Multidimensional Compositional Models

Part I: Introduction

Radim Jiroušek

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic
radim@utia.cas.cz

April 9, 2006

Contents

1 Introduction **9**

2 Notions of probability theory – notation **13**
2.1 Conditional independence of variables 15
2.2 Extensions of distributions 26
2.3 Information-theoretic notions 30

3 Operators of composition **35**
3.1 Definition of operators 35
3.2 Basic properties 40
3.3 I-geometry of composition operators 50
3.4 Iterations of the operators of composition 53

4 Generating sequences **73**
4.1 Perfect sequences 74
4.2 Commutable sets 90
4.3 Flexible sequences 107

Bibliography **113**

Notice

This text has not been subjected to linguistic corrections. Nevertheless, in view of the author's intention to improve, complete and extend the text, he wants to assure all the readers that he will highly appreciate receiving any kind of comments, proposals and questions that can help him to correct (both from the linguistic and the mathematical point of view) the text and also to make it easier to understand. All such comments and proposals should be sent to radim@utia.cas.cz.

Foreword

The first version of this text, which summarizes results published mainly in papers [14] – [20], was assembled as an accompanying material to my lectures at the summer school REASON PARK held in Foligno (Perugia), Italy, 26th August - 14th September, 2002. In this context I want to express my thanks to the organizers for the invitation, since this was an impulse for starting preparation of this text.

The goal of this text is to coherently explicate the apparatus of probabilistic operators of composition, which is, in a way, an alternative approach to multidimensional probability distributions representation and processing. When saying "alternative" approach, I mean that it is an alternative to widely used *graphical Markov modelling*.

The text consists of two parts. The presented first part is a detailed description of the basic theoretical properties of operators of composition and, above all, different types of generating sequences that are models of multidimensional distribution. Since this approach is for most of the readers new, the concepts and their properties are illustrated by numerous examples. The second part, which is under preparation, will be devoted to relation of compositional models to classical Graphical Markov Models and to some other advanced issues like reading probability independence relations from generating sequences, or heuristic approaches to model construction. In future, the third part could appear describing a program system MUDIM, which is an open source system determined to be an experimental environment for compositional model constructions.

Though all the necessary concepts are introduced, the reader is expected to be familiar with basic notions of (finite) probability theory. With respect to the facts that a couple of notions of information theory are employed, and that a relation of the compositional approach to the mostly used graphical models is explained, knowledge of both information theory and graphical Markov modelling is advantageous, but not anything like necessary.

Chapter 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; many others appeared in the second half of the last century, though, from many-valued and fuzzy logics, through rough sets theory to approaches based on non-additive measures, e.g., possibility theory.

In this text 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 [24, 32]. That means they can also be applied to situations when the assumption of additivity is not adequate [8][1]. Nevertheless, in this text we shall restrict our consideration only to probabilistic models.

The basic idea of the approach is the same as that on which expert systems are based: it is beyond human capabilities to represent or express global knowledge of an application area - one has always to work only with pieces of local knowledge. Such a local knowledge can be, within probability theory, easily represented by an oligodimensional (low-dimensional) distribution. For example, the statement *directors are usually older persons*

[1] For example, in a situation when you consider uncertainty connected with your subjective estimate of a person's age, using additive measure (probability) could lead you into troubles. In this case, insisting on additivity could lead to a situation that uncertainty connected with estimating age of two persons would have to be a sum of uncertainties connected with the individual persons. This means that the larger group of persons is observed, the more precise estimates of individual persons would have to be, otherwise the total uncertainty would have to get above the maximum value – one.

Table 1.1: Distribution representing knowledge: *directors are usually older persons*

age	director of an enterprise		
	small up to 20 employees	medium up to 150 employees	large 150 + employees
20 - 30	0.032	0.014	0.004
31 - 40	0.061	0.086	0.034
41 - 50	0.102	0.118	0.051
51 +	0.114	0.183	0.201

can be well (and moreover, quite precisely) expressed by a 2-dimensional distribution from Table 1.1. From this table one can easily get that among directors there are only 5 percent of persons younger 31 while 27.1 percent are between 41 and 50 and almost half of them (49.8 %) are older than 50. Notice, that such a table can yield even more *information* than that contained in the statement *directors are usually older persons*. For example, from this table one can also get that usually *the larger enterprize, the older director*.

Analogously, a 3-dimensional distribution (an example of which is in Table 1.2) can easily express the knowledge *relationship between sex, age and occurrence of diabetes*. A great advantage of this type of local knowledge representation is the fact that in a majority of situations, low-dimensional distributions can be obtained from various data sources by classical statistical estimates. What should be stressed, however, is the fact that in such situations the dimensionality of the estimated distributions is strictly limited because of the size of available data. Whatever size of data is at our disposal we can hardly assume to obtain reliable estimates of probabilities of a 20-dimensional distribution (even for binary variables). Typically, one can assume that a dimensionality of the considered distributions is $2-8$. Therefore, we will 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

Table 1.2: Distribution representing knowledge: *relationship between sex, age and occurrence of diabetes*

age	diabetes			
	(-)		(+)	
	sex			
	F	M	F	M
20 - 40	0.025	0.024	0.003	0.003
41 - 50	0.075	0.072	0.020	0.018
51 - 60	0.096	0.086	0.048	0.045
61 - 70	0.067	0.052	0.044	0.037
71 +	0.084	0.059	0.086	0.056

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 catalyzed by success, which was achieved during the last twenty years in the field that is often called *graphical Markov modeling*. This term is used as a general term describing 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 others. What is common to all these models 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[2], represent distributions with special dependence structures which are described by acyclic directed graphs. Some other models, like decomposable models, use for the dependence structure representation undirected graphs, and special models need even more complicated graphical tools like chain graphs, hypergraphs, or, annotated graphs.

[2]We shall discuss Bayesian networks and decomposable models from the point of view of this text in more detail in Chapter 5.

The approach presented herein abandons the necessity to describe the dependence structure of a modelled distribution by 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 is a (usually great) number of pieces of local knowledge represented by a system of low-dimensional distributions. The task we will address in this text resembles a jig-saw puzzle that has a great number of parts, each bearing a local piece of a picture, and the goal is to find how to assemble them in such a way that the global picture makes sense, reflecting all 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 the local knowledge carried by the oligodimensional distributions

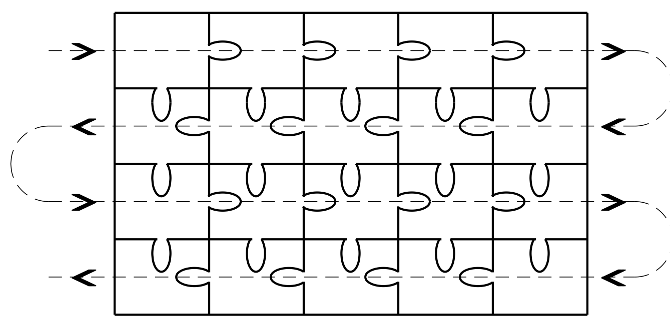


Figure 1.1: Ordering of pieces of jig-saw puzzle.

Chapter 2

Notions of probability theory – notation

In this text, 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 $\mathbf{X}_i$. Thus, in Table 1.2 there occur three variables *age*, *diabetes* and *sex* having 5, 2 and 2 values, respectively.

Distributions of these variables will be denoted by Greek letters (usually π, κ, ν, μ); 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 $\pi(x_K)$ will denote both a $|K|$-dimensional distribution and a value of a probability distribution π (when several distributions will be considered, we shall distinguish them by indices), which is defined for variables $(X_i)_{i \in K}$ at a combination of values x_K; x_K thus represents a $|K|$-dimensional vector of values of variables $\{X_i\}_{i \in K}$. Analogously, we shall also denote the set of all these vectors $\mathbf{X}_K$:

$$\mathbf{X}_K = \times_{i \in K} \mathbf{X}_i.$$

A distribution $\pi(x_K)$ is represented by a $|K|$-dimensional table containing numbers from the interval $[0, 1]$, and all the numbers of this distribution have to sum up to one:

$$\sum_{x_K \in \mathbf{X}_K} \pi(x_K) = 1.$$

An example of such a table/distribution is in Table 1.2. In this case $x_{\{1,2,3\}}$ represents a combination of values of the considered variables *age*, *diabetes*, *sex*. Thus, $x_{\{1,2,3\}}$ is one of the 20 combinations: $(20 - 40, (-), F)$, $(20 -$

$40, (+), M)$, $(20 - 40, (-), F)$, ..., $(71+, (+), M)$, and, for example

$$\pi(61 - 70, (+), F) = 0.044.$$

For a probability distribution $\pi(x_K)$ and $J \subset K$ we will often consider a *marginal distribution* $\pi(x_J)$ of $\pi(x_K)$, which can be computed by

$$\pi(x_J) = \sum_{x_{K \setminus J} \in \mathbf{X}_{K \setminus J}} \pi(x_K) = \sum_{x_{K \setminus J} \in \mathbf{X}_{K \setminus J}} \pi(x_{K \setminus J}, x_J).$$

An example of a marginal distribution for distribution from Table 1.2 is in Table 2.1.

Table 2.1: A marginal distribution to distribution from Table 1.2

	diabetes	
	(-)	(+)
20 - 40	0.049	0.006
41 - 50	0.147	0.038
51 - 60	0.182	0.093
61 - 70	0.119	0.081
71 +	0.143	0.142

In the above simple formula defining a marginal distribution we have implicitly introduced a notation, which will be used in the sequel. A vector x_K is split into two parts: vectors $x_{K \setminus J}$ and x_J, where x_J is a *projection* of x_K into $\mathbf{X}_J$, and, analogously $x_{K \setminus J}$ is a projection of x_K into $\mathbf{X}_{K \setminus J}$. For computation of marginal distributions we need not exclude situations when $J = \emptyset$. In accordance with the above introduced formula we get $\pi(x_\emptyset) = 1$.

In some situations, when we will want to stress that we are dealing with a marginal distribution of a distribution π, we will use symbol $\pi^{\downarrow J}$ to denote the marginal distribution of π for variables $(X_i)_{i \in J}$; i.e., for $J \subseteq K$ and a distribution $\pi(x_K)$[1]

$$\pi^{\downarrow J} = \pi(x_J).$$

[1]This notation is taken over from G. Shachter and P.P. Shenoy. Their notation will also enable us to denote variables, which will be deleted during marginalization process. This special notation will be introduced later.

Table 2.2: Conditional distribution computed from a distribution from Table 1.2

age	diabetes			
	(-)		(+)	
	sex			
	F	M	F	M
20 - 40	0.072	0.082	0.015	0.019
41 - 50	0.216	0.246	0.099	0.113
51 - 60	0.277	0.294	0.239	0.283
61 - 70	0.193	0.177	0.219	0.233
71 +	0.242	0.201	0.428	0.352

For a distribution $\pi(x_K)$ and two disjoint subsets $L_1, L_2 \subseteq K$ we will often speak about a *conditional distribution* $\pi(x_{L_1}|x_{L_2})$, which is, for each fixed $x_{L_2} \in \mathbf{X}_{L_2}$ an $|L_1|$-dimensional probability distribution, for which

$$\pi(x_{L_1}|x_{L_2})\pi(x_{L_2}) = \pi(x_{L_1 \cup L_2}).$$

It is important to realize that, if $\pi(x_{L_2}) = 0$ for some combination(s) of values $x_{L_2} \in \mathbf{X}_{L_2}$, this definition is ambiguous. Nevertheless, the advantage of this definition is that conditional distribution is always defined. The reader can immediately see that if $L_1 = \emptyset$ then

$$\pi(x_{L_1}|x_{L_2}) = 1,$$

and if $L_2 = \emptyset$ then

$$\pi(x_{L_1}|x_{L_2}) = \pi(x_{L_1}).$$

An example of a conditional distribution $\pi(age|diabetes, sex)$ computed from Table 1.2 is in Table 2.2

2.1 Conditional independence of variables

In this section, we shall introduce one of the most important notions of this text, a concept of *conditional independence* that generalizes well-known independence of variables. Since this notion does not belong among the basic subjects notoriously repeated in all textbooks on probability theory, we shall illustrate the notion by several simple examples. For more examples the reader is referred to basic textbooks on Bayesian networks like books by F.V. Jensen [11] and [12].

Definition 2.1.1 Consider a probability distribution $\pi(x_K)$ and three disjoint subsets $L_1, L_2, L_3 \subseteq K$ such that $L_1 \neq \emptyset \neq L_2$. We say that groups of variables X_{L_1} and X_{L_2} are *conditionally independent* given X_{L_3} (in symbol $X_{L_1} \perp\!\!\!\perp X_{L_2}|X_{L_3}[\pi]$) if

$$\pi(x_{L_1\cup L_2\cup L_3})\pi(x_{L_3}) = \pi(x_{L_1\cup L_3})\pi(x_{L_2\cup L_3}) \tag{2.1}$$

for all $x_{L_1\cup L_2\cup L_3} \in \mathbf{X}_{L_1\cup L_2\cup L_3}$.

Remark 2.1.1 Equality (2.1) certainly holds for all $x_{L_1\cup L_2\cup L_3} \in \mathbf{X}_{L_1\cup L_2\cup L_3}$, for which $\pi(x_{L_3}) = 0$. This is because $\pi(x_{L_1\cup L_3}) \leq \pi(x_{L_3})$ due to the way how marginal distributions are defined. For those $x_{L_1\cup L_2\cup L_3}$, for which $\pi(x_{L_3}) > 0$, we can divide both sides of equality (2.1) by $\pi(x_{L_3})$, which gives us

$$\pi(x_{L_1\cup L_2\cup L_3}) = \pi(x_{L_1\cup L_3})\pi(x_{L_2}|x_{L_3}). \tag{2.2}$$

Since equation (2.2) holds true also for all $x_{L_1\cup L_2\cup L_3}$, for which $\pi(x_{L_3}) > 0$, we could define conditional independence by requirement that equality (2.2) holds true for all $x_{L_1\cup L_2\cup L_3} \in \mathbf{X}_{L_1\cup L_2\cup L_3}$. ○

Remark 2.1.2 Another way how to define conditional independence is to require that the expression

$$\pi(x_{L_1\cup L_2}|x_{L_3}) = \pi(x_{L_1}|x_{L_3})\pi(x_{L_2}|x_{L_3})$$

holds true for all vectors $x_{L_1\cup L_2\cup L_3}$, for which $\pi(x_{L_3}) > 0$. Moreover, since

$$\pi(x_{L_1\cup L_2}|x_{L_3}) = \pi(x_{L_1}|x_{L_3})\pi(x_{L_2}|x_{L_1\cup L_3})$$

is valid for all $x \in \mathbf{X}_{L_1\cup L_2\cup L_3}$, for which $\pi(x_{L_3})$ is positive, and for any distribution π (regardless it meets the property of conditional independence), we can see that the conditional independence can also be expressed in the following way:

$$X_{L_1} \perp\!\!\!\perp X_{L_2}|X_{L_3}[\pi]$$
$$\iff \forall x \in \mathbf{X}_{L_1\cup L_2\cup L_3} : \pi(x_{L_3}) > 0 \;\; (\pi(x_{L_2}|x_{L_3}) = \pi(x_{L_2}|x_{L_1\cup L_3})).$$

The last formula is often used to explain the concept of conditional independence. It says that conditional probability of variables X_{L_2} given variables X_{L_3} is the same as conditional probability of these variables given variables $X_{L_1 \cup L_3}$. In other words, if we know values of variables X_{L_3}, the conditional probability of variables X_{L_2} does not change if we learn also values of variables X_{L_1}. ○

Since this notion is of great importance for multidimensional model construction, let us illustrate it on a couple of examples.

Example 2.1.1 *From the point of view of this text interesting are those situations when groups of variables X_{L_1} and X_{L_2} are dependent but conditionally (given variables X_{L_3}) independent. We will not be interested in trivial situations when all variables are mutually independent like, for example, when considering three independent tosses of a coin: any two tosses are not only (unconditionally) independent but they are also conditionally independent given the third toss. It holds even if the coin is unfair (i.e., probability of one side is higher than probability of the other side). This can easily be proven for any 3-dimensional product distribution, i.e. $\pi(x_1, x_2, x_3) = \pi(x_1)\pi(x_2)\pi(x_3)$.* ♢

Example 2.1.2 *Realize that interesting situations of conditional independence are sometimes connected with (non-deterministic) causality. For example, consider three variables X_1, X_2, X_3: first reflects effort of a student to learn a subject (e.g. number of days of learning), second represents result of a test (number of received points) and X_3 is just a binary variable with 0 and 1 corresponding to failure and success in the examination, respectively. In this case it is quite natural to expect that variables X_1 and X_3 are dependent ($X_1 \not\perp\!\!\!\perp X_3$) but, simultaneously, they are conditionally independent given X_2 ($X_1 \perp\!\!\!\perp X_3 | X_2$). This corresponds to the fact that two students with the same result of a test should be equally treated in spite of how much time they spent by learning the subject.* ♢

Example 2.1.3 *Imagine a statistician, discovering by a thorough statistical analysis a high correlation (i.e. dependence) between daily harvest of honey and beer consumption. It would be naïve to explain this dependence by trying to find a way how froth-blowers influence bees or vice versa. The natural way how to explain this dependence stems from the fact that both the considered events (load of honey and beer consumption) depend on weather: the warmer wether, the higher harvest of honey and the higher beer consumption. So, this natural example explains, in a slightly different* *way, the situation when two variables (beer and honey) are dependent but turn independent when considering their conditional independence given weather.*

Let us repeat the preceding intuitive considerations using formal probabilistic tools. In the considered situation we assume there is a probability distribution $\pi(weather)$ *describing how often individual types of weather occurs (as we assume all the variables are finite-valued, we are allowed to consider only a finite number of weather classes). Then it is natural to assume that load of honey is dependent on weather, it means we can describe it quite reasonably by a distribution* $\kappa(honey|weather)$. *Similarly, beer consumption also changes under different weather conditions, so it may be described by* $\mu(beer|weather)$. *Our assumption that there is no direct dependence between variables beer and honey reflects in the way how we define the 3-dimensional distribution representing relationships among the considered variables:*

$$\nu(weather, honey, beer) = \pi(weather)\kappa(honey|weather)\mu(beer|weather).$$

Marginalizing this expression we get

$$\nu(weather, honey) = \pi(weather)\kappa(honey|weather),$$

and therefore

$$\nu(beer|weather, honey) = \mu(beer|weather).$$

This means that $\nu(beer|weather, honey)$ *does not depend on variable honey, and therefore, due to Remark 2.1.2, variables honey and beer are conditionally independent given weather.* ◊

Example 2.1.4 *It should be stressed, however, that conditional independence need not be connected with causality. In some situations it can be just a coincidence. Let us present a (slightly modified) example taken over from [21].*

Consider a family with two children: a son and a daughter. Define three variables:

X_s – *son goes to visit grandmother,*
X_d – *daughter goes to visit grandmother,*
X_f – *father goes to visit grandmother,*

and assume the situation is described by the 3-dimensional distribution from Table 2.3. To verify conditional independence $X_s \perp\!\!\!\perp X_d | X_f$ *we have to show that*

Table 2.3: Distribution $\pi(x_s, x_d, x_f)$ from Example 2.1.4

		X_d			
		yes		no	
		X_s			
		yes	no	yes	no
X_f	yes	$\frac{4}{18}$	$\frac{2}{18}$	$\frac{2}{18}$	$\frac{1}{18}$
	no	$\frac{1}{18}$	$\frac{2}{18}$	$\frac{2}{18}$	$\frac{4}{18}$

$$\pi(x_s, x_d, x_f)\pi(x_f) = \pi(x_s, x_f)\pi(x_d, x_f)$$

holds true for all eight possible combinations. In case that all three variables achieve value "yes" we are getting:

$\pi(X_s = yes, X_d = yes, X_f = yes) = \frac{4}{18}$,
$\pi(X_s = yes, X_f = yes) = \frac{6}{18}$,
$\pi(X_d = yes, X_f = yes) = \frac{6}{18}$,
$\pi(X_f = yes) = \frac{9}{18}$,

and since

$$\frac{4}{18}\frac{9}{18} = \frac{6}{18}\frac{6}{18}$$

we see that the required equality is met for this combination of values. When the reader shows that the equality holds also for the remaining 7 combinations of values, the conditional independence $X_s \perp\!\!\!\perp X_d | X_f$ will be verified. Simultaneously, any effort to introduce a causality in this example would be rather artificial. ♢

Example 2.1.5 *Up to now, all the examples illustrating conditional independence described situations when two variables were dependent but conditionally independent given the third variable:*

$$X_1 \not\!\perp\!\!\!\perp X_2 \quad \& \quad X_1 \perp\!\!\!\perp X_2 | X_3.$$

Here we present an example of an opposite situation. Consider three variables X_1, X_2, X_3 corresponding to three coins with 0 and 1 on their sides. Two of them are randomly tossed, the third one is laid on the table so that

Table 2.4: Distribution describing example with 3 coins

<table>
<tr><td colspan="2" rowspan="4">$\kappa(x_1, x_2, x_3)$</td><td colspan="4">X_1</td></tr>
<tr><td colspan="2">0</td><td colspan="2">1</td></tr>
<tr><td colspan="4">X_2</td></tr>
<tr><td>0</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td rowspan="2">X_3</td><td>0</td><td>0</td><td>$\frac{1}{4}$</td><td>$\frac{1}{4}$</td><td>0</td></tr>
<tr><td>1</td><td>$\frac{1}{4}$</td><td>0</td><td>0</td><td>$\frac{1}{4}$</td></tr>
</table>

each experiment results with even number of 0*s. The reader can easily check that the situation is well described by the distribution in Table 2.4.*

It is clear that variables X_1 and X_2 are independent. In fact, it is really trivial to show that any pair X_i, X_j $(i \neq j)$ of variables are independent. Now, let us consider a definition of conditional independence

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

for example for $(x_1, x_2, x_3) = (0, 0, 0)$. Marginalizing distribution from Table 2.4 one gets that any 2-dimensional distribution is uniform, *equals $\frac{1}{4}$ for any combinations of values of variables. Therefore $\kappa(x_1 = 0, x_3 = 0)\kappa(x_2 = 0, x_3 = 0) = \frac{1}{16}$, but $\kappa(x_1 = 0, x_2 = 0, x_3 = 0)\kappa(x_3 = 0) = 0$, which contradicts the definition of conditional independence. Therefore, in this example*

$$X_1 \perp\!\!\!\perp X_2 \quad \& \quad X_1 \not\perp\!\!\!\perp X_2 | X_3. \qquad \Diamond$$

Remark 2.1.3 The reader certainly noticed that, since the definition of conditional independence is symmetric with respect to groups of variables X_{L_1} and X_{L_2}, also the following three conditions are equivalent

- $X_{L_1} \perp\!\!\!\perp X_{L_2} | X_{L_3}[\pi]$,
- $\forall x \in \mathbf{X}_{L_1 \cup L_2 \cup L_3} \pi(x_{L_1 \cup L_2 \cup L_3}) = \pi(x_{L_2 \cup L_3})\pi(x_{L_1}|x_{L_3})$,
- $\forall x \in \mathbf{X}_{L_1 \cup L_2 \cup L_3} : \pi(x_{L_3}) > 0 \pi(x_{L_1}|x_{L_3}) = \pi(x_{L_1}|x_{L_2 \cup L_3})$. ○

Let us conclude this section by presenting three lemmata expressing important results regarding conditional independence of variables. They can be found in several books on probabilistic multidimensional models like for example in [28]. The proofs presented here are taken over from [21]. The purpose of presenting the proofs of famous assertions is to offer the reader some exercise in computations with probability distributions.

The first assertion is often called *factorization rule* or *factorization lemma*.

Lemma 2.1.1 *Let $K, L \subset N$ be such that $K \setminus L \neq \emptyset \neq L \setminus K$. Then for any probability distribution $\pi(x_{K\cup L})$*

$$X_{K\setminus L} \perp\!\!\!\perp X_{L\setminus K} | X_{K\cap L}[\pi]$$

if and only if there exists functions

$$\psi_1 : \times_{i\in K} \mathbf{X}_i \longrightarrow [0, +\infty)$$
$$\psi_2 : \times_{i\in L} \mathbf{X}_i \longrightarrow [0, +\infty)$$

such that

$$\pi(x_{K\cup L}) = \psi_1(x_K)\psi_2(x_L).$$

Proof. To make this proof (and the following two ones) transparent, we will keep the notation as simple as possible. Therefore within this proof x will always be an element of $\mathbf{X}_{K\setminus L}$, $y \in \mathbf{X}_{K\cap L}$ and $z \in \mathbf{X}_{L\setminus K}$. Therefore, for example, computation of $\pi(x_{K\cap L})$ from $\pi(x_K)$ is presented in this notation simply as

$$\pi(y) = \sum_x \pi(x, y).$$

If $X_{K\setminus L} \perp\!\!\!\perp X_{L\setminus K} | X_{K\cap L}[\pi]$ then, due to expression (2.2) from Remark 2.1.1, it is possible to define

$$\psi_1(x, y) = \pi(x, y)$$
$$\psi_2(y, z) = \pi(z|y).$$

Therefore, to prove the assertion it is enough to show that the existence of the functions ψ_1 and ψ_2 guarantees the respective conditional independence.

In the definition of the conditional independence $X_{K\setminus L} \perp\!\!\!\perp X_{L\setminus K} | X_{K\cap L}[\pi]$, three marginal distributions of $\pi(x_{K\cup L}) = \pi(x, y, z)$ occur; namely: $\pi(y)$, $\pi(x, y)$ and $\pi(y, z)$. Let us compute them.

$$\begin{aligned}\pi(y) &= \sum_x \sum_z \pi(x, y, z) = \sum_x \sum_z \psi_1(x, y)\psi_2(y, z)\\ &= \sum_x \psi_1(x, y) \sum_z \psi_2(y, z) = \left(\sum_x \psi_1(x, y)\right)\left(\sum_z \psi_2(y, z)\right).\end{aligned}$$

Analogously

$$\pi(x,y)=\sum_z \pi(x,y,z)=\sum_z \psi_1(x,y)\psi_2(y,z)=\psi_1(x,y)\sum_z \psi_2(y,z),$$

and

$$\pi(y,z)=\sum_x \pi(x,y,z)=\sum_x \psi_1(x,y)\psi_2(y,z)=\psi_2(y,z)\sum_x \psi_1(x,y).$$

Therefore

$$\begin{aligned}\pi(x,y,z)\pi(y) &= \psi_1(x,y)\psi_2(y,z)\left(\sum_x \psi_1(x,y)\right)\left(\sum_z \psi_2(y,z)\right)\\ &= \left(\psi_1(x,y)\sum_z \psi_2(y,z)\right)\left(\psi_2(y,z)\sum_x \psi_1(x,y)\right)\\ &= \pi(x,y)\pi(y,z),\end{aligned}$$

which is exactly the definition of the required conditional independence. □

Remark 2.1.4 Let us stress that the preceding assertion holds even for $X_{K\cap L}=\emptyset$. In this case, namely, $\pi(y)=1$ and $\pi(x,y)=\pi(x), \pi(y,z)=\pi(z)$, and, analogously $\psi_1(x,y)=\psi_1(x), \psi_2(y,z)=\psi_2(z)$. ○

The following property is usually called *block independence lemma.*

Lemma 2.1.2 *Let $L_1,L_2,L_3,L_4\subset N$ be disjoint and $L_1\neq\emptyset$, $L_2\neq\emptyset$, $L_3\neq\emptyset$. Then for any probability distribution $\pi(x_{L_1\cup L_2\cup L_3\cup L_4})$ the following two expressions are equivalent*

(A) $X_{L_1} \perp\!\!\!\perp X_{L_2\cup L_3}|X_{L_4}[\pi]$,

(B) $X_{L_1} \perp\!\!\!\perp X_{L_3}|X_{L_4}[\pi]$ *and* $X_{L_1} \perp\!\!\!\perp X_{L_2}|X_{L_3\cup L_4}[\pi]$.

