Probabilistic Compositional Models: solution of an equivalence problem☆

Václav Kratochvíl[a,b]

[a] *Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic*
[b] *University of Economics, Prague, Czech Republic*

Abstract

Probabilistic compositional models, similarly to graphical Markov models, are able to represent multidimensional probability distributions using factorization and closely related concept of conditional independence. Compositional models represent an algebraic alternative to the graphical models. The system of related conditional independencies is not encoded explicitly (e.g. using a graph) but it is hidden in a model structure itself. This paper provides answers to the question how to recognize whether two different compositional model structures are equivalent - i.e. whether they induce the same system of conditional independencies. Above that, it provides an easy way to convert one structure into an equivalent one in terms of some elementary operations on structures, closely related ability to generate all structures equivalent with a given one, and a unique representative of a class of equivalent structures.

Keywords:
probabilistic model, compositional model, conditional independence, equivalence, independence structure, equivalence problem

1. Introduction

One of the many ways to handle uncertainty in decision making is to express our knowledge about a real world problem using a multidimensional probability distribution and base the reasoning process on probability theory.

The efficient representation of multidimensional probability distribution (the size of probability distributions grows exponentially with the number of involved variables) is possible if the concept of probabilistic *conditional independence* (CI) [11] is taken into consideration. Since every CI-statement can be interpreted as a certain qualitative relationship among involved variables, the dimensionality of the problem can be reduced and a more effective way of storing the knowledge base can be found.

System of CI statements can be encoded in many different ways and e.g. in case of wide spread graphical models, this is done by graphs. The idea of compositional models (CM) is to abandon the necessity to use graphs to describe the CI structure of a modeled distribution. In contrast, CM describe directly how the multidimensional distribution is computed - *composed* - from a system of low-dimensional distributions and, therefore, need not represent respective CI structure explicitly. See [5] for CM basic properties and [10] (the logical but unofficial prequel of this paper) for more details about CI statements in CM.

The considerable advantage of CM is that they can also be developed in possibility theory [14] and Dempster-Shafer theory of belief functions [6],[7] equally efficiently. This means that they can also be applied to situations when the assumption of additivity is not adequate [9]. Nevertheless, in this paper, we

☆The research was partially supported by the *Institutional support of research of Faculty of Management, Prague University of Economics.*

Email address: velorex@utia.cas.cz (Václav Kratochvíl)
URL: http://www.utia.cas.cz/people/kratochvil (Václav Kratochvíl)

shall restrict our consideration to probabilistic case only. Another advantage is that CM appear to be less computationally demanding for the frequent task of computing marginal probability distributions [5].

The thorough investigation of the induced system of CI statements is essential not only for deeper understanding of CM but also for the further development of learning algorithms [1], etc. In case of CM, the system of presumed CI-statements is not very obvious at the first glance. It is encoded in the compositional model structure. However, the structure as a tool for representing such a system is imperfect. Two different structures may induce the same structural independencies; we say they are *equivalent*. Then, the question how to recognize two equivalent structures makes sense and it is of special importance to have a simple rule to do that (the notion of a rule's simplicity may differ when considering whether people or a computer will use it). Another very important aspect is the ability to generate all structures equivalent with a given one and an easy way to convert a structure into an equivalent one in terms of some elementary operations on structures. All these questions/problems are known as the *equivalence problem*.

This text covers the solution of all parts of the equivalence problem and it is closely tied to [10] where a partial solution of the equivalence problem was given. Let us say, this is its continuation.

2. Basic concepts

In this text, we will deal with a non-empty finite set of finite-valued variables $\{u, v, w, x, y, z, \ldots\}$, subsets of which will be denoted by upper-case Roman characters such as U, V, W, Z and specials K, R, S. The symbol $\pi(U)$ will be used for a probability distribution defined over variables U. Ordered sequences of variable sets will be denoted by calligraphic characters like $\mathcal{P} = (U_1, \ldots, U_n)$, $\mathcal{P}' = (U_1, U_2, U_3, U_4, U_5)$, or $\mathcal{P}'' = (U_2, U_1, U_3, U_4, U_5)$. Notice here that $\mathcal{P}' \neq \mathcal{P}''$ because $\mathcal{P}''$ is a reordering (permutation) of $\mathcal{P}'$. If not specified otherwise, $\mathcal{P} = (U_1, \ldots, U_n)$ in the following text. I.e. the number of sets in the sequence $\mathcal{P}$ is n. The symbol $\bigcup_{\mathcal{P}}$ will denote set of all variables from the sequence - $\bigcup_{\mathcal{P}'} = \bigcup_{\mathcal{P}''} = U_1 \cup \ldots \cup U_5$ and $\bigcup_{\mathcal{P}} = U_1 \cup \ldots \cup U_n$.

2.1. Permutations

For frequent reorderings we will use the concept of permutation [2] and we employ the well-known fact that every permutation can be expressed as a *product of transpositions*. Let us have a set and i, j be two position indices. The transposition $\sigma = (i\ j) = (j\ i)$ is the permutation which maps i-th element to j-th position $i \mapsto j$, j-th element to i-th position $j \mapsto i$, and fixes all other elements in the set. We also use r-cycles: r-cycle $(i_1\ i_2\ \ldots\ i_r)$ is the permutation which maps $i_1 \mapsto i_2$, $i_2 \mapsto i_3$, $\ldots$, $i_{r-1} \mapsto i_r$, $i_r \mapsto i_1$ and fixes all other elements in the set. Logically, every r-cycle can be expressed as a product of transpositions (2-cycles). Note that for example,

$$(1\ 2\ 3\ 4\ 5) = (5\ 4)(4\ 3)(3\ 2)(2\ 1)$$

does what we want for a cycle of length 5. Analogous calculations establish the same for other lengths.

If σ is a permutation, we shall write $i\sigma$ for the image of the element $i \in X$ under σ (rather then $\sigma(i)$). The principal reason for doing this is that it makes composition of permutations much easier: $\sigma_1\sigma_2$ will mean we apply σ_1 first and then apply σ_2, rather than the other way around. E.g. $\mathcal{P}(1\ 2) = (U_2, U_1, U_3, U_4)$ for $\mathcal{P} = (U_1, U_2, U_3, U_4)$, in case of natural numbers $2(1\ 2) = 1$, $3(1\ 2\ 4) = 3$.

2.2. R and S parts

Assume a structure $\mathcal{P} = (U_1, \ldots, U_n)$. We use the symbol $U_i \in \mathcal{P}$ to express the fact that U_i belongs to $\mathcal{P}$. Such a set will be called a *column* of $\mathcal{P}$ to distinguish it from a general set of variables $U \subseteq \bigcup_{\mathcal{P}}$. Moreover, we recognize the auxiliary notation $K_i^{\mathcal{P}}$ which reflects the ordering in $\mathcal{P}$: $K_i^{\mathcal{P}}$ is the i-th column in $\mathcal{P}$; e.g., for $\mathcal{P} = (U_1, \ldots, U_n)$ it holds that $K_i^{\mathcal{P}} = U_i$ for all $i = 1, \ldots, n$. The reason for double notation of the same column within $\mathcal{P}$ is as follows: Consider $\mathcal{P} = (U_1, U_2, U_3, U_4)$ and let $\mathcal{P}'$ be its reordering - for example - $\mathcal{P}' = (U_4, U_3, U_2, U_1) = \mathcal{P}(1\ 4)(2\ 3)$. In this case, U_3 is the second column in $\mathcal{P}'$ - i.e. $K_2^{\mathcal{P}'} = U_3$; on the contrary, U_2 is the third column in $\mathcal{P}'$: $K_3^{\mathcal{P}'} = U_2$. Note that $K_i^{\mathcal{P}} = K_{i\sigma}^{\mathcal{P}\sigma}$ for every column and every permutation σ.

In addition, each column $K_i^{\mathcal{P}}$ can be divided into two disjoint parts with respect to its position in $\mathcal{P}$. We denote them $R_i^{\mathcal{P}}$ and $S_i^{\mathcal{P}}$, where

$$R_1^{\mathcal{P}} = K_1^{\mathcal{P}} \quad and \quad R_i^{\mathcal{P}} = K_i^{\mathcal{P}} \setminus (K_1^{\mathcal{P}} \cup \ldots \cup K_{i-1}^{\mathcal{P}}) \; \forall i = 2, \ldots, n$$

and

$$S_1^{\mathcal{P}} = \emptyset \quad and \quad S_i^{\mathcal{P}} = K_i^{\mathcal{P}} \cap (K_1^{\mathcal{P}} \cup \ldots \cup K_{i-1}^{\mathcal{P}}) \; \forall i = 2, \ldots, n$$

It has the following meaning: $R_i^{\mathcal{P}}$ denotes the variables first occurring in the i-th column of the sequence $\mathcal{P}$ (taken from left to right). Conversely, $S_i^{\mathcal{P}}$ denotes variables from the i-th column of $\mathcal{P}$ which have been already used in a previous one. Observe that $K_i^{\mathcal{P}} = R_i^{\mathcal{P}} \cup S_i^{\mathcal{P}}$ and $\{R_i^{\mathcal{P}}\}_{i=1,\ldots,n}$ is a disjoint covering of $\bigcup_{\mathcal{P}}$. We say that a column $K_i^{\mathcal{P}}$ is *trivial* in $\mathcal{P}$ if $R_i^{\mathcal{P}} = \emptyset$ (i.e. $K_i^{\mathcal{P}} = S_i^{\mathcal{P}}$). Otherwise, $K_i^{\mathcal{P}}$ is *non-trivial* in $\mathcal{P}$. Trivial column does not introduce any new variable to the sequence $\mathcal{P}$.

