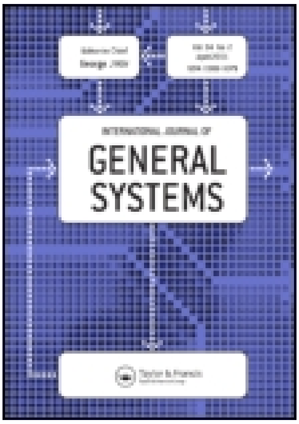


International Journal of General Systems
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ggen20

Foundations of compositional model theory

Radim Jiroušek [a] [b]

[a] Faculty of Management, University of Economics, Prague, Jarošovská 1117/II, 377 01, Jindřichu˚v Hradec, Czech Republic

[b] Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Czech Republic

Available online: 30 Jun 2011

To cite this article: Radim Jiroušek (2011): Foundations of compositional model theory, International Journal of General Systems, 40:6, 623-678

To link to this article: http://dx.doi.org/10.1080/03081079.2011.562627

International Journal of General Systems
Vol. 40, No. 6, August 2011, 623–678

Foundations of compositional model theory

Radim Jiroušek[a,b]*

[a]*Faculty of Management, University of Economics, Prague, Jarošovská 1117/II, 377 01 Jindřichův Hradec, Czech Republic;* [b]*Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Czech Republic*

(*Received 3 September 2010; final version received 1 January 2011*)

Graphical Markov models, most of all Bayesian networks, have become a very popular way for multidimensional probability distribution representation and processing. What makes representation of a very-high-dimensional probability distribution possible is its independence structure, i.e. a system of conditional independence relations valid for the distribution in question. The fact that some of independence systems can be successfully represented with the help of graphs is reflected in the general title: graphical modelling. However, graphical representation of independence structures is also associated with some disadvantages: only a small part of different independence structures can be faithfully represented by graphs; and still one structure is usually equally well represented by several graphs. These reasons, among others, initiated development of an alternative approach, called here *theory of compositional models*, which enables us to represent exactly the same class of distributions as Bayesian networks. This paper is a survey of the most important basic concepts and results concerning compositional models necessary for reading advanced papers on computational procedures and other aspects connected with this (relatively new) approach for multidimensional distribution representation.

Keywords: multidimensional probability distribution; conditional independence; graphical Markov model; composition of distributions

1. Introduction

A number of different models for knowledge representation have been developed. When uncertain knowledge is considered – and in our opinion, deterministic knowledge applies to very specific situations only – one has to consider models based on some of the calculi proposed specifically for this purpose. The oldest one is probability theory but many others appeared in the second half of the last century, from many-valued and fuzzy logics, through rough sets theory to approaches based on non-additive measures, e.g. possibility theory.

In this paper, we shall discuss one class of models built within the framework of probability theory. However, it should be stressed that these models can also be developed equally efficiently in possibility theory (Vejnarová 1998a, 1998b, Dubois and Prade 2001, Klir 2006) and in the framework of Dempster–Shafer theory of evidence (Dempster 1967, Shafer 1976, Jiroušek *et al.* 2007). This means that they can also be applied to situations when the assumption of additivity is not adequate (Jiroušek and Vejnarová 2003). Nevertheless, in this paper, we shall restrict our consideration only to probabilistic models.

*Email: radim@utia.cas.cz

ISSN 0308-1079 print/ISSN 1563-5104 online

DOI: 10.1080/03081079.2011.562627
http://www.informaworld.com

The basic idea of the approach described here is the same as that on which expert systems are based; it is beyond human capabilities to *represent/express/understand* global knowledge of an application area – one always has to work only with pieces of local knowledge. Such local knowledge can be, within probability theory, easily represented by an oligodimensional (low-dimensional) distribution.

What should be stressed, however, is the fact that in such situations the dimensionality of the estimated distributions is strictly limited also because of the size of available data. Whatever size of data is at our disposal, we can hardly expect to obtain reliable estimates of probabilities of a 20-dimensional distribution (even for binary variables). Typically, one can assume that dimensionality of the considered distributions is between two and eight. This is why we call them *oligodimensional distributions*.

When pieces of local knowledge are represented by oligodimensional distributions, the global knowledge should be represented by a multidimensional probability distribution. In artificial intelligence, application of the whole class of methods based on knowledge modelling by multidimensional probability distributions – and here, we have in mind distributions of hundreds rather than tens of variables – was catalysed by successes achieved during the last three decades in the field known as *graphical Markov modelling* (Pearl 1988, Lauritzen 1996). This term is used to describe any of the approaches representing multidimensional probability distributions by means of graphs and systems of quantitative parameters. These parameters are usually oligodimensional, sometimes conditional, probability distributions. Therefore, graphical Markov modelling includes influence diagrams, decomposable and graphical models, chain graph models, and many other models. What is common to all of them is the capability to represent and process distributions of very high dimensionality, which cannot be otherwise handled because of the exponential growth of the number of necessary parameters. Perhaps the most famous representative of these models, Bayesian networks (Jensen 2001), represents distributions having special dependence structures, which are described by acyclic directed graphs. Some other models, like decomposable models, use undirected graphs for the dependence structure representation, and special models need even more complicated graphical tools like chain graphs, hypergraphs, or annotated (di)graphs.

The approach presented herein abandons the necessity to describe the dependence structure of a modelled distribution using a graph. In contrast to this, the presented technique of compositional models describes directly how the multidimensional distribution is computed – *composed* – from a system of low-dimensional distributions and, therefore, need not represent the dependence structure explicitly. Thus, we start describing our model with an assumption that there are a (usually great) number of pieces of local knowledge represented by a system of low-dimensional distributions. The task we will address in this paper resembles a jig-saw puzzle that has a large number of parts, each bearing a local piece of a picture, and the goal is to figure out how to assemble them in such a way that the global picture makes sense, reflecting all of the individual small parts. The only difference is that, in our case, we will look for a linear ordering of oligodimensional distributions in the way that, when composed together, the resulting multidimensional distribution optimally reflects all of the local knowledge carried by the oligodimensional distributions (Figure 1).

This paper is a survey making an introduction to compositional models in probability theory. Here, we will define the most important notion of this approach, the operators of composition, which were originally introduced by Jiroušek (1997), and selected results published by Jiroušek (1998, 2000), especially those describing how these operators are used to construct multidimensional models and what the properties are of the resulting models. The paper includes parts that have not yet been published (e.g. flexible sequences

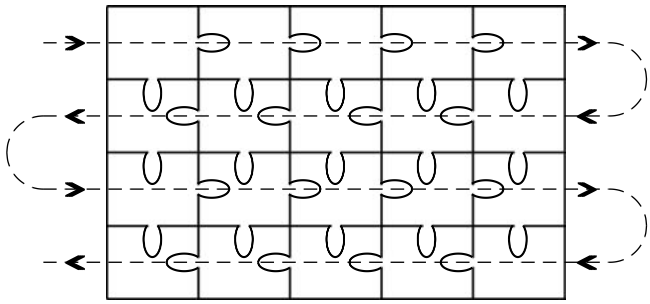


Figure 1. Ordering of pieces of a jig-saw puzzle.

and their properties), but due to space limitations we could not include some other advanced results. For example, we could not include sections describing identification of the conditional independence structure of the models, which can be performed with the help of so-called *persegrams* (Jiroušek 2008), or description of computationally efficient algorithms that were designed for compositional models (Jiroušek 2000).

The paper is organized as follows. Section 2 introduces notation and general basic concepts used from probability theory. In Section 3, operator of composition and its basic properties are introduced. More advanced properties of this operator are also studied in Sections 4, 5, 7, and 8. In connection with this, we want to stress that the reader cannot expect new results of probability theory; just using the well-known properties of this theory, we show (algebraic) properties of the operator of composition. We show that, despite the fact that it is generally neither commutative nor associative, in some special situations these properties are observed. Lack of associativity also called for definition of an anticipating operator, whose definition and basic properties are described in Section 9. Its purpose and importance manifest mainly in studies of computational procedures; yet, the basic notions of compositional models would not be complete if this operator was omitted. So, although it is not frequently used in this paper, we devote a short section to it. An application of the operators of composition to representation of multidimensional probability distributions is described in Sections 10–13, with each of them applied to one special class of models.

A special role is designated to Section 6. We believe that by presenting Imre Csiszár's point of view, we can help some readers to form their own intuitive image that helps them to understand the approach of compositional models.

While all the necessary concepts are introduced in this paper, the reader is expected to be familiar with basic notions of (finite) probability theory. With respect to the fact that the described approach forms an algebraic alternative to Bayesian networks, some knowledge of graphical Markov modelling is advantageous, but not necessary to fully understand the text (a formalized comparison of compositional models and graphical Markov models can be found in Jiroušek's (2004a, 2004b) papers).

2. Basic notions of probability theory and notation

In this paper, we will deal with a finite system of finite-valued random variables. Let N be an arbitrary finite index set, $N \neq \emptyset$. Each variable from $\{X_i\}_{i \in N}$ is assumed to have a finite (non-empty) set of values $\mathbf{X}_i$. The set of all combinations of the considered values will be denoted $\mathbf{X}_N = \times_{i \in N} \mathbf{X}_i$. Analogously, for $K \subset N$, $\mathbf{X}_K = \times_{i \in K} \mathbf{X}_i$.

Distributions of the considered variables will be denoted by Greek letters (π, κ, ν, and μ) with possible indices; thus for $K \subseteq N$, we can consider a distribution $\pi((X_i)_{i \in K})$.

To make the formulae more lucid, the following simplified notation will be used: symbol $\pi(K)$ will denote a $|K|$-dimensional distribution and $\pi(x)$ is a value of probability distribution π for point $x \in \mathbf{X}_K$.

For a probability distribution $\pi(K)$ and $J \subset K$, we will often consider a *marginal distribution* $\pi^{\downarrow J}$ of π, which can be computed for all $x \in \mathbf{X}_J$ by

$$\pi^{\downarrow J}(x) = \sum_{y \in \mathbf{X}_K : y^{\downarrow J} = x} \pi(y),$$

where $y^{\downarrow J}$ denotes the *projection* of $y \in \mathbf{X}_K$ into $\mathbf{X}_J$. For computation of marginal distributions, we need not exclude situations when $J = \emptyset$. By definition, we get $\pi^{\downarrow \emptyset} = 1$.

Having two distributions $\pi(K)$ and $\kappa(K)$, we say that κ dominates π (in symbol $\pi \ll \kappa$) if for all $x \in \mathbf{X}_K$,

$$\kappa(x) = 0 \Longrightarrow \pi(x) = 0$$

(notice that some authors say that π is absolutely continuous with respect to κ; the latter notion is used mainly when considering continuous spaces).

2.1 *Conditional independence of variables*

One of the most important notions of this paper, a concept of *conditional independence*, generalizes the well-known independence of variables. Since it does not belong among the basic subjects notoriously repeated in all textbooks on probability theory, we introduce it here in the form of a definition, along with two of the most important properties that can be found in several books on probabilistic multidimensional models: for example, Lauritzen (1996) or Studený (2005).

DEFINITION 2.1. Consider a probability distribution $\pi(K)$ and three disjoint subsets $L, M, R \subseteq K$ such that both $L, M \neq \emptyset$. We say that groups of variables X_L and X_M are *conditionally independent* given X_R (in symbol $X_L \perp\!\!\!\perp X_M | X_R[\pi]$) if

$$\pi^{\downarrow L \cup M \cup R} \pi^{\downarrow R} = \pi^{\downarrow L \cup R} \pi^{\downarrow M \cup R}. \tag{1}$$

LEMMA 2.2 (FACTORIZATION LEMMA). Let $K, L, R \subset N$ be disjoint such that $K \neq \emptyset \neq L$. Then for any probability distribution $\pi(K \cup L \cup R)$,

$$X_K \perp\!\!\!\perp X_L | X_R[\pi],$$

if and only if there exist functions

$$\psi_1 : \mathbf{X}_{K \cup R} \longrightarrow [0, +\infty), \quad \psi_2 : \mathbf{X}_{L \cup R} \longrightarrow [0, +\infty),$$

such that for all $x \in \mathbf{X}_{K \cup L \cup R}$ $\pi(x) = \psi_1(x^{\downarrow K \cup R}) \psi_2(x^{\downarrow L \cup R})$.

LEMMA 2.2 (BLOCK INDEPENDENCE LEMMA). Let $K, L, M, R \subset N$ be disjoint and $K \neq \emptyset$, $L \neq \emptyset$, and $M \neq \emptyset$. Then for any probability distribution $\pi(K \cup L \cup M \cup R)$, the following two expressions are equivalent

(A) $X_K \perp\!\!\!\perp X_{L \cup M} | X_R[\pi]$,

(B) $X_K \perp\!\!\!\perp X_M | X_R[\pi]$ and $X_K \perp\!\!\!\perp X_L | X_{M \cup R}[\pi]$.

2.2 *Extensions of distributions*

Consider $K \subseteq L \subseteq N$ and a probability distribution $\pi(K)$. By $\Pi^{(L)}$, we shall denote the set of all probability distributions defined for variables X_L. Similarly, $\Pi^{(L)}(\pi)$ will denote the system of all *extensions* of the distribution π to L-dimensional distributions

$$\Pi^{(L)}(\pi) = \left\{\kappa \in \Pi^{(L)} : \kappa^{\downarrow K} = \pi\right\}.$$

Having a system $\Xi = \{\pi_1(K_1), \pi_2(K_2), \cdots, \pi_n(K_n)\}$, of oligodimensional distributions ($K_1 \cup \cdots \cup K_n \subseteq L$), the symbol $\Pi^{(L)}(\Xi)$ denotes the system of distributions that are extensions of all the distributions from Ξ,

$$\Pi^{(L)}(\Xi) = \left\{\kappa \in \Pi^{(L)} : \kappa^{\downarrow K_i} = \pi_i \forall i = 1, \cdots, n\right\} = \bigcap_{i=1}^{n} \Pi^{(L)}(\pi_i).$$

It is obvious that the set of extensions $\Pi^{(L)}(\Xi)$ is either empty or convex (naturally, a one-point set is convex, too).

3. Definition of operators of composition

To be able to compose low-dimensional distributions to get a distribution of a higher dimension, we will introduce two *operators of composition*.

First, let us introduce an operator $\triangleright$ of *right composition*. To make it clear from the very beginning, let us stress that it is just a generalization of the idea of computing the three-dimensional distribution from two two-dimensional ones introducing the conditional independence

$$\pi(x_1, x_2) \triangleright \kappa(x_2, x_3) = \frac{\pi(x_1, x_2)\kappa(x_2, x_3)}{\kappa(x_2)} = \pi(x_1, x_2)\kappa(x_3|x_2),$$

where $\kappa(x_2)$ and $\kappa(x_3|x_2)$ denote the corresponding marginal and conditional distributions, respectively.

Consider two probability distributions $\pi(K)$ and $\kappa(L)$, whose composition we want to define. Notice that we do not pose any condition on the relationship of the two sets of variables: X_K and X_L. Nevertheless, if these sets are not disjoint, it may happen (as will be illustrated in an example below) that the composition $\pi \triangleright \kappa$ does not exist. Therefore, we will assume that $\kappa^{\downarrow K\cap L}$ dominates $\pi^{\downarrow K\cap L}$ and the *right composition* of these two distributions is given by the formula

$$\pi \triangleright \kappa = \frac{\pi\kappa}{\kappa^{\downarrow K\cap L}}.$$

Since we assume $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$, if for any $x \in \mathbf{X}_{L\cup K}$ $\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L}) = 0$, then both $\pi(x^{\downarrow K})$ and $\kappa(x^{\downarrow L})$ equal 0, too. Hence, there is a product of two zeros in the numerator of this formula and we, quite naturally, take

$$\frac{0.0}{0} = 0.$$

If $K \cap L = \emptyset$, then $\kappa^{\downarrow K\cap L} = 1$, and the formula degenerates to a simple product of π and κ (obviously, since in this case $\pi^{\downarrow K\cap L} = \kappa^{\downarrow K\cap L} = 1$, the condition $\pi(x_{K\cap L}) \ll \kappa(x_{K\cap L})$ holds true).

Let us stress that the expression $\pi \rhd \kappa$ remains undefined in the case that $\pi^{\downarrow L\cap K} \ll \kappa^{\downarrow K\cap L}$.

Thus, the formal definition of the operator $\rhd$ is as follows.

Definition 3.1. For two arbitrary distributions $\pi(K)$ and $\kappa(L)$, for which $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$, their *right composition* is, for each $x \in \mathbf{X}_{(L\cup K)}$, given by the following formula:

$$(\pi \rhd \kappa)(x) = \frac{\pi(x^{\downarrow K})\kappa(x^{\downarrow L})}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})}.$$

In a case where $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$, the composition remains undefined.

The following simple assertion answers the question: what is the result of composition of two distributions?

Lemma 3.2. Let π and κ be probability distributions from $\Pi^{(K)}$ and $\Pi^{(L)}$, respectively. If $\pi^{\downarrow L\cap K} \ll \kappa^{\downarrow L\cap K}$ (i.e. if $\pi \rhd \kappa$ is defined), then $\pi \rhd \kappa$ is a probability distribution from $\Pi^{(L\cup K)}(\pi)$, i.e. it is a probability distribution and its marginal distribution for variables X_K equals π:

$$(\pi \rhd \kappa)^{\downarrow K} = \pi.$$

Proof. To show that $\pi \rhd \kappa$ is a probability distribution from $\Pi^{(L\cup K)}$, we have to show that

$$\sum_{y\in\mathbf{X}_{K\cup L}} (\pi \rhd \kappa)(y) = 1.$$

Therefore, to prove the whole assertion it is enough to show its second part; that is, to show that for $x \in \mathbf{X}_K$,

$$(\pi \rhd \kappa)^{\downarrow K}(x) = \sum_{y\in\mathbf{X}_{K\cup L}:y^{\downarrow K}=x} (\pi \rhd \kappa)(y) = \pi(x),$$

because then the required equality is guaranteed by the fact that $\pi(K)$ is a probability distribution.

$$\sum_{y\in\mathbf{X}_{K\cup L}:y^{\downarrow K}=x} (\pi \rhd \kappa)(y) = \sum_{y\in\mathbf{X}_{K\cup L}:y^{\downarrow K}=x} \frac{\pi(y^{\downarrow K})\kappa(y^{\downarrow L})}{\kappa^{\downarrow K\cap L}(y^{\downarrow K\cap L})} = \sum_{y\in\mathbf{X}_{K\cup L}:y^{\downarrow K}=x} \frac{\pi(x)\kappa(y^{\downarrow L})}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})}$$

$$= \frac{\pi(x)}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})} \sum_{z\in\mathbf{X}_L:z^{\downarrow K\cap L}=x^{\downarrow K\cap L}} \kappa(z) = \pi(x).$$

Moreover, due to the assumption

$$\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L},$$

if $\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L}) = 0$, then also $\pi(x) = 0$, and we defined $\pi \rhd \kappa = 0$ in these points. Therefore, $(\pi \rhd \kappa)^{\downarrow K}(x) = \pi(x)$ for all $x \in \mathbf{X}_K$, which finishes the proof. □

Example 3.3. Let us illustrate difficulties which can occur when $\pi^{\downarrow L\cap K} \ll \kappa^{\downarrow K\cap L}$, using a simple example.

Consider the distributions $\pi(x_1, x_2)$ and $\kappa(x_2, x_3)$ given in Table 1, for which $\pi(x_2 = 0) > 0$ and $\kappa(x_2 = 0) = 0$.

If the composition of these two distributions was computed according to the expression

$$(\pi \rhd \kappa)(x_1, x_2, x_3) = \frac{\pi(x_1, x_2)\kappa(x_2, x_3)}{\kappa^{\downarrow\{2\}}(x_2)},$$

for all $(x_1, x_2, x_3) \in \mathbf{X}_{\{1,2,3\}}$, the reader could easily see that, for any (x_1, x_2, x_3),

$$\pi(x_1, x_2)\kappa(x_2, x_3) = 0,$$

since for $x_2 = 1$ $\pi(x_2, x_3) = 0$, and for $x_2 = 0$ $\kappa(x_1, x_2) = 0$.

Notice also that it can easily happen that $\pi \rhd \nu$ is well defined, whereas $\nu \rhd \pi$ remains undefined. For this, consider the distribution ν from Table 2 and π from Table 1. Computation of $\pi \rhd \nu$ and $\nu \rhd \pi$ is in Table 3.

Analogously to $\rhd$, we can also introduce the operator of *left composition*.

Definition 3.4. For two arbitrary distributions $\pi(K)$ and $\kappa(L)$, for which $\kappa^{\downarrow K\cap L} \ll \pi^{\downarrow K\cap L}$, their *left composition* for each $x \in \mathbf{X}_{(L\cup K)}$ is given by the following formula:

$$(\pi \lhd \kappa)(x) = \frac{\pi(x^{\downarrow K})\kappa(x^{\downarrow L})}{\pi^{\downarrow K\cap L}(x^{\downarrow K\cap L})}.$$

In a case where $\kappa^{\downarrow K\cap L} \ll \pi^{\downarrow K\cap L}$, the composition remains undefined.

The reader most likely noticed that $\pi \rhd \kappa = \kappa \lhd \pi$, so it seems quite unnecessary to introduce two operators. The advantage of having both of them will become fully apparent in the second part of this paper, where multidimensional models will be studied. Nevertheless, let us reiterate here that either of the expressions $\pi(K) \rhd \kappa(L)$ and $\pi(K) \lhd \kappa(L)$, if defined, is a probability distribution of variables $X_{K\cup L}$. Let us now begin discussing properties of these composed distributions.

Table 1. Probability distributions π and κ.

π	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$\frac{1}{2}$	$\frac{1}{2}$
$x_2 = 1$	0	0
κ	$x_3 = 0$	$x_3 = 1$
$x_2 = 0$	0	0
$x_2 = 1$	$\frac{1}{2}$	$\frac{1}{2}$

Table 2. Uniform probability distribution ν.

ν	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$\frac{1}{4}$	$\frac{1}{4}$
$x_2 = 1$	$\frac{1}{4}$	$\frac{1}{4}$

Table 3. Computation of $\pi \triangleright \nu$ and $\nu \triangleright \pi$.

x_1	x_2	x_3	$\pi \triangleright \nu$	$\nu \triangleright \pi$
0	0	0	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$	$\frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$
0	0	1	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$	$\frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$
0	1	0	$0 \cdot \frac{1}{2} = 0$	$\frac{1}{4} \cdot \frac{0}{0} = ?$
0	1	1	$0 \cdot \frac{1}{2} = 0$	$\frac{1}{4} \cdot \frac{0}{0} = ?$
1	0	0	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$	$\frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$
1	0	1	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$	$\frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$
1	1	0	$0 \cdot \frac{1}{2} = 0$	$\frac{1}{4} \cdot \frac{0}{0} = ?$
1	1	1	$0 \cdot \frac{1}{2} = 0$	$\frac{1}{4} \cdot \frac{0}{0} = ?$

4. Composition of consistent distributions (commutativity)

The following assertion presents a trivial but rather important property.

LEMMA 4.1. For any $\pi(K)$ and $M \subset K$, $\pi = \pi^{\downarrow M} \triangleright \pi$.

Proof. First, realize that this composition is always defined. The assertion then directly follows from the definition

$$\pi^{\downarrow M} \triangleright \pi = \frac{\pi^{\downarrow M}\pi}{\pi^{\downarrow M}} = \pi. \qquad \square$$

DEFINITION 4.2. We say that distributions $\pi(K)$ and $\kappa(L)$ are *consistent* if

$$\pi^{(K\cap L)} = \kappa^{(K\cap L)}.$$

Remark 4.3. Notice that if $K \cap L = \emptyset$, the distributions π and κ are always consistent.

Directly from the definition of the operators $\triangleleft$ and $\triangleright$, we also get the following trivial assertion.