Proof. In analogy to the previous proof, here we will consider x to be an element of $\mathbf{X}_{L_1}$, $y\in\mathbf{X}_{L_2}$, $z\in\mathbf{X}_{L_3}$ and $w\in\mathbf{X}_{L_4}$. Therefore, we are proving equivalence of the expressions

(A) $\pi(x,y,z,w)\pi(w)=\pi(x,w)\pi(y,z,w)$,

(B) $\pi(x,z,w)\pi(w) = \pi(x,w)\pi(z,w)$ &
$\pi(x,y,z,w)\pi(z,w) = \pi(x,z,w)\pi(y,z,w)$.

(A) $\Rightarrow$ **(B)** - Let us compute $\pi(x,z,w)\pi(w)$ under the assumption that independence (A) holds, i.e. that

$$\pi(x,y,z,w)\pi(w) = \pi(x,w)\pi(y,z,w).$$

Using simple marginalization we get

$$\begin{aligned}
\pi(x,z,w)\pi(w) &= \left(\sum_y \pi(x,y,z,w)\right)\pi(w) \\
&= \sum_y \pi(x,y,z,w)\pi(w) = \sum_y \pi(x,w)\pi(y,z,w) \\
&= \pi(x,w)\sum_y \pi(y,z,w) = \pi(x,w)\pi(z,w)\,,
\end{aligned}$$

which corresponds to $X_{L_1} \perp\!\!\!\perp X_{L_3}|X_{L_4}$.

To prove the second part of **(B)**, we have to show that

$$\pi(x,y,z,w)\pi(z,w) = \pi(x,z,w)\pi(y,z,w).$$

Since this equality holds for all

$$(x,y,z,w) \in \mathbf{X}_{L_1\cup L_2\cup L_3\cup L_4}$$

for which

$$\pi(w) = 0$$

(in this case both sides of the equality are equal to 0), it is enough to show that the required equality holds also for the other (x,y,z,w), for which it is equivalent to

$$\pi(x,y,z,w)\pi(z,w)\pi(w) = \pi(x,z,w)\pi(y,z,w)\pi(w).$$

To prove this, we need only to apply first the assumed independence $X_{L_1} \perp\!\!\!\perp X_{L_2\cup L_3}|X_{L_4}$ and then the just proven independence $X_{L_1} \perp\!\!\!\perp X_{L_3}|X_{L_4}$, so that we obtain

$$\begin{aligned}
\pi(x,y,z,w)\pi(z,w)\pi(w) &= \pi(x,w)\pi(y,z,w)\pi(z,w) \\
&= \pi(x,z,w)\pi(y,z,w)\pi(w)\,.
\end{aligned}$$

(B) $\Rightarrow$ **(A)** - Now, our goal is to show that under the conditional independence relations from **(B)** the equality

$$\pi(x,y,z,w)\pi(w) = \pi(x,w)\pi(y,z,w)$$

holds true for all quadruples

$$(x,y,z,w) \in \mathbf{X}_{L_1 \cup L_2 \cup L_3 \cup L_4}.$$

For those quadruples (x,y,z,w) for which

$$\pi(z,w) = 0$$

this equality holds because both sides are equal to 0, therefore we can prove the equality only for the remaining combinations of values (x,y,z,w). For them, it is equivalent to

$$\pi(x,y,z,w)\pi(w)\pi(z,w) = \pi(x,w)\pi(y,z,w)\pi(z,w)\,.$$

In fact, assuming **(B)** amounts to assume the following two equalities:

$$\begin{aligned} \pi(x,z,w)\pi(w) &= \pi(x,w)\pi(z,w) \\ \pi(x,y,z,w)\pi(z,w) &= \pi(x,z,w)\pi(y,z,w)\,, \end{aligned}$$

from which it follows

$$\begin{aligned} \pi(x,y,z,w)\pi(w)\pi(z,w) &= \pi(x,z,w)\pi(y,z,w)\pi(w) \\ &= \pi(x,w)\pi(z,w)\pi(y,z,w)\,. \end{aligned}$$ □

The following assertion is an analogy of the previous one but it holds only for strictly positive distributions. Therefore it is sometimes called either *symmetric block independence lemma*, or, *block independence lemma for positive distributions*.

Lemma 2.1.3 *Let $L_1, L_2, L_3, L_4 \subset N$ be disjoint and $L_1 \neq \emptyset$, $L_2 \neq \emptyset$, $L_3 \neq \emptyset$. Then for any strictly positive probability distribution $\pi(x_{L_1 \cup L_2 \cup L_3 \cup L_4})$ the following two expressions are equivalent*

(A) $X_{L_1} \perp\!\!\!\perp X_{L_2 \cup L_3} | X_{L_4} [\pi]$,

(B) $X_{L_1} \perp\!\!\!\perp X_{L_3} | X_{L_2 \cup L_4} [\pi]$ *and* $X_{L_1} \perp\!\!\!\perp X_{L_2} | X_{L_3 \cup L_4} [\pi]$.

Proof. As in the previous proof we will again consider x to be an element of $\mathbf{X}_{L_1}$, $y \in \mathbf{X}_{L_2}$, $z \in \mathbf{X}_{L_3}$ and $w \in \mathbf{X}_{L_4}$. Therefore, we are proving equivalence of the expressions

(A) $\pi(x, y, z, w)\pi(w) = \pi(x, w)\pi(y, z, w)$,

(B) $\pi(x, y, z, w)\pi(y, w) = \pi(x, y, w)\pi(y, z, w)$ &
$\pi(x, y, z, w)\pi(z, w) = \pi(x, z, w)\pi(y, z, w)$.

The first half of the proof - **(A)** $\Rightarrow$ **(B)** - follows immediately from Lemma 2.1.2. Let us remark that for this part of the assertion the strict positivity of the distribution is not necessary.

(B) $\Rightarrow$ **(A)** - Due to Factorization rule (Lemma 2.1.1), the independence relations from **(B)** entail the existence of functions $\psi_1, \psi_2, \psi_3, \psi_4$ such that

$$\begin{aligned} \pi(x, y, z, w) &= \psi_1(x, y, w)\psi_2(y, z, w)\,, \\ \pi(x, y, z, w) &= \psi_3(x, z, w)\psi_4(y, z, w)\,, \end{aligned} \tag{2.3}$$

and therefore also

$$\pi(y, z, w) = \psi_2(y, z, w) \sum_x \psi_1(x, y, w),$$
$$\pi(y, z, w) = \psi_4(y, z, w) \sum_x \psi_3(x, z, w).$$

Since the distribution π is assumed to be strictly positive, all functions ψ_1, ψ_2, ψ_3 and ψ_4 are also strictly positive and $\pi(x|y, z, w)$ can be expressed in the form of a ratio $\pi(x, y, z, w)/\pi(y, z, w)$ and therefore

$$\pi(x|y, z, w) = \frac{\psi_1(x, y, w)\psi_2(y, z, w)}{\psi_2(y, z, w) \sum_x \psi_1(x, y, w)} = \frac{\psi_1(x, y, w)}{\sum_x \psi_1(x, y, w)} = \frac{\psi_1(x, y, w)}{\varphi_1(y, w)}\,,$$

$$\pi(x|y, z, w) = \frac{\psi_3(x, z, w)\psi_4(y, z, w)}{\psi_4(y, z, w) \sum_x \psi_3(x, z, w)} = \frac{\psi_3(x, z, w)}{\sum_x \psi_3(x, z, w)} = \frac{\psi_3(x, z, w)}{\varphi_3(z, w)}\,.$$

From the first equality

$$\pi(x|y, z, w) = \frac{\psi_1(x, y, w)}{\varphi_1(y, w)}$$

we see that $\pi(x|y, z, w)$ does not depend on z, from the other equality

$$\pi(x|y, z, w) = \frac{\psi_3(x, z, w)}{\varphi_3(z, w)}$$

we can see that it does not depend on y (more precisely, $\pi(x|y,z,w) = \pi(x|y',z',w)$ for all quadruples $(x,y,z,w),(x,y',z',w) \in \mathbf{X}_{L_1\cup L_2\cup L_3\cup L_4}$). Denoting

$$\frac{\psi_1(x,y,w)}{\sum_x \psi_1(x,y,w)} = \frac{\psi_3(x,z,w)}{\sum_x \psi_3(x,z,w)} = \eta_1(x,w)$$

one gets

$$\psi_1(x,y,w) = \left(\sum_x \psi_1(x,y,w)\right)\eta_1(x,w) = \zeta(y,w)\eta_1(x,w).$$

Substituting this into (2.3) we get

$$\begin{aligned}\pi(x,y,z,w) &= \psi_1(x,y,w)\psi_2(y,z,w) = \zeta(y,w)\eta_1(x,w)\psi_2(y,z,w) \\ &= \eta_1(x,w)\eta_2(y,z,w)\,,\end{aligned}$$

which yields, due to Factorization Lemma 2.1.1, the conditional independence **(A)** $X_{L_1} \perp\!\!\!\perp X_{L_@\cup L_3}|X_{L_4}$. □

2.2 Extensions of distributions

Consider $K \subseteq L \subseteq N$ and a probability distribution $\pi(x_K)$. 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:

$$\Pi^{(L)}(\pi) = \left\{\kappa \in \Pi^{(L)} : \kappa(x_K) = \pi(x_K)\right\},$$

(recall that $\kappa(x_K)$ is the marginal distribution of κ for variables X_K). Having a system

$$\Xi = \{\pi_1(x_{K_1}), \pi_2(x_{K_2}), \ldots, \pi_n(x_{K_n})\}\,,$$

of oligodimensional distributions ($K_1 \cup \ldots \cup K_n \subseteq L$), the symbol $\Pi^{(L)}(\Xi)$ denotes the system of distributions that are extensions of all the distributions from Ξ:

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

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

Table 2.5: 3-dimensional distribution

π	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	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

Lemma 2.2.1 *For any system of oligodimensional distributions*

$$\Xi = \{\pi_1(x_{K_1}), \pi_2(x_{K_2}), \ldots, \pi_n(x_{K_n})\}$$

the set of its extensions $\Pi^{(L)}(\Xi)$ *is either empty or convex.*

Proof. To prove this assertion we have to show that if $\Pi^{(L)}(\Xi) \neq \emptyset$ than

$$\nu, \kappa \in \Pi^{(L)}(\Xi) \Longrightarrow (\alpha\nu + (1-\alpha)\kappa) \in \Pi^{(L)}(\Xi),$$

for any $\alpha \in [0,1]$. Consider arbitrary such distributions $\nu(x_L), \kappa(x_L) \in \Pi^{(L)}(\Xi)$ and $\pi_i \in \Xi$. Since both $\nu^{\downarrow K_i} = \pi_i$, and $\kappa^{\downarrow K_i} = \pi_i$, it is clear that also $\mu_\alpha^{\downarrow K_i} = \pi_i$ for any distribution

$$\mu_\alpha(x_L) = \alpha\nu(x_L) + (1-\alpha)\kappa(x_L).$$

As this must hold for all $i = 1, \ldots, n$ and therefore Ξ is a convex set. □

Example 2.2.1 *Consider a 3-dimensional distribution* $\pi(x_{\{1,2,3\}})$ *from Table 2.5. Its marginal distributions* $\pi^{\downarrow\{1,2\}}$, $\pi^{\downarrow\{1,3\}}$ *and* $\pi^{\downarrow\{2,3\}}$ *are in Table 2.6.*

Since $\pi(x_3) > 0$ *for all* $x_3 = 0, 1, 2$, *the conditional distributions* $\pi(x_1, x_2|x_3)$, $\pi(x_1|x_3)$ *and* $\pi(x_2|x_3)$ *are uniquely defined by*

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

and (for $j = 1, 2$*)*

$$\pi(x_j|x_3) = \frac{\pi(x_j, x_3)}{\pi(x_3)}.$$

Table 2.6: 2-dimensional marginal distributions

$\pi^{\downarrow\{1,2\}}$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0.3	0.2
$x_2 = 1$	0.2	0.3
$\pi^{\downarrow\{1,3\}}$	$x_1 = 0$	$x_1 = 1$
$x_3 = 0$	0.2	0.3
$x_3 = 1$	0.1	0.1
$x_3 = 2$	0.2	0.1
$\pi^{\downarrow\{2,3\}}$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	0.3	0.2
$x_3 = 1$	0.0	0.2
$x_3 = 2$	0.2	0.1

Therefore, if we want to find out whether the variables X_1 and X_2 are conditional independent given X_3, it is enough to check the equality

$$\pi(x_1, x_2|x_3) = \pi(x_1|x_3)\pi(x_2|x_3)$$

for all 12 possible combinations of values $x_1, x_2 \in \{0, 1\}$ and $x_3 \in \{0, 1, 2\}$. In this way we are getting that $X_1 \not\perp\!\!\!\perp X_2|X_3[\pi]$ because (for example) for $x_1 = x_2 = x_3 = 0$

$$\pi(x_1 = 0, x_2 = 0|x_3 = 0) = \frac{1}{5},$$

and

$$\pi(x_1 = 0|x_3 = 0) = \frac{2}{5}, \quad \pi(x_2 = 0|x_3 = 0) = \frac{3}{5}.$$

In contrary to conditioning by single variable X_3, conditional distribution $\pi(x_1|x_2, x_3)$ is not defined uniquely; the definition is met by any distribution from Table 2.7 (for $\alpha \in [0, 1]$).

Now, let us raise a question what are the classes of extensions of the 2-dimensional distributions from Table 2.6. First, let us consider all three distributions $\pi^{\downarrow\{1,2\}}$, $\pi^{\downarrow\{1,3\}}$ and $\pi^{\downarrow\{2,3\}}$. The set of their extensions is, in

Table 2.7: Conditional distribution $\pi(x_1|x_2, x_3)$

	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$1/3$	$1/2$	$2/3$	$1/2$
$x_3 = 1$	α	$1/2$	$1-\alpha$	$1/2$
$x_3 = 2$	1	0	0	1

correspondence with the above used notation, denoted by the symbol

$$\Pi^{(\{1,2,3\})}(\{\pi^{\downarrow\{1,2\}}, \pi^{\downarrow\{1,3\}}, \pi^{\downarrow\{2,3\}}\}).$$

Since $\pi \in \Pi^{(\{1,2,3\})}(\{\pi^{\downarrow\{1,2\}}, \pi^{\downarrow\{1,3\}}, \pi^{\downarrow\{2,3\}}\})$, *it is clear that this set must be nonempty. Due to Lemma 2.2.1 this set must be also convex. The reader can verify that the set of extensions is the set of all distributions described in Table 2.8 for* $\beta \in [.1, .2]$.

Table 2.8: Set $\Pi^{(\{1,2,3\})}(\{\pi^{\downarrow\{1,2\}}, \pi^{\downarrow\{1,3\}}, \pi^{\downarrow\{2,3\}}\})$

	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	β	$0.2-\beta$	$0.3-\beta$	β
$x_3 = 1$	0	0.1	0	0.1
$x_3 = 2$	$0.3-\beta$	$\beta-0.1$	$\beta-0.1$	$0.2-\beta$

Considering extensions of only two distributions, let us say $\pi^{\downarrow\{1,2\}}$ *and* $\pi^{\downarrow\{2,3\}}$, *we are getting a wider class of distributions, since*

$$\Pi^{(\{1,2,3\})}(\{\pi^{\downarrow\{1,2\}}, \pi^{\downarrow\{1,3\}}, \pi^{\downarrow\{2,3\}}\}) = \bigcap_{i=1}^{3} \Pi^{(\{1,2,3\})}(\pi^{\downarrow\{i,(i \bmod 3)+1\}})$$
$$\subseteq \Pi^{(\{1,2,3\})}(\pi^{\downarrow\{1,2\}}) \cap \Pi^{(\{1,2,3\})}(\pi^{\downarrow\{2,3\}}) = \Pi^{(\{1,2,3\})}(\{\pi^{\downarrow\{1,2\}}, \pi^{\downarrow\{2,3\}}\}).$$

This class is described in Table 2.9 for $\gamma \in [.1, .3]$, $\delta \in [0, .2]$ *and* $\varepsilon \in [.1-\delta, .2-\delta]$. $\diamondsuit$

Table 2.9: Set $\Pi^{(\{1,2,3\})}(\{\pi^{\downarrow\{1,2\}}, \pi^{\downarrow\{2,3\}}\})$

	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	γ	δ	$0.3 - \gamma$	$0.2 - \delta$
$x_3 = 1$	0	ε	0	$0.2 - \varepsilon$
$x_3 = 2$	$0.3 - \gamma$	$0.2 - (\delta + \varepsilon)$	$\gamma - 0.1$	$(\delta + \varepsilon) - 0.1$

2.3 Information-theoretic notions

In the sequel, several notions characterizing probability distributions and their relationship will be used. The first is the well-known *Shannon entropy* defined (for $\pi \in \Pi^{(N)}$)

$$H(\pi) = - \sum_{x \in \mathbf{X}_N : \pi(x) > 0} \pi(x) \log \pi(x).$$

Remark 2.3.1 In this section we shall see a close relation of the Shannon entropy to (mutual) information. If we want to measure the information in *bits* then in the above definition must be the binary logarithm ($\log_2$), otherwise we can consider a logarithm of an arbitrary basis. We have only keep in mind that all the logarithms must be the same. ○

Recall that for two disjoint index sets $K, L \subset N$ one can also define a *conditional entropy* $H(\pi(x_K|x_L)$ using the expression:

$$H(\pi(x_K|x_L)) = - \sum_{x \in \mathbf{X}_{K \cup L} : \pi(x) > 0} \pi(x) \log \pi(x_K|x_L).$$

Notice that for $L = \emptyset$ $H(\pi(x_K|x_L) = H(\pi(x_K))$. It is well known that (conditional) Shannon entropy is always nonnegative and equals 0 only for "degenerate" distribution, ie., for distribution that equals 1 for one x. Therefore, it is quite natural to define also for $K = \emptyset$

$$H(\pi(x_K|x_L) = 0.$$

To compare two distributions defined for the same system of variables (i.e. $\pi, \kappa \in \Pi^N$) we will use *Kullback-Leibler divergence* (in literature called also I-divergence, or cross-entropy). It is in fact a relative entropy of the first distribution with respect to the other:

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

where symbol $\pi \ll \kappa$ denotes the fact that κ *dominates* π. Dominance (or absolute continuity) is an important concept that will be used also when introducing other basic notions. In the considered finite case, this property can be defined by the following condition

$$\pi \ll \kappa \quad \Longleftrightarrow \quad \forall x \in \mathbf{X} \quad (\kappa(x) = 0 \Longrightarrow \pi(x) = 0)\,.$$

Let us go back to the Kullback-Leibler divergence. The reader can immediately see that if $\pi = \kappa$ then $Div(\pi\|\kappa) = 0$. It is a well-known property of Kullback-Leibler divergence (and not too difficult to be proven) that its value is always non-negative and equals 0 if and only if $\pi = \kappa$. This is the reason why we shall use it to "measure a distance" between two probability distributions defined for the same set of variables. But we will have to keep in mind that this divergence is not a distance in a mathematical sense, because it is not symmetric, i.e., generally $Div(\pi\|\kappa) \neq Div(\kappa\|\pi)$. For a trivial example of this inequality consider two distributions π and κ of 3-valued variable X from Table 2.10. As the reader can easily verify, when considering a binary logarithm, we get $Div(\pi\|\kappa) = 1$ and $Div(\kappa\|\pi) = +\infty$.

Table 2.10: Distributions π and κ

	0	1	2
π	0.5	0	0.5
κ	0.25	0.5	0.25

One of the fundamental notions of information theory is a *mutual information*. Having a distribution $\pi(x_N)$ and two disjoint subsets $K, L \subset N$, it expresses how much one group of variables X_K influences the other one – X_L. It is defined

$$MI_\pi(X_K; X_L) = \sum_{x\in\mathbf{X}_{K\cup L}:\pi(x)>0} \pi(x)\log\frac{\pi(x)}{\pi(x_K)\pi(x_L)},$$

and equals 0 if and only if variables X_K are independent with variables X_L under the distribution π. Otherwise, the mutual information $MI_\pi(X_K; X_L)$ is always positive.

For disjoint subsets $K, L, M \subset N$ *conditional mutual information* is defined by the formula

$$MI_\pi(X_K; X_L|X_M) = \sum_{x \in \mathbf{X}_{K\cup L\cup M}:\pi(x)>0} \pi(x) \log \frac{\pi(x)\pi(x_M)}{\pi(x_{K\cup M})\pi(x_{L\cup M})}.$$

As the reader can see, analogously to Shannon entropy, mutual information is just a special case of a conditional mutual information for $M = \emptyset$. Therefore, it is not surprising that also conditional mutual information is always nonnegative and equals zero if and only if the respective groups of variables are conditionally independent:

$$MI_\pi(X_K; X_L|X_M) = 0 \iff X_K \perp\!\!\!\perp X_L|X_M[\pi].$$

The last notion, which will be of great importance, but which is not as famous as Shannon entropy or mutual information, is an *informational content*[2] of a distribution defined (for $\pi \in \Pi^{(N)}$):

$$IC(\pi) = \sum_{x \in \mathbf{X}_N:\pi(x)>0} \pi(x) \log \frac{\pi(x)}{\prod\limits_{j\in N} \pi(x_j)},$$

and its conditional version (for disjoint $K, L \subset N$)

$$IC(\pi(x_K|x_L)) = \sum_{x \in \mathbf{X}_{K\cup L}:\pi(x)>0} \pi(x) \log \frac{\pi(x_K|x_L)}{\prod\limits_{j\in K} \pi(x_j)}.$$

Notice that $IC(\pi)$ is nothing but a Kullback-Leibler divergence of two distributions: $\pi(x_N)$ and $\prod_{j\in N} \pi(x_j)$. Since $\pi(x) \ll \prod\limits_{j\in N} \pi(x_j)$, this value is always finite and equals 0 if and only if $\pi(x) = \prod\limits_{j\in N} \pi(x_j)$. In fact, this value expresses how much individual variables are dependent under the distribution π. Therefore the higher this value, the more dependent the variables, and consequently, the greater amount of information carried by the distribution.

One can also immediately see that for a 2-dimensional distribution $\pi(x_1, x_2)$

$$IC(\pi) = MI_\pi(X_1; X_2).$$

[2]Some authors (Milan Studený) call this notion *multiinformation*.

Let us present a couple of expressions describing relationship among the introduced notions (all of the proofs can be found in any textbook on Information Theory like the classical textbook by Gallager [9], however we recommend the reader to prove all these formulae as an exercise).

Lemma 2.3.1 *Let $K, L, M \subseteq N$ be disjoint. For any probability distribution $\pi(x_N)$ the following expressions hold true.*

1. $0 \leq H(\pi(x_K)) \leq H(\pi(x_{K \cup L}))$;
2. $0 \leq MI_\pi(X_K; X_L) \leq \min(H(\pi(x_K)), H(\pi(x_L)))$;
3. $H(\pi(x_K | x_L)) = H(\pi(x_{K \cup L})) - H(\pi(x_L))$;
4. $MI_\pi(X_K; X_L) = H(\pi(x_K)) + H(\pi(x_L)) - H(\pi(x_{K \cup L}))$
   $= H(\pi(x_K)) - H(\pi(x_K | x_L))$;
5. $MI_\pi(X_K; X_L | X_M) = H(\pi(x_K | x_M)) + H(\pi(x_L | x_M)) - H(\pi(x_{K \cup L} | x_M))$
   $= H(\pi(x_K | x_M)) - H(\pi(x_K | x_{L \cup M}))$;
6. $MI_\pi(X_K; X_{L \cup M}) = MI_\pi(X_K; X_M) + MI_\pi(X_K; X_L | X_M)$;
7. $IC(\pi(x_K)) = \sum_{i \in K} H(\pi(x_i)) - H(\pi(x_K))$;
8. $IC(\pi(x_{K \cup L})) = IC(\pi(x_K)) + IC(\pi(x_L)) + MI_\pi(X_K; X_L)$;
9. $IC(\pi(x_K | x_L)) = IC(\pi(x_{K \cup L})) - IC(\pi(x_L))$;
10. $X_K \perp\!\!\!\perp X_L | X_M [\pi] \iff MI_\pi(X_K; X_L | X_M) = 0$;
11. $X_K \perp\!\!\!\perp X_L | X_M [\pi] \implies MI_\pi(X_K; X_M) \geq MI_\pi(X_K; X_L)$;
12. $X_K \perp\!\!\!\perp X_L | X_M [\pi] \iff MI_\pi(X_K; X_{L \cup M}) = MI_\pi(X_K; X_M)$.