Example 2.1. *Consider $\mathcal{P} = (U_1, U_2, U_3, U_4) = (\{u\}, \{v, w\}, \{u, v, x\}, \{w, x, y\})$ and its permutation $\mathcal{P}' = \mathcal{P}(1\ 4)(2\ 3) = (U_4, U_3, U_2, U_1)$. For the respective R and S-parts see Table 1. Notice that U_3 and U_4*

i	$K_i^{\mathcal{P}}$	$R_i^{\mathcal{P}}$	$S_i^{\mathcal{P}}$	$K_i^{\mathcal{P}'}$	$R_i^{\mathcal{P}'}$	$S_i^{\mathcal{P}'}$
1	$U_1 = \{u\}$	$\{u\}$	$\emptyset$	$U_4 = \{w, x, y\}$	$\{w, x, y\}$	$\emptyset$
2	$U_2 = \{v, w\}$	$\{v, w\}$	$\emptyset$	$U_3 = \{u, v, x\}$	$\{u, v\}$	$\{x\}$
3	$U_3 = \{u, v, x\}$	$\{x\}$	$\{u, v\}$	$U_2 = \{v, w\}$	$\emptyset$	$\{v, w\}$
4	$U_4 = \{w, x, y\}$	$\{x, y\}$	$\{w\}$	$U_1 = \{u\}$	$\emptyset$	$\{u\}$

Table 1: R- and S-parts of $\mathcal{P}'$ and $\mathcal{P}' = \mathcal{P}(1\ 4)(2\ 3)$.

are trivial in $\mathcal{P}'$ but not in $\mathcal{P}$.

3. Compositional model

Compositional model is a multidimensional distribution assembled from a sequence of low-dimensional *unconditional distributions*, with the aid of an *operator of composition*. The binary *operator of composition* $\triangleright$ used during the compositioning process is basically a normalized product of its parameters designed to create a probability distribution over the union of variables for which the input distributions are defined:

Definition 3.1. *For two arbitrary distributions $\pi_1(U)$ and $\pi_2(V)$ their* composition *is given by the formula*

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

if $\pi_1(U \cap V) \ll \pi_2(U \cap V)$ [1]*, otherwise the composition remains undefined.*

As stated above, the result of the composition (if defined) is a new distribution. We can iteratively repeat the process of composition to obtain a multidimensional distribution. Consider a sequence of low-dimensional probability distributions $\pi_1(U_1)$, $\pi_2(U_2)$, $\ldots$, $\pi_n(U_n)$. If all compositions are defined, then multidimensional distribution (compositional model) $(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)(U_1 \cup U_2 \cup \ldots \cup U_n))$ is represented by the sequence $\pi_1, \pi_2, \ldots, \pi_n$. Regarding the fact that the operator $\triangleright$ is neither commutative nor associative, we always apply the operator from left to right; e.g.,

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \ldots \triangleright \pi_n = (((\pi_1 \triangleright \pi_2) \triangleright \pi_3) \triangleright \ldots) \triangleright \pi_n.$$

Note that the operator of composition is closely related to the notion of conditional probability independence; the operator of composition introduces conditional independence among the variables:

[1] $\pi_1(U) \ll \pi_2(U)$ denotes that the distribution $\pi_1(U)$ is absolutely continues with respect to distribution $\pi_2(U)$, which, in our finite settings, means that whenever $\pi_1(U)$ is positive also $\pi_2(U)$ must be positive.

Assertion 3.2. *Let $\pi = \pi_1 \triangleright \pi_2$ be defined for $\pi_1(U)$ and $\pi_2(V)$ and $U \setminus V \neq \emptyset \neq V \setminus U$. Then $(U \setminus V) \perp\!\!\!\perp (V \setminus U) | (U \cap V)[\pi]$.*

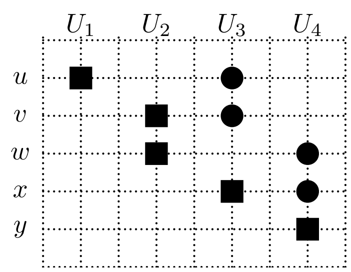
Having a model defined by sequence of probability distributions $\pi_1(U_1)$, $\pi_2(U_2)$, $\ldots, \pi_n(U_n)$ then the sequence of sets of variables $\mathcal{P} = (U_1, U_2, \ldots, U_n)$ is said to be its *structure*. Considering both a probability distribution represented by a compositional model and Assertion 3.2, we can see that there are independence statements induced fully by the model structure. Above that, every probability distribution represented by a model with the same structure has to meet identical independence statements. We will call them *structural independencies*. (To read more about compositional models we refer the reader to the readable survey [5] written by R. Jiroušek.)

3.1. Persegram

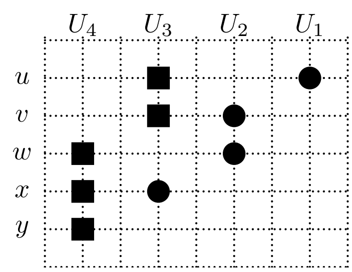
To read structural independencies, we use structure visualization - the so-called *persegram*. Let $\mathcal{P}$ be a structure of a compositional model. Its persegram is a table in which rows correspond to variables from $\bigcup_{\mathcal{P}}$ (in an arbitrary order) and columns to all sets $U_i \in \mathcal{P}$; the ordering of the columns corresponds to the structure. 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 marker in each row) are *box-markers*, and for other occurrences there are *bullets*. Observe that bullets in the i-th column correspond to variables from $S_i^{\mathcal{P}}$ while box-markers to variables from $R_i^{\mathcal{P}}$.

Remark 3.3. *The word "column" has two different interpretations now. It is either a set from a structure or a part of a persegram. Note that these two concepts are closely related.*

Example 3.4. *Consider two structures $\mathcal{P}$ and $\mathcal{P}' = \mathcal{P}(1\ 4)(2\ 3)$ from Example 2.1 once more. See respective persegrams in Figure 1. Note that the shape of markers may variate during permutation - e.g. while $[U_1, u]$ is a box-marker in $\mathcal{P}$, it is a bullet in $\mathcal{P}'$.*



(a) $\mathcal{P}$



(b) $\mathcal{P}(1\ 4)(2\ 3)$

Figure 1: Persegram of a structure and one of its permutation

3.2. Z-avoiding trails

Here we shall demonstrate how to read structural independencies from respective persegram. Such an independence statement is indicated by the absence of a *trail connecting relevant markers and avoiding others* defined below. This technique was originally published in [8] altogether with this theorem: "*Every independence statement read from the structure (or its persegram) of a compositional model corresponds to probability independence statement valid for every multidimensional probability distribution represented by a compositional model with this structure.*"

Definition 3.5. *A sequence of markers $m_0, \ldots, m_t$ in a persegram of a structure $\mathcal{P}$ is called a Z-*avoiding trail *($Z \subseteq \bigcup_{\mathcal{P}}$) that connects m_0 and m_t if it meets the following five conditions:*

0. neither m_0 nor m_t corresponds to a variable from Z

1. *for each $s = 1, \ldots, t$, the couple (m_{s-1}, m_s) is either in the same row (i.e., horizontal connection) or in the same column (vertical connection);*

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

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

4. *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 will be denoted by $u \not\perp\!\!\!\perp v|Z[\mathcal{P}]$.*

Example 3.6. *Consider structures $\mathcal{P} = (U_1, U_2, U_3, U_4)$ and $\mathcal{P}' = (U_4, U_3, U_2, U_1)$ from the previous examples. There are two different sequences of markers highlighted in persegrams of both structures in Figure 2. In order to illustrate vertical and horizontal connections and to highlight the ordering, each two consecutive markers are connected with a line, either solid or dashed.*

By the solid line we depict sequence $\tau_1 = [U_2, w], [U_2, v], [U_3, v], [U_3, x]$ and by the dashed one sequence $\tau_2 = [U_3, u], [U_3, x], [U_4, x], [U_4, y], [U_4, w]$. Note that τ_1 is Z-avoiding trail in $\mathcal{P}$ if $Z = \emptyset$ or $Z = \{u\}$. (There are horizontal and regular vertical connections, regularly alternating.) Similarly τ_2 is a Z-avoiding trail if $\{y\} \subseteq Z \subseteq \{y, v\}$ (y has to be a part of Z in this case because of two consecutive vert. connections and Condition 4 of Definition 3.5). Hence for $\mathcal{P}$: $w \not\perp\!\!\!\perp x|\emptyset[\mathcal{P}]$, $w \not\perp\!\!\!\perp x|u[\mathcal{P}]$, $u \not\perp\!\!\!\perp w|y[\mathcal{P}]$ (and many others).

