.11}$$

It is easy to also see that $(\bar{V} \cap S) = \emptyset$. Indeed, if not then $v \leftrightsquigarrow_Z r$ (consisting of just one regular connection) for any $v \in (\bar{V} \cap S)$ and any $r \in R_U$ contradicting $(\bar{U} \cup R_U) \perp\!\!\!\perp \bar{V} | \bar{Z}[\mathcal{P}]$. Note that $R \perp\!\!\!\perp (\widehat{K}(\mathcal{P}) \setminus K_n) | S[\pi_1 \triangleright \ldots \triangleright \pi_n]$ by Lemma 2.4. By the (multiple) application of decomposition (2.2) it follows that

$$R_U \perp\!\!\!\perp (\bar{V} \cup (\bar{U} \setminus S) \cup (\bar{Z} \setminus S)) | S[\pi_1 \triangleright \ldots \triangleright \pi_n],$$

which can be, using the weak union property (2.3), further rewritten into

$$R_U \perp\!\!\!\perp \bar{V} | (\bar{Z} \cup \bar{U} \cup (S \setminus \bar{Z}))[\pi_1 \triangleright \ldots \triangleright \pi_n]. \tag{3.12}$$

Now, applying the contraction property (2.4), statements (3.11) and (3.12) yield

$$(\bar{U} \cup R_U \cup (S \setminus \bar{Z})) \perp\!\!\!\perp \bar{V} | \bar{Z}[\pi_1 \triangleright \ldots \triangleright \pi_n], \tag{3.13}$$

from which the desired conditional independence $(\bar{U} \cup R_U) \perp\!\!\!\perp \bar{V} | \bar{Z}[\pi_1 \triangleright \ldots \triangleright \pi_n]$ is obtained by using the decomposition property (2.2).

Ad (iii) Assume the independence statement in the form of

$$\bar{U} \perp\!\!\!\perp \bar{V} | (\bar{Z} \cup R_Z)[\mathcal{P}], \tag{3.14}$$

where, again, $\bar{U}, \bar{V}, \bar{Z}, R$ are disjoint.

Now, let us show by contradiction that either $\bar{U} \perp\!\!\!\perp R_Z | \bar{Z}[\mathcal{P}]$ or $R_Z \perp\!\!\!\perp \bar{V} | \bar{Z}[\mathcal{P}]$. Assuming that neither of these two independence relations hold, there must exist trails $u \leftrightsquigarrow_{\bar{Z}} r_1[\mathcal{P}]$ and $r_2 \leftrightsquigarrow_{\bar{Z}} v[\mathcal{P}]$ (consider the shortest possible) for some $u \in \bar{U}, v \in \bar{V}$, and $r_1, r_2 \in R_Z$. Each variable from $r \in R$ has only one marker in the respective persegram, and therefore both trails contain only one marker from R – the one at the end. It means that changing just the last marker in trail $u \leftrightsquigarrow_{\bar{Z}} r_1[\mathcal{P}]$ one gets $u \leftrightsquigarrow_{\bar{Z}} r_2[\mathcal{P}]$, and by concatenating trails $u \leftrightsquigarrow_{\bar{Z}} r_2[\mathcal{P}]$ and $r_2 \leftrightsquigarrow_{\bar{Z}} v[\mathcal{P}]$ one gets $u \leftrightsquigarrow_{\bar{Z} \cup R_Z} [\mathcal{P}]v$. Since the last trail contradicts our assumption, we proved that really either $\bar{U} \perp\!\!\!\perp R_Z | \bar{Z}[\mathcal{P}]$ or $R_Z \perp\!\!\!\perp \bar{V} | \bar{Z}[\mathcal{P}]$.

Without a loss of generality, assume that $R_Z \perp\!\!\!\perp \bar{V} | \bar{Z}[\mathcal{P}]$. This independence statement along with the relation (3.14) meets the assumption of the contraction property (3.5), which yields that $(\bar{U} \cup R_Z) \perp\!\!\!\perp \bar{V} | \bar{Z}[\mathcal{P}]$. This structural independence statement meets the assumptions already solved in the proof in case (ii). It means that the corresponding probabilistic conditional independence

$$(\bar{U} \cup R_Z) \perp\!\!\!\perp \bar{V} | \bar{Z}[\pi_1 \triangleright \ldots \triangleright \pi_n]$$

must also hold true, from which the required probabilistic conditional independence $\bar{U} \perp\!\!\!\perp \bar{V} | (R_Z \cup \bar{Z})[\pi_1 \triangleright \ldots \triangleright \pi_n]$ can be obtained by the simple application of the weak union rule (2.3).

Ad (iv): Now, we assume $(\bar{U} \cup R_U) \perp\!\!\!\perp \bar{V} | (\bar{Z} \cup R_Z)[\mathcal{P}]$ with disjoint $\bar{U}, \bar{V}, \bar{Z}, R$.

First, let us show by contradiction that $R_U \perp\!\!\!\perp \bar{V} | \bar{Z}[\mathcal{P}]$. Assuming the opposite there must exist a trail $r_u \leftrightsquigarrow_{\bar{Z}} v[\mathcal{P}]$ for some $r_u \in R_U$ and $v \in \bar{V}$. In this trail, there is no horizontal connection corresponding to $r \in R_Z$ since there is only one marker for every $r \in R$ in $\mathcal{P}$. Therefore, this trail is also $\bar{Z} \cup R_Z$-avoiding trail. However, the existence of such a trail contradicts our assumption, which completes this step of the proof.

Following the idea from the previous part of the proof, we know that any trail $r \leftrightsquigarrow_{\bar{Z}} v[\mathcal{P}]$ for some $r \in R, v \in \bar{V}$ can be modified into $r_u \leftrightsquigarrow_{\bar{Z}} v[\mathcal{P}]$ (for $r_u \in R_U$) just by substituting the marker corresponding to r by the marker corresponding to r_u. Therefore we can see that the relation $R_U \perp\!\!\!\perp \bar{V} | \bar{Z}[\mathcal{P}]$ proven in the preceding paragraph implies $(R_Z \cup R_U) \perp\!\!\!\perp \bar{V} | \bar{Z}[\mathcal{P}]$. The last structural independence statement can be treated in the same way as case (ii), which means that the probabilistic independence statement

$$(R_Z \cup R_U) \perp\!\!\!\perp \bar{V} | \bar{Z}[\pi_1 \triangleright \ldots \triangleright \pi_n] \tag{3.15}$$

also holds true.

On the other hand, $\bar{U} \perp\!\!\!\perp \bar{V} | (\bar{Z} \cup R_Z \cup R_U)[\mathcal{P}]$ can be obtained from the given independence statement using the weak union property (3.4) and it can be treated in the same way as in case (iii). Hence

$$\bar{U} \perp\!\!\!\perp \bar{V} | (\bar{Z} \cup R_Z \cup R_U)[\pi_1 \triangleright \ldots \triangleright \pi_n]. \tag{3.16}$$

Applying the contraction property (2.4) on (3.15) and (3.16), one gets

$$(\bar{U} \cup R_Z \cup R_U) \perp\!\!\!\perp \bar{V} | \bar{Z}[\pi_1 \triangleright \ldots \triangleright \pi_n],$$

from which the desired statement can be obtained by application of the weak union property (2.3). □

3.3 Structural equivalence

An inherited part of the notion of structural independence is implied by its structural properties that uniquely determine the induced structural independence relations. To illustrate such properties, let us summarize the most important results from [Kratochvíl, 2011, 2013]. Two structures $\mathcal{P}$ and $\mathcal{P}'$ will be said to be *equivalent* if they induce the same structural independence relations, i.e., if

$$U \perp\!\!\!\perp V | Z[\mathcal{P}] \iff U \perp\!\!\!\perp V | Z[\mathcal{P}']$$

for all disjoint sets of variables.

Naturally, a trivial necessary condition for two structures $\mathcal{P}$ and $\mathcal{P}'$ to be equivalent is that they are defined over the same set of variables: $\widehat{K}(\mathcal{P}) = \widehat{K}(\mathcal{P}')$.