Chapter 3

Operators of composition

3.1 Definition of operators

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:

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

Example 3.1.1 *Let us illustrate this formula by computing*

$$\pi^{\downarrow\{1,2\}}(x_1, x_2) \triangleright \pi^{\downarrow\{2,3\}}(x_2, x_3) = \frac{\pi^{\downarrow\{1,2\}}(x_1, x_2)\pi^{\downarrow\{2,3\}}(x_2, x_3)}{\pi^{\downarrow\{2,3\})}(x_2)},$$

where the 2*-dimensional distributions involved are the marginal distributions from Example 2.2.1. The computation results in a distribution presented in Table 3.1. One can immediately see that this distribution belongs to* $\Pi^{(\{1,2,3\})}(\{\pi^{\downarrow\{1,2\}}, \pi^{\downarrow\{2,3\}}\})$ *(it is a distribution from Table 2.9 with* $\gamma = .18$ *and* $\delta = .08$*) but not to* $\Pi^{(\{1,2,3\})}(\{\pi^{\downarrow\{1,2\}}, \pi^{\downarrow\{1,3\}}, \pi^{\downarrow\{2,3\}}\})$. $\diamondsuit$

Consider two probability distributions $\pi(x_K)$ and $\kappa(x_L)$, for which we want to define their composition. At this moment 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 (under a rather special condition) that the composition $\pi \triangleright \kappa$ does not exist. In case that

Table 3.1: Composed 3-dimensional distribution $\pi^{(\{1,2\})} \triangleright \pi^{(\{2,3\})}$

	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	0.18	0.08	0.12	0.12
$x_3 = 1$	0	0.08	0	0.12
$x_3 = 2$	0.12	0.04	0.08	0.06

$\pi(x_{L\cap K}) \ll \kappa(x_{L\cap K})$, the *right composition* of these two distributions is given by the formula

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

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

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

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

Let us stress that in case $\pi^{\downarrow L\cap K} \not\ll \kappa^{\downarrow L\cap K}$ the expression $\pi \triangleright \kappa$ remains undefined.

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

Definition 3.1.1 For arbitrary two distributions $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$ their *right composition* is given by the following formula

$$\pi(x_K) \triangleright \kappa(x_L) = \begin{cases} \frac{\pi(x_K)\kappa(x_L)}{\kappa(x_{K\cap L})} & \text{if } \pi(x_{K\cap L}) \ll \kappa(x_{K\cap L}), \\ \text{undefined} & \text{otherwise.} \end{cases}$$

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

Lemma 3.1.1 *Let π, κ be probability distributions from $\Pi^{(K)}, \Pi^{(L)}$, respectively. If $\pi(x_{L\cap K}) \ll \kappa(x_{L\cap K})$ (i.e., if $\pi(x_K) \triangleright \kappa(x_L)$ is defined) then*

$\pi(x_K) \triangleright \kappa(x_L)$ 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:

$$(\pi \triangleright \kappa)(x_K) = \pi(x_K)$$

for all $x_K \in \mathbf{X}_K$.

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

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

Therefore, to prove the whole assertion it is enough to show the second part, that is to show that

$$(\pi \triangleright \kappa)(x_K) = \sum_{x_{L \setminus K} \in \mathbf{X}_{L \setminus K}} (\pi \triangleright \kappa)(x) = \pi(x_K),$$

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

$$\begin{aligned}
\sum_{x_{L \setminus K} \in \mathbf{X}_{L \setminus K}} (\pi \triangleright \kappa)(x) &= \sum_{x_{L \setminus K} \in \mathbf{X}_{L \setminus K}} \frac{\pi(x_K)\kappa(x_L)}{\kappa(x_{K \cap L})} \\
&= \sum_{x_{L \setminus K} \in \mathbf{X}_{L \setminus K}} \frac{\pi(x_K)\kappa(x_{K \cap L}, x_{L \setminus K})}{\kappa(x_{K \cap L})} \\
&= \sum_{x_{L \setminus K} \in \mathbf{X}_{L \setminus K}} \frac{\pi(x_K)\kappa(x_{K \cap L})\kappa(x_{L \setminus K} | x_{K \cap L})}{\kappa(x_{K \cap L})} \\
&= \pi(x_K)\frac{\kappa(x_{K \cap L})}{\kappa(x_{K \cap L})} \sum_{x_{L \setminus K} \in \mathbf{X}_{L \setminus K}} \kappa(x_{L \setminus K} | x_{K \cap L}).
\end{aligned}$$

In the last expression $\kappa(x_{L \setminus K} | x_{K \cap L})$ is a conditional distribution and therefore

$$\sum_{x_{L \setminus K} \in \mathbf{X}_{L \setminus K}} \kappa(x_{L \setminus K} | x_{K \cap L}) = 1.$$

Moreover, due to the assumption

$$\pi(x_{L \cap K}) \ll \kappa(x_{L \cap K}),$$

if $\kappa(x_{L \cap K}) = 0$ then also $\pi(x_{L \cap K}) = 0$, and we defined $\pi \triangleright \kappa = 0$ in these points. Therefore

$$\pi(x_K)\frac{\kappa(x_{K \cap L})}{\kappa(x_{K \cap L})} = \pi(x_K)$$

for all $x_K \in \mathbf{X}_K$, which finishes the proof. $\square$

Example 3.1.2 *Let us illustrate difficulties, which can occur when $\pi^{\downarrow L\cap K} \not\ll \kappa^{\downarrow K\cap L}$ by a simple example.*

Consider the distributions $\pi(x_1, x_2)$ and $\kappa(x_2, x_3)$ given in Tables 3.2 and 3.3, for which $\pi(x_2 = 0) > 0$ and $\kappa(x_2 = 0) = 0$.

Table 3.2: Probability distribution π

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

Table 3.3: Probability distribution κ

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

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

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

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 $\pi \triangleright \nu$ is well defined whereas $\nu \triangleright \pi$ remains undefined. For this, consider the distribution ν from Table 3.4 and π from Table 3.2. Computation of $\pi \triangleright \nu$ and $\nu \triangleright \pi$ is in Table 3.5. $\diamondsuit$

Table 3.4: 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 3.5: 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	$0 \cdot \frac{1}{2} = 0$	$\frac{1}{4} \cdot \frac{0}{0} = ?$

Let us state here also a property of a dominance that will help us to prove some of the assertions in the following chapters.

Lemma 3.1.2 *Let $M \subseteq K \cap L$. For arbitrary two distributions $\pi(x_K)$ and $\kappa(x_L)$ the following two expressions are equivalent:*

(A) $\pi^{\downarrow K\cap L} \not\ll \kappa^{\downarrow K\cap L}$;

(B) $\pi^{\downarrow M} \not\ll \kappa^{\downarrow M}$ *or* $\pi \not\ll \pi^{\downarrow (K\setminus L)\cup M} \triangleright \kappa^{\downarrow K\cap L}$.

Proof. Let us recall the meaning of dominance. Condition **(A)** is true *iff* there exists $x \in \mathbf{X}_{K\cap L}$ for which $\kappa(x) = 0$ and simultaneously $\pi(x) > 0$. Since $M \subseteq K \cap L$,

$$\kappa(x) = \kappa(x_M)\kappa(x_{(K\cap L)\setminus M}|x_M),$$

and therefore **(A)** holds true *iff* there exists $x \in \mathbf{X}_{K\cap L}$ such that $\pi(x) > 0$ and simultaneously either

$$\kappa(x_M) = 0,$$

or

$$\kappa(x_{(K\cap L)\setminus M}|x_M) = 0.$$

The former case occurs if $\pi^{\downarrow M} \not\ll \kappa^{\downarrow M}$ is true, whereas the latter case is equivalent to $\pi \not\ll \pi^{\downarrow (K\setminus L)\cup M} \triangleright \kappa^{\downarrow K\cap L}$. $\square$

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

Definition 3.1.2 For arbitrary two distributions $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$ their *left composition* is given by the following formula

$$\pi(x_K) \triangleleft \kappa(x_L) = \begin{cases} \frac{\pi(x_K)\kappa(x_L)}{\pi(x_{K\cap L})} & \text{if } \kappa(x_{K\cap L}) \ll \pi(x_{K\cap L}), \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Let us repeat that either of the expressions $\pi(x_K) \triangleright \kappa(x_L), \pi(x_K) \triangleleft \kappa(x_L)$, if defined, is a probability distribution of variables $X_{K\cup L}$. Properties of these composed distributions will be discussed in the next section.

3.2 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 others are rather technical and will be used to simplify proofs in subsequent sections.

Lemma 3.2.1 *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*

$$\pi \triangleright \kappa \ll \kappa$$

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

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

Proof. The assertion directly follows from the definition of the operator $\triangleright$ which can be for the current situation written

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

From this formula it follows evidently that for any $x \in \mathbf{X}_L$

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

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

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

$$(\pi \triangleright \kappa)(x) = \frac{\pi(x_K)\kappa(x)}{\kappa(x_K)} = 0.$$

That means that either $\pi(x_K) = 0$ or $\kappa(x) = 0$ (or both). If $\pi(x_K) = 0$ then $\nu(x_K) = 0$ as $\nu^{\downarrow K}$ equals π since $\nu \in \Pi^L(\pi)$. 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

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

The opposite implication

$$\nu \ll \pi \triangleright \kappa \Longrightarrow \nu \ll \kappa$$

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

$$\nu \ll \pi \triangleright \kappa \quad \& \quad \pi \triangleright \kappa \ll \kappa \Longrightarrow \nu \ll \kappa.$$ □

Definition 3.2.1 We shall say that distributions $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$ are *consistent* if

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

Remark 3.2.1 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 get the following trivial assertion.

Lemma 3.2.2 *Let $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$. If π and κ are consistent 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 also the reverse implication holds true:

$$\pi \triangleright \kappa = \pi \triangleleft \kappa \quad \Longrightarrow \quad \pi \text{ and } \kappa \text{ are consistent.}$$

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 assume also that either $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$ or $\kappa^{\downarrow K\cap L} \ll \pi^{\downarrow K\cap L}$, it means that either $\pi \triangleright \kappa$ or $\pi \triangleleft \kappa$ is defined, and because these compositions equal each other, both of them must be defined. Using twice Lemma 3.1.1, assumed equivalence and definition of the operators one gets

$$\begin{aligned} \pi(x_{K\cap L}) &= (\pi \triangleright \kappa)(x_{K\cap L}) = (\pi \triangleleft \kappa)(x_{K\cap L}) \\ &= (\kappa \triangleright \pi)(x_{K\cap L}) = \kappa(x_{K\cap L}). \end{aligned}$$

□

As said at the beginning of this chapter, application of the operator of composition introduces conditional independence among the variables. What is the exact meaning of this statement can be seen from the following simple but important assertion.

Lemma 3.2.3 *Let $\nu(x_{K\cup L}) = \pi(x_K) \triangleright \kappa(x_L)$ be defined. 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 $\nu = \pi \triangleright \kappa$

$$\nu(x_{K\cup L})\nu(x_{K\cap L}) = \nu(x_K)\nu(x_L) \tag{3.1}$$

for all $x \in \mathbf{X}_{K\cup L}$. If for $x \in \mathbf{X}_{K\cup L}$ $\kappa(x_{K\cap L}) = 0$, then also $\pi(x_{K\cap L}) = 0$ (because ν is defined only when $\pi(x_{K\cap L}) \ll \kappa(x_{K\cap L})$), and therefore also $\nu(x_{K\cap L}) = 0$ (and thus $\nu(x_K) = \nu(x_L) = \nu(x_{K\cup L}) = 0$, too). From this we immediately get that equality (3.1) holds because both its sides equal 0.

Consider now $x \in \mathbf{X}_{K\cup L}$, for which $\kappa(x_{K\cap L}) > 0$. Lemma 3.1.1 says that $\nu(x_K) = \pi(x_K)$. Let us compute $\nu(x_L)$:

$$\begin{aligned}\nu(x_L) &= \sum_{x_{K\setminus L}\in\mathbf{X}_{K\setminus L}} \frac{\pi(x_{K\setminus L}, x_{K\cap L})\kappa(x_L)}{\kappa(x_{K\cap L})} \\ &= \frac{\pi(x_{K\cap L})\kappa(x_L)}{\kappa(x_{K\cap L})} \sum_{x_{K\setminus L}\in\mathbf{X}_{K\setminus L}} \pi(x_{K\setminus L}|x_{K\cap L}),\end{aligned}$$

where

$$\sum_{x_{K\setminus L}\in\mathbf{X}_{K\setminus L}} \pi(x_{K\setminus L}|x_{K\cap L}) = 1.$$

Therefore

$$\nu(x_K)\nu(x_L) = \pi(x_K)\frac{\pi(x_{K\cap L})\kappa(x_L)}{\kappa(x_{K\cap L})} = \frac{\pi(x_K)\kappa(x_L)}{\kappa(x_{K\cap L})}\pi(x_{K\cap L}),$$

where

$$\frac{\pi(x_K)\kappa(x_L)}{\kappa(x_{K\cap L})} = \nu(x_{K\cup L})$$

from the definition of operator $\triangleright$, and $\pi(x_{K\cap L}) = \nu(x_{K\cap L})$ due to Lemma 3.1.1. □

Remark 3.2.2 In the following proof we shall use a standard trick, which will be later repeated quite often. Let $M \subset K$ and $\nu \in \Pi^{(K)}$. Let us compute

$$\begin{aligned}\sum_{\substack{x\,\in\,\mathbf{X}_K\\ \nu(x)>0}} \nu(x)\log\nu(x_M) &= \sum_{\substack{x\,\in\,\mathbf{X}_K\\ \nu(x)>0}} \nu(x_M)\nu(x_{K\setminus M}|x_M)\log\nu(x_M) \\ &= \sum_{\substack{x\,\in\,\mathbf{X}_M\\ \nu(x)>0}} \nu(x_M)\log\nu(x_M) \sum_{\substack{y\,\in\,\mathbf{X}_{K\setminus M}\\ \nu(y|x)>0}} \nu(y|x) = H(\nu^{\downarrow M})\end{aligned}$$

because

$$\sum_{\substack{y\,\in\,\mathbf{X}_{K\setminus M}\\ \nu(y|x)>0}} \nu(y|x) = 1.$$

○

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, as expressed in the following assertion.

Theorem 3.2.1 *If probability distributions $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$ are consistent then*

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

and

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

Proof. The first part of the proof is trivial:

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

because both π and κ are marginal to $\pi \triangleright \kappa$ (this holds due to consistency of π and κ, and Lemmata 3.2.2 and 3.1.1).

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

$$\begin{aligned}
H(\nu) &= -\sum_{\substack{x \in \mathbf{X}_{K\cup L}\\ \nu(x)>0}} \nu(x) \log \nu(x) \\
&= -\sum_{\substack{x \in \mathbf{X}_{K\cup L}\\ \nu(x)>0}} \nu(x) \log \frac{\nu(x_K)\nu(x_L)\nu(x)\nu(x_{K\cap L})}{\nu(x_{K\cap L})\nu(x_K)\nu(x_L)}
\end{aligned}$$

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

which concludes the proof because the conditional mutual information is always nonnegative (and equals 0 if and only if $\nu = \pi \triangleright \kappa$). □

Remark 3.2.3 First notice that $H(\kappa) - H(\kappa^{\downarrow K \cap L})$ is a conditional entropy $H(\kappa(x_L | x_{K \cap L}))$ (see property 3 on page 33). In this context the reader should realize that the equality

$$H(\pi(x_K) \triangleright \kappa(x_L)) = H(\pi(x_K)) + H(\kappa(x_L | x_{K \cap L}))$$

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

Example 3.2.1 *Consider distribution κ from Table 3.6.*

Table 3.6: Probability distribution κ

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

Taking binary logarithm for computation of Shannon entropy we easily get

$$H(\kappa(x_2|x_1)) = H(\kappa(x_1,x_2)) - H(\kappa(x_1)) = \frac{3}{2} - 1 = \frac{1}{2}.$$

Let us compute entropy of the distribution

$$\pi(x_1) \triangleright \kappa(x_1,x_2),$$

for $\pi(x_1 = 0) = 0.1$, $\pi(x_1 = 1) = 0.9$. *This composed distribution is in Table 3.7, and its entropy equals*

$$\begin{aligned}\sum_{x_1=0,1}\sum_{x_2=0,1} \pi(x_1)\kappa(x_2|x_1) &= 0\log_2 0 + .1\log_2 .1 + .45\log_2 .45 + .45\log_2 .45 \\ &= 1.369,\end{aligned}$$

which certainly differs from

$$H(\pi(x_1)) + H(\kappa(x_2|x_1)) = 0.469 + 0.5 = 0.969.$$

Table 3.7: Probability distribution $\pi \triangleright \kappa$

$\pi \triangleright \kappa$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0	.1
$x_2 = 1$	.45	.45

Similarly, when composing κ *with distribution* $\hat{\pi}$, *for which* $\hat{\pi}(x_1 = 0) = 0.9$, $\hat{\pi}(x_1 = 1) = 0.1$ *(evidently* $H(\hat{\pi}) = H(\pi)$*), we get the distribution* $\hat{\pi} \triangleright \kappa$ *(see Table 3.8), whose entropy equals*

$$\begin{aligned}\sum_{x_1=0,1}\sum_{x_1=0,1} \hat{\pi}(x_1)\kappa(x_2|x_1) &= 0\log_2 0 + .9\log_2 .9 + .05\log_2 .05 + .05\log_2 .05 \\ &= 0.569.\end{aligned}$$

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

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

Table 3.8: Probability distribution $\hat{\pi} \triangleright \kappa$

$\hat{\pi} \triangleright \kappa$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	0	.9
$x_2 = 1$	.05	.05

Table 3.9: Probability distribution $\hat{\kappa}$

$\hat{\kappa}$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	.1	.4
$x_2 = 1$	.4	.1

holds in case of inconsistent distributions. For example, for $\hat{\kappa}$ from Table 3.9 all distributions $\nu \triangleright \hat{\kappa}$ achieve the same value of entropy. This is because

$$H(\nu \triangleright \hat{\kappa}) = H(\nu) + \sum_{x_1=0,1} \nu(x_1) H(\hat{\kappa}(\cdot|x_1)),$$

and $H(\hat{\kappa}(\cdot|x_1 = 0)) = H(\hat{\kappa}(\cdot|x_1 = 1))$. $\Diamond$

Example 3.2.2 *In the proofs, we shall often compute a marginal distribution from a distribution defined as a composition of two (or several) oligodimensional distributions. Therefore, it is important to realize that generally for $M \subset K \cup L$*

$$(\pi \triangleright \kappa)^{\downarrow M} \neq \pi^{\downarrow K \cap M} \triangleright \kappa^{\downarrow L \cap M}. \tag{3.2}$$

To illustrate the situation when equality in formula (3.2) does not hold consider $\pi^{\downarrow\{1,2\}} \triangleright \pi^{\downarrow\{2,3\}}$ from Example 3.1.1 (see Table 3.1 on page 36) and its marginal distribution $(\pi^{\downarrow\{1,2\}} \triangleright \pi^{\downarrow\{2,3\}})^{\downarrow\{1,3\}}$, which is in Table 3.10. Examining this marginal distribution we see that variables X_1 and X_3 are not independent. Therefore

$$\begin{aligned}(\pi(x_1, x_2) \triangleright \pi(x_2, x_3))^{\downarrow\{1,3\}} &\neq (\pi(x_1, x_2))^{\downarrow\{1\}} \triangleright (\pi(x_2, x_3))^{\downarrow\{3\}} \\ &= \pi(x_1) \triangleright \pi(x_3) = \pi(x_1)\pi(x_3).\end{aligned}$$

$\Diamond$

Table 3.10: Marginal distribution $(\pi^{\downarrow\{1,2\}} \triangleright \pi^{\downarrow\{2,3\}})^{\downarrow\{1,3\}}$

$\pi^{\downarrow\{1,3\}}$	$x_1 = 0$	$x_1 = 1$
$x_3 = 0$	0.26	0.24
$x_3 = 1$	0.08	0.12
$x_3 = 2$	0.16	0.14

The following simple assertion presents a sufficient condition under which the equality in expression (3.2) holds.

Lemma 3.2.4 *Let $K, L, M \subseteq N$. If $K \cup L \supseteq M \supseteq K \cap L$ then for any probability distributions $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$*

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

Proof. Let us first mention that $\pi \triangleright \kappa$ is not defined only if $\pi^{\downarrow K\cap L} \not\ll \kappa^{\downarrow K\cap L}$. However, because of the assumption laid on M, $K \cap L = (K\cap M)\cap(L\cap M)$, and therefore it holds true if and only if $\pi^{(K\cap M)} \triangleright \kappa^{(L\cap M)}$ is not defined, too. Therefore, if one composition is not defined then neither the other composition is defined.

To prove the assertion in case that $\pi \triangleright \kappa$ is defined, let us first compute

$$\begin{aligned}(\pi \triangleright \kappa)^{\downarrow K\cup M} &= \sum_{x_{L\setminus M} \in \mathbf{X}_{L\setminus M}} \frac{\pi(x_K)\kappa(x_{L\cap M}, x_{L\setminus M})}{\kappa(x_{L\cap K})} \\ &= \frac{\pi(x_K)\kappa(x_{L\cap M})}{\kappa(x_{L\cap K})} \sum_{x_{L\setminus M} \in \mathbf{X}_{L\setminus M}} \kappa(x_{L\setminus M}|x_{L\cap M}) = \pi \triangleright \kappa^{\downarrow L\cap M}.\end{aligned}$$

Now we can compute the required marginal distribution

$$\begin{aligned}(\pi \triangleright \kappa)^{\downarrow M} &= ((\pi \triangleright \kappa)^{\downarrow K\cup M})^{\downarrow M} = (\pi \triangleright \kappa^{\downarrow L\cap M})^{\downarrow M} \\ &= \sum_{x_{K\setminus M} \in \mathbf{X}_{K\setminus M}} \frac{\pi(x_{K\cap M}, x_{K\setminus M})\kappa(x_{L\cap M})}{\kappa(x_{L\cap K})} \\ &= \frac{\pi(x_{K\cap M})\kappa(x_{L\cap M})}{\kappa(x_{L\cap K})} \sum_{x_{K\setminus M} \in \mathbf{X}_{K\setminus M}} \pi(x_{K\setminus M}|x_{K\cap M}) \\ &= \pi^{\downarrow K\cap M} \triangleright \kappa^{\downarrow L\cap M}.\end{aligned}$$

□

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 3.2.5 *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 3.2.6 *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*

$$\pi \triangleright \kappa = \pi \triangleright (\pi \triangleright \kappa)^{\downarrow M}. \tag{3.3}$$

Proof First notice that if $\pi \triangleright \kappa$ is not defined then also the right hand side of formula (3.3) is not defined. In opposite case, if $\pi \triangleright \kappa$ is defined, π and $\pi \triangleright \kappa$ are consistent (see Lemma 3.1.1) and therefore $\pi \triangleright (\pi \triangleright \kappa)^{\downarrow M}$ is also defined.

Therefore, consider only situations when $\pi \triangleright \kappa$ is defined. We have to prove that the required equality holds for all $x \in \mathbf{X}_{K \cup L}$. If for such x, $\pi(x_{K \cap M}) = 0$, equality (3.3) holds true, because in these points $\pi(x_K) = 0$ and therefore also both $\pi \triangleright \kappa = 0$ and $\pi \triangleright (\pi \triangleright \kappa)^{\downarrow M} = 0$. For all other $x \in \mathbf{X}_{K \cup L}$, for which $\pi(x_{K \cap M}) > 0$ the assertion follows from Lemma 3.2.4, the definition of the composition operator $\triangleright$ and Lemma 3.1.1 saying that $(\pi^{\downarrow K \cap M} \triangleright \kappa)^{\downarrow K \cap M} = \pi^{\downarrow K \cap M}$:

$$\begin{aligned} \pi \triangleright (\pi \triangleright \kappa)^{\downarrow M} &= \pi \triangleright \left(\pi^{\downarrow K \cap M} \triangleright \kappa\right) = \frac{\pi \left(\pi^{\downarrow K \cap M} \triangleright \kappa\right)}{\left(\pi^{\downarrow K \cap M} \triangleright \kappa\right)^{\downarrow K \cap M}} \\ &= \frac{\pi \left(\pi^{\downarrow K \cap M} \triangleright \kappa\right)}{\pi^{\downarrow K \cap M}} = \frac{\pi}{\pi^{\downarrow K \cap M}} \frac{\pi^{\downarrow K \cap M} \kappa}{\kappa^{\downarrow K \cap M \cap L}} = \frac{\pi \kappa}{\kappa^{\downarrow K \cap L}} = \pi \triangleright \kappa. \end{aligned}$$

□

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

Lemma 3.2.7 *Let* M *be such that* $K \cap L \subseteq M \subseteq L$*; then*

$$\pi(x_K) \triangleright \kappa(x_L) = (\pi(x_K) \triangleright \kappa(x_M)) \triangleright \kappa(x_L).$$

Proof. It is, again, trivial to show that left hand side of this equality is undefined if and only if also its right hand side is undefined. Therefore we will prove it under an assumption the expression is defined. We have to distinguish two situations. For $x \in \mathbf{X}_{K\cup L}$, for which $\kappa(x_{L\cap M}) = 0$, the property holds, because both sides of the equality in question equal 0. To prove it for $x \in \mathbf{X}_{K\cup L}$, for which $\kappa(x_{L\cap M}) > 0$, realize that, under the given assumptions, $M \cap K = K \cap L$ and $L \cap (K \cup M) = (L \cap K) \cup M = M$. Then the assertion follows immediately from the definition of the composition operator $\triangleright$:

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

□

3.3 I-geometry of composition operators

This section is based on the results of Imre Csiszár ([6]) and therefore we use also his terminology (including the term *I-geometry* in the section title).

Definition 3.3.1 Consider any $\pi \in \Pi^{(L)}$ and an arbitrary subset Θ of $\Pi^{(L)}$. Distribution

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

is called an *I-projection* of π into Θ, and, similarly,

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

is called a *reverse I-projection* of π into Θ.

Thus, according to this definition, both I-projection and reverse I-projection are distributions from $\Theta \subset \Pi^{(L)}$, which are, in a sense, closest to π. As a measure of distance we take the Kullback-Leibler divergence[1]

$$Div(\kappa\|\pi) = \sum_{x\in\mathbf{X}_L:\kappa(x)>0} \kappa(x) \log \frac{\kappa(x)}{\pi(x)}.$$

Generally, it may happen that for given π and Θ neither of the projections exists. However, we will always consider Θ to be a set of distributions with given marginal(s), which is always a convex compact set of

[1] Recall, that if $\kappa \not\ll \pi$ then $Div(\kappa\|\pi) = +\infty$ by definition.

distributions. For $\Theta = \Pi^{(L)}(\Xi)$ the existence of an I-projection (reverse I-projection) is guaranteed just by the existence of one $\nu \in \Theta$, for which $Div(\nu\|\pi)$ ($Div(\pi\|\nu)$, respectively) is finite. Instructions how to find these I-projections are given by the following two assertions.

Theorem 3.3.1 *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 $\Pi^{(L)}(\pi)$. Moreover,*

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

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

Proof. $\pi \triangleright \kappa \in \Pi^{(L)}(\pi)$ and since $\pi \triangleright \kappa \ll \kappa$ (this holds due to Lemma 3.2.1), $Div(\pi \triangleright \kappa\|\kappa)$ is finite. Therefore the I-projection ν^* of κ into $\Pi^{(L)}(\pi)$ must be dominated by κ (otherwise $Div(\nu^*\|\kappa) = +\infty$ and ν^* cannot be an I-projection of κ into $\Pi^{(L)}(\pi)$).