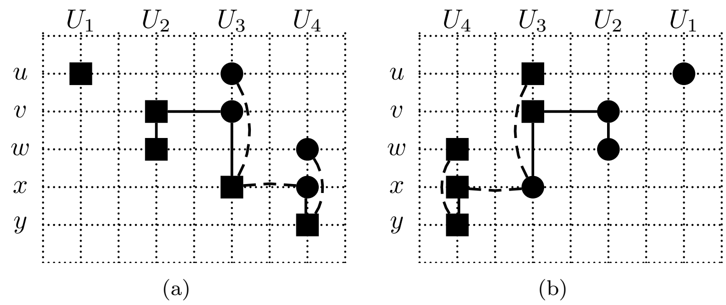


Figure 2: Two different sequence of markers in a structure and its reordering

On the contrary, τ_1 cannot form a Z-avoiding trail in $\mathcal{P}'$ because its vertical connection in U_2 is not regular - required by Condition 2. of Definition 3.5. However, τ_2 represents a Z-avoiding trail under the same assumptions as in $\mathcal{P}$. Moreover, observe that $[U_4, y]$ can be excluded from τ_2 in this case because vertical connection $[U_4, w], [U_4, x]$ is regular. I.e. τ_2 is not the shortest possible trail (by number of markers) representing $u \not\perp\!\!\!\perp v|Z[\mathcal{P}']$.

3.3. Structural independencies

As it was stated in the previous subsection, Z-avoiding trails can be used to read structural independencies induced by respective structure.

Definition 3.7. *By* structural independencies *induced by structure $\mathcal{P}$ we understand a system of disjoint triples $U, V, Z \subset \bigcup_{\mathcal{P}}$ such that $U, V \neq \emptyset$ and no $u \in U$ is connected with any $v \in V$ by a Z-avoiding trail in the structure (in symbol $U \perp\!\!\!\perp V|Z[\mathcal{P}]$). We say that sets of variables U and V are* conditionally independent given Z in $\mathcal{P}$.

Remark 3.8. *Considering the previous definition, realize that for any structure $\mathcal{P}$ its induced system of structural independencies is uniquely determined by the set of "elementary relations" $u \not\perp\!\!\!\perp v|Z[\mathcal{P}]$ representing the existence of a Z-avoiding trail in $\mathcal{P}$ connecting u with v, where u and v are singletons.*

4. Equivalence problem

Structure - as a tool for representing CI statements (the so-called structural independencies) - is imperfect. Two or more different structures can induce the same set of structural independencies and we say that they are *equivalent* in that case.

Remark 4.1. *Another definition of the equivalence can sound like this: "Two compositional model structures are said to be* equivalent *if the set of probability distributions that can be represented by compositional models with one of those structures is identical to the set of distributions that can be represented by models with the other one."*

In the prequel of this paper - in [10] - we described several structure properties invariant within a class of equivalence.

4.1. Invariants

One of the most important features introduced in [10] is the concept of the so-called *non-trivial sets*. We have already defined what is meant by a trivial and non-trivial column of a structure. (Recall that column $K_i^{\mathcal{P}}$ is non-trivial in $\mathcal{P}$ if $R_i^{\mathcal{P}} \neq \emptyset$.) Non-trivial set is its generalization for an arbitrary set of variables - not only for columns of the respective structure.

4.1.1. Non-trivial set

We say that a *set* U *is non-trivial* in $\mathcal{P}$ if $\exists K_i^{\mathcal{P}}$ such that $U \subseteq K_i^{\mathcal{P}}$ and $U \cap R_i^{\mathcal{P}} \neq \emptyset$. Otherwise, U is *trivial* in $\mathcal{P}$. To read more about non-trivial sets - see [10]. Note that every non-trivial column is a non-trivial set in respective structure simultaneously. The collection of all non-trivial sets in a structure $\mathcal{P}$ is denoted by $\mathcal{N}(\mathcal{P})$ and it has been proven in [10] that it is invariant within a class of equivalence:

Assertion 4.2. *If structures $\mathcal{P}$ and $\mathcal{P}'$ are equivalent then $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$*

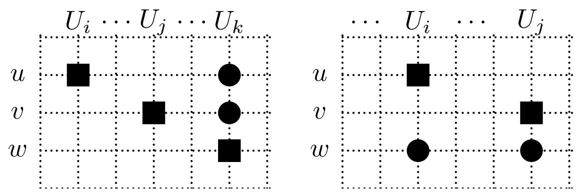
In this paper we also add the opposite implication:

Lemma 4.3. *If $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$ then $\mathcal{P}$ and $\mathcal{P}'$ are equivalent.*

To prove the lemma, it is enough to realize the close relation between non-trivial sets of cardinality 2 and 3 and Z-avoiding trails.

Observation 4.4. *In the language of persegrams, a set is non-trivial if there exists a column containing all respective markers and at least of of them is a box-marker. Hence, every regular vertical connection corresponds to a non-trivial set of cardinality 2 and vice-versa. I.e. $\{u, v\} \in \mathcal{N}(\mathcal{P}) \Leftrightarrow u \not\perp\!\!\!\perp v | Z[\mathcal{P}]$ for all $Z \subseteq \bigcup_{\mathcal{P}} \setminus \{u, v\}$*

Observation 4.5. *Assume $\{u, w\}, \{v, w\} \in \mathcal{N}(\mathcal{P})$ and $\{u, v\} \notin \mathcal{N}(\mathcal{P})$. There are basically two types of structures satisfying these assumptions. See Figure 3 for a simplified situation where columns and rows out of focus are omitted. Then*



(a) $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$ (b) $\{u, v, w\} \notin \mathcal{N}(\mathcal{P})$

Figure 3: Illustration of Observation 4.5

- $\{u, w, v\} \in \mathcal{N}(\mathcal{P}) \Leftrightarrow u \not\perp\!\!\!\perp v | Z[\mathcal{P}]$ *for all* $Z \subseteq \bigcup_{\mathcal{P}} \setminus \{u, v\}$ *such that* $w \in Z$. *I.e. the trail can be composed from two regular vertical connections in direct succession with common box-marker in* w. *(Figure 3a)*
- $\{u, w, v\} \notin \mathcal{N}(\mathcal{P}) \Leftrightarrow u \not\perp\!\!\!\perp v | Z[\mathcal{P}]$ *for all* $Z \subseteq \bigcup_{\mathcal{P}} \setminus \{u, v\}$ *such that* $w \notin Z$. *I.e. the trail can be composed from two regular vertical connections in different columns with a horizontal connection in* w. *(Figure 3b)*

Proof. (*Lemma 4.3*) The proof of this lemma is based on inductive step showing that every Z-avoiding trail in one structure has a corresponding Z-avoiding trail in the other one (Remark 3.8). Suppose $u \not\perp\!\!\!\perp v | Z[\mathcal{P}]$. Without loss of generality, choose a trail τ representing $u \not\perp\!\!\!\perp v | Z[\mathcal{P}]$ which involves the minimal number of vertical connections. Denote the number of involved vertical connections as $nv(\tau)$.

It is evident for τ with $nv(\tau) = 1$ by Observation 4.4 and assumption of $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$. In case of τ with $nv(\tau) = 2$ we have three variables involved: u, v and e.g. w. The choice of τ implies that u cannot be connected with v directly - there is no shorter trail than τ and therefore $\{u, v\} \notin \mathcal{N}(\mathcal{P})$ by Observation 4.4. On the other hand $\{u, w\}, \{v, w\} \in \mathcal{N}(\mathcal{P})$ from the same reason. Then this is a simple consequence of Observation 4.5 and assumption of $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$.