3.3.1 Non-trivial sets

A necessary and sufficient condition for structures to be equivalent is closely connected with the notion of a *non-trivial set.*

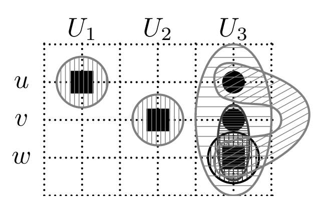
Definition 3.6. *We say that U is* non-trivial with respect to $\mathcal{P}$ *if there exists $K_i \in \mathcal{P}$ such that $U \subseteq K_i$ and $U \cap R(\mathcal{P}, K_i) \neq \emptyset$. The collection of all sets U that are non-trivial with respect to a structure $\mathcal{P}$ is denoted by $\mathcal{N}(\mathcal{P})$.*

The notion of non-trivial sets was introduced in [Kratochvíl, 2011] where it was identified as a property invariable within a class of equivalent structures. Later, in [Kratochvíl, 2013], it was proven that correspondence of these sets is not only necessary but also sufficient to guarantee the equivalence of two given structures.

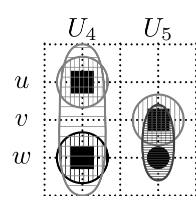
Theorem 3.7. *Structures $\mathcal{P}$ and $\mathcal{P}'$ are equivalent* iff $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$.

Consider two structures $\mathcal{P}_1$ and $\mathcal{P}_2$ whose persegrams are shown in Figure 3, where all the respective non-trivial sets are highlighted. Considering Theorem 3.7, the respective structures are not equivalent because

$$\mathcal{N}(\mathcal{P}_1) = \{\{u\}, \{v\}, \{w\}, \{u, w\}, \{v, w\}, \{u, v, w\}\} \neq \mathcal{N}(\mathcal{P}_2) = \{\{u\}, \{v\}, \{w\}, \{u, w\}, \{v, w\}\} .$$



(a) $\mathcal{P}_1$



(b) $\mathcal{P}_2$

Fig. 3: Non-trivial sets in different structures

Remark 3.8. Note that there is a close connection between non-trivial sets of cardinality 2 and regular vertical connections; similarly, there is a close connection between non-trivial sets of cardinality 3 and alternating of vertical and horizontal connections.

A special combination of both these cardinalities deserves our attention. Consider a triplet $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$ such that $\{u, v\} \notin \mathcal{N}(\mathcal{P})$. Let K_k be the set guaranteeing the non-triviality of the triplet in $\mathcal{P}$ (i.e., $\{u, v, w\} \subseteq K_k$; $\{u, v, w\} \cap R(\mathcal{P}, K_k) \neq \emptyset$). Obviously, $\{u, v\} \subseteq S(\mathcal{P}, K_k)$ because we assume it is trivial.

Moreover, an analogous reasoning yields that u and v have to be introduced in different sets preceding K_k in $\mathcal{P}$, i.e., $u \in R(\mathcal{P}, K_i)$, $v \in R(\mathcal{P}, K_j)$, $i \neq j$, and both indices $i, j < k$. So, one can easily see that a specific combination of non-trivial and trivial sets may put restrictions on the ordering of sets in the respective structure, and, by Theorem 3.7, in every structure equivalent with it. This property will be frequently used in Section 7, where we will use a symbol $\mathcal{N}_{3-2}(\mathcal{P})$ to denote a system of non-trivial triplets with a trivial subset of cardinality 2 - i.e. $\mathcal{N}_{3-2}(\mathcal{P}) = \{\{u, v, w\} \in \mathcal{N}(\mathcal{P}) : \{u, v\} \notin \mathcal{N}(\mathcal{P})\}$.

Remark 3.9. The reader familiar with the *imsets* of Milan Studený [Hemmecke, Lindner, Studený, 2012] can see a close connection between this famous apparatus and the above-introduced concept of non-trivial sets. Recall that a characteristic imset is a unique representative of an independence structure induced (represented) by an acyclic directed graph of a Bayesian Network (BN). In the case of a graph $G(N, E)$, it is a $\{0, 1\}$-vector indexed by subsets of N. It is easy to show that every probability distribution that can be represented by a compositional model with a structure $\mathcal{P}$ can be equivalently represented using a BN with a graph G. Moreover, $\mathcal{P}$ and G are equivalent in the sense that they encode the same system of structural independencies. It turns out that 1-components of characteristic imset of G corresponds to $\mathcal{N}(\mathcal{P})$.

In fact, it was shown in [Kratochvíl, 2013] that non-trivial sets of cardinality 2 and 3 are sufficient to guarantee the equivalence. The algorithm generating the complete $\mathcal{N}(\mathcal{P})$ from the respective non-trivial sets of cardinality 2 and 3 was published in [Studený, Hemmecke, and Lindner, 2012] in the case of characteristic imsets. It is based on the fact that a set of cardinality $c \geq 4$ is non-trivial if there are at least three different non-trivial subsets of cardinality $c - 1$.

3.3.2 Formal ratio

Since the number of non-trivial sets grows exponentially with the number of variables, they are not very useful for characterization of the structural properties of compositional models. This is why another closely related tool has been derived.

It appears that an efficient test of equivalence of structures can be based on a concept of a *formal ratio* that was introduced in [Kratochvíl, 2013]. Informally stated, one can write a formal ratio $\mathcal{F}(\mathcal{P})$ for a structure $\mathcal{P}$ as a ratio in which the numerator contains all of the sets K_i for $K_i \in \mathcal{P}$, and the denominator contains all of the sets $S(K_i, \mathcal{P})$ for $K_i \in \mathcal{P}$. If there are sets contained in both the numerator and denominator then these sets are "canceled" with each other: one occurrence of a set $U \in \mathcal{P}$ in the numerator is canceled with one occurrence of the same set in the denominator.

Using a formal notation of multisets, $[S(K_i, \mathcal{P})]_{K_i \in \mathcal{P}}$, which are sets in the sense that the ordering of the included sets is irrelevant but in which one element may appear several times, we may express this idea precisely by the following definition:

Definition 3.10. *A* formal ratio $\mathcal{F}(\mathcal{P})$ *corresponding to a structure* $\mathcal{P}$ *is*

$$\frac{[K_i]_{K_i\in\mathcal{P}} \setminus [S(\mathcal{P},K_i)]_{K_i\in\mathcal{P}}}{[S(\mathcal{P},K_i)]_{K_i\in\mathcal{P}} \setminus [K_i]_{K_i\in\mathcal{P}}}.$$

Consider structure $\mathcal{P} = (K_1, K_2, K_3, K_4) = (\{u\}, \{v, w\}, \{u, v, x\}, \{w, x, y\})$ and its reordering $\mathcal{P}' = (K_4, K_3, K_2, K_1) = (\{w, x, y\}, \{u, v, x\}, \{v, w\}, \{u\})$. For these structures

$$\begin{aligned} [S(\mathcal{P},K_i)]_{K_i\in\mathcal{P}} &= \emptyset, \emptyset, \{u,v\}, \{w,x\}, \\ [S(\mathcal{P}',K_i)]_{K_i\in\mathcal{P}} &= \emptyset, \{x\}, \{v,w\}, \{u\}, \end{aligned}$$

and therefore the respective formal ratios are the following

$$\mathcal{F}(\mathcal{P}) = \frac{\{u\}, \{v,w\}, \{u,v,x\}, \{w,x,y\}}{\emptyset, \emptyset, \{u,v\}, \{w,x\}},$$

$$\mathcal{F}(\mathcal{P}') = \frac{\{u,v,x\}, \{w,x,y\}}{\emptyset, \{x\}}.$$

Let us stress once more that the ordering of sets in both numerator and denominator is irrelevant.

