International Journal of General Systems

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ggen20

Foundations of compositional models: inference

Vl. Bína, R. Jiroušek & V. Kratochvíl

INTERNATIONAL JOURNAL OF GENERAL SYSTEMS
2021, VOL. 50, NO. 4, 409–433
https://doi.org/10.1080/03081079.2021.1895142

Foundations of compositional models: inference

Vl. Bína[a], R. Jiroušek[a,b] and V. Kratochvíl[a,b]

[a]Faculty of Management, Prague University of Economics and Business, Jindřichův Hradec, Czech Republic; [b]Institute of Information Theory and Automation, Prague, Czech Republic

ABSTRACT

Compositional models, as an alternative to Bayesian networks, are assembled from a system of low-dimensional distributions. Thus the respective apparatus falls fully into probability theory. The present paper surveys the results supporting the design of computational procedures, without which the application of these models to practical problems would be impossible.

The methods of inference cannot do without a possibility to focus on a part of the considered multidimensional model and to incorporate data describing the actual situation. Thus the paper shows how to compute marginals and conditionals of multidimensional models. Also, the paper briefly solves the problem of computing the effect of an intervention, in case the model is interpreted as a causal model.

ARTICLE HISTORY
Received 24 April 2020
Accepted 28 January 2021

KEYWORDS
Probability distribution; multidimensionality; marginalization; conditioning; causality; intervention

Inscription

It was in 2008 when being on a visit at the Binghamton University (SUNY), I promised my distinguished friend Prof. George J. Klir, a founder and the editor in chief of the International Journal of General Systems, a series of three papers summarizing results on theory of probabilistic compositional models. Since that time, a long time has passed and many things have changed. Compositional models stepped out beyond the scope of probability theory into several generalized uncertainty theories and gained the capability of representing causal relations. However, from the intended series of papers, I published just two: on basic properties (IJGS 2011) and structural properties of compositional models (IJGS 2015, coauthored by Václav Kratochvíl). I was postponing the last paper until some open problems are solved, I did not expect there might be a reason to rush. So it happened that when George passed away in May 2016, I had not even started preparing the promised last paper. So, I am keeping my pledge only now and devote this paper to the memory of Prof. George J. Klir.

Radim Jiroušek

1. Introduction

Though this paper is a follow-up of the previous two papers on the theory of compositional models published in *International Journal of General Systems* (Jiroušek 2011; Jiroušek and Kratochvíl 2015), we still owe the readers a more intuitive explanation of why the described

CONTACT R. Jiroušek radim@utia.cas.cz

models are called "compositional". This is because a simple answer, *they are based on the application of the operator of composition*, though fully true, does not satisfy the inquisitive reader, but evokes a natural continuation: *why is the operator called compositional?* So, let us start the paper, rather unconventionally, by answering the latter question.

We assume that a notion of "decomposition" is a generally intuitively accepted notion used in many fields (and not only) of sciences. In artificial intelligence, problems are decomposed into its subproblems, in mathematics, positive integers are unambiguously decomposed into the powers of prime numbers. In Jiroušek (2020), we studied how to decompose knowledge if formalized, say, using logical expressions. One can even decompose a big wardrobe into pieces to move it from one house to another. Let us illustrate the matter of decomposition with an example of graphs (Golumbic 1980; Lauritzen, Speed, and Vijayan 1984).

Two (simple) graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ form a decomposition of a graph $G = (V, E)$ if

(1) G_1, G_2 *are induced subgraphs of G, and* $V_1 \cup V_2 = V, E_1 \cup E_2 = E$,
(2) $V_1 \neq V \neq V_2$,
(3) $V_1 \cap V_2$ *is complete in G.*

Notice that condition (1) guarantees that the original object can be fully reconstructed from the decomposed parts (nothing is lost). Condition (2) guarantees that the subobjects are simpler than the decomposed object. Eventually, the last condition (3) guarantees that, in a way, the subobjects fit each other; it guarantees the existence of the inverse operation. It means that when reconstructing the original object the *operation of composition* can be applied. Notice also that not all objects can be decomposed into simpler parts and that pieces from different decompositions need not fit each other.

Naturally, the decomposition of objects into smaller parts is usually done on purpose. The considered wardrobe, being decomposed, is easier to move, and subproblems are, usually, easier to solve. From this point of view, the present paper is the most important from the above-mentioned series of papers on compositional models. Namely, it shows that the properties of compositional models (studied in the previous two articles) allow for efficient computational procedures, they make the application of these models to inference possible. Principally, to make an inference from the knowledge represented in the form of a multidimensional compositional model, we need a possibility to compute its marginals and the necessary conditionals. Marginalization corresponds to "focusing" on those variables, values of which we are interested in (either we know their values, or we want to update their probabilities). And the conditioning may be used for updating.

Like the preceding ones, this is a survey paper summarizing important computational methods for compositional models as well as presenting some unpublished results. To do it, in the next section we briefly recall the notation and terminology used in the previous papers as well as the basic theoretical properties of the operator of composition and compositional models. Then, Section 2 will be devoted to procedures making the efficient computation of marginal distributions possible, and Section 3 will deal with computation of conditionals. In the latter section, we will also show how to compute the effect of interventions in case that the considered compositional model is interpreted as a causal model.

2. Basic notions and notation

Holding the notation from the preceding papers, we denote the considered finite-valued variables by lower-case Roman characters $\{u, v, w, \ldots\}$. Their sets are denoted by upper-case Roman characters such as $K, L, M, N, \ldots$, with possible indices. $\mathbf{X}_u$ denote the set of values of variable u, and, analogously, $\mathbf{X}_K$ denote the set of all combinations (vectors) of values of variables from K. Thus if $K = \{u, v, w\}$, then $\mathbf{X}_K = \mathbf{X}_u \times \mathbf{X}_v \times \mathbf{X}_w$. Elements of $\mathbf{X}_u$ (values of variables) and $\mathbf{X}_K$ (vectors of variable values) are also called *states* and will be denoted by bold lower-case Roman characters $\{\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots\}$.

Lower-case Greek characters denote probability distributions, e.g. $\pi(K)$ is a probability distribution defined for variables from K. Its marginal distribution for variables from $L \subset K$ is denoted either simply $\pi(L)$ or $\pi^{\downarrow L}$ ($\pi^{\downarrow \emptyset} = 1$). An analogous notation is also used for states: for $\mathbf{a} \in \mathbf{X}_K$, $\mathbf{a}^{\downarrow L}$ is a state from $\mathbf{X}_L$, which is a projection of $\mathbf{a}$. Thus for $K = \{u, v, w\}$ and $\mathbf{a} = (\mathbf{a}_u, \mathbf{a}_v, \mathbf{a}_w) \in \mathbf{X}_K$, and $L = \{u, w\}$, $\mathbf{a}^{\downarrow L} = (\mathbf{a}_u, \mathbf{a}_w)$.

Two distributions $\kappa(K)$ and $\lambda(L)$ are said to be *consistent* if[1] $\kappa^{\downarrow K \cap L} = \lambda^{\downarrow K \cap L}$. Notice that if K and L are disjoint, then $\kappa(K)$ and $\lambda(L)$ are always consistent because $\kappa^{\downarrow \emptyset} = \lambda^{\downarrow \emptyset} = 1$.

Consider a probability distribution $\pi(N)$. If for three disjoint subsets $K, L, M \subseteq N$ (assume that both $K, L \neq \emptyset$) it holds that

$$\pi^{\downarrow K \cup L \cup M} \pi^{\downarrow M} = \pi^{\downarrow K \cup M} \pi^{\downarrow L \cup M}, \tag{1}$$

we say that groups of variables K and L are *conditionally independent given M for probability distribution π*, and denote it $K \perp\!\!\!\perp L \mid M\ [\pi]$. In case that $M = \emptyset$, we say that groups of variables K and L are *independent* and denote it $K \perp\!\!\!\perp L\ [\pi]$. If for two distributions $\pi(K)$ and $\kappa(K)$ it holds that for all $\mathbf{a} \in \mathbf{X}_K$

$$\kappa(\mathbf{a}) = 0 \quad \Longrightarrow \quad \pi(\mathbf{a}) = 0,$$

then we say that κ *dominates* π and express this situation in symbols $\pi \ll \kappa$.

2.1. Operator of composition and its anticipating generalization

Definition 2.1: For two arbitrary distributions $\kappa(K)$ and $\lambda(L)$, for which $\kappa^{\downarrow K \cap L} \ll \lambda^{\downarrow K \cap L}$, their *composition* is given by the following formula:[2]

$$(\kappa \triangleright \lambda) = \frac{\kappa \cdot \lambda}{\lambda^{\downarrow K \cap L}}.$$

In case $\kappa^{\downarrow K \cap L} \not\ll \lambda^{\downarrow K \cap L}$ the composition remains undefined.

Remark 2.2: The reader certainly noticed that in this definition we relaxed one of the conditions mentioned in the Introduction. Namely, we do not require that both κ and λ are simpler than $\kappa \triangleright \lambda$. It means that we admit situations when $K \subseteq L$, or $L \subseteq K$, which, as we will see later, will appear to be useful.

In Jiroušek (2011), we presented simple examples showing that the operator is neither commutative nor associative. In the cited paper and in Jiroušek (2002), we proved its basic properties, which are summarized in the following assertion.

Theorem 2.3: *Suppose $\kappa(K)$, $\lambda(L)$ and $\mu(M)$ are probability distributions. The following statements hold under the assumption that the respective compositions are defined:*

(1) *(Domain):* $\kappa \triangleright \lambda$ *is a probability distribution for* $K \cup L$.
(2) *(Composition preserves first marginal):* $(\kappa \triangleright \lambda)^{\downarrow K} = \kappa$.
(3) *(Reduction): If* $L \subseteq K$ *then* $\kappa \triangleright \lambda = \kappa$.
(4) *(Extension): If* $M \subseteq K$ *then* $\kappa^{\downarrow M} \triangleright \kappa = \kappa$.
(5) *(Perfectization):* $\kappa \triangleright \lambda = \kappa \triangleright (\kappa \triangleright \lambda)^{\downarrow L}$.
(6) *(Commutativity under consistency):* κ *and* λ *are consistent if and only if* $\kappa \triangleright \lambda = \lambda \triangleright \kappa$.
(7) *(Associativity under RIP): If* $K \supseteq (L \cap M)$ *or* $L \supseteq (K \cap M)$ *then* $(\kappa \triangleright \lambda) \triangleright \mu = \kappa \triangleright (\lambda \triangleright \mu)$.
(8) *(Stepwise composition): If* $(K \cap L) \subseteq M \subseteq L$ *then* $(\kappa \triangleright \lambda^{\downarrow M}) \triangleright \lambda = \kappa \triangleright \lambda$.
(9) *(Exchangeability): If* $K \supseteq (L \cap M)$ *then* $(\kappa \triangleright \lambda) \triangleright \mu = (\kappa \triangleright \mu) \triangleright \lambda$.
(10) *(Simple marginalization): If* $(K \cap L) \subseteq M \subseteq K \cup L$ *then* $(\kappa \triangleright \lambda)^{\downarrow M} = \kappa^{\downarrow K \cap M} \triangleright \lambda^{\downarrow L \cap M}$.
(11) *(Conditional independence):* $(K \setminus L) \perp\!\!\!\perp (L \setminus K) \mid (K \cap L)[\kappa \triangleright \lambda]$.
(12) *(Factorization): Let* $M \supseteq K \cup L$. $(K \setminus L) \perp\!\!\!\perp (L \setminus K) \mid (K \cap L)[\mu]$ *if and only if* $\mu^{\downarrow K \cup L} = \mu^{\downarrow K} \triangleright \mu^{\downarrow L}$.

As we will see in the next section, the lack of associativity of the operator of composition can be sufficiently compensated by its generalization called an anticipating operator.