LEMMA 4.4. If $\pi(K)$ and $\kappa(L)$ are consistent distributions, then $\pi \triangleright \kappa = \pi \triangleleft \kappa$. If either $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$ or $\kappa^{\downarrow K\cap L} \ll \pi^{\downarrow K\cap L}$, then the reverse implication also holds true,

$$\pi \triangleright \kappa = \pi \triangleleft \kappa \Longrightarrow \pi^{\downarrow K\cap L} = \kappa^{\downarrow K\cap L}.$$

Proof. If π and κ are consistent, then

$$\pi \triangleright \kappa = \frac{\pi\kappa}{\kappa^{\downarrow K\cap L}} = \frac{\pi\kappa}{\pi^{\downarrow K\cap L}} = \pi \triangleleft \kappa.$$

To prove the other side of the equivalence, assume $\pi \triangleright \kappa = \pi \triangleleft \kappa$. Since we also assume that either $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$ or $\kappa^{\downarrow K\cap L} \ll \pi^{\downarrow K\cap L}$, meaning that either $\pi \triangleright \kappa$ or $\pi \triangleleft \kappa$ is defined, and since these compositions equal each other, both of them must be defined. Using Lemma 3.2 twice and then the assumed equivalence, one gets

$$\pi^{\downarrow K\cap L} = (\pi \triangleright \kappa)^{\downarrow K\cap L} = (\kappa \triangleright \pi)^{\downarrow K\cap L} = \kappa^{\downarrow K\cap L}. \qquad \square$$

5. Basic properties

In this section, a number of basic properties of operators of composition are presented. Some of them are quite intuitive and help us to understand more complex properties necessary for multidimensional model construction; some are rather technical and will be used to simplify proofs in subsequent sections.

5.1 *Dominance*

Lemma 5.1. Let $K \subseteq L \subseteq N$. For any probability distributions $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$ such that $\pi \ll \kappa^{\downarrow K}$, the relation

$$\pi \rhd \kappa \ll \kappa$$

holds true, and for any $\nu \in \Pi^{(L)}(\pi)$,

$$\nu \ll \kappa \Longleftrightarrow \nu \ll \pi \rhd \kappa.$$

Proof. The assertion follows from the definition of the operator $\rhd$, which can, under the given assumption, be written as

$$\pi \rhd \kappa = \frac{\pi\kappa}{\kappa^{\downarrow K}}.$$

From this formula, it is obvious that for any $x \in \mathbf{X}_L$,

$$\kappa(x) = 0 \Longrightarrow (\pi \rhd \kappa)(x) = 0,$$

which proves that $\pi \rhd \kappa \ll \kappa$.

Analogously, let $\nu \in \Pi^{(L)}(\pi)$ be dominated by κ. Consider an $x \in \mathbf{X}_L$ for which

$$(\pi \rhd \kappa)(x) = \frac{\pi(x^{\downarrow K})\kappa(x)}{\kappa^{\downarrow K}(x^{\downarrow K})} = 0.$$

That means that either $\pi(x^{\downarrow K}) = 0$ or $\kappa(x) = 0$ (or both). If $\pi(x^{\downarrow K}) = 0$, then $\nu^{\downarrow K}(x^{\downarrow K}) = 0$ as $\nu^{\downarrow K} = \pi$, because $\nu \in \Pi^{L}(\pi)$. Therefore, also $\nu(x) = 0$. On the other hand, if $\kappa(x) = 0$, then $\nu(x) = 0$, because ν is dominated by κ. This proves that

$$\nu \ll \kappa \Longrightarrow \nu \ll \pi \rhd \kappa.$$

The opposite implication,

$$\nu \ll \pi \rhd \kappa \Longrightarrow \nu \ll \kappa,$$

follows immediately from the first part of the proof due to transitivity of dominance

$$\nu \ll \pi \rhd \kappa \quad \text{and} \quad \pi \rhd \kappa \ll \kappa \Longrightarrow \nu \ll \kappa.$$

□

5.2 *Conditional independence*

As said above, application of the operator of composition introduces conditional independence among the variables. The exact meaning of this statement is expressed by the following simple, but important, assertion.

LEMMA 5.2. Let $\nu(K \cup L) = \pi(K) \rhd \kappa(L)$ be defined and $K \setminus L \neq \emptyset \neq L \setminus K$. Then $X_{K\setminus L} \perp\!\!\!\perp X_{L\setminus K} | X_{K\cap L}[\nu]$.

Proof. To prove this assertion, we have to show that for $\nu = \pi \rhd \kappa$,

$$\nu(x)\nu^{\downarrow K\cap L}(x^{\downarrow K\cap L}) = \nu^{\downarrow K}(x^{\downarrow K})\nu^{\downarrow L}(x^{\downarrow L}) \tag{2}$$

for all $x \in \mathbf{X}_{K\cup L}$. If, for $x \in \mathbf{X}_{K\cup L}$, $\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L}) = 0$, then also $\pi^{\downarrow K\cap L}(x^{\downarrow K\cap L}) = 0$ (because ν is defined only when $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$), and therefore, also $\nu^{\downarrow K\cap L}(x^{\downarrow K\cap L}) = 0$ (and thus $\nu^{\downarrow K}(x^{\downarrow K}) = \nu^{\downarrow L}(x^{\downarrow L}) = \nu(x) = 0$, too). From this, we immediately get that equality (2) holds, because both its sides equal 0.

Consider now such $x \in \mathbf{X}_{K\cup L}$ that $\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L}) > 0$. Lemma 3.2 says that $\nu^{\downarrow K}(x^{\downarrow K}) = \pi(x^{\downarrow K})$. Let us compute $\nu^{\downarrow L}(x^{\downarrow L})$:

$$\begin{aligned}\nu^{\downarrow L}(x^{\downarrow L}) &= \sum_{y\in\mathbf{X}_{K\cup L}: y^{\downarrow L}=x^{\downarrow L}} \frac{\pi(y^{\downarrow K})\kappa(y^{\downarrow L})}{\kappa^{\downarrow K\cap L}(y^{\downarrow K\cap L})} = \frac{\kappa(x^{\downarrow L})}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})} \sum_{y\in\mathbf{X}_{K\cup L}: y^{\downarrow L}=x^{\downarrow L}} \pi(y^{\downarrow K}) \\ &= \frac{\kappa(x^{\downarrow L})\pi^{\downarrow K\cap L}(x^{\downarrow K\cap L})}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})}.\end{aligned}$$

Therefore,

$$\nu^{\downarrow K}(x^{\downarrow K})\nu^{\downarrow L}(x^{\downarrow L}) = \frac{\pi(x^{\downarrow K})\kappa(x^{\downarrow L})\pi^{\downarrow K\cap L}(x^{\downarrow K\cap L})}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})},$$

where

$$\frac{\pi(x^{\downarrow K})\kappa(x^{\downarrow L})}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})} = \nu(x)$$

from the definition of operator $\rhd$, and $\pi^{\downarrow K\cap L}(x^{\downarrow K\cap L}) = \nu^{\downarrow K\cap L}(x^{\downarrow K\cap L})$ due to Lemma 3.2. So we get $\nu^{\downarrow K}\nu^{\downarrow L} = \nu\nu^{\downarrow K\cap L}$. □

COROLLARY 5.3. Consider a distribution $\pi(K)$ and two subsets $L, M \subset K$ such that $L\setminus M \neq \emptyset$, $M\setminus L \neq \emptyset$. Then

$$X_{L\setminus M} \perp\!\!\!\perp X_{M\setminus L}|X_{L\cap M}[\pi] \iff \pi^{\downarrow L\cup M} = \pi^{\downarrow L} \rhd \pi^{\downarrow M}.$$

Proof. First notice that $\pi^{\downarrow L} \rhd \pi^{\downarrow M}$ is always defined, because the marginal distributions from this expression are consistent.

If $\pi^{\downarrow L\cup M} = \pi^{\downarrow L} \rhd \pi^{\downarrow M}$, then validity of the required independence follows from the preceding Lemma 5.2; therefore, we have only to show that if $\pi^{\downarrow L\cup M} \cdot \pi^{\downarrow L\cap M} = \pi^{\downarrow L} \cdot \pi^{\downarrow M}$ (which is a definition of the conditional independence in question), then also $\pi^{\downarrow L\cup M} = \pi^{\downarrow L} \rhd \pi^{\downarrow M}$. But this is trivial, as

$$\pi^{\downarrow L} \rhd \pi^{\downarrow M} = \frac{\pi^{\downarrow L} \cdot \pi^{\downarrow M}}{\pi^{\downarrow L\cap M}} = \frac{\pi^{\downarrow L\cup M} \cdot \pi^{\downarrow L\cap M}}{\pi^{\downarrow L\cap M}} = \pi^{\downarrow L\cup M}.$$

All the above modifications are correct because, if for some x $\pi^{\downarrow L\cap M}(x^{\downarrow L\cap M}) = 0$, then also $\pi^{\downarrow L}(x^{\downarrow L}) = \pi^{\downarrow M}(x^{\downarrow M}) = \pi^{\downarrow L\cup M}(x^{\downarrow L\cup M}) = 0$ (recall that we defined $0.0/0 = 0$). □

5.3 Shannon entropy

It is well known that the conditional independence of variables introduced by the operator of composition is closely connected with the fact that the composed distribution achieves maximal Shannon entropy,[1] as expressed in the following assertion (its proof is based on a simple technique that can be found in any textbook on information theory). Recall first that it follows directly from Lemmas 3.2 and 4.4 that for consistent distributions $\pi(K)$ and $\kappa(L)$, their composition $\pi \triangleright \kappa$ is their common extension from $\Pi^{(K\cup L)}$, i.e.

$$\pi \triangleright \kappa \in \Pi^{(K\cup L)}(\{\pi,\kappa\})\} = \Pi^{(K\cup L)}(\pi)\} \cap \Pi^{(K\cup L)}(\kappa)\}.$$

Theorem 5.4. If probability distributions $\pi(K)$ and $\kappa(L)$ are consistent, then

$$H(\pi \triangleright \kappa) = H(\pi) + H(\kappa) - H(\kappa^{\downarrow K\cap L}),$$

and

$$\pi \triangleright \kappa = \arg \max_{\nu \in \Pi^{(K\cup L)}(\{\pi,\kappa\})} H(\nu).$$

Proof. The first part of the proof is trivial

$$\begin{aligned} H(\pi \triangleright \kappa) &= -\sum_{\substack{x\in\mathbf{X}_{K\cup L}\\ (\pi\triangleright\kappa)(x)>0}} (\pi \triangleright \kappa)(x)\log(\pi \triangleright \kappa)(x)\\ &= -\sum_{\substack{x\in\mathbf{X}_{K\cup L}\\ (\pi\triangleright\kappa)(x)>0}} (\pi \triangleright \kappa)(x)\log\pi(x^{\downarrow K}) - \sum_{\substack{x\in\mathbf{X}_{K\cup L}\\ (\pi\triangleright\kappa)(x)>0}} (\pi \triangleright \kappa)(x)\log\kappa(x^{\downarrow L})\\ &\quad + \sum_{\substack{x\in\mathbf{X}_{K\cup L}\\ (\pi\triangleright\kappa)(x)>0}} (\pi \triangleright \mathrm{t}\kappa)(x)\log\kappa(x^{\downarrow K\cap L}) = H(\pi) + H(\kappa) - H(\kappa^{\downarrow K\cap L}), \end{aligned}$$

because both π and κ are marginal to $\pi \triangleright \kappa$ (this holds due to consistency of π and κ, and Lemmas 4.4 and 3.2).

Now, let us compute the Shannon entropy for an arbitrary distribution $\nu \in \Pi^{(K\cup L)}(\{\pi,\kappa\})$.

$$\begin{aligned} H(\nu) &= -\sum_{\substack{x\in\mathbf{X}_{K\cup L}\\ \nu(x)>0}} \nu(x)\log\nu(x) = -\sum_{\substack{x\in\mathbf{X}_{K\cup L}\\ \nu(x)>0}} \nu(x)\log\frac{\nu^{\downarrow K}(x^{\downarrow K})\nu^{\downarrow L}(x^{\downarrow L})\nu(x)\nu^{\downarrow K\cap L}(x^{\downarrow K\cap L})}{\nu^{\downarrow K\cap L}(x^{\downarrow K\cap L})\nu^{\downarrow K}(x^{\downarrow K})\nu^{\downarrow L}(x^{\downarrow L})}\\ &= -\sum_{\substack{x\in\mathbf{X}_{K\cup L}\\ \nu(x)>0}} \nu(x)\log\nu^{\downarrow K}(x^{\downarrow K}) - \sum_{\substack{x\in\mathbf{X}_{K\cup L}\\ \nu(x)>0}} \nu(x)\log\nu^{\downarrow L}(x^{\downarrow L})\\ &\quad + \sum_{\substack{x\in\mathbf{X}_{K\cup L}\\ \nu(x)>0}} \nu(x)\log\nu^{\downarrow K\cap L}(^{\downarrow K\cap L}) - \sum_{\substack{x\in\mathbf{X}_{K\cup L}\\ \nu(x)>0}} \nu(x)\log\frac{\nu(x)\nu^{\downarrow K\cap L}(x^{\downarrow K\cap L})}{\nu^{\downarrow K}(x^{\downarrow K})\nu^{\downarrow L}(x^{\downarrow L})}\\ &= H(\nu^{\downarrow K}) + H(\nu^{\downarrow L}) - H(\nu^{\downarrow K\cap L}) - \mathrm{MI}_\nu(X_{K\setminus L}; X_{L\setminus K}|X_{K\cap L})\\ &= H(\pi) + H(\kappa) - H(\kappa^{\downarrow K\cap L}) - \mathrm{MI}_\nu(X_{K\setminus L}; X_{L\setminus K}|X_{K\cap L})\\ &= H(\pi \triangleright \kappa) - \mathrm{MI}_\nu(X_{K\setminus L}; X_{L\setminus K}|X_{K\cap L}), \end{aligned}$$

where $\mathrm{MI}_\nu(X_{K\setminus L}; X_{L\setminus K}|X_{K\cap L})$ is *conditional mutual information* (see, e.g. Gallager 1968), which is known to always be non-negative (and equal to 0 if and only if $X_{K\setminus L} \perp\!\!\!\perp X_{L\setminus K}|X_{K\cap L}|[\nu]$, i.e. when $\nu = \pi \triangleright \kappa$). □

Remark 5.5. In this context, the reader should realize that the equality

$$H(\pi \triangleright \kappa) = H(\pi) + H(\kappa) - H(\kappa^{\downarrow K\cap L})$$

is guaranteed only for consistent distributions. As we shall see from the following example, in a case where π and κ are inconsistent, the entropy of their composition can be lower or higher than this sum.

Example 5.6. Consider distribution κ from Table 4.

Taking the binary logarithm for computation of the Shannon entropy, we get

$$H(\kappa) - H(\kappa^{\downarrow\{1\}}) = \frac{3}{2} - 1 = \frac{1}{2}.$$

Let us compute entropy values for the distributions $\pi \triangleright \kappa$ and $\hat{\pi} \triangleright \kappa$, where $\pi(x_1 = 0) = \hat{\pi}(x_1 = 1) = 0.1$ and $\pi(x_1 = 1) = \hat{\pi}(x_1 = 0) = 0.9$. These composed distributions are shown in Table 5 and their entropy values equal

$$H(\pi \triangleright \kappa) = 0 \log_2 0 + 0.1 \log_2 0.1 + 0.45 \log_2 0.45 + 0.45 \log_2 0.45 = 1.369,$$
$$H(\hat{\pi} \triangleright \kappa) = 0 \log_2 0 + 0.9 \log_2 0.9 + 0.05 \log_2 0.05 + 0.05 \log_2 0.05 = 0.569,$$

which certainly differ from

$$H(\pi) + H(\kappa) - H(\kappa^{\downarrow\{1\}}) = H(\hat{\pi}) + H(\kappa) - H(\kappa^{\downarrow\{1\}}) = 0.469 + 0.5 = 0.969.$$

To make the situation even more complicated, let us mention that it may happen that the equality

$$H(\pi(K) \triangleright \kappa(L)) = H(\pi(K)) + H(\kappa(L)) - H(\kappa^{\downarrow K\cap L})$$

holds even in the case of inconsistent distributions. For simplicity, consider uniform distribution $\hat{\kappa}(x_1, x_2) = 0.25$. Then

$$H(\hat{\kappa}) - H(\hat{\kappa}^{\downarrow\{1\}}) = 2 - 1 = 1,$$

Table 4. Probability distribution κ.

κ	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0	$\frac{1}{4}$
$x_2 = 1$	$\frac{1}{2}$	$\frac{1}{4}$

Table 5. Probability distributions $\pi \triangleright \kappa$ and $\hat{\pi} \triangleright \kappa$.

$\pi \triangleright \kappa$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0	0.45
$x_2 = 1$	0.1	0.45
$\hat{\pi} \triangleright \kappa$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0	0.05
$x_2 = 1$	0.9	0.05

and for $\nu(x_1)$,

$$H(\nu \triangleright \hat{\kappa}) = -2 \cdot \left(\frac{\nu(x_1 = 0)}{2} \log \frac{\nu(x_1 = 0)}{2}\right) - 2 \cdot \left(\frac{\nu(x_1 = 1)}{2} \log \frac{\nu(x_1 = 1)}{2}\right)$$
$$= -\nu(x_1 = 0)((\log \nu(x_1 = 0)) - 1) - \nu(x_1 = 1)((\log \nu(x_1 = 1)) - 1) = H(\nu) + 1,$$

from which we immediately get that

$$H(\nu \triangleright \hat{\kappa}) = H(\nu) + H(\hat{\kappa}) - H(\hat{\kappa}^{\downarrow\{1\}})$$

for any one-dimensional distribution ν.

5.4 *Basic exchange lemma*

In many proofs, we will need the following important assertion. It is perhaps worth mentioning that its proof would be quite simple for strictly positive distributions (namely, all the compositions would be defined); however, it holds even in the presented general form.

Lemma 5.7. Consider three distributions $\pi(K)$, $\kappa(L)$, and $\nu(M)$. If $K \supseteq (L \cap M)$, then

$$(\pi \triangleright \kappa) \triangleright \nu = (\pi \triangleright \nu) \triangleright \kappa. \tag{3}$$

Proof. First, let us show that the left-hand side expression in (3) is not defined *iff* the right-hand side of this formula is not defined. From the definition of the operators, we know that $(\pi \triangleright \kappa) \triangleright \nu$ is defined *iff*

$$\pi^{\downarrow K \cap L} \ll \kappa^{\downarrow K \cap L} \quad \text{and} \quad (\pi \triangleright \kappa)^{\downarrow (K \cup L) \cap M} \ll \nu^{\downarrow (K \cup L) \cap M}.$$

Analogously, $(\pi \triangleright \nu) \triangleright \kappa$ is defined *iff*

$$\pi^{\downarrow K \cap M} \ll \nu^{\downarrow K \cap M} \quad \text{and} \quad (\pi \triangleright \nu)^{\downarrow (K \cup M) \cap L} \ll \kappa^{\downarrow (K \cup M) \cap L}.$$

Under the given assumption $K \supseteq (L \cap M)$, these two conditions coincide because

$$((K \cup L) \cap M) = (K \cap M) \quad \text{and} \quad ((K \cup M) \cap L) = (K \cap L), \tag{4}$$

and

$$(\pi \triangleright \kappa)^{\downarrow(K\cup L)\cap M} = \pi^{\downarrow K\cap M},$$
$$(\pi \triangleright \nu)^{\downarrow(K\cup M)\cap L} = \pi^{\downarrow K\cap L}.$$

Now, let us assume that both the expressions in formula (3) are defined. Because of equalities (4), the expressions

$$(\pi \triangleright \kappa) \triangleright \nu = \frac{\pi\kappa\nu}{\kappa^{\downarrow K\cap L}\nu^{\downarrow M\cap(K\cup L)}},$$
$$(\pi \triangleright \nu) \triangleright \kappa = \frac{\pi\kappa\nu}{\nu^{\downarrow K\cap M}\kappa^{\downarrow L\cap(K\cup M)}},$$

are equivalent to each other, which finishes the proof. □

5.5 *Marginalization*

Example 5.8. Quite often we have to compute a marginal from a distribution defined as a composition of oligodimensional distributions. Therefore, it is important to realize that generally for $\pi(K)$, $\kappa(L)$, and $M \subset K \cup L$,

$$(\pi \triangleright \kappa)^{\downarrow M} \neq \pi^{\downarrow K\cap M} \triangleright \kappa^{\downarrow L\cap M}. \tag{5}$$

To illustrate a situation when equality in formula (5) does not hold, consider composition $\pi \triangleright \kappa$ of distributions from Table 6 and its marginal distribution $(\pi \triangleright \kappa)^{\downarrow\{1,3\}}$, which is depicted in Table 7. At first glance, we see that variables X_1 and X_3 are not independent for this marginal distribution. Therefore,

$$(\pi(x_1,x_2) \triangleright \kappa(x_2,x_3))^{\downarrow\{1,3\}} \neq (\pi(x_1,x_2))^{\downarrow\{1\}} \triangleright (\kappa(x_2,x_3))^{\downarrow\{3\}} = \pi^{\downarrow\{1\}}(x_1) \triangleright \kappa^{\downarrow\{3\}}(x_3)$$
$$= \pi^{\downarrow\{1\}}(x_1)\cdot\kappa^{\downarrow\{3\}}(x_3).$$

Nevertheless, as it will be formulated in Lemma 5.10, in special situations the equality in expression (5) holds. Before presenting this assertion, let us formulate a simpler, more specific, assertion.

Lemma 5.9. Consider two distributions $\pi(K)$, $\kappa(L)$, and $M \subseteq N$ such that $K \cup L \supseteq M \supseteq K$. Then

$$(\pi \triangleright \kappa)^{\downarrow M} = \pi \triangleright \kappa^{\downarrow L\cap M}.$$

Table 6. Probability distributions π and κ.

π	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$\frac{1}{2}$	0
$x_2 = 1$	0	$\frac{1}{2}$
κ	$x_3 = 0$	$x_3 = 1$
$x_2 = 0$	$\frac{1}{2}$	0
$x_2 = 1$	0	$\frac{1}{2}$

Table 7. Marginal $(\pi \triangleright \kappa)^{\downarrow\{1,3\}}$ of a composed distribution.

$(\pi \triangleright \kappa)^{\downarrow\{1,3\}}$	$x_1 = 0$	$x_1 = 1$
$x_3 = 0$	$\frac{1}{2}$	0
$x_3 = 1$	0	$\frac{1}{2}$

Proof. Let us first mention that both $\pi \triangleright \kappa$ and $\pi \triangleright \kappa^{\downarrow L\cap M}$ are not defined *iff* $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$. So, to prove the assertion, we can restrict our considerations to a case when $\pi \triangleright \kappa$ is defined. Let us compute (for $x \in \mathbf{X}_M$)

$$(\pi \triangleright \kappa)^{\downarrow M}(x) = \sum_{y\in\mathbf{X}_{K\cup L}:y^{\downarrow M}=x} \frac{\pi(y^{\downarrow K})\kappa(y^{\downarrow L})}{\kappa^{\downarrow K\cap L}(y^{\downarrow K\cap L})} = \frac{\pi(x^{\downarrow K})}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})} \sum_{y\in\mathbf{X}_{K\cup L}:y^{\downarrow M}=x} \kappa(y^{\downarrow L})$$

$$= \frac{\pi(x^{\downarrow K})\kappa^{\downarrow L\cap M}(x^{\downarrow L\cap M})}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})} = (\pi \triangleright \kappa^{\downarrow L\cap M})(x).$$

□

Lemma 5.10. Let $K, L, M \subseteq N$. If $K \cup L \supseteq M \supseteq K \cap L$, then for any probability distributions $\pi(K)$ and $\kappa(L)$, $(\pi \triangleright \kappa)^{\downarrow M} = \pi^{\downarrow K\cap M} \triangleright \kappa^{\downarrow L\cap M}$.