The importance of the formal ratio follows from the following assertion proven in [Kratochvíl, 2013]:

Theorem 3.11. *Structures* $\mathcal{P}$ *and* $\mathcal{P}'$ *are equivalent* iff *their formal ratios coincide.*

The proof in [Kratochvíl, 2013] is based on the following idea. Assume a zero-one vector $u_{\mathcal{P}}$ whose coordinates correspond to all subsets of $\widehat{K}(\mathcal{P})$ such that $u_{\mathcal{P}}[U] = 1$ if $U \in \mathcal{N}(\mathcal{P})$ and $u_{\mathcal{P}}[U] = 0$ otherwise. Similarly, let an integer vector $c_{\mathcal{P}}$ (of the same length) be such that $c_{\mathcal{P}}(U) = 1$ if U is in the numerator of $\mathcal{F}(\mathcal{P})$, $c_{\mathcal{P}}[U] = -k$ if U is in the denominator of $\mathcal{F}(\mathcal{P})$ k-times, and $c_{\mathcal{P}}[U] = 0$ otherwise. Obviously, $u_{\mathcal{P}}$ uniquely characterizes $\mathcal{N}(\mathcal{P})$. Similarly, $c_{\mathcal{P}}$ uniquely characterizes $\mathcal{F}(\mathcal{P})$. It can be shown that $u_{\mathcal{P}}$ is a Möbius transform of $c_{\mathcal{P}}$ and vice-versa; the proof is completed by employing Theorem 3.7. Note that there is again a close connection between vectors $u_{\mathcal{P}}$ and $c_{\mathcal{P}}$ and Studený's imsets mentioned in Remark 3.9

Observe that structures $\mathcal{P}$ and $\mathcal{P}'$ from Example 3 are equivalent. They both induce the following formal ratio:

$$\mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}') = \frac{\{u,v,w\}, \{u,v,x\}, \{w,y\}, \{u,y,z\}}{\emptyset, \{w\}, \{u,v\}, \{u,y\}}.$$

To conclude this Section, let us summarize its results in an easy form. The following three statements are equivalent:

- $\mathcal{P}$ and $\mathcal{P}'$ are equivalent;
- $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$;
- $\mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}')$.

4 Operations on structures

Assume that $\mathcal{P}$ and $\mathcal{P}'$ are equivalent. The question to be answered in this Section is how to get from $\mathcal{P}'$ to $\mathcal{P}$ in terms of some elementary operations on the structures. Another very important aspect is our ability to generate all structures equivalent to a given one. In other words, we are looking for operations that, applied to a structure, yield other structures within the same class of equivalence. Regarding Theorem 3.11, we can see that we are looking for operations that do not change the formal ratio of a structure, and the respective definition gives us a clear hint about the form of possible operations.

These operations can be divided into two groups:

1. *Adding/removing sets* – we can add/remove a set whose impact disappears during formal ratio cancelation (i.e., adding or deleting a reducible set).

2. *Reordering* keeping the system of S-parts, we can apply changes that do not modify the denominator of the formal ratio. Hence, for both $\mathcal{P}$ and its reordering $\mathcal{P}'$ $[S(\mathcal{P}, K_i)]_{K_i \in \mathcal{P}} = [S(\mathcal{P}', K_j)]_{K_j \in \mathcal{P}'}$.

Before we investigate all these elementary operations in detail, note that all of the mentioned operations were introduced in [Kratochvíl, 2013] together with the proof of their completeness.

4.1 Adding/removing sets

It has been shown that we can restrict ourselves to adding/removing of only reducible sets. Recall that K_i is reducible in $\mathcal{P}$ if $K_i = S(K_i, \mathcal{P})$.

Definition 4.1. *By* simple extension/reduction *of $\mathcal{P}$ we understand a structure that differs from $\mathcal{P}$ in adding/removing of one set reducible in $\mathcal{P}$.*

Recall the structure $\mathcal{P}' = (K_3, K_1, K_2, K_4, K_5)$ introduced in Example 3. One can easily see in Figure 1b that K_1 is reducible in $\mathcal{P}'$, and hence (K_2, K_4, K_3, K_5) is a simple reduction of $\mathcal{P}'$.

Theorem 4.2. *A structure $\mathcal{P}$ and its simple extension/reduction are equivalent.*

Remark 4.3. Theorem 4.2 was proven in [Kratochvíl, 2013] by showing that this type of transformation does not influence the respective formal ratio. However, the reader may find it interesting to see how to prove the above assertion using the notions of Z-avoiding trails. Consider the persegram of $\mathcal{P}$ and the corresponding system of all the induced Z-avoiding trails. There is no box-marker in the columns corresponding to reducible sets. Therefore, there can be no regular vertical connection in such columns. That is why adding/removing of a reducible set cannot affect the system of Z-avoiding trails, which means that the respective structures are equivalent.

4.2 Reordering

Let $\mathcal{P}'$ be a reordering of $\mathcal{P}$ (i.e., $K_i \in \mathcal{P}' \Leftrightarrow K_i \in \mathcal{P}$). As mentioned above, $\mathcal{P}'$ and $\mathcal{P}$ are equivalent *iff* $(\mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}'))$, and therefore also *iff* $[S(\mathcal{P}', K_i)]_{K_i \in \mathcal{P}'} = [S(\mathcal{P}, K_j)]_{K_j \in \mathcal{P}}$.

In the following, we restrict ourselves to simple reorderings – transpositions of two successive sets K_k, K_{k+1}. In the introduced notation, such a reordered structure $\mathcal{P}$ can be denoted by $\mathcal{P}(k \curvearrowright k{-}1)$ or, equivalently, $\mathcal{P}(k{-}1 \curvearrowright k)$. The reason for this restriction is clear: recall that both R- and S-parts of a set $K_i \in \mathcal{P}$ are fully given by the set itself and the union of the sets preceding K_i. Hence, by swapping positions of two consecutive sets, the R- and S-parts remain the same for all of the other sets not being swapped. In a more formal way, in the case of $\mathcal{P} = (K_1, \dots, K_n)$ and $\mathcal{P}' = \mathcal{P}(k{-}1 \curvearrowright k)$, it holds that $S(\mathcal{P}', K_i) = S(\mathcal{P}, K_i)$ and $R(\mathcal{P}', K_i) = R(\mathcal{P}, K_i)$ for all $i \in \{1, \dots, n\}$ such that $i \neq k, i \neq k-1$. Thus, to check whether $\mathcal{F}(\mathcal{P}(k{-}1 \curvearrowright k)) = \mathcal{F}(\mathcal{P})$, it is enough to check only the swapped sets.

4.2.1 Constant transposition

Constant transposition is designed to preserve R- and S-parts of involved sets.

Definition 4.4. *Consider a structure $\mathcal{P} = (K_1, \dots, K_n)$. Its reordering $\mathcal{P}(k{-}1 \curvearrowright k)$ is called its* constant transposition *if $R(\mathcal{P}, K_{k-1}) \cap K_k = \emptyset$.*

Considering structure $\mathcal{P}$ from Example 3 (see Figure 1a), one can easily see that $\mathcal{P}(4 \curvearrowright 3)$ is a constant transposition of $\mathcal{P}$. Indeed, $R(\mathcal{P}, K_3) \cap K_4 = \{x\} \cap \{w, y\} = \emptyset$. Similarly, $\mathcal{P}(4 \curvearrowright 3)(5 \curvearrowright 4) = (K_1, K_2, K_4, K_5, K_3)$ is a constant transposition of $\mathcal{P}(4 \curvearrowright 3)$.

To show that a structure and its constant transposition are equivalent, it is enough to show that $[S(\mathcal{P}', K_i)]_{K_i \in \mathcal{P}'} = [S(\mathcal{P}, K_j)]_{K_j \in \mathcal{P}}$. In fact, it has thus been shown that even a stronger property holds true:

Theorem 4.5. *If $\mathcal{P}'$ is a constant transposition of $\mathcal{P}$, then $S(\mathcal{P}', K_i) = S(\mathcal{P}, K_i)$ for all $K_i \in \mathcal{P}$.*

The proof has been published in [Kratochvíl, 2011].

Remark 4.6. We can generalize constant transposition for long distance moves as well. Here, we show just a very special case needed in Section 7.