Definition 2.4: Consider an arbitrary set of variables M and two distributions $\kappa(K)$, $\lambda(L)$. Their *anticipating composition* is given by the formula

$$\kappa \circledtriangleright_M \lambda = (\lambda^{\downarrow (M \setminus K) \cap L} \cdot \kappa) \triangleright \lambda = (\lambda^{\downarrow (M \setminus K) \cap L} \triangleright \kappa) \triangleright \lambda.$$

The operator $\circledtriangleright_M$ is called an *anticipating operator of composition*.

Notice that $\kappa \circledtriangleright_{\emptyset} \lambda = \kappa \triangleright \lambda$. Thus it is clear that the result of the composition may remain undefined. However, it follows immediately from the respective definitions that if $\kappa \triangleright \lambda$ is defined then also $\kappa \circledtriangleright_M \lambda$ is defined. Both $\kappa \triangleright \lambda$ and $\kappa \circledtriangleright_M \lambda$ are distributions defined for the same set of variables.

Let us also note that the computations corresponding to these two operators are of the same computational complexity. So, the main difference between the anticipating operator and the operator $\triangleright$ is that the generalized operator is parameterized by an index set. In the following theorem (proved in Jiroušek 2011), we articulate the main purpose for which this operator is introduced. Namely, operator $\triangleright$ can be substituted by an anticipating operator simultaneously with changing the ordering of operations.

Theorem 2.5: *If* $\kappa(K)$, $\lambda(L)$ *and* $\mu(M)$ *are such that* $\mu \triangleright (\kappa \circledtriangleright_M \lambda)$ *is defined, then*

$$(\mu \triangleright \kappa) \triangleright \lambda = \mu \triangleright (\kappa \circledtriangleright_M \lambda).$$

2.2. Compositional models

By a compositional model, we understand a multidimensional probability distribution assembled from low-dimensional distributions by a repetitive application of the operator of composition:

$$\pi = \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m \quad (2)$$

To understand this expression properly, we have to make three conventions, which were also made in previously cited papers (Jiroušek 2011; Jiroušek and Kratochvíl 2015). All of them are of technical nature and make the following exposition more lucid.

(1) To avoid the necessity to repeat it all the time, let us assume in the rest of the paper that each distribution κ_i is defined for variables K_i: i.e. $\kappa_i(K_i)$.
(2) Since the operator of composition is not generally associative, we should use parentheses to control the ordering, in which the operator is to be applied. It would lead to rather intricate formulas, and therefore we omit the parentheses whenever the operator is to be performed from left to right, i.e.

$$\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3 \triangleright \ldots \triangleright \kappa_m = (\ldots((\kappa_1 \triangleright \kappa_2) \triangleright \kappa_3) \triangleright \ldots) \triangleright \kappa_m.$$

(3) The last convention is connected with the fact that the operator is not always defined. To avoid repeating all the time that the respective statements are valid in the case that the particular formulas are defined, let us assume in what follows that all the presented formulas are defined.

Now, based on Property (1) of Theorem 2.3 (and the above-mentioned conventions), we see that the compositional model π defined by formula (2) is a probability distribution of variables $K_1 \cup K_2 \cup \ldots \cup K_m$. As shown in Jiroušek and Kratochvíl (2015), some important properties of the compositional model $\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m$ are encoded in the sequence of sets $K_1, K_2, \ldots, K_m$. For example, from this sequence, one can read the system of conditional independence relations holding for the model. Therefore, we usually refer to the ordered sequence $K_1, K_2, \ldots, K_m$ as to the *structure* of the model.

In Jiroušek (2011), we considered the following special classes of compositional models.

Definition 2.6: Compositional model $\pi = \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m$ is called:

- *perfect*[3] if for all $i = 1, \ldots, m$, $\pi^{\downarrow K_i} = \kappa_i$;
- *flexible*[4] if for all $u \in K_1 \cup \ldots \cup K_m$ there exists a permutation $j_1, j_2, \ldots, j_m$ of indices $1, 2, \ldots, m$ such that $u \in K_{j_1}$, and $\pi = \kappa_{j_1} \triangleright \kappa_{j_2} \triangleright \ldots \triangleright \kappa_{j_m}$;
- *decomposable*[5] if its structure $K_1, K_2, \ldots, K_m$ meets the *running intersection property* (RIP), i.e. if

$$\forall j = 3, \ldots, m \;\; \exists k \in \{1, \ldots, j-1\} \; : \; K_j \cap (K_1 \cup \ldots \cup K_{j-1}) \subseteq K_k.$$

In the next sections, we will need the properties, which are summarized in the following assertion. The proofs of Properties (1)–(4) can be found in Jiroušek (2011), the proof of Property (5) is in Jiroušek (2017).

Theorem 2.7: *For an arbitrary compositional model $\pi = \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m$ the following statements hold true:*

(1) *(Prefix marginalization): For all $j = 1, 2, \ldots, m$ $\pi^{\downarrow K_1 \cup \ldots \cup K_j} = \pi^{\downarrow K_1} \triangleright \ldots \triangleright \pi^{\downarrow K_j}$.*

(2) *(Perfectization):* $\pi = \pi^{\downarrow K_1} \triangleright \pi^{\downarrow K_2} \triangleright \dots \triangleright \pi^{\downarrow K_m}$, *which means that* $\pi^{\downarrow K_1} \triangleright \pi^{\downarrow K_2} \triangleright \dots \triangleright \pi^{\downarrow K_m}$ *is a perfect compositional model, which equals* $\kappa_1 \triangleright \kappa_2 \triangleright \dots \triangleright \kappa_m$.
(3) *(Strong flexibility): If* $\kappa_1 \triangleright \kappa_2 \triangleright \dots \triangleright \kappa_m$ *is perfect and decomposable, then it is flexible.*
(4) *(Perfect model permutation): If* $\kappa_1 \triangleright \kappa_2 \triangleright \dots \triangleright \kappa_m$ *and its permutation* $\kappa_{j_1} \triangleright \kappa_{j_2} \triangleright \dots \triangleright \kappa_{j_m}$ *are both perfect, then* $\kappa_1 \triangleright \kappa_2 \triangleright \dots \triangleright \kappa_m = \kappa_{j_1} \triangleright \kappa_{j_2} \triangleright \dots \triangleright \kappa_{j_m}$.
(5) *(Perfectization preserves flexibility): If* $\kappa_1 \triangleright \kappa_2 \triangleright \dots \triangleright \kappa_m$ *is flexible, then* $\pi^{\downarrow K_1} \triangleright \pi^{\downarrow K_2} \triangleright \dots \triangleright \pi^{\downarrow K_m}$ *is flexible, too.*

Notice that property (2) says that each compositional model can be transformed into a perfect compositional model. In other words, when considering perfect compositional models only, we do not restrict the class of considered probability distributions. Both flexibility and decomposability are useful for conditioning (see Section 3). This is because the respective computational procedures become efficient if the conditioning variable appears among the variables of the first probability distribution from the respective compositional model. Nevertheless, it is also important to realize what generally does not hold for the specified subclasses of models:

(a) Not all perfect models are flexible and not all flexible models are perfect.
(b) Though flexibility guarantees that a flexible model may be rearranged (permuted) in many ways without changing the resulting multidimensional probability distribution, Property (4) does not hold for flexible sequences. Example 12.6 from Jiroušek (2011) presents a system of low-dimensional distributions, from which two different flexible models may be set up: all the permutations of these distributions split into two groups, one part of permutations define one flexible model, permutations from the other group define a different flexible model.

Remark 2.8: In this paper, we concentrate our attention on computations with compositional models. A more general problem, marginalization of compositional expressions, is solved by Malvestuto (2015, 2016), where the author considers compositional expressions. In contrast to compositional models, compositional expressions are defined recursively[6]:

- any probability distribution $\pi(K)$ is a compositional expression;
- if θ_1 and θ_2 are compositional expressions, then $(\theta_1) \triangleright (\theta_2)$ is also a compositional expression.

Thus one can immediately see that any compositional model, which is expressed in the form of Expression (2), is a compositional expression, but not vice versa. For example, $(\pi(u,v) \triangleright \kappa(u,v,w)) \triangleright ((\lambda(v,w) \triangleright \mu(u,v,w)) \triangleright \nu(v,w,x))$ is a compositional expression but not a compositional model, because the parentheses cannot be removed without changing the resulting distribution. The reason why we do not follow Malvestuto's more general approach is purely pragmatic. Having a compositional model composed from low-dimensional distributions, we are sure that all the computations are tractable. When computing the respective compositions, we need only the marginals of low-dimensional distributions. For multidimensional compositional expression, the application of the operator of composition (which may need the computation of a marginal from a subexpression) may easily become intractable.

3. Marginalization

Recall that based on the conventions accepted in Section 2.2, we consider low-dimensional probability distributions $\kappa_1(K_1), \kappa_2(K_2), \ldots, \kappa_m(K_m)$. Using this, the goal of this section is the following: Having a compositional model $\pi = \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m$ and a set of variables $N \subset K_1 \cup \ldots \cup K_m$, find a compositional model representing marginal distribution $\pi^{\downarrow N}$.

To describe the marginalization procedure, we have borrowed the style R. Shachter used in Shachter (1986, 1988) to describe the marginalization in Bayesian networks. He uses just two simple rules: *node deletion* and *edge reversal*. The rules presented below are theoretically supported by the following two lemmas and the properties from Theorem 2.7.

The first lemma enables us to decrease the dimensionality of a compositional model (i.e. the number of variables) by one. It can be done only in situations when a variable appears only in one K_i.

Lemma 3.1: *Consider a compositional model $\pi = \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m$, and denote $K = K_1 \cup \ldots \cup K_m$. If, for variable u, there exists index j such that $u \in K_j$, and $u \notin K_i$ for all $i \in \{1, \ldots, m\} \setminus \{j\}$, then*

$$\pi^{\downarrow K \setminus \{u\}} = \kappa_1 \triangleright \ldots \triangleright \kappa_{j-1} \triangleright \kappa_j^{\downarrow K_j \setminus \{u\}} \triangleright \kappa_{j+1} \triangleright \ldots \triangleright \kappa_m.$$

Proof: The proof consists only in a repetitive application of Property (10) of Theorem 2.3:

$$\begin{aligned}
\pi^{\downarrow K \setminus \{u\}} &= (\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m)^{\downarrow K \setminus \{u\}} = (\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_{m-1})^{\downarrow (K_1 \cup \ldots \cup K_{m-1}) \setminus \{u\}} \triangleright \kappa_m \\
&= (\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_{m-2})^{\downarrow (K_1 \cup \ldots \cup K_{m-2}) \setminus \{u\}} \triangleright \kappa_{m-1} \triangleright \kappa_m = \ldots \\
&= (\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_{j-1} \triangleright \kappa_j)^{\downarrow (K_1 \cup \ldots \cup K_j) \setminus \{u\}} \triangleright \kappa_{j+1} \triangleright \ldots \triangleright \kappa_m \\
&= (\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_{j-1}) \triangleright \kappa_j^{\downarrow K_j \setminus \{u\}} \triangleright \kappa_{j+1} \triangleright \ldots \triangleright \kappa_m.
\end{aligned}$$

■

There is no direct way to delete a variable appearing among the arguments of more than one κ_i. Therefore, we need a tool enabling us to decrease the number of appearances of a variable among the arguments of distributions, from which the model is composed. This is made possible by the following assertion.

Lemma 3.2: *Consider a compositional model $\pi = \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m$, and variable u. Let j, k ($j < k$) be indices, for which $u \in K_j \cap K_k$, and $u \notin K_i$ for all $i \in \{1, \ldots, k-1\} \setminus \{j\}$ (i.e. K_j, K_k are the first two sets from sequence $K_1, \ldots, K_m$ containing u), then*

$$\pi = \kappa_1 \triangleright \ldots \triangleright \kappa_{j-1} \triangleright \kappa_j^{\downarrow K_j \setminus \{u\}} \triangleright \kappa_{j+1} \triangleright \ldots \triangleright \kappa_{k-1} \triangleright \left(\kappa_j \circledtriangleright_M \kappa_k\right) \triangleright \kappa_{k+1} \triangleright \ldots \triangleright \kappa_m,$$

where $M = (K_1 \cup \ldots \cup K_{k-1}) \setminus \{u\}$.