Consider any $\nu \in \Pi^{(L)}(\pi)$ that is dominated by κ. First, realize that because of Lemma 3.2.1 $\nu \ll \pi \triangleright \kappa$. Therefore one can compute

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

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

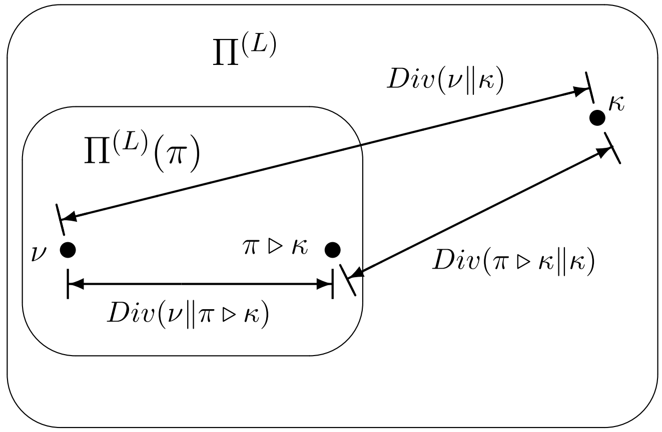


Figure 3.1: I-projection of κ into $\Pi^{(L)}(\pi)$

and thus

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

The equality

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

holds also when $\nu \not\ll \kappa$ because, according to Lemma 3.2.1, also $\nu \not\ll \pi \triangleright \kappa$ and therefore both $Div(\nu\|\kappa)$ and $Div(\nu\|\pi \triangleright \kappa)$ equal $+\infty$. $\square$

Theorem 3.3.2 *Let $K \subseteq L \subseteq N$. If probability distributions $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$ are such that $\pi \ll \kappa^{\downarrow K}$ and $\kappa \ll \pi \triangleright \kappa$, then $\pi \triangleright \kappa$ is the reverse I-projection of κ into $\Pi^{(L)}(\pi)$.*

Proof. The assumption $\kappa \ll \pi \triangleright \kappa$ guarantees that there exists at least one distribution ν from $\Pi^{(L)}(\pi)$ for which $Div(\kappa\|\nu) < +\infty$ (namely $\nu = \pi \triangleright \kappa$). Therefore we are sure that the reverse I-projection of κ into $\Pi^{(L)}(\pi)$ dominates κ. According to Lemma 3.2.1, κ dominates $\pi \triangleright \kappa$ and so does, due to transitivity of a dominance, also the considered reverse I-projection.

Consider any $\nu \in \Pi^{(L)}(\pi)$ such that $\pi \triangleright \kappa \ll \nu$. Then we can compute (the extension of the argument of the logarithm is possible because we assume that $\kappa \ll \pi \triangleright \kappa$):

$$\begin{aligned}
Div(\kappa\|\nu) &= \sum_{x\in\mathbf{X}_L:\kappa(x)>0} \kappa(x)\log\frac{\kappa(x)}{\nu(x)} \\
&= \sum_{x\in\mathbf{X}_L:\kappa(x)>0} \kappa(x)\log\left(\frac{\kappa(x)}{(\pi\triangleright\kappa)(x)}\frac{(\pi\triangleright\kappa)(x)}{\nu(x)}\right) \\
&= Div(\kappa\|\pi\triangleright\kappa) + \sum_{x\in\mathbf{X}_L:\kappa(x)>0} \kappa(x)\log\frac{(\pi\triangleright\kappa)(x)}{\nu(x)} \\
&= Div(\kappa\|\pi\triangleright\kappa) + \sum_{x\in\mathbf{X}_L:\kappa(x)>0} \kappa(x)\log\frac{\pi(x_K)\kappa(x)}{\kappa(x_K)\nu(x)} \\
&= Div(\kappa\|\pi\triangleright\kappa) \\
&\qquad + \sum_{x\in\mathbf{X}_L:\kappa(x)>0}\left(\kappa(x)\log\frac{\pi(x_K)\kappa(x_K)\kappa(x_{L\setminus K}|x_K)}{\kappa(x_K)\nu(x_K)\nu(x_{L\setminus K}|x_K)}\right) \\
&= Div(\kappa\|\pi\triangleright\kappa) + \sum_{\substack{z\in\mathbf{X}_K\\ \kappa^{(K)}(z)>0}}\ \sum_{\substack{y\in\mathbf{X}_{L\setminus K}\\ \kappa(y|z)>0}} \kappa^{\downarrow K}(z)\kappa(y|z)\log\frac{\kappa(y|z)}{\nu(y|z)} \\
&= Div(\kappa\|\pi\triangleright\kappa) + \sum_{\substack{z\in\mathbf{X}_K\\ \kappa^{(K)}(z)>0}} \kappa^{\downarrow K}(z) \sum_{\substack{y\in\mathbf{X}_{L\setminus K}\\ \kappa(y|z)>0}} \kappa(y|z)\log\frac{\kappa(y|z)}{\nu(y|z)} \\
&= Div(\kappa\|\pi\triangleright\kappa) + \sum_{\substack{z\in\mathbf{X}_K\\ \kappa^{(K)}(z)>0}} \kappa^{\downarrow K}(z) Div(\kappa(\cdot|z)\|\nu(\cdot|z)).
\end{aligned}$$

As $Div(\kappa(\cdot|z)\|\nu(\cdot|z))$ cannot be negative for any $z \in \mathbf{X}_K$, $Div(\kappa\|\nu)$ achieves its minimum for $\nu = \pi \triangleright \kappa$. □

3.4 Iterations of the operators of composition

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, from which we will also see the reason why we defined two operators $\triangleright$ and $\triangleleft$.

To make the formulae more lucid, let us make the following convention: if not specified otherwise by brackets, operators $\triangleright$ and $\triangleleft$ are always in expressions applied from left to right. It means that

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

and

$$\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \ldots \triangleleft \pi_{n-1} \triangleleft \pi_n = (\ldots((\pi_1 \triangleleft \pi_2) \triangleleft \pi_3) \triangleleft \ldots \triangleleft \pi_{n-1}) \triangleleft \pi_n.$$

This section will be for most of the readers rather technical. We shall present ten assertions describing elementary situations when the operators are applied (as a rule) twice. Nevertheless, we will see that even in these simple situations there arise quite interesting problems. This is because the operators are neither commutative nor associative. In the following example we shall show that generally

$$\begin{array}{llcl}
(\mathbf{a}) & \pi_1 \triangleright \pi_2 \triangleright \pi_3 & \neq & \pi_1 \triangleright (\pi_2 \triangleright \pi_3) = \pi_2 \triangleright \pi_3 \triangleleft \pi_1 = \pi_3 \triangleleft \pi_2 \triangleleft \pi_1, \\
(\mathbf{b}) & \pi_1 \triangleright \pi_2 \triangleright \pi_3 & \neq & \pi_1 \triangleright \pi_3 \triangleright \pi_2, \\
(\mathbf{c}) & \pi_1 \triangleright \pi_2 \triangleright \pi_3 & \neq & \pi_1 \triangleleft \pi_2 \triangleleft \pi_3, \\
(\mathbf{d}) & \pi_1 \triangleleft \pi_2 \triangleleft \pi_3 & \neq & \pi_1 \triangleleft \pi_3 \triangleleft \pi_2.
\end{array}$$

Nevertheless, let us keep in mind that in special situations the equality in these expressions may occur. For example, when all the distributions π_1, π_2, π_3 are uniform, then all the expressions result in a uniform distribution, too.

Example 3.4.1
(a) *Consider*

$$\pi_1(x_1) \triangleright \pi_2(x_2) \triangleright \pi_3(x_1, x_2) = \pi_1(x_1)\pi_2(x_2), \tag{3.4}$$

which evidently differs from

$$\pi_1(x_1) \triangleright (\pi_2(x_2) \triangleright \pi_3(x_1, x_2)) = \frac{\pi_1(x_1)(\pi_2(x_2)\pi_3(x_1|x_2))}{\sum_{y \in \mathbf{X}_1} \pi_2(x_2)\pi_3(y|x_2)}. \tag{3.5}$$

Namely, in (3.4), the variables X_1 and X_2 are independent:

$$X_1 \perp\!\!\!\perp X_2[\pi_1 \triangleright \pi_2 \triangleright \pi_3],$$

which need not be true, generally, for (3.5). To see it, take an example where both $\pi_1(x_1)$ *and* $\pi_2(x_2)$ *are uniform distributions and* $\pi_3(0,0) = \pi_3(1,1) = \frac{1}{2}$, $\pi_3(0,1) = \pi_3(1,0) = 0$. *In this case both the marginal distributions* $\pi_3(x_1)$ *and* $\pi_3(x_2)$ *are uniform and therefore* π_1 *and* π_3, *as well as* π_2 *and* π_3 *are consistent. Therefore (due to Lemma 3.2.2)*

$$\pi_1 \triangleright \pi_3 = \pi_1 \triangleleft \pi_3 = \pi_3,$$

and also

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

From this we get

$$\pi_1 \triangleright (\pi_2 \triangleleft \pi_3) = \pi_1 \triangleright \pi_3 = \pi_1 \triangleleft \pi_3 = \pi_3,$$

which obviously differ from $\pi_1 \triangleright \pi_2$, *which, as a product of two* 1*-dimensional uniform distributions, is a uniform distribution, too.*

(b) *To illustrate the second inequality consider three* 1*-dimensional distributions* $\pi_1(x_1)$, $\pi_2(x_2)$, $\pi_3(x_2)$, *such that* $\pi_2(x_2) \neq \pi_3(x_2)$. *Then*

$$\begin{aligned}\pi_1(x_1) \triangleright \pi_2(x_2) \triangleright \pi_3(x_2) &= \pi_1(x_1)\pi_2(x_2) \\ &\neq \pi_1(x_1)\pi_3(x_2) = \pi_1(x_1) \triangleright \pi_3(x_2) \triangleright \pi_2(x_2).\end{aligned}$$

(c) *Consider three distributions* $\pi_1(x)$, $\pi_2(x)$, $\pi_3(x)$, *for which* $\pi_1(x) \neq \pi_3(x)$. *Then*

$$\pi_1(x) \triangleright \pi_2(x) \triangleright \pi_3(x) = \pi_1(x) \neq \pi_3(x) = \pi_1(x) \triangleleft \pi_2(x) \triangleleft \pi_3(x).$$

(d) *Consider again three distributions* $\pi_1(x)$, $\pi_2(x)$, $\pi_3(x)$, *this time such that* $\pi_2(x) \neq \pi_3(x)$. *Then it is clear that*

$$\pi_1(x) \triangleleft \pi_2(x) \triangleleft \pi_3(x) = \pi_3(x) \neq \pi_2(x) = \pi_1(x) \triangleleft \pi_3(x) \triangleleft \pi_2(x). \qquad \diamondsuit$$

Now, let us present lemmata saying under which conditions some of these equalities hold. Whenever in this section we shall use probability distributions π_1, π_2, π_3, we shall assume that $\pi_i \in \Pi^{(K_i)}$ for $i = 1, 2, 3$.

Lemma 3.4.1 *If* $K_1 \supseteq (K_2 \cap K_3)$ *then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleright \pi_2. \tag{3.6}$$

Proof. First, let us show that the left hand side expression in (3.6) is not defined *iff* the right hand side of this formula is not defined. From the definition of the operators we know that $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is not defined *iff*

$$\pi_1^{\downarrow K_1 \cap K_2} \not\ll \pi_2^{\downarrow K_1 \cap K_2}$$

or

$$(\pi_1 \triangleright \pi_2)^{\downarrow (K_1 \cup K_2) \cap K_3} \not\ll \pi_3^{\downarrow (K_1 \cup K_2) \cap K_3}.$$

Analogously, $\pi_1 \triangleright \pi_3 \triangleright \pi_2$ is not defined *iff*

$$\pi_1^{\downarrow K_1 \cap K_3} \not\ll \pi_3^{\downarrow K_1 \cap K_3}$$

or

$$(\pi_1 \triangleright \pi_3)^{\downarrow (K_1 \cup K_3) \cap K_2} \not\ll \pi_2^{\downarrow (K_1 \cup K_3) \cap K_2}.$$

Under the given assumption $K_1 \supseteq (K_2 \cap K_3)$, these two conditions coincide because

$$((K_1 \cup K_2) \cap K_3) = (K_1 \cap K_3), \tag{3.7}$$

$$((K_1 \cup K_3) \cap K_2) = (K_1 \cap K_2), \tag{3.8}$$

and

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

Now, let us assume that both the expressions in formula (3.6) are defined. Because of (3.7) and (3.8) the expressions

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

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

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

Lemma 3.4.2 *If π_1 and π_2 are consistent then*

$$K_2 \supseteq (K_1 \cap K_3) \Longrightarrow \pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleleft \pi_2.$$

Proof. Let us start, again, by showing that, under the given assumptions, $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is undefined *iff* $\pi_1 \triangleright \pi_3 \triangleleft \pi_2$ is undefined.

Since we assume that π_1 and π_2 are consistent, the former expression is undefined only if

$$(\pi_1 \triangleright \pi_2)^{\downarrow K_3 \cap (K_1 \cup K_2)} \not\lll \pi_3^{\downarrow K_3 \cap (K_1 \cup K_2)}.$$

Moreover, since $K_2 \supseteq (K_1 \cap K_3) \Longrightarrow K_3 \cap (K_1 \cup K_2) = K_3 \cap K_2$,

$$(\pi_1 \triangleright \pi_2)^{\downarrow K_3 \cap (K_1 \cup K_2)} = (\pi_1 \triangleright \pi_2)^{\downarrow K_3 \cap K_2} = (\pi_1 \triangleleft \pi_2)^{\downarrow K_3 \cap K_2} = \pi_2^{\downarrow K_3 \cap K_2}.$$

Thus we got that $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is not defined *iff*

$$\pi_2^{\downarrow K_2 \cap K_3} \not\lll \pi_3^{\downarrow K_2 \cap K_3}. \tag{3.9}$$

$\pi_1 \triangleright \pi_3 \triangleleft \pi_2$ is undefined *iff*

$$\pi_1^{\downarrow K_1 \cap K_3} \not\lll \pi_3^{\downarrow K_1 \cap K_3}, \tag{3.10}$$

or

$$\pi_2^{\downarrow K_2 \cap (K_1 \cup K_3)} \not\lll (\pi_1 \triangleright \pi_3)^{\downarrow K_2 \cap (K_1 \cup K_3)}. \tag{3.11}$$

Since $K_1 \cap K_3 \subseteq K_2 \cap (K_1 \cup K_3)$ we can apply Lemma 3.2.4 getting

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

where the last equality follows from the consistency of π_1 and π_2. Thus we got that (3.11) is equivalent to

$$\pi_2^{\downarrow K_2 \cap (K_1 \cup K_3)} \not\lll \pi_2^{\downarrow K_2 \cap K_1} \triangleright \pi_3^{\downarrow K_2 \cap K_3}. \tag{3.12}$$

Regarding the fact that in our case $K_1 \cap K_3 \subseteq K_2$, and π_1 and π_2 are consistent, (3.10) is equivalent to

$$\pi_2^{\downarrow K_1 \cap K_3} \not\lll \pi_3^{\downarrow K_1 \cap K_3}. \tag{3.13}$$

The fact that (3.9) occurs *iff* (3.12) or (3.13) holds true, immediately follows from Lemma 3.1.2 (to check it assign: $\pi \leftarrow \pi_2^{\downarrow K_2 \cap (K_1 \cup K_3)}$, $\kappa \leftarrow \pi_3^{\downarrow K_2 \cap K_3}$ and $M \leftarrow K_1 \cap K_3$), which finishes the first part of the proof.

Now, it has remained to be shown that

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

in case that both sides of the equality are defined.

$$\pi_1 \triangleright \pi_3 \triangleleft \pi_2 = \frac{\pi_1\pi_2\pi_3}{\pi_3^{(K_1\cap K_3)}(\pi_1 \triangleright \pi_3)^{(K_2\cap(K_1\cup K_3))}},$$

and since we assume $K_2 \supseteq (K_1 \cap K_3)$, we can apply Lemma 3.2.4

$$(\pi_1 \triangleright \pi_3)^{(K_2\cap(K_1\cup K_3))} = \pi_1^{(K_1\cap K_2)} \triangleright \pi_3^{(K_3\cap K_2)} = \frac{\pi_1^{(K_1\cap K_2)}\pi_3^{(K_3\cap K_2)}}{\pi_3^{(K_1\cap K_3)}},$$

and therefore

$$\pi_1 \triangleright \pi_3 \triangleleft \pi_2 = \frac{\pi_1\pi_2\pi_3}{\pi_3^{(K_1\cap K_3)}\frac{\pi_1^{(K_1\cap K_2)}\pi_3^{(K_3\cap K_2)}}{\pi_3^{(K_1\cap K_3)}}} = \pi_1 \triangleright \pi_2 \triangleright \pi_3$$

because of $\pi_1^{(K_1\cap K_2)} = \pi_2^{(K_1\cap K_2)}$ and $K_3 \cap (K_1 \cup K_2) = K_3 \cap K_2$. □

Lemma 3.4.3 *If π_1 and π_3 are consistent then*

$$K_1 \supseteq (K_2 \cap K_3) \Longrightarrow \pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_2 \triangleleft \pi_3.$$

Proof. Both expressions $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ and $\pi_1 \triangleright \pi_2 \triangleleft \pi_3$ are not defined if $\pi_1 \triangleright \pi_2$ is undefined (i.e. if $\pi_1^{\downarrow K_1\cap K_2} \not\ll \pi_2^{\downarrow K_1\cap K_2}$).

In case that $\pi_1 \triangleright \pi_2$ is defined, then, under the given assumptions

$$K_3 \cap (K_1 \cup K_2) = K_3 \cap K_1,$$

we get that $(\pi_1 \triangleright \pi_2)^{\downarrow K_3\cap(K_1\cup K_2)} = \pi_1^{\downarrow K_3\cap K_1} = \pi_3^{\downarrow K_3\cap K_1}$, and therefore $\pi_1 \triangleright \pi_2$ and π_3 are consistent. Therefore both expressions $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ and $\pi_1 \triangleright \pi_2 \triangleleft \pi_3$ are defined and equivalent each to other (due to Lemma 3.2.2). □

Lemma 3.4.4 *If π_2 and π_3 are consistent then*

$$K_3 \supseteq (K_1 \cap K_2) \Longrightarrow \pi_1 \triangleleft \pi_2 \triangleleft \pi_3 = \pi_1 \triangleleft \pi_3 \triangleleft \pi_2.$$

Proof. Let us start with discussing the conditions under which the respective expressions are not defined. $\pi_1 \triangleleft \pi_3 \triangleleft \pi_2$ is not defined *iff* $\pi_1 \triangleleft \pi_3$ is not defined, i.e. when

$$\pi_3^{\downarrow K_1\cap K_3} \not\ll \pi_1^{\downarrow K_1\cap K_3}, \tag{3.14}$$

because in opposite case $(\pi_1 \triangleleft \pi_3) \triangleleft \pi_2$ is always defined. It follows from the fact that under the given assumptions $K_2 \cap (K_1 \cup K_3) = K_2 \cap K_3$ and $(\pi_1 \triangleleft \pi_3)^{\downarrow K_2\cap K_3} = \pi_2^{\downarrow K_2\cap K_3}$.

The other expression $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3$ is not defined *iff* at least one of the operators is not defined. The first one is not defined if $\pi_2^{K_1 \cap K_2} \not\ll \pi_1^{K_1 \cap K_2}$, or, which is, due to the given assumptions, equivalent to

$$\pi_3^{K_1 \cap K_2} \not\ll \pi_1^{K_1 \cap K_2}. \tag{3.15}$$

In case that (3.15) does not hold, then the second operator is not defined if

$$\pi_3^{\downarrow K_3 \cap (K_1 \cup K_2)} \not\ll (\pi_1 \triangleleft \pi_2)^{\downarrow K_3 \cap (K_1 \cup K_2)}.$$

Using Lemma 3.2.4, the given assumptions and the definition of the operators we can perform the following modifications

$$\begin{aligned}(\pi_1 \triangleleft \pi_2)^{\downarrow K_3 \cap (K_1 \cup K_2)} &= \pi_1^{\downarrow K_3 \cap K_1} \triangleleft \pi_2^{\downarrow K_3 \cap K_2} = \pi_1^{\downarrow K_3 \cap K_1} \triangleleft \pi_3^{\downarrow K_3 \cap K_2} \\ &= \pi_3^{\downarrow K_3 \cap K_2} \triangleright \pi_1^{\downarrow K_3 \cap K_1}\end{aligned}$$

getting an equivalent condition

$$\pi_3^{\downarrow K_3 \cap (K_1 \cup K_2)} \not\ll \pi_3^{\downarrow K_3 \cap K_2} \triangleright \pi_1^{\downarrow K_3 \cap K_1}. \tag{3.16}$$

Now, applying Lemma 3.1.2 we get that (3.14) occurs *iff* either (3.15) or (3.16) holds true. (For application of Lemma 3.1.2 we assign $\pi \leftarrow \pi_3^{\downarrow K_3 \cap (K_1 \cup K_2)}$, $\kappa \leftarrow \pi_1^{\downarrow K_1 \cap K_3}$ and $M \leftarrow K_1 \cap K_2$.)

Now let us start proving that if the respective expressions are defined they equal each other. For this, express $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3$ in the form of a ratio

$$\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 = \frac{\pi_1 \pi_2 \pi_3}{\pi_1^{(K_1 \cap K_2)} (\pi_1 \triangleleft \pi_2)^{(K_3 \cap (K_1 \cup K_2))}}.$$

Since we assume that $K_3 \supseteq (K_1 \cap K_2)$ we can apply Lemma 3.2.4 according to which

$$(\pi_1 \triangleleft \pi_2)^{(K_3 \cap (K_1 \cup K_2))} = \pi_1^{(K_3 \cap K_1)} \triangleleft \pi_2^{(K_3 \cap K_2)}.$$

In the following computations we apply the fact that under our assumptions $K_1 \cap K_2 \cap K_3 = K_1 \cap K_2$, and then the assumed consistency of π_2, π_3, the fact that $\pi_3^{(K_3\cap K_2)} = (\pi_1 \triangleleft \pi_3)^{(K_3\cap K_2)}$ and the equality $K_2 \cap (K_1 \cup K_3) = K_2 \cap K_3$:

$$\begin{aligned}
\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 &= \frac{\pi_1\pi_2\pi_3}{\pi_1^{(K_1\cap K_2)}\left(\pi_1^{(K_3\cap K_1)} \triangleleft \pi_2^{(K_3\cap K_2)}\right)} \\
&= \frac{\pi_1\pi_2\pi_3\pi_1^{(K_1\cap K_2\cap K_3)}}{\pi_1^{(K_1\cap K_2)}\pi_1^{(K_3\cap K_1)}\pi_2^{(K_3\cap K_2)}} = \frac{\pi_1\pi_2\pi_3}{\pi_1^{(K_1\cap K_3)}\pi_2^{(K_3\cap K_2)}} \\
&= \frac{\pi_1\pi_2\pi_3}{\pi_1^{(K_1\cap K_3)}\pi_3^{(K_3\cap K_2)}} = \frac{\pi_1\pi_2\pi_3}{\pi_1^{(K_1\cap K_3)}\,(\pi_1 \triangleleft \pi_3)^{(K_3\cap K_2)}} \\
&= \frac{\pi_1\pi_2\pi_3}{\pi_1^{(K_1\cap K_3)}\,(\pi_1 \triangleleft \pi_3)^{(K_2\cap(K_1\cup K_3)}} = \pi_1 \triangleleft \pi_3 \triangleleft \pi_2.
\end{aligned}$$

□

Lemma 3.4.5 *If $K_3 \supseteq (K_1 \cap K_2)$ then*

$$\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 = \pi_1 \triangleleft \pi_3 \triangleleft \pi_2 \triangleleft \pi_3. \tag{3.17}$$

if the right hand formula is defined.

Proof. To show that, under the given assumptions, $\pi_1 \triangleleft \pi_2$ is defined is simple. Namely, existence of $\pi_1 \triangleleft \pi_3 \triangleleft \pi_2 \triangleleft \pi_3$ guarantees that $\pi_3^{\downarrow K_1\cap K_3} \ll \pi_1^{\downarrow K_1\cap K_3}$ and $\pi_2^{\downarrow K_2\cap(K_1\cup K_3)} \ll (\pi_1 \triangleleft \pi_3)^{\downarrow K_2\cap(K_1\cup K_3)}$. Applying transitivity of the dominance and the assumption $K_3 \supseteq (K_1 \cap K_2)$ we immediately get that $\pi_2^{\downarrow K_1\cap K_2} \ll \pi_1^{\downarrow K_1\cap K_2}$.

Now, assume that $(\pi_1 \triangleleft \pi_2) \triangleleft \pi_3$ is not defined. It means that there exists $x \in \mathbf{X}_{K_3\cap(K_1\cup K_2)}$ such that $(\pi_1 \triangleleft \pi_2)(x) = 0 < \pi_3(x)$. Regarding the assumption that $\pi_1 \triangleleft \pi_3$ is defined it means that $\pi_2(x_{K_2\cap K_3}) = 0$, and therefore also $(\pi_1 \triangleleft \pi_3 \triangleleft \pi_2)(x) = 0$, which contradicts to the assumption that $\pi_1 \triangleleft \pi_3 \triangleleft \pi_2 \triangleleft \pi_3$ is defined. So, we have proven that $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3$ is also defined.