Assume $\mathcal{P} = (K_1, \dots, K_n), 2 \leq k \leq n$ such that $S(\mathcal{P}, K_k) \subseteq K_1$. We can see that $\mathcal{P}(k \curvearrowright 2) = \mathcal{P}(k \curvearrowright k{-}1)(k{-}1 \curvearrowright k{-}2) \dots (3 \curvearrowright 2)$, and that each of these subsequent transpositions corresponds to a constant transposition: $R(\mathcal{P}, K_i) \cap K_k = R(\mathcal{P}, K_i) \cap S(\mathcal{P}, K_k) \subseteq R(\mathcal{P}, K_i) \cap K_1 = \emptyset$ by the definition of the $R(\mathcal{P}, K_i)$ for all $i = 2, \dots, k-1$. Therefore, $\mathcal{P}(k \curvearrowright k{-}1)$ is a constant transposition of $\mathcal{P}$ and, by iterative application of Theorem 4.5, $\mathcal{P}(k \curvearrowright k{-}1)(k{-}1 \curvearrowright k{-}2)$ is a constant transposition of $\mathcal{P}(k \curvearrowright k{-}1)$, etc.

4.2.2 Box transposition

A constant transposition preserves S-parts of the involved sets. On the contrary, *box transposition* was designed to interchange these S-parts.

Definition 4.7. *Consider a structure $\mathcal{P} = (K_1, \ldots, K_n)$. Its reordering $\mathcal{P}(k-1 \curvearrowright k)$ is called its* box transposition *if $S(\mathcal{P}, K_{k-1}) = S(\mathcal{P}, K_k) \setminus R(\mathcal{P}, K_{k-1})$.*

To prove that a structure $\mathcal{P}$ and its box transposition $\mathcal{P}'$ are equivalent ($\mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}')$), it is enough to prove that $[S(\mathcal{P}', K_i)]_{K_i \in \mathcal{P}'} = [S(\mathcal{P}, K_j)]_{K_j \in \mathcal{P}}$.

Theorem 4.8. *If $\mathcal{P}' = \mathcal{P}(k-1 \curvearrowright k)$ is a box transposition of $\mathcal{P}$, then $S(\mathcal{P}', K_k) = S(\mathcal{P}, K_{k-1})$ and $S(\mathcal{P}', K_{k-1}) = S(\mathcal{P}, K_k)$.*

The proof has been published in [Kratochvíl, 2011].

Remark 4.9. Let $\mathcal{P} = (K_1, \ldots, K_n), n \geq 2$. Note that $\mathcal{P}(2 \curvearrowright 1)$ is always either a box or a constant transposition of $\mathcal{P}$. It is a constant transposition if $K_1 \cap K_2 = \emptyset$ and a box transposition otherwise.

Assume a structure $\mathcal{P}$ from Example 3 once again (see Figure 1a). Note that $\mathcal{P}(3 \curvearrowright 2)$ is a box transposition of $\mathcal{P}$. Indeed, $S(\mathcal{P}, K_3) \setminus R(\mathcal{P}, K_2) = \{u, v\} \setminus \{v, w\} = \{u\} = S(\mathcal{P}, K_2)$. Considering Remark 4.9, transposition $(2 \curvearrowright 1)$ always corresponds to either a box or a constant transposition. Specifically, if applied to $\mathcal{P}(3 \curvearrowright 2)$, i.e., $\mathcal{P}(3 \curvearrowright 2)(2 \curvearrowright 1)$ is a box transposition of $\mathcal{P}(3 \curvearrowright 2)$. Hence $\mathcal{P}'$ from Example 3 (see Figure 1b), where $\mathcal{P}' = (K_3, K_1, K_2, K_4, K_5) = \mathcal{P}(3 \curvearrowright 2)(2 \curvearrowright 1)$ is obtained from $\mathcal{P}$ by two constant transpositions.

The completeness of the above-mentioned operations was shown in [Kratochvíl, 2013]:

Theorem 4.10. *Two structures $\mathcal{P}_A$ and $\mathcal{P}_B$ are equivalent* iff *there exists a sequence $\mathcal{P}_1, ..., \mathcal{P}_m$, $m \geq 1$ of structures such that $\mathcal{P}_1 = \mathcal{P}_A, \mathcal{P}_m = \mathcal{P}_B$ and $\mathcal{P}_{i+1}$ is a simple reduction/extension, constant transposition, or box transposition of $\mathcal{P}_i$ for all $i = 1, \ldots, (m-1)$.*

5 Relation to graphical models

This section, included into the paper at the instigation of the anonymous reviewer, is intended for the reader familiar with probabilistic graphical models, who requires to see the relation between graphical and compositional approaches to multidimensional probability distributions representation. It means that, among others, the results described in this section will not be used in subsequent parts of this text, and therefore the section may be skipped without depriving legibility of the rest of the paper.

It is known that the class of compositional models is, in a way, equivalent to a class of distributions representable in a form of Bayesian networks [Jiroušek, 2004], which can be introduced in two different ways. One possibility is to define

Bayesian network (BN) as a couple of an acyclic directed graph (DAG) and a respective system of conditional probability distributions. Here, we will use an alternative approach that defines a BN as a probability distribution factorizing with respect to a DAG. In any case we have to use a couple of symbols from graph theory.

Let us consider a DAG $G = (V, \mathcal{E})$ with nodes from a set of variables $V = \{x_1, x_2, \ldots, x_q\}$ and the set of oriented edges $\mathcal{E}$. If $(x_i \rightarrow x_j) \in \mathcal{E}$ then we say that x_i is a *parent* of x_j, and $pa(x_j)$ denotes the set of all parents of x_j. For all $j = 1, \ldots, n$ let $K_j = pa(x_j) \cup \{x_j\}$. We say that a probability distribution $\kappa(V)$ is a BN with DAG $G = (V, \mathcal{E})$ if it factorizes with respect to G, i.e., if

$$\kappa(V) = \frac{\prod_{i=1}^{q} \kappa^{\downarrow K_i}}{\prod_{i=1}^{q} \kappa^{\downarrow pa(x_i)}}. \tag{5.1}$$

5.1 Transformation of a BN into a compositional model and vice versa

To get a compositional model representing a distribution $\kappa(V)$, which is a BN with graph $G = (V, \mathcal{E})$, is a simple task. First, one has to realize that nodes of a DAG can be ordered in the way that parents are always before their children. Without loss of generality assume it is the ordering $(x_1, x_2, \ldots, x_q)$, i.e.,

$$x_i \in pa(x_j) \quad \Longrightarrow \quad i < j.$$

This ordering guarantees the fact that $pa(x_j) \subseteq \{x_1, x_2, \ldots, x_{j-1}\}$, and therefore formula (5.1) rewrites into the form

$$\kappa(V) = \kappa^{\downarrow K_1} \triangleright \kappa^{\downarrow K_2} \triangleright \ldots \triangleright \kappa^{\downarrow K_q},$$

where $K_i = \{i\} \cup pa(i)$ for all $i = 1, 2, \ldots, q$. Notice that the structure of this model $(K_1, K_2, \ldots, K_q)$ is unambiguously specified by the graph of BN.

The opposite transformation of a compositional model into a BN is a little bit more complicated. Consider a general compositional model $\kappa(V) = \pi_1(L_1) \triangleright \pi_2(L_2) \triangleright \ldots \triangleright \pi_n(L_n)$ with structure $\mathcal{P} = (L_1, L_2, \ldots, L_n)$, such that $\bigcup_{i=1}^{n} L_i = V$. This model is, as a rule, equivalent to several BNs. To get a unique DAG we have to choose an ordering of variables from V. Let this ordering be defined by the relation $\prec$. Now, the definition of the required DAG is simple: $G = (V, \mathcal{E})$, where

> $(x_i \rightarrow x_j) \in \mathcal{E}$ *iff* there exists $k \in \{1, \ldots, n\}$, such that $\{x_i, x_j\} \subseteq L_k$, $j \in R(\mathcal{P}, L_k)$, and either $i \in S(\mathcal{P}, L_k)$, or $i \prec j$. (5.2)

The reader certainly noticed that we do not need the ordering on the whole set V but just on all subsets $R(\mathcal{P}, K_k)$. Given these orderings the graph of the resulting BN is uniquely given by the structure of the considered compositional model. In general, however, the respective DAG is not unique. It follows from (5.2) that the orientation of edges is from S to R-part of the involved set. If both variables are from R-part of the set, then the edge orientation may be arbitrary and it is determined by relation $\prec$, randomly chosen before.