Proof: Due to Property (4) of Theorem 2.3 $\kappa_j = \kappa_j^{\downarrow K_j \setminus \{u\}} \triangleright \kappa_j$, and therefore

$$\kappa_1 \triangleright \ldots \triangleright \kappa_j = (\kappa_1 \triangleright \ldots \triangleright \kappa_{j-1}) \triangleright \left(\kappa_j^{\downarrow K_j \setminus \{u\}} \triangleright \kappa_j\right) = \kappa_1 \triangleright \ldots \triangleright \kappa_{j-1} \triangleright \kappa_j^{\downarrow K_j \setminus \{u\}} \triangleright \kappa_j,$$

where the last equality follows from Property (7) of Theorem 2.3, because $K_j \setminus \{u\} \supseteq (K_1 \cup \ldots \cup K_{j-1}) \cap K_j$. The proof can be finished by a multiple application of Property (9) of

Theorem 2.3 (it may be applied because $(K_1 \cup \ldots \cup K_{j-1} \cup (K_j \setminus \{u\})) \supseteq K_j \cap K_i$ for all $i = j+1, \ldots, k-1$) followed by the application of Theorem 2.5:

$$\begin{aligned}
\kappa_1 \triangleright \ldots \triangleright \kappa_k &= \kappa_1 \triangleright \ldots \triangleright \kappa_{j-1} \triangleright \kappa_j^{\downarrow K_j \setminus \{u\}} \triangleright \kappa_j \triangleright \kappa_{j+1} \triangleright \ldots \triangleright \kappa_{k-1} \triangleright \kappa_k \\
&= \kappa_1 \triangleright \ldots \triangleright \kappa_{j-1} \triangleright \kappa_j^{\downarrow K_j \setminus \{u\}} \triangleright \kappa_{j+1} \triangleright \kappa_j \triangleright \ldots \triangleright \kappa_{k-1} \triangleright \kappa_k \\
&= \ldots = \kappa_1 \triangleright \ldots \triangleright \kappa_{j-1} \triangleright \kappa_j^{\downarrow K_j \setminus \{u\}} \triangleright \kappa_{j+1} \triangleright \ldots \triangleright \kappa_{k-1} \triangleright \kappa_j \triangleright \kappa_k \\
&= \kappa_1 \triangleright \ldots \triangleright \kappa_{j-1} \triangleright \kappa_j^{\downarrow K_j \setminus \{u\}} \triangleright \kappa_{j+1} \triangleright \ldots \triangleright \kappa_{k-1} \triangleright \left(\kappa_j \circledtriangleright_M \kappa_k\right).
\end{aligned}$$

■

As the reader can see, these two assertions guarantee that any variable can be marginalized out of a compositional model. If the considered variable is among the arguments of only one distribution κ_i, then its elimination is made possible by Lemma 3.1. If this variable is among the arguments of several distributions forming the compositional model, then one has first successively to decrease the number of occurrences of the variable in a model until it appears among the arguments of only one κ_i. This decrease is made possible by the application of Lemma 3.2. The final elimination of the variable from the model is, again, realized using Lemma 3.1.

Thus from multiple applications of Lemma 3.2 one immediately gets the following corollary, which was proved already in Jiroušek (2000).

Corollary 3.3: *Consider a compositional model* $\pi = \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m$, *and variable* $u \in K_{i_1} \cap K_{i_2} \cap \ldots \cap K_{i_k}$, *for which* $u \notin K_j$ *for all* $j \in \{1, 2, \ldots, m\} \setminus \{i_1, \ldots, i_k\}$. *Assume* $i_1 < i_2 < \ldots < i_k$. *Then*

$$\pi^{\downarrow (K_1 \cup \ldots \cup K_m) \setminus \{u\}} = \lambda_1 \triangleright \lambda_2 \triangleright \ldots \triangleright \lambda_m,$$

where

$$\begin{aligned}
\lambda_j &= \kappa_j \quad \text{for all } j \in \{1, \ldots, m\} \setminus \{i_1, \ldots, i_k\}, \\
\lambda_{i_1} &= \kappa_{i_1}^{\downarrow K_{i_1} \setminus \{u\}}, \\
\lambda_{i_2} &= \left(\kappa_{i_1} \circledtriangleright_{L_{i_2}} \kappa_{i_2}\right)^{\downarrow (K_{i_1} \cup K_{i_2}) \setminus \{u\}}, \\
\lambda_{i_3} &= \left(\kappa_{i_1} \circledtriangleright_{L_{i_2}} \kappa_{i_2} \circledtriangleright_{L_{i_3}} \kappa_{i_3}\right)^{\downarrow (K_{i_1} \cup K_{i_2} \cup K_{i_3}) \setminus \{u\}}, \\
&\;\vdots \\
\lambda_{i_k} &= \left(\kappa_{i_1} \circledtriangleright_{L_{i_2}} \kappa_{i_2} \circledtriangleright_{L_{i_3}} \ldots \circledtriangleright_{L_{i_k}} \kappa_{i_k}\right)^{\downarrow (K_{i_1} \cup \ldots \cup K_{i_k}) \setminus \{u\}},
\end{aligned}$$

where $L_{i_j} = (K_1 \cup K_2 \cup \ldots \cup K_{i_j - 1}) \setminus \{u\}$.

3.1. Marginalization rules

The process of marginalization of a compositional model can be enhanced by the application of two properties presented in previous theorems. Property (3) from Theorem 2.3

advises to delete a distribution from a model if all its arguments appear among the arguments of previous distributions, and Property (1) of Theorem 2.7 advises to cut off the tail of the model if it is unnecessary. Thus the application of the following four rules enables us to find a marginal to an arbitrary compositional model rather efficiently. The notation corresponds to the computation of $(\kappa_1 \triangleright \ldots \triangleright \kappa_m)^{\downarrow N}$ for $N \subset K_1 \cup \ldots \cup K_m$.

Tail cut off. If $N \subseteq K_1 \cup \ldots \cup K_j$ then cut off the tail of the model with $\kappa_{j+1}, \ldots, \kappa_m$, i.e. redefine $m: = j$.

Model reduction. If there exists index j such that $K_j \subseteq K_1 \cup \ldots \cup K_{j-1}$ then delete κ_j from the model, i.e. for all $\ell = j, j+1, \ldots, m-1$ redefine $\kappa_\ell := \kappa_{\ell+1}$; redefine $m: = m-1$.

Variable deletion. If there exists variable $u \notin N$ and index j such that $u \in K_j$, and $u \notin K_i$ for all $i = 1, 2, \ldots, j-1, j+1, \ldots, m$, then marginalize variable u out of distribution κ_j, i.e. redefine $\kappa_j := \kappa_j^{\downarrow K_j \setminus \{u\}}$.

Decrease of variable occurrences. For variable $u \in (K_1 \cup \ldots \cup K_m) \setminus N$ find indices j and k such that $u \in K_j \cap K_k$, and $u \notin K_i$ for all $i = 1, 2, \ldots, j-1, j+1, \ldots, k-1$, then set $M := (K_1 \cup \ldots \cup K_{k-1}) \setminus \{u\}$, and redefine $\kappa_j := \kappa_j^{\downarrow K_j \setminus \{u\}}$; redefine $\kappa_k := \kappa_j \circledtriangleright_M \kappa_k$.

The marginalization procedure consists of proper applications of the above-presented rules. The simplest way is to start with the Tail-cut-off rule, and then check whether Variable-deletion rule is applicable. The application of the latter rule may induce the applicability of Model-reduction rule and vice versa. So, these two rules should be applied in turns as long as one of them is applicable. The Decrease-of-variable-occurrences rule should be applied only when all other rules are not applicable. When applying this last rule, one should start with the variable with the lowest number of occurrences in the model, and apply it to this variable until the variable is among the argument of only one distribution of the model, which means that it can be deleted using Variable-deletion rule.

Naturally, the above-presented sketch of a marginalization procedure just highlights that one should start with the application of simple rules, and only if necessary one should start applying the Decrease-of-variable-occurrences rule. When designing an efficient algorithm one can employ other properties of the introduced rules as those described in this section.

It is obvious that the application of the first three marginalization rules (i.e. Tail-cut-off, Model-reduction and Variable-deletion) preserves perfectness of the model. This means that if a model is perfect before these rules are applied, it is perfect also after the rules are performed. This does not hold for Decrease-of-variable-occurrences rule. However, thanks to Property (5) of Theorem 2.3, one can "easily" modify this rule to preserve the perfectness of the models, too. It suffices to change the way how κ_k is redefined:

$$\kappa_k := (\kappa_1 \triangleright \ldots \triangleright \kappa_{k-1})^{\downarrow (K_1 \cup \ldots \cup K_{k-1}) \cap K_k \setminus \{u\}} \triangleright (\kappa_j \circledtriangleright_M \kappa_k).$$

The reason, why we put the word "easily" into quotation marks is that the computations of the necessary marginal distribution may be computationally expensive.

Theorem 3.4: *The application of any of the following three rules:*

- *Tail-cut-off rule,*
- *Model-reduction rule,*
- *Variable-deletion rule*

preserves decomposability of the model.

Proof: *Tail-cut-off rule.* This part of the proof is trivial because the definition of decomposable model $\kappa_1 \triangleright \ldots \triangleright \kappa_m$ (see Definition 2.6) guarantees that all its "prefix" models $\kappa_1 \triangleright \ldots \triangleright \kappa_j$, for $j < m$, are decomposable, too.

Model-reduction rule. Consider a model $\kappa_1 \triangleright \ldots \triangleright \kappa_{j-1} \triangleright \kappa_{j+1} \triangleright \ldots \triangleright \kappa_m$ resulting by the deletion of κ_j from the original decomposable model $\kappa_1 \triangleright \ldots \triangleright \kappa_m$ (the application of the Model-reduction rule). This rule is applicable only if $K_j \subseteq K_1 \cup \ldots \cup K_{j-1}$. In combination with the assumption that the model is decomposable, it means that there must exist $i < j$ such that $K_j \subseteq K_i$.

To show that $K_1, \ldots, K_{j-1}, K_{j+1}, \ldots, K_m$ meets RIP we have to show that for each $\ell = 3, \ldots, j-1, j+1, \ldots, m$ there exists $k < \ell$, $k \neq j$, such that

$$K_\ell \cap \left(K_1 \cup \ldots \cup K_{\ell-1}\right) \subseteq K_k. \tag{3}$$

First, notice that because of $K_j \subseteq K_1 \cup \ldots \cup K_{j-1}$, for $\ell > j$ $(K_1 \cup \ldots \cup K_{\ell-1}) = (K_1 \cup \ldots \cup K_{j-1} \cup K_{j+1} \cup \ldots \cup K_{\ell-1})$. Then realize that the decomposability of the original model $\kappa_1 \triangleright \ldots \triangleright \kappa_m$ guarantees the existence of $k < \ell$, for which Equation (3) holds but it may happen that for some ℓ this $k = j$, which is from the sequence deleted. However, we showed above that $K_j \subseteq K_i$, and therefore

$$K_\ell \cap \left(K_1 \cup \ldots \cup K_{\ell-1}\right) \subseteq K_i,$$

for $i < j$, which finishes the proof that the Model-reduction rule preserves decomposability of the model.