Proof. Let us compute the required marginal distribution in two steps. In the first step, we will employ Lemma 5.9, then Lemma 4.1 and finally Lemma 5.7:

$$(\pi \triangleright \kappa)^{\downarrow K\cup M} = \pi \triangleright \kappa^{\downarrow L\cap M} = (\pi^{\downarrow K\cap L} \triangleright \pi) \triangleright \kappa^{\downarrow L\cap M} = (\pi^{\downarrow K\cap L} \triangleright \kappa^{\downarrow L\cap M}) \triangleright \pi.$$

The last expression will be further marginalized with the help of Lemma 5.9 and afterwards the final form will be received with application of Lemma 5.7 and Lemma 4.1.

$$(\pi \triangleright \kappa)^{\downarrow M} = \left((\pi^{\downarrow K\cap L} \triangleright \kappa^{\downarrow L\cap M}) \triangleright \pi\right)^{\downarrow M} = (\pi^{\downarrow K\cap L} \triangleright \kappa^{\downarrow L\cap M}) \triangleright \pi^{\downarrow K\cap M}$$

$$= (\pi^{\downarrow K\cap L} \triangleright \pi^{\downarrow K\cap M}) \triangleright \kappa^{\downarrow L\cap M} = \pi^{\downarrow K\cap M} \triangleright \kappa^{\downarrow L\cap M}.$$

□

The following assertion shows that any composition of two distributions can be expressed as a composition of two consistent distributions, each of which is defined for the same group of variables as the original ones. □

Lemma 5.11. Let $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$. If $\pi \triangleright \kappa$ is defined, then

$$\pi \triangleright \kappa = \pi \triangleright (\pi \triangleright \kappa)^{\downarrow L}.$$

Proof. The assertion is a trivial consequence of the next, more general assertion. □

Lemma 5.12. Let $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$. If $\pi \triangleright \kappa$ is defined and $L \subseteq M \subseteq K \cup L$, then

$$\pi \triangleright \kappa = \pi \triangleright (\pi \triangleright \kappa)^{\downarrow M}. \tag{6}$$

Proof. First, notice that if $\pi \triangleright \kappa$ is defined, then also $\pi \triangleright (\pi \triangleright \kappa)^{\downarrow M}$ is defined (namely, π and $(\pi \triangleright \kappa)^{\downarrow M}$ are consistent). The required equality (6) follows immediately from

Lemmas 5.10, 4.4, 5.7, and 4.1:

$$\pi \triangleright (\pi \triangleright \kappa)^{\downarrow M} = \pi \triangleright (\pi^{\downarrow K\cap M} \triangleright \kappa) = (\pi^{\downarrow K\cap M} \triangleright \kappa) \triangleright \pi = (\pi^{\downarrow K\cap M} \triangleright \pi) \triangleright \kappa = \pi \triangleright \kappa. \quad \square$$

6. I-geometry of composition operators

This section is based on the results of Csiszár (1975), and therefore, we use also his terminology (such as the term *I-geometry* in the section title or using Kullback–Leibler divergence 'Div' as a 'measure' of distance).

Definition 6.1. Consider any $\pi \in \Pi^{(L)}$ and an arbitrary subset Θ of $\Pi^{(L)}$. Distribution

$$\kappa = \arg\min_{\nu \in \Theta} \mathrm{Div}(\nu \| \pi)$$

is called an *I-projection* of π into Θ.

According to this definition, I-projection is a distribution from $\Theta \subset \Pi^{(L)}$, which is, in a sense, closest to π. As a measure of distance, we take the *Kullback–Leibler divergence*[2]

$$\mathrm{Div}(\kappa \| \pi) = \begin{cases} \sum\limits_{x \in \mathbf{X}_L : \kappa(x) > 0} \kappa(x) \log \frac{\kappa(x)}{\pi(x)} & \text{if } \kappa \ll \pi, \\ +\infty & \text{otherwise.} \end{cases}$$

Generally, it may happen that for given π and Θ, the I-projection does not exist. However, considering Θ to be a set of distributions with given marginal(s), which is always a convex compact set of distributions, the existence and uniqueness of the I-projection are guaranteed just by the existence of one $\nu \in \Theta$ for which $\mathrm{Div}(\nu \| \pi)$ is finite. Instructions for finding this I-projection are given by the following assertion.

Theorem 6.2. Let $K \subseteq L \subseteq N$. For arbitrary probability distributions $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$ such that $\pi \ll \kappa^{\downarrow K}$, $\pi \triangleright \kappa$ is the I-projection of κ into $\Pi^{(L)}(\pi)$ (Figure 2). Moreover,

$$\mathrm{Div}(\nu \| \kappa) = \mathrm{Div}(\nu \| \pi \triangleright \kappa) + \mathrm{Div}(\pi \triangleright \kappa \| \kappa)$$

for any $\nu \in \Pi^{(L)}(\pi)$.

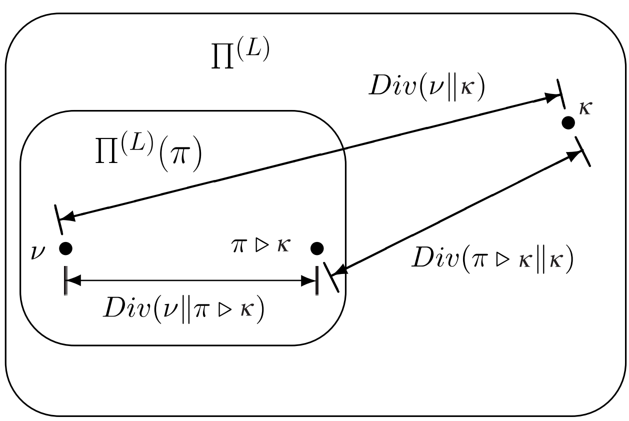


Figure 2. I-projection of κ into $\Pi^{(L)}(\pi)$.

Proof. $\pi \triangleright \kappa \in \Pi^{(L)}(\pi)$ and, since $\pi \triangleright \kappa \ll \kappa$ (this holds due to Lemma 5.1), $\mathrm{Div}(\pi \triangleright \kappa\|\kappa)$ is finite. Therefore, the I-projection ν^* of κ in $\Pi^{(L)}(\pi)$ must be dominated by κ (otherwise $\mathrm{Div}(\nu^*\|\kappa) = +\infty$ and ν^* cannot be an I-projection of κ in $\Pi^{(L)}(\pi)$).

Consider any $\nu \in \Pi^{(L)}(\pi)$ that is dominated by κ. First, realize that, because of Lemma 5.1, $\nu \ll \pi \triangleright \kappa$. Therefore, we can compute

$$
\begin{aligned}
\mathrm{Div}(\nu\|\kappa) &= \sum_{x\in\mathbf{X}_L:\nu(x)>0} \nu(x)\log\frac{\nu(x)}{\kappa(x)} = \sum_{x\in\mathbf{X}_L:\nu(x)>0} \nu(x)\log\left(\frac{\nu(x)}{(\pi \triangleright \kappa)(x)}\frac{(\pi \triangleright \kappa)(x)}{\kappa(x)}\right)\\
&= \mathrm{Div}(\nu\|\pi \triangleright \kappa) + \sum_{x\in\mathbf{X}_L:\nu(x)>0} \nu(x)\log\frac{(\pi \triangleright \kappa)(x)}{\kappa(x)}\\
&= \mathrm{Div}(\nu\|\pi \triangleright \kappa) + \sum_{x\in\mathbf{X}_L:\nu(x)>0} \nu(x)\log\frac{\pi(x^{\downarrow K})\kappa(x)}{\kappa^{\downarrow K}(x^{\downarrow K})\kappa(x)}\\
&= \mathrm{Div}(\nu\|\pi \triangleright \kappa) + \sum_{z\in\mathbf{X}_K:\nu^{\downarrow K}(z)>0} \nu^{\downarrow K}(z)\log\frac{\pi(z)}{\kappa^{\downarrow K}(z)}\\
&= \mathrm{Div}(\nu\|\pi \triangleright \kappa) + \sum_{z\in\mathbf{X}_K:\pi(z)>0} \pi(z)\log\frac{\pi(z)}{\kappa^{\downarrow K}(z)} = \mathrm{Div}(\nu\|\pi \triangleright \kappa) + Div(\pi\|\kappa^{\downarrow K}).
\end{aligned}
$$

As it is known that the divergence $\mathrm{Div}(\nu\|\pi \triangleright \kappa)$ cannot be negative, $\mathrm{Div}(\nu\|\kappa)$ achieves its minimum for $\nu = \pi \triangleright \kappa$ (since $\mathrm{Div}(\pi \triangleright \kappa\|\pi \triangleright \kappa) = 0$) and thus

$$\mathrm{Div}(\pi \triangleright \kappa\|\kappa) = \mathrm{Div}(\pi\|\kappa^{\downarrow K}).$$

The equality

$$\mathrm{Div}(\nu\|\kappa) = \mathrm{Div}(\nu\|\pi \triangleright \kappa) + \mathrm{Div}(\pi \triangleright \kappa\|\kappa)$$

also holds when $\nu \ll \kappa$ because, according to Lemma 5.1, also $\nu \ll \pi \triangleright \kappa$, and, therefore, both $\mathrm{Div}(\nu\|\kappa)$ and $\mathrm{Div}(\nu\|\pi \triangleright \kappa)$ equal $+\infty$. □

7. Associativity

Having a binary operator, mathematicians usually ask questions regarding its mathematical properties like commutativity, associativity, and idempotence. Up to now, we have shown that composition is commutative only for consistent distributions (cf. Lemma 4.4). The idempotence of composition follows immediately from Lemma 3.2, because from this assertion one can see that

$$\pi \triangleright \pi = \pi \triangleleft \pi = \pi.$$

In this section, we will show that composition is generally not associative, but, similarly to commutativity, associativity holds under some special conditions.

The importance of the operators of composition stems from the fact that they can form multidimensional distributions from a system of oligodimensional (low-dimensional) distributions. When these operators are iteratively applied to a sequence of distributions, the result, if defined, is a multidimensional distribution. This resulting distribution is defined for all the variables which appear among the arguments of at least one distribution from the considered sequence. And it is this iterative application of operators, together with the fact that they are neither commutative nor associative, that led us to define two operators $\triangleright$ and $\triangleleft$.

Let us start this section with an example, in which we will show that generally

(a) $(\pi \triangleright \kappa) \triangleright \nu \neq \pi \triangleright (\kappa \triangleright \nu)$,
(b) $(\pi \triangleright \kappa) \triangleright \nu \neq (\pi \triangleright \nu) \triangleright \kappa$,
(c) $(\pi \triangleright \kappa) \triangleright \nu \neq (\pi \triangleleft \kappa) \triangleleft \nu$, and
(d) $(\pi \triangleright \kappa) \triangleleft \nu \neq (\pi \triangleleft \nu) \triangleleft \kappa$.

Nevertheless, let us keep in mind that the equality in these expressions may occur in special situations. For example, when all the distributions π, κ, and ν are uniform, then all the expressions result in a uniform distribution, too.

Example 7.1. Suppose all the expressions appearing in this example are defined.

(a) Consider

$$(\pi(x_1) \triangleright \kappa(x_2)) \triangleright \nu(x_1, x_2) = \pi(x_1)\kappa(x_2), \tag{7}$$

which evidently differs from

$$\pi(x_1) \triangleright (\kappa(x_2) \triangleright \nu(x_1, x_2)) = \frac{\pi(x_1)(\kappa(x_2)\nu(x_1|x_2))}{\sum_{y \in \mathbf{X}_1} \kappa(x_2)\nu(y|x_2)}. \tag{8}$$

Namely, in (7), the variables X_1 and X_2 are independent; $X_1 \perp\!\!\!\perp X_2[(\pi \triangleright \kappa) \triangleright \nu]$, which need not be generally true for (8). To see it, take an example, where both $\pi(x_1)$ and $\kappa(x_2)$ are uniform distributions and $\nu(0,0) = \nu(1,1) = 1/2$ and $\nu(0,1) = \nu(1,0) = 0$. In this case, both the marginal distributions $\nu(x_1)$ and $\nu(x_2)$ are uniform and, therefore, π and ν, as well as κ and ν, are consistent. Therefore (due to Lemma 4.4), $\pi \triangleright \nu = \pi \triangleleft \nu = \nu$ and also $\kappa \triangleright \nu = \kappa \triangleleft \nu = \nu$. From this, we get

$$\pi \triangleright (\kappa \triangleleft \nu) = \pi \triangleright \nu = \pi \triangleleft \nu = \nu,$$

which obviously differs from $\pi \triangleright \kappa$. As a product of two one-dimensional uniform distributions, the latter is a uniform distribution, too.

(b) To illustrate the second inequality, consider three one-dimensional distributions $\pi(x_1)$, $\kappa(x_2)$, and $\nu(x_2)$, such that $\kappa(x_2) \neq \nu(x_2)$. Then

$$(\pi(x_1) \triangleright \kappa(x_2)) \triangleright \nu(x_2) = \pi(x_1)\kappa(x_2) \neq \pi(x_1)\nu(x_2) = (\pi(x_1) \triangleright \nu(x_2)) \triangleright \kappa(x_2).$$

(c) Consider three distributions $\pi(x)$, $\kappa(x)$, and $\nu(x)$ for which $\pi(x) \neq \nu(x)$. Then

$$(\pi(x) \triangleright \kappa(x)) \triangleright \nu(x) = \pi(x) \neq \nu(x) = (\pi(x) \triangleleft \kappa(x)) \triangleleft \nu(x).$$

(d) Consider again distributions $\pi(x)$, $\kappa(x)$, and $\nu(x)$, this time such that $\kappa(x) \neq \nu(x)$. Then it is clear that

$$(\pi(x) \lhd \kappa(x)) \lhd \nu(x) = \nu(x) \neq \kappa(x) = (\pi(x) \lhd \nu(x)) \lhd \kappa(x).$$

Let us start with the two most important assertions of this section. They formulate sufficient conditions under which the associativity of $\rhd$ holds.

Theorem 7.2. Let $\pi(K)$, $\kappa(L)$, and $\nu(M)$ be distributions for which $\pi \rhd (\kappa \rhd \nu)$ is defined. If $K \supseteq (L \cap M)$, then

$$(\pi \rhd \kappa) \rhd \nu = \pi \rhd (\kappa \rhd \nu) = (\nu \lhd \kappa) \lhd \pi.$$

Proof. We begin by showing that under the given assumptions, $(\pi \rhd \kappa) \rhd \nu$ is also defined. The fact that $\pi \rhd (\kappa \rhd \nu)$ is defined enforces $\pi^{\downarrow K \cap (L \cup M)} \ll (\kappa \rhd \nu)^{\downarrow K \cap (L \cup M)}$, from which, due to Lemma 3.2, one gets that $\pi^{\downarrow K \cap L} \ll \kappa^{\downarrow K \cap L}$, which guarantees that $\pi \rhd \kappa$ is defined.

So, it remains to also be shown that $(\pi \rhd \kappa)^{\downarrow M \cap (K \cup L)} \ll \nu^{\downarrow M \cap (K \cup L)}$, which is under the given assumptions regarding K equivalent to $\pi^{\downarrow M \cap K} \ll \kappa^{\downarrow M \cap K}$. This follows from the transitivity of the dominance. We have already said that $\pi^{\downarrow K \cap (L \cup M)} \ll (\kappa \rhd \nu)^{\downarrow K \cap (L \cup M)}$, because $\pi \rhd (\kappa \rhd \nu)$ is defined, and $(\kappa \rhd \nu)^{\downarrow M} \ll \nu$ due to Lemma 5.1.

The fact that under the given assumptions (see Lemma 5.10),

$$(\kappa \rhd \nu)^{\downarrow K \cap (L \cup M)} = \kappa^{\downarrow K \cap L} \rhd \nu^{\downarrow K \cap M} = \frac{\kappa^{\downarrow K \cap L} \cdot \nu^{\downarrow K \cap M}}{\nu^{\downarrow L \cap M}}$$

will be used in the following computations:

$$\pi \rhd (\kappa \rhd \nu) = \pi \rhd \left(\frac{\kappa \cdot \nu}{\nu^{\downarrow L \cap M}}\right) = \frac{\pi \cdot \kappa \cdot \nu}{\nu^{\downarrow L \cap M} \cdot (\kappa \rhd \nu)^{\downarrow K \cap (L \cup M)}} = \frac{\pi \cdot \kappa \cdot \nu}{\nu^{\downarrow L \cap M}} \cdot \frac{\nu^{\downarrow L \cap M}}{\kappa^{\downarrow K \cap L} \cdot \nu^{\downarrow K \cap M}}$$

$$= \frac{\pi \cdot \kappa}{\kappa^{\downarrow K \cap L}} \cdot \frac{\nu}{\nu^{\downarrow K \cap M}} = (\pi \rhd \kappa) \rhd \nu,$$

where the last modification is valid, because under the given assumption, $M \cap (K \cup L) = M \cap K$. □

Theorem 7.3. Let $\pi(K)$, $\kappa(L)$, and $\nu(M)$ be distributions for which $\pi \rhd (\kappa \rhd \nu)$ is defined. If $L \supseteq (K \cap M)$, then

$$(\pi \rhd \kappa) \rhd \nu = \pi \rhd (\kappa \rhd \nu) = (\nu \lhd \kappa) \lhd \pi.$$

Proof. From the fact that $\pi \rhd (\kappa \rhd \nu)$ is defined, one gets that $\kappa^{\downarrow L \cap M} \ll \nu^{\downarrow L \cap M}$ and $\pi^{\downarrow K \cap (L \cup M)} \ll (\kappa \rhd \nu)^{\downarrow K \cap (L \cup M)}$. However, assuming that $L \supseteq (K \cap M)$, we know that $K \cap (L \cup M) = K \cap L$, and therefore, the latter dominance can be, in fact, expressed as $\pi^{\downarrow K \cap L} \ll \kappa^{\downarrow K \cap L}$ (notice that $(\kappa \rhd \nu)^{\downarrow K \cap L} = \kappa^{\downarrow K \cap L}$ due to Lemma 3.2), which guarantees that $\pi \rhd \kappa$ is defined. Therefore, to show that $(\pi \rhd \kappa) \rhd \nu$ is defined, it is enough to prove that $(\pi \rhd \kappa)^{\downarrow M \cap L} \ll \nu^{\downarrow M \cap L}$ (here, we used the fact that $L \supseteq (K \cap M)$ implies $M \cap (K \cup L) = M \cap L$). But this follows from the above-mentioned dominance $\kappa^{\downarrow L \cap M} \ll \nu^{\downarrow L \cap M}$ and the dominance $(\pi \rhd \kappa)^{\downarrow L} \ll \kappa$, guaranteed by Lemma 5.1 by the transitivity of dominance. So, we know that since $\pi \triangleright (\kappa \triangleright \nu)$ is defined, $(\pi \triangleright \kappa) \triangleright \nu$ must be defined, too.

To finish the proof, in the following computations, we will only employ the definition of the operator of composition, equalities $K \cap (L \cup M) = K \cap L$ and $M \cap (K \cup L) = M \cap L$, and Lemma 3.2:

$$\pi \triangleright (\kappa \triangleright \nu) = \frac{\kappa \cdot \nu}{\nu^{\downarrow L \cap M}} \cdot \frac{\pi}{(\kappa \triangleright \nu)^{\downarrow K \cap (L \cup M)}} = \frac{\pi \cdot \kappa \cdot \nu}{\kappa^{\downarrow K \cap L} \cdot \nu^{\downarrow L \cap M}}$$
$$= \frac{\pi \cdot \kappa}{\kappa^{\downarrow K \cap L}} \cdot \frac{\nu}{\nu^{\downarrow M \cap (K \cup L)}} = (\pi \triangleright \kappa) \triangleright \nu. \qquad \square$$

Remark 7.4. If the readers expected that there would be a third theorem saying that $(\pi \triangleright \kappa) \triangleright \nu = \pi \triangleright (\kappa \triangleright \nu)$ under the assumption that $M \supseteq K \cap L$, then we are afraid, they will be disappointed. Going back to Example 7.1, in case (a) we presented a counterexample to such an assertion.

Remark 7.5. Notice that it may happen that $(\pi \triangleright \kappa) \triangleright \nu$ is defined, while $\pi \triangleright (\kappa \triangleright \nu)$ is undefined for the same distributions, as an example of the following distributions shows:

$$\pi(x) \sim \left(\frac{1}{2}, \frac{1}{2}, 0\right), \quad \kappa(x) \sim \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \quad \nu(x) \sim \left(\frac{1}{2}, \frac{1}{2}, 0\right).$$

(Hint: $\pi \triangleright \kappa = \pi$, $\pi \triangleright \nu = \pi$, and $\kappa \triangleright \nu$ are undefined.)

8. Advanced exchange properties

The following simple assertion introduces a property that will be used in several proofs.

LEMMA 8.1. Let $\pi(K)$ and $\kappa(L)$ be distributions for which $K \cap L \subseteq M \subseteq L$, then

$$\pi \triangleright \kappa = (\pi \triangleright \kappa^{\downarrow M}) \triangleright \kappa.$$

Proof. Obviously, $\pi \triangleright \kappa$ is defined *iff* $\pi \triangleright \kappa^{\downarrow M}$ is defined, and therefore, the assertion trivially follows from Theorem 7.3 and Lemma 4.1:

$$(\pi \triangleright \kappa^{\downarrow M}) \triangleright \kappa = \pi \triangleright (\kappa^{\downarrow M} \triangleright \kappa) = \pi \triangleright \kappa. \qquad \square$$

In the rest of this section, we present some other lemmata that state under which conditions one can change ordering of distributions when applying the operator of composition twice. Nevertheless, associativity of the operator of composition will also be discussed in the next section.

LEMMA 8.2. If π and κ are consistent, then for $\pi(K)$, $\kappa(L)$, and $\nu(M)$,

$$L \supseteq (K \cap M) \Longrightarrow (\pi \triangleright \kappa) \triangleright \nu = (\pi \triangleright \nu) \triangleleft \kappa = \kappa \triangleright (\pi \triangleright \nu).$$

Proof. Again we start by showing that, under the given assumptions, $(\pi \triangleright \kappa) \triangleright \nu$ is defined *iff* $(\pi \triangleright \nu) \triangleleft \kappa$ is defined.

Since we assume that π and κ are consistent, the former expression is defined if and only if $(\pi \triangleright \kappa)^{\downarrow M\cap(K\cup L)} \ll \nu^{\downarrow M\cap(K\cup L)}$. Moreover, since $L \supseteq (K \cap M)$, then $M \cap (K \cup L) = M \cap L$, and therefore,

$$(\pi \triangleright \kappa)^{\downarrow M\cap(K\cup L)} = (\pi \triangleright \kappa)^{\downarrow M\cap L} = (\pi \triangleleft \kappa)^{\downarrow M\cap L} = \kappa^{\downarrow M\cap L}.$$

Thus, we get that $(\pi \triangleright \kappa) \triangleright \nu$ is defined *iff*

$$\kappa^{\downarrow L\cap M} \ll \nu^{\downarrow L\cap M}. \tag{9}$$

Notice that because of the consistency of π and κ, dominance (9) guarantees that $\pi \triangleright \nu$ is also defined.

$(\pi \triangleright \nu) \triangleleft \kappa$ is defined *iff* $\pi \triangleright \nu$ is defined and

$$\kappa^{\downarrow L\cap(K\cup M)} \ll (\pi \triangleright \nu)^{\downarrow L\cap(K\cup M)}. \tag{10}$$

Since $K \cap M \subseteq L \cap (K \cup M)$, we can apply Lemma 5.10 getting

$$(\pi \triangleright \nu)^{\downarrow L\cap(K\cup M)} = \pi^{\downarrow L\cap K} \triangleright \nu^{\downarrow L\cap M} = \kappa^{\downarrow L\cap K} \triangleright \nu^{\downarrow L\cap M},$$

where the last equality follows from the consistency of π and κ. Thus, we get that dominance (10) is equivalent to

$$\kappa^{\downarrow L\cap(K\cup M)} \ll \kappa^{\downarrow L\cap K} \triangleright \nu^{\downarrow L\cap M},$$

from which the required dominance (9) is derived by the application of Lemma 5.1 and the transitivity of dominance, which finishes the first part of the proof.

Now, it remains to be shown that $(\pi \triangleright \kappa) \triangleright \nu = \kappa \triangleright (\pi \triangleright \nu)$ in a case where both sides of the equality are defined. For this, in the following computations, we will, respectively, use Theorem 7.3, consistency of π and κ together with Lemma 4.4, Lemma 5.7 and eventually Theorem 7.2.

$$(\pi \triangleright \kappa) \triangleright \nu = \pi \triangleright (\kappa \triangleright \nu) = (\kappa \triangleright \nu) \triangleright \pi = (\kappa \triangleright \pi) \triangleright \nu = \kappa \triangleright (\pi \triangleright \nu).$$

LEMMA 8.3. Consider distributions $\pi(K)$, $\kappa(L)$, and $\nu(M)$. If π and ν are consistent, then

$$K \supseteq (L \cap M) \Longrightarrow (\pi \triangleright \kappa) \triangleright \nu = (\pi \triangleright \kappa) \triangleleft \nu.$$

Proof. This assertion is almost obvious. Both expressions $(\pi \triangleright \kappa) \triangleright \nu$ and $(\pi \triangleright \kappa) \triangleleft \nu$ are not defined if $\pi \triangleright \kappa$ is undefined (i.e. if $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$).