Assume τ with $nv(\tau) = m \geq 3$ and that the implication holds for trails τ' with $nv(\tau') < m$. Note that τ involves $m + 1$ variables. Denote them $u_0, u_1, \ldots, u_m$ where $u_0 = u$ and $u_m = v$ ordered with respect to τ. Then we can split τ into two parts $u \not\perp\!\!\!\perp u_k | Z[\mathcal{P}]$ and $u_k \not\perp\!\!\!\perp v | Z[\mathcal{P}]$ in a way $u_k \notin Z$ (this is always possible since there can be at most two vertical connections in a direct succession by the definition, $m \geq 3$, and no horizontal connection can correspond to a variable from Z). Both these trails also exist in $\mathcal{P}'$ by induction hypothesis. Obviously, $\{u_{k-1}, u_k\}, \{u_k, u_{k+1}\} \in \mathcal{N}(\mathcal{P})$ and $\{u_{k-1}, u_{k+1}\} \notin \mathcal{N}(\mathcal{P})$ by the choice of τ. Above that $\{u_{k-1}, u_k, u_{k+1}\} \notin \mathcal{N}(\mathcal{P})$ by Observation 4.5. Since $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$ then, using Observation 4.5, both trails end in u_k in different columns and one can connect them by horizontal connection in u_k and create $u \not\perp\!\!\!\perp v | Z[\mathcal{P}']$. Since the role of $\mathcal{P}$ and $\mathcal{P}'$ is interchangeable, the proof is done. □

Remark 4.6. *It became apparent that there is a close connection between the system of non-trivial sets and* characteristic imsets *introduced by Studený, Hemmecke, and Lindner [13] as a unique algebraic representative of a Bayesian network structure. Characteristic imset is a 0-1 vector indexed by subsets of the set of variables. It appears that in the case of Bayesian network structure – acyclic directed graph – inducing the same independence model as a given CM structure, the characteristic imset of the acyclic directed graph takes value 1 on components corresponding to non-trivial sets only in the respective CM structure.*

Note that the basic idea for introducing such an algebraic representative lies in possibility of using classical linear programming methods for learning the Bayesian network structure. For example, we refer to [13] for their solution in the case of undirected forests.

To brighten the connection, the alternative definition of characteristic imset c_G *of graph* G *states:* $c_G(U) = 1$ *iff there exists* $u \in U$ *with* $U \setminus \{u\} \subseteq pa_G(u)$ *where* $pa_G(u)$ *denotes the parent set of node* u*;* $pa_G(u) = \{v \in N | v \to u\}$ *(Theorem 1 in [13]). On the other hand, it had been shown that Bayesian networks and CM represent the same class of probability distributions and, above that, a conversion algorithm from CM to Bayesian network and vice-versa was published in [4]. In the language of this text, every non-trivial set of cardinality two (recall that in case of a persegram it is a pair of markers in one column where at least one of them is a box-marker) represents an arrow whose orientation is in the direction to the box-marker. Hence, having a box-marker, parent set of the respective variable is represented by all bullets and some box-markers in the column. Hence, every non-trivial set corresponds to a subset of a parent set of a variable.*

The number of sets to be possibly non-trivial is exponential in the number of variables. That is why we do not consider collection of induced non-trivial sets as a handy tool for practical use. Here, we will derive another closely related feature - the so-called *formal ratio*.

4.1.2. Formal ratio

Another test of equivalence related to non-trivial sets is the so-called *formal ratio*. One writes a formal ratio $\mathcal{F}(\mathcal{P})$ for every structure $\mathcal{P}$ as follows: in the numerator one lists all sets $K_i^{\mathcal{P}}$ while in the denominator one lists all sets $S_i^{\mathcal{P}}$ for $i = 1, \ldots, n$. Then cancelation is performed: one occurrence of a set $U \subseteq \bigcup_{\mathcal{P}}$ in the denominator is canceled against one occurrence of U in the numerator. For example the structure $\mathcal{P}'$ from Example 2.1 induces the following "ratio" (See Table 1):

$$\mathcal{P}' : \frac{\{w, x, y\} * \{u, v, x\} * \{v, w\} * \{u\}}{\emptyset * \{x\} * \{v, w\} * \{u\}}.$$

Then, after cancelation, its formal ratio is

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

Lemma 4.7. *Two structures $\mathcal{P}_1, \mathcal{P}_2$ are equivalent iff they lead to the same formal ratio $\mathcal{F}(\mathcal{P}_1) = \mathcal{F}(\mathcal{P}_2)$.*

To prove the lemma we will employ the fact that system of non-trivial sets characterizes equivalence (Assertion 4.2, Lemma 4.3) and we introduce affine transformation of a system of non-trivial sets into formal ratio.

Consider set U. When checking $U \in \mathcal{N}(\mathcal{P})$ then, by definition of non-trivial set, we are looking for columns $K_i^{\mathcal{P}}$ such that $U \subseteq K_i^{\mathcal{P}}$ and simultaneously $U \cap R_i^{\mathcal{P}} \neq \emptyset$. If there are several columns in $\mathcal{P}$ containing U then, obviously, U has non-empty intersection with respective R part at most once. It is the first column containing U. Indeed, variables U were already introduced for successive columns and therefore they are in their S-parts. Hence, for function $c_{\mathcal{P}} : 2^{\bigcup_{\mathcal{P}}} \to \{0, 1\}$ such that $c_{\mathcal{P}}(U) = |\{K_i^{\mathcal{P}} : U \subseteq K_i^{\mathcal{P}}\}| - |\{S_j^{\mathcal{P}} : U \subseteq S_j^{\mathcal{P}}\}|$ holds

$$U \in \mathcal{N}(\mathcal{P}) \iff c_{\mathcal{P}}(U) = 1. \tag{4.1}$$

Because of the fact that equal sets $K_i^{\mathcal{P}} = S_j^{\mathcal{P}}$ are canceled in the formal ratio, the result of function $c_{\mathcal{P}}$ would be the same if one uses sets from numerator of respective formal ratio instead of sets $K_i^{\mathcal{P}}$ and sets from its denominator instead of sets $S_j^{\mathcal{P}}$ in formula (4.1). I.e. the same characterization of non-trivial sets can be got using function $u_{\mathcal{P}} : 2^{\bigcup_{\mathcal{P}}} \to \{\ldots, -2, -1, 0, 1\}$ based on respective formal ratio $\mathcal{F}(\mathcal{P})$. Its value for set U is got as the number of occurences of set U in the numerator minus the number of its occurrence in the denominator. Therefore

$$c_{\mathcal{P}}(U) = \sum_{V, V \supseteq U} u_{\mathcal{P}}(V) \tag{4.2}$$

and the formula (4.1) can be reformulated as

$$U \in \mathcal{N}(\mathcal{P}) \Leftrightarrow c_{\mathcal{P}}(U) = \sum_{V, V \supseteq U} u_{\mathcal{P}}(V) = 1. \tag{4.3}$$

Note that the function $u_{\mathcal{P}}$ uniquely characterizes respective formal ratio $\mathcal{F}(\mathcal{P})$. Indeed, one can easily reconstruct $\mathcal{F}(\mathcal{P})$ using the following process: If $u_{\mathcal{P}}(U) = 0$ do nothing, if $u_{\mathcal{P}}(U) = 1$ put U into the numerator of $\mathcal{F}(\mathcal{P})$, and if $u_{\mathcal{P}}(U) = -k$ put U k-times into its denominator.

Proof. (*Lemma 4.7*) $\Leftarrow$: For two structures $\mathcal{P}_1$ and $\mathcal{P}_2$ with identical formal ratio evidently $u_{\mathcal{P}_1}(U) = u_{\mathcal{P}_2}(U)$ for all $U \subseteq \bigcup_{\mathcal{P}_1} = \bigcup_{\mathcal{P}_2}$. Thus, by (4.3), $\mathcal{N}(\mathcal{P}_1) = \mathcal{N}(\mathcal{P}_2)$ which finishes this part of the proof by Lemma 4.3.

$\Rightarrow$ Assuming the equivalence, $\mathcal{N}(\mathcal{P}_1) = \mathcal{N}(\mathcal{P}_2)$ by Assertion 4.2 and we will show $u_{\mathcal{P}_1} = u_{\mathcal{P}_2}$. This will conclude the proof since respective formal ratios are uniquely reconstructible from these functions. Note that $\bigcup_{\mathcal{P}_1} = \bigcup_{\mathcal{P}_2}$. Put $W = \bigcup_{\mathcal{P}_1}$.

Using (4.3), one can rewrite lemma assumption $\mathcal{N}(\mathcal{P}_1) = \mathcal{N}(\mathcal{P}_2)$ into

$$c_{\mathcal{P}_1}(U) = c_{\mathcal{P}_2}(U) \textit{ for all } U \subseteq W \tag{4.4}$$

Let us also prove that functions $u_{\mathcal{P}_1}$ and $u_{\mathcal{P}_1}$ coincide using a proof by induction on $|U|$. For $U = W$: $u_{\mathcal{P}_1}(W) = c_{\mathcal{P}_1}(W) = c_{\mathcal{P}_2}(W) = u_{\mathcal{P}_2}(W)$ by (4.4) (there is no superset of W in W in the sum of (4.3)).