Variable-deletion rule. Assume that $K_1, \ldots, K_m$ meets RIP, and u is the element of only one K_j from this sequence. This assumption says that for any $k > 1$ $(k \leq m)$ there exists $\ell < k$ such that $K_k \cap (K_1 \cup \ldots \cup K_{k-1}) \subseteq K_\ell$. Since we assume that u is contained only in one K_j, it is clear that for all $k = 2, 3, \ldots, m$ it holds that $u \notin K_k \cap (K_1 \cup \ldots \cup K_{k-1})$. Therefore, if for some K and ℓ the inclusion $K_k \cap (K_1 \cup \ldots \cup K_{k-1}) \subseteq K_\ell$ holds for the sequence $K_1, \ldots, K_m$, the same inclusion holds with the same indices K and ℓ also for the sequence $K_1, K_2, \ldots, K_{j-1}$, $(K_j \setminus \{u\}), K_{j+1}, \ldots, K_m$. Thus the latter sequence meets RIP, too. ■

3.2. Example

The goal of this section is to illustrate how the application of the above-presented rules eliminates variables without damaging the mutual dependence of the remaining variables (we do not care about the efficiency of the procedure). All the rules presented in the preceding section are employed.

Consider 10 variables $u_1, u_2, \ldots, u_{10}$ and the following model composed from six distributions

$$\begin{aligned} &\kappa_1(u_1, u_2, u_3) \triangleright \kappa_2(u_1, u_3, u_4) \triangleright \kappa_3(u_4, u_5, u_6) \triangleright \kappa_4(u_1, u_4, u_7) \\ &\quad \triangleright \kappa_5(u_5, u_6, u_8, u_9) \triangleright \kappa_6(u_6, u_9, u_{10}). \end{aligned} \tag{4}$$

Let us compute its marginal for variables $N = \{u_3, u_4, u_7, u_8\}$.

First, Cut-off-tail rule advises to delete the last distribution κ_6. Then, one can notice that variables u_2 and u_9 are among the arguments of only κ_1, and κ_5, respectively (the latter fact is true after κ_6 is deleted), and therefore these two variables may be marginalized out using the Variable-deletion rule. After these three steps, we get the following model for 7 variables

$$\kappa_1(u_1,u_3) \triangleright \kappa_2(u_1,u_3,u_4) \triangleright \kappa_3(u_4,u_5,u_6) \triangleright \kappa_4(u_1,u_4,u_7) \triangleright \kappa_5(u_5,u_6,u_8), \tag{5}$$

which is marginal to the original model defined by Expression (4). With the given goal (to compute marginal for variables $N = \{u_3,u_4,u_7,u_8\}$) only the Decrease-of-variable-occurrences is applicable to the model from Expression (5). The application of this rule to variable u_5 yields (notice that $(\{u_1,u_3,u_4,u_6,u_7\} \setminus \{u_4,u_6\}) \cap \{u_5,u_6,u_8\} = \emptyset$)

$$\begin{aligned}
&\kappa_1(u_1,u_3) \triangleright \kappa_2(u_1,u_3,u_4) \triangleright \kappa_3(u_4,u_6) \triangleright \kappa_4(u_1,u_4,u_7)\\
&\quad \triangleright \left(\kappa_3(u_4,u_5,u_6) \circledtriangleright_{\{u_1,u_3,u_4,u_6,u_7\}} \kappa_5(u_5,u_6,u_8)\right)\\
&= \kappa_1(u_1,u_3) \triangleright \kappa_2(u_1,u_3,u_4) \triangleright \kappa_3(u_4,u_6) \triangleright \kappa_4(u_1,u_4,u_7)\\
&\quad (\kappa_3(u_4,u_5,u_6) \triangleright\ \kappa_5(u_5,u_6,u_8))\,.
\end{aligned}$$

In this model, variable u_5 is among the arguments only for the last four-dimensional distribution $(\kappa_3(u_4,u_5,u_6) \triangleright \kappa_5(u_5,u_6,u_8))$, and therefore it can be eliminated using the Variable-deletion rule. Therefore, denoting

$$\lambda(u_4,u_6,u_8) = (\kappa_3(u_4,u_5,u_6) \triangleright \kappa_5(u_5,u_6,u_8))^{\downarrow\{u_4,u_6,u_8\}},$$

we got the following six-dimensional model:

$$\kappa_1(u_1,u_3) \triangleright \kappa_2(u_1,u_3,u_4) \triangleright \kappa_3(u_4,u_6) \triangleright \kappa_4(u_1,u_4,u_7) \triangleright \lambda(u_4,u_6,u_8),$$

from which u_6 can be eliminated in an analogous way:

$$\begin{aligned}
&\kappa_1(u_1,u_3) \triangleright \kappa_2(u_1,u_3,u_4) \triangleright \kappa_3(u_4,u_6) \triangleright \kappa_4(u_1,u_4,u_7) \triangleright \lambda(u_4,u_6,u_8)\\
&\quad = \kappa_1(u_1,u_3) \triangleright \kappa_2(u_1,u_3,u_4) \triangleright \kappa_3(u_4) \triangleright \kappa_4(u_1,u_4,u_7)\\
&\qquad \triangleright \left(\kappa_3(u_4,u_6) \circledtriangleright_{\{u_1,u_3,u_4,u_7\}} \lambda(u_4,u_6,u_8)\right)\\
&\quad = \kappa_1(u_1,u_3) \triangleright \kappa_2(u_1,u_3,u_4) \triangleright \kappa_3(u_4) \triangleright \kappa_4(u_1,u_4,u_7) \triangleright \mu(u_4,u_6,u_8),
\end{aligned}$$

the marginal of which is got by the application of Variable-deletion rule to variable u_6, which is an argument of the only distribution $\mu = (\kappa_3(u_4,u_6) \circledtriangleright_{\{u_1,u_3,u_4,u_7\}} \lambda(u_4,u_6,u_8))$. Thus we are getting the following model:

$$\kappa_1(u_1,u_3) \triangleright \kappa_2(u_1,u_3,u_4) \triangleright \kappa_3(u_4) \triangleright \kappa_4(u_1,u_4,u_7) \triangleright \mu(u_4,u_8).$$

Now, we can apply the so far unused Model-reduction rule that deletes $\kappa_3(u_4)$ from the model. The reader might also have a temptation to apply Property (4) of Theorem 2.3 to $\kappa_1(u_1,u_3) \triangleright \kappa_2(u_1,u_3,u_4)$, but it is applicable only when we know that κ_1 is a marginal of κ_2. Not assuming this, we have to proceed carefully and eliminate the last remaining variable u_1 using the Decrease-of-variable-occurrences rule twice:

$$\begin{aligned}
&\kappa_1(u_1,u_3) \triangleright \kappa_2(u_1,u_3,u_4) \triangleright \kappa_4(u_1,u_4,u_7) \triangleright \mu(u_4,u_8)\\
&\quad = \kappa_1(u_3) \triangleright \left(\kappa_1(u_1,u_3) \circledtriangleright_{\{u_3\}} \kappa_2(u_1,u_3,u_4)\right) \triangleright \kappa_4(u_1,u_4,u_7) \triangleright \mu(u_4,u_8)
\end{aligned}$$

$$= \kappa_1(u_3) \rhd \left(\kappa_1(u_1,u_3) \circledtriangleright_{\{u_3\}} \kappa_2(u_1,u_3,u_4)\right)^{\downarrow\{u_3,u_4\}}$$
$$\rhd \left(\left(\kappa_1(u_1,u_3) \circledtriangleright_{\{u_3\}} \kappa_2(u_1,u_3,u_4)\right) \circledtriangleright_{\{u_3,u_4\}} \kappa_4(u_1,u_4,u_7)\right) \rhd \mu(u_4,u_8).$$

Thus the resulting model is composed from four distributions. Denoting

$$\nu(u_3,u_4) = \left(\kappa_1(u_1,u_3) \circledtriangleright_{\{u_3\}} \kappa_2(u_1,u_3,u_4)\right)^{\downarrow\{u_3,u_4\}}$$

and

$$\rho(u_1,u_3,u_4,u_7) = \left(\left(\kappa_1(u_1,u_3) \circledtriangleright_{\{u_3\}} \kappa_2(u_1,u_3,u_4)\right) \circledtriangleright_{\{u_3,u_4\}} \kappa_4(u_1,u_4,u_7)\right),$$

it is

$$\kappa_1(u_3) \rhd \nu(u_3,u_4) \rhd \rho(u_1,u_3,u_4,u_7) \rhd \mu(u_4,u_8),$$

from which u_1 can easily be eliminated using Variable-deletion rule. Thus the resulting compositional model $\kappa_1(u_3) \rhd \nu(u_3,u_4) \rhd \rho(u_3,u_4,u_7) \rhd \mu(u_4,u_8)$ is the required four-dimensional marginal $\pi^{\downarrow\{u_3,u_4,u_7,u_8\}}$ of the compositional model defined by Expression (4).

Remark 3.5: In this paper, only theoretical foundations of computational procedures are presented. The reader interested in more sophisticated rules supporting the design of efficient marginalization processes is referred to Bína and Jiroušek (2006); Malvestuto (2015).

4. Conditioning

The application of multidimensional probabilistic models to problems of inference can hardly be realized without the possibility to compute conditional probability distributions. For this purpose, we need a degenerate one-dimensional distribution expressing certainty. Consider variable u and its value $\mathbf{a} \in \mathbf{X}_u$. The distribution $\delta^u_{\mathbf{a}}$ expressing for certain that variable $u = \mathbf{a}$ is defined for each $\mathbf{e} \in \mathbf{X}_u$ as

$$\delta^u_{\mathbf{a}}(\mathbf{e}) = \begin{cases} 1, & \text{if } \mathbf{e} = \mathbf{a}; \\ 0, & \text{otherwise.} \end{cases}$$

The following assertion shows that a conditional distribution can be computed using the operation of composition.

Theorem 4.1: *Consider a distribution $\kappa(K)$, variable $u \in K$, its value $\mathbf{a} \in \mathbf{X}_u$, and $L \subseteq K \setminus \{u\}$. If $\kappa^{\downarrow\{u\}}(\mathbf{a}) > 0$, then the corresponding conditional distribution can be computed*

$$\kappa^{L|u=\mathbf{a}} = \left(\delta^u_{\mathbf{a}} \rhd \kappa\right)^{\downarrow L}.$$

Proof: Due to Property (1) of Theorem 2.3, $\delta^u_{\mathbf{a}} \rhd \kappa$ is a probability distribution of K. Thus, for any $\mathbf{c} \in \mathbf{X}_K$

$$\left(\delta^u_{\mathbf{a}} \rhd \kappa\right)(\mathbf{c}) = \begin{cases} \dfrac{\kappa(\mathbf{c})}{\kappa^{\downarrow\{u\}}(\mathbf{a})}, & \text{if } \mathbf{c}^{\downarrow\{u\}} = \mathbf{a}; \\ 0, & \text{otherwise,} \end{cases}$$

and therefore, for any $\mathbf{e} \in \mathbf{X}_L$

$$\left(\delta^u_{\mathbf{a}} \rhd \kappa\right)^{\downarrow L}(\mathbf{e}) = \frac{\kappa^{\downarrow L\cup\{u\}}(\mathbf{e}, \mathbf{a})}{\kappa^{\downarrow\{u\}}(\mathbf{a})}.$$

■

Theorem 4.1 shows that simple conditional probability distribution can be expressed in a form of a compositional model. However, a computational problem may arise, when one wants to compute a conditional distribution from a multidimensional distribution represented in a form of a compositional model. To study this possibility, from now on, we keep considering compositional model $\kappa_1(K_1) \rhd \kappa_2(K_2) \rhd \ldots \rhd \kappa_m(K_m)$.

Due to Theorem 4.1, the computation of a conditional distribution from a model $\kappa_1 \rhd \ldots \rhd \kappa_m$ means to compute

$$\delta^u_{\mathbf{a}} \rhd (\kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_m),$$

which may be difficult. Nevertheless, the computation of the considered conditional distribution is an easy task if $u \in K_1$ because, in this special case, the following assertion may be used.