In a case where $\pi \triangleright \kappa$ is defined, then, under the given assumptions $M \cap (K \cup L) = M \cap K$, we get that $(\pi \triangleright \kappa)^{\downarrow M\cap(K\cup L)} = \pi^{\downarrow M\cap K} = \nu^{\downarrow M\cap K}$, and therefore, $\pi \triangleright \kappa$ and ν are consistent. Therefore, both expressions $(\pi \triangleright \kappa) \triangleright \nu$ and $(\pi \triangleright \kappa) \triangleleft \nu$ are defined and equivalent to each other (due to Lemma 4.4). □

LEMMA 8.4. Let $\kappa(L)$ and $\nu(M)$ be consistent. If, for $\pi(K)$, the expression $(\pi \triangleleft \nu) \triangleleft \kappa$ is defined, then

$$M \supseteq (K \cap L) \Longrightarrow (\pi \triangleleft \kappa) \triangleleft \nu = (\pi \triangleleft \nu) \triangleleft \kappa.$$

Proof. Assuming that $(\pi \lhd \nu) \lhd \kappa$ is defined, we know that $\pi \lhd \nu$ is defined, and therefore,

$$\nu^{\downarrow K \cap M} \ll \pi^{\downarrow K \cap M}. \tag{11}$$

The other expression $(\pi \lhd \kappa) \lhd \nu$ is defined *iff* $\kappa^{K \cap L} \ll \pi^{K \cap L}$, which is, due to the given assumptions, equivalent to

$$\nu^{\downarrow K \cap L} \ll \pi^{\downarrow K \cap L}, \tag{12}$$

and

$$\nu^{\downarrow M \cap (K \cup L)} \ll (\pi \lhd \kappa)^{\downarrow M \cap (K \cup L)}. \tag{13}$$

However, (11) $\Longrightarrow$(12) is obvious and the dominance (13) can be reformulated, using the equivalence following from Lemma 5.10, the given assumptions and the definition of the operators

$$(\pi \lhd \kappa)^{\downarrow M \cap (K \cup L)} = \pi^{\downarrow M \cap K} \lhd \kappa^{\downarrow M \cap L} = \pi^{\downarrow M \cap K} \lhd \nu^{\downarrow M \cap L} = \nu^{\downarrow M \cap L} \rhd \pi^{\downarrow M \cap K}$$

into an equivalent form

$$\nu^{\downarrow M \cap L} = \nu^{\downarrow M \cap (K \cup L)} \ll \nu^{\downarrow M \cap L} \rhd \pi^{\downarrow M \cap K} = \nu^{\downarrow M \cap L},$$

which always holds. So, we have shown that if $(\pi \lhd \nu) \lhd \kappa$ is defined, then $(\pi \lhd \kappa) \lhd \nu$ is also defined.

Now let us start proving that if the respective expressions are defined, they are equal to each other. For this, we will use the definitions of the operators of composition, Theorems 7.2 and 7.3, and Lemma 4.4.

$$(\pi \lhd \kappa) \lhd \nu = \nu \rhd (\kappa \rhd \pi) = (\nu \rhd \kappa) \rhd \pi = (\kappa \rhd \nu) \rhd \pi = \kappa \rhd (\nu \rhd \pi)$$
$$= (\pi \lhd \nu) \lhd \kappa. \qquad \square$$

LEMMA 8.5. If $M \supseteq (K \cap L)$, then for $\pi(K)$, $\kappa(L)$ and $\nu(M)$,

$$(\pi \lhd \kappa) \lhd \nu = ((\pi \lhd \nu) \lhd \kappa) \lhd \nu \tag{14}$$

holds true if the right-hand side of the formula is defined.

Proof. Knowing that $((\pi \lhd \nu) \lhd \kappa) \lhd \nu$ is defined, $(\pi \lhd \nu) \lhd \kappa$ must be defined, too, and, due to preceding Lemma 8.4, $(\pi \lhd \nu) \lhd \kappa = (\pi \lhd \kappa) \lhd \nu$. Therefore,

$$((\pi \lhd \nu) \lhd \kappa) \lhd \nu = ((\pi \lhd \kappa) \lhd \nu) \lhd \nu = \nu \rhd ((\pi \lhd \kappa) \lhd \nu) = (\pi \lhd \kappa) \lhd \nu,$$

where the last modification is simply an application of Lemma 4.1. $\square$

9. Anticipating operator

In this section, we will introduce and study the properties of another operator, which will define a special type of composition of two distributions.

DEFINITION 9.1. Consider an arbitrary subset $K \subset N$ and two distributions $\kappa(L)$ and $\nu(M)$. Their *anticipating composition* is given by the formula:

$$\kappa \circledRightTriangle_K \nu = (\nu^{\downarrow(K\setminus L)\cap M} \cdot \kappa) \vartriangleright \nu.$$

The operator $\circledRightTriangle_K$ will be called an *anticipating operator of composition*.

Remark 9.2. Notice that this operator is parameterized by an index set, which is the main difference with respect to the previously defined operators $\vartriangleright$ and $\vartriangleleft$. In Theorem 9.4, we will articulate the main purpose for which this operator is introduced. Namely, operator $\vartriangleright$ can be substituted by an anticipating operator simultaneously with changing the ordering of operations. The purpose of the parameter K will be intuitively explained in Remark 9.8.

Remark 9.3. It should be stressed that since the anticipating operator is defined with the help of the operator of right composition, it may happen that the result remains undefined. It follows immediately from the respective definitions that $\kappa \circledRightTriangle_K \nu$ is defined *iff* $\kappa \vartriangleright \nu$ is defined, i.e. if $\kappa^{\downarrow L\cap M} \ll \nu^{\downarrow L\cap M}$. Moreover, notice that distribution $\kappa \circledRightTriangle_K \nu$, if defined, is defined for the same set of variables as the distribution $\kappa \vartriangleright \nu$, and that $(\kappa(L) \circledRightTriangle_K \nu(M))^{\downarrow L} = \kappa(L)$.

THEOREM 9.4. If $\pi(K)$, $\kappa(L)$, and $\nu(M)$ are such that $\pi \vartriangleright (\kappa \circledRightTriangle_K \nu)$ is defined, then

$$(\pi \vartriangleright \kappa) \vartriangleright \nu = \pi \vartriangleright (\kappa \circledRightTriangle_K \nu).$$

To prove this theorem, we need the following simple auxiliary assertion.

LEMMA 9.5. Consider distributions $\pi(K)$ and $\kappa(L)$. If $\pi^{\downarrow M} \ll \nu^{\downarrow M}$ for $M \subseteq K\setminus L$, then

$$\pi \vartriangleright \kappa = \pi \vartriangleright (\nu^{\downarrow M} \cdot \kappa),$$

if any of the expressions is defined.

Proof. In the following computations, everything is correct, because we assume $\pi^{\downarrow M} \ll \nu^{\downarrow M}$ (i.e. if there is a zero in a denominator, then there are at least two zeroes in the respective numerator and the result is considered to be zero):

$$\pi \vartriangleright (\nu^{\downarrow M} \cdot \kappa) = \frac{\pi \cdot \nu^{\downarrow M} \cdot \kappa}{(\nu^{\downarrow M} \cdot \kappa)^{\downarrow K\cap(L\cup M)}} = \frac{\pi \cdot \nu^{\downarrow M} \cdot \kappa}{\nu^{\downarrow M} \cdot \kappa^{\downarrow K\cap L}} = \frac{\pi \cdot \kappa}{\kappa^{\downarrow K\cap L}} = \pi \vartriangleright \kappa. \qquad \square$$

Proof of Theorem 9.4. Assume that $\pi \vartriangleright (\kappa \circledRightTriangle_K \nu)$ is defined. It means that

$$\pi^{\downarrow K\cap(L\cup M)} \ll (\kappa \circledRightTriangle_K \nu)^{\downarrow K\cap(L\cup M)}, \tag{15}$$

and, as a consequence of the fact that dominance also holds for the respective marginal distributions, $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$. This guarantees that $\pi \vartriangleright \kappa$ is defined.

Let us now show by contradiction that $(\pi \vartriangleright \kappa) \vartriangleright \nu$ must also be defined. Assume it is not. It means that there exists $x \in \mathbf{X}_{K\cup L\cup M}$ such that $\nu(x^{\downarrow M}) = 0$ and simultaneously

$(\pi \triangleright \kappa)(x^{\downarrow K\cup L}) > 0$ (and thus also $\kappa(x^{\downarrow L}) > 0$), but this contradicts our assumption that $\kappa \circledtriangleright_K \nu$ is defined.

Now, knowing that $(\pi \triangleright \kappa) \triangleright \nu$ is defined, we can apply Lemma 9.5, getting

$$(\pi \triangleright \kappa) \triangleright \nu = (\pi \triangleright (\nu^{\downarrow M} \cdot \kappa)) \triangleright \nu.$$

This enables us to complete the proof by applying Theorem 7.3:

$$\pi \triangleright (\kappa \circledtriangleright_K \nu) = \pi \triangleright ((\nu^{\downarrow M} \cdot \kappa) \triangleright \nu) = (\pi \triangleright (\nu^{\downarrow M} \cdot \kappa)) \triangleright \nu = (\pi \triangleright \kappa) \triangleright \nu. \qquad \square$$

Remark 9.6. Notice that, in the same way as Theorems 7.2 and 7.3, the assertion does not claim that the equality holds true when $(\pi \triangleright \kappa) \triangleright \nu$ is defined. This is because it may also happen here that $(\pi \triangleright \kappa) \triangleright \nu$ is defined and $\pi \triangleright (\kappa \circledtriangleright_K \nu)$ is not. The reader can show it with the distributions π, κ, and ν from Remark 7.5.

Remark 9.7. It should be noted here that the computational complexity of the composition $\kappa \circledtriangleright_K \nu$ does not differ substantially from the complexity of the computation of $\kappa \triangleright \nu$. It follows, namely, from the fact that both of these distributions are of the same dimensionality; and both are defined for variables $X_{L\cup M}$. In other words, in both cases, we have to compute the same number of probability values.

Remark 9.8. As we have already noted above, the operator is parameterized by the index set K. The purpose of the operator is, namely, to compose the distributions (in our case distributions κ and ν), but to simultaneously introduce the necessary independence of variables $X_{(K\setminus L)\cap M}$ and X_L that would otherwise be omitted. If we want to compose distributions κ and ν before π is considered, we have to 'anticipate' the independence which was originally introduced by the previous operator. This also explains why the operator $\circledtriangleright_K \nu$ is called an *anticipating operator*.

Example 9.9. As stated previously, the specific purpose of the anticipating operator is to introduce the necessary conditional independence that would otherwise be omitted. To illustrate the point, let us consider three distributions $\pi(x_1)$, $\kappa(x_2)$, and $\nu(x_1, x_2)$ for which $(\pi(x_1) \triangleright \kappa(x_2)) \triangleright \nu(x_1, x_2) = \pi(x_1)\kappa(x_2)$. If we used the operator $\triangleright$ instead of $\circledtriangleright_K$, we would get

$$\pi(x_1) \triangleright \left(\kappa(x_2) \triangleright \nu(x_1, x_2)\right) = \frac{\pi(x_1)(\kappa(x_2)\nu(x_1|x_2))}{\sum_{y\in\mathbf{X}_1} \kappa(x_2)\nu(y|x_2)},$$

which evidently differs from $\pi(x_1)\kappa(x_2)$, because $\pi \triangleright (\kappa \triangleright \nu)$ inherits the dependence of variables X_1 and X_2 from ν. Nevertheless, considering

$$\begin{aligned}\pi(x_1) \triangleright \left(\kappa(x_2) \circledtriangleright_{\{1\}} \nu(x_1, x_2)\right) &= \pi(x_1) \triangleright \left(\nu(x_1)\kappa(x_2) \triangleright \nu(x_1, x_2)\right)\\ &= \pi(x_1) \triangleright \nu(x_1)\kappa(x_2) = \pi(x_1)\kappa(x_2),\end{aligned}$$

we get the desired result.

Perhaps it is also worth mentioning that, in this example, $\kappa(x_2) \circledtriangleright_{\{1\}} \nu(x_1, x_2) = \nu(x_1)\kappa(x_2)$ is not a marginal distribution of the resulting $(\pi(x_1) \triangleright \kappa(x_2)) \triangleright \nu(x_1, x_2)$.

Example 9.10. Let us present another, slightly more complex, example illustrating the application of the anticipating operator. This time consider distributions $\pi(x_1, x_2, x_3, x_4)$, $\kappa(x_2, x_3, x_5)$, and $\nu(x_3, x_4, x_6)$. In this case, according to Theorem 9.4,

$$\begin{aligned}\big(\pi(x_1, x_2, x_3, x_4) \triangleright \kappa(x_2, x_3, x_5)\big) &\triangleright \nu(x_3, x_4, x_6)\\ &= \pi(x_1, x_2, x_3, x_4) \triangleright \big(\kappa(x_2, x_3, x_5) \circledtriangleright_{\{1,2,3,4\}} \nu(x_3, x_4, x_6)\big).\end{aligned}$$

According to the definition of the anticipating operator, the expression in parentheses equals

$$\nu(x_4)\kappa(x_2, x_3, x_5) \triangleright \nu(x_3, x_4, x_6) = \nu(x_4)\kappa(x_2, x_3, x_5)\nu(x_6|x_3, x_4). \tag{16}$$

The reader most likely notices that, thanks to the anticipating operator, $\nu(x_6|x_3, x_4)$ appears in this formula, which is exactly the form at which ν occurs in

$$\begin{aligned}\big(\pi(x_1, x_2, x_3, x_4) \triangleright \kappa(x_2, x_3, x_5)\big) &\triangleright \nu(x_3, x_4, x_6)\\ &= \pi(x_1, x_2, x_3, x_4)\kappa(x_5|x_2, x_3)\nu(x_6|x_3, x_4).\end{aligned}$$

Moreover, formula (16) allows for simple computation of the marginal required in the next step:

$$\big(\kappa(x_2, x_3, x_5) \circledtriangleright_{\{1,2,3,4\}} \nu(x_3, x_4, x_6)\big)^{\downarrow \{2,3,4\}} = \nu(x_4)\kappa(x_2, x_3).$$

Therefore,

$$\begin{aligned}\pi(x_1, x_2, x_3, x_4) &\triangleright \big(\kappa(x_2, x_3, x_5) \circledtriangleright_{\{1,2,3,4\}} \nu(x_3, x_4, x_6)\big)\\ &= \pi(x_1, x_2, x_3, x_4)\frac{\nu(x_4)\kappa(x_2, x_3, x_5)\nu(x_6|x_3, x_4)}{\nu(x_4)\kappa(x_2, x_3)}\\ &= \pi(x_1, x_2, x_3, x_4)\kappa(x_5|x_2, x_3)\nu(x_6|x_3, x_4).\end{aligned}$$

10. Generating sequences

Beginning with this section, the reader will learn how the operators of composition are used to construct *multidimensional compositional models* and about the properties of these models. Therefore, it may be useful to summarize the most important properties of the operators of composition that we studied in the previous sections.

- Composing two distributions, we can define a distribution of a higher dimensionality than any of the original ones.
- One of the distributions is always a marginal of the composed distribution (for $\triangleright$ it is the first one and for $\triangleleft$ it is the second one).
- The operator is neither commutative nor associative. Nevertheless, there are special situations under which both commutativity and associativity hold: the commutativity holds for consistent distributions, and the associativity holds when the first (or second) distribution is defined for the set of variables containing an intersection of arguments of the remaining two distributions.

- Operations of marginalization and composition commute only if the resulting marginal distribution is defined for all variables which appear among the arguments of both the original distributions.
- There are many 'special situations' under which an order of operators of composition can be changed without influencing the resulting composed distribution. Since we will use them quite often from now on, we present a list of the most important assertions in Table 8.

To avoid some technical problems and the necessity of repeating some assumptions to excess, let us make three important conventions.

In this and all the remaining sections, we will consider a system of n oligodimensional distributions $\pi_1(K_1)$, $\pi_2(K_2), \cdots, \pi_n(K_n)$. Therefore, whenever we speak about a distribution π_k, if not explicitly specified otherwise (usually in examples), the distribution π_k will always be assumed to be a distribution from $\Pi^{(K_k)}$, which means it will be a distribution $\pi_k(K_k)$. Thus, formulae $\pi_1 \triangleright \pi_2 \triangleright \cdots \triangleright \pi_n$ and $\pi_1 \triangleleft \pi_2 \triangleleft \cdots \triangleleft \pi_n$, if they are defined, will determine the distributions of variables $X_{K_1 \cup K_2 \cup \cdots \cup K_n}$.

Our second convention pertains to the fact that (as we know from the preceding sections) the operators of composition are neither commutative nor associative. To avoid having to write too many parentheses in the formulae, in the rest of the paper, we will apply the operators from left to right. Thus

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \cdots \triangleright \pi_n = (\cdots((\pi_1 \triangleright \pi_2) \triangleright \pi_3) \triangleright \cdots \triangleright \pi_n),$$

and analogously also

$$\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \cdots \triangleleft \pi_n = (\cdots((\pi_1 \triangleleft \pi_2) \triangleleft \pi_3) \triangleleft \cdots \triangleleft \pi_n).$$

So the parentheses will be used only when we want to change this default ordering. Therefore, to construct a multidimensional distribution, it is sufficient to determine a sequence – we will call it a *generating sequence* – of oligodimensional distributions. However, since generally

$$\pi_1 \triangleright \pi_2 \triangleright \cdots \triangleright \pi_n \neq \pi_1 \triangleleft \pi_2 \triangleleft \cdots \triangleleft \pi_n,$$

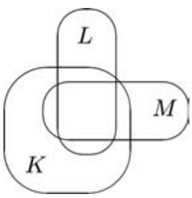
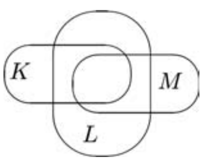
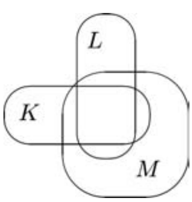
it is necessary to determine which of the operators, $\triangleright$ or $\triangleleft$, is used for composition. If not explicitly stated otherwise, we will usually consider operator $\triangleright$. This is because these two operators, when applied iteratively to a generating sequence, substantially differ from a computational point of view. To realize it, consider application of the kth operator in the sequences $\pi_1 \triangleright \pi_2 \triangleright \cdots \triangleright \pi_n$ and $\pi_1 \triangleleft \pi_2 \triangleleft \cdots \triangleleft \pi_n$. In the former case, when computing

$$(\pi_1 \triangleright \cdots \triangleright \pi_k) \triangleright \pi_{k+1} = \frac{(\pi_1 \triangleright \cdots \triangleright \pi_k)\pi_{k+1}}{\pi_{k+1}^{\downarrow K_{k+1} \cap (K_1 \cup \cdots \cup K_k)}},$$

one has to marginalize distribution π_{k+1}, which is assumed to be a low-dimensional distribution. On the other hand, computation of

$$(\pi_1 \triangleleft \cdots \triangleleft \pi_k) \triangleleft \pi_{k+1} = \frac{(\pi_1 \triangleleft \cdots \triangleleft \pi_k)\pi_{k+1}}{(\pi_1 \triangleleft \cdots \triangleleft \pi_k)^{\downarrow K_{k+1} \cap (K_1 \cup \cdots \cup K_k)}}$$

Table 8. Survey of assertions enabling alteration of order of compositions.

Structural property	Requirement on consistency	Statement	Reference
–	–	$(\pi \triangleright \kappa) \triangleright \nu$ $= \pi \triangleright (\kappa \bigcirc\!\!\!\!\!\triangleright_{K} \nu)$	Theorem 9.4
$K \supseteq (L \cap M)$	–	$(\pi \triangleright \kappa) \triangleright \nu$ $= (\pi \triangleright \nu) \triangleright \kappa$	Lemma 5.7
	π, ν	$(\pi \triangleright \kappa) \triangleright \nu$ $= (\pi \triangleright \kappa) \triangleleft \nu$	Lemma 8.3
	–	$(\pi \triangleright \kappa) \triangleright \nu$ $= \pi \triangleright (\kappa \triangleright \nu)$	Theorem 7.2
$L \supseteq (K \cap M)$	π, κ	$(\pi \triangleright \kappa) \triangleright \nu$ $= (\pi \triangleright \nu) \triangleleft \kappa$	Lemma 8.2
	–	$(\pi \triangleright \kappa) \triangleright \nu$ $= \pi \triangleright (\kappa \triangleright \nu)$	Theorem 7.3
$M \supseteq (K \cap L)$	κ, ν	$(\pi \triangleleft \kappa) \triangleleft \nu$ $= (\pi \triangleleft \nu) \triangleleft \kappa$	Lemma 8.4
	–	$(\pi \triangleleft \kappa) \triangleleft \nu$ $= ((\pi \triangleleft \nu) \triangleleft \kappa) \triangleleft \nu$	Lemma 8.5

can only be done when computation of

$$(\pi_1 \lhd \cdots \lhd \pi_k)^{\downarrow K_{k+1} \cap (K_1 \cup \cdots \cup K_k)}$$

is tractable; since the distribution $(\pi_1 \lhd \cdots \lhd \pi_k)$ can be of a very high dimension, its marginalization may become computationally expensive.

The third aforementioned convention is of a rather technical nature. The reader may have noticed that in the previous sections, we spent a substantial part of some of the proofs clarifying situations when the result of composition is not defined. Since in the remaining part of the paper we are interested in a construction of multidimensional models, it is quite natural that we will always assume that all the models we speak about are well defined.

10.1 *Perfect sequences*

Not all generating sequences are equally efficient in their representations of multidimensional distributions. Among them, so-called perfect sequences hold an important position (Jiroušek 1997).

Definition 10.1. A generating sequence of probability distributions $\pi_1, \pi_2, \cdots, \pi_n$ is called *perfect* if $\pi_1 \rhd \cdots \rhd \pi_n$ is defined and

$$\begin{aligned}
\pi_1 \rhd \pi_2 &= \pi_1 \lhd \pi_2, \\
\pi_1 \rhd \pi_2 \rhd \pi_3 &= \pi_1 \lhd \pi_2 \lhd \pi_3, \\
&\vdots \\
\pi_1 \rhd \pi_2 \rhd \cdots \rhd \pi_n &= \pi_1 \lhd \pi_2 \lhd \cdots \lhd \pi_n.
\end{aligned}$$

From this definition, one can hardly see the importance of perfect sequences. This importance becomes clearer from the characterization theorem that follows. First, however, let us present a technical property, which, being an immediate consequence of an inductive application of Lemma 4.4, is presented without a proof.

Lemma 10.2. A sequence $\pi_1, \pi_2, \cdots, \pi_n$ is perfect if and only if the pairs of distributions $(\pi_1 \rhd \cdots \rhd \pi_{m-1})$ and π_m are consistent for all $m = 2, 3, \cdots, n$.

Theorem 10.3. A sequence of distributions $\pi_1, \pi_2, \cdots, \pi_n$ is perfect *iff* all the distributions from this sequence are marginals of the distribution $(\pi_1 \rhd \pi_2 \rhd \cdots \rhd \pi_n)$.

Proof. The fact that all distributions π_k from a perfect sequence are marginals of $(\pi_1 \rhd \pi_2 \rhd \cdots \rhd \pi_n)$ follows from the fact that $(\pi_1 \rhd \cdots \rhd \pi_k)$ is marginal to $(\pi_1 \rhd \cdots \rhd \pi_n)$ and π_k is marginal to $(\pi_1 \lhd \cdots \lhd \pi_k)$ (see Lemma 3.2).