It remains to be shown that both the expressions in (3.17) equal each other. In the following computations we shall use only the definition of $\triangleleft$, relations following from the assumption $K_3 \supseteq (K_1 \cap K_2)$ and Lemma 3.2.4.

$$\begin{aligned}
\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 &= \frac{\pi_1\pi_2}{\pi_1^{(K_1\cap K_2)}}\frac{\pi_3}{(\pi_1 \triangleleft \pi_2)^{(K_3\cap(K_1\cup K_2))}} \\
&= \frac{\pi_1\pi_2\pi_3}{\pi_1^{(K_1\cap K_2)}(\pi_1^{(K_3\cap K_1)} \triangleleft \pi_2^{(K_3\cap K_2)})} \\
&= \frac{\pi_1\pi_2\pi_3\pi_1^{(K_1\cap K_2)}}{\pi_1^{(K_1\cap K_2)}\pi_1^{(K_3\cap K_1)}\pi_2^{(K_3\cap K_2)}} = \frac{\pi_1\pi_2\pi_3}{\pi_1^{(K_3\cap K_1)}\pi_2^{(K_3\cap K_2)}}.
\end{aligned}$$

Now, we shall use also the fact that $(\pi_1 \triangleleft \pi_3)^{(K_3)} = \pi_3$.

$$
\begin{aligned}
&\pi_1 \triangleleft \pi_3 \triangleleft \pi_2 \triangleleft \pi_3 \\
&\quad = \frac{\pi_1\pi_3}{\pi_1^{(K_1\cap K_3)}} \frac{\pi_2}{(\pi_1 \triangleleft \pi_3)^{(K_2\cap(K_1\cup K_3))}} \frac{\pi_3}{(\pi_1 \triangleleft \pi_3 \triangleleft \pi_2)^{(K_3\cap(K_1\cup K_2\cup K_3))}} \\
&\quad = \frac{\pi_1\pi_3}{\pi_1^{(K_1\cap K_3)}} \frac{\pi_2}{(\pi_1 \triangleleft \pi_3)^{(K_2\cap K_3)}} \frac{\pi_3}{(\pi_1 \triangleleft \pi_3 \triangleleft \pi_2)^{(K_3)}} \\
&\quad = \frac{\pi_1\pi_3}{\pi_1^{(K_1\cap K_3)}} \frac{\pi_2}{(\pi_1 \triangleleft \pi_3)^{(K_2\cap K_3)}} \frac{\pi_3}{(\pi_1 \triangleleft \pi_3)^{(K_3)} \triangleleft \pi_2^{(K_2\cap K_3)}} \\
&\quad = \frac{\pi_1\pi_3}{\pi_1^{(K_1\cap K_3)}} \frac{\pi_2}{(\pi_1 \triangleleft \pi_3)^{(K_2\cap K_3)}} \frac{\pi_3}{\pi_3 \triangleleft \pi_2^{(K_2\cap K_3)}} \\
&\quad = \frac{\pi_1\pi_3}{\pi_1^{(K_1\cap K_3)}} \frac{\pi_2}{\pi_3^{(K_2\cap K_3)}} \frac{\pi_3\pi_3^{(K_2\cap K_3)}}{\pi_3\pi_2^{(K_2\cap K_3)}} = \frac{\pi_1\pi_3\pi_2}{\pi_1^{(K_1\cap K_3)}\pi_2^{(K_2\cap K_3)}}.
\end{aligned}
$$

□

The following lemma is the only assertion presented here expressing a condition allowing to change the ordering of operations of composition that does not require a special form of sets K_i. It is based only on a special form of conditional independence of distributions π_2 and π_3 and coincidence of respective conditional distributions.

Lemma 3.4.6 *If*

$$\pi_2^{(K_2\cap K_3)}\pi_2^{(K_2\cap K_1)} = \pi_2^{(K_2\cap(K_1\cup K_3))}\pi_2^{(K_1\cap K_2\cap K_3)},$$

$$\pi_3^{(K_3\cap K_2)}\pi_3^{(K_3\cap K_1)} = \pi_3^{(K_3\cap(K_1\cup K_2))}\pi_3^{(K_1\cap K_2\cap K_3)},$$

and

$$\frac{\pi_2^{(K_2\cap K_3)}}{\pi_2^{(K_1\cap K_2\cap K_3)}} = \frac{\pi_3^{(K_3\cap K_2)}}{\pi_3^{(K_1\cap K_2\cap K_3)}}$$

then

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

Proof.

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \frac{\pi_1\pi_2\pi_3}{\pi_2^{(K_2\cap K_1)}\pi_3^{(K_3\cap(K_1\cup K_2))}}$$

and

$$\pi_1 \triangleright \pi_3 \triangleright \pi_2 = \frac{\pi_1\pi_2\pi_3}{\pi_3^{(K_3\cap K_1)}\pi_2^{(K_2\cap(K_1\cup K_3))}}.$$

Now, we shall show that the denominators in both these expressions equal each to other.

$$\begin{aligned}
\pi_2^{(K_2\cap K_1)}\pi_3^{(K_3\cap(K_1\cup K_2))} &= \pi_2^{(K_2\cap K_1)}\pi_3^{(K_3\cap(K_1\cup K_2))}\frac{\pi_3^{(K_1\cap K_2\cap K_3)}}{\pi_3^{(K_3\cap K_2)}}\frac{\pi_2^{(K_2\cap K_3)}}{\pi_2^{(K_1\cap K_2\cap K_3)}}\\
&= \frac{\pi_2^{(K_2\cap K_1)}\pi_3^{(K_3\cap K_2)}\pi_3^{(K_3\cap K_1)}\pi_2^{(K_2\cap K_3)}}{\pi_3^{(K_3\cap K_2)}\pi_2^{(K_1\cap K_2\cap K_3)}}\\
&= \frac{\pi_3^{(K_3\cap K_2)}\pi_3^{(K_3\cap K_1)}\pi_2^{(K_2\cap(K_1\cup K_3))}\pi_2^{(K_1\cap K_2\cap K_3)}}{\pi_3^{(K_3\cap K_2)}\pi_2^{(K_1\cap K_2\cap K_3)}}\\
&= \pi_3^{(K_3\cap K_1)}\pi_2^{(K_2\cap(K_1\cup K_3))}.
\end{aligned}$$

Therefore if one of the expressions $\pi_1 \triangleright \pi_2 \triangleright \pi_3$, $\pi_1 \triangleright \pi_3 \triangleright \pi_2$ is defined then the other one must be defined, too, and then

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

□

Corollary 3.4.1 *If π_2 and π_3 are consistent and for both $i = 2, 3$*

$$\pi_i^{(K_i\cap(K_1\cup K_{5-i}))} = \pi_i^{((K_i\cap K_1)\setminus K_{5-i})}\pi_i^{(K_2\cap K_3)},$$

then

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

Proof. We shall only verify that all the assumptions of the preceding Lemma are fulfilled. The equality

$$\frac{\pi_2^{(K_2\cap K_3)}}{\pi_2^{(K_1\cap K_2\cap K_3)}} = \frac{\pi_3^{(K_3\cap K_2)}}{\pi_3^{(K_1\cap K_2\cap K_3)}}$$

follows immediately from the consistency of π_2, π_3.

Moreover, under the given assumptions

$$\begin{aligned}
\pi_i^{(K_i\cap K_1)} &= (\pi_i^{(K_i\cap(K_1\cup K_{5-i}))})^{(K_i\cap K_1)} = (\pi_i^{((K_i\cap K_1)\setminus K_{5-i})}\pi_i^{(K_2\cap K_3)})^{(K_i\cap K_1)}\\
&= \pi_i^{((K_i\cap K_1)\setminus K_{5-i})}\pi_i^{(K_1\cap K_2\cap K_3)}
\end{aligned}$$

and therefore

$$\begin{aligned}
\pi_i^{(K_i\cap K_1)}\pi_i^{(K_2\cap K_3)} &= \pi_i^{((K_i\cap K_1)\setminus K_{5-i})}\pi_i^{(K_1\cap K_2\cap K_3)}\pi_i^{(K_2\cap K_3)}\\
&= \pi_i^{(K_i\cap(K_1\cup K_{5-i}))}\pi_i^{(K_1\cap K_2\cap K_3)}
\end{aligned}$$

for both $i = 2, 3$. □

To prove the remaining assertions of this section we will need still another operator (it will be called an *anticipating* operator), which will define a special type of composition of two distributions. Notice, that this operator is parametrized by an index set, which is the main difference with respect to the previously defined operators $\triangleright$ and $\triangleleft$. In Theorem 3.4.1 we will articulate the main purpose, why this operators is introduced. Namely, operator $\triangleright$ 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 3.4.5.

Definition 3.4.1 Let K, L, M be subsets of N. By application of the *anticipating operator* $\circledcirc_M$ to distributions $\pi \in \Pi^{(K)}$ and $\kappa \in \Pi^{(L)}$ we understand computation of the following distribution

$$\pi(x_K) \circledcirc_M \kappa(x_L) = (\kappa(x_{(M \setminus K) \cap L}) \cdot \pi(x_K)) \triangleright \kappa(x_L).$$

regardless of the choice of set M.

Remark 3.4.1 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 $\pi(x_K) \circledcirc_M \kappa(x_L)$ is not defined *iff* $\pi(x_K) \triangleright \kappa(x_L)$ is undefined. Moreover notice that distribution $\pi(x_K) \circledcirc_M \kappa(x_L)$, if defined, is defined for the same set of variables as the distribution $\pi(x_K) \triangleright \kappa(x_L)$, and that $(\pi(x_K) \circledcirc_M \kappa(x_L))^{\downarrow K} = \pi$. ○

Theorem 3.4.1 *If π_1, π_2 and π_3 are such that $\pi_1 \triangleright (\pi_2 \circledcirc_{K_1} \pi_3)$ is defined then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright (\pi_2 \circledcirc_{K_1} \pi_3) = \pi_2 \circledcirc_{K_1} \pi_3 \triangleleft \pi_1.$$

Proof. Assume that $\pi_1 \triangleright (\pi_2 \circledcirc_{K_1} \pi_3)$ is defined. It means that

$$\pi_1^{\downarrow K_1 \cap (K_2 \cup K_3)} \ll (\pi_2 \circledcirc_{K_1} \pi_3)^{\downarrow K_1 \cap (K_2 \cup K_3)}, \tag{3.18}$$

and, as a consequence of the fact that dominance holds also for the respective marginal distributions, $\pi_1^{\downarrow K_1 \cap K_2} \ll \pi_2^{\downarrow K_1 \cap K_2}$. This guarantees that $\pi_1 \triangleright \pi_2$ is defined.

Let us now show by contradiction that also $(\pi_1 \triangleright \pi_2) \triangleright \pi_3$ must be defined. Assume it is not defined. It means that there exists $x_{K_1 \cup K_2 \cup K_3}$ such that in the expression

$$(\pi_1 \triangleright \pi_2 \triangleright \pi_3)(x) = \frac{\pi_1(x_{K_1}) \cdot \pi_2(x_{K_2}) \cdot \pi_3(x_{K_3})}{\pi_2(x_{K_2 \cap K_1}) \cdot \pi_3(x_{K_3 \cap (K_1 \cup K_2)})}$$

$\pi_3(x_{K_3 \cap (K_1 \cup K_2)}) = 0$ and simultaneously $\pi_1(x_{K_1}) \cdot \pi_2(x_{K_2}) > 0$. This, however, contradicts to our assumption that $\pi_2 \circledtriangleright_{K_1} \pi_3$ is defined: as we can see from the respective formula

$$(\pi_2 \circledtriangleright_{K_1} \pi_3)(x) = \frac{\pi_3(x_{(K_1 \setminus K_2) \cap K_2}) \cdot \pi_2(x_{K_2}) \cdot \pi_3(x_{K_3})}{\pi_3(x_{K_3 \cap (K_1 \cup K_2)})},$$

$$\pi_3(x_{K_3 \cap (K_1 \cup K_2)}) = 0 \implies \pi_2(x_{K_2}) = 0.$$

Now, assuming $\pi_1 \triangleright (\pi_2 \circledtriangleright_{K_1} \pi_3)$ be defined let us compute (using the definition of the operator $\circledtriangleright$ and Lemma 3.2.4):

$$\begin{aligned}
\pi_1 \triangleright (\pi_2 \circledtriangleright_{K_1} \pi_3) &= \frac{\pi_1 \frac{\pi_3^{\downarrow (K_1 \setminus K_2) \cap K_3} \pi_2 \pi_3}{\pi_3^{\downarrow (K_1 \cup K_2) \cap K_3}}}{\left(\frac{\pi_3^{\downarrow (K_1 \setminus K_2) \cap K_3} \pi_2 \pi_3}{\pi_3^{\downarrow (K_1 \cup K_2) \cap K_3}} \right)^{\downarrow (K_2 \cup K_3) \cap K_1}} \\
&= \frac{\pi_3^{\downarrow (K_1 \setminus K_2) \cap K_3} \frac{\pi_1 \pi_2 \pi_3}{\pi_3^{\downarrow (K_1 \cup K_2) \cap K_3}}}{\pi_3^{\downarrow (K_1 \setminus K_2) \cap K_3} \left(\frac{\pi_2 \pi_3}{\pi_3^{\downarrow (K_1 \cup K_2) \cap K_3}} \right)^{\downarrow (K_2 \cup K_3) \cap K_1}} \\
&= \frac{\frac{\pi_1 \pi_2 \pi_3}{\pi_3^{\downarrow (K_1 \cup K_2) \cap K_3}}}{\left(\frac{\pi_2 \pi_3}{\pi_3^{\downarrow (K_1 \cup K_2) \cap K_3}} \right)^{\downarrow (K_2 \cup K_3) \cap K_1}},
\end{aligned}$$

where the second modification is feasible because

$$(K_1 \setminus K_2) \cap K_3 \subseteq (K_2 \cup K_3) \cap K_1.$$

Notice here that the last modification is just an elimination of the auxiliary distribution $\pi_3^{\downarrow (K_1 \setminus K_2) \cap K_3)}$ introduced in the definition of the operator $\circledtriangleright_{K_1}$. Therefore, we can see that instead of $\pi_3^{\downarrow (K_1 \setminus K_2) \cap K_3}$ we could use (almost) an arbitrary distribution $\kappa(x_{(K_1 \setminus K_2) \cap K_3})$ (see Remark 3.4.4).

Let us focus our attention on the denominator of the last fraction. It is a marginal of a product of π_2 with a conditional distribution

$$\pi_3(x_{K_3\setminus(K_1\cup K_2)}|x_{K_3\cap(K_1\cup K_2)}).$$

When computing this marginal, we have to sum up over all combinations of values of variables $X_{(K_2\cup K_3)\setminus K_1}$. In the following computations we will separate these variables into two groups: $X_{K_2\setminus K_1}$ and $X_{K_3\setminus(K_1\cup K_2)}$. $x_{K_2} \in \mathbf{X}_{K_2}$ is thus a vector of values of variables X_{K_2} which can be split into two parts:

$$x_{K_2} = (x_{K_2\setminus K_1}, x_{K_2\cap K_1}).$$

Analogously, for $x_{K_3\cap(K_1\cup K_2)} \in \mathbf{X}_{K_3\cap(K_1\cup K_2)}$ we will consider parts

$$x_{K_3\cap(K_1\cup K_2)} = (x_{K_3\cap K_1}, x_{(K_3\cap K_2)\setminus K_1}).$$

Using this notation, we can compute:

$$\left(\pi_2(x_{K_2})\pi_3(x_{K_3\setminus(K_1\cup K_2)}|x_{K_3\cap(K_1\cup K_2)})\right)^{\downarrow(K_2\cup K_3)\cap K_1}$$

$$= \sum_{x_{K_2\setminus K_1}\in\mathbf{X}_{K_2\setminus K_1}} \sum_{x_{K_3\setminus(K_1\cup K_2)}\in\mathbf{X}_{K_3\setminus(K_1\cup K_2)}} \pi_2(x_{K_2\cap K_1}, x_{K_2\setminus K_1})$$

$$\cdot\pi_3(x_{K_3\setminus(K_1\cup K_2)}|x_{(K_3\cap K_2)\setminus K_1}, x_{K_3\cap K_1})$$

$$= \pi_2(x_{K_2\cap K_1}) \sum_{x_{K_2\setminus K_1}} \pi_2(x_{K_2\setminus K_1}|x_{K_2\cap K_1})$$

$$\sum_{x_{K_3\setminus(K_1\cup K_2)}} \pi_3(x_{K_3\setminus(K_1\cup K_2}|x_{(K_3\cap K_2)\setminus K_1}, x_{K_3\cap K_1})$$

$$= \pi_2((X_i)_{i\in K_2\cap K_1}).$$

Substituting this result back into the denominator of the fraction, we get

$$\pi_1 \triangleright (\pi_2 \circledtriangleright_{K_1} \pi_3) \quad = \quad \frac{\frac{\pi_1\pi_2\pi_3}{\pi_3^{\downarrow(K_1\cup K_2)\cap K_3}}}{\pi_2^{\downarrow K_2\cap K_1}} = \frac{\pi_1\pi_2\pi_3}{\pi_2^{\downarrow K_2\cap K_1}\pi_3^{\downarrow(K_1\cup K_2)\cap K_3}} = \pi_1 \triangleright \pi_2 \triangleright \pi_3.$$

which completes the proof. □

Remark 3.4.2 Notice that the assertion does not claim that the equality holds true when $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is defined. This is because it can easily happen that $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is defined and $\pi_1 \triangleright (\pi_2 \circledtriangleright_{K_1} \pi_3)$ not. To show it consider a simple situation with the following distributions:

$$\pi_1(x) = \left(\frac{1}{5}, \frac{1}{2}, 0\right)$$
$$\pi_2(x) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$$
$$\pi_3(x) = \left(\frac{1}{5}, \frac{1}{2}, 0\right).$$

○

Corollary 3.4.2 *If $K_2 \supseteq (K_1 \cap K_3)$ then*

$$\pi_2 \circledtriangleright_{K_1} \pi_3 = \pi_2 \triangleright \pi_3$$

and thus also

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

Proof. The assertion is an immediate consequence of the fact

$$K_2 \supseteq (K_1 \cap K_3) \Longrightarrow (K_1 \setminus K_2) \cap K_3 = \emptyset.$$ □

Remark 3.4.3 It should be highlighted here that the computational complexity of the composition $\pi_2 \circledtriangleright_{K_1} \pi_3$ does not differ substantially from the complexity of computation of $\pi_2 \triangleright \pi_3$. It follows, namely, from the fact that both of these distributions are of the same dimensionality; both are defined for variables $X_{K_2 \cup K_3}$. In other words, in both cases we have to compute the same number of probability values. ○

Remark 3.4.4 As mentioned in the proof, we could define the $\circledtriangleright_K$ operator with the aid of an (almost) arbitrary distribution ν

$$\pi_2 \circledtriangleright_{K_1} \pi_3 = (\nu^{((K_1 \setminus K_2) \cap K_3)} \pi_2) \triangleright \pi_3.$$

For example, an arbitrary positive distribution which is defined for the respective variables will serve well. For the sake of simplicity, it seems reasonable to consider a uniform distribution. ○

Remark 3.4.5 As the reader can see, the operator is parameterized by the index set K_1. The purpose of the operator is, namely, to compose the distributions (in our case distributions π_2 and π_3) but, simultaneously, to introduce the necessary independence of variables $X_{(K_1 \setminus K_2) \cap K_3}$ and X_{K_2} that would otherwise be omitted. If we want to compose distributions π_2 and π_3 before π_1 is considered, we have to "anticipate" the independence which was originally introduced by the previous operator. This also explains why the operator $\textcircled{\triangleright}_K$ is called an anticipating operator. ○

Example 3.4.2 *As said before, 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* $\pi_1(x_1), \pi_2(x_2), \pi_3(x_1, x_2)$ *for which*

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

If we used the operator $\triangleright$ *instead of* $\textcircled{\triangleright}_{K_1}$, *we would get*

$$\pi_1(x_1) \triangleright (\pi_2(x_2) \triangleright \pi_3(x_1, x_2)) = \frac{\pi_1(x_1)(\pi_2(x_2)\pi_3(x_1|x_2))}{\sum\limits_{x_1 \in \mathbf{X}_1} \pi_2(x_2)\pi_3(x_1|x_2)},$$

which evidently differs from $\pi_1(x_1)\pi_2(x_2)$ *because* $\pi_1 \triangleright (\pi_2 \triangleright \pi_3)$ *inherits the dependence of variables* X_1 *and* X_2 *from* π_3. *Nevertheless, considering*

$$\begin{aligned} \pi_1(x_1) \triangleright (\pi_2(x_2) \textcircled{\triangleright}_{\{1\}} \pi_3(x_1, x_2)) &= \pi_1(x_1) \triangleright (\pi_3(x_1)\pi_2(x_2) \triangleright \pi_3(x_1, x_2)) \\ &= \pi_1(x_1) \triangleright \pi_3(x_1)\pi_2(x_2) = \pi_1(x_1)\pi_2(x_2) \end{aligned}$$

we get the desired result.

Perhaps, it is also worth of mentioning that in this example $\pi_2(x_2) \textcircled{\triangleright}_{\{1\}} \pi_3(x_1, x_2)) = \pi_3(x_1)\pi_2(x_2)$ *is not a marginal distribution of the resulting* $\pi_1(x_1) \triangleright \pi_2(x_2) \triangleright \pi_3(x_1, x_2)$. ◇

Example 3.4.3 *Let us present another, a little bit more complex example illustrating application of the anticipating operator. This time consider distributions* $\pi_1(x_1, x_2, x_3, x_4), \pi_2(x_2, x_3, x_5), \pi_3(x_3, x_4, x_6)$. *In this case according to Theorem 3.4.1*

$$\begin{aligned} &\pi_1(x_1, x_2, x_3, x_4) \triangleright \pi_2(x_2, x_3, x_5) \triangleright \pi_3(x_3, x_4, x_6) \\ &\quad = \pi_1(x_1, x_2, x_3, x_4) \triangleright \Big(\pi_2(x_2, x_3, x_5) \textcircled{\triangleright}_{\{1,2,3,4\}} \pi_3(x_3, x_4, x_6)\Big). \end{aligned}$$

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

$$\pi_3(x_4)\pi_2(x_2,x_3,x_5) \triangleright \pi_3(x_3,x_4,x_6) \\ = \pi_3(x_4)\pi_2(x_2,x_3,x_5)\pi_3(x_6|x_3,x_4). \tag{3.19}$$

The reader certainly noticed that, thanks to the anticipating operator, there appears $\pi_3(x_6|x_3,x_4)$ in this formula, which is exactly the form at which π_3 occurs in

$$\pi_1(x_1,x_2,x_3,x_4) \triangleright \pi_2(x_2,x_3,x_5) \triangleright \pi_3(x_3,x_4,x_6) \\ = \pi_1(x_1,x_2,x_3,x_4)\pi_2(x_5|x_2,x_3)\pi_3(x_6|x_3,x_4).$$

Moreover, formula (3.19) allows for simple computation of the (in the next step) required marginal:

$$\left(\pi_2(x_2,x_3,x_5) \ \circledcirc\!\!\triangleright_{\{1,2,3,4\}} \pi_3(x_3,x_4,x_6)\right)^{\downarrow\{2,3,4\}} = \pi_3(x_4)\pi_2(x_2,x_3).$$

Therefore

$$\pi_1(x_1,x_2,x_3,x_4) \triangleright \left(\pi_2(x_2,x_3,x_5) \ \circledcirc\!\!\triangleright_{\{1,2,3,4\}} \pi_3(x_3,x_4,x_6)\right) \\ = \pi_1(x_1,x_2,x_3,x_4)\frac{\pi_3(x_4)\pi_2(x_2,x_3,x_5)\pi_3(x_6|x_3,x_4)}{\pi_3(x_4)\pi_2(x_2,x_3)} \\ = \pi_1(x_1,x_2,x_3,x_4)\pi_2(x_5|x_2,x_3)\pi_3(x_6|x_3,x_4). \qquad \Diamond$$

The last assertion of this section resembles Corollary 3.4.2. Notice, however, that in this case we cannot say that $\pi_2 \circledcirc\!\!\triangleright_{K_1} \pi_3 = \pi_2 \triangleright \pi_3$.

Lemma 3.4.7 *If $K_1 \supseteq (K_2 \cap K_3)$ then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright (\pi_2 \triangleright \pi_3) = \pi_2 \triangleright \pi_3 \triangleleft \pi_1$$

if the right hand side formula is defined.

Proof. First let us show that if $\pi_2 \triangleright \pi_3 \triangleleft \pi_1$ is defined then also $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ must be defined.

Considering definition of the operator of composition we know that $\pi_2 \triangleright \pi_3 \triangleleft \pi_1$ is defined *iff*

$$\pi_2^{\downarrow K_2 \cap K_3} \ll \pi_3^{\downarrow K_2 \cap K_3}, \tag{3.20}$$

$$\pi_1^{\downarrow K_1 \cap (K_2 \cup K_3)} \ll (\pi_2 \triangleright \pi_3)^{\downarrow K_1 \cap (K_2 \cup K_3)}. \tag{3.21}$$

A trivial consequence of the definition says that the dominance must hold true also for the respective marginal distributions, and therefore (3.21) yields

$$\pi_1^{\downarrow K_1\cap K_2} \ll (\pi_2 \triangleright \pi_3)^{\downarrow K_1\cap K_2} = \pi_2^{\downarrow K_1\cap K_2}, \tag{3.22}$$

which guarantees that $\pi_1 \triangleright \pi_2$ is defined.

To show that also the second operator in $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is defined we have to prove that

$$(\pi_1 \triangleright \pi_2)^{\downarrow K_3\cap(K_1\cup K_2)} \ll \pi_3^{\downarrow K_3\cap(K_1\cup K_2)}$$

holds, which is, because of the assumption $K_1 \supseteq (K_2 \cap K_3)$, equivalent to

$$\pi_1^{\downarrow K_1\cap K_3} \ll \pi_3^{\downarrow K_1\cap K_3}.$$

Assume the opposite: It means that there exists $x \in \mathbf{X}_{K_1\cap K_3}$ such that $\pi_1(x) > 0$ and $\pi_3(x) = 0$, or, which is the same (because $K_2 \cap K_3 \subseteq K_1 \cap K_3$),

$$\pi_3(x_{K_2\cap K_3})\pi_3(x_{(K_1\setminus K_2)\cap K_3}|x_{K_2\cap K_3}) = 0.$$

However, from (3.20) and (3.22) we get by transitivity of the dominance that

$$\pi_1^{\downarrow K_2\cap K_3} \ll \pi_3^{\downarrow K_2\cap K_3},$$

and therefore $\pi_3(x_{K_2\cap K_3}) > 0$. Therefore, we are getting that

$$\pi_3(x_{(K_1\setminus K_2)\cap K_3}|x_{K_2\cap K_3}) = 0,$$

which contradicts to (3.21).