Theorem 4.2: *Consider a compositional model $\kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_m$, variable $u \in K_1$ and its value $\mathbf{a} \in \mathbf{X}_u$. Then,*

$$\delta^u_{\mathbf{a}} \rhd (\kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_m) = (\delta^u_{\mathbf{a}} \rhd \kappa_1) \rhd \kappa_2 \rhd \ldots \rhd \kappa_m.$$

Proof: The proof is based on a multiple application of Property (7) (Associativity under RIP) of Theorem 2.3. Namely,

$$\delta^u_{\mathbf{a}} \rhd \Big((\kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_{m-1}) \rhd \kappa_m\Big) = \delta^u_{\mathbf{a}} \rhd (\kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_{m-1}) \rhd \kappa_m,$$

because $(K_1 \cup \ldots \cup K_{m-1}) \supseteq \{u\} \cap K_m$, and therefore the mentioned Property (7) is applicable. Since we assume that $u \in K_1$, $(K_1 \cup \ldots \cup K_{j-1}) \supseteq \{u\} \cap K_j$ for all $j = m-1$, $m-2, \ldots, 2$, and we can repeat this idea getting

$$\begin{aligned} \delta^u_{\mathbf{a}} \rhd (\kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_{m-1} \rhd \kappa_m) &= \delta^u_{\mathbf{a}} \rhd (\kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_{m-1}) \rhd \kappa_m \\ &= \delta^u_{\mathbf{a}} \rhd (\kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_{m-2}) \rhd \kappa_{m-1} \rhd \kappa_m \\ &= \ldots = \delta^u_{\mathbf{a}} \rhd \kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_{m-2} \rhd \kappa_{m-1} \rhd \kappa_m. \end{aligned}$$

■

4.1. Conditioning in flexible models

Theorem 4.2 shows that if $u \in K_1$, all the necessary computations are local. In case that variable u is not among the variables, for which κ_1 is defined, one cannot employ Theorem 4.2, and the computation of $\delta^u_{\mathbf{a}} \triangleright (\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n)$ may be space and time-demanding. It is why we want the conditioning variable to be among the arguments of the first distribution. It is why we are so much interested in models, in which the ordering of distributions in a compositional model may be changed without modifying the represented distribution. Recall, these are the flexible models (see Definition 2.6), for which there exist as many re-orderings as necessary to get any variable at the very front part of the model. The reader certainly noticed that it is the widest class of models possessing this property. As expressed in Theorem 2.7, its proper subclass is formed by perfect decomposable models.

The flexible models do not allow only for the efficient computation of simple conditionals but also for conditionals with a multiple condition like $\kappa^{L|u=\mathbf{a},v=\mathbf{c}}$. This is possible due to Theorem 4.5 presented below. It states that the computation of a conditional from a flexible model does not spoil the flexibility of the model, and therefore one can repeat the computations successively several times. To prove this important assertion, we need the following Lemma and its Corollary.

Lemma 4.3: *Consider $\kappa(K), \lambda(L)$, and $\mu(M)$. If either $M \subseteq K$, or $M \subseteq L$, then*

$$\mu \triangleright (\kappa \triangleright \lambda) = (\mu \triangleright (\kappa \triangleright \lambda))^{\downarrow K} \triangleright (\mu \triangleright (\kappa \triangleright \lambda))^{\downarrow L}. \tag{6}$$

Proof: Let us denote $\pi(K \cup L) = \kappa(K) \triangleright \lambda(L)$. First, assume that $K \supseteq M$. Then,

$$\begin{aligned}
\mu \triangleright \pi &= (\mu \triangleright \pi)^{\downarrow K} \triangleright (\mu \triangleright \pi) && \text{Property (4), Th. 2.3}\\
&= (\mu \triangleright \pi)^{\downarrow K} \triangleright \mu \triangleright \pi && \text{Property (7), Th. 2.3}\\
&= (\mu \triangleright \pi)^{\downarrow K} \triangleright \mu \triangleright \left(\pi^{\downarrow K} \triangleright \pi^{\downarrow L}\right) && \text{Properties (11) and (12), Th. 2.3}\\
&= (\mu \triangleright \pi)^{\downarrow K} \triangleright \mu \triangleright \pi^{\downarrow K} \triangleright \pi^{\downarrow L} && \text{Property (7), Th. 2.3}\\
&= (\mu \triangleright \pi)^{\downarrow K} \triangleright \mu \triangleright \pi^{\downarrow L} && \text{Property (3), Th. 2.3}\\
&= (\mu \triangleright \pi)^{\downarrow K} \triangleright \left(\mu \triangleright \pi^{\downarrow L}\right) && \text{Property (7), Th. 2.3}\\
&= (\mu \triangleright \pi)^{\downarrow K} \triangleright (\mu \triangleright \pi)^{\downarrow L}. && \text{Property (10), Th. 2.3}
\end{aligned}$$

In case that $M \subseteq L$, using the same computations as above we get

$$\mu \triangleright \pi = (\mu \triangleright \pi)^{\downarrow L} \triangleright (\mu \triangleright \pi)^{\downarrow K} = (\mu \triangleright \pi)^{\downarrow K} \triangleright (\mu \triangleright \pi)^{\downarrow L},$$

where the last equality holds because of Property (6) of Theorem 2.3. ■

Corollary 4.4: *Let $\pi(K \cup L) = \kappa(K) \triangleright \lambda(L)$. Then for any $u \in K \cup L$ and $\mathbf{a} \in \mathbf{X}_u$ such that $\pi^{\downarrow\{u\}}(\mathbf{a}) > 0$, the conditional $\vartheta(K \cup L) = \delta^u_{\mathbf{a}} \triangleright \pi(K \cup L)$ can be expressed in the form of a composition*

$$\vartheta = \vartheta^{\downarrow K} \triangleright \vartheta^{\downarrow L}.$$

Theorem 4.5: *Consider a flexible model $\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m$, variable $u \in K_1$, and its value $\mathbf{a} \in \mathbf{X}_u$ such that $\kappa_1^{\downarrow\{X\}}(\mathbf{a}) > 0$. Denote $\pi = \delta^u_{\mathbf{a}} \triangleright (\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m)$. Then the perfect model $\pi^{\downarrow K_1} \triangleright \pi^{\downarrow K_2} \triangleright \ldots \triangleright \pi^{\downarrow K_m}$ is also flexible.*

Proof: We will show that for each permutation $i_1, i_2, \ldots, i_m$ for which

$$\kappa_{i_1} \rhd \kappa_{i_2} \rhd \ldots \rhd \kappa_{i_m} = \kappa_1 \rhd \kappa_2 \rhd \ldots \rhd \kappa_m,$$

and for each $j = 1, \ldots, m$, distribution $\pi^{\downarrow K_{i_1}} \rhd \pi^{\downarrow K_{i_2}} \rhd \ldots \rhd \pi^{\downarrow K_{i_j}}$ is marginal to π, which means that

$$\pi^{\downarrow K_{i_1}} \rhd \pi^{\downarrow K_{i_2}} \rhd \ldots \rhd \pi^{\downarrow K_{i_j}} = \pi^{\downarrow K_{i_1} \cup \ldots \cup K_{i_j}}. \tag{7}$$

For $j = 1$ Equality (7) is trivial. To conclude the induction we will show that if $\pi^{\downarrow K_{i_1}} \rhd \ldots \rhd \pi^{\downarrow K_{i_{j-1}}}$ is the marginal of π, the same must hold also for

$$\pi^{\downarrow K_{i_1}} \rhd \ldots \rhd \pi^{\downarrow K_{i_j}} = (\pi^{\downarrow K_{i_1}} \rhd \ldots \rhd \pi^{\downarrow K_{i_{j-1}}}) \rhd \pi^{\downarrow K_{i_j}}.$$

Assume

$$\pi^{\downarrow K_{i_1}} \rhd \ldots \rhd \pi^{\downarrow K_{i_{j-1}}} = \pi^{\downarrow K_{i_1} \cup \ldots \cup K_{i_{j-1}}}. \tag{8}$$

The considered permutation is selected in the way that $(\kappa_{i_1} \rhd \ldots \rhd \kappa_{i_{j-1}}) \rhd \kappa_{i_j}$ is marginal to $\kappa_1 \rhd \ldots \rhd \kappa_m$, and therefore, due to Property (11) of Theorem 2.3,

$$(K_{i_1} \cup \ldots \cup K_{i_{j-1}}) \setminus K_{i_j} \perp\!\!\!\perp K_{i_j} \setminus (K_{i_1} \cup \ldots \cup K_{i_{j-1}}) \mid K_{i_j} \cap (K_{i_1} \cup \ldots \cup K_{i_{j-1}}) \qquad [\kappa_1 \rhd \ldots \rhd \kappa_m].$$

Due to Corollary 4.4, the same conditional independence relation holds also for distribution π, and therefore, due to Property (12) of Theorem 2.3,

$$\pi^{\downarrow K_{i_1} \cup \ldots \cup K_{i_j}} = \pi^{\downarrow K_{i_1} \cup \ldots \cup K_{i_{j-1}}} \rhd \pi^{\downarrow K_{i_j}}.$$

■

Perhaps, it is not necessary to highlight the importance of Theorem 4.5. It follows from the fact that it makes conditioning by several variables possible. When computing $\pi^{L|u=\mathbf{a},v=\mathbf{c}}$ for flexible compositional model $\pi = \kappa_1 \rhd \ldots \rhd \kappa_m$, and some $u, v \in K_1 \cup \ldots \cup K_m$, $L \subseteq (K_1 \cup \ldots \cup K_m) \setminus \{u, v\}$, and $\mathbf{a} \in \mathbf{X}_u$, $\mathbf{c} \in \mathbf{X}_v$, one can proceed in two steps. In the first step, one finds the ordering $K_{i_1}, \ldots, K_{i_m}$, for which the variable $u \in K_{i_1}$, and computes $(\delta^u_{\mathbf{a}} \rhd \kappa_{i_1}) \rhd \ldots \rhd \kappa_{i_m}$ (recall that $\pi^{L|u=\mathbf{a}} = ((\delta^u_{\mathbf{a}} \rhd \kappa_{i_1}) \rhd \ldots \rhd \kappa_{i_m})^{\downarrow L}$). Before performing the second step one has to become aware that Theorem 4.5 does not state that $(\delta^u_{\mathbf{a}} \rhd \kappa_{i_1}) \rhd \ldots \rhd \kappa_{i_m}$ is flexible. It states that its perfectized form is flexible (all distributions must be marginals of the multidimensional model). Therefore, one has to compute the perfect model (see Theorem 10.9 of (Jiroušek 2011)):

$$\begin{aligned}
\lambda_{i_1} &= (\delta^u_{\mathbf{a}} \rhd \kappa_{i_1}), \\
\lambda_{i_2} &= (\kappa_{i_1})^{\downarrow K_{i_2} \cap K_{i_1}} \rhd \kappa_{i_2}, \\
\lambda_{i_3} &= (\kappa_{i_1} \rhd \kappa_{i_2})^{\downarrow K_{i_3} \cap (K_{i_1} \cup K_{i_2})} \rhd \kappa_{i_3}, \\
&\vdots \\
\lambda_{i_m} &= (\kappa_{i_1} \rhd \ldots \rhd \kappa_{i_{m-1}})^{\downarrow K_{i_m} \cap (K_{i_1} \cup \ldots \cup K_{i_{m-1}})} \rhd \kappa_{i_m}.
\end{aligned}$$

Now, having a flexible model $\lambda_{i_1} \triangleright \ldots \triangleright \lambda_{i_m}$ (keep in mind that $\pi^{L|u=\mathbf{a}} = (\lambda_{i_1} \triangleright \ldots \triangleright \lambda_{i_m})^{\downarrow L}$), one can find its permutation $\lambda_{j_1} \triangleright \ldots \triangleright \lambda_{j_m}$, for which $v \in K_{j_1}$, and compute

$$\pi^{L|u=\mathbf{a},v=\mathbf{c}} = \left((\delta^{v}_{\mathbf{c}}(v) \triangleright \lambda_{i_1}) \triangleright \lambda_{i_2} \triangleright \ldots \triangleright \lambda_{i_m}\right)^{\downarrow L}.$$

4.2. Decomposable models

In practical situations, it may be very difficult to verify the flexibility of a compositional model. It is the widest class of models, in which all variables may be moved to the beginning of the model, but because of computational problems, we usually have to restrict our attention to its proper subclass, to perfect decomposable models. Namely, both the perfectness and decomposability can be easily checked. Moreover, it is an old result of Kellerer (1964a, 1964b) guaranteeing that if a decomposable model is composed from pairwise consistent distributions, then it is also perfect. This simplifies the verification procedure to pairwise consistency, which is for low-dimensional distributions a trivial task.