Suppose that for all $k = 1, \cdots, n$, π_k are marginal distributions of $(\pi_1 \rhd \cdots \rhd \pi_n)$. It means that all the distributions from the sequence are pairwise consistent, and that each π_k is consistent with any marginal distribution of $(\pi_1 \rhd \cdots \rhd \pi_n)$. Therefore, π_1 and π_2 are consistent, and due to Lemma 4.4

$$\pi_1 \rhd \pi_2 = \pi_1 \lhd \pi_2.$$

Since $\pi_1 \rhd \pi_2$ is marginal to $(\pi_1 \rhd \cdots \rhd \pi_n)$ (Lemma 3.2), it must be consistent with π_3, too. Using Lemma 4.4 again, we get

$$\pi_1 \rhd \pi_2 \rhd \pi_3 = \pi_1 \lhd \pi_2 \lhd \pi_3.$$

However, $\pi_1 \rhd \pi_2 \rhd \pi_3$, being marginal to $(\pi_1 \rhd \cdots \rhd \pi_n)$, must also be consistent with π_4 and we can continue in this manner until we conclude that for all $k = 2, \cdots, n$

$$\pi_1 \rhd \pi_2 \rhd \cdots \rhd \pi_k = \pi_1 \lhd \pi_2 \lhd \cdots \lhd \pi_k. \qquad \square$$

Example 10.4. The theorem presented above states that a model defined by a generating sequence preserves all the given marginals *iff* the model is defined by a perfect sequence. If the considered generating sequence is not perfect, then some of the marginal distributions differ from the given ones. In this example, we show that a non-perfect generating sequence need not preserve one-dimensional marginal distributions – even in the case when the given oligodimensional distributions are pairwise consistent.

Consider distribution $\pi(x_1, x_2, x_3)$ from Table 9. It is obvious that three distributions $\pi^{\downarrow\{1\}}$, $\pi^{\downarrow\{2\}}$, and π must be pairwise consistent. Let us deal with the distribution defined by generating sequence $\pi^{\downarrow\{1\}}$, $\pi^{\downarrow\{2\}}$, and π, i.e. with the distribution

$$\pi^{\downarrow\{1\}}(x_1) \rhd \pi^{\downarrow\{2\}}(x_2) \rhd \pi(x_1, x_2, x_3).$$

Since both the considered one-dimensional marginal distributions $\pi^{\downarrow\{1\}}$ and $\pi^{\downarrow\{2\}}$ are uniform, their composition $\pi^{\downarrow\{1\}} \rhd \pi^{\downarrow\{2\}}$ is also uniform. Thus, it is an easy task to compute distribution $\kappa = \pi^{\downarrow\{1\}}(x_1) \rhd \pi^{\downarrow\{2\}}(x_2) \rhd \pi(x_1, x_2, x_3)$, which is shown in Table 10.

Summing up entries in the rows of Tables 9 and 10, we get the respective one-dimensional marginal distributions $\pi^{\downarrow\{3\}}$ and $\kappa^{\downarrow\{3\}}$, respectively, from which we see that these distributions are different:

$$\begin{aligned}
&\pi^{\downarrow\{3\}}(x_3 = 0) = 0.5, \quad \pi^{\downarrow\{3\}}(x_3 = 1) = 0.2, \quad \pi^{\downarrow\{3\}}(x_3 = 2) = 0.3,\\
&\kappa^{\downarrow\{3\}}(x_3 = 0) = \tfrac{13}{24}, \quad \kappa^{\downarrow\{3\}}(x_3 = 1) = \tfrac{5}{24}, \quad \kappa^{\downarrow\{3\}}(x_3 = 2) = \tfrac{6}{24}.
\end{aligned}$$

Remark 10.5. What is the main message conveyed by the characterization Theorem 10.3? Considering that low-dimensional distributions π_k are carriers of local information, the constructed multidimensional distribution, if it is a perfect sequence model, represents global information, faithfully reflecting all of the local input. This is why we will be so interested in perfect sequence models.

Table 9. Three-dimensional distribution.

	$x_1 = 0$		$x_1 = 1$	
π	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	0.1	0.1	0.2	0.1
$x_3 = 1$	0.0	0.1	0.0	0.1
$x_3 = 2$	0.2	0.0	0.0	0.1

Table 10. Distribution $\kappa(x_1,x_2,x_3) = \pi^{\downarrow\{1\}}(x_1) \triangleright \pi^{\downarrow\{2\}}(x_2) \triangleright \pi(x_1,x_2,x_3)$.

	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{6}{24}$	$\frac{2}{24}$
$x_3 = 1$	0	$\frac{3}{24}$	0	$\frac{2}{24}$
$x_3 = 2$	$\frac{4}{24}$	0	0	$\frac{2}{24}$

Remark 10.6. From Theorem 10.3 and the definition of a perfect sequence, it is obvious that for perfect sequence $\pi_1, \cdots, \pi_n$, all the distributions $\pi_k\,(k = 1, \cdots, n)$ are marginals of $\pi_1 \triangleleft \cdots \triangleleft \pi_n$. It should, however, be stressed that, as we will show in the following example, it does not mean that if all $\pi_1, \cdots, \pi_n$ are marginal to $\pi_1 \triangleleft \cdots \triangleleft \pi_n$ that the considered sequence is perfect.

Example 10.7. Consider a sequence $\pi_1(x_1,x_2), \pi_2(x_2,x_3), \pi_3(x_3,x_4)$ and assume it is perfect. Thus, we know that all π_1, π_2, and π_3 are marginal distributions of $\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleleft \pi_2 \triangleleft \pi_3$ and all three distributions π_1, π_2, and π_3 are pairwise consistent. Since $\{2,3\} \supset \{1,2\} \cap \{3,4\}$, we can apply Lemma 8.2, from which we get

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleleft \pi_2 = \pi_1 \triangleleft \pi_3 \triangleleft \pi_2.$$

(The last modification is possible because of Lemma 4.4.) Thus, we got that all π_1, π_2, and π_3 are also marginal distributions of $\pi_1 \triangleleft \pi_3 \triangleleft \pi_2$.

The question is whether the sequence π_1, π_3, π_2 is also perfect. Using Lemma 10.2, we see it is perfect only if distributions $\pi_1 \triangleright \pi_3$ and π_2 are consistent. However, they are consistent only when $(\pi_1 \triangleright \pi_3)^{\downarrow\{2,3\}} = \pi_2$, which generally need not be true, because $(\pi_1 \triangleright \pi_3)^{\downarrow\{2,3\}} = \pi_1 \cdot \pi_3$. Therefore, $\pi_1 \triangleright \pi_3$ and π_2 are consistent only when we consider π_2 to be a distribution of two independent variables ($X_2 \perp\!\!\!\perp X_3[\pi_2]$). We thus see that all distributions π_1, π_2, and π_3 are marginals of $\pi_1 \triangleleft \pi_3 \triangleleft \pi_2$ and yet the sequence π_1, π_3, π_2 need not be perfect.

Remark 10.8. Notice that when defining a perfect sequence, let alone a generating sequence, we have not imposed any conditions on sets of variables for which the distributions were defined. For example, considering a generating sequence where one distribution is defined for a subset of variables of another distribution (i.e. $K_j \subset K_k$), is fully sensible, and may provide some information about the distribution. If $\pi^{\downarrow\{1\}}, \pi^{\downarrow\{2\}}, \pi$ is a perfect sequence, it is quite obvious that

$$\pi^{\downarrow\{1\}} \triangleright \pi^{\downarrow\{2\}} \triangleright \pi = \pi$$

(because all the elements of a perfect sequence are marginals of the resulting distribution, and therefore, π must be marginal to $\pi^{\downarrow\{1\}} \triangleright \pi^{\downarrow\{2\}} \triangleright \pi$). Nevertheless, it can happen that, for some reason or another, it may be more advantageous to work with the model defined by the perfect sequence than just with the distribution π. From this model, one can immediately see that variables X_1 and X_2 are independent, which, not knowing the numbers defining the distribution, one cannot say about distribution π. (In Jiroušek's

(2008) paper, it was described how to read all the conditional independence relations from a compositional model.)

10.2 *Perfectization*

The following assertion shows that each generating sequence (recall that we continually assume that $\pi_1 \triangleright \cdots \triangleright \pi_n$ is defined) can be transformed into a perfect sequence (it is, in a way, a generalization of Lemma 5.11).

THEOREM 10.9. For any generating sequence $\pi_1, \pi_2, \cdots, \pi_n$, the sequence $\kappa_1, \kappa_2, \cdots, \kappa_n$ computed by the following process

$$\begin{aligned}
\kappa_1 &= \pi_1, \\
\kappa_2 &= \kappa_1^{\downarrow K_2 \cap K_1} \triangleright \pi_2, \\
\kappa_3 &= (\kappa_1 \triangleright \kappa_2)^{\downarrow K_3 \cap (K_1 \cup K_2)} \triangleright \pi_3, \\
&\vdots \\
\kappa_n &= (\kappa_1 \triangleright \cdots \triangleright \kappa_{n-1})^{\downarrow K_n \cap (K_1 \cup \cdots \cup K_{n-1})} \triangleright \pi_n
\end{aligned}$$

is perfect and

$$\pi_1 \triangleright \cdots \triangleright \pi_n = \kappa_1 \triangleright \cdots \triangleright \kappa_n.$$

Proof. The perfectness of the sequence $\kappa_1, \cdots, \kappa_n$ follows immediately from Lemma 10.2 and from the definition of this sequence, as

$$\kappa_i^{\downarrow K_i \cap (K_1 \cup \cdots \cup K_{i-1})} = (\kappa_1 \triangleright \cdots \triangleright \kappa_{i-1})^{\downarrow K_i \cap (K_1 \cup \cdots \cup K_{i-1})}$$

yields consistency of $(\kappa_1 \triangleright \cdots \triangleright \kappa_{i-1})$ and κ_i.

Let us prove that $\pi_1 \triangleright \cdots \triangleright \pi_n = \kappa_1 \triangleright \cdots \triangleright \kappa_n$ by mathematical induction. Since $\pi_1 = \kappa_1$ by definition, it is enough to show that $\pi_1 \triangleright \cdots \triangleright \pi_i = \kappa_1 \triangleright \cdots \triangleright \kappa_i$ implies also $\pi_1 \triangleright \cdots \triangleright \pi_{i+1} = \kappa_1 \triangleright \cdots \triangleright \kappa_{i+1}$. In the following computations, we will use the fact that, due to Lemma 5.10,

$$(\kappa_1 \triangleright \cdots \triangleright \kappa_i)^{\downarrow K_{i+1} \cap (K_1 \cup \cdots K_i)} \triangleright \pi_{i+1} = ((\kappa_1 \triangleright \cdots \triangleright \kappa_i) \triangleright \pi_{i+1})^{\downarrow K_{i+1}},$$

and afterwards we will employ Lemma 5.11.

$$\begin{aligned}
\kappa_1 \triangleright \cdots \triangleright \kappa_{i+1} &= \kappa_1 \triangleright \cdots \triangleright \kappa_i \triangleright ((\kappa_1 \triangleright \cdots \triangleright \kappa_i)^{\downarrow K_{i+1} \cap (K_1 \cup \cdots K_i)} \triangleright \pi_{i+1}) \\
&= \kappa_1 \triangleright \cdots \triangleright \kappa_i \triangleright \left((\kappa_1 \triangleright \cdots \triangleright \kappa_i) \triangleright \pi_{i+1}\right)^{\downarrow K_{i+1}} \\
&= \kappa_1 \triangleright \cdots \triangleright \kappa_i \triangleright \pi_{i+1} = \pi_1 \triangleright \cdots \triangleright \pi_i \triangleright \pi_{i+1}. \qquad \square
\end{aligned}$$

Remark 10.10. From the theoretical point of view, the process of perfectization described by Theorem 10.9 is simple. Unfortunately, it is not valid from the point of view of computational complexity. The process requires marginalization of models, which are distributions, represented by generating sequences. As we have already said, this problem is not be studied in this paper but it may be computationally very expensive (Jiroušek 2000).

10.3 *Running intersection property*

Having a generating sequence, one should apply Lemma 10.2 to verify whether the sequence is perfect or not. Unfortunately, it may happen that this verification is not easy. Nevertheless, in some special situations, one can see the answer immediately. For example, the reader can easily prove that an arbitrary sequence of uniform distributions is perfect. More important situations in which verification of perfectness is simple are described in Lemma 10.13. It is, in fact, just a reformulation of a classical result of Kellerer (1964) into the language of this paper. To formulate it, let us recall an important concept that is not new to the reader familiar with decomposable models (see, e.g. Pearl 1988).

Definition 10.11. A sequence of sets $K_1, K_2, \cdots, K_n$ is said to meet a *running intersection property* (RIP, in the sections that follow), if

$$\forall i = 2, \cdots, n \;\; \exists j (1 \leq j < i) \left(K_i \cap \left(\bigcup_{k=1}^{i-1} K_k \right) \subseteq K_j \right).$$

In the field of graphical Markov models, the notion of a RIP is quite frequent. Therefore, it is not surprising that we will also meet with it several times in the following text, and we will also use the following famous property (which follows, for example, from the properties of acyclic hypergraphs).

Lemma 10.12. If a sequence of sets $K_1, K_2, \cdots, K_n$ meets RIP, then for each $\ell \in \{1, 2, \cdots, n\}$ there exists a permutation $i_1, i_2, \cdots, i_n$ such that $\ell = i_1$ and $K_{i_1}, K_{i_2}, \cdots, K_{i_n}$ meets RIP, too.

Lemma 10.13. If $\pi_1, \pi_2, \cdots, \pi_n$ is a sequence of pairwise consistent oligodimensional probability distributions such that $K_1, \cdots, K_n$ meets RIP, then this sequence is perfect.

Proof. The proof is performed by mathematical induction. Since we assume that all the distributions are pairwise consistent, $\pi_1 \triangleright \pi_2 = \pi_1 \triangleleft \pi_2$, and the sequence π_1, π_2 is perfect. Therefore, assuming that the assertion is valid for $m - 1$, the proof will be finished by showing it holds also for m.

Consider pairwise consistent $\pi_1, \pi_2, \cdots, \pi_m$, for which $K_1, K_2, \cdots, K_m$ meets RIP and $\pi_1, \pi_2, \cdots, \pi_{m-1}$ is perfect. Thus to show that $\pi_1, \pi_2, \cdots, \pi_m$ is perfect, it is enough to show that $(\pi_1 \triangleright \cdots \triangleright \pi_{m-1})$ and π_m are consistent.

Since $K_1, \cdots, K_n$ meets RIP, $K_m \cap (K_1 \cup \cdots \cup K_{m-1})$ must be a subset of K_k for some $k \leq m - 1$. Therefore, $K_m \cap (K_1 \cup \cdots \cup K_{m-1}) = K_m \cap K_k$. The assumption of induction says that, due to Theorem 10.3, all π_ℓ $(1 \leq \ell < m)$ are marginal to $\pi_1 \triangleright \cdots \triangleright \pi_{m-1}$ and thus

$$(\pi_1 \triangleright \cdots \triangleright \pi_{m-1})^{\downarrow K_m \cap (K_1 \cup \cdots \cup K_{m-1})} = \pi_k^{\downarrow K_m \cap K_k} = \pi_m^{\downarrow K_m \cap K_k},$$

where the last equality follows from the fact that π_k and π_m are assumed to be consistent. Thus, we have shown that $(\pi_1 \triangleright \cdots \triangleright \pi_{m-1})$ and π_m are consistent, which finishes the proof. □

10.4 *Shannon entropy*

If a generating sequence $\pi_1, \pi_2, \cdots, \pi_n$ (with $n > 1$) is perfect, then it can always be reordered in a way that its permutation $\pi_{i_1}, \pi_{i_2}, \cdots, \pi_{i_n}$ is also perfect. Trivially, if $\pi_1, \pi_2, \pi_3, \cdots, \pi_n$ is perfect, then $\pi_2, \pi_1, \pi_3, \cdots, \pi_n$ must be perfect, too. To be able to show that all such perfect sequences define the same multidimensional distribution, we will need the following assertion proving that perfect sequence models always achieve maximum entropy, in a certain sense (it is, in fact, a generalization of Theorem 5.4).

THEOREM 10.14. Denote $\Xi = \{\pi_1, \pi_2, \cdots, \pi_n\}$ a system of probability distributions. If the sequence $\pi_1, \pi_2, \cdots, \pi_n$ is perfect, then

$$H(\pi_1 \rhd \pi_2 \rhd \cdots \rhd \pi_n) \geq H(\kappa),$$

for any $\kappa \in \Pi^{(K_1 \cup \cdots \cup K_n)}(\Xi) = \bigcap_{i=1}^{n} \Pi^{(K_1 \cup \cdots \cup K_n)}(\pi_i)$.

Proof. To make the following computations more transparent, assume that $N = K_1 \cup \cdots \cup K_n$ and for each $i = 1, \cdots, n$, denote $S_i = K_i \cap (K_1 \cup \cdots \cup K_{i-1})$ (naturally $S_1 = \emptyset$). Using this, we can compute (the summation is performed only over those points x or y for which the respective probabilities are positive)

$$\begin{aligned}
H(\pi_1 \rhd \pi_2 \rhd \cdots \rhd \pi_n) &= -\sum_{x \in \mathbf{X}_N} (\pi_1 \rhd \cdots \rhd \pi_n)(x) \log (\pi_1 \rhd \cdots \rhd \pi_n)(x) \\
&= -\sum_{x \in \mathbf{X}_N} (\pi_1 \rhd \cdots \rhd \pi_n)(x) \log \prod_{i=1}^{n} \frac{\pi_i(x^{\downarrow K_i})}{\pi_i^{\downarrow S_i}(x^{\downarrow S_i})} \\
&= -\sum_{i=1}^{n} \sum_{x \in \mathbf{X}_N} (\pi_1 \rhd \cdots \rhd \pi_n)(x) \log \frac{\pi_i(x^{\downarrow K_i})}{\pi_i^{\downarrow S_i}(x^{\downarrow S_i})} \\
&= -\sum_{i=1}^{n} \sum_{y \in \mathbf{X}_{K_i}} (\pi_1 \rhd \cdots \rhd \pi_n)^{\downarrow K_i}(y) \log \frac{\pi_i(y)}{\pi_i^{\downarrow S_i}(y^{\downarrow S_i})} \\
&= -\sum_{i=1}^{n} \sum_{y \in \mathbf{X}_{K_i}} \pi_i(y) \log \left(\frac{\pi_i(y)}{\pi_i^{\downarrow S_i}(y^{\downarrow S_i})} \cdot \frac{\pi_i^{\downarrow K_i \setminus S_i}(y^{\downarrow K_i \setminus S_i})}{\pi_i^{\downarrow K_i \setminus S_i}(y^{\downarrow K_i \setminus S_i})} \right) \\
&= \sum_{i=1}^{n} \left(H(\pi_i^{\downarrow K_i \setminus S_i}) - \mathrm{MI}_{\pi_i}(X_{S_i}; X_{K_i \setminus S_i}) \right),
\end{aligned}$$

where the next to last modification is possible, because $\pi_1, \cdots, \pi_n$ is perfect, and therefore, π_i is a marginal distribution of $\pi_1 \rhd \cdots \rhd \pi_n$. The expression

$$\mathrm{MI}_{\pi_i}(X_{S_i}; X_{K_i \setminus S_i}) = \sum_{y \in \mathbf{X}_{K_i}} \pi_i(y) \log \frac{\pi_i(y)}{\pi_i^{\downarrow S_i}(y^{\downarrow S_i}) \cdot \pi_i^{\downarrow K_i \setminus S_i}(y^{\downarrow K_i \setminus S_i})}$$

is the well-known *mutual information* between groups of variables X_{S_i} and $X_{K_i \setminus S_i}$ for distribution π_i (in the previous section, we mentioned the fact that it is always non-negative (see, e.g. Gallager 1968).

Let us now compute the Shannon entropy of an arbitrary distribution $\kappa \in \Pi^{(K_1 \cup \cdots \cup K_n)}(\Xi)$. For this, we will utilize the fact that for $x \in \mathbf{X}_N$ such that $\kappa(x) > 0$ the following equality holds

$$\kappa(x) = \kappa^{\downarrow K_1}(x^{\downarrow K_1}) \cdot \frac{\kappa^{\downarrow K_1 \cup K_2}(x^{\downarrow K_1 \cup K_2})}{\kappa^{\downarrow K_1}(x^{\downarrow K_1})} \cdot \ \cdots \ \cdot \frac{\kappa^{\downarrow K_1 \cup \cdots \cup K_n}(x^{\downarrow K_1 \cup \cdots \cup K_n})}{\kappa^{\downarrow K_1 \cup \cdots \cup K_{n-1}}(x^{\downarrow K_1 \cup \cdots \cup K_{n-1}})}.$$

Therefore,

$$
\begin{aligned}
H(\kappa) &= -\sum_{x\in\mathbf{X}_N}\kappa(x)\log\kappa(x) = -\sum_{x\in\mathbf{X}_N}\kappa(x)\log\prod_{i=1}^{n}\frac{\kappa^{\downarrow K_1\cup\cdots\cup K_i}(x^{\downarrow K_1\cup\cdots\cup K_i})}{\kappa^{\downarrow K_1\cup\cdots\cup K_{i-1}}(x^{\downarrow K_1\cup\cdots\cup K_{i-1}})}\\
&= -\sum_{i=1}^{n}\sum_{x\in\mathbf{X}_N}\kappa(x)\log\left(\frac{\kappa^{\downarrow K_1\cup\cdots\cup K_i}(x^{\downarrow K_1\cup\cdots\cup K_i})}{\kappa^{\downarrow K_1\cup\cdots\cup K_{i-1}}(x^{\downarrow K_1\cup\cdots\cup K_{i-1}})}\cdot\frac{\kappa^{\downarrow K_i\setminus S_i}(x^{\downarrow K_i\setminus S_i})}{\kappa^{\downarrow K_i\setminus S_i}(x^{\downarrow K_i\setminus S_i})}\right)\\
&= \sum_{i=1}^{n}\Big(H(\kappa^{\downarrow K_i\setminus S_i}) - \mathrm{MI}_\kappa(X_{K_1\cup\cdots\cup K_{i-1}};X_{K_i\setminus S_i})\Big).
\end{aligned}
$$

Now, we apply the famous property of mutual information (Gallager 1968) saying that it is monotonous in the sense that

$$\mathrm{MI}_\kappa(X_{K_1\cup\cdots\cup K_{i-1}};X_{K_i\setminus S_i}) \geq \mathrm{MI}_\kappa(X_{S_i};X_{K_i\setminus S_i}),$$

and therefore, assuming that all π_is are marginal to κ, we get that $\mathrm{MI}_\kappa(X_{S_i};X_{K_i\setminus S_i}) = \mathrm{MI}_{\pi_i}(X_{S_i};X_{K_i\setminus S_i})$ and $H(\kappa^{\downarrow K_i\setminus S_i}) = H(\pi_i^{\downarrow K_i\setminus S_i})$, and therefore, also

$$H(\kappa) \leq \sum_{i=1}^{n}\Big(H(\pi_i^{\downarrow K_i\setminus S_i}) - \mathrm{MI}_{\pi_i}(X_{S_i};X_{K_i\setminus S_i})\Big) = H(\pi_1 \triangleright \pi_2 \triangleright \cdots \triangleright \pi_n). \qquad \square$$

10.5 *Uniqueness*

Now, we are ready to prove an important assertion claiming that if a system of low-dimensional distributions can form a perfect sequence, it defines (as a perfect sequence) a unique distribution.

THEOREM 10.15. If a sequence $\pi_1, \pi_2, \cdots, \pi_n$ and its permutation $\pi_{i_1}, \pi_{i_2}, \cdots, \pi_{i_n}$ are both perfect, then

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

Proof. Applying previous Theorem 10.14 to both of these sequences, we see that

$$H(\pi_1 \triangleright \pi_2 \triangleright \cdots \triangleright \pi_n) = H(\pi_{i_1} \triangleright \pi_{i_2} \triangleright \cdots \triangleright \pi_{i_n}) = \max_{\kappa\in\bigcap_{i=1}^{n}\Pi^{(K_1\cup\cdots\cup K_n)}(\pi_i)} H(\kappa).$$

Since the entropy is a continuous and strictly convex function on the convex and compact set $\bigcap_{i=1}^{n}\Pi^{(K_1\cup\cdots\cup K_n)}(\pi_i)$, it achieves its maximum at a single point and, therefore,

$$\pi_1 \triangleright \pi_2 \triangleright \cdots \triangleright \pi_n = \pi_{i_1} \triangleright \pi_{i_2} \triangleright \cdots \triangleright \pi_{i_n}. \qquad \square$$

Remark 10.16. Theorem 10.14 is only an implication; if there exists a perfect sequence formed by the distributions from $\Xi = \{\pi_1, \pi_2, \cdots, \pi_n\}$, then it achieves the maximum Shannon entropy among the distributions from $\Pi^{(K_1\cup\cdots\cup K_n)}(\Xi)$. However, as can be seen from the following example (which the author learned from Steffen Lauritzen), it does not mean that the maximum entropy distribution from $\Pi^{(K_1\cup\cdots\cup K_n)}(\Xi)$ must be a compositional model.