Now assume that the induction holds for all $U \subseteq W$ such that $|U| \geq |W| - k$. For $U \subset W$ such that $|U| = |W| - k - 1$ holds by induction hypothesis

$$c_{\mathcal{P}_1}(U) = \sum_{V, V \supseteq U} u_{\mathcal{P}_1}(V) = u_{\mathcal{P}_1}(U) + \sum_{V, V \supset U} u_{\mathcal{P}_1}(V) = u_{\mathcal{P}_1}(U) + \sum_{V, V \supset U} u_{\mathcal{P}_2}(U).$$

Then $u_{\mathcal{P}_1}(U) = u_{\mathcal{P}_2}(U)$ by (4.4) which finishes the proof. □

Example 4.8. *Assume a simple structure $\mathcal{P} = (\{u, v\}, \{v, w\}, \{u, v, x\})$ and its permutation $\mathcal{P}' = (\{u, v, x\}, \{v, w\}, \{u, v\})$. Their 'ratios' are*

$$\mathcal{P} : \frac{\{u, v\} * \{v, w\} * \{u, v, x\}}{\emptyset * \{v\} * \{u, v\}}, \ \mathcal{P}' : \frac{\{u, v, x\} * \{v, w\} * \{u, v\}}{\emptyset * \{v\} * \{u, v\}}.$$

The structures are equivalent since their formal ratios coincide:

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

4.2. Elementary operations

The fact that formal ratio characterizes equivalence gives us a specific notion how two equivalent structures look like. They both have to contain columns corresponding to numerator of respective formal ratio. Moreover, columns are in an ordering induced by S-sets listed in the denominator. Considering Example 2.1 and Table 1, realize that respective structures $\mathcal{P}$ and $\mathcal{P}'$ are not equivalent.

Equivalence problem (as formulated in the Introduction) also includes the subproblem of an easy way to get from $\mathcal{P}$ to an equivalent $\mathcal{P}'$ in terms of some elementary operations invariant the the equivalence. Considering formal ratio, two types of operations can be considered only: (i) changing the structure ordering (permutation) in a way that the set of induced S-parts remains the same, or (ii) adding/removing columns corresponding to "cancelation" from formal ratio creation process.

Now we will derive two special transpositions. We will restrict ourselves to transpositions of successive columns only. The reason for such a restriction is simple and, for case of referencing, it is summarized in the following remark:

Remark 4.9. *Recall that both R and S-parts are defined by respective column and the union of previous ones in the structure. Imagine a transposition of two successive columns, e.g. $(k–1\ k)$. Then these parameters are the same for all other columns in both the original and the permutated structure because the union operator is commutative. Hence, a transposition of two adjacent columns affects R and S-parts of those columns only.*

4.2.1. Constant transposition

Considering the previous remark, we will distinguish two transpositions. The first one is the so-called *constant transposition* It is defined in a way that S-parts of affected column do not change. Its idea is the following: S-part of a column represents variables already introduced in the previous columns of the structure. Thus, when switching two adjacent columns, it is enough to guarantee that variables from S-part of the latter one are not introduced in the former one. I.e. in case of transposition $(k-1\ k)$ to guarantee that $R^{\mathcal{P}}_{k-1} \cap S^{\mathcal{P}}_k = \emptyset$.

Definition 4.10. *For a structure $\mathcal{P}$ of length $n \geq 2$ and $k \in \{2, \dots, n\}$ a transposition $\sigma = (k–1\ k) = (k\ k–1)$ is said to be* constant *in $\mathcal{P}$ if $R^{\mathcal{P}}_{k-1} \cap S^{\mathcal{P}}_k = \emptyset$. We say that $\mathcal{P}\sigma$ is a* constant transposition *of $\mathcal{P}$.*

Recall the structure $\mathcal{P}$ from Example 2.1. Note that the transposition (1 2) is constant in $\mathcal{P}$ - see Table 1.

Lemma 4.11. *Consider a structure $\mathcal{P}$ and a transposition σ which is* constant *in $\mathcal{P}$. Then $R^{\mathcal{P}}_i = R^{\mathcal{P}\sigma}_{i\sigma}$ and $S^{\mathcal{P}}_i = S^{\mathcal{P}\sigma}_{i\sigma}$ for all $i \in \{1, \dots, n\}$.*

Proof. Let $\sigma = (k{-}1\ k)$. Since $K_i^{\mathcal{P}} = K_{i\sigma}^{\mathcal{P}\sigma}$ by definition of permutation and $S_i = K_i \setminus R_i$, it is enough to prove that $R_i^{\mathcal{P}} = R_{i\sigma}^{\mathcal{P}\sigma}$ but only for indices $i = k{-}1, k$ by Remark 4.9.

For index $k-1$:

$$\begin{aligned} R_{(k-1)\sigma}^{\mathcal{P}\sigma} &= R_k^{\mathcal{P}\sigma} \\ &= K_k^{\mathcal{P}\sigma} \setminus (K_1^{\mathcal{P}\sigma} \cup \ldots \cup K_{k-2}^{\mathcal{P}\sigma} \cup K_{k-1}^{\mathcal{P}\sigma}) \\ &= K_{k-1}^{\mathcal{P}} \setminus (K_1^{\mathcal{P}} \cup \ldots \cup K_{k-2}^{\mathcal{P}} \cup K_k^{\mathcal{P}}) \\ &= R_{k-1}^{\mathcal{P}} \setminus K_k^{\mathcal{P}} = R_{k-1}^{\mathcal{P}} \end{aligned}$$

where the first and second equations are given by definition of σ and $R_k^{\mathcal{P}\sigma}$, respectively. The third equation is given by the way how any permutation σ works, while the last equation is guaranteed by the fact that σ is a constant transposition in $\mathcal{P}$: $R_{k-1}^{\mathcal{P}} \cap K_k^{\mathcal{P}} = R_{k-1}^{\mathcal{P}} \cap S_k^{\mathcal{P}} = \emptyset$.

Considering index k: $R_{k\sigma}^{\mathcal{P}\sigma} = R_k^{\mathcal{P}}$ is a direct consequence of the fact that $\{R_i^{\mathcal{P}\sigma}\}_{i=1,\ldots,n}$ is a disjoint partition of $\bigcup_{\mathcal{P}} = \bigcup_{\mathcal{P}'}$. □

Note that it is the previous lemma that states that S-parts of affected columns do not change. It is a simple consequence of the fact that $K_i^{\mathcal{P}} = K_{i\sigma}^{\mathcal{P}\sigma}$ for every column and every permutation σ.

Remark 4.12. *Notice that if σ is a constant transposition in $\mathcal{P}$, then σ is a constant transposition in $\mathcal{P}\sigma$ as well. Indeed, considering $\sigma = (k{-}1\ k)$, then $R_{k-1}^{\mathcal{P}\sigma} \cap S_k^{\mathcal{P}\sigma} = R_k^{\mathcal{P}} \cap S_{k-1}^{\mathcal{P}} = \emptyset$ by Lemma 4.11 and definition of $R_k^{\mathcal{P}}$. This also means that the role of $\mathcal{P}$ and $\mathcal{P}\sigma$ is interchangeable in a way that if $\mathcal{P}' = \mathcal{P}\sigma$ then $\mathcal{P} = \mathcal{P}'\sigma$.*

To simplify some of the following proofs, we introduce two special generalizations of constant transposition - cycles - that can be easily replaced by a composition of constant transpositions.

Let $\mathcal{P}$ be a structure of length$n \geq 2$ and $i \in \{1, \ldots, n{-}1\}$, $k \in \{1, \ldots, n{-}i\}$. If $R_i^{\mathcal{P}} \cap (S_{i+1}^{\mathcal{P}} \cup \ldots \cup S_{i+k}^{\mathcal{P}}) = \emptyset$ then $\sigma_1 = (i\ i{+}1)$ is constant transposition in $\mathcal{P}$. Similarly, since $R_{i+1}^{\mathcal{P}\sigma_1} = R_i^{\mathcal{P}}$ by Lemma 4.11, then $(i{+}1\ i{+}2)$ is constant transposition in $\mathcal{P}\sigma_1$, etc. Hence, for $\sigma = (i{+}k\ \ i{+}k{-}1\ \ \ldots\ i) = (i\ i{+}1)(i{+}1\ i{+}2)\ldots(i{+}k{-}1\ i{+}k)$ holds that $\mathcal{P}\sigma$ may be obtained from $\mathcal{P}$ by a sequence of constant transpositions.