Now, we are able to start proving the required equivalence. With respect to Theorem 3.4.1 it is enough to show that, under the given assumption,

$$\pi_1 \triangleright (\pi_2 \triangleright \pi_3) = \pi_1 \triangleright (\pi_2 \circledtriangleright_{K_1} \pi_3).$$

Therefore, regarding the definition of operator $\triangleright$, we shall show that

$$\frac{\pi_2 \circledtriangleright_{K_1} \pi_3}{(\pi_2 \circledtriangleright_{K_1} \pi_3)^{\downarrow K_1\cap(K_2\cup K_3)}} = \frac{\pi_2 \triangleright \pi_3}{(\pi_2 \triangleright \pi_3)^{\downarrow K_1\cap(K_2\cup K_3)}}.$$

In the following modifications we will employ the relations following from the assumption $K_1 \supseteq (K_2 \cap K_3)$, Lemma 3.2.4 and definitions of the respective operators:

$$\frac{\pi_2 \circledtriangleright_{K_1} \pi_3}{(\pi_2 \circledtriangleright_{K_1} \pi_3)^{\downarrow K_1\cap(K_2\cup K_3)}} = \frac{\left(\pi_3^{\downarrow(K_1\setminus K_2)\cap K_3}\pi_2\right) \triangleright \pi_3}{\left(\left(\pi_3^{\downarrow(K_1\setminus K_2)\cap K_3}\pi_2\right) \triangleright \pi_3\right)^{\downarrow K_1\cap(K_2\cup K_3)}}$$

$$= \frac{\left(\pi_3^{\downarrow(K_1\setminus K_2)\cap K_3}\pi_2\right) \triangleright \pi_3}{\left(\pi_3^{\downarrow(K_1\setminus K_2)\cap K_3}\pi_2\right)^{\downarrow K_1\cap(K_2\cup K_3)} \triangleright \pi_3^{\downarrow K_1\cap K_3}}$$

$$= \frac{\left(\pi_3^{\downarrow(K_1\setminus K_2)\cap K_3}\pi_2\right) \triangleright \pi_3}{\left(\pi_3^{\downarrow(K_1\setminus K_2)\cap K_3}\pi_2^{\downarrow K_1\cap K_2}\right) \triangleright \pi_3^{\downarrow K_1\cap K_3}} = \frac{\left(\pi_3^{\downarrow(K_1\setminus K_2)\cap K_3}\pi_2\right) \triangleright \pi_3}{\pi_3^{\downarrow(K_1\setminus K_2)\cap K_3}\pi_2^{\downarrow K_1\cap K_2}}$$

$$= \frac{\pi_3^{\downarrow(K_1\setminus K_2)\cap K_3}\pi_2}{\pi_3^{\downarrow(K_1\setminus K_2)\cap K_3}\pi_2^{\downarrow K_1\cap K_2}} \frac{\pi_3}{\pi_3^{\downarrow K_3\cap(K_1\cup K_2)}}$$

$$= \frac{\pi_2\pi_3}{\pi_2^{\downarrow K_1\cap K_2}\pi_3^{\downarrow K_3\cap(K_1\cup K_2)}} \frac{\pi_3^{(K_1\cap K_2\cap K_3)}}{\pi_3^{(K_2\cap K_3)}}$$

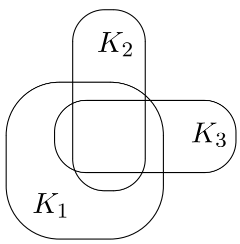
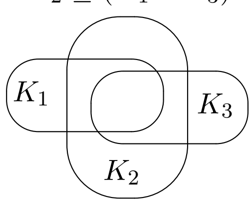
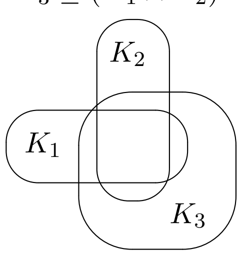
$$= \frac{\pi_2 \triangleright \pi_3}{\pi_2^{\downarrow K_1\cap K_2} \triangleright \pi_3^{\downarrow K_1\cap K_2\cap K_3}} = \frac{\pi_2 \triangleright \pi_3}{(\pi_2 \triangleright \pi_3)^{\downarrow K_1\cap(K_2\cup K_3)}}.$$

□

Remark 3.4.6 Notice that it may happen that $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is defined and for the same distributions, $\pi_2 \triangleright \pi_3 \triangleleft \pi_1$ is undefined. The reader can show it for the same distributions as mentioned in Remark 3.4.2. ○

To conclude this section, we shall make an overview of the most important assertions presented in this section. Since they describe situations under which an order of operators of composition can be changed without influencing the resulting ("more-dimensional") distribution we shall use them quite often in the following proofs. For this, the table contains the assumptions in an easy to check way.

Table 3.11: Survey of assertions enabling alteration of order of compositions

<table>
<tr><th>structural property</th><th>requirement on consistency</th><th>statement</th><th>reference</th></tr>
<tr><td>-</td><td>-</td><td>$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright (\pi_2 \textcircled{$\triangleright$}_{K_1} \pi_3)$</td><td>Theorem 3.4.1</td></tr>
<tr><td rowspan="3">$K_1 \supseteq (K_2 \cap K_3)$

</td><td>-</td><td>$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleright \pi_2$</td><td>Lemma 3.4.1</td></tr>
<tr><td>π_1, π_3</td><td>$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_2 \triangleleft \pi_3$</td><td>Lemma 3.4.3</td></tr>
<tr><td>-</td><td>$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright (\pi_2 \triangleright \pi_3)$</td><td>Lemma 3.4.7</td></tr>
<tr><td rowspan="2">$K_2 \supseteq (K_1 \cap K_3)$

</td><td>π_1, π_2</td><td>$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleleft \pi_2$</td><td>Lemma 3.4.2</td></tr>
<tr><td>-</td><td>$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright (\pi_2 \triangleright \pi_3)$</td><td>Corollary 3.4.2</td></tr>
<tr><td rowspan="2">$K_3 \supseteq (K_1 \cap K_2)$

</td><td>π_2, π_3</td><td>$\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 = \pi_1 \triangleleft \pi_3 \triangleleft \pi_2$</td><td>Lemma 3.4.4</td></tr>
<tr><td>-</td><td>$\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 = \pi_1 \triangleleft \pi_3 \triangleleft \pi_2 \triangleleft \pi_3$</td><td>Lemma 3.4.5</td></tr>
</table>

Chapter 4

Generating sequences

In this chapter we will start considering *multidimensional compositional models*, i.e. multidimensional probability distributions assembled from sequences of oligodimensional distributions with the help of operators of composition. To avoid some technical problems and necessity of repeating some assumptions too many times, let us make two important conventions. In this and the following chapters we will consider a system of n oligodimensional distributions $\pi_1(x_{K_1}),\ \pi_2(x_{K_2}), \ldots, \pi_n(x_{K_n})$. Therefore, whenever we will speak about a distribution π_k, if not specified explicitly 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 $\pi_k(x_{K_k})$. Thus, formulae $\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n$ and $\pi_1 \triangleleft \pi_2 \triangleleft \ldots \triangleleft \pi_n$, if they are defined, will determine the distributions of variables $X_{K_1 \cup K_2 \cup \ldots \cup K_n}$. And our second convention concerns this very condition that the models in question are defined. Namely, since now, we shall always assume that the models we shall speak about, which will be constructed by application of operators of composition, will be defined.

Recall that regarding the fact that neither of these operators is commutative or associative, we always apply the operators from left to right; e.g.

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

Therefore, in order to construct a distribution it is sufficient to determine a sequence – we will call it a *generating sequence* – of oligodimensional distributions. Moreover, since generally

$$\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n \neq \pi_1 \triangleleft \pi_2 \triangleleft \ldots \triangleleft \pi_n$$

it is necessary to determine which of the operators, $\triangleright$ or $\triangleleft$, is used for composition. If not said explicitly otherwise, we will usually consider operator $\triangleright$. This is because these two operators substantially differ from the computational point of view. To realize it, consider application of the k-th operator in the sequences $\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n$ and $\pi_1 \triangleleft \pi_2 \triangleleft \ldots \triangleleft \pi_n$. In the first case, when computing

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

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

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

can only be done when computation of

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

is tractable; since the distribution $(\pi_1 \triangleleft \ldots \triangleleft \pi_k)$ can be of a very high dimension it is not always the case. We will see that marginalization in compositional models is not a simple task; we will devote to it Section 7.1.

In agreement with what has just been said, for example, the generating sequence

$$\pi_1(x_1, x_3), \pi_2(x_3, x_5), \pi_3(x_1, x_4, x_5, x_6), \pi_4(x_2, x_5, x_6)$$

defines the distribution

$$\begin{aligned}
&(\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4)(x_1, x_2, x_3, x_4, x_5, x_6) \\
&\qquad = ((\pi_1(x_1, x_3) \triangleright \pi_2(x_3, x_5))\, \pi_3(x_1, x_4, x_5, x_6)) \triangleright \pi_4(x_2, x_5, x_6) \\
&\qquad = \pi_1(x_1, x_3)\pi_2(x_5|x_3)\pi_3(x_4, x_6|x_1, x_5)\pi_4(x_2|x_5, x_6).
\end{aligned}$$

4.1 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.

Definition 4.1.1 A generating sequence of probability distributions $\pi_1, \pi_2, \ldots, \pi_n$ is called *perfect* if $\pi_1 \triangleright \ldots \triangleright \pi_n$ is defined and

$$\begin{aligned}
&\pi_1 \triangleright \pi_2 = \pi_1 \triangleleft \pi_2, \\
&\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleleft \pi_2 \triangleleft \pi_3, \\
&\qquad \vdots \\
&\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n = \pi_1 \triangleleft \pi_2 \triangleleft \ldots \triangleleft \pi_n.
\end{aligned}$$

From this definition one can hardly see the importance of perfect sequences. This importance becomes clearer from the following characterization theorem. First, however, let us present a technical property, which, being an immediate consequence of an inductive application of Lemma 3.2.2, is presented without a proof.

Lemma 4.1.1 *A sequence $\pi_1, \pi_2, \ldots, \pi_n$ is perfect if and only if the pairs of distributions $(\pi_1 \triangleright \ldots \triangleright \pi_{m-1})$ and π_m are consistent for all $m = 2, 3, \ldots, n$.*

Theorem 4.1.1 *A sequence of distributions $\pi_1, \pi_2, \ldots, \pi_n$ is perfect* iff *all the distributions from this sequence are marginals of the distribution $(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)$.*

Proof. The fact that all distributions π_k from a perfect sequence are marginals of $(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)$ follows from the fact that $(\pi_1 \triangleright \ldots \triangleright \pi_k)$ is marginal to $(\pi_1 \triangleright \ldots \triangleright \pi_n)$ and π_k is marginal to $(\pi_1 \triangleleft \ldots \triangleleft \pi_k)$.

Suppose that for all $k = 1, \ldots, n$, π_k are marginal distributions of $(\pi_1 \triangleright \ldots \triangleright \pi_n)$. It means that all the distributions from the sequence are pairwise consistent, and that each π_k is consistent with any marginal distribution of $(\pi_1 \triangleright \ldots \triangleright \pi_n)$. Therefore, π_1 and π_2 are consistent, and due to Lemma 3.2.2

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

Since $\pi_1 \triangleright \pi_2$ is marginal to $(\pi_1 \triangleright \ldots \triangleright \pi_n)$ (Lemma 3.1.1), it must be consistent with π_3, too. Using Lemma 3.2.2 again, we get

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

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

$$\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_k = \pi_1 \triangleleft \pi_2 \triangleleft \ldots \triangleleft \pi_k. \qquad \square$$

Example 4.1.1 *The above presented theorem claims 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 non-perfect generating sequence need not preserve one-dimensional marginal distributions – even in case that the given oligodimensional distributions are pairwise consistent.*

Consider distribution $\pi(x_1, x_2, x_3)$ *from Table 2.5 on page 27. It is obvious that* $\pi(x_1), \pi(x_2)$ *and* $\pi(x_1, x_2, x_3)$ *must be pairwise consistent. Let us deal with the distribution defined by generating sequence* $\pi(x_1), \pi(x_2)$, $\pi(x_1, x_2, x_3)$*, i.e. with the distribution*

$$\pi(x_1) \triangleright \pi(x_2) \triangleright \pi(x_1, x_2, x_3).$$

Since both the considered 1*-dimensional marginal distributions* $\pi(x_1), \pi(x_2)$ *are uniform, the composition* $\pi(x_1) \triangleright \pi(x_2)$ *is also uniform. Thus it is an easy task to compute distribution* $\kappa = \pi(x_1) \triangleright \pi(x_2) \triangleright \pi(x_1, x_2, x_3)$*, which is in Table 4.1.*

Table 4.1: Distribution $\pi(x_1) \triangleright \pi(x_2) \triangleright \pi(x_1, x_2, x_3)$

	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{2}{24}$	$\frac{3}{24}$	$\frac{6}{24}$	$\frac{2}{24}$
$x_3 = 1$	0.0	$\frac{3}{24}$	0.0	$\frac{2}{24}$
$x_3 = 2$	$\frac{4}{24}$	0.0	0.0	$\frac{2}{24}$

Summarizing entries in rows of Tables 2.5 and 4.1 we get the respective one-dimensional marginal distributions $\pi(x_3)$ *and* $\kappa(x_3)$*, respectively, from which we see that these distributions are different:*

$$\begin{aligned} \pi(x_3 = 0) &= 0.5 \qquad & \kappa(x_3 = 0) &= \frac{13}{24}, \\ \pi(x_3 = 1) &= 0.2 \qquad & \kappa(x_3 = 1) &= \frac{5}{24}, \\ \pi(x_3 = 2) &= 0.3 \qquad & \kappa(x_3 = 2) &= \frac{6}{24}. \end{aligned}$$

◊

Remark 4.1.1 What is the main message conveyed by the characterization Theorem 4.1.1? 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 much interested in perfect sequence models. ○

Remark 4.1.2 From Theorem 4.1.1 and the definition of a perfect sequence we also immediately get that for perfect sequence $\pi_1, \ldots, \pi_n$ all the distributions π_k ($k = 1, \ldots, n$) are marginals of $\pi_1 \triangleleft \ldots \triangleleft \pi_n$. It should be, however, stressed that in this case it does not mean that if all $\pi_1, \ldots, \pi_n$ are marginal to $\pi_1 \triangleleft \ldots \triangleleft \pi_n$ that the considered sequence is perfect. We will illustrate this fact by the following example. ○

Example 4.1.2 *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, 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, π_3 are pairwise consistent. Since $\{2, 3\} \supset \{1, 2\} \cap \{3, 4\}$ we can apply Lemma 3.4.2, 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 was possible because of Lemma 3.2.2.) Thus we got that all π_1, π_2, π_3 are marginal distributions of $\pi_1 \triangleleft \pi_3 \triangleleft \pi_2$.

The question is whether also the sequence π_1, π_3, π_2 is perfect. Using Lemma 4.1.1 it happens if and only if the following two pairs of distributions are consistent: (π_1, π_3) and $(\pi_1 \triangleright \pi_3, \pi_2)$. The former pair of distributions is consistent, however, the latter one is consistent only when $(\pi_1 \triangleright \pi_3)(x_2, x_3) = \pi_2(x_2, x_3)$, which generally need not be true, because $(\pi_1 \triangleright \pi_3)(x_2, x_3) = \pi_1(x_2)\pi_3(x_3)$. 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

$$X_2 \perp\!\!\!\perp X_3[\pi_2].$$

So we see that all distributions π_1, π_2, 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 4.1.3 Notice that when defining perfect sequence, let alone 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 (ie., $K_j \subset K_k$) is fully sensible and may carry an information about the distribution. If $\pi(x_1), \pi(x_2), \pi(x_1, x_2, x_3)$ is a perfect sequence modelling a 3-dimensional distribution, it is quite obvious that

$$\pi(x_1) \triangleright \pi(x_2) \triangleright \pi(x_1, x_2, x_3) = \pi(x_1, x_2, x_3)$$

(because all the elements of a perfect sequence are marginals of the resulting distribution and therefore $\pi(x_1, x_2, x_3)$ must be marginal to $\pi(x_1) \triangleright \pi(x_2) \triangleright \pi(x_1, x_2, x_3)$). Nevertheless, it may happen that for some reason or another it may be advantageous to work with the model defined by the perfect sequence than just with the distribution $\pi(x_1, x_2, x_3)$. From the 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 $\pi(x_1, x_2, x_3)$. (How to read all the conditional independence relations from a compositional model will be presented in Section 6.) ○

The following assertion shows that each generating sequence, for which $\pi_1 \triangleright \ldots \triangleright \pi_n$ is defined, can be transformed into a perfect sequence (it is, in a way, a generalization of Lemma 3.2.5).

Theorem 4.1.2 *For any generating sequence* $\pi_1, \pi_2, \ldots, \pi_n$ *the sequence* $\kappa_1, \kappa_2, \ldots, \kappa_n$ *computed by the following process*

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

is perfect and

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

Proof The perfectness of the sequence $\kappa_1, \ldots, \kappa_n$ follows immediately from Lemma 4.1.1 and from the definition of this sequence as

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

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

Let us prove

$$\pi_1 \triangleright \ldots \triangleright \pi_n = \kappa_1 \triangleright \ldots \triangleright \kappa_n$$

by mathematical induction. Since $\pi_1 = \kappa_1$ by definition, it is enough to show that

$$\pi_1 \triangleright \ldots \triangleright \pi_i = \kappa_1 \triangleright \ldots \triangleright \kappa_i$$

implies also

$$\pi_1 \triangleright \ldots \triangleright \pi_{i+1} = \kappa_1 \triangleright \ldots \triangleright \kappa_{i+1}.$$

In the following computations we will use the fact that due to Lemma 3.2.4

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

and afterwards we will employ Lemma 3.2.5.

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

□

Example 4.1.3 *From the theoretical point of view the process of perfectization described by Theorem 4.1.2 is simple. Unfortunately, it does not hold for its computational complexity. Namely, the process requires marginalization of models, which may be multidimensional distributions, represented by generating sequences. We have already mentioned before that a whole section will be devoted to this topic (Section 7.1). Nevertheless, just to illustrate this process consider the following example. In it we will also see that a generating sequence consisting of (even only three) pairwise consistent distributions need not be perfect.*

Consider three 3-dimensional distributions presented in Table 4.2. The reader can easily verify that these distributions are pairwise consistent (i.e., $\pi_1(x_2, x_3) = \pi_2(x_2, x_3)$, $\pi_1(x_1) = \pi_3(x_1)$ and $\pi_2(x_4) = \pi_3(x_4)$) and that sequence π_1, π_2, π_3 is not perfect (hint: since π_1 and π_2 are consistent it is obvious that to show that π_1, π_2, π_3 is not a perfect sequence one has to show that distributions $\pi_1 \triangleright \pi_2$ and π_3 are not consistent; e.g. $(\pi_1 \triangleright \pi_2)(x_1 = 0, x_4 = 0) = 0.30275$, whilst $\pi_3(x_1 = 0, x_4 = 0) = 0.28$ – see also the following computations).

Table 4.2: 3-dimensional distributions estimated from data

π_1	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	0.18	0.14	0.12	0.14
$x_3 = 1$	0.04	0.18	0.06	0.14

π_2	$x_4 = 0$		$x_4 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	0.20	0.16	0.10	0.12
$x_3 = 1$	0.06	0.14	0.04	0.18

π_3	$x_1 = 0$		$x_1 = 1$	
	$x_4 = 0$	$x_4 = 1$	$x_4 = 0$	$x_4 = 1$
$x_5 = 0$	0.20	0.08	0.20	0.06
$x_5 = 1$	0.08	0.18	0.08	0.12

In this case, the process of perfectization consists of computation of the following three distributions

$$\begin{aligned}
\kappa_1(x_1, x_2, x_3) &= \pi_1(x_1, x_2, x_3), \\
\kappa_2(x_2, x_3, x_4) &= \kappa_1(x_2, x_3) \triangleright \pi_2(x_2, x_3, x_4), \\
\kappa_3(x_1, x_4, x_5) &= (\kappa_1(x_1, x_2, x_3) \triangleright \kappa_2(x_2, x_3, x_4))^{\downarrow\{1,4\}} \triangleright \pi_3(x_1, x_4, x_5).
\end{aligned}$$

The first step is trivial and also the second one is quite simple:

$$\begin{aligned}
\kappa_2(x_2, x_3, x_4) &= \kappa_1(x_2, x_3) \triangleright \pi_2(x_2, x_3, x_4) \\
&= \pi_1(x_2, x_3) \triangleright \pi_2(x_2, x_3, x_4) \\
&= \pi_2(x_2, x_3) \triangleright \pi_2(x_2, x_3, x_4) \\
&= \pi_2(x_2, x_3, x_4)
\end{aligned}$$

However, the third step is much more complicated. Let us compute κ_3

at point $(x_1 = 0, x_4 = 0, x_5 = 0)$. *Since*

$$\begin{aligned}(\kappa_1 \triangleright \kappa_2)^{\downarrow\{1,4\}}(x_1 = 0, x_4 = 0) &= (\kappa_1 \triangleright \kappa_2)(x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 0) \\ &\quad +(\kappa_1 \triangleright \kappa_2)(x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 0) \\ &\quad +(\kappa_1 \triangleright \kappa_2)(x_1 = 0, x_2 = 1, x_3 = 0, x_4 = 0) \\ &\quad +(\kappa_1 \triangleright \kappa_2)(x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 0)\end{aligned}$$

we have to compute $\kappa_1 \triangleright \kappa_2$ *at these four points:*

$$\begin{aligned}(\kappa_1 \triangleright \kappa_2)(0,0,0,0) &= \frac{\pi_1(0,0,0)\pi_2(0,0,0)}{\pi_2(x_2 = 0, x_3 = 0)} = \frac{0.18 \times 0.20}{0.30} = 0.12, \\ (\kappa_1 \triangleright \kappa_2)(0,0,1,0) &= \frac{\pi_1(0,0,1)\pi_2(0,1,0)}{\pi_2(x_2 = 0, x_3 = 1)} = \frac{0.04 \times 0.06}{0.10} = 0.024, \\ (\kappa_1 \triangleright \kappa_2)(0,1,0,0) &= \frac{\pi_1(0,1,0)\pi_2(1,0,0)}{\pi_2(x_2 = 1, x_3 = 0)} = \frac{0.14 \times 0.16}{0.28} = 0.08, \\ (\kappa_1 \triangleright \kappa_2)(0,1,0,0) &= \frac{\pi_1(0,1,1)\pi_2(1,1,0)}{\pi_2(x_2 = 1, x_3 = 1)} = \frac{0.18 \times 0.14}{0.32} = 0.07875.\end{aligned}$$

From this we get

$$(\kappa_1 \triangleright \kappa_2)^{\downarrow\{1,4\}}(x_1 = 0, x_4 = 0) = 0.30275,$$

and therefore

$$\begin{aligned}\kappa_3(0,0,0) &= (\kappa_1 \triangleright \kappa_2)^{\downarrow\{1,4\}}(0,0) \triangleright \pi_3(0,0,0) = \frac{(\kappa_1 \triangleright \kappa_2)^{\downarrow\{1,4\}}(0,0)\pi_3(0,0,0)}{\pi_3(x_1 = 0, x_4 = 0)} \\ &= \frac{0.30275 \times 0.20}{0.28} = 0.21625.\end{aligned}$$

Table 4.3: κ_3

κ_3	$x_1 = 0$		$x_1 = 1$	
	$x_4 = 0$	$x_4 = 1$	$x_4 = 0$	$x_4 = 1$
$x_5 = 0$	0.2162	0.0730	0.1838	0.0676
$x_5 = 1$	0.0865	0.1642	0.0734	0.1352

The distribution κ_3 *is in Table 4.3 (notice that* π_3 *and* κ_3 *do not differ too much; their Kullback-Leibler divergence* $Div(\pi_3 \| \kappa_3) = 0.004239$*).* $\Diamond$

Having a generating sequence, one should apply Lemma 4.1.1 to verify whether the sequence is perfect or not. It is rather a strong property and it may happen that its 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, when verification of perfectness is simple, are described in Lemma 4.1.3. It is, in fact, just a reformulation of a classical result of Kellerer [25] into the language of this text. To formulate it, let us recall an important concept that is not new to the reader familiar with decomposable models (see e.g. [10]).

Definition 4.1.2 A sequence of sets $K_1, K_2, \ldots, K_n$ is said to meet *running intersection property* (RIP, in the sequel), if

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

In the field of graphical Markov models the notion of running intersection property is one of the most important concepts. Therefore, it is not surprising that we shall also meet with it several times in the following text. Then, we shall need the following interesting property.

Lemma 4.1.2 *If a sequence of sets $K_1, K_2, \ldots, K_n$ meets RIP, then for each $\ell \in \{1, 2, \ldots, n\}$ there exists a permutation $i_1, i_2, \ldots, i_n$ such that $\ell = i_1$ and $K_{i_1}, K_{i_2}, \ldots, K_{i_n}$ meets RIP, too.*

Proof. To prove this we shall employ the mathematical induction. Since the property is trivial for $n = 2$, it is enough to show that it holds also for sequences of length n under the assumption that it has been proven for sequences of length $n - 1$.

Consider an arbitrary sequence $K_1, \ldots, K_n$ meeting RIP. Choose any $\ell \in \{1, 2, \ldots, n\}$. If $\ell < n$ we can get the required permutation in the following way: Since also $K_1, \ldots, K_{n-1}$ meets RIP we can find, thanks to the inductive assumption, a permutation $i_1, \ldots, i_{n-1}$, for which $\ell = i_1$ and $K_{i_1}, \ldots, K_{i_{n-1}}$ meets RIP. Then $K_{i_1}, \ldots, K_{i_{n-1}}, K_n$ must meet this property, too (for the last term of the sequence it follows from the fact that RIP holds for $K_1, \ldots, K_n$).

Now, we have to show that there exists a permutation $i_1, i_2, \ldots, i_n$ of the required properties with $n = i_1$. For this we shall take advantage of the fact that there exists $j \in \{1, 2, \ldots, n-1\}$ such that $K_n \cap (K_1 \cup \ldots \cup K_{n-1}) \subseteq K_j$.

$j < n$ and therefore, due to the assumption of induction, there exists a permutation $K_{i_1}, \ldots, K_{i_{n-1}}$ starting with $K_{i_1} = K_j$, for which RIP holds true. Let us show, now, that a sequence $K_n, K_{i_1}, \ldots, K_{i_{n-1}}$ meets RIP, too. Consider $k \in \{2, \ldots, n-1\}$. For this there exists $\hat{k} < k$ such that

$$K_{i_k} \cap (K_{i_1} \cup \ldots \cup K_{i_{k-1}}) \subseteq K_{i_{\hat{k}}}.$$

However, since $K_n \cap K_{i_k} \subseteq K_j = K_{i_1}$ we get also that

$$K_{i_k} \cap (K_n \cup K_{i_1} \cup \ldots \cup K_{i_{k-1}}) \subseteq K_{i_{\hat{k}}}$$

holds true. Thus we have shown that for each $\ell \in \{1, 2, \ldots, n\}$ there exists a permutation $i_1, i_2, \ldots, i_n$ such that $\ell = i_1$ and $K_{i_1}, K_{i_2}, \ldots, K_{i_n}$ meets RIP. □

Lemma 4.1.3 *If $\pi_1, \pi_2, \ldots, \pi_n$ is a sequence of pairwise consistent oligodimensional probability distributions such that $K_1, \ldots, K_n$ 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 $\pi_1, \pi_2, \ldots, \pi_m$, for which $K_1, K_2, \ldots, K_m$ meet RIP and $\pi_1, \pi_2, \ldots, \pi_{m-1}$ is perfect. Thus to show that $\pi_1, \pi_2, \ldots, \pi_m$ is perfect it is enough to show that $(\pi_1 \triangleright \ldots \triangleright \pi_{m-1})$ and π_m are consistent.