Table 11. Two-dimensional distributions.

$\pi_1(x_1, x_2)$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$\frac{2}{6}$	$\frac{1}{6}$
$x_2 = 1$	$\frac{1}{6}$	$\frac{2}{6}$
$\pi_2(x_1, x_3)$	$x_1 = 0$	$x_1 = 1$
$x_3 = 0$	$\frac{2}{6}$	$\frac{1}{6}$
$x_3 = 1$	$\frac{1}{6}$	$\frac{2}{6}$
$\pi_3(x_2, x_3)$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{1}{6}$	$\frac{2}{6}$
$x_3 = 1$	$\frac{2}{6}$	$\frac{1}{6}$

Example 10.17. It is not difficult to show that for two-dimensional distributions from Table 11 there exists the only common extension – the distribution from Table 12. (Hint: consider an arbitrary distribution having the given three marginals and show that none of its probabilities can be greater than $\frac{1}{6}$, then all the couples of probabilities which contribute to marginal probabilities equalling $\frac{2}{6}$ – those, which are positive in Table 12 – must equal $\frac{1}{6}$.) Therefore, this extension is also the maximum entropy extension.

Since all the considered two-dimensional distributions are positive, all possible compositional models constructed from them must also be positive, which means that the distribution from Table 12 cannot be obtained as a compositional model of distributions from Table 11.

11. Commutable sets

This section is the exception proving the rule; here, we will be interested in generating sequences whose distributions are connected by the operator of left composition. Specifically, we will be interested in the sequences defining a unique distribution regardless of their ordering. This is also the reason we deal with *sets* of distribution rather than with their sequences – the ordering of the distributions is irrelevant.

DEFINITION 11.1. A set of distributions $\{\pi_1, \pi_2, \cdots, \pi_n\}$ is said to be *commutable* if $\pi_1 \lhd \pi_2 \lhd \cdots \lhd \pi_n$ is defined and

$$\pi_1 \lhd \pi_2 \lhd \cdots \lhd \pi_n = \pi_{i_1} \lhd \pi_{i_2} \lhd \cdots \lhd \pi_{i_n}$$

for all permutations of indices $i_1, i_2, \cdots, i_n$.

Table 12. Extension of the distributions from Table 11.

	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$
$x_3 = 1$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$

Let us start discussing properties of commutable sets of oligodimensional distributions. First, two lemmata will be formulated that are almost direct consequences of the definition. The first one states that as it is true for perfect sequences, it also holds that all π_is are marginal to $\pi_1 \triangleleft \cdots \triangleleft \pi_n$ for commutable sets.

LEMMA 11.2. If $\{\pi_1, \pi_2, \cdots, \pi_n\}$ is a commutable set of probability distributions, then

$$\pi_1 \triangleleft \pi_2 \triangleleft \cdots \triangleleft \pi_n \in \bigcap_{i=1}^{n} \Pi^{(K_1 \cup \cdots \cup K_n)}(\pi_i),$$

which means that all of the distributions π_is are marginals of $\pi_1 \triangleleft \pi_2 \triangleleft \cdots \triangleleft \pi_n$.

Proof. The assertion is a direct consequence of the facts that for any $j \in \{1, 2, \cdots, n\}$ there are permutations $i_1, \cdots, i_n$ in which index $j = i_n$ is the last one, and thus, π_j is a marginal of the distribution $\pi_{i_1} \triangleleft \pi_{i_2} \triangleleft \cdots \triangleleft \pi_{i_n}$ (Lemma 3.2). □

Remark 11.3. The reader already knows (Remark 10.6) that the assertion expresses only a necessary condition; and the opposite assertion does not hold. If this were the case, all perfect sequences would form commutable sets, because for perfect sequences all distributions π_i are marginals of $\pi_1 \triangleleft \pi_2 \triangleleft \cdots \triangleleft \pi_n$.

Example 11.4. Let us present a non-trivial example of a generating sequence which is not perfect, and yet its distributions form a commutable set. Consider four distributions π_1, π_2, π_3, and π_4, for which $K_1 = \{1, 2, 4\}$, $K_2 = \{2, 3, 5\}$, $K_3 = \{1, 3, 6\}$, and $K_4 = \{1, 2, 3\}$. This situation is illustrated in Figure 3.

Let π_4 be the distribution from Table 13, and all three remaining distributions $\pi_1(x_1, x_2, x_4)$, $\pi_2(x_2, x_3, x_5)$, and $\pi_3(x_1, x_3, x_6)$ be uniform distributions of the respective sets of variables. The reader can immediately see that the distributions are pairwise consistent, because all their two-dimensional marginal distributions are uniform.

First, let us show that the considered distributions really form a commutable set. Since π_1, π_2, and π_3 are uniform, it is obvious that $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3$ is also the uniform six-dimensional distribution. The same holds for any permutation of these three distributions. Therefore,

$$\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4 = \pi_{i_1} \triangleleft \pi_{i_2} \triangleleft \pi_{i_3} \triangleleft \pi_{i_4}$$

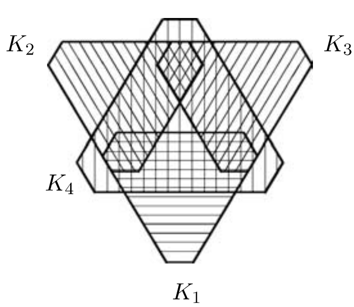


Figure 3. Star-like system of sets.

Table 13. Probability distribution π_4.

π_1	$x_1=0$		$x_1=1$	
	$x_2=0$	$x_2=1$	$x_2=0$	$x_2=1$
$x_3=0$	0	$\frac{1}{4}$	$\frac{1}{4}$	0
$x_3=1$	$\frac{1}{4}$	0	0	$\frac{1}{4}$

holds true for any permutation for which $i_4=4$. Now, applying Lemma 8.4 twice and Lemma 4.4 once, we get

$$\pi_{i_1} \triangleleft \pi_{i_2} \triangleleft \pi_{i_3} \triangleleft \pi_{i_4} = \pi_{i_1} \triangleleft \pi_{i_2} \triangleleft \pi_{i_4} \triangleleft \pi_{i_3} = \pi_{i_1} \triangleleft \pi_{i_4} \triangleleft \pi_{i_2} \triangleleft \pi_{i_3}$$
$$= \pi_{i_4} \triangleleft \pi_{i_1} \triangleleft \pi_{i_2} \triangleleft \pi_{i_3}.$$

Therefore, all permutations yield the same six-dimensional distribution.

Let us now show that the sequence $\pi_1, \pi_2, \pi_3, \pi_4$ is not perfect.

For uniform distributions π_1, π_2, and π_3, the distribution $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 = \pi_1 \triangleright \pi_2 \triangleright \pi_3$ is obviously also uniform. Since this distribution is from $\Pi^{(K_1 \cup K_2 \cup K_3)}$, and $K_4 \subset K_1 \cup K_2 \cup K_3$, we get that $\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4 = \pi_1 \triangleright \pi_2 \triangleright \pi_3$. Thus, $\pi_1, \pi_2, \pi_3, \pi_4$ is not perfect, because π_4 is not a marginal of $\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4$.

11.1 *Star-like systems*

Let us repeat once more that it may happen that all the distributions π_i are marginals of $\pi_1 \triangleleft \pi_2 \triangleleft \cdots \triangleleft \pi_n$ and still the sequence may neither be perfect (Example 11.4), nor the corresponding set commutable (it will be shown for sequence $\pi_1, \pi_2, \pi_3, \pi_5, \pi_4$ in Example 11.14). In what follows, we will present sufficient conditions describing special situations of perfect sequences and commutable sets, as well as examples illustrating the described theoretical properties.

Lemma 11.5. Let $\{\pi_1, \pi_2, \cdots, \pi_n\}$ be a commutable set of distributions. When $K_{i_1}, K_{i_2}, \cdots, K_{i_n}$ meets RIP then the sequence $\pi_{i_1}, \pi_{i_2}, \cdots, \pi_{i_n}$ is perfect.

Proof. Lemma 11.2 guarantees the pairwise consistency of the distributions from $\{\pi_1, \pi_2, \cdots, \pi_n\}$. Therefore, the sequence $\pi_{i_1}, \cdots, \pi_{i_n}$ is perfect due to Lemma 10.13. □

As we shall see in the following theorem, it is not surprising that the set of distributions from Example 11.4 is commutable. As a matter of fact, the system of the sets of variables (or their indices) has a special structural property by which the commutability is guaranteed.

Definition 11.6. A system of sets $\{K_1, K_2, \cdots, K_n\}$ is called *star-like* if there exists an index $\ell \in \{1, 2, \cdots, n\}$ such that for any couple of different indices $i, j \in \{1, 2, \cdots, n\}$ $K_i \cap K_j \subseteq K_\ell$. The set K_ℓ is called a *centre* of the system.

For examples of star-like systems of sets, see Figures 3–5. It is a trivial consequence of the definition that any star-like system of sets can be ordered to meet RIP. In fact, any ordering in which the centre of the system is at the first or second position meets RIP. Thus, if the

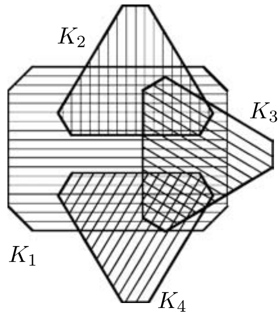


Figure 4. Strongly decomposable star-like system of sets.

system of sets $K_1, K_2, \cdots, K_n$ is star-like, then the distributions $\pi_1, \pi_2, \cdots, \pi_n$ can be reordered into a perfect sequence if they are pairwise consistent.

THEOREM 11.7. If, for a set of pairwise consistent distributions $\{\pi_1, \pi_2, \cdots, \pi_n\}$, the system $\{K_1, K_2, \cdots, K_n\}$ is star-like, then $\{\pi_1, \pi_2, \cdots, \pi_n\}$ is commutable.

Proof. We will show that any permutation $\pi_{i_1}, \pi_{i_2}, \cdots, \pi_{i_n}$ is either perfect, or may be transformed into a perfect sequence $\pi_{j_1}, \pi_{j_2}, \cdots, \pi_{j_n}$ in the way that

$$\pi_{i_1} \lhd \pi_{i_2} \lhd \cdots \lhd \pi_{i_n} = \pi_{j_1} \lhd \pi_{j_2} \lhd \cdots \lhd \pi_{j_n}.$$

Without loss of generality, assume that K_1 is the centre, and consider an arbitrary permutation $i_1, i_2, \cdots i_n$. Let $1 = i_k$. Apparently, if $k \leq 2$, the sequence $K_{i_1}, K_{i_2}, \cdots, K_{i_n}$ meets RIP. If $k > 2$, we can apply Lemma 8.4 ($(k-2)$ times), getting

$$\begin{aligned}(\pi_{i_1} \lhd \cdots \lhd \pi_{i_{k-2}}) \lhd \pi_{i_{k-1}} \lhd \pi_{i_k} &= (\pi_{i_1} \lhd \cdots \lhd \pi_{i_{k-2}}) \lhd \pi_{i_k} \lhd \pi_{i_{k-1}} \\ &= \cdots = \pi_{i_1} \lhd\lhd \pi_{i_k} \lhd \pi_{i_2} \lhd \cdots \lhd \pi_{i_{k-1}}.\end{aligned}$$

Thus, we see that $\pi_{i_1} \lhd \cdots \lhd \pi_{i_n}$ equals $\pi_{j_1} \lhd \cdots \lhd \pi_{j_n}$, where the sequence $K_{j_1}, \cdots, K_{j_n}$ meets RIP. Therefore, due to Lemma 10.13 and Theorem 10.15, for all permutations $i_1, i_2, \cdots i_n$, the expressions $\pi_{i_1} \lhd \cdots \lhd \pi_{i_n}$ define the same multidimensional distribution, which means that $\{\pi_1, \pi_2, \cdots, \pi_n\}$ is commutable. □

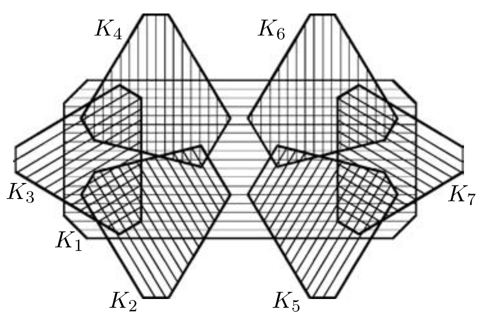


Figure 5. Star-like system of seven sets.

The previous theorem presents a *structural* condition under which a set of pairwise consistent distributions is commutable. The same property holds also for other special systems of sets of variables.

11.2 *Strongly decomposable systems*

Definition 11.8. A system of sets $\{K_1, K_2, \cdots, K_n\}$ is called *strongly decomposable* if each of its subsets can be ordered to meet RIP.

Example 11.9. First, notice that a star-like system in Figure 3 is not strongly decomposable. As shown in Example 11.4, the system can be ordered to meet RIP, but there is a subsystem (the reader can easily show that in this case it is the only one) – K_1, K_2, K_3 – which cannot be ordered to meet RIP.

Nevertheless, it does not mean that star-like systems are not strongly decomposable. For example, the system of four sets in Figure 4 is star-like and simultaneously strongly decomposable. In the next lemma, we shall prove that all strongly decomposable systems can be ordered in the way that both the whole sequence $K_1, K_2, K_3, \cdots, K_n$ and its 'suffix' $K_2, K_3, \cdots, K_n$ meet RIP. Let us first show that, although the assertion seems to be rather simple, it is not so obvious.

Consider the system K_1, K_2, K_3, K_4 from Figure 4. The sequence K_1, K_2, K_4, K_3 meets RIP, while sequence K_2, K_4, K_3 does not. On the other hand, sequence K_2, K_1, K_3 meets RIP and sequence K_4, K_2, K_1, K_3 does not, although there exists a sequence meeting RIP which starts with K_4 (it is a sequence K_4, K_1, K_2, K_3).

Lemma 11.10. For any strongly decomposable system of sets $K_1, K_2, \cdots, K_n$ and any $\ell \in \{1, \cdots, n\}$, there exists a permutation of indices $i_1, i_2, \cdots, i_n$ such that $i_1 = \ell$ and both the sequences $K_{i_1}, K_{i_2}, K_{i_3}, \cdots, K_{i_n}$ and $K_{i_2}, K_{i_3}, \cdots, K_{i_n}$ meet RIP.

Proof. Without loss of generality, assume $\ell = 1$. Let us start constructing the required permutation from an arbitrary ordering of all sets starting with K_1 and meeting RIP. Such an ordering is guaranteed by Lemma 10.12. Let it be $K_1, K_2, \cdots, K_n$.

Now, starting with K_n we will group the sets $K_2, \cdots, K_n$ into one or several clusters. At the beginning, each of these sets forms one cluster. According to RIP, there exists $j_n < n$ such that

$$K_n \cap (K_1 \cup \cdots \cup K_{n-1}) \subseteq K_{j_n}.$$

If there are more such j_n take the largest of them and, in case $j_n > 1$, put K_n and K_{j_n} into one cluster. If $j_n = 1$, do nothing. Then consider $\ell = n-1,\ n-2, \cdots, 2$ and at each step find (the largest) j_ℓ for which the RIP condition holds true, and, if the respective $j_\ell > 1$, connect the clusters holding K_ℓ and K_{j_ℓ} into one cluster.

After this process, we get a resulting partition of sets $K_2, \cdots, K_n$ into several, let us say m, clusters. The set with the lowest index in a cluster will be, in this proof, called the *cluster representative*. Having m clusters, we have a system of m representatives which can be ordered to meet RIP, because we assume that $K_1, K_2, \cdots, K_n$ is strongly decomposable. Denote the RIP ordering of these cluster representatives $K_{j_1}, K_{j_2}, \cdots, K_{j_m}$. Similarly, we will also order the sets in each cluster to meet RIP; each of these orderings must start with the cluster representative. Let such an ordering of sets from the kth cluster be $K_{j_k}, K_{j_{k,2}}, K_{j_{k,3}}, \cdots, K_{j_{k,r(k)}}$.

Then the required permutation $K_{i_1}, K_{i_2}, \cdots, K_{i_n}$ is the following:

$$\begin{aligned}
&K_1,\\
&K_{j_1}, K_{j_{1,2}}, K_{j_{1,3}}, \cdots, K_{j_{1,r(1)}},\\
&K_{j_2}, K_{j_{2,2}}, K_{j_{2,3}}, \cdots, K_{j_{2,r(2)}},\\
&K_{j_3}, K_{j_{3,2}}, K_{j_{3,3}}, \cdots, K_{j_{3,r(3)}},\\
&\vdots\\
&K_{j_m}, K_{j_{m,2}}, K_{j_{m,3}}, \cdots, K_{j_{m,r(m)}}.
\end{aligned}$$

What can be said about this construction? Apparently, the system $K_1, K_{j_1}, K_{j_2}, \cdots, K_{j_m}$ is a star-like system (its centre is K_1). Another property, which can be seen from the way clusters were constructed, is that an intersection of any two sets from different clusters is contained in an intersection of the respective cluster representatives (and therefore, also in K_1). These two properties are sufficient to show that the two required sequences meet RIP.

First, consider the shorter sequence $K_{i_2}, K_{i_3}, \cdots, K_{i_n}$ and any $k \in \{2, 3, \cdots, n\}$. If K_{i_k} is one of the cluster representatives, then the existence of $j < k$ required by RIP condition is guaranteed by the fact that the representatives have been ordered to meet RIP (other 'non-representative' sets cannot interfere with this fact, because intersections of sets from different clusters are contained in the intersections of the respective cluster representatives). If K_{i_k} is a 'non-representative' set, then the existence of the needed $j < k$ follows from the fact that each cluster has been ordered to meet RIP (again, sets from other clusters cannot interfere with it, because intersections of sets from different clusters are contained in the intersection of the respective cluster representatives).

Considering the longer sequence $K_{i_1}, K_{i_2}, K_{i_3}, \cdots, K_{i_n}$ can change the way of seeking for the indices i_j required by RIP condition only when the cluster representatives are considered. However, since $K_1, K_{j_1}, K_{j_2}, \cdots, K_{j_m}$ is a star-like system, starting with K_1, which is a centre of the system, cannot spoil the validity of the RIP condition, because *any* ordering of a star-like system starting with the centre meets RIP. □

Remark 11.11. At the beginning of Example 11.9, we mentioned that a star-like system in Figure 3 is not strongly decomposable. The same also holds for the larger system in Figure 5. It is easy to show that all star-like systems $K_1, \cdots, K_n$ can be ordered in such a way that the full ordering and its 'suffix' ordering of length $n-1$ meet RIP. Nevertheless, Lemma 11.10 guarantees that for strongly decomposable systems there are many such orderings; for each $i = 1, \cdots, n$ there exists at least one such ordering starting with K_i. For star-like systems, this property holds for sequences at which the centre is at the second position. For systems from Figures 3 and 5, no sequence starting with the centres K_4 and K_1, respectively, meets this condition. Therefore, the proof of the following Theorem 11.12 cannot be applied to Theorem 11.7.

THEOREM 11.12. If, for a set of pairwise consistent distributions $\{\pi_1, \pi_2, \cdots, \pi_n\}$, the system $\{K_1, K_2, \cdots, K_n\}$ is strongly decomposable, then the considered set of distributions is commutable.

Proof. Assuming that $\{K_1, K_2, \cdots, K_n\}$ is strongly decomposable, we will show that for any permutation of indices $i_1, i_2, \cdots, i_n$

$$\pi_{i_1} \lhd \pi_{i_2} \lhd \cdots \lhd \pi_{i_n} = \pi_{j_1} \rhd \pi_{j_2} \rhd \cdots \rhd \pi_{j_n},$$

for some permutation $j_1, j_2, \cdots, j_n$, for which $K_{j_1}, K_{j_2}, \cdots, K_{j_n}$ meets RIP. Then the commutability of $\{\pi_1, \pi_2, \cdots, \pi_n\}$ will be a direct consequence of Lemma 10.13 and Theorem 10.15.

Consider a permutation $i_1, i_2, \cdots, i_n$. If $n = 2$, then, due to Lemma 4.4, $\pi_1 \lhd \pi_2 = \pi_1 \rhd \pi_2$, and the required condition holds true. Therefore, applying mathematical induction, it is enough to show the required property under the assumption that it holds for $n - 1$.

Let us consider permutations of all indices starting with i_n. Due to Lemma 11.10, there exists a permutation $i_n, j_1, \cdots, j_{n-1}$ among them such that $K_{i_n}, K_{j_1}, \cdots, K_{j_{n-1}}$ and $K_{j_1}, \cdots, K_{j_{n-1}}$ meet RIP. We will show that

$$\pi_{i_1} \lhd \pi_{i_2} \lhd \cdots \lhd \pi_{i_n} = \pi_{i_n} \rhd \pi_{i_1} \rhd \pi_{i_2} \rhd \cdots \rhd \pi_{i_{n-1}}.$$

Applying the assumption of mathematical induction, we get

$$\begin{aligned}\pi_{i_1} \lhd \pi_{i_2} \lhd \cdots \lhd \pi_{i_n} &= \pi_{i_n} \rhd (\pi_{i_1} \lhd \pi_{i_2} \lhd \cdots \lhd \pi_{i_{n-1}}) \\ &= \pi_{i_n} \rhd (\pi_{j_1} \rhd \pi_{j_2} \rhd \cdots \rhd \pi_{j_{n-1}}) \\ &= \pi_{i_n} \rhd ((\pi_{j_1} \rhd \cdots \rhd \pi_{j_{n-2}}) \rhd \pi_{j_{n-1}}).\end{aligned}$$

Since $i_n, j_1, \cdots, j_{n-1}$ meets RIP,

$$K_{j_{n-1}} \cap (K_{i_n} \cup K_{j_1} \cup \cdots \cup K_{j_{n-2}})$$

must be a subset of either K_{i_n} or $(K_{j_1} \cup \cdots \cup K_{j_{n-2}})$, or, in other words, either

$$K_{i_n} \supseteq ((K_{j_1} \cup \cdots \cup K_{j_{n-2}}) \cap K_{j_{n-1}})$$

or

$$(K_{j_1} \cup \cdots \cup K_{j_{n-2}}) \supseteq (K_{i_n} \cap K_{j_{n-1}})$$

holds true. Therefore, applying either Theorem 7.2 or 7.3, respectively, we get

$$\begin{aligned}\pi_{i_n} \rhd ((\pi_{j_1} \rhd \cdots \rhd \pi_{j_{n-2}}) \rhd \pi_{j_{n-1}}) &= \pi_{i_n} \rhd (\pi_{j_1} \rhd \cdots \rhd \pi_{j_{n-2}}) \rhd \pi_{j_{n-1}} \\ &= \pi_{i_n} \rhd ((\pi_{j_1} \rhd \cdots \rhd \pi_{j_{n-3}}) \rhd \pi_{j_{n-2}}) \rhd \pi_{j_{n-1}}.\end{aligned}$$

However, regarding that $i_n, j_1, \cdots, j_{n-2}$ meets RIP, too, we can repeat the previous step, getting

$$\pi_{i_n} \rhd (\pi_{j_1} \rhd \pi_{j_2} \rhd \cdots \rhd \pi_{j_{n-1}}) = \pi_{i_n} \rhd (\pi_{j_1} \rhd \cdots \rhd \pi_{j_{n-3}}) \rhd \pi_{j_{n-2}} \rhd \pi_{j_{n-1}}.$$

In this way, we can eliminate all parentheses, getting eventually that

$$\pi_{i_n} \rhd (\pi_{j_1} \rhd \pi_{j_2} \rhd \cdots \rhd \pi_{j_{n-1}}) = \pi_{i_n} \rhd \pi_{j_1} \rhd \pi_{j_2} \rhd \cdots \rhd \pi_{j_{n-1}},$$

which finishes the proof. □

Remark 11.13. The reader familiar with the iterative proportional fitting procedure (IPFP) (Deming and Stephan 1940) has certainly noticed a close relation of the studied commutable sets with this famous computational procedure. In principle, this iterative procedure is nothing else, but an infinite application of the operator of left composition

$$\pi_1 \lhd \pi_2 \lhd \; \cdots \; \lhd \pi_n \lhd \pi_1 \lhd \; \cdots \; \lhd \pi_n \lhd \pi_1 \lhd \cdots$$

Therefore, the commutable sets are exactly those for which IPFP converges after the first cycle (n steps) regardless of the ordering of the distributions. This problem was studied by Habermann (1974).