5.2 Impact of basic operations

In this subsection we answer the question how the graph of the corresponding BN changes when we apply a basic operation (constant or box transpositions) to a compositional model. Realize that the existence of reducible sets does not influence the respective graphs constructed in the preceding paragraph, and therefore adding/removing reducible sets does not change the structure of the equivalent BN.

Recalling definition of non-trivial sets we immediately see from (5.2) that $(x_i \to x_j) \in \mathcal{E}$ *iff* $(x_i, x_j) \in \mathcal{N}(\mathcal{P})$, or, in other words, every edge corresponds to a non-trivial set of cardinality two and vice-versa. Therefore, using Theorems 3.7 and 4.10 guaranteeing that constant and box transposition does not change the set of non-trivial sets, we see that constant and box transpositions neither introduce nor delete an edge in the DAG of the respective BN. The question remains whether these transformation can change the orientation of the edges. As mentioned above, the orientation of every edge is from S to R, or, if both variables are in an R-part of the set, then the orientation is given by the previously chosen ordering $\prec$.

5.2.1 Constant transposition

Consider a structure $\mathcal{P} = (L_1, \ldots, L_n)$. Recall that its reordering $\mathcal{P}' = \mathcal{P}(\ell{-}1 \curvearrowright \ell)$ is called constant transposition if $R(\mathcal{P}, L_{\ell-1}) \cap L_\ell = \emptyset$. Using Theorem 4.5, It means that

$$S(\mathcal{P}, L_j) = S(\mathcal{P}', L_j) \quad \text{and} \quad R(\mathcal{P}, L_j) = R(\mathcal{P}', L_j)$$

for all $j = 1, \ldots, n$. Therefore, regarding the rule (5.2) we can immediately see that the constant transposition does not change the DAG of the corresponding BN.

5.2.2 Box transposition

Recall that the reordering $\mathcal{P}(\ell{-}1 \curvearrowright \ell)$ of a structure $\mathcal{P} = (L_1, \ldots, L_n)$ is called a box transposition if $S(\mathcal{P}, L_{\ell-1}) = S(\mathcal{P}, L_\ell) \setminus R(\mathcal{P}, L_{\ell-1})$. Theorem 4.8 says that if $\mathcal{P}' = \mathcal{P}(\ell - 1 \curvearrowright \ell)$ is a box transposition of $\mathcal{P}$, then $S(\mathcal{P}', L_\ell) = S(\mathcal{P}, L_{\ell-1})$ and $S(\mathcal{P}', L_{\ell-1}) = S(\mathcal{P}, L_\ell)$. Therefore

$$R(\mathcal{P}', L_\ell) = L_\ell \setminus S(\mathcal{P}', L_\ell) = L_\ell \setminus S(\mathcal{P}, L_{\ell-1}) = R(\mathcal{P}, L_\ell) \cup (R(\mathcal{P}, L_{\ell-1}) \cap L_\ell).$$

So, the following situations may happen if $R(\mathcal{P}, L_{\ell-1}) \cap L_\ell \neq \emptyset$. Consider $x_i \in (R(\mathcal{P}, L_{\ell-1}) \cap L_\ell)$ and $x_j \in R(\mathcal{P}, L_\ell)$. In this case, naturally, $x_i \in S(\mathcal{P}, L_\ell)$, and therefore the graph of a BN corresponding to structure $\mathcal{P}$ has an edge $(x_i \to x_j)$. On the other hand side, both $x_i, x_j \in R(\mathcal{P}', L_\ell)$. and therefore the orientation of the respective edge in the graph corresponding to structure $\mathcal{P}'$ depends on relation $\prec$. An analogous situations happens if $x_i \in (R(\mathcal{P}, L_{\ell-1}) \cap L_\ell)$ and $x_k \in (R(\mathcal{P}, L_{\ell-1}) \setminus L_\ell)$. In this case the orientation of the edge connecting these two nodes in the graph of a BN corresponding to structure $\mathcal{P}$ is given by relation $\prec$, whereas the graph corresponding to structure $\mathcal{P}'$ contains the edge $(x_i \to x_j)$ regardless relation $\prec$.

So, we see that a box transposition can change the orientation of an edge of the corresponding BN but only in situations when there are several equivalent BNs corresponding to the given structure.

6 Operations on generating sequences

Up to now we have studied the impact of elementary operations (transposition, and adding/deleting sets) on structural independence. From this point forward, we will study what happens when performing the introduced elementary operations with distributions – elements of a respective generating sequence (a sequence of low-dimensional probability distributions that represent a compositional model). What is the impact of each of these operations on the respective compositional model? What are sufficient conditions for two compositional models with equivalent structures to be the same? These are the questions to be answered in this Section.

In order to simplify the following lemmata, we will work with a model whose generating sequence consists of only three distributions $\pi_1(K_1), \pi_2(K_2)$, and $\pi_3(K_3)$. Thus, we will consider a generating sequence with a structure $\mathcal{P} = (K_1, K_2, K_3)$, and will apply respective operations on π_2 and π_3. Notice that this simplification is not at the expense of generality. Indeed, realize that $\pi_1(K_1)$ may be a compositional model itself – it may be composed from several distributions. Similarly, if $\pi_1(K_1), \pi_2(K_2), \pi_3(K_3)$ is from the beginning of a much longer generating sequence, Lemma 2.3 says that we in fact study properties of a marginal of a multidimensional distribution represented by a long generating sequence.

The first lemma deals with the simple extension/reduction of a structure and of the respective compositional model.

Lemma 6.1. *Consider three distributions $\pi_1(K_1), \pi_2(K_2)$, and $\pi_3(K_3)$ such that $\pi_1 \triangleright \pi_2$ is defined. If K_2 is reducible in $\mathcal{P} = (K_1, K_2, K_3)$ then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3. \tag{6.1}$$

Proof. K_2 is reducible if $K_2 = S(\mathcal{P}, K_2)$, i.e., if $K_2 \subseteq K_1$. Therefore $K_1 = K_1 \cup K_2$ and therefore $(\pi_1 \triangleright \pi_2)^{\downarrow K_1} = \pi_1$ by Lemma 2.3, which completes the proof. □

Lemma 6.2. *Consider three distributions $\pi_1(K_1), \pi_2(K_2)$, and $\pi_3(K_3)$. If (K_1, K_3, K_2) is a constant transposition of (K_1, K_2, K_3) then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleright \pi_2. \tag{6.2}$$

Proof. Recall that (K_1, K_3, K_2) is a constant transposition of (K_1, K_2, K_3) if $R(\mathcal{P}, K_2) \cap K_3 = \emptyset$. It means that $K_2 \cap K_3 \subseteq S(\mathcal{P}, K_2) \subseteq K_1$. Therefore, we can apply Lemma (5.7) from [Jiroušek, 2011], which yields that $\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleright \pi_2$. □

Lemma 6.3. *Consider three distributions $\pi_1(K_1), \pi_2(K_2)$, and $\pi_3(K_3)$ such that π_2 and π_3 are consistent. If (K_1, K_3, K_2) is a box transposition of (K_1, K_2, K_3) then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleright \pi_2. \tag{6.3}$$

Proof. Consider $\mathcal{P} = (K_1, K_2, K_3)$ and denote $\mathcal{P}' = (K_1, K_3, K_2)$. Let us start by showing that, under the given assumption, $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is defined *iff* $\pi_1 \triangleright \pi_3 \triangleright \pi_2$ is defined. Recall that $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is defined *iff*

(i) $\pi_1^{\downarrow S(\mathcal{P},K_2)} \ll \pi_2^{\downarrow S(\mathcal{P},K_2)}$, and

(ii) $(\pi_1 \triangleright \pi_2)^{\downarrow S(\mathcal{P},K_3)} \ll \pi_3^{\downarrow S(\mathcal{P},K_3)}$.