Since $K_1, \ldots, K_n$ meets RIP, $K_m \cap (K_1 \cup \ldots \cup K_{m-1})$ must be part of K_k for some $k \leq m-1$. Therefore, $K_m \cap (K_1 \cup \ldots \cup K_{m-1}) = K_m \cap K_k$. The assumption of induction says that, due to Theorem 4.1.1, all π_ℓ $(1 \leq \ell < m)$ are marginal to $\pi_1 \triangleright \ldots \triangleright \pi_{m-1}$ and thus

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

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

An arbitrary perfect sequence $\pi_1, \pi_2, \ldots, \pi_n$ (with $n > 1$) can always be reordered in the way that its permutation $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_n}$ is also perfect.

Trivially, if $\pi_1, \pi_2, \pi_3, \ldots, \pi_n$ is perfect then $\pi_2, \pi_1, \pi_3, \ldots, \pi_n$ 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 showing that perfect sequence models always achieve maximum entropy, in a sense (it is, in fact, a generalization of Theorem 3.2.1).

Theorem 4.1.3 *Denote $\Xi = \{\pi_1, \pi_2, \ldots, \pi_n\}$ a system of oligodimensional probability distributions. If the sequence $\pi_1, \pi_2, \ldots, \pi_n$ is perfect then*

$$H(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n) = \sum_{i=1}^{n} H(\pi_i) - \sum_{i=2}^{n} H(\pi_i(x_{K_i \cap (K_1 \cup \ldots \cup K_{i-1})})) \geq H(\kappa)$$

for any

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

Proof To make the following computations more transparent we will use the following notation: for each $i = 1, \ldots, n$ set K_i is split into two disjoint parts

$$R_i = K_i \setminus (K_1 \cup \ldots \cup K_{i-1}), \qquad S_i = K_i \cap (K_1 \cup \ldots \cup K_{i-1}).$$

(Naturally, $R_1 = K_1$ and $S_1 = \emptyset$.) Using this, we can compute (the summation is performed only over those points $x \in \mathbf{X}_{K_1 \cup \ldots \cup K_n}$ for which the respective probabilities are positive)

$$\begin{aligned}
H(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n) &= -\sum_{x} (\pi_1 \triangleright \ldots \triangleright \pi_n)(x) \log(\pi_1 \triangleright \ldots \triangleright \pi_n)(x) \\
&= -\sum_{x} (\pi_1 \triangleright \ldots \triangleright \pi_n)(x) \log \prod_{i=1}^{n} \pi_i(x_{R_i} | x_{S_i}) \\
&= -\sum_{i=1}^{n} \sum_{x} (\pi_1 \triangleright \ldots \triangleright \pi_n)(x) \log \pi_i(x_{R_i} | x_{S_i}) \\
&= -\sum_{i=1}^{n} \sum_{x_{K_i}} (\pi_1 \triangleright \ldots \triangleright \pi_n)(x_{K_i}) \log \pi_i(x_{R_i} | x_{S_i}) \\
&= -\sum_{i=1}^{n} \sum_{x_{K_i}} \pi_i(x_{K_i}) \log \pi_i(x_{R_i} | x_{S_i}),
\end{aligned}$$

where the last modification is possible because $\pi_1, \ldots, \pi_n$ is perfect and therefore π_i is a marginal distribution of $\pi_1 \triangleright \ldots \triangleright \pi_n$ (notice that when changing summation over x to summation over x_{K_i} we have employed again the standard trick described in Remark 3.2.2 – see page 43). Therefore, using point 3 of Lemma 2.3.1 we get

$$\begin{aligned} H(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n) &= \sum_{i=1}^{n} H(\pi_i(x_{R_i}|x_{S_i})) \\ &= \sum_{i=1}^{n} H(\pi_i(x_{K_i})) - \sum_{i=2}^{n} H(\pi_i(x_{S_i})) \\ &= \sum_{i=1}^{n} H(\pi_i) - \sum_{i=2}^{n} H(\pi_i(x_{K_i \cap (K_1 \cup \ldots \cup K_{i-1})})). \end{aligned}$$

(The second summation starts from $i = 2$ because $S_1 = \emptyset$.)

Let us, now, compute the Shannon entropy of an arbitrary distribution $\kappa \in \Pi^{(K_1 \cup \ldots \cup K_n)}(\Xi)$. In this, we shall use the fact that, since $R_1, R_2, \ldots, R_n$ forms a partition of $K_1 \cup K_2 \cup \ldots \cup K_n$, any distribution $\kappa \in \Pi^{(K_1 \cup \ldots \cup K_n)}$ can be expressed as

$$\kappa(x) = \prod_{i=1}^{n} \kappa(x_{R_i}|x_{R_1 \cup \ldots \cup R_{i-1}}) = \prod_{i=1}^{n} \kappa(x_{R_i}|x_{K_1 \cup \ldots \cup K_{i-1}}).$$

Therefore

$$\begin{aligned} H(\kappa) &= -\sum_{x} \kappa(x) \log \kappa(x) = -\sum_{x} \kappa(x) \log \prod_{i=1}^{n} \kappa(x_{R_i}|x_{K_1 \cup \ldots \cup K_{i-1}}) \\ &= \sum_{i=1}^{n} \sum_{x} \kappa(x) \log \kappa(x_{R_i}|x_{K_1 \cup \ldots \cup K_{i-1}}) \\ &= \sum_{i=1}^{n} H(\kappa(x_{R_i}|x_{K_1 \cup \ldots \cup K_{i-1}})) \\ &= \sum_{i=1}^{n} \Big(H(\kappa(x_{R_i}|x_{S_i})) - MI_\kappa(X_{R_i}; X_{(K_1 \cup \ldots \cup K_{i-1}) \setminus S_i}|X_{S_i}) \Big) \\ &\leq H(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n). \end{aligned}$$

In the last but one modification we used the relation between conditional entropy and conditional mutual information presented in point 5 of Lemma 2.3.1 (see page 33), and the last inequality follows from nonnegativity of a mutual information and the assumption that $\pi_i(xK_i) = \kappa_i(xK_i)$ for all $i = 1, \ldots, n$.

□

Remark 4.1.4 Let us note that a more elegant way how to prove the preceding assertion can be based on properties of the Iterative proportional fitting procedure, which will be studied in Section 5.4. We presented above a rather technical proof, which does not require knowledge introduced in the next sections. ○

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

Theorem 4.1.4 *If a sequence* $\pi_1, \pi_2, \ldots, \pi_n$ *and its permutation* π_{i_1}, π_{i_2}, $\ldots, \pi_{i_n}$ *are both perfect then*

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

Proof Applying previous Theorem 4.1.3 to both these sequences we see that

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

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

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

□

Remark 4.1.5 Theorem 4.1.3 is only an implication: if there exists a perfect sequence formed by the distributions from $\Xi = \{\pi_1, \pi_2, \ldots, \pi_n\}$ then it achieves the maximum Shannon entropy among the distributions from $\Pi^{(K_1 \cup \ldots \cup K_n)}(\Xi)$. However, as it can be seen from the following example, it does not mean that the maximum entropy distribution from $\Pi^{(K_1 \cup \ldots \cup K_n)}(\Xi)$ must be a compositional model. ○

Example 4.1.4 *It is not difficult to show that for* 2*-dimensional distributions from Table 4.4 there exists the only common extension – the distribution from Table 4.5. (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 equaling* $\frac{2}{6}$ *– those, which are positive in Table 4.5 – must equal* $\frac{1}{6}$*.) Therefore, this extension is also the maximum entropy extension.*

Since all the considered 2*-dimensional distributions are positive, all possible compositional models constructed from them must also be positive, which means that the distribution from Table 4.5 cannot be obtained as a*

Table 4.4: 2-dimensional distributions

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

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

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

compositional model of distributions from Table 4.4. ◊

Table 4.5: Extension of the distributions from Table 4.4

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

Theorem 4.1.3 gives an instruction how to compute entropy of a distribution represented by perfect sequence model. The next assertion of this section presents an instruction how to compute informational content of these distributions.

Theorem 4.1.5 *If the sequence $\pi_1, \pi_2, \ldots, \pi_n$ is perfect then*

$$IC(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n) = \sum_{i=1}^{n} IC(\pi_i) - \sum_{i=2}^{n} IC(\pi_i(x_{K_i \cap (K_1 \cup \ldots \cup K_{i-1})})).$$

Proof In the proof, property 7 of Lemma 2.3.1 presented on page 33, Theorem 4.1.3, as well as the fact that all π_i are marginals of $\pi_1 \triangleright \ldots \triangleright \pi_n$, and eventually property 9 of Lemma 2.3.1 will be used.

$$
\begin{aligned}
IC(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n) &= \sum_{j \in K_1 \cup \ldots \cup K_n} H((\pi_1 \triangleright \ldots \triangleright \pi_n)(x_j)) - H(\pi_1 \triangleright \ldots \triangleright \pi_n) \\
&= \left(\sum_{i=1}^{n} \sum_{j \in K_i \setminus (K_1 \cup \ldots \cup K_{i-1})} H((\pi_1 \triangleright \ldots \triangleright \pi_n)(x_j)) \right) \\
&\qquad - \left(\sum_{i=1}^{n} H(\pi_i) - \sum_{i=2}^{n} H(\pi_i^{\downarrow K_i \cap (K_1 \cup \ldots \cup K_{i-1})}) \right) \\
&= - \sum_{i=1}^{n} \left(H(\pi_i) - \sum_{j \in K_i \setminus (K_1 \cup \ldots \cup K_{i-1})} H((\pi_1 \triangleright \ldots \triangleright \pi_n)(x_j)) \right) \\
&\qquad + \sum_{i=2}^{n} H(\pi_i^{\downarrow K_i \cap (K_1 \cup \ldots \cup K_{i-1})}) \\
&= - \sum_{i=1}^{n} \left(H(\pi_i) - \sum_{j \in K_i} H((\pi_1 \triangleright \ldots \triangleright \pi_n)(x_j)) \right) \\
&\qquad - \sum_{i=2}^{n} \sum_{j \in K_i \cap (K_1 \cup \ldots \cup K_{i-1})} H((\pi_1 \triangleright \ldots \triangleright \pi_n)(x_j)) \\
&\qquad + \sum_{i=2}^{n} H(\pi_i^{\downarrow K_i \cap (K_1 \cup \ldots \cup K_{i-1})}) \\
&= - \sum_{i=1}^{n} \left(H(\pi_i) - \sum_{j \in K_i} H(\pi_i(x_j)) \right) \\
&\qquad + \sum_{i=2}^{n} \left(\sum_{j \in K_i \cap (K_1 \cup \ldots \cup K_{i-1})} H(\pi_i(x_j)) - H(\pi_i^{\downarrow K_i \cap (K_1 \cup \ldots \cup K_{i-1})}) \right) \\
&= \sum_{i=1}^{n} IC(\pi_i) - \sum_{i=2}^{n} IC(\pi_i^{\downarrow K_i \cap (K_1 \cup \ldots \cup K_{i-1})}) \\
&= \sum_{i=1}^{n} IC(\pi_i(x_{K_i \setminus (K_1 \cup \ldots \cup K_{i-1})} | x_{K_i \cap (K_1 \cup \ldots \cup K_{i-1})})).
\end{aligned}
$$

□

Remark 4.1.6 Computation of the expression

$$JC(\pi_1,\ldots,\pi_n) = \sum_{i=1}^{n} IC(\pi_i) - \sum_{i=2}^{n} IC(\pi_i^{\downarrow K_i\cap(K_1\cup\ldots\cup K_{i-1})}) = \sum_{i=1}^{n} IC(\pi_i(x_{K_i\setminus(K_1\cup\ldots\cup K_{i-1})}|x_{K_i\cap(K_1\cup\ldots\cup K_{i-1})})) \quad (4.1)$$

is simple because each term of the summation can be computed just from a respective (oligodimensional) distribution π_i. However, it is important to realize that it equals information content of a generating sequence only for perfect sequences. If the sequence is not perfect then computation of the informational content of the respective model is usually more complex (the process requires computation of a system of marginal distributions). Roughly speaking, from the algorithmical point of view it is as complex as realization of the perfectization procedure described in Theorem 4.1.2. Moreover, for non-perfect sequences the value of expression (4.1) may be both higher and lower than the corresponding informational content. This can be easily shown by the reader according to the following hint: Considering distributions $\pi_1(x_1), \pi_2(x_1)$ and $\kappa_1(x_1,x_2), \kappa_2(x_1,x_2)$ from Table 4.6 it is obvious that π_1, κ_1 and π_2, κ_2 are perfect sequences such that

$$\kappa_1 = \pi_1 \triangleright \kappa_1 = \pi_1 \triangleright \kappa_2,$$
$$\kappa_2 = \pi_2 \triangleright \kappa_2 = \pi_2 \triangleright \kappa_1,$$

and that $JC(\pi_1,\kappa_1) = JC(\pi_2,\kappa_1)$ and $JC(\pi_2,\kappa_2) = JC(\pi_1,\kappa_2)$. The required inequalities are then directly received from the fact that $IC(\kappa_1) = 0.025 < IC(\kappa_2) = 0.028$. ○

Table 4.6: Distributions $\pi_1(x_1), \pi_2(x_1), \kappa_1(x_1,x_2), \kappa_2(x_1,x_2)$

	π_1	κ_1		π_2	κ_2	
		$x_2 = 0$	$x_2 = 1$		$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	0.1	0.08	0.02	0.9	0.72	0.18
$x_1 = 1$	0.9	0.45	0.45	0.1	0.05	0.05

4.2 Commutable sets

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

Definition 4.2.1 A set of distributions $\{\pi_1, \pi_2, \ldots, \pi_n\}$ is said to be *commutable* if $\pi_1 \triangleleft \pi_2 \triangleleft \ldots \triangleleft \pi_n$ is defined and

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

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

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 likewise for perfect sequences, for commutable sets it holds that all π_i's are marginal to $\pi_1 \triangleleft \ldots \triangleleft \pi_n$.

Lemma 4.2.1 *If $\{\pi_1, \pi_2, \ldots, \pi_n\}$ is a commutable set of probability distributions then*

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

which means that all the distributions $\{\pi_1, \pi_2, \ldots, \pi_n\}$ are marginals of $\pi_1 \triangleleft \pi_2 \triangleleft \ldots \triangleleft \pi_n$.

Proof. The assertion is a direct consequence of the facts that for any $j \in \{1, 2, \ldots, n\}$ there are permutations $i_1, \ldots, i_n$ in which index $j = i_n$ is the last one, and thus π_j is a marginal of the distribution

$$\pi_{i_1} \triangleleft \pi_{i_2} \triangleleft \ldots \triangleleft \pi_{i_n}.$$

□

Remark 4.2.1 The reader should realize that the assertion expresses only a necessary condition; 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 $\pi_1 \triangleleft \pi_2 \triangleleft \ldots \triangleleft \pi_n$. ○

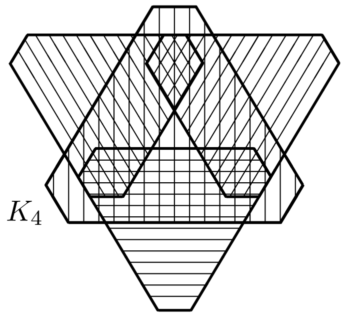


Figure 4.1: Star-like system of sets

Example 4.2.1 *Let us present a nontrivial example of a generating sequence, which is not perfect, and yet its distributions form a commutable set. Consider four distributions* $\pi_1, \pi_2, \pi_3, \pi_4$ *for which*

$$\begin{aligned} K_1 &= \{1,2,4\}, \\ K_2 &= \{2,3,5\}, \\ K_3 &= \{1,3,6\}, \\ K_4 &= \{1,2,3\}. \end{aligned}$$

This situation is illustrated in Figure 4.1.

Let π_4 *be the distribution from Table 4.7, and all three remaining distributions* $\pi_1(x_1,x_2,x_4), \pi_2(x_2,x_3,x_5), \pi_3(x_1,x_3,x_6)$ *be uniform distributions of the respective sets of variables. The reader can immediately see that the distributions are pairwise consistent because all their 2-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 6-dimensional distribution. The same holds for any permutation of these three distributions. Therefore*

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

holds true for any permutation, for which $i_4 = 4$*. Now, applying twice Lemma 3.4.4 and once Lemma 3.2.2 we get*

$$\begin{aligned} \pi_{i_1} \triangleleft \pi_{i_2} \triangleleft \pi_{i_3} \triangleleft \pi_{i_4} &= \pi_{i_1} \triangleleft \pi_{i_2} \triangleleft \pi_{i_4} \triangleleft \pi_{i_3} = \pi_{i_1} \triangleleft \pi_{i_4} \triangleleft \pi_{i_2} \triangleleft \pi_{i_3} \\ &= \pi_{i_4} \triangleleft \pi_{i_1} \triangleleft \pi_{i_2} \triangleleft \pi_{i_3}. \end{aligned}$$

Table 4.7: 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	$\frac{1}{4}$

Therefore, all permutations yield the same 6-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, 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

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \pi_4 = \pi_1 \triangleright \pi_2 \triangleright \pi_3.$$

Thus, $\pi_1, \pi_2, \pi_3, \pi_4$ is not perfect because π_4 is not marginal of $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4$. $\diamondsuit$

Let us repeat once more that it may happen that all the distributions π_i are marginals of $\pi_1 \triangleleft \pi_2 \triangleleft \ldots \triangleleft \pi_n$ and still the sequence may be neither perfect (Example 4.2.1), nor the corresponding set commutable (sequence $\pi_1, \pi_2, \pi_3, \pi_5, \pi_4$ of Example 4.2.3). In what follows we will present some sufficient conditions describing special situations of perfect sequences and commutable sets, as well as examples illustrating the described theoretical properties.

Lemma 4.2.2 *Let $\{\pi_1, \pi_2, \ldots, \pi_n\}$ be a commutable set of distributions. When $K_{i_1}, K_{i_2}, \ldots, K_{i_n}$ meets RIP then the sequence $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_n}$ is perfect.*

Proof. Lemma 4.2.1 guarantees the pairwise consistency of the distributions from $\{\pi_1, \pi_2, \ldots, \pi_n\}$. Therefore the sequence $\pi_{i_1}, \ldots, \pi_{i_n}$ is perfect due to Lemma 4.1.3. $\square$

As we shall see in the following theorem, it is not surprising that the set of distributions from the Example 4.2.1 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 4.2.2 A system of sets $\{K_1, K_2, \ldots, K_n\}$ is called *star-like* if their exists and index $\ell \in \{1, 2, \ldots, n\}$ such that for any couple of different indices $i, j \in \{1, 2, \ldots, n\}$ $K_i \cap K_j \subseteq K_\ell$. The set K_ℓ is called a *centre* of the system.

For examples of star-like systems of sets see Figures 4.1, 4.2 and 4.3. 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 system of sets $K_1, K_2, \ldots, K_n$ is star-like then the distributions $\pi_1, \pi_2, \ldots, \pi_n$, if they are pairwise consistent, can be reordered into a perfect sequence.

Theorem 4.2.1 *If for a set of pairwise consistent distributions* $\{\pi_1, \pi_2, \ldots, \pi_n\}$ *the system* $\{K_1, K_2, \ldots, K_n\}$ *is star-like then* $\{\pi_1, \pi_2, \ldots, \pi_n\}$ *is commutable.*

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

$$\pi_{i_1} \triangleleft \pi_{i_2} \triangleleft \ldots \triangleleft \pi_{i_n} = \pi_{j_1} \triangleleft \pi_{j_2} \triangleleft \ldots \triangleleft \pi_{j_n}.$$

Without loss of generality assume that $K_1 \supseteq K_i \cap K_j$ for all couples of different $i, j \in \{1, 2, \ldots, n\}$ and consider an arbitrary permutation $i_1, i_2, \ldots i_n$. Let $1 = i_k$. Apparently, if $k \leq 2$ the sequence $K_1, K_2, \ldots, K_n$ meets RIP. If $k > 2$ we can apply Lemma 3.4.4 ($(k-2)$-times) getting

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

In both cases we see that $\pi_{i_1} \triangleleft \ldots \triangleleft \pi_{i_n}$ equals $\pi_{j_1} \triangleleft \ldots \triangleleft \pi_{j_n}$, where the sequence $K_{j_1}, \ldots, K_{j_n}$ meets RIP. Therefore, due to Lemma 4.1.3 and Theorem 4.1.4, for all permutations $i_1, i_2, \ldots i_n$ the expressions $\pi_{i_1} \triangleleft \ldots \triangleleft \pi_{i_n}$ define the same multidimensional distribution, which means that $\{\pi_1, \pi_2, \ldots, \pi_n\}$ is commutable. □

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

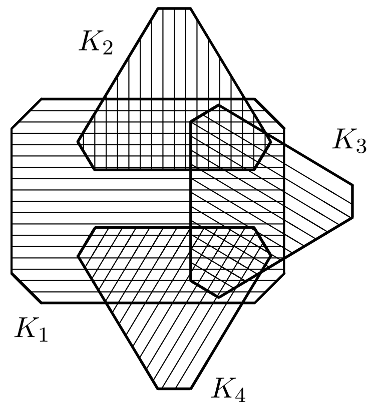


Figure 4.2: Strongly decomposable star-like system of sets

Definition 4.2.3 A system of sets $\{K_1, K_2, \ldots, K_n\}$ is called *strongly decomposable* if each its subset can be ordered to meet RIP.

Example 4.2.2 *First notice that a star-like system in Figure 4.1 is not strongly decomposable. As shown in Example 4.2.1, the system can be ordered to meet RIP, but there is a subsystem (the reader can easily show that in this case only one) – K_2, K_3, K_4 – 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 4.2 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 $K_1, K_2, K_3, \ldots, K_n$ and its "suffix" $K_2, K_3, \ldots, K_n$ meet RIP. Let us show in this example that, though 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 4.2. The sequence K_1, K_2, K_4, K_3 meets RIP whilst sequence K_2, K_4, K_3 not. On the other hand, sequence K_3, K_2, K_1 meets RIP and sequence K_4, K_3, K_2, K_1 not, though there exists a sequence meeting RIP, which starts with K_4 (it is a sequence K_4, K_1, K_2, K_3). $\diamondsuit$

Lemma 4.2.3 *For any strongly decomposable system of sets $K_1, K_2, \ldots, K_n$ and any $\ell \in \{1, \ldots, n\}$ there exists a permutation of indices $i_1, i_2, \ldots, i_n$ such that $i_1 = \ell$ and both the sequences $K_{i_1}, K_{i_2}, K_{i_3}, \ldots, K_{i_n}$ and $K_{i_2}, K_{i_3}, \ldots, K_{i_n}$ 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 4.1.2. Let it be $K_1, K_2, \ldots, K_n$.

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

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

If there are more such j_n take the highest one and, in case $j_n > 1$, put K_n and K_{j_n} into one cluster. If $j_n = 1$ do nothing. Then consider $\ell = n-1, n-2, \ldots, 2$ and at each step find (the highest) 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 have a partition of sets $K_2, \ldots, K_n$ into several, let us say m, clusters. The set with the lowest index in a cluster will be, in this proof, called *cluster representative*. Having m clusters we have a system of m representatives, which can be ordered to meet RIP, because we assume that $K_1, K_2, \ldots, K_n$ is strongly decomposable. Denote the RIP ordering of these cluster representatives $K_{j_1}, K_{j_2}, \ldots, K_{j_m}$. Similarly, we will order also the sets in each cluster to meet RIP; each of this orderings must start with the cluster representative. Let such an ordering of sets from the k-th cluster be $K_{j_k}, K_{j_{k,2}}, K_{j_{k,3}}, \ldots, K_{j_{k,r(k)}}$.

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

$$\begin{array}{l}
K_1, \\
K_{j_1}, K_{j_{1,2}}, K_{j_{1,3}}, \ldots, K_{j_{1,r(1)}}, \\
K_{j_2}, K_{j_{2,2}}, K_{j_{2,3}}, \ldots, K_{j_{2,r(2)}}, \\
K_{j_3}, K_{j_{3,2}}, K_{j_{3,3}}, \ldots, K_{j_{3,r(3)}}, \\
\quad \vdots \\
K_{j_m}, K_{j_{m,2}}, K_{j_{m,3}}, \ldots, K_{j_{m,r(m)}}.
\end{array}$$

What can be said about this construction? Apparently, the system $K_1, K_{j_1}, K_{j_2}, \ldots, K_{j_m}$ is a star-like system. Another property, which can be seen from the way how 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 $K_{i_2}, K_{i_3}, \ldots, K_{i_n}$ and any $k \in \{2, 3, \ldots, n\}$. 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 were ordered to meet RIP (other "non-representative" sets cannot interfere with this fact, because intersection of sets from different clusters are contained in intersection 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 was ordered to meet RIP (again, sets from other clusters cannot interfere with it, because intersection of sets from different clusters are contained in intersection of the respective cluster representatives).

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

Remark 4.2.2 At the beginning of Example 4.2.2 we mentioned that a star-like system in Figure 4.1 is not strongly decomposable. The same holds also for the larger system in Figure 4.3. It is easy to show that all star-like systems $K_1, \ldots, K_n$ can be ordered in the way that the full ordering and its "suffix" ordering of length $n - 1$ meet RIP. Nevertheless, Lemma 4.2.3 guarantees that for strongly decomposable systems there are many such ordering; for each $i = 1, \ldots, n$ 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 4.1 and 4.3 no sequence starting with the centre K_4 and K_1, respectively, meets this condition. Therefore the proof of the following Theorem 4.2.2 cannot be applied to Theorem 4.2.1. ○

Now, we are ready to prove an important assertion expressing the other sufficient condition for a system of distribution to be commutable.

Theorem 4.2.2 *If for a set of pairwise consistent distributions* $\{\pi_1, \pi_2, \ldots, \pi_n\}$ *the system* $\{K_1, K_2, \ldots, K_n\}$ *is strongly decomposable then the considered set of distributions is commutable.*

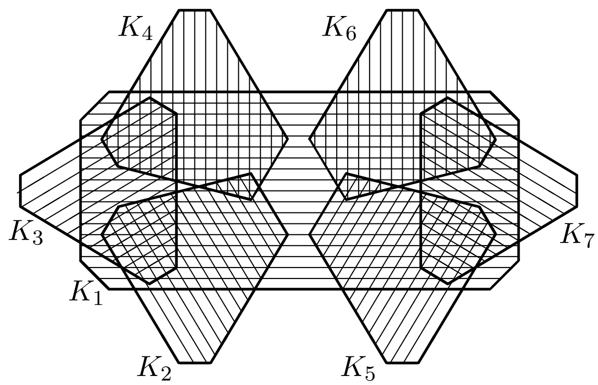


Figure 4.3: Star-like system of 7 sets

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

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

for some permutation $j_1, j_2, \ldots, j_n$, for which $K_{j_1}, K_{j_2}, \ldots, K_{j_n}$ meets RIP. Then the commutability of $\{\pi_1, \pi_2, \ldots, \pi_n\}$ will be a direct consequence of Lemma 4.1.3 and Theorem 4.1.4.

Consider a permutation $i_1, i_2, \ldots, i_n$. If $n = 2$ then, due to Lemma 3.2.2, $\pi_1 \lhd \pi_2 = \pi_1 \rhd \pi_2$ 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. Among them, due to Lemma 4.2.3, there exists a permutation $i_n, j_1, \ldots, j_{n-1}$, such that $K_{i_n}, K_{j_1}, \ldots, K_{j_{n-1}}$ and $K_{j_1}, \ldots, K_{j_{n-1}}$ meet RIP. We shall show that

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

Applying the assumption of mathematical induction, we get

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

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

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

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

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

or

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

hold true. Therefore, applying either Lemma 3.4.7 or Corollary 3.4.2, respectively, we get

$$\begin{aligned}\pi_{i_n} \triangleright (\pi_{j_1} \triangleright \pi_{j_2} \triangleright \ldots \triangleright \pi_{j_{n-1}}) &= \pi_{i_n} \triangleright (\pi_{j_1} \triangleright \ldots \triangleright \pi_{j_{n-2}}) \triangleright \pi_{j_{n-1}} \\ &= \pi_{i_n} \triangleright ((\pi_{j_1} \triangleright \ldots \triangleright \pi_{j_{n-3}}) \triangleright \pi_{j_{n-2}}) \triangleright \pi_{j_{n-1}}.\end{aligned}$$

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

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

In this way we can eliminate all brackets getting eventually that

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

which finishes the proof. □

Example 4.2.3 *In this example it will be shown that the assumption in the previous theorem cannot be weakened in the sense that instead of strong decomposability of a system* $\{K_1, K_2, \ldots, K_n\}$ *one would assume just an existence of its ordering meeting RIP.*

Let us consider 5 *oligodimensional distributions* $\pi_1, \ldots, \pi_5$ *with a structure of variables as depicted in Figure 4.4. For the purpose of this example it will suffice to consider only* 5 *binary variables which are arguments of the distributions* $\pi_1, \ldots, \pi_5$ *according to the following pattern:*

$$\begin{aligned} K_1 &= \{1, 2, 4\} \\ K_2 &= \{2, 3\} \\ K_3 &= \{1, 3\} \\ K_4 &= \{4, 5\} \\ K_5 &= \{1, 2, 3\} \end{aligned}$$

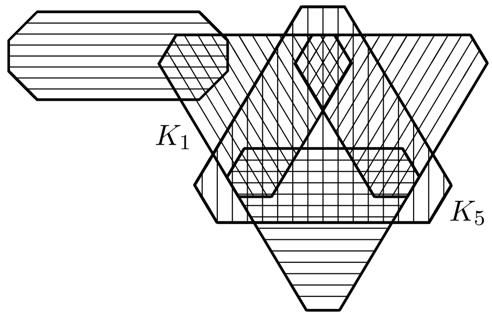


Figure 4.4: System of sets that can be ordered to meet RIP

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

Let us consider distributions, whose values are in Tables 4.8 and 4.9.

Notice that because x_4 is an argument of neither π_2 nor π_3 it is possible to "isolate" variable x_4 from the expression $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3$ in the following sense

$$\begin{aligned}\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 &= \frac{\pi_1\pi_2\pi_3}{\pi_1(x_2)(\pi_1 \triangleleft \pi_2)(x_1, x_2)} = \frac{(\pi_1(x_1, x_2)\pi_2\pi_3)\pi_1(x_4|x_1, x_2)}{\pi_1(x_2)(\pi_1(x_1, x_2) \triangleleft \pi_2)(x_1, x_2)} \\ &= (\pi_1(x_1, x_2) \triangleleft \pi_2 \triangleleft \pi_3)\pi_1(x_4|x_1, x_2),\end{aligned}$$

and therefore

$$(\pi_1 \triangleleft \pi_2 \triangleleft \pi_3)(x_1, x_2, x_4) = (\pi_1 \triangleleft \pi_2 \triangleleft \pi_3)(x_1, x_2)\pi_1(x_4|x_1, x_2).$$

Thus we see that since $(\pi_1 \triangleleft \pi_2 \triangleleft \pi_3)(x_1, x_2) \neq \pi_1(x_1, x_2)$ (cf. Table 4.11) that also $(\pi_1 \triangleleft \pi_2 \triangleleft \pi_3)(x_4) \neq \pi_1(x_4)$. (For the distribution

$$(\pi_1 \triangleleft \pi_2 \triangleleft \pi_3)(x_1, x_2, x_4) = (\pi_1 \triangleleft \pi_2 \triangleleft \pi_3)(x_1, x_2)\pi_1(x_4|x_1, x_2)$$

see Table 4.12.)

Computing now

$$(\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4)(x_1, x_2, x_4) = (\pi_1 \triangleleft \pi_2 \triangleleft \pi_3)(x_1, x_2)\pi_1(x_4|x_1, x_2) \triangleleft \pi_4(x_4)$$

Table 4.8: 2-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	$\frac{1}{2}$

we get the distribution in Table 4.13. From this distribution we see that

$$(\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4)(x_4|x_1,x_2) \neq \pi_1(x_4|x_1,x_2)$$

and therefore we are not surprised that also

$$\begin{aligned}(\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4 \triangleleft \pi_5)(x_4) &= \sum_{x_1,x_2} (\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4)(x_4|x_1,x_2)\pi_5(x_1,x_2) \\ &\neq \pi_1(x_4)\end{aligned}$$

(cf. Table 4.14 for $(\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4 \triangleleft \pi_5)(x_1,x_2,x_4)$*). This means that* π_1 *is not a marginal of* $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4 \triangleleft \pi_5$ *and therefore* $\pi_1,\pi_2,\pi_3,\pi_4,\pi_5$ *is not commutable (otherwise it would be in contradiction with Lemma 4.2.1).* ◊

Remark 4.2.3 We have shown that the considered distributions are not marginals of $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4 \triangleleft \pi_5$ (more precisely π_1 is not a marginal of this distribution), nevertheless, the reader can easily show that they are marginals of $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3 \triangleleft \pi_4 \triangleleft \pi_5 \triangleleft \pi_1$. We will recall this example in Section 5.4 where we will speak about Iterative Proportional Fitting Procedure. Commutable sets are namely those, for which this iterative procedure finishes independently of the ordering of distributions always after n steps (one cycle). ○

Table 4.9: 3-dimensional distributions

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

π_5	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$
$x_3 = 1$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$

Table 4.10: Distribution $(\pi_1 \triangleleft \pi_2)(x_1, x_2, x_3)$ of Example 4.2.3.

π_5	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{4}{18}$	$\frac{1}{18}$	$\frac{2}{18}$	$\frac{2}{18}$
$x_3 = 1$	$\frac{2}{18}$	$\frac{2}{18}$	$\frac{1}{18}$	$\frac{4}{18}$

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

Example 4.2.4 *Consider again four distributions $\pi_1, \pi_2, \pi_3, \pi_4$ defined for the same groups of variables as in Example 4.2.1:*

$$\begin{aligned} K_1 &= \{1, 2, 3\}, \\ K_2 &= \{1, 2, 4\}, \\ K_3 &= \{2, 3, 5\}, \\ K_4 &= \{1, 3, 6\}. \end{aligned}$$

Table 4.11: Distribution $(\pi_1 \triangleleft \pi_2 \triangleleft \pi_3)(x_1, x_2, x_3)$ of Example 4.2.3.

π_5	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{4}{30}$	$\frac{1}{30}$	$\frac{5}{30}$	$\frac{5}{30}$
$x_3 = 1$	$\frac{5}{30}$	$\frac{5}{30}$	$\frac{1}{30}$	$\frac{4}{30}$

Table 4.12: Distribution $(\pi_1 \triangleleft \pi_2 \triangleleft \pi_3)(x_1, x_2, x_4)$ of Example 4.2.3.

π_5	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_4 = 0$	$\frac{1}{10}$	0	$\frac{2}{10}$	$\frac{1}{10}$
$x_4 = 1$	$\frac{2}{10}$	$\frac{2}{10}$	0	$\frac{2}{10}$

This time, values of distributions $\pi_1, \pi_2, \pi_3, \pi_4$ are given in Table 4.15.

To show that $\{\pi_1, \pi_2, \pi_3, \pi_4\}$ is commutable it is enough, due to Theorem 4.2.1, 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 4.16).

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 4.2.1, all π_2, π_3 and π_4 are marginal to

$$\kappa(x_1, x_2, x_3, x_4, x_5, x_6) = \pi_2(x_1, x_2, x_4) \triangleleft \pi_3(x_2, x_3, x_5) \triangleleft \pi_4(x_1, x_3, x_6).$$

Therefore, under this assumption,

$$\kappa(x_2 = 0, x_3 = 0) = \pi_3(x_2 = 0, x_3 = 0) = \frac{1}{3}.$$

Since

$$\kappa(x_1 = 0, x_2 = 0, x_3 = 0) \leq \kappa(x_1 = 0, x_3 = 0) = \pi_4(x_1 = 0, x_3 = 0) = \frac{1}{6},$$

Table 4.13: Distribution $(\pi_1 \lhd \pi_2 \lhd \pi_3 \lhd \pi_4)(x_1, x_2, x_4)$ of Example 4.2.3.

π_5	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_4 = 0$	$\frac{21}{216}$	0	$\frac{42}{216}$	$\frac{21}{216}$
$x_4 = 1$	$\frac{44}{216}$	$\frac{44}{216}$	0	$\frac{44}{216}$

Table 4.14: Distribution $(\pi_1 \lhd \pi_2 \lhd \pi_3 \lhd \pi_4 \lhd \pi_5)(x_1, x_2, x_4)$ of Example 4.2.3.

π_5	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_4 = 0$	$\frac{42}{390}$	0	$\frac{65}{390}$	$\frac{42}{390}$
$x_4 = 1$	$\frac{88}{390}$	$\frac{65}{390}$	0	$\frac{88}{390}$

and

$$\kappa(x_1 = 1, x_2 = 0, x_3 = 0) \leq \kappa(x_1 = 1, x_2 = 0) = \pi_2(x_1 = 1, x_2 = 0) = \frac{1}{6},$$

we are sure that

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

Thus, because

$$\kappa(x_1 = 0, x_2 = 0, x_3 = 0) = \kappa(x_1 = 0, x_3 = 0) = \pi_4(x_1 = 0, x_3 = 0) = \frac{1}{6},$$

we get

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

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

$\diamondsuit$

Table 4.15: Probability distributions $\pi_1, \pi_2, \pi_3, \pi_4$

π_1	$x_1 = 0$		$x_1 = 1$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_3 = 0$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{6}$
$x_3 = 1$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$

π_2	$x_4 = 0$		$x_4 = 1$	
	$x_1 = 0$	$x_1 = 1$	$x_1 = 0$	$x_1 = 1$
$x_2 = 0$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
$x_2 = 1$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$

π_3	$x_5 = 0$		$x_5 = 1$	
	$x_3 = 0$	$x_3 = 1$	$x_3 = 0$	$x_3 = 1$
$x_2 = 0$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
$x_2 = 1$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$

π_4	$x_6 = 0$		$x_6 = 1$	
	$x_3 = 0$	$x_3 = 1$	$x_3 = 0$	$x_3 = 1$
$x_1 = 0$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$
$x_1 = 1$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$

Even in case that 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 a larger commutable set. More exactly: consider $\pi_i \in \Pi^{(K_i)}, i = 1, \ldots, n$; if both $\{\pi_1, \ldots, \pi_n\}$ and $\{\pi_1, \ldots, \pi_k\}$ (for some $1 \leq k < n$) are commutable then it may happen that

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

An example of this situation are sets $\{\pi_1, \pi_2, \pi_3, \pi_4\}$ and $\{\pi_2, \pi_3, \pi_4\}$ from Example 4.2.1. Both these sets are commutable and both define 6-dimensional distributions of variables $X_1, X_2, \ldots, X_6$. Whilst the distribution $\pi_2 \triangleleft \pi_3 \triangleleft \pi_4$

Table 4.16: 2-dimensional marginal distributions of π_1

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

is uniform, the distribution

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

is the one for which

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

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 reordered in the way that the new permutation 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.

Example 4.2.5 *All the variables in this example are binary with values* $\{0, 1\}$*. We shall consider three* 6*-dimensional distributions, each of which will be an independent product of three* 2*-dimensional distributions:*

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

where distribution μ is a uniform distribution and

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

It is not difficult to show that distributions π_1, π_2, π_3 are pairwise consistent. Distributions π_1 and π_3 share only variables x_{12} and x_{23} and both the respective 2-dimensional marginal distributions are for this pair of variables uniform. Similarly, distributions π_2 and π_3 have common arguments x_{13}, x_{32} and π_1, π_3 share x_{22}, x_{33}. Again, all the respective 2-dimensional marginal distributions are uniform.

Notice that in this case

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

because

$$\pi_2(x_{13}, x_{24}, x_{31}, x_{32} | x_{12}, x_{23}) = \kappa(x_{13}|x_{12})\mu(x_{24}|x_{23})\mu(x_{31}, x_{32}).$$

Analogously

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

Thus $\pi_1 \triangleleft \pi_2 \triangleleft \pi_3$ is a product of three terms, each of which is a composition of three 2-dimensional distributions. Moreover, each of these terms is a composition of distributions whose variables form a star-like system, and therefore, due to Theorem 4.2.1, one can see that $\{\pi_1, \pi_2, \pi_3\}$ is commutable.

Now, we shall 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), it is enough to show it for π_1, π_2, π_3.

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

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

From this we see that

$$(\pi_1 \triangleright \pi_2 \triangleright \pi_3)(x_{32}, x_{33})$$

is a uniform distribution, which is not true for π_3*. This means that* π_3 *cannot be a marginal distribution of* $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ *and therefore* π_1, π_2, π_3 *is not perfect.* $\diamondsuit$

Remark 4.2.4 In the following text we will also speak about *commutable sequences*. By this we will understand a generating sequence whose elements form a commutable set. ○

4.3 Flexible sequences

Definition 4.3.1 A generating sequence $\pi_1, \pi_2, \ldots, \pi_n$ is called *flexible* if for all $j \in K_1 \cup \ldots \cup K_n$ there exists a permutation $i_1, i_2, \ldots, i_n$ such that $j \in K_{i_1}$ and

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

In other words, flexible sequences are those, which can be reordered in many ways so that each variable can appear among arguments of the first distribution. It does not mean, however, that each distribution appears at the beginning of the generating sequence. If this would be the case, then flexible sequences would be just a subclass of perfect sequences (since each distribution would be a marginal of the composed distribution – see Lemma 4.3.2).

Example 4.3.1 *Obviously, any triplet of distributions* $\pi_1(x_1, x_2), \pi_2(x_1, x_3)$, $\pi_3(x_2, x_3)$*, 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 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 opposite case because of Theorem 4.1.1. 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 easily prove that perfectized the sequence* $\kappa_1, \kappa_2, \kappa_3$ *from Example 4.1.3 is not flexible).* $\diamondsuit$

Flexibility of generating sequences will be used in computational algorithms, especially when computing conditional distributions in Section 7.2. Let us mention for the reader familiar with Bayesian networks that flexible sequences play for compositional models the same role as decomposable models for Bayesian networks. Namely, local computations ([29]) are based on the fact that cliques of a decomposable graph can always be ordered in the way that the ordering meets RIP and the sequence starts with an arbitrarily pre-selected clique. Therefore, we will study properties of flexible sequences also in Section 5.3, which is devoted to investigation of relation of compositional and decomposable models. At this place we present only a couple of basic properties and examples illuminating relation between flexible and perfect sequences. The first one shows that if a generating sequence meets the condition of Lemma 4.1.3 then this sequence is not only perfect but also flexible.

Lemma 4.3.1 *If $\pi_1, \pi_2, \ldots, \pi_n$ is a sequence of pairwise consistent oligodimensional probability distributions such that $K_1, \ldots, K_n$ meets RIP then this sequence is flexible.*

Proof The assertion is a direct consequence of Lemma 4.1.2, according to which we can find a permutation meeting RIP and starting with an arbitrary K_ℓ. Then it is enough to realize that, due to Lemma 4.1.3, all the RIP permutations yield perfect sequences, which define according to Theorem 4.1.4 the same multidimensional distribution. □

Remark 4.3.1 Notice that there exist non-trivial flexible sequences $\pi_1(x_{K_1})$, $\pi_2(x_{K_2}), \ldots, \pi_n(x_{K_n})$, for which no permutation $K_{i_1}, K_{i_2}, \ldots, K_{i_n}$ meets RIP – see e.g. Example 4.3.2. (Trivially, any sequence of uniform distributions is flexible.) ○

Remark 4.3.2 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 from a system of distributions one can create different flexible sequences - see the following Example. ○

Example 4.3.2 *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:*

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

All the corresponding sequences, which are perfect due to Lemma 4.1.3, define the same distribution and therefore all of them are flexible. Nevertheless, the generating sequences corresponding to the remaining two permutations

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

are also flexible (though not perfect! – to verify it show that π_2 *is not a marginal of the resulting* 4*-dimensional distribution) because both of them define the same distribution*

$$\pi_1 \triangleright \pi_3 \triangleright \pi_2 = \pi_3 \triangleright \pi_1 \triangleright \pi_2 = \pi_1 \pi_3,$$

and each variable appears among the arguments of the first distribution in a sequence. Thus we have shown that from the considered three distributions one can set up two different 4*-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 side, flexible sequence* π_1, π_3, π_2 *has a special (for the first sight may be rather strange) property: Its distributions form a commutable set* $\{\pi_1, \pi_3, \pi_2\}$ *but this defines the distribution, which differs from the distribution defined by the flexible sequence* π_1, π_3, π_2*:*

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

$\diamondsuit$

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

Lemma 4.3.2 *If for all* $i = 1, \ldots, n$ *of a flexible sequence* $\pi_1, \ldots, \pi_n$ *there exists an index*

$$j \in K_i \setminus (K_1 \cup \ldots \cup K_{i-1} \cup K_{i+1} \cup \ldots \cup K_n)$$

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

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

and therefore all π_i are marginal distributions of $\pi_1 \triangleright \ldots \triangleright \pi_n$. From this, perfectness of $\pi_1, \ldots, \pi_n$ is guaranteed by Theorem 4.1.1. $\square$

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 can be simplified. Somehow these sequences resemble the *decomposable models* and therefore, as already said above, we will learn more about perfect flexible sequences in Section 5.3.

Example 4.3.3 *Consider a situation when* $K_1 = \{1,2\}$, $K_2 = \{2,3\}$, $K_3 = \{3,4\}$, $K_4 = \{1,4,5\}$ *(see Figure 4.5 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 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 is guaranteed by the fact that* $K_4 \cup K_2 = \{1,2,3,4,5\}$ *and* $\pi_4 \triangleright \pi_2 \triangleright \pi_{i_2} = \pi_2 \triangleright \pi_4 \triangleright \pi_{i_2}$ *because of consistency of* π_2 *and* π_4 *(Lemma 3.2.2). Moreover, in this situation* $K_{i_2} \subset K_4 \cup K_2$ *and therefore* $\pi_4 \triangleright \pi_2 = \pi_4 \triangleright \pi_2 \triangleright \pi_{i_2}$.

If $i_1 \neq 2$, *then* K_4, K_{i_1}, K_{i_2} *meets RIP and the flexibility follows from Lemma 4.1.2, which says that there are RIP orderings starting with an arbitrary set, and Theorem 4.1.4, according to which all these RIP orderings define the same distribution.* ◊

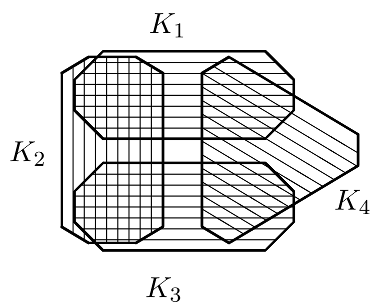


Figure 4.5: System of index sets from Example 4.3.3

Bibliography

[1] R.R. Boukaert, *Bayesian belief networks – from construction to inference*, PhD. thesis, University of Utrecht (Netherlands), 1995.

[2] P. Cheeseman, A method of computing generalized Bayesian probability values for expert systems. In: *Proc. 6th Int. Conf. on AI (IJCAI 83)*, Karlsruhe, Germany, 1983, pp. 198-202.

[3] G. de Cooman, Possibility theory I: The measure- and integral-theoretic groundwork. *Int. J. General Systems* **25** (1997), pp. 291–323.

[4] G. de Cooman, Possibility theory II: Conditional possibility. *Int. J. General Systems* **25** (1997), pp. 325–351.

[5] G. de Cooman, Possibility theory III: Possibilistic independence. *Int. J. General Systems* **25** (1997), pp. 353–371.

[6] I. Csiszár, I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.*, **3** (1975), pp.146-158.

[7] W.E. Deming and F.F. Stephan, On a least square adjustment of a sampled frequency table when the expected marginal totals are known, *Ann. Math. Stat.* **11** (1940), pp. 427-444.

[8] D. Dubois and H. Prade, *Possibility theory.* Plenum Press, New York, 1988.

[9] R.G. Gallager, *Information theory and reliable communication.* J. Wiley, New York, 1968.

[10] P. Hájek, T. Havránek, and R. Jiroušek, *Uncertain Information Processing in Expert Systems*. CRC Press, Inc., Boca Raton, 1992.

[11] F.V. Jensen, *Introduction to Bayesian Network*. UCL Press, London, 1996.

[12] F.V. Jensen, *Bayesian Networks and Decision Graphs*. Springer Verlag, Series: Information Science and Statistics, New York 2001.

[13] R. Jiroušek, A survey of methods used in probabilistic expert systems for knowledge integration, *Knowledge-Based Systems* **3** (1990), 7-12.

[14] R. Jiroušek, Composition of probability measures on finite spaces, in: *Proc. of the 13th Conf. Uncertainty in Artificial Intelligence UAI'97*, eds. D. Geiger and P. P. Shenoy, Morgan Kaufmann Publ., San Francisco, California, 1997, pp. 274-281.

[15] Jiroušek, R. (1998). Graph Modelling without Graphs. In *Proc. of the 17th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems IPMU'98*, Paris (B. Bouchon-Meunier, R.R. Yager, eds.). Editions E.D.K. Paris, pp. 809-816.

[16] R. Jiroušek, Marginalization in composed probabilistic models. In: *Proc. of the 16th Conf. Uncertainty in Artificial Intelligence UAI'00* (C. Boutilier and M. Goldszmidt eds.), Morgan Kaufmann Publ., San Francisco, California, 2000, pp. 301-308.

[17] R. Jiroušek, Decomposition of multidimensional distributions represented by perfect sequences. *Annals of Mathematics and Artificial Intelligence*, **35**(2002) pp. 215-226.

[18] R. Jiroušek, Detection of independence relations from persegrams. In *Proc. of the 19th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems IPMU'02*, Annecy (B. Bouchon-Meunier, L. Foulloy, R.R. Yager, eds.). ESIA, Framce, pp. 1261-1268.

[19] R. Jiroušek, On computational procedures for probabilistic compositional models. In: *Proceedings of the 5th Czech-Japan Seminar* (J. Ramík, ed.), 2002.

[20] R. Jiroušek, On Approximating Multidimensional Probability Distributions by Compositional Models. In: *Proceedings of the 3rd Internationa Symposium on Imprecise Probabilities and Their Application*, (J.M. Bernard, T. Seidenfeld, M. Zeffalon, eds.), Carleton Scientific, Otawa, Canada, 2003, pp. 305-320.

[21] R. Jiroušek and R. Scozzafava, *Basic probability*, Lecture Notes for PhD. studies 1/2003, Faculty of Management, Jindřichův Hradec, Oeconomica, Praha, 2003.

[22] R. Jiroušek and J. Vejnarová, Construction of Multidimesional Models by Operators of Composition: current state of art. em Soft Computing, 7(2003), pp.328-335.

[23] R. Jiroušek and J. Vejnarová, General framework for multidimensional models. *Int. J. of Intelligent Systems* 18(2003), pp. 107-127.

[24] R. Jiroušek, J. Vejnarová and J. Gemel'a, Possibilistic belief network constructed by operators of composition and its application to financial analysis. In: Srivastava R.P. and Mock T, eds. Belief Functions in Business and Finance. Physica-Verlag, Studies in Fuzzines and Soft Computing, 2002; pp. 252-280.

[25] H.G. Kellerer, Verteilungsfunktionen mit gegeben Marginalverteilungen. *Z. Warhsch. Verw. Gebiete*, 3(1964), pp. 247- 270.

[26] K, H.G. (1964). Telefoonverkeersrekening. *De Ingenieur*, 52(1937), E15 – E25.

[27] S. Kullback, An information-theoretic derivation of certain limit relations for a stationary Markov chain. *J. SIAM Control*, 4(1966), pp. 454-459.

[28] S.L. Lauritzen, *Graphical Models*. Clarendon Press, Oxford, 1996.

[29] S. L. Lauritzen and D. J. Spiegelhalter, Local computation with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society series B* **50** (1988), pp. 157–224.

[30] A. Perez, ε-admissible simplification of the dependence structure of a set of random variables, *Kybernetika* **13** (1977), pp. 439–450.

[31] R.D. Shachter, Evaluating Influence Diagrams, Operations Res. 34 (1986), 871-890.

[32] J. Vejnarová, Possibilistic independence and operators of composition of possibility measures. In: *Prague Stochastics'98* (M. Hušková, J. Á. Víšek, P. Lachout eds.) JČMF, 1998, pp. 575–580.

[33] J. Vomlel, Methods probabilistic knowledge of integration. PhD. thesis. FEL ČVUT, Praha, 2000.