Corollary 4.13. *If for $\mathcal{P}$ holds that $R_i^{\mathcal{P}} \cap (S_{i+1}^{\mathcal{P}} \cup \ldots \cup S_{i+k}^{\mathcal{P}}) = \emptyset$ then $\mathcal{P}\sigma$, where $\sigma = (i{+}k\ i{+}k{-}1\ \ldots\ i)$, may be obtained from $\mathcal{P}$ by iterative application of constant transposition.*

Similarly, consider a structure $\mathcal{P}$ of length $n \geq 3$, $i \in \{1, \ldots, n-2\}$, $k \in \{2, \ldots, n-i\}$ such that $K_i^{\mathcal{P}} \supseteq S_{i+k}^{\mathcal{P}}$. Observe that $\forall j$ such that $i < j < i{+}k$ holds that: $R_j^{\mathcal{P}} \cap S_{i+k}^{\mathcal{P}} = \emptyset$. Then, similarly, permutation $\sigma_L = (i{+}1\ i{+}2\ \ldots\ i{+}k) = (i{+}k\ i{+}k{-}1)(i{+}k{-}1\ i{+}k{-}2)\ldots(i{+}2\ i{+}1)$ may be replaced by a sequence of constant transpositions in $\mathcal{P}$ by iterative application of Lemma 4.11, and one can conclude:

Corollary 4.14. *If for $\mathcal{P}$ holds that $K_i^{\mathcal{P}} \supseteq S_{i+k}^{\mathcal{P}}$ then $\mathcal{P}\sigma$, where $\sigma = (i{+}1\ i{+}2\ \ldots\ i{+}k)$, may be obtained from $\mathcal{P}$ by iterative application of constant transposition.*

4.2.2. Box transposition

The second transposition is the so-called *box transposition.* It is designed in a way that involved columns exchange their S-parts during the transposition. I.e. considering the following Lemma 4.16, while two adjacent columns switch their position, respective S-parts seems to hold their position.

Definition 4.15. *For $\mathcal{P}$ with length $n \geq 2$ and $k \in \{2, \ldots, n\}$ we call transposition $\sigma = (k{-}1\ k) = (k\ k{-}1)$ a* box transposition *in $\mathcal{P}$ if $S_k^{\mathcal{P}} \setminus R_{k-1}^{\mathcal{P}} = S_{k-1}^{\mathcal{P}}$. We say that $\mathcal{P}\sigma$ is* box transposition *of $\mathcal{P}$.*

Recall the structure $\mathcal{P}'$ from Example 2.1. Note that (1 2) is a box transposition in $\mathcal{P}'$. Check Table 1.

Lemma 4.16. *If σ is a box transposition in $\mathcal{P}$ then $S_i^{\mathcal{P}} = S_i^{\mathcal{P}\sigma}$ for all $i = 1, \ldots, n$.*

Proof. Suppose $\sigma = (k{-}1\ k)$. It will be enough to show the lemma for $i = k{-}1, k$ by Remark 4.9. Using the definition of box transposition, it holds for index $k{-}1$:

$$S^{\mathcal{P}\sigma}_{k-1} = K^{\mathcal{P}\sigma}_{k-1} \cap (K^{\mathcal{P}\sigma}_1 \cup \ldots \cup K^{\mathcal{P}\sigma}_{k-2}) = K^{\mathcal{P}}_k \cap (K^{\mathcal{P}}_1 \cup \ldots \cup K^{\mathcal{P}}_{k-2}) = S^{\mathcal{P}}_k \setminus R^{\mathcal{P}}_{k-1} = S^{\mathcal{P}}_{k-1}.$$

To prove the same for index k, one has to realize that by definition of box transposition

$$S^{\mathcal{P}}_{k-1} \subseteq S^{\mathcal{P}}_k \text{ and } S^{\mathcal{P}}_k \subseteq R^{\mathcal{P}}_{k-1} \cup S^{\mathcal{P}}_{k-1} = K^{\mathcal{P}}_{k-1}. \tag{4.5}$$

Then, employing distributive law for set operations and (4.5), one can finish the proof with

$$\begin{aligned} S^{\mathcal{P}\sigma}_k &= K^{\mathcal{P}}_{k-1} \cap (K^{\mathcal{P}}_1 \cup \ldots \cup K^{\mathcal{P}}_{k-2} \cup K^{\mathcal{P}}_k) \\ &= S^{\mathcal{P}}_{k-1} \cup (K^{\mathcal{P}}_{k-1} \cap K^{\mathcal{P}}_k) = S^{\mathcal{P}}_{k-1} \cup (K^{\mathcal{P}}_{k-1} \cap S^{\mathcal{P}}_k) \\ &= K^{\mathcal{P}}_{k-1} \cap S^{\mathcal{P}}_k = S^{\mathcal{P}}_k. \end{aligned}$$

□

In case of a persegram sets $R^{\mathcal{P}}_i$ and $S^{\mathcal{P}}_i$ correspond to box-markers and bullets in corresponding column, respectively. Considering previous lemma, S-parts (and corresponding sets of bullets) seems to keep their positions during box transposition and box-markers are the only thing that seems to move. If the reader realizes this link, the adjective "box" (in "box transposition") makes sense.

Remark 4.17. *Observe that if $\sigma = (k{-}1\ k)$ is a box transposition in $\mathcal{P}$, then it is a box transposition in $\mathcal{P}\sigma$ as well. To prove this, check the validity of $S^{\mathcal{P}\sigma}_k \setminus R^{\mathcal{P}\sigma}_{k-1} = S^{\mathcal{P}\sigma}_{k-1}$ which can be rewritten into $S^{\mathcal{P}}_k \setminus (K^{\mathcal{P}}_k \setminus S^{\mathcal{P}}_{k-1}) = S^{\mathcal{P}}_{k-1}$ using Lemma 4.16. Since $U \setminus (V \setminus W) = (U \setminus V) \cup (U \cap W)$ for arbitrary sets U, V, W, the equation is guaranteed by $S^{\mathcal{P}}_{k-1} \subseteq S^{\mathcal{P}}_k$ following from the fact that σ is box transposition in $\mathcal{P}$ - recall (4.5). Moreover, since any transposition is its own inversion then $\mathcal{P}\sigma\sigma = \mathcal{P}$. Hence the roles of $\mathcal{P}$ and $\mathcal{P}\sigma$ are interchangeable with respect to σ. ($\mathcal{P}' = \mathcal{P}\sigma$ if and only if $\mathcal{P} = \mathcal{P}'\sigma$)*

4.2.3. Reduction/extension

Assume a structure $\mathcal{P}$. Recall its formal ratio $\mathcal{F}(\mathcal{P})$ creation process. One lists sets $K^{\mathcal{P}}_i$ in the numerator and sets $S^{\mathcal{P}}_i$ in the denominator for all $i \in \{1 \ldots, n\}$. Then cancelation is performed: one occurrence of a set U in the denominator is canceled against one occurrence of U in the numerator. One can say that the ratio has been reduced. If there is no "cancelation" in the process - we say that the structure is *reduced* as well.

Definition 4.18. *A structure $\mathcal{P}$ is said to be* reduced *if $\nexists i, j \in \{1, \ldots, n\}$ such that $K^{\mathcal{P}}_i = S^{\mathcal{P}}_j$.*

As an example or reduced structure recall the structure $\mathcal{P}$ from Example 2.1. Structures that are not reduced can be found in Example 4.8.

Remark 4.19. *An arbitrary reduced structure has several interesting and convenient properties:*

- *there is no trivial column in it (recall that a column $K^{\mathcal{P}}_i$ is trivial in $\mathcal{P}$ if $K^{\mathcal{P}}_i = S^{\mathcal{P}}_i$),*
- *reduced structure consists from columns listed in numerator of respective formal ratio exactly,*
- *if two reduced structures are equivalent, they are each other's permutation (their formal ratios coincide by Lemma 4.7).*

Assume a structure $\mathcal{P}$ that is not reduced. I.e. $\exists i, j \in \{1, \ldots, n\}$ such that $K^{\mathcal{P}}_i = S^{\mathcal{P}}_j$. If $i = j$ then the column $K^{\mathcal{P}}_i$ is trivial in $\mathcal{P}$ and it may be deleted. And this is the idea on which the following definition is based:

Definition 4.20. Simple reduction *means a change of a structure $\mathcal{P}$ into a structure $\mathcal{P}'$ by removing a trivial column.*

Simple extension *means a change of structure a $\mathcal{P}$ into a structure $\mathcal{P}''$ by adding a trivial column.*

Remark 4.21. *Note that formal ratio is invariant with respect to elementary reduction/extension and constant and box transpositions. Indeed, while this is guaranteed for constant and box transposition by Lemma 4.11 and 4.16, respectively; application of simple reduction/extension does not represent any threat due to the cancelation in formal ratio creation process.*

Lemma 4.22. *Every non-reduced structure can be transformed into its reduced and equivalent form using simple reduction, constant transposition, and box transposition.*