Analogously, $\pi_1 \triangleright \pi_3 \triangleright \pi_2$ is defined *iff*

(iii) $\pi_1^{\downarrow S(\mathcal{P}',K_3)} \ll \pi_3^{\downarrow S(\mathcal{P}',K_3)}$, and

(iv) $(\pi_1 \triangleright \pi_3)^{\downarrow S(\mathcal{P}',K_2)} \ll \pi_2^{\downarrow S(\mathcal{P}',K_2)}$.

Using Theorem 4.8 and consistency of π_2 and π_3 we see that (i) and (iii) coincide. Let us prove that (ii) and (iv) also coincide. By the definition of box transposition, observe that $S(\mathcal{P}, K_2) \subseteq S(\mathcal{P}, K_3)$. Denoting $S = S(\mathcal{P}, K_3)$ we can easily see (just using the definition of $S(\mathcal{P}, K_i)$) that

$$(K_1 \cap K_2) = S(\mathcal{P}, K_2) \subseteq S \subseteq K_2 \subseteq (K_1 \cup K_2), \tag{6.4}$$

from which we immediately see that

$$(S \cap K_2) = S = (S \cap K_3). \tag{6.5}$$

This enables us to compute

$$(\pi_1 \triangleright \pi_2)^{\downarrow S} = \pi_1^{\downarrow S \cap K_1} \triangleright \pi_2^{\downarrow S \cap K_2} = \pi_1^{\downarrow S \cap K_1} \triangleright \pi_3^{\downarrow S \cap K_3} = (\pi_1 \triangleright \pi_3)^{\downarrow S},$$

where the first equation is guaranteed by relationship (6.4) and Lemma 2.3, the second equation follows from equality (6.5) and the assumption of consistency on π_2 and π_3, and the last equation is guaranteed again by Lemma 2.3. Thus we have shown that (ii) is equivalent to (iv).

Let us now assume that both expressions in equality (6.3) are defined. Because of Lemma 4.8 and the fact that π_2 and π_3 are assumed to be consistent, the expressions

$$\pi_1 \rhd \pi_2 \rhd \pi_3 = \frac{\pi_1\pi_2\pi_3}{\pi_2^{\downarrow S(\mathcal{P},K_2)}\pi_3^{\downarrow S(\mathcal{P},K_3)}},$$
$$\pi_1 \rhd \pi_3 \rhd \pi_2 = \frac{\pi_1\pi_2\pi_3}{\pi_3^{\downarrow S(\mathcal{P}',K_3)}\pi_2^{\downarrow S(\mathcal{P}',K_3)}}$$

are mutually equivalent, which completes the proof. □

7 Conditioning and flexible sequences

Knowledge of structural properties of a compositional model helps us, among other things, when computing conditional distributions. Namely, it can be shown that computation of a conditional distribution $\pi(\cdot|u = \alpha)$, for distribution π represented in the form of a compositional model $\pi = \pi_1 \rhd \ldots \rhd \pi_n$, is granted to be easy only if the conditioning variable u appears among the arguments of the first distribution π_1. This property is more precisely expressed in the following assertion.

Theorem 7.1. *Let $\pi_1, \pi_2, \ldots, \pi_n$ be a generating sequence with structure $\mathcal{P} = (K_1, K_2, \ldots, K_n)$ and $u \in K_1$. Then, for any value α of variable u for which $\pi_1(u = \alpha) > 0$,*

$$(\pi_1 \rhd \pi_2 \rhd \ldots \rhd \pi_n)\left(\widehat{K}(\mathcal{P}) \setminus \{u\}|u = \alpha\right) = \kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_n,$$

where for all $i = 1, 2, \ldots, n$

$$\kappa_i(K_i \setminus \{u\}) = \begin{cases} \pi_i(K_i) & \text{if } u \notin K_i \\ \pi_i(K_i \setminus \{u\}|u = \alpha) & \text{if } u \in K_i. \end{cases}$$

Proof. Let us show that the assertion holds for $n = 2$. For $n = 1$ the assertion is trivial, and for $n > 2$ it can easily be proven by the technique of mathematical induction based on the fact it holds for $n = 2$.

Let us distinguish between two situations: $u \in K_2$ and $u \notin K_2$. If $u \in K_2$

then

$$
\begin{aligned}
(\pi_1 \triangleright \pi_2)((K_1 \cup K_2) \setminus \{u\} | u = \alpha) &= \frac{(\pi_1 \triangleright \pi_2)((K_1 \cup K_2) \setminus \{u\}, u = \alpha)}{(\pi_1 \triangleright \pi_2)^{\downarrow \{u\}}(u = \alpha)} \\
&= \frac{\pi_1(K_1 \setminus \{u\}, u = \alpha) \triangleright \pi_2(K_2 \setminus \{u\}, u = \alpha)}{\pi_1^{\downarrow \{u\}}(u = \alpha)} \\
&= \frac{\pi_1(K_1 \setminus \{u\}, u = \alpha)}{\pi_1^{\downarrow \{u\}}(u = \alpha)} \cdot \frac{\pi_2(K_2 \setminus \{u\}, u = \alpha)}{\pi_2^{\downarrow K_1 \cap K_2}((K_1 \cap K_2) \setminus \{u\}, u = \alpha)} \\
&= \pi_1(K_1 \setminus \{u\} | u = \alpha) \cdot \frac{\pi_2(K_2 \setminus \{u\}, u = \alpha)}{\pi_2^{\downarrow \{u\}}(u = \alpha) \cdot \pi_2^{\downarrow K_1 \cap K_2}((K_1 \cap K_2) \setminus \{u\} | u = \alpha)} \\
&= \frac{\pi_1(K_1 \setminus \{u\} | u = \alpha) \cdot \pi_2(K_2 \setminus \{u\} | u = \alpha)}{\pi_2^{\downarrow K_1 \cap K_2}((K_1 \cap K_2) \setminus \{u\} | u = \alpha)} \\
&= \pi_1(K_1 \setminus \{u\} | u = \alpha) \triangleright \pi_2(K_2 \setminus \{u\} | u = \alpha)
\end{aligned}
$$

If $u \notin K_2$ the computation, though analogous, is even simpler.

$$
\begin{aligned}
(\pi_1 \triangleright \pi_2)((K_1 \cup K_2) \setminus \{u\} | u = \alpha) &= \frac{\pi_1(K_1 \setminus \{u\}, u = \alpha) \triangleright \pi_2(K_2)}{\pi_1^{\downarrow \{u\}}(u = \alpha)} \\
&= \frac{\pi_1(K_1 \setminus \{u\}, u = \alpha)}{\pi_1^{\downarrow \{u\}}(u = \alpha)} \cdot \frac{\pi_2(K_2)}{\pi_2^{\downarrow K_1 \cap K_2}((K_1 \cap K_2))} \\
&= \pi_1(K_1 \setminus \{u\} | u = \alpha) \triangleright \pi_2(K_2)
\end{aligned}
$$

□

In light of Theorem 7.1, it seems reasonable to study this question: When and how can a given generating sequence be reordered so that a desired variable is among the arguments of the first distribution? However, not knowing which variable will be the conditioning one, we will solve this problem for all variables from $\mathcal{P}$ at once. This is why we will be interested in sequences for which any variable may appear among the arguments of the first distribution (naturally, after a necessary reordering). This property is met by the so-called *flexible sequences* that, in addition to a stronger concept of decomposable generating sequences, were already defined in [Jiroušek, 2011].