Another advantage of decomposable models resides in the fact that there is a simple way to find out whether a system of sets $\{K_1, \ldots, K_m\}$ can be ordered to meet RIP. If such a reordering exists, then one of such possible permutations $(K_{j_1}, \ldots, K_{j_m})$ is obtained by the following *Maximum cardinality search* (adapted for sets from Tarjan and Yannakakis (1984)):

(1) Place any set K_i at the beginning of the sequence, i.e. $K_{j_1} = K_i$.
(2) For $\ell > 1$ assign $K_{j_\ell} = K_r$ for

$$r \in \left\{\arg \max_{k \in \{1,\ldots,m\} \setminus \{j_1,\ldots,j_{\ell-1}\}} |K_k \cap (K_{j_1} \cup \ldots \cup K_{j_{\ell-1}})|\right\}$$

arbitrarily. If the resulting ordering does not meet RIP, then $\{K_1, \ldots, K_m\}$ cannot be ordered to meet RIP.

As said in Theorem 2.7, all compositional models can be transformed into perfect models without influencing the represented multidimensional distribution (see the perfectization procedure described at the end of the preceding section). Moreover, both the models (before and after the transformation) are defined by the same number of parameters (probabilities). Unfortunately, no similar assertion holds for the transformation of a compositional into a decomposable model. For this transformation, one has to pay by the increase of necessary parameters, and simultaneously, by the loss of some information encoded in the structure of the model. Let us now describe the process of such a transformation and its influence on the model structure in more detail. For this, we employ the apparatus and results introduced in Jiroušek and Kratochvíl (2015).

First, recall the definition of non-trivial sets.

Definition 4.6: Let $\mathcal{P} = (K_1, \ldots, K_m)$ be the structure of a compositional model $\pi = \kappa_1 \triangleright \ldots \triangleright \kappa_m$. We say that set of variables U is non-trivial with respect to $\mathcal{P}$ if there exists $K_i \in \mathcal{P}$ (we will say that K_i generates U) such that $U \subseteq K_i$ and $U \setminus (K_1 \cup \ldots \cup K_{i-1}) \neq \emptyset$. Denote

$$\mathcal{N}(\mathcal{P}) = \{U \text{ is non-trivial with respect to } \mathcal{P} : 2 \leq |U| \leq 3\}.$$

It was shown in Kratochvíl (2013) that $\mathcal{N}(\mathcal{P})$ bears all the information about the structural conditional independence relations holding for the respective model. Among others, one special combination deserves our attention: a triplet $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$ such that $\{u, v\} \notin \mathcal{N}(\mathcal{P})$. Note that the existence of such a triplet prevents $\mathcal{P}$ to meet running intersection property. Therefore, a compositional model with such a triplet among its non-trivial sets cannot be decomposable. This is the key property on which the transformation process into a decomposable form is based. If there is triplet $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$ such that $\{u, v\} \notin \mathcal{N}(\mathcal{P})$, we will add $\{u, v\}$ into $\mathcal{N}(\mathcal{P})$ using the following assertion. This modification complies with our effort to minimize the changes in the model structure, though other non-trivial sets of cardinality 3 can simultaneously be added to $\mathcal{N}(\mathcal{P})$. For each variable $u \in K_1 \cup \ldots \cup K_m$, define $\lfloor u \rfloor = \min\{i : u \in K_i\}$.

Lemma 4.7: *Let $\mathcal{P} = (K_1, \ldots, K_m)$ be a compositional model structure, and a triplet $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$ such that $\{u, v\} \notin \mathcal{N}(\mathcal{P})$ and $\lfloor u \rfloor < \lfloor v \rfloor$. Let $R = K_{\lfloor v \rfloor} \setminus (K_1 \cup \ldots \cup K_{\lfloor v \rfloor - 1} \cup \{v\})$. Then,*

$$\mathcal{N}(\mathcal{P}') = \mathcal{N}(\mathcal{P}) \cup \{\{u, v\}\} \cup \{\{u, v, x\} : x \in K_{\lfloor v \rfloor} \setminus R, x \neq v\},$$

for structure $\mathcal{P}' = (K_1, \ldots, K_{\lfloor v \rfloor - 1}, (K_{\lfloor v \rfloor} \setminus R) \cup \{u\}, K_{\lfloor v \rfloor}, \ldots, K_m)$.

Proof: Note that $\mathcal{P}'$ is created from $\mathcal{P}$ by inserting

$$L = (K_{\lfloor v \rfloor} \setminus R) \cup \{u\} = \big(K_{\lfloor v \rfloor} \cap (K_1 \cup \ldots \cup K_{\lfloor v \rfloor - 1})\big) \cup \{u, v\}$$

between $K_{\lfloor v \rfloor - 1}$ and $K_{\lfloor v \rfloor}$. Therefore,

$$L \setminus (K_1 \cup \ldots \cup K_{\lfloor v \rfloor - 1}) = \{v\}. \tag{9}$$

To prove this Lemma, realize that non-trivial sets generated by $K_i, i \neq \lfloor v \rfloor$ are the same in both $\mathcal{P}$ and $\mathcal{P}'$. Indeed, this is clear for $i < \lfloor v \rfloor$ and to show it for $i > \lfloor v \rfloor$ it is enough to realize that $L \cup K_{\lfloor v \rfloor} \subseteq (K_1 \cup \ldots \cup K_{\lfloor v \rfloor})$. Thus one can concentrate on non-trivial sets generated by $K_{\lfloor v \rfloor}$ in $\mathcal{P}$ and non-trivial sets generated by L and $K_{\lfloor v \rfloor}$ in $\mathcal{P}'$.

First, let us show that $\mathcal{N}(\mathcal{P}') \supseteq \mathcal{N}(\mathcal{P}) \cup \{\{u, v\}\} \cup \{\{u, v, x\} \mid x \in K_{\lfloor v \rfloor} \setminus R, x \neq v\}$. Note that $v \in L \setminus (K_1 \cup \ldots \cup K_{\lfloor v \rfloor - 1})$, $\{u, v\} \subseteq L$, and therefore $\{u, v\} \subseteq \mathcal{N}(\mathcal{P}')$, since it is generated by L. Similarly, $\{u, v, x\}$ for any $x \in K_{\lfloor v \rfloor} \setminus (R \cup \{v\})$ is also generated by L in $\mathcal{P}'$.

Consider any $U \in \mathcal{N}(\mathcal{P})$ generated by $K_{\lfloor v \rfloor}$ in $\mathcal{P}$.

- If $U \subseteq L$, then $U \setminus (K_1 \cup \ldots \cup K_{\lfloor v \rfloor - 1}) = \{v\}$ because of $U \in \mathcal{N}(\mathcal{P})$ and Equation (9). Hence, U is generated by L in $\mathcal{P}'$, and therefore $U \in \mathcal{N}(\mathcal{P}')$.
- If $U \not\subseteq L$, then $U \cap R \neq \emptyset$ (because U is generated by $K_{\lfloor v \rfloor}$ in $\mathcal{P}$), which means that U is generated by $K_{\lfloor v \rfloor}$ also in $\mathcal{P}'$.

To finish the proof, we have also to show that the opposite inclusion holds. Assume $U \in \mathcal{N}(\mathcal{P}')$ generated either by L or $K_{\lfloor v \rfloor}$ in $\mathcal{P}'$.

- If U is generated by L in $\mathcal{P}'$, Equation (9) implies that $v \in U$. If $u \notin U$, then $U \subseteq K_{\lfloor v \rfloor}$ and U is generated by $K_{\lfloor v \rfloor}$ in $\mathcal{P}$. If $u \in U$, then either $U = \{u, v\}$, or $U = \{u, v, x\}$ for some $x \in (K_{\lfloor v \rfloor} \setminus R)$ such that $x \neq v$.

- If U is generated by $K_{\lfloor v \rfloor}$ in $\mathcal{P}'$, then it has to be generated by $K_{\lfloor v \rfloor}$ in $\mathcal{P}$ as well, which finishes the proof. ■

Lemma 4.7 constitutes a theoretical background, on which a simple heuristic[7] procedure is based, the procedure transforming a general compositional model into a decomposable one. This assertion gives instructions how to modify a structure $\mathcal{P}$ of a model if it is non-decomposable. It is obvious that a subsequent application of this modification to a general structure results in a decomposable structure. Namely, the process cannot cycle, because the application of Lemma 4.7 increases $\mathcal{N}(\mathcal{P})$ by one couple $\{u, v\}$, which cannot be removed in subsequent steps. Therefore (considering finite models) one must reach a situation that there is no triplet $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$ such that $\{u, v\} \notin \mathcal{N}(\mathcal{P})$. And the non-existence of such a triplet characterizes decomposable models – see Lemma 7.6 in (Jiroušek and Kratochvíl 2015).

4.3. Example

John does not use municipal transport regularly, and from time to time, he does not buy a ticket. Thus if being caught by a ticket inspector as a fare-dodger, he has to pay a fine. All possible situations are described with the help of five binary variables with values $\{yes, no\}$. Their meaning is: pay a fine f, a day off (holiday) h, meet a ticket inspector i, public transport t, and traveling to university u.

Consider the following compositional model:

$$\pi(f, h, i, t, u) = \kappa_1(h, i) \triangleright \kappa_2(h, u) \triangleright \kappa_3(t, u) \triangleright \kappa_4(f, i, t), \qquad (10)$$

where $\kappa_1(h, i)$ is a probability distribution describing the different activities of ticket inspectors on working days and holidays. Similarly, $\kappa_2(h, u)$ expresses the fact that on working days John's destination is usually the university, whilst on holidays he quite often goes to the countryside. The dependence of the destination and the mean of transport is described by $\kappa_3(t, u)$. The last probability distribution $\kappa_4(i, t, f)$ represents how often he has to pay a fine (naturally, for this he has to travel by means of public transport and he has to meet a ticket inspector). The respective probability distributions are in the upper part of Table 1, from which the reader can easily verify that the respective model is perfect. Let the goal be to compute a conditional distribution of variable h under the assumption that John pays a fine on his journey to the university, i.e. $\pi^{\{h\}|f=yes,u=yes}$.

Not knowing anything about the flexibility of this model, let us first convert it into a decomposable form (the reader can test its non-decomposability using *Maximum cardinality search* algorithm described above). To apply the idea from Lemma 4.7 one has to make a list of all non-trivial sets from $\mathcal{N}(\mathcal{P})$ for the structure of the considered model. In this example, the considered structure is $\mathcal{P} = (\{h, i\}, \{h, u\}, \{t, u\}, \{f, i, t\})$. $\mathcal{N}(\mathcal{P})$ contains all the sets from the considered structure (each of them contains a variable, which in not among the preceding sets, and the cardinalities of all these sets are either two or three) plus two more pairs:

$$\mathcal{N}(\mathcal{P}) = \big\{\{h, i\}, \{h, u\}, \{t, u\}, \{f, i, t\}, \{f, i\}, \{f, t\}\big\}.$$

From this, we immediately see that the model is not decomposable because $\{f, i, t\} \in \mathcal{N}(\mathcal{P})$ and $\{i, t\} \notin \mathcal{N}(\mathcal{P})$. Thus Lemma 4.7 advises to consider a new structure $\mathcal{P}' =$

$(\{h,i\},\{h,u\},\{i,t,u\},\{t,u\},\{f,i,t\})$ created from $\mathcal{P}$ by inserting set

$$L = (\{t,u\} \setminus \emptyset\}) \cup \{i\} = \{i,t,u\}$$

between $\{h,u\}$ and $\{t,u\}$. For this new structure $\mathcal{P}'$,

$$\mathcal{N}(\mathcal{P}') = \big\{\{h,i\},\{h,u\},\{i,t,u\},\{i,t\},\{t,u\},\{f,i,t\},\{f,i\},\{f,t\}\big\}.$$

Neither structure $\mathcal{P}'$ is decomposable; this is disproved by the existence of $\{i,t,u\} \in \mathcal{N}(\mathcal{P}')$ and $\{i,u\} \notin \mathcal{N}(\mathcal{P}')$. Therefore, applying Lemma 4.7 once more, we consider a new structure $\mathcal{P}'' = (\{h,i\},\{h,i,u\},\{h,u\},\{i,t,u\},\{t,u\},\{f,i,t\})$ created from $\mathcal{P}'$ by inserting set

$$L' = (\{h,u\} \setminus \emptyset\}) \cup \{i\} = \{h,i,u\}$$

between $\{h,i\}$ and $\{h,u\}$. The decomposability of this new structure can be seen from the fact that in the system

$$\mathcal{N}(\mathcal{P}'') = \big\{\{h,i\},\{h,i,u\},\{i,u\},\{h,u\},\{i,t,u\},\{i,t\},\{u,t\},\{f,i,t\},\{f,i\},\{f,t\}\big\}$$

there is no a three-element set, two-element subset of which would not appeared in the system.