Proof. Assume a structure $\mathcal{P}$ that is not reduced. I.e. $\exists i, j \in \{1, \ldots, n\}$ such that $K_i^{\mathcal{P}} = S_j^{\mathcal{P}}$. If $i = j$ then the column $K_i^{\mathcal{P}}$ is trivial in $\mathcal{P}$ and it may be removed by simple reduction. If $i > j$ then, however, $K_i^{\mathcal{P}}$ must be trivial column as well since it is contained in previous column $K_j^{\mathcal{P}}$ and therefore $K_i^{\mathcal{P}} = S_i^{\mathcal{P}}$. This case has been already treated. If $i < j$ put $k = j - i$, i.e. $K_i^{\mathcal{P}} = S_{i+k}^{\mathcal{P}}$. Then we can use special permutation - k-cycle $\sigma_C = (i{+}1\ i{+}2\ \ldots\ i{+}k)$ - which can be replaced by a sequence of constant transpositions by Corollary 4.14. Considering the fact $(i + k)\sigma_C = i + 1$, it implies that $K_i^{\mathcal{P}\sigma_C} = K_i^{\mathcal{P}} = S_{i+k}^{\mathcal{P}} = S_{i+1}^{\mathcal{P}\sigma_C}$ by Lemma 4.11. Then, however, $\sigma_b = (i\ i{+}1)$ is a box transposition in $\mathcal{P}\sigma_C$ and $K_{i+1}^{\mathcal{P}\sigma_C\sigma_b} = K_i^{\mathcal{P}\sigma_C} = S_{i+1}^{\mathcal{P}\sigma_C} = S_{i+1}^{\mathcal{P}\sigma_C\sigma_b}$. Hence $K_{i+1}^{\mathcal{P}\sigma_C\sigma_b}$ is trivial and it can be removed using simple reduction. By iterative application of the above process one obtains a reduced structure which is equivalent with the given one by Remark 4.21. □

Note that every k-cycle $(i{+}1\ \ldots\ i{+}k)$ simply moves $i{+}k$-th column to the $i{+}1$-th position and shifts the other columns accordingly. Hence, we can consider every k-cycle as one operation, easily. Thus - following the previous proof - when reducing a structure, one needs at most 3 operations for every column causing that structure is not reduced. If the length of the reduced structure is m and the length of non-reduced is n, then we need at most $3(n - m)$ operations. Nevertheless, since one has to scan the whole structure to find a pair $K_i^{\mathcal{P}} = S_j^{\mathcal{P}}$ for every $i \in \{1, \ldots, n\}$, the reduction is polynomial in n.

5. Solution of the Equivalence problem

In this last section we present the complete solution of the so-called equivalence problem in all its parts described in the Introduction. We list characteristic properties of equivalence - i.e. properties necessary and sufficient to guarantee equivalence of given structures. Similarly, we show how to transform a given structure into an arbitrary equivalent one structure using elementary operations like constant transposition, box transposition, simple extension, and simple reduction. This process is described in the proof of the most important assertion of this paper - Theorem 5.3. First, let us state several auxiliary assertions.

Lemma 5.1. *Let $\mathcal{P}, \mathcal{P}'$ be two reduced structures such that $\mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}')$. Then one can transform $\mathcal{P}'$ to have the same last column as $\mathcal{P}$ (including its R and S-part) with the help of box and constant transpositions only.*

Proof. The assumption of $\mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}')$, together with the fact that $\mathcal{P}, \mathcal{P}'$ are reduced, implies that $\mathcal{P}$ and $\mathcal{P}'$ are of the same length n and $\mathcal{P}' = \mathcal{P}\sigma$ for some permutation σ by Remark 4.19. Then, similarly, $K_n^{\mathcal{P}} = K_{n\sigma}^{\mathcal{P}\sigma}$ and we will show how to permute $\mathcal{P}\sigma$ to move $K_{n\sigma}^{\mathcal{P}\sigma}$ to the n-th position. Recall that since both $\mathcal{P}$ and $\mathcal{P}\sigma$ are reduced, $R_n^{\mathcal{P}} \neq \emptyset$ and $K_n^{\mathcal{P}}, K_{n\sigma}^{\mathcal{P}\sigma}$ have to be the only columns containing variables from $R_n^{\mathcal{P}}$. Then

$$R_n^{\mathcal{P}} \cap S_i^{\mathcal{P}\sigma} = \emptyset \text{ for all } i \neq n\sigma. \tag{5.1}$$

This implies that $R_{n\sigma}^{\mathcal{P}\sigma} \supseteq R_n^{\mathcal{P}}$. We can distinguish two cases:

I. $R_{n\sigma}^{\mathcal{P}\sigma} = R_n^{\mathcal{P}}$

II. $R_{n\sigma}^{\mathcal{P}\sigma} \supset R_n^{\mathcal{P}}$

If $R_{n\sigma}^{\mathcal{P}\sigma} = R_n^{\mathcal{P}}$ then $\sigma_R = (n\ n{-}1\ n{-}2\ \ldots\ n\sigma)$ (which can be replaced by a sequence of constant transpositions by (5.1) and Corollary 4.13) moves $K_{n\sigma}^{\mathcal{P}\sigma}$ to the n-th position in $\mathcal{P}\sigma\sigma_R$. I.e. $K_n^{\mathcal{P}\sigma\sigma_R} = K_n^{\mathcal{P}}$ and $R_n^{\mathcal{P}\sigma\sigma_R} = R_n^{\mathcal{P}}$ by Lemma 4.11 which finishes the proof.

If $R^{\mathcal{P}\sigma}_{n\sigma} \supset R^{\mathcal{P}}_n$ then (using $K^{\mathcal{P}}_n = K^{\mathcal{P}\sigma}_{n\sigma}$)

$$S^{\mathcal{P}}_n \setminus R^{\mathcal{P}\sigma}_{n\sigma} = S^{\mathcal{P}\sigma}_{n\sigma}. \tag{5.2}$$

Since the formal ratios of $\mathcal{P}$ and $\mathcal{P}\sigma$ coincide, then their denominators coincide as well. Hence $\exists k > 0$ such that $S^{\mathcal{P}\sigma}_{n\sigma+k} = S^{\mathcal{P}}_n$. Then, $K^{\mathcal{P}\sigma}_{n\sigma} = K^{\mathcal{P}}_n \supset S^{\mathcal{P}}_n = S^{\mathcal{P}\sigma}_{n\sigma+k}$ and $\mathcal{P}\sigma\sigma_L$ can be, for $\sigma_L = (n\sigma{+}1 \;\; n\sigma{+}2 \;\ldots\; n\sigma{+}k)$, obtained from $\mathcal{P}\sigma$ by a sequence of constant transpositions by Corollary 4.14 - which, by iterative application of Lemma 4.11, implies that

$$S^{\mathcal{P}\sigma\sigma_L}_{n\sigma+1} = S^{\mathcal{P}}_n. \tag{5.3}$$

After applying this observation to (5.2), one can see that $\sigma_b = (n\sigma \;\; n\sigma{+}1)$ is a box transposition in $\mathcal{P}\sigma\sigma_L$ and $S^{\mathcal{P}\sigma\sigma_L\sigma_b}_{n\sigma+1} = S^{\mathcal{P}}_n$ by (5.3) and Lemma 4.16. Moreover, since $K^{\mathcal{P}\sigma\sigma_L\sigma_b}_{n\sigma+1} = K^{\mathcal{P}}_n$ by $n\sigma\sigma_L\sigma_b = n\sigma + 1$ then $R^{\mathcal{P}\sigma\sigma_L\sigma_b}_{n\sigma\sigma_L\sigma_b} = R^{\mathcal{P}}_n$. Then, the case I. occurs for the pair $\mathcal{P}$ and $\mathcal{P}\sigma\sigma_L\sigma_b$ which was already treated. This concludes the proof. □

To simplify the following, we introduce the concept of *substructure*. Consider structure $(U_1, \ldots, U_n)$. Then (U_1), (U_1, U_2), $(U_1, U_2, U_3), \ldots, (U_1, \ldots, U_k)$ are its substructures for $k < n$. In other words, substructure of a structure is its left part containing first k columns.

Remark 5.2. *Each of elementary reduction/extension, constant transposition, and box transposition is defined with respect to previous columns. Hence, if an operation is elementary reduction/extension, constant transposition, or box transposition in a substructure, then it is elementary reduction/extension, constant transposition, or box transposition, in corresponding structure, respectively.*

Theorem 5.3. *Supposing $\mathcal{P}_A$ and $\mathcal{P}_B$ are two structures, the following four conditions are mutually equivalent:*

(1) $\mathcal{P}_A$ and $\mathcal{P}_B$ are equivalent

(2) $\mathcal{N}(\mathcal{P}_A) = \mathcal{N}(\mathcal{P}_B)$

(3) $\mathcal{F}(\mathcal{P}_A) = \mathcal{F}(\mathcal{P}_B)$

(4) there exists a sequence $\mathcal{P}_1, \ldots, \mathcal{P}_m$, $m \geq 1$ of structures over N such that $\mathcal{P}_1 = \mathcal{P}_A, \mathcal{P}_m = \mathcal{P}_B$ and $\mathcal{P}_{i+1}$ is obtained from $\mathcal{P}_i$ using one of the elementary reduction/extension, constant transposition, and box transposition for $i = 1, \ldots, (m{-}1)$.