Definition 7.2. *A generating sequence $\pi_1, \pi_2, \ldots, \pi_n$ with structure $\mathcal{P} = (K_1, K_2, \ldots, K_n)$ is called* flexible *if for each $u \in \widehat{K}(\mathcal{P})$ there exists a permutation $i_1, i_2, \ldots, i_n$ of $1, 2, \ldots, n$ such that $u \in K_{i_1}$ and*

$$\pi_{i_1} \triangleright \pi_{i_2} \triangleright \ldots \triangleright \pi_{i_n} = \pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n.$$

In other words, flexible sequences are those which can be reordered in many ways so that each variable can appear among the arguments of the first distribution. However, this does not mean that each distribution appears at the beginning of the generating sequence. If this were the case, then flexible sequences would form a subclass of the so-called *perfect sequences* [Jiroušek, 2011] (every distribution from a perfect sequence is a marginal of the represented distribution – Theorem (10.5) in [Jiroušek, 2011]).

It seems natural that if a generating sequence and its reordering represent the same probability distribution, they should also induce the same system of structural independencies. In other words, their structures should be equivalent. This is why we are going to define and study the concept of *structural flexibility* as well.

Definition 7.3. *A structure $\mathcal{P}$ is called* flexible *if for all $u \in \widehat{K}(\mathcal{P})$ there exists its equivalent reordering $\mathcal{P}'$ such that u appears in the first set of $\mathcal{P}'$.*

Let us stress that similar to the flexibility of generating sequences, the structure flexibility definition does not require that each set from the structure should appear at the beginning of an equivalent structure. And yet, we will show that structure flexibility is much stronger than flexibility for generating sequences; namely, it coincides with decomposability defined by the well-known *running intersection property* (RIP) - see the definition below.

Remark 7.4. As already mentioned in Remark 3.8, a structure $\mathcal{P}$ of a compositional model defines an order on the set of the respective variables from $\widehat{K}(\mathcal{P})$. These variables are ordered with respect to their first appearance in $\mathcal{P}$. There is a strong relationship between special patterns from $\mathcal{N}(\mathcal{P})$ and this partial order. For example, if $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$ then there exists a set $K_i \in \mathcal{P}$ such that $\{u, v, w\} \subseteq K_i$ and at least one of u, v, w lies in its R-part. If simultaneously $\{u, v\} \notin \mathcal{N}(\mathcal{P})$ then $\{u, v\} \subseteq S(K_i, \mathcal{P})$, and, necessarily, $w \in R(K_i, \mathcal{P})$. Therefore, both u and v have to be introduced before w in $\mathcal{P}$. Since the system of non-trivial sets $\mathcal{N}(\mathcal{P})$ is one of the characteristics of the equivalence relation on structures, it means that u and v have to be introduced before w in every structure equivalent with $\mathcal{P}$ (see Theorem 3.7).

Denote $\mathcal{N}_{3-2}(\mathcal{P}) = \{\{u, v, w\} \in \mathcal{N}(\mathcal{P}) | \{u, v\} \notin \mathcal{N}(\mathcal{P})\}$. It will follow from Lemma 7.6 and Theorem 7.7 that $\mathcal{N}_{3-2}(\mathcal{P}) = \emptyset$ for flexible $\mathcal{P}$. First, we will prove that $\mathcal{N}_{3-2}(\mathcal{P}) = \emptyset$ is equivalent to the running intersection property (RIP), which is defined as follows for a structure $\mathcal{P}$ (recall that set $K_i \in \mathcal{P}$ is irreducible in $\mathcal{P}$ if $R(\mathcal{P}, K_i) \neq \emptyset$).

Definition 7.5. *We say that structure $\mathcal{P} = (K_1, \ldots, K_n)$ satisfies the running intersection property (RIP) if for every irreducible set $K_i \in \mathcal{P}$ there exists an irreducible set $K_j \in \mathcal{P}$ such that $j < i$, and $S(K_i, \mathcal{P}) \subseteq K_j$.*

Lemma 7.6. $\mathcal{N}_{3-2}(\mathcal{P}) = \emptyset$ iff *$\mathcal{P}$ satisfies RIP.*

Proof. First, assume that $\mathcal{P} = (K_1, \ldots, K_n)$ satisfies RIP and there exists $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$ such that $\{u, v\} \notin \mathcal{N}(\mathcal{P})$. Let K_i be the set for which

$w \in R(\mathcal{P}, K_i)$. Then $\{u, v\} \subseteq S(K_i, \mathcal{P})$ (because $\{u, v\} \notin \mathcal{N}(\mathcal{P})$), and employing the definition of RIP, we know that there must exist irreducible $K_j \in \mathcal{P}$, $j < i$ such that $\{u, v\} \subseteq K_j$. Since we assume that $\{u, v\} \notin \mathcal{N}(\mathcal{P})$, neither of these two variables may lie in $R(\mathcal{P}, K_i)$, and therefore $\{u, v\} \subseteq S(K_j, \mathcal{P})$, which further implies the existence of another irreducible $K_k \in \mathcal{P}$, $k < j$ such that $\{u, v\} \subseteq K_k$. This reasoning process can be endlessly repeated, which contradicts with the fact that the number of the sets in $\mathcal{P}$ is finite.

To prove the opposite implication, i.e., $\mathcal{N}_{3-2}(\mathcal{P}) = \emptyset \Rightarrow \mathcal{P}$ satisfies RIP, we will use the induction on the length of the considered structure n. The assertion is trivial for $n = 2$: $S(K_2, \mathcal{P}) = K_1 \cap K_2 \subseteq K_1$.

Suppose the assertion holds for all structures of a length smaller than n. This assumption trivially implies that the assertion also holds for K_n reducible, and for K_n such that $|S(K_n, \mathcal{P})| = 1$. Now, we will show that it also holds for irreducible K_n for which $|S(K_n, \mathcal{P})| \geq 2$. Denote

$$v = \arg \max_{\bar{v} \in S(K_n, \mathcal{P})} \{j : \bar{v} \in R(K_j, \mathcal{P})\}, \tag{7.1}$$

i.e., v is the variable from $S(K_n, \mathcal{P})$ that is introduced last in $\mathcal{P}$. Since K_n is irreducible we can choose an arbitrary $w \in R(K_n, \mathcal{P})$. Let us show that $S(K_n, \mathcal{P}) \subseteq K_j$ for which $v \in R(K_j, \mathcal{P})$. Consider any $u \in S(K_n, \mathcal{P}), u \neq v$. Since $u, v \in S(K_n, \mathcal{P})$, we see that $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$, and therefore also $\{u, v\} \in \mathcal{N}(\mathcal{P})$ (because we assume $\mathcal{N}_{3-2}(\mathcal{P}) = \emptyset$). Therefore there must exist K_i such that $u, v \in K_i$, and at least one of them must lie in $R(K_i, \mathcal{P})$. But, neither $i < j$ ($v \notin K_1 \cup \ldots \cup K_{j-1}$ because $v \in R(K_j, \mathcal{P})$) nor $i > j$ (because of (7.1)) and therefore $u \in K_j$, which concludes the proof. $\square$

Now, we can prove that the notion of structural flexibility coincides with RIP.

Theorem 7.7. *A structure $\mathcal{P}$ is flexible* iff *it satisfies RIP.*

Proof. First, assume that $\mathcal{P}$ does not satisfy RIP, which means $\mathcal{N}_{3-2}(\mathcal{P}) \neq \emptyset$ due to Lemma 7.6. Therefore there exists a triplet $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$ such that $\{u, v\} \notin \mathcal{N}(\mathcal{P})$ in every equivalent structure, and therefore w cannot appear in the first set of any structure equivalent with $\mathcal{P}$ because u and v have to be introduced first.

To prove the opposite implication of the desired equivalence, we will prove a little bit stronger assertion: If structure $\mathcal{P}$ satisfies RIP then every irreducible set can be moved to the first position in the structure using two elementary operations: box and constant transpositions. It means that the resulting structure is equivalent with $\mathcal{P}$.

Let us proceed using the induction on the length of the structure. It is evident that the assertion holds for a structure consisting of only one set. Now, supposing it holds for all structures of a length smaller than n we will prove it also holds for structure $\mathcal{P} = (K_1, \ldots, K_n)$.