To get a decomposable model corresponding to distribution π we have to compute distributions corresponding to the received decomposable structure, i.e.

$$\kappa_5(h,i,u) = \kappa_1(h,i) \triangleright \kappa_2(h,u),$$
$$\kappa_6(i,t,u) = \kappa_5^{\downarrow\{i,u\}}(i,u) \triangleright \kappa_3(t,u),$$

(see the lower part of Table 1) and the decomposable model corresponding to the original model given in Equation (10) is

$$\begin{aligned}\pi(f,h,i,t,u) &= \kappa_1(h,i) \triangleright \kappa_5(h,i,u) \triangleright \kappa_2(h,u) \triangleright \kappa_6(i,t,u) \triangleright \kappa_3(t,u) \triangleright \kappa_4(f,i,t)\\ &= \kappa_5(h,i,u) \triangleright \kappa_6(i,t,u) \triangleright \kappa_4(f,i,t). \end{aligned} \tag{11}$$

The realized simplification is made possible due to Properties (4) and (3) of Theorem 2.3.

Table 1. Probability distributions defining model from Expression (10).

	$\mathbf{a} = yes$		$\mathbf{a} = no$	
	$\mathbf{b} = yes$	$\mathbf{b} = no$	$\mathbf{b} = yes$	$\mathbf{b} = no$
$\kappa_1(h = \mathbf{a}, i = \mathbf{b})$	.04	.36	.16	.44
$\kappa_2(h = \mathbf{a}, u = \mathbf{b})$	.10	.30	.45	.15
$\kappa_3(t = \mathbf{a}, u = \mathbf{b})$	.40	.20	.15	.25
$\kappa_4(f = \mathbf{a}, i = \mathbf{b}, t = yes)$	.20	0	0	.40
$\kappa_4(f = \mathbf{a}, i = \mathbf{b}, t = no)$	0	0	0	.40
$\kappa_5(h = \mathbf{a}, i = \mathbf{b}, u = yes)$	.01	.09	.12	.33
$\kappa_5(h = \mathbf{a}, i = \mathbf{b}, u = no)$	.03	.27	.04	.11
$\kappa_6(i = \mathbf{a}, t = \mathbf{b}, u = yes)$	.13	0	.27	.15
$\kappa_6(i = \mathbf{a}, t = \mathbf{b}, u = no)$	.07	0	.13	.25

Table 2. Conditional probability distributions.

	$\mathbf{a} = yes$		$\mathbf{a} = no$	
	$\mathbf{b} = yes$	$\mathbf{b} = no$	$\mathbf{b} = yes$	$\mathbf{b} = no$
$\lambda_1(h = \mathbf{a}, i = \mathbf{b}, u = yes)$	.018	.164	.218	.600
$\lambda_1(h = \mathbf{a}, i = \mathbf{b}, u = no)$	0	0	0	0
$\lambda_2(i = \mathbf{a}, t = \mathbf{b}, u = yes)$	.236	0	.491	.273
$\lambda_2(i = \mathbf{a}, t = \mathbf{b}, u = no)$	0	0	0	0
$\lambda_3(f = \mathbf{a}, i = \mathbf{b}, t = yes)$	.236	0	0	.491
$\lambda_3(f = \mathbf{a}, i = \mathbf{b}, t = no)$	0	0	0	.273

Table 3. Resulting conditional probability distribution.

	$\mathbf{a} = yes$	$\mathbf{a} = no$
$\pi^{\{h=\mathbf{a}\}\|f=yes,u=yes}$	.074	.924

Having a decomposable model defined by Expression (11), to compute the required $\pi^{\{h\}|f=yes,u=yes}$, proceed in the following two steps. First realize that even though the compositional model in Equation (11) is perfect, its conditional form

$$\left(\delta^u_{yes} \triangleright \kappa_5(h,i,u)\right) \triangleright \kappa_6(i,t,u) \triangleright \kappa_4(f,i,t)$$

is not perfect any more. Since decomposable models are flexible only when they are also perfect, we have to perfectize it. Thus consider a perfect model $\lambda_1 \triangleright \lambda_2 \triangleright \lambda_3$ with

$$\lambda_1(h,i,u) = \delta^u_{yes} \triangleright \kappa_5(h,i,u),$$
$$\lambda_2(i,t,u) = \lambda_1^{\downarrow\{i,u\}}(i,u) \triangleright \kappa_6(i,t,u),$$
$$\lambda_3(f,i,t) = \lambda_2^{\downarrow\{i,t\}}(i,t) \triangleright \kappa_4(f,i,t),$$

the values of which are in Table 2. This model, being perfect and decomposable, may be reordered (keeping the RIP property), so that the required conditional distribution

$$\pi^{\{h\}|f=yes,u=yes} = \left((\delta^f_{yes} \triangleright \lambda_3(f,i,t)) \triangleright \lambda_2(i,t,u) \triangleright \lambda_1(h,i,u)\right)^{\downarrow\{h\}},$$

the values of which are in Table 3.

4.4. Conditioning and intervention in causal models

In this section, we briefly show how to handle compositional models when they are interpreted as Markovian causal models. Originally, Markovian causal models were introduced in a graphical form by Pearl (2009). The results presented in this section are taken over from Bína and Jiroušek (2015), where the interested reader can find a more detailed explanation of the topic and some illustrative examples.

As in previous sections, we keep considering compositional model $\kappa_1(K_1) \triangleright \kappa_2(K_2) \triangleright \ldots \triangleright \kappa_m(K_m)$. For each variable $u \in K_1 \cup \ldots \cup K_m$, define

$$\mathfrak{C}(u) = K_{\lfloor u \rfloor} \cap (K_1 \cup \ldots \cup K_{\lfloor u \rfloor - 1}),$$

(recall that $\lfloor u \rfloor = \min\{i : u \in K_i\}$), which obviously means that $\mathfrak{C}(u) = \emptyset$ for all $u \in K_1$.

$\mathfrak{C}(u)$ is interpreted as a set of *causes* of variable u, and the considered model $\kappa_1(K_1) \triangleright \kappa_2(K_2) \triangleright \ldots \triangleright \kappa_m(K_m)$ is called *causal compositional model*. Thus, as the reader certainly noticed, we consider only causal models without feedback.

When using a causal compositional model for inference, we can take advantage of most of the properties presented in previous sections. The only difference is that it is not possible to use operations changing the ordering of distributions in a model, one cannot employ Property (9) of Theorem 2.3. Namely, changing the ordering of distributions in a model may result in swapping the roles of cause and effect. Thus the flexibility of a model cannot be used, either.

As explained by Pearl (2009), in causal models one has an additional possibility that does not have a sense in non-causal models. It, in a way, resembles conditioning but its semantic is quite different. It is a possibility of intervention. Let us explain this notion with the following simple example.

Example 4.8: Consider two binary variables: s indicating whether there is smoke in a room ($s = 1$) or not ($s = 0$), and variable a, which equals 1 if the fire alarm is on, and equals 0 if it is off. The corresponding causal compositional model is $\pi(a, s) = \mu(s) \triangleright \nu(a, s)$. We can see that $\mathfrak{C}(s) = \emptyset$, and $\mathfrak{C}(a) = \{s\}$, which corresponds with our intuition. For illustration, consider probabilities $\mu(0) = \mu(1) = 0.5$, and $\nu(0, 0) = \nu(1, 1) = 0.45$, $\nu(0, 1) = \nu(1, 0) = 0.05$. From this, we can easily compute conditional distribution[8]

$$\pi^{a|s=0} = \left(\delta_0^s \triangleright (\mu(s) \triangleright \nu(a, s))\right)^{\downarrow\{a\}} = \left(\delta_0^s \triangleright \nu(a, s)\right)^{\downarrow\{a\}}, \tag{12}$$

for which $\pi^{a|s=0}(0) = 0.9$, and $\pi^{a|s=0}(1) = 0.1$. Analogously, we can compute $\pi^{a|s=1}(0) = 0.1$, and $\pi^{a|s=1}(1) = 0.9$. Naturally, the model also makes the computation of conditionals $\pi^{s|a=0}$ and $\pi^{s|a=1}$ possible (notice, the simplification performed in Formula (12) cannot be done, now):

$$\pi^{s|a=0} = \left(\delta_0^a \triangleright (\mu(s) \triangleright \nu(a, s))\right)^{\downarrow\{s\}}.$$

From this, we get $\pi^{s|a=0}(0) = 0.9$, $\pi^{s|a=0}(1) = 0.1$, and $\pi^{s|a=1}(0) = 0.1$, $\pi^{s|a=1}(1) = 0.9$.

In addition to the above-computed conditionals, causal models also enable us to compute the effect of *interventions*. By (external) intervention, we understand a forced change of the modeled system. In our example, we can realize an intervention to variable s, for example, by smoking a cigar, which fills the room with smoke. An intervention to variable a means that we disconnect the alarm from the smoke sensors and make it wail, for example, by pushing a test button. It makes the alarm wailing regardless of whether there is smoke in the room or not. In the mathematical description, such interventions are denoted using Perl's *do operator*. Thus smoking a cigar in the room means that we set the value of variable s to 1; in notation, $do(s = 1)$. Analogously, an intervention that makes the alarm wail is denoted $do(a = 1)$.

In the rest of this section we show, how to compute the effect of intervention, i.e. how to compute expressions like $\pi^{a|do(s=1)}$ and $\pi^{s|do(a=1)}$. We show that, while the intervention $do(s = 1)$ causes the alarm wail, the intervention to alarm, i.e. $do(a = 1)$, does not fill the room with smoke.