11.3 *Examples*

Example 11.14. From this example, which is by Vomlel (1999), it can be seen that the assumption in the previous theorem cannot be weakened in the sense that instead of strong decomposability of a system $\{K_1, K_2, \cdots, K_n\}$, one would assume just an existence of its ordering meeting RIP. In fact, Vomlel used this example to refute Habermann's conjecture that if IPFP is applied to a set of distributions, which can be ordered in the way that $K_1, K_2, \cdots, K_n$ meets RIP, then the procedure converges in a finite number of steps regardless of the ordering of the distributions.

Vomlel's example considers five distributions $\pi_1, \cdots, \pi_5$ with a structure of variables depicted in Figure 6. For the purpose of this example, we will consider: $K_1 = \{1, 2, 4\}$, $K_2 = \{2, 3\}$, $K_3 = \{1, 3\}$, $K_4 = \{4, 5\}$, and $K_5 = \{1, 2, 3\}$.

It is easy to show that although the sequence K_1, K_2, K_3, K_4, K_5 does not meet RIP, it can be reordered so that the RIP holds.

We will not repeat Vomlel's computations here, but for the distributions in Tables 14 and 15, the reader can show that $(\pi_1 \lhd \pi_2 \lhd \pi_3 \lhd \pi_4 \lhd \pi_5)^{\downarrow\{4\}} \neq \pi_1^{\downarrow\{4\}}$, which means that this set of distributions is not commutable.

Example 11.15. Let us now highlight a substantial difference between perfect sequences and commutable sets of probability distributions. For any perfect sequence $\pi_1, \cdots, \pi_n$, its initial subsequence $\pi_1, \cdots, \pi_k$ is again perfect. In contrast with this, from the following example, we will see that there are commutable sets whose subsets are not commutable.

Consider again four distributions π_1, π_2, π_3, and π_4 with $K_1 = \{1, 2, 3\}$, $K_2 = \{1, 2, 4\}$, $K_3 = \{2, 3, 5\}$, and $K_4 = \{1, 3, 6\}$. Values of these distributions are given in Table 16.

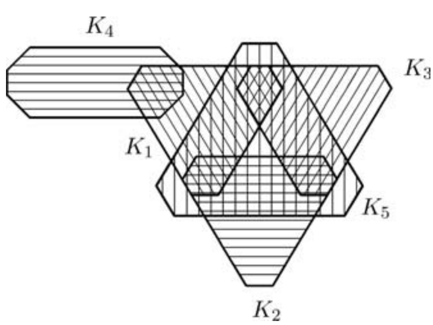


Figure 6. System of sets that can be ordered to meet RIP.

Table 14. Two-dimensional distributions.

$\pi_2(x_2, x_3)$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{1}{3}$	$\frac{1}{6}$
$x_3 = 1$	$\frac{1}{6}$	$\frac{1}{3}$
$\pi_3(x_1, x_3)$	$x_1 = 0$	$x_1 = 1$
$x_3 = 0$	$\frac{1}{6}$	$\frac{1}{3}$
$x_3 = 1$	$\frac{1}{3}$	$\frac{1}{6}$
$\pi_4(x_4, x_5)$	$x_4 = 0$	$x_4 = 1$
$x_5 = 0$	$\frac{1}{2}$	0
$x_5 = 1$	0	$\frac{1}{2}$

To show that $\{\pi_1, \pi_2, \pi_3, \pi_4\}$ is commutable, it is enough, due to Theorem 11.7, to show that these distributions are pairwise consistent, which follows immediately from the consistency of π_1 with π_2, π_3, and π_4 (see Table 17).

Now, we shall prove by contradiction that $\{\pi_2, \pi_3, \pi_4\}$ is not commutable. Assume that $\{\pi_2, \pi_3, \pi_4\}$ forms a commutable set. Then, due to Lemma 11.2, all π_2, π_3, and π_4 are marginal to $\kappa = \pi_2 \lhd \pi_3 \lhd \pi_4$. Therefore, under this assumption,

$$\kappa^{\downarrow\{2,3\}}(x_2 = 0, x_3 = 0) = \pi_3^{\downarrow\{2,3\}}(x_2 = 0, x_3 = 0) = \tfrac{1}{3},$$
$$\kappa^{\downarrow\{1,2,3\}}(x_1 = 0, x_2 = 0, x_3 = 0) \leq \pi_4^{\downarrow\{1,3\}}(x_1 = 0, x_3 = 0) = \tfrac{1}{6},$$
$$\kappa^{\downarrow\{1,2,3\}}(x_1 = 1, x_2 = 0, x_3 = 0) \leq \pi_2^{\downarrow\{1,2\}}(x_1 = 1, x_2 = 0) = \tfrac{1}{6},$$

and therefore, we are sure that

$$\kappa^{\downarrow\{1,2,3\}}(x_1 = 0, x_2 = 0, x_3 = 0) = \kappa^{\downarrow\{1,2,3\}}(x_1 = 1, x_2 = 0, x_3 = 0) = \frac{1}{6}.$$

Table 15. Three-dimensional distributions.

	$x_1 = 0$		$x_1 = 1$	
π_1	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_4 = 0$	$\frac{1}{9}$	0	$\frac{1}{6}$	$\frac{1}{9}$
$x_4 = 1$	$\frac{2}{9}$	$\frac{1}{6}$	0	$\frac{2}{9}$
	$x_1 = 0$		ν	
π_5	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$
$x_3 = 1$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$

Table 16. Probability distributions π_1, π_2, π_3, and π_4.

	$x_1 = 0$		$x_1 = 1$	
π_1	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$
$x_3 = 1$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$
	$x_4 = 0$		$x_4 = 1$	
π_2	$x_1 = 0$	$x_1 = 1$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
$x_2 = 1$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$
	$x_5 = 0$		$x_5 = 1$	
π_3	$x_3 = 0$	$x_3 = 1$	$x_3 = 0$	$x_3 = 1$
$x_2 = 0$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
$x_2 = 1$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$
	$x_6 = 0$		$x_6 = 1$	
π_4	$x_3 = 0$	$x_3 = 1$	$x_3 = 0$	$x_3 = 1$
$x_1 = 0$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$
$x_1 = 1$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$

Thus, because

$$\kappa^{\downarrow\{1,3\}}(x_1 = 0, x_3 = 0) = \pi_4^{\downarrow\{1,3\}}(x_1 = 0, x_3 = 0) = \frac{1}{6},$$

we get

$$\kappa(x_1 = 0, x_2 = 1, x_3 = 0) = 0,$$

which contradicts with the obvious fact that $\pi_2 \triangleleft \pi_3 \triangleleft \pi_4$ is strictly positive (namely, all π_2, π_3, and π_4 are positive). Therefore, $\{\pi_2, \pi_3, \pi_4\}$ cannot be commutable.

Table 17. Two-dimensional marginal distributions of π_1.

$\pi_1^{\downarrow\{1,2\}}(x_1, x_2)$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$\frac{2}{6}$	$\frac{1}{6}$
$x_2 = 1$	$\frac{1}{6}$	$\frac{2}{6}$
$\pi_1^{\downarrow\{1,3\}}(x_1, x_3)$	$x_1 = 0$	$x_1 = 1$
$x_3 = 0$	$\frac{1}{6}$	$\frac{2}{6}$
$x_3 = 1$	$\frac{2}{6}$	$\frac{1}{6}$
$\pi_1^{\downarrow\{2,3\}}(x_2, x_3)$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{2}{6}$	$\frac{1}{6}$
$x_3 = 1$	$\frac{1}{6}$	$\frac{2}{6}$

Example 11.16. Even if a subset of a commutable set is also commutable, it does not mean that this commutable subset defines a distribution marginal to the distribution defined by the larger commutable set. More exactly, if both $\{\pi_1, \cdots, \pi_n\}$ and $\{\pi_1, \cdots, \pi_k\}$ (for some $1 \le k < n$) are commutable, then it may happen that

$$(\pi_1 \triangleleft \cdots \triangleleft \pi_n)^{(K_1 \cup \cdots \cup K_k)} \neq \pi_1 \triangleleft \cdots \triangleleft \pi_k.$$

An example of this situation is given by sets $\{\pi_1, \pi_2, \pi_3, \pi_4\}$ and $\{\pi_1, \pi_2, \pi_3\}$ from Example 11.4. Both of these sets are commutable and both define six-dimensional distributions of variables $X_1, X_2, \cdots, X_6$. While the distribution $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3$ is uniform, the distribution

$$\kappa = \pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4$$

is the one for which

$$\kappa(x_1, x_2, \cdots, x_6) = \begin{cases} \frac{1}{32} & \text{if } x_1 + x_2 + x_3 \text{ is odd,} \\ 0 & \text{otherwise.} \end{cases}$$

Example 11.17. There may still be one more question regarding commutable sets of distributions. All the examples of commutable sets presented up to now had one common property: the distributions could be ordered in the way that the sequence was perfect. The natural question arises whether there always exists an ordering of a commutable set that forms a perfect sequence. As it will be shown in the following example, the answer to this question is negative.

Consider three six-dimensional distributions of binary variables (with values $\{0, 1\}$), each of which will be an independent product of three two-dimensional distributions (to stress a symmetry in the model, we use double-digit indices):

$$\pi_1(x_{11}, x_{12}, x_{22}, x_{23}, x_{33}, x_{34}) = \mu(x_{11}, x_{12})\kappa(x_{22}, x_{23})\mu(x_{33}, x_{34}),$$
$$\pi_2(x_{12}, x_{13}, x_{23}, x_{24}, x_{31}, x_{32}) = \kappa(x_{12}, x_{13})\mu(x_{23}, x_{24})\mu(x_{31}, x_{32}),$$
$$\pi_3(x_{13}, x_{14}, x_{21}, x_{22}, x_{32}, x_{33}) = \mu(x_{13}, x_{14})\mu(x_{21}, x_{22})\kappa(x_{32}, x_{33}),$$

where distribution μ is a uniform distribution and

$$\kappa(y, z) = \begin{cases} \frac{1}{2} & \text{iff } y + z = 1, \\ 0 & \text{otherwise.} \end{cases}$$

It is not difficult to show that distributions π_1, π_2, and π_3 are pairwise consistent. Distributions π_1 and π_3 share only variables x_{12} and x_{23}, and both the respective two-dimensional marginal distributions are uniform for this pair of variables. Similarly, distributions π_2 and π_3 have common arguments x_{13} and x_{32}, while π_1 and π_3 share x_{22} and x_{33}. Again, all the respective two-dimensional marginal distributions are uniform.

Notice that in this case

$$\pi_1 \triangleleft \pi_2 = (\mu(x_{11}, x_{12}) \triangleleft \kappa(x_{12}, x_{13}))(\kappa(x_{22}, x_{23}) \triangleleft \mu(x_{23}, x_{24}))(\mu(x_{33}, x_{34}) \triangleleft \mu(x_{31}, x_{32})),$$

and analogously

$$\pi_1 \lhd \pi_2 \lhd \pi_3 = (\mu(x_{11}, x_{12}) \lhd \kappa(x_{12}, x_{13}) \lhd \mu(x_{13}, x_{14}))$$
$$(\kappa(x_{22}, x_{23}) \lhd \mu(x_{23}, x_{24}) \lhd \mu(x_{21}, x_{22}))$$
$$(\mu(x_{33}, x_{34}) \lhd \mu(x_{31}, x_{32}) \lhd \kappa(x_{32}, x_{33})).$$

Thus, $\pi_1 \lhd \pi_2 \lhd \pi_3$ is a product of three terms, each of which is a composition of three two-dimensional distributions. Moreover, each of these terms is a composition of distributions whose variables form a star-like system, and therefore, due to Theorem 11.7, one can see that $\{\pi_1, \pi_2, \pi_3\}$ is commutable.

Now, we will show that $\pi_{i_1}, \pi_{i_2}, \pi_{i_3}$ is not perfect for any permutation i_1, i_2, i_3. However, since the situation is symmetric, in a sense (each distribution π_i is a product of a uniform distribution with κ), we will show it just for π_1, π_2, π_3.

Similarly to application of the operator of left composition, using operator $\rhd$ leads to a product of three terms

$$\pi_1 \rhd \pi_2 \rhd \pi_3 = (\mu(x_{11}, x_{12}) \rhd \kappa(x_{12}, x_{13}) \rhd \mu(x_{13}, x_{14}))$$
$$(\kappa(x_{22}, x_{23}) \rhd \mu(x_{23}, x_{24}) \rhd \mu(x_{21}, x_{22}))$$
$$(\mu(x_{33}, x_{34}) \rhd \mu(x_{31}, x_{32}) \rhd \kappa(x_{32}, x_{33})).$$

From this, we see that $(\pi_1 \rhd \pi_2 \rhd \pi_3)^{\downarrow\{32,33\}}$ is a uniform distribution, which is not true for π_3. This means that π_3 cannot be a marginal distribution of $\pi_1 \rhd \pi_2 \rhd \pi_3$, and therefore, π_1, π_2, π_3 is not perfect.

12. Flexible sequences

Consider a (perfect) sequence $\pi_1(x_1, x_2), \pi_2(x_2, x_3), \pi_3(x_3, x_4)$ representing distribution $\kappa(x_1, x_2, x_3, x_4) = \pi_1(x_1, x_2) \rhd \pi_2(x_2, x_3) \rhd \pi_3(x_3, x_4)$. Having a need to compute a conditional probability $\kappa(x_4|x_1 = a)$, one can do it simply by marginalizing[3]

$$\kappa(x_2, x_3, x_4|x_1 = a) = \pi_1(x_2|x_1 = a) \rhd \pi_2(x_2, x_3) \rhd \pi_3(x_3, x_4).$$

However, when the desire is to compute conditional distribution $\kappa(x_1|x_4 = b)$, the situation becomes much more complicated. In this case, one cannot take advantage of the fact that κ is represented in the form of a compositional model; instead, one has to compute the full four-dimensional distribution and only afterwards compute the required conditional probability. In fact, this is the very problem of asymmetry which is for Bayesian networks usually solved by application of the famous computational procedure of *local computations* (Lauritzen and Spiegelhalter 1988) based on a transformation of a Bayesian network into a decomposable model. And this is also the reason why we are going to study flexible sequences, which may play for compositional models a role similar to the one played by decomposable models for Bayesian networks.

At this point, we present only a couple of basic properties and examples illuminating the relationship between flexible and perfect sequences. The first one shows that if a generating sequence meets the condition of Lemma 10.13, this sequence is not only perfect but also flexible. But first, we have to introduce the definition.

Definition 12.1. A generating sequence $\pi_1, \pi_2, \cdots, \pi_n$ is called *flexible* if for all $j \in K_1 \cup \cdots \cup K_n$ there exists a permutation $i_1, i_2, \cdots, i_n$ such that $j \in K_{i_1}$ and

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

As we said above, flexible sequences are those, which can be reordered in many ways so that each variable can appear among the arguments of the first distribution. It does not mean, however, that each distribution appears at the beginning of the generating sequence. If this were the case, flexible sequences would just form a subclass of perfect sequences (since each distribution would be a marginal of the composed distribution – see Lemma 12.7).

Example 12.2. Obviously, any triplet of distributions $\pi_1(x_1, x_2)$, $\pi_2(x_1, x_3)$, and $\pi_3(x_2, x_3)$ for which π_1 and π_2 are consistent is flexible, since in this case

$$\pi_1(x_1, x_2) \triangleright \pi_2(x_1, x_3) \triangleright \pi_3(x_2, x_3) = \pi_2(x_1, x_3) \triangleright \pi_1(x_1, x_2) \triangleright \pi_3(x_2, x_3).$$

Let us stress that sequence π_1, π_2, π_3, as well as sequence π_2, π_1, π_3, is flexible regardless of the values of distribution π_3. Therefore, if

$$\pi_3(x_2, x_3) = (\pi_1(x_1, x_2) \triangleright \pi_2(x_1, x_3))^{\downarrow\{2,3\}},$$

then both π_1, π_2, π_3 and π_2, π_1, π_3 are also perfect, which is not true in the opposite case because of Theorem 10.3. Thus, we see that not all flexible sequences are perfect.

It is also easy to show that there exist perfect sequences which are not flexible. For example, the reader can show that the sequence $\kappa_1, \kappa_2, \kappa_3$ of distributions from Table 18 is perfect. The fact that it is not flexible is a consequence of $\kappa_3 \triangleright \kappa_1 \neq \kappa_1 \triangleright \kappa_2 \triangleright \kappa_3 \neq \kappa_3 \triangleright \kappa_2$. Obviously $(\kappa_3 \triangleright \kappa_1)(x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1) > 0$ and $(\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3)(x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1) = 0$ and similarly $(\kappa_3 \triangleright \kappa_2)(x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 1) > 0$ and $(\kappa_1 \triangleright \kappa_2 \triangleright \kappa_3)(x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 1) = 0$.

Lemma 12.3. If $\pi_1, \pi_2, \cdots, \pi_n$ is a sequence of pairwise consistent probability distributions such that $K_1, \cdots, K_n$ meets RIP, then this sequence is flexible.

Table 18. Distributions κ_1, κ_2, and κ_3.

$\kappa_1(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	$\frac{1}{4}$	$\frac{1}{4}$
$x_1 = 1$	$\frac{1}{2}$	0

$\kappa_2(x_1, x_3)$	$x_3 = 0$	$x_3 = 1$
$x_1 = 0$	$\frac{1}{4}$	$\frac{1}{4}$
$x_1 = 1$	$\frac{1}{2}$	0

	$x_2 = 0$		$x_2 = 1$	
$\kappa_3(x_2, x_3, x_4)$	$x_3 = 0$	$x_3 = 1$	$x_3 = 0$	$x_3 = 1$
$x_4 = 0$	$\frac{5}{8}$	0	0	$\frac{1}{8}$
$x_4 = 1$	0	$\frac{1}{8}$	$\frac{1}{8}$	0

Proof. The assertion follows from the well-known fact that for all $\ell \in \{1, 2, \cdots, n\}$, one can find a permutation of $K_1, \cdots, K_n$ meeting RIP and starting with K_ℓ. Then it is enough to realize that, due to Lemma 10.13, all the RIP permutations yield perfect sequences, which define the same multidimensional distribution according to Theorem 10.15. □

Remark 12.4. Notice that there exist non-trivial flexible sequences $\pi_1(x_{K_1}), \pi_2(x_{K_2}), \cdots, \pi_n(x_{K_n})$ for which no permutation $K_{i_1}, K_{i_2}, \cdots, K_{i_n}$ meets RIP – see, e.g. Example 12.6. (Realize that any sequence of uniform distributions is flexible.)

Remark 12.5. It should be stressed that, when speaking about a flexible sequence, the ordering of distributions is substantial in spite of the fact that it allows a number of different reorderings not changing the resulting multidimensional distribution. Notice, however, that one can create different flexible sequences from a system of distributions – see the following example.

Example 12.6. Consider three pairwise consistent distributions $\{\pi_1(x_1, x_2), \pi_2(x_2, x_3), \pi_3(x_3, x_4)\}$, and assume that $X_2 \not\perp\!\!\!\perp X_3[\pi_2]$. Obviously, three sets $\{1, 2\}$, $\{2, 3\}$, and $\{3, 4\}$ can be ordered in $3! = 6$ ways, four of which meet RIP

$$\{1, 2\}\ \{2, 3\}\ \{3, 4\},\ \ \{2, 3\}\ \{1, 2\}\ \{3, 4\},$$
$$\{2, 3\}\ \{3, 4\}\ \{1, 2\},\ \ \{3, 4\}\ \{2, 3\}\ \{1, 2\},$$

All the corresponding sequences, which are perfect due to Lemma 10.13, define the same distribution, and therefore, all of them are flexible. Nevertheless, the generating sequences corresponding to the two remaining permutations

$$\{1, 2\}\ \{3, 4\}\ \{2, 3\},\ \ \{3, 4\}\ \{1, 2\}\ \{2, 3\},$$

are also flexible (though not perfect! – to verify it show that π_2 is not a marginal of the resulting four-dimensional distribution), because both of them define the same distribution

$$\pi_1 \rhd \pi_3 \rhd \pi_2 = \pi_3 \rhd \pi_1 \rhd \pi_2 = \pi_1 \cdot \pi_3,$$

and each variable appears among the arguments of the first distribution in one of the sequences. Thus we have shown that from the considered three distributions one can set up two different four-dimensional distributions, each of which is defined by a flexible sequence. Additionally, let us remark that the considered set $\{\pi_1, \pi_2, \pi_3\}$ is also commutable defining the same distribution as the flexible perfect sequence π_1, π_2, π_3. On the other hand, flexible sequence π_1, π_3, π_2 has a special property (which may seem rather strange at first sight): its distributions form a commutable set $\{\pi_1, \pi_3, \pi_2\}$, but this set defines a distribution which differs from the distribution defined by the flexible sequence π_1, π_3, π_2,

$$\pi_1 \lhd \pi_3 \lhd \pi_2 \neq \pi_1 \rhd \pi_3 \rhd \pi_2.$$

The next assertion introduces a simple sufficient condition under which a flexible sequence is also perfect.

LEMMA 12.7. If, for all $i = 1, \cdots, n$ of a flexible sequence $\pi_1, \cdots, \pi_n$, there exists an index

$$j \in K_i \backslash (K_1 \cup \cdots \cup K_{i-1} \cup K_{i+1} \cup \cdots \cup K_n),$$

then this sequence is perfect, too.

Proof. In other words, the assumption says that each set K_i contains at least one index which is not included in any other set K_j. Therefore, the assumption of flexibility in this case requires that for each π_i there must exist a permutation of indices such that $\pi_i = \pi_{i_1}$ and $\pi_{i_1} \triangleright \cdots \triangleright \pi_{i_n} = \pi_1 \triangleright \cdots \triangleright \pi_n$, and therefore, all π_is are marginal distributions of $\pi_1 \triangleright \cdots \triangleright \pi_n$. From this, perfectness of $\pi_1, \cdots, \pi_n$ is guaranteed by Theorem 10.3. $\square$

In the following example, we will show that the requirement for a generating sequence to be both perfect and flexible is rather strong, and in many situations such sequences may be simplified. In some ways, these sequences resemble the decomposable distributions, and therefore, as previously stated, we will learn more about perfect flexible sequences in Section 13.

Example 12.8. Consider a situation when $K_1 = \{1, 2\}$, $K_2 = \{2, 3\}$, $K_3 = \{3, 4\}$, $K_4 = \{1, 4, 5\}$ (see Figure 7) and assume that the sequence

$$\pi_1(x_1, x_2), \pi_2(x_2, x_3), \pi_3(x_3, x_4), \pi_4(x_1, x_4, x_5)$$

is perfect and flexible. We will show that in this case at least one of the distributions π_1, π_2, or π_3 can be deleted without changing the distribution represented by this flexible perfect sequence.

Since x_5 appears among the arguments of only π_4, due to the flexibility of the considered sequence there must exist an ordering $\pi_4, \pi_{i_1}, \pi_{i_2}, \pi_{i_3}$ such that

$$\pi_4 \triangleright \pi_{i_1} \triangleright \pi_{i_2} \triangleright \pi_{i_3} = \pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4.$$

Whatever the permutation i_1, i_2, i_3 is, $K_{i_3} \subset K_4 \cup K_{i_1} \cup K_{i_2}$, and therefore, $\pi_4 \triangleright \pi_{i_1} \triangleright \pi_{i_2} \triangleright \pi_{i_3} = \pi_4 \triangleright \pi_{i_1} \triangleright \pi_{i_2}$.