If K_n is reducible in $\mathcal{P}$ then the assertion holds because of the induction assumption. If it is irreducible, due to RIP there exists irreducible $K_j, j < n$ such that $S(K_n, \mathcal{P}) \subseteq K_j$. Moreover, if $(K_1, \ldots, K_n)$ meets RIP then $(K_1, \ldots, K_{n-1})$ meets RIP, too, and therefore, using the induction hypothesis, we can find $\mathcal{P}' = (K_{i_1}, \ldots, K_{i_{n-1}}, K_n)$, which is the reordering of $\mathcal{P}$ (that can be obtained using only box and constant transpositions) such that $K_j = K_{i_1}$. It is evident that $S(K_n, \mathcal{P}') = S(K_n, \mathcal{P})$. Hence, as shown in Remark 4.6, $\mathcal{P}'(n \curvearrowright 2)$ can be obtained from $\mathcal{P}'$ by a sequence of constant transpositions. Since the transposition $(2 \curvearrowright 1)$ is always either a box or a constant transposition (see Remark 4.9), the structure $\mathcal{P}'(n \curvearrowright 2)(2 \curvearrowright 1)$ meets the required property and can be obtained from $\mathcal{P}$ using only box and constant transpositions. □

Remark 7.8. The reader familiar with decomposable graphs knows that there are many different ways in which these graphs can be characterized. It follows from the existence of a *join tree* [Beeri, Fagin, Maier, and Yannakakis, 1983] that if a sequence meets RIP then it can be reordered into another RIP sequence so that the new sequence starts with an arbitrarily selected set. From this point of view, the preceding theorem is not surprising. The originality of the message contained in this assertion is twofold. For one thing, the respective reordering can be done with only the help of constant and box transpositions, and for another, RIP is guaranteed by a weaker property, which is the structural flexibility of a sequence.

Based on Theorem 7.7 we can immediately conclude the following assertion, which is the same as in Lemma 12.3. in [Jiroušek, 2011] (notice that for the application of box transposition by Lemma 6.3 we have to assume that the swapped distributions are consistent).

Corollary 7.9. *Let $\pi_1, \ldots, \pi_n$ be a generating sequence of pairwise-consistent distributions with flexible structure $(K_1, \ldots, K_n)$. Then the generating sequence is flexible.*

8 Conclusions

Compositional models provide a tool for efficient representation of multidimensional probability distributions. Note that an arbitrary probability distribution can be characterized by many properties. One of them is the system of probabilistic conditional independence statements induced by this distribution. It means that every compositional model – as a probability distribution – also induces a system of probabilistic conditional independence statements. A significant part of these statements is given by the respective model structure, which is the same for all the distributions represented in the form of compositional models with this structure.

This paper introduces a compositional model structure as a bearer of the information about these conditional independence statements. The first part of the paper recalls the basic properties of compositional models relevant to the notion of probabilistic conditional independence. In the second part, a separation criterion is presented, based on nonexistence of a Z-avoiding trail, enabling us to read the respective conditional independence statements from the given structure. It is worth repeating that two different structures may induce the same system of conditional independence statements; in this case we say they are equivalent. This issue is treated in the main part of the paper: we present two ways to characterize equivalent structures and describe transformations converting a given structure into another equivalent one. In the last part of the paper we also reveal the impact of these operations on probability distributions represented in the form of compositional models.

There are many other important questions that are not answered in this paper. First, let us stress that we dealt only with sequential models. Very interesting and important results concerning structures of more general (non-sequential) compositional expressions were achieved by Malvestuto [2014]. Nevertheless, even for sequential models some basic questions remained beyond the scope of this paper. For example: What is the number of structures equivalent with a given one? How to find a structure corresponding to a given system of conditional independence statements? And in case a system of conditional independence statements cannot be perfectly represented by any structure, how to find a structure inducing its maximal subsystem? So, one can see that there is still an interesting part of compositional model theory open for further research.

Acknowledgement

The paper is a survey of results on structural properties of compositional models that were achieved with the financial support of several research grants. During the past two years there were grants of Czech Science Foundation No. 403/12/2175 (first author) and No. 13-20012S (second author).

The authors also appreciate the valuable comments of the anonymous reviewer who suggested, among others, to include parts clarifying the relation between graphical and compositional models. This led to inclusion of Section 5.

References

Andersson, S.A. Madigan, D. Perlman,M.D. 1997. “A characterization of Markov equivalence classes for acyclic digraphs”, *Annals of Statistics*, 25: 505-541

Beeri, C. Fagin, R. Maier D., and Yannakakis, M. 1983. “On the Desirability of Acyclic Database Schemes”, *Journal of the ACM* 30 (3):479-513.

Bína, V. Jiroušek, R. 2006. “Marginalization in multidimensional compositional models” *Kybernetika* 42 (4): 405–422.

Chickering, D.M. 1995. “A transformational characterization of equivalent Bayesian network structures” *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence (UAI'95)* Philippe Besnard and Steve Hanks (Eds.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA: 87–98.

Cowell, R.G. Dawid, P.A. Lauritzen, S.L. Spiegelhalter, D.J. 1999. "Probabilistic networks and expert systems" Berlin: Springer.

Geiger, D., Pearl, J. 1988. "On the Logic of Causal Models" *Proceedings of the Fourth Conference on Uncertainty in Artificial Intelligence*:1304.2355.

Hemmecke, R., Lindner, S., Studený, M. 2012. "Characteristic imsets for learning Bayesian network structure" *International Journal of Approximate Reasoning* 53: 1336–1349.

Jiroušek R. 2004. "What is the difference between Bayesian networks and compositional models" *Proceedings of the 7th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty.* Awaji, Japan: 191–196.

Jiroušek R. 2008. "Persegrams of Compositional Models Revisited: conditional independence" *Proceedings of the 12th International Conference on Information Processing and Management of Uncertianlty in Konowledge-based Systems* L. Magdalena, M. Ojeda-Aciego, J. L. Verdegay (Eds.). Málaga, Spain: 915–922.

Jiroušek, R., 2011. "Foundations of compositional model theory". *International journal of general systems* 40: 623–678.

Kratochvíl V. 2011. "Characteristic Properties of Equivalent Structures in Compositional Models", *International Journal of Approximate Reasoning* 52 (5): 599–612.

Kratochvíl, V. 2013. "Probabilistic Compositional Models: solution of an equivalence problem", *International Journal of Approximate Reasoning* 54 (5): 590-601.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N. and Leimer, H.-G. 1990. "Independence properties of directed markov fields" *Networks* 20: 491-505.

Malvestuto, F. M. 2014. "Equivalence of compositional expressions and marginalization in compositional models". To appear in *Kybernetika* 50, (3).

Pearl, J. 1988. "Probabilistic Reasoning in Intelligent Systems" Morgan Kaufmann, San Francisco, California.

Pearl, J. Paz, A. 1987. "Graphoids: a graph-based logic for reasoning about relevance relations" *Advances in Artificial Intelligence II* B. du Boulay, D. Hogg, and L. Steels (Eds.). North-Holland, Amsterdam: 357–363.

Studený M. 2005. "Probabilistic Conditional Independence Structures", Springer London.

Studený, M. Hemmecke, R. Lindner, S. 2012."Characteristic imset: a simple algebraic representative of a Bayesian network structure" *Proceedings of the 5th European Workshop on Probabilistic Graphical Models* P. Myllymaki, T. Roos and T. Jaakkola (Eds.). HIIT Publications 2010-2: 257-264.

Studený, M. Vomlel, J. Hemmecke, R. 2010. “A geometric view on learning Bayesian network structures” *International Journal of Approximate Reasoning* 51: 578–586.

Verma, T.S. 1987. “Some Mathematical Properties of Dependency Models” Technical Reports R-103 (UCLA Cognitive Systems Laboratory, Los Angeles).

Verma, T.S. Pearl, J. 1990. “Causal networks: semantics and expressiveness” *Uncertainty in Artificial Intelligence* 4.

Verma, T.S. Pearl, J. 1991 “Equivalence and synthesis of Causal models” *Uncertainty in Artificial Intelligence* 6.

Wright, S. 1921. “Systems of mating I. The biometric relations between offspring and parent” *Genetics* 6: 111–123.