An intervention realizes an external change. It changes forcefully the value of an intervened variable regardless the causal relations encoded in the model. So, it breaks all the links of the intervened variable with its causes. How does such intervention to variable u influence a causal compositional model? As shown in Bína and Jiroušek (2015), the interruption of all the respective links of u with its causes in the compositional model representation means that the model $\pi = \kappa_1(K_1) \triangleright \kappa_2(K_2) \triangleright \ldots \triangleright \kappa_m(K_m)$ changes to

$$\kappa_{\lfloor u \rfloor}(u) \triangleright \kappa_1(K_1) \triangleright \kappa_2(K_2) \triangleright \ldots \triangleright \kappa_m(K_m). \tag{13}$$

Thus the effect of intervention $do(u = \mathbf{a})$ computed from a causal model $\kappa_1(K_1) \triangleright \kappa_2(K_2) \triangleright \ldots \triangleright \kappa_m(K_m)$ corresponds to the computation of a conditional distribution from the model represented by Formula (13). That is,

$$\begin{aligned}
\pi^{L|do(u=\mathbf{a})} &= \left(\delta^u_{\mathbf{a}} \triangleright \left(\kappa_{\lfloor u \rfloor}(u) \triangleright \kappa_1(K_1) \triangleright \kappa_2(K_2) \triangleright \ldots \triangleright \kappa_m(K_m)\right)\right)^{\downarrow L} \\
&= \left(\delta^u_{\mathbf{a}} \triangleright \kappa_{\lfloor u \rfloor}(u) \triangleright \kappa_1(K_1) \triangleright \kappa_2(K_2) \triangleright \ldots \triangleright \kappa_m(K_m)\right)^{\downarrow L} \\
&= \left(\delta^u_{\mathbf{a}} \triangleright \kappa_1(K_1) \triangleright \kappa_2(K_2) \triangleright \ldots \triangleright \kappa_m(K_m)\right)^{\downarrow L}.
\end{aligned} \tag{14}$$

Thus comparing Expression (14) with the statement of Theorem 4.1, one can see that the computation of the conditional distribution and the computation of the effect of intervention, when done in causal compositional models, differ from each other just in a pair of parentheses:

$$\pi^{L|u=\mathbf{a}} = \left(\delta^u_{\mathbf{a}} \triangleright \left(\kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m\right)\right)^{\downarrow L}, \tag{15}$$

$$\pi^{L|do(u=\mathbf{a})} = \left(\delta^u_{\mathbf{a}} \triangleright \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_m\right)^{\downarrow L}. \tag{16}$$

Example 4.9 (continued): To finish this example, let us compute the effects of the interventions $do(s = 1)$ and $do(a = 1)$. For the former case, we get

$$\pi^{a|do(s=1)} = \left(\delta^s_1 \triangleright \mu(s) \triangleright \nu(a,s)\right)^{\downarrow\{a\}} = \left(\delta^s_1 \triangleright \nu(a,s)\right)^{\downarrow\{a\}}, \tag{17}$$

for which $\pi^{a|do(s=1)}(0) = 0.1$, and $\pi^{a|do(s=1)}(1) = 0.9$. However, computing the effect of the intervention $do(a = 1)$ yields

$$\pi^{s|do(a=1)} = \left(\delta^a_1 \triangleright \mu(s) \triangleright \nu(a,s)\right)^{\downarrow\{s\}} = \left(\delta^a_1 \triangleright \mu(s)\right)^{\downarrow\{s\}},$$

and thus we get $\pi^{s|a=1}(0) = 0.5, \pi^{s|a=1}(1) = 0.5$.

When speaking about conditioning, we also paid attention to answering a question about how to proceed when one wants to compute a conditional with multiple conditions. An analogous question may be raised in connection to multiple interventions. But it turns out it is an easy task. Consider a causal compositional model $\pi = \kappa_1(K_1) \triangleright \kappa_2(K_2) \triangleright \ldots \triangleright \kappa_m(K_m)$, and two variables $u, v \in K_1 \cup \ldots \cup K_m$ and their values $\mathbf{a} \in \mathbf{X}_u$, $\mathbf{c} \in \mathbf{X}_v$. Then for $L \subseteq (K_1 \cup \ldots \cup K_m) \setminus \{u, v\}$

$$\pi^{L|do(u=\mathbf{a}),do(v=\mathbf{c})} = \left(\delta^u_{\mathbf{a}} \triangleright \delta^v_{\mathbf{c}} \triangleright \kappa_1 \triangleright \ldots \triangleright \kappa_m\right)^{\downarrow L} = \left(\delta^{(u,v)}_{(\mathbf{a},\mathbf{c})} \triangleright \kappa_1 \triangleright \ldots \triangleright \kappa_m\right)^{\downarrow L}.$$

5. Conclusion

Compositional models provide a tool for efficient representation of multidimensional probability distributions. This paper is the last from a series of three papers surveying the theoretical background of this theory – the preceding ones are Jiroušek (2011); Jiroušek and Kratochvíl (2015). As shown in this paper, the respective theory supports also the design of computational procedures, the efficiency of which is comparable with those designed for Bayesian network handling. The advantage of compositional models manifests best when being applied to causal models (described in Section 4.4). When computing the effect of an intervention, one need not change the model and gets the result simply using Formula (16).

Another advantage follows from the fact that these models can also be used in the framework of other uncertainty theories. For example, within all those falling under the unifying Shenoy's theory of Valuation-based systems (i.e. a version of possibility theory where the combination is the product t-norm, Spohn's epistemic belief theory, Dempster–Shafer belief function theory, and others) (Shenoy 1989; Jiroušek and Shenoy 2012). For the implementation of compositional model theory, it is enough to find a binary operator meeting Properties (1), (2), (6) and (7) of Theorem 2.3. In this way, we partially answer the question asked by the anonymous referee of Jiroušek (2011): what are the axioms that must be met by a general operator of composition? The above-mentioned properties are sufficient, but what is the minimum system is still an open problem. For example, it can be shown that Property (7) may be substituted by Property (8) and a Simplified associativity rule (under the assumption of Theorem 2.3): Let $L \supseteq (K \cap M)$ then, $(\kappa \triangleright \lambda) \triangleright \mu = \kappa \triangleright (\lambda \triangleright \mu)$.

Notes

1. This and similar formulas could be more precisely expressed: for all $\mathbf{a} \in \mathbf{X}_{K\cap L}$, $\kappa^{\downarrow K\cap L}(\mathbf{a}) = \lambda^{\downarrow K\cap M}(\mathbf{a})$. In what follows we prefer simpler, more lucid formulas like, e.g. (1).
2. In this paper, we consider $\frac{0\cdot 0}{0} = 0$.
3. In Jiroušek (2011), a different definition was introduced. The presented definition is equivalent to the original one due to Theorem 10.3 of the cited paper.
4. In Jiroušek and Kratochvíl (2015), there are two definitions of flexibility. The flexibility studied in this paper correspond to Definition 7.2 of the cited paper. The other flexibility, called *structural flexibility* (Definition 7.3 of the cited paper) is much stronger.
5. These models are equivalent to decomposable models known from Bayesian network theory (Lauritzen, Speed, and Vijayan 1984).
6. Malvestuto considers a much more general framework, he does not restrict his consideration only to probability distributions, so we are simplifying his general definition.
7. It is known (Arnborg, Corneil, and Proskurowski 1987; Kjærulff 1990) that optimal triangulation of a general graph is an NP-hard problem. Thus, finding an optimal decomposable model is an NP-hard problem, too. Nevertheless, similarly to Bayesian network theory, where several heuristic algorithms for decomposable model construction were designed (see e.g. Cano and Moral (1994)) other heuristics for this purpose may be applied.
8. The following simplification is possible due to Properties (7) and (3) of Theorem 2.3.

Acknowledgments

The paper was written under the support from the Czech Science Foundation – Grant No. 19-06569S.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The paper was written under the support from the Czech Science Foundation – Grant No. 19-06569S.

Notes on contributors

V. Bína graduated in Computer Physics at Czech Technical University in Prague in 2003. He received his PhD in Management at Prague University of Economics and Business in 2011 where he currently holds an associate professor position at Faculty of Management. Among his research interests ranks analysis of an income situation and poverty in EU countries and probabilistic and generalized approaches to modelling of uncertainty together with other statistical approaches supporting managerial decision making.

R. Jiroušek graduated in Mathematics from Charles University (Prague) in 1969. He received his PhD and DrSc degrees from the Czech Academy of Sciences respectively in 1979 and 1993. His main scientific interests relate to decision-making under uncertainty (complex models of uncertain knowledge). Though he retired in 2016, he still keeps research contacts with the Faculty of Management, located in Jindřichův Hradec, and with the Institute of Information Theory and Automation of the Czech Academy. Currently, he serves also as the Editor in Chief of the *Kybernetika* journal.

V. Kratochvíl obtained his PhD degree in Artificial Intelligence from the Czech Technical University in Prague. He spent 1 year at Concordia University in Montréal, Canada, where he worked as a postdoc in the team of Vašek Chvátal. He serves currently as a research associate at the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic and at the Faculty of Management, University of Economics, Jindřichův Hradec. His researcher interests are Bayesian Networks, Belief Functions, and various problems in algorithmization.

References

Arnborg, Stefan, Derek G. Corneil, and Andrzej Proskurowski. 1987. "Complexity of Finding Embeddings in a *k*-Tree." *SIAM Journal on Algebraic Discrete Methods* 8 (2): 277–284.

Bína, Vladislav, and Radim Jiroušek. 2006. "Marginalization in Multidimensional Compositional Models." *Kybernetika* 42 (4): 405–422.

Bína, Vladislav, and Radim Jiroušek. 2015. "On Computations with Causal Compositional Models." *Kybernetika* 51 (3): 525–539.

Cano, Andrés, and Serafin Moral. 1994. "Heuristic Algorithms for The Triangulation of Graphs." In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 98–107. New York: Springer.

Golumbic, Martin Charles. 1980. *Algorithmic Graph Theory and Perfect Graphs*. New York: ACAD Press.

Jiroušek, Radim. 2000. "Marginalization in Composed Probabilistic Models." In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, 301–308. Burlington, MA: Morgan Kaufmann Publishers Inc.

Jiroušek, Radim. 2002. "Decomposition of Multidimensional Distributions Represented by Perfect Sequences." *Annals of Mathematics and Artificial Intelligence* 35 (1–4): 215–226.
Jiroušek, Radim. 2011. "Foundations of Compositional Model Theory." *International Journal of General Systems* 40 (6): 623–678.
Jiroušek, Radim. 2017. "On Conditioning in Multidimensional Probabilistic Models." In *Robustness in Econometrics*, 201–216. New York, NY: Springer.
Jiroušek, Radim. 2020. "A Short Note on Decomposition and Composition of Knowledge." *International Journal of Approximate Reasoning* 120: 24–32.
Jiroušek, Radim, and Václav Kratochvíl. 2015. "Foundations of Compositional Models: Structural Properties." *International Journal of General Systems* 44 (1): 2–25.
Jiroušek, Radim, and Prakash P. Shenoy. 2012. "Compositional Models in Valuation-Based Systems." In Belief Functions: Theory and Applications, 221–228. New York, NY: Springer.
Kellerer, Hans G. 1964a. "Masstheoretische Marginalprobleme." *Mathematische Annalen* 153 (3): 168–198.
Kellerer, Hans G. 1964b. "Verteilungsfunktionen Mit Gegebenen Marginalverteilungen." *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete* 3 (3): 247–270.
Kjærulff, Uffe. 1990. "Triangulation of Graphs-Algorithms Giving Total State Space. r90-09, Department of Mathematics and Computer Science. *Institute for Electronic Systems, Aalborg University*.
Kratochvíl, Václav. 2013. "Probabilistic Compositional Models: Solution of An Equivalence Problem." *International Journal of Approximate Reasoning* 54 (5): 590–601.
Lauritzen, Steffen L., T. P. Speed, and K. Vijayan. 1984. "Decomposable Graphs and Hypergraphs." *Journal of the Australian Mathematical Society* 36 (1): 12–29.
Malvestuto, Francesco M. 2015. "Marginalization in Models Generated by Compositional Expressions." *Kybernetika* 51 (4): 541–570.
Malvestuto, Francesco M. 2016. "Compositional Models, Bayesian Models and Recursive Factorization Models." *Kybernetika* 52 (5): 696–723.
Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. New York, NY: Cambridge University Press.
Shachter, Ross D. 1986. "Evaluating Influence Diagrams." *Operations Research* 34 (6): 871–882.
Shachter, Ross D. 1988. "Probabilistic Inference and Influence Diagrams." *Operations Research* 36 (4): 589–604.
Shenoy, Prakash P. 1989. "A Valuation-Based Language for Expert Systems." *International Journal of Approximate Reasoning* 3 (5): 383–411.
Tarjan, Robert E., and Mihalis Yannakakis. 1984. "Simple Linear-Time Algorithms to Test Chordality of Graphs, Test Acyclicity of Hypergraphs, and Selectively Reduce Acyclic Hypergraphs." *SIAM Journal on Computing* 13 (3): 566–579.