Proof. We show (1) ⇔ (2) and (1) ⇔ (3) ⇔ (4). Note that the (1) ⇔ (2) is stated in Assertion 4.2 and Lemma 4.3. Similarly, (1) ⇔ (3) is in Lemma 4.7.

To prove (3) ⇒ (4), first assume that both $\mathcal{P}_A$ and $\mathcal{P}_B$ are reduced. This, combined with (3) guarantees that $\mathcal{P}_A$ and $\mathcal{P}_B$ are of the same length n and we prove the implication (3) ⇒ (4) by induction on n. The induction statement for $n \geq 1$ is that (3) ⇒ (4) holds for any pair of structures $\mathcal{P}_A, \mathcal{P}_B$ of length $m \leq n$. The implication is evident for $n = 1$. Assume $n \geq 2$.

Observe that (3) altogether with our assumption of reduced structures implies the existence of a sequence of structures $\mathcal{P}_A = \mathcal{P}_1, \ldots, \mathcal{P}_k$ such that $K^{\mathcal{P}_k}_n = K^{\mathcal{P}_B}_n$, $S^{\mathcal{P}_k}_n = S^{\mathcal{P}_B}_n$ and $\mathcal{P}_{i+1}$ is obtained from $\mathcal{P}_i$ using one of constant or box transposition for $i = 1, \ldots, (k{-}1)$ according to Lemma 5.1. Then introduce $\mathcal{P}'_k$ and $\mathcal{P}'_B$ as the substructures of $\mathcal{P}_k$ and $\mathcal{P}_B$, respectively, by removing their last column (which is the same including its R and S-part). Note that $\mathcal{F}({\mathcal{P}'}_k) = \mathcal{F}({\mathcal{P}'}_B)$ by Remark 4.21 and the definition of substructures. By the induction hypothesis, there exists a desired sequence of $\mathcal{P}'_k, \ldots, \mathcal{P}'_{k+m} = \mathcal{P}'_B$ where $m \geq 1$. Introduce $\mathcal{P}_{k+i}$ as a structure obtained from $\mathcal{P}'_{k+i}$ by adding a column $K^{\mathcal{P}_B}_n$ at the last position for $i = 1, \ldots, m$. Using Remark 5.2, $\mathcal{P}_{k+i+1}$ is obtained from $\mathcal{P}_{k+i}$ using one of elementary reduction/extension, constant transposition, and box transposition for $i = 1, \ldots, (m{-}1)$. Hence $\mathcal{P}_{k+m} = \mathcal{P}_B$ which concludes the induction step.

If $\mathcal{P}_A$ or $\mathcal{P}_B$ is not reduced, then one may easily create sequences of structures $\mathcal{P}_A = \mathcal{P}_{A_1}, \ldots, \mathcal{P}_{A_k}$ and $\mathcal{P}_B = \mathcal{P}_{B_1}, \ldots, \mathcal{P}_{B_l}$ where both $\mathcal{P}_{A_k}$ and $\mathcal{P}_{B_l}$ are reduced and $\mathcal{P}_{A_{i+1}}$ or $\mathcal{P}_{B_{j+1}}$ is obtained from $\mathcal{P}_{A_i}$ or $\mathcal{P}_{B_j}$, respectively, by an elementary reduction/extension, constant transposition, or box transposition for

$i = 1, \ldots, (k-1)$ and $j = 1, \ldots, (l-1)$ by Lemma 4.22. Since formal ratios of both $\mathcal{P}_{A_k}, \mathcal{P}_{B_l}$ coincide by Remark 4.21, the case where both structures are reduced occurs for the pair $(\mathcal{P}_{A_k}, \mathcal{P}_{B_l})$ which was already treated.

The proof of (4) $\Rightarrow$ (3) follows from Remark 4.21 □

Note that the above proof is constructive when transforming one structure into another equivalent one. We need at most $3n$ operations for the transformation when having two reduced structures of length n (let us consider k-cycle as one operation) - we need at most two k-cycles and one box transposition for every column. Considering non-reduced structures of length m_1 and m_2 that have length n in their reduced case, then we need at most $3(m_1 - n) + 3(m_2 - n) + 3n$ operations. However, the complexity of the algorithm (when including resources for finding identical columns and columns with the same S-part) is polynomial.

Conclusions

This paper deals with the so-called *equivalence problem* whose solution occupies a large part of this text. The equivalence problem is understood as a problem of how to recognize whether two given structures $\mathcal{P}$ and $\mathcal{P}'$ over the same set of variables induce the same set of structural independencies - i.e. whether they can represent the same class of probability distributions. It is of special importance to have a simple rule to recognize that two structures are equivalent in this sense, and an easy way to convert $\mathcal{P}$ into $\mathcal{P}'$ in terms of some elementary operations on structures. Another very important aspect is the ability to generate all structures which are equivalent to a given structure.

In this text, we present the solution of all above-mentioned subproblems. We introduced and described two structure properties - characteristics - necessary and sufficient to guarantee the equivalence of respective structures. They are the so-called *non-trivial sets* and *formal ratio*. It should be stressed that using formal ratio, the equivalence can be tested in polynomial time, while neither Z-avoiding trails nor non-trivial sets are operational. Moreover, formal ratio represents a unique representative of an equivalence class. We also introduce four elementary operations and we shown the way how to transform a structure into an arbitrary equivalent one using these operations. (See Proof of Theorem 5.3, which is constructive.) Moreover, using this set of operations one can generate the complete class of equivalent structures - nevertheless, there is still no efficient algorithm to do that and this area is a problem open to further research.

Aknowledgement

This article summarizes part of the knowledge gained when working on my dissertation thesis that was finally submitted in September 2011 and defended in February 2012. This work was supported by GAČR under the grant no. 13-20012S.

Let me thank my supervisor Radim Jiroušek for his cooperation in solving various problems, for reading the manuscript in its various stages and numerous ideas that improved the paper. One of the greatest gratitude belongs to Milan Studený. Thanks to his design of Lemma 4.7, the whole work could be significantly simplified - if compared with the procedure published in the thesis. I would also like to thank both anonymous reviewers for their careful reading and valuable suggestions.

References

[1] V. Bína: *Multidimensional probability distributions: Structure and learning.* (Doctoral disertation) University of Economics, Prague (2011).

[2] M. Bona: *Combinatorics of Permutations*, Chapman Hall-CRC, (2004). ISBN 1-58488-434-7.

[3] D. Geiger, T. Verma, J. Pearl: *Identifying independence in bayesian networks.* Networks 20:5 (1990), 507–534.

[4] R. Jiroušek: *Decomposition of Bayesian Networks Induced by Compositional Models.* In: P. Lucas (ed.) Proceedings of the 2nd European Workshop on Graphical Models. Lorentz Center, Leiden, Netherlands, 2004, pp 129-136.

[5] R. Jiroušek: *Foundations of compositional model theory.* International Journal of General Systems 40:6 (2011), 623–678.

[6] R. Jiroušek: *Conditional independence and factorization of multidimensional models* Fuzzy Systems, IEEE World Congress on Computational Intelligence, Hong Kong, (2008), pp. 2359-2366.

[7] R. Jiroušek and J. Vejnarová: *Compositional models and conditional independence in evidence theory*, International Journal of Approximate Reasoning vol.52, 3 (2011), p. 316-334.
[8] R. Jiroušek: *Persegrams of Compositional Models Revisited: conditional independence.* In Proceedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, Malaga, (2008).
[9] R. Jiroušek and J. Vejnarová: *General framework for multidimensional models.* International journal of intelligent systems, 18 (2003), 107127.
[10] V. Kratochvíl: *Characteristic Properties of Equivalent Structures in Compositional Models.* International Journal of Approximate Reasoning 52:5 (2011), 599–612.
[11] J. Pearl: *Probabilistic Reasoning in Intelligent systems: Networks of Plausible Inference*, Margan Kaufmann, San Mateo, CA, (1988).
[12] J. Pearl: *Causality: Models, Reasoning, and Inference*, Cambridge University Press, (2000).
[13] M. Studený, R. Hemmecke, S. Lindner: *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), pp. 257-264.
[14] J. Vejnarová: *Composition of possibility measures on finite spaces: preliminary results.* In: B. Bouchon-Meunier and R.R. Yager, eds. Proceedings of 7th international conference on information processing and management of uncertainty in knowledge-based systems, IPMU98. Paris: Editions E.D.K., (1998) p. 2530.