Now, it is an easy task to show that $\pi_4, \pi_{i_1}, \pi_{i_2}$ is perfect and flexible. Perfectness is an immediate consequence of the perfectness of the original sequence $\pi_1, \pi_2, \pi_3, \pi_4$ (all the distributions are marginals of the distribution $\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4 = \pi_4 \triangleright \pi_{i_1} \triangleright \pi_{i_2}$). Regarding the flexibility, we will consider two separate situations. If $i_1 = 2$, then flexibility

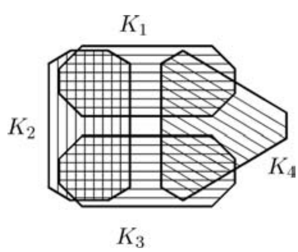


Figure 7. System of index sets from Example 12.8.

is guaranteed by the fact that $K_4 \cup K_2 = \{1, 2, 3, 4, 5\}$ and $\pi_4 \rhd \pi_2 \rhd \pi_{i_2} = \pi_2 \rhd \pi_4 \rhd \pi_{i_2}$ because of consistency of π_2 and π_4 (Lemma 4.4). Moreover, in this situation $K_{i_2} \subset K_4 \cup K_2$, and therefore, $\pi_4 \rhd \pi_2 = \pi_4 \rhd \pi_2 \rhd \pi_{i_2}$. If $i_1 \neq 2$, then K_4, K_{i_1}, K_{i_2} meets RIP and the flexibility follows from Lemma 12.3.

13. Decomposable models

To make a complete list of the considered special cases, we cannot omit sequences defining decomposable models. But the question 'What are the sequences defining decomposable models?' is not as simple as it looks at first sight.

In the field of graphical Markov models, decomposable distributions are those which factorize with respect to decomposable graphs, i.e. distributions ν which are uniquely given[4] by a system of their marginal distributions $\nu^{\downarrow L_1}, \nu^{\downarrow L_2}, \cdots, \nu^{\downarrow L_m}$ such that the index sets $L_1, L_2, \cdots, L_m$ can be ordered to meet RIP. Therefore, it seems quite natural to define a generating sequence $\pi(K_1), \pi(K_2), \cdots, \pi(K_n)$ to be decomposable, if the distributions are pairwise consistent and $K_1, K_2, \cdots, K_n$ meets RIP. Adopting this type of definition, we would get a proper subclass of flexible perfect sequences. The fact that these sequences are perfect and flexible follows from Lemmas 10.13 and 12.3. Obviously, there are flexible perfect sequences $\pi(K_1), \pi(K_2), \cdots, \pi(K_n)$ for which $K_1, K_2, \cdots, K_n$ does not meet RIP: for example, sequence $\pi_1(x_1, x_2), \pi_2(x_1, x_3), \pi_3(x_2, x_3)$ when π_1 and π_2 are consistent, and $\pi_3 = (\pi_1 \rhd \pi_2)^{\downarrow\{2,3\}}$. As another trivial example one can consider a sequence of uniform distributions, which is also flexible perfect regardless whether the respective sequence $K_1, \cdots, K_n$ meets RIP or not. This is why we propose the following, rather non-standard definition.

DEFINITION 13.1. A generating sequence $\pi_1(K_1), \pi_2(K_2), \cdots, \pi_n(K_n)$ is called *decomposable* if there exists a generating sequence $\nu_1(L_1), \nu_2(L_2), \cdots, \nu_r(L_r)$ such that each ν_k $(k = 1, 2, \cdots, r)$ is a marginal of some π_ℓ $(L_k \subseteq K_\ell,\ \nu_k = \pi_\ell^{\downarrow(L_k)})$,

$$\pi_1 \rhd \pi_2 \rhd \cdots \rhd \pi_n = \nu_1 \rhd \nu_2 \rhd \cdots \rhd \nu_r,$$

and $L_1, L_2, \cdots, L_r$ meets RIP.

Let us first illustrate the definition of decomposable sequences using two simple examples.

Example 13.2. In this example, we shall show that if $\pi_1(x_1, x_2), \pi_2(x_3, x_4), \pi_3(x_2, x_4, x_5)$ is perfect, then it is also decomposable.

Applying Theorem 9.4 to the considered sequence, we obtain $\pi_1 \rhd \pi_2 \rhd \pi_3 = \pi_1 \rhd (\pi_2 \circledcirc_{\{1,2\}} \pi_3)$. Due to the assumed perfectness of π_1, π_2, π_3 both $\pi_1^{\downarrow\{2\}}$ and

$$\left(\pi_2 \circledcirc_{\{1,2\}} \pi_3\right)^{\downarrow\{2\}} = \left(\pi_3^{\downarrow\{2\}}\, \pi_2 \rhd \pi_3\right)^{\downarrow\{2\}} = \pi_3^{\downarrow\{2\}}$$

equal to each other and therefore (cf. Lemma 4.4)

$$\pi_1 \rhd \pi_2 \rhd \pi_3 = \pi_1 \rhd \left(\pi_2 \circledcirc_{\{1,2\}} \pi_3\right) = \pi_2 \circledcirc_{\{1,2\}} \pi_3 \rhd \pi_1.$$

From this, we can see that

$$(\pi_1 \rhd \pi_2 \rhd \pi_3)^{\downarrow\{2,3,4,5\}} = \pi_2 \circledcirc_{\{1,2\}} \pi_3 = \pi_3^{\downarrow\{2\}} \pi_2 \rhd \pi_3$$

and therefore, $\pi_3^{\downarrow\{2\}}\pi_2 = (\pi_1 \rhd \pi_2 \rhd \pi_3)^{\downarrow\{2,3,4\}}$. In this way, we have deduced a consistency of $\pi_3^{\downarrow\{2\}}\pi_2$ and π_3 (both are marginals of $\pi_1 \rhd \pi_2 \rhd \pi_3$), and therefore,

$$\pi_3^{\downarrow\{2\}}\pi_2 \rhd \pi_3 = \pi_3 \rhd \pi_3^{\downarrow\{2\}}\pi_2 = \pi_3 \rhd \pi_2$$

(the last equality follows just from the definition of the operator $\rhd$). Thus, based on perfectness of $\pi_1 \rhd \pi_2 \rhd \pi_3$, we have obtained that $\pi_1 \rhd \pi_2 \rhd \pi_3 = \pi_3 \rhd \pi_2 \rhd \pi_1$. To show that π_1, π_2, π_3 is decomposable, it is enough to realize an obvious fact that the index sets corresponding to π_3, π_2, π_1 (i.e. $\{2, 4, 5\}$, $\{3, 4\}$, $\{1, 2\}$) meet RIP.

Example 13.3. This example presents a decomposable generating sequence that is neither flexible nor perfect. Let distributions from Tables 19 and 20 form a generating sequence $\pi_1, \pi_2, \pi_3, \pi_4$.

To verify that this sequence is not perfect, although all the distributions are pairwise consistent, it is enough to employ Lemma 10.2, compute $(\pi_1 \rhd \pi_2 \rhd \pi_3)^{\downarrow\{3,4\}}$ (see Table 21), and compare it with $\pi_4^{\downarrow\{3,4\}}$. Moreover, it means that π_4 is not a marginal of $\pi_1 \rhd \pi_2 \rhd \pi_3 \rhd \pi_4$, from which one can deduce that $\pi_1, \pi_2, \pi_3, \pi_4$ is not flexible (notice, namely, that the only distribution having X_5 among its arguments is π_4). Nevertheless, it is not difficult to show that

$$\pi_1 \rhd \pi_2 \rhd \pi_3 \rhd \pi_4 = \pi_1 \rhd \pi_2 \rhd \pi_3 \rhd \pi_4^{\downarrow\{3,5\}}$$

(we recommend the reader to do it; hint: notice that $X_5 \perp\!\!\!\perp X_4|X_3[\pi_4]$) and therefore, the sequence $\pi_1, \pi_2, \pi_3, \pi_4$ meets the condition of Definition 13.1.

Remark 13.4. The presented definition of decomposable sequences is so broad that it is really difficult to deduce interesting results on them. The reader certainly noticed that to show decomposability of the perfect sequence from Example 13.2, we showed that it was also flexible. As a rule, most flexible perfect sequences are decomposable. It had been an open problem for some time whether flexible perfect sequences exist which are not decomposable. And this is the main non-trivial result of this section: to show that it really is possible to find a perfect flexible sequence which is not decomposable.

Table 19. Probability distributions π_1, π_2, and π_3 of Example 13.3.

$\pi_1(x_1, x_2)$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$\frac{1}{6}$	$\frac{2}{6}$
$x_2 = 1$	$\frac{2}{6}$	$\frac{1}{6}$
$\pi_2(x_1, x_3)$	$x_1 = 0$	$x_1 = 1$
$x_3 = 0$	$\frac{1}{4}$	$\frac{1}{4}$
$x_3 = 1$	$\frac{1}{4}$	$\frac{1}{4}$
$\pi_3(x_2, x_4)$	$x_2 = 0$	$x_2 = 1$
$x_4 = 0$	$\frac{1}{2}$	0
$x_4 = 1$	0	$\frac{1}{2}$

Table 20. Probability distribution π_4 of Example 13.3.

	$x_3 = 0$		$x_3 = 1$	
π_4	$x_4 = 0$	$x_4 = 1$	$x_4 = 0$	$x_4 = 1$
$x_5 = 0$	$\frac{6}{24}$	$\frac{2}{24}$	$\frac{1}{24}$	$\frac{3}{24}$
$x_5 = 1$	$\frac{3}{24}$	$\frac{1}{24}$	$\frac{2}{24}$	$\frac{6}{24}$

Table 21. Probability distribution $(\pi_1 \triangleright \pi_2 \triangleright \pi_3)^{(\{3,4\})}$.

$x_3 = 0$		$x_3 = 1$	
$x_4 = 0$	$x_4 = 1$	$x_4 = 0$	$x_4 = 1$
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Example 13.5. Consider a sequence $\pi_1(x_1, x_2)$, $\pi_2(x_2, x_3, x_6)$, $\pi_3(x_3, x_4)$, $\pi_4(x_1, x_4, x_5)$ (see Figure 8) and assume it is perfect and flexible. First notice that $\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4 = \pi_2 \triangleright \pi_1 \triangleright \pi_3 \triangleright \pi_4$, because the first two distributions of any perfect sequence may be swapped. Then, since x_5 appears among the arguments of only π_4, due to the flexibility of the considered sequence, there must exist an ordering $\pi_4, \pi_{i_1}, \pi_{i_2}, \pi_{i_3}$ such that

$$\pi_4 \triangleright \pi_{i_1} \triangleright \pi_{i_2} \triangleright \pi_{i_3} = \pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4.$$

If $i_1 = 2$, then $\pi_4 \triangleright \pi_2 = \pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4$, and K_4, K_2 meets RIP. Similarly, if $i_2 = 2$, then $\pi_4 \triangleright \pi_{i_1} \triangleright \pi_2 = \pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4$, and K_4, K_{i_1}, K_2 meets RIP. Therefore, these situations resemble Example 12.8 – one or two distributions may be omitted without loss of possibility to define the required distribution $\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4$. Therefore, these situations cannot yield the required example and we have to analyze the situation when $i_3 = 2$, i.e. the situation when

$$\pi_4 \triangleright \pi_1 \triangleright \pi_3 \triangleright \pi_2 = \pi_2 \triangleright \pi_1 \triangleright \pi_3 \triangleright \pi_4. \tag{17}$$

($\pi_4 \triangleright \pi_3 \triangleright \pi_1 = \pi_4 \triangleright \pi_1 \triangleright \pi_3$ because of Lemma 5.7.) Denote the six-dimensional distribution from (17) by ν.

Application of Lemma 5.2 to $\pi_2 \triangleright \pi_1 = \nu^{(\{1,2,3,6\})}$ yields $X_3, X_6 \perp\!\!\!\perp X_1 | X_2[\nu]$, and, due to Lemma 2.3, also

$$X_3 \perp\!\!\!\perp X_1 | X_2[\nu].$$

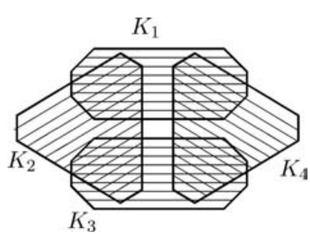


Figure 8. System of index sets from Example 13.5.

From this independence, employing the well-known property of mutual information (Gallager 1968), one gets $\mathrm{MI}_\nu(X_1;X_2) \geq \mathrm{MI}_\nu(X_1;X_3)$ and also $\mathrm{MI}_\nu(X_2;X_3) \geq \mathrm{MI}_\nu(X_1;X_3)$. Analogously, from $\pi_2 \triangleright \pi_1 \triangleright \pi_3 = \nu^{(\{1,2,3,4,6\})}$, we are getting $X_4 \perp\!\!\!\perp X_1, X_2, X_6 | X_3$, and therefore, also $\mathrm{MI}_\nu(X_1;X_3) \geq \mathrm{MI}_\nu(X_1;X_4)$.

Considering now $\pi_4 \triangleright \pi_1 = \nu^{(\{1,2,4,5\})}$ and $\pi_4 \triangleright \pi_1 \triangleright \pi_3 = \nu^{(\{1,2,3,4,5\})}$, we get, in an analogic way, $\mathrm{MI}_\nu(X_1;X_4) \geq \mathrm{MI}_\nu(X_2;X_4)$, $\mathrm{MI}_\nu(X_2;X_4) \geq \mathrm{MI}_\nu(X_2;X_3)$, and $\mathrm{MI}_\nu(X_3;X_4) \geq \mathrm{MI}_\nu(X_2;X_3)$. Combining all six of these inequalities, one immediately gets the following necessary condition that must hold true for a distribution ν defined by both generating sequences $\pi_2, \pi_1, \pi_3, \pi_4$ and $\pi_4, \pi_1, \pi_3, \pi_2$:

$$\begin{aligned}\mathrm{MI}_\nu(X_1;X_2) \geq \mathrm{MI}_\nu(X_1;X_3) = \mathrm{MI}_\nu(X_2;X_3) = \mathrm{MI}_\nu(X_1;X_4)\\ = \mathrm{MI}_\nu(X_2;X_4) \leq \mathrm{MI}_\nu(X_3;X_4).\end{aligned} \tag{18}$$

This is a rather strong condition, which is met, for example, when all these expressions for mutual information equal 0, i.e. when variables X_1, X_2, X_3, X_4 are independent. In this case, however, distribution ν may be expressed just as a composition $\pi_2 \triangleright \pi_4$ and we would not get anything new in comparison with Example 12.8.

However, the inequalities (18) are also satisfied when

$$X_1 \perp\!\!\!\perp X_4[\pi_4] \quad \text{and} \quad X_2 \perp\!\!\!\perp X_3[\pi_2], \tag{19}$$

and X_2 and X_4 are copies of X_1 and X_3, respectively, by which we understand that

$$\begin{aligned}\pi_1(x_1, x_2) = 0 \quad \text{for } x_1 \neq x_2,\\ \pi_3(x_3, x_4) = 0 \quad \text{for } x_3 \neq x_4.\end{aligned} \tag{20}$$

Thus, if distributions π_1, π_2, π_3, and π_4 are pairwise consistent and expressions (19) and (20) hold true, the sequence $\pi_1, \pi_2, \pi_3, \pi_4$ is perfect and flexible (to prove it, verify that both $\pi_2, \pi_1, \pi_3, \pi_4$ and $\pi_4, \pi_1, \pi_3, \pi_2$ satisfy conditions of Lemma 10.2). In addition to this, if $X_6 \perp\!\!\!\perp X_2[\pi_2]$, $X_6 \perp\!\!\!\perp X_3[\pi_2]$, $X_5 \perp\!\!\!\perp X_1[\pi_4]$, and $X_5 \perp\!\!\!\perp X_4[\pi_4]$, then this sequence is not decomposable. To show it, realize that the distribution $\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4$ is defined only by the generating sequences, where either π_2 or π_4 is the last.

Remark 13.6. Perhaps the importance of flexible perfect sequences can be emphasized even more by mentioning that the above presented example has its weak spot. We do not know whether the inequalities (18) can be met by any types of distributions other than those mentioned in the example: X_1, X_2, X_3, X_4 independent, or, distributions for which (19) and (20) hold true. If not, we have to admit that the example is not fully convincing because, although it does not meet conditions of Definition 13.1, the considered distribution $\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4$ can be represented in a decomposable form. This is because X_2 and X_4 are copies of X_1 and X_3, respectively, and therefore, we can see that

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4 = \pi_2 \triangleright (\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4)^{\downarrow\{2,3,5\}} \triangleright \pi_1 \triangleright \pi_3,$$

which is a sequence of distributions whose sets of arguments meet RIP. We see that the dimensionality of new distributions (in fact, we consider only one new distribution, namely $(\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4)^{\downarrow\{2,3,5\}}$) is the same as the dimensionality of the original distributions, but the necessary computations of $(\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4)^{\downarrow\{2,3,5\}}$ cannot be performed locally.

14. Summary

This survey paper summarizes basic theoretical properties on compositional models. The first part of the paper introduces operators of composition and their properties, and the second part shows how these operators are used for multidimensional probability distribution representation. It is worth repeating that among all generating sequences, perfect sequences play an important role, because they faithfully reflect the information contained in the individual distributions from the generating sequence. This property is important from the point of view of potential applications; when the individual oligodimensional distributions π_i represent pieces of local knowledge, then $\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \cdots \triangleright \pi_n$ is a proper representative of global knowledge. Though we have not dealt with this question, decomposable sequences are most prosperous from a computational point of view, although flexible perfect sequences seem to have almost equally advantageous properties.

Acknowledgements

The author wants to express his appreciation to an anonymous referee, whose relevant comments helped to improve the final version of the paper. The referee also suggested an original and interesting idea to define the operator of composition in the way that the resulting distribution $\pi(K) \triangleright \kappa(L)$ would be such a distribution $\nu(K \cup L)$ that $\nu^{\downarrow K} = \pi$ and $\nu^{\downarrow L}(L \setminus K | K \cap L) = \kappa^{\downarrow L}(L \setminus K | K \cap L)$. This alternative definition would possess an obvious advantage: the composition would always be defined (though sometimes ambiguously) and it seems that even some proofs would be simplified, because one would not have to pay attention to situations when the composition is not defined. The reason we have not accepted this suggestion is that the whole apparatus of compositional models was designed with the goal of getting a *constructive* tool for multidimensional distribution construction and processing. Therefore, we are convinced that it should be based on principles of intuitionistic logic and it seems to us that the presented definition (with all its disadvantageous properties) better complies with this requirement.

This paper is the culmination of a long-term research on compositional models in probability theory. Therefore, it is quite natural that during its development the research was supported by a number of research grants. Here, the author wants to mention only those which have been active during the last 3 or 4 years: GA AV ČR no. A2075302, GAČR no. ICC/08/E010 and 201/09/1891, and MŠMT ČR no. 1M0572 and 2C06019.

Notes

1. $H(\nu)$ denotes the Shannon entropy of ν defined $H(\nu) = -\sum_{x:\nu(x)>0} \nu(x) \log \nu(x)$.
2. The reader can find it in the literature often under names *I-divergence* or *cross-entropy*.
3. Notice that $\pi_1(x_2 | x_1 = a)$ is a one-dimensional distribution of X_2.
4. More precisely, ν is a maximum entropy extension of the given system of its marginal distributions.

Notes on contributor

Radim Jiroušek graduated in Mathematics from Charles University (Prague) in 1969. He received his PhD and DrSc degrees from the Czech Academy of Sciences in 1979 and 1993, respectively. Currently, he is with the University of Economics, Prague, where he serves as a Dean of the Faculty of Management, which is located in Jindřichův Hradec, and with the Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic.

References

Csiszár, I., 1975. I-divergence geometry of probability distributions and minimization problems. *Annals of probability*, 3, 146–158.

Deming, W.E. and Stephan, F.F., 1940. On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of mathematical statistics*, 11, 427–444.

Dempster, A., 1967. Upper and lower probabilities induced by a multi-valued mapping. *Annals of mathematical statistics*, 38, 325–339.

Dubois, D. and Prade, H., 2001. Possibility theory, probability theory and multiple-valued logics: a clarification. *Annals of mathematics and artificial intelligence*, 32, 35–66.

Gallager, R.G., 1968. *Information theory and reliable communications*. New York: John Wiley.

Habermann, S.J., 1974. *The analysis of frequency data*. Chicago, IL/London: The University of Chicago Press.

Jensen, F.V., 2001. *Bayesian networks and decision graphs*. New York: IEEE Computer Society Press.

Jiroušek, R., 1997. Composition of probability measures on finite spaces. *In*: D. Geiger and P.P. Shenoy, eds. *Proceedings of the 13th conference uncertainty in artificial intelligence UAI'97*. San Francisco, CA: Morgan Kaufmann Publ, 274–281.

Jiroušek, R., 1998. Graph modelling without graphs. *In*: B. Bouchon-Meunier and R.R. Yager, eds. *Proceedings of the 7th international conference on information processing and management of uncertainty in knowledge-based systems IPMU'98*. Paris: Editions E.D.K., 809–816.

Jiroušek, R., 2000. Marginalization in composed probabilistic models. *In*: C. Boutilier and M. Goldszmidt, eds. *Proceedings of the 16th conference uncertainty in artificial intelligence UAI'00*. San Francisco, CA: Morgan Kaufmann Publ., 301–308.

Jiroušek, R., 2004a. What is the difference between Bayesian networks and compositional models? *In*: H. Noguchi, H. Ishii and M. Inuiguchi, eds. *Proceedings of the 7th Czech-Japan seminar on data analysis and decision making under uncertainty*. Japan: Awaji Yumebutai ICC, 191–196.

Jiroušek, R., 2004b. On relation of compositional and decomposable models. *In*: B. Bouchon-Meunier, G. Coletti and R.R. Yager eds. *Proceedings of the 10th international conference on information processing and management of uncertainty in knowledge-based system*. Roma, Italy: Editrice Universita La Sapienza, 1057–1064.

Jiroušek, R., 2008. Persegrams of compositional models revisited: conditional independence. *In*: L. Magdalena, M. Ojeda-Aciego and J.L. Verdegay, eds. *Proceedings of the 12th international conference on information processing and management of uncertianlty in konowledge-based systems IPMU'08*, Málaga, Spain, 915–922.

Jiroušek, R. and Vejnarová, J., 2003. General framework for multidimensional models. *International journal of intelligent systems*, 18, 107–127.

Jiroušek, R., Vejnarová, J. and Daniel, M., 2007. Compositional models of belief functions. *In*: G. De Cooman, J. Vejnarová and M. Zaffalon, eds. *Proceedings of the 5th symposium on imprecise probabilities and their applications, ISIPTA'07*. Prague: Action M Agency, 243–252.

Kellerer, H.G., 1964. Verteilungsfunktionen mit gegeben Marginalverteilungen. *Z. Warhsch. Verw. Gebiete*, 3, 247–270.

Klir, G.J., 2006. *Uncertainty and information. Foundations of generalized information theory*. Hoboken, NJ: Wiley.

Lauritzen, S.L., 1996. *Graphical models*. Oxford, UK: Oxford University Press.

Lauritzen, S.L. and Spiegelhalter, D.J., 1988. Local computation with probabilities on graphical structures and their application to expert systems. *Journal of the royal statistical society series B*, 50, 157–224.

Pearl, J., 1988. *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufmann.

Shafer, G., 1976. *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.

Studený, M., 2005. *Probabilistic conditional independence structures*. London: Springer.

Vejnarová, J., 1998a. 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, IPMU'98*. Paris: Editions E.D.K., 25–30.

Vejnarová, J., 1998b. Possibilistic independence and operators of composition of possibility measures. *In*: M. Hušková, J.Á. Víšek and P. Lachout, eds. *Prague stochastics'98*. JČMF, 575–580.

Vomlel, J., 1999. *Methods of probabilistic knowledge integration*. Thesis (PhD). FEL ČVUT, Praha, Czech Republic.