Czech Technical University in Prague

Faculty of Nuclear Sciences and Physical Engineering

Department of Mathematics

Probabilistic Compositional Models: solution of an equivalence problem

DOCTORAL THESIS

September 2011

Ing. Václav Kratochvíl

Czech Technical University in Prague

Faculty of Nuclear Sciences and Physical Engineering

Department of Mathematics

Probabilistic Compositional Models: solution of an equivalence problem

by

Ing. Václav Kratochvíl

Supervisor: Prof. Radim Jiroušek DrSc.

Prague 2011

Acknowledgements

First of all, I would like to thank my supervisor Radim Jiroušek for his support and encouragement during my studies; his reading on an early draft of this text and commentary substantially improved the exposition. Many thanks belongs also to my colleagues from the Institute of Information Theory and Automation of the Academy of Sciences who provided me the background for my work and gave me many advices and criticisms.

Last but not least, I am grateful to my family and wife Pavla for their neverending patience, help, and love.

Concerning the financial arrangements, this work was supported by National Science Foundation of the Czech Republic under Grants No. ICC/08/E010, and 201/09/1891, and by Ministry of Education under Grants No. 1M0572 and 2C06019.

Václav Kratochvíl

Czech Technical University in Prague
Prague, September 2011

Probabilistic Compositional Models: solution of an equivalence problem

Ing. Václav Kratochvíl

Czech Technical University in Prague, Prague, September 2011

Supervisor: Prof. Radim Jiroušek DrSc.

Compositional model theory (originally developed by Radim Jiroušek) represents an alternative approach to probabilistic models, mainly graphical ones. A compositional model may be defined as a multidimensional distribution assembled from a sequence of low-dimensional unconditional distributions (the so-called generating sequence), with help from the operator of composition. The main advantage lies in the fact that those low-dimensional distributions may be easily stored in computer memory – the size of a joint probability distribution grows exponentially with the number of variables of interest. Fragmenting the multidimensional distribution into a generating sequence brings forth several complications. While a model is put together, a system of (un)conditional independencies is simultaneously introduced by the structure of the generating sequence. This system of independencies - the so-called induced independence model – is valid for any compositional model defined by a generating sequence with this structure.

This text should familiarize the reader with new results in this theory, namely with complete solution of the equivalence problem. The equivalence problem is the problem of recognizing whether two given structures over the same set of variables induce the same independence model. In this case, we present three different simple rules to recognize that two structures are equivalent. We also present three elementary operations – IE operations – on structures such that we can easily convert one structure into an equivalent one in terms of these operations. Moreover, we show that one can generate all structures equivalent to a given one using these operations, and the impact of IE operations on generating sequences is explored.

Using IE operations and our knowledge of equivalence problem solution, we were able to examine the problem of conditioning a probability distribution represented by a compositional model, as well as a connected problem of generating sequence flexibility, from a different perspective. A partial solution of this problem is published.

Contents

List of Figures

Notation

Symbol	**Meaning**
$\mathbb{N}$	set of natural numbers
N	non-empty set of finite-valued random variables
U, V, W, Z	subsets of N; sets of variables
2^U	powerset of U – set of all subsets of U
u, v, w, x, y, z	singletons; variables from N
$\mathbf{X}_u$	finite set of values of variable u
$\mathbf{X}_U$	$\times_{u \in U} \mathbf{X}_u$ – all combination of considered values
$=$	equality
$\equiv$	identity
$\subseteq$	the usual (non-strict) case of inclusion
$\subset$	strict inclusion only

Logical connectives (symbol)

$\wedge$	logical and
$\vee$	logical or
$\Rightarrow$	implication (only if)
$\Leftrightarrow$	shorthand for $\Rightarrow \wedge \Leftarrow$ (if and only if)

Permutations

σ	permutation
T_n	set of all permutations $1..n$
$(i_1\ i_2\ \ldots\ i_r)$	r-cycle permutation
$(i\ j)$	transposition, 2-cycle permutation

Compositional model description

$\ll$	dominance
π, κ	probability distributions
$\pi(U)$	distribution defined over set of variables U
$\Pi^{(U)}$	set of all probability distributions defined for variables U
$\Pi^{(U)}(\pi(U))$	set of all extensions of the distribution π to $\lvert V\rvert$-dimensional distribution
$\pi^{\downarrow U}$	marginal distribution over set of variables U
$\triangleright$	operator of composition
$\pi_1(U_1), \ldots, \pi_n(U_n)$	generating sequence of a compositional model $\pi_1(U_1) \triangleright \ldots \triangleright \pi_n(U_n)$

Structure description

$\mathcal{P} = U_1, \ldots, U_n$	structure of sequence $\pi_1(U_1), \ldots, \pi_n(U_n)$
$\lvert\mathcal{P}\rvert$	number of sets in the structure, its cardinality
$K_i^{\mathcal{P}}$	i-th set (column) of structure $\mathcal{P}$
$R_i^{\mathcal{P}}, S_i^{\mathcal{P}}$	partition of $K_i^{\mathcal{P}}$ with respect to the structure $\mathcal{P}$
$\mathcal{P}[U]$	substructure of $\mathcal{P}$ induced by set U
$]u[_{\mathcal{P}}$	i such that $u \in R_i^{\mathcal{P}}$
$u \prec_{\mathcal{P}} v$	$]u[_{\mathcal{P}} <]v[_{\mathcal{P}}$
$u \preceq_{\mathcal{P}} v$	$]u[_{\mathcal{P}} \leq]v[_{\mathcal{P}}$

Structure properties (invariants)

$u \leftrightarrow_{\mathcal{P}} v$	u and v are connected in $\mathcal{P}$
$\mathcal{E}(\mathcal{P})$	set of all connections in $\mathcal{P}$
$\langle u, v \rvert w\rangle \in \mathcal{F}(\mathcal{P})$	u, v, w form an F-condition in $\mathcal{P}$
$\mathcal{F}(\mathcal{P})$	set of all F-conditions in $\mathcal{P}$
$\mathcal{C}^*(\mathcal{P})$	strong structure core of $\mathcal{P}$
$\mathcal{C}(\mathcal{P})$	weak structure core of $\mathcal{P}$
$ntriv(\mathcal{P})$	set of non-trivial columns of $\mathcal{P}$
$red(\mathcal{P})$	structure obtained from $\mathcal{P}$ where all trivial columns were removed
$\mathcal{N}(\mathcal{P})$	set of all non-trivial sets induced by $\mathcal{P}$

Induced independence relations

$U \perp\!\!\!\perp V \rvert Z[\mathcal{P}]$	conditional independence induced by structure $\mathcal{P}$
$U \perp\!\!\!\perp V \rvert Z[\pi]$	conditional independence induced by distribution π
$\mathcal{I}(\mathcal{P})$	independence model induced by structure $\mathcal{P}$
$\mathcal{D}(\mathcal{P})$	dependence model induced by structure $\mathcal{P}$

Chapter 1

Introduction

The ability to represent and process multidimensional probability distributions is necessary for application of probabilistic methods in Artificial Intelligence and computer-aided reasoning. Among the most popular approaches nowadays are the methods based on Graphical Markov Models, e.g., Bayesian Networks. An alternative approach to Graphical Markov Models is represented by the so-called compositional models, which seem to be more efficient than Bayesian networks (more efficient in computations, etc.). Nevertheless, many substantial problems remain to be solved. One of them, the so-called equivalence problem, is a primary focus of this text.

Artificial Inteligence

The interest in computer-aided reasoning within computer science dates back to the very early days of Artificial Inteligence (AI), when much work had been initiated for developing computer programs to solve problems that require a high degree of intelligence. However, aside from computer science, the intellectual roots of AI and the concept of intelligent machines may be found even in Greek mythology. Intelligent artifacts appear in literature published since that time, along with real mechanical devices actually demonstrating some degree of intelligence. Some authors of texts about AI history like [6],[7], or [58] mention clocks, the first modern measuring machines, as well as mechanical animals and other toys created by clockmakers. For example, see *DaVinchi's walking lion* (1515) on youtube.com – [61]. Another, much older, example of AI can be found in *Aristotle's syllogistic logic* – the first formal deductive reasoning system. In the Middle Ages, there were rumors of secret mystical or alchemical means of placing mind into matter, such as Prague's famous *Rabbi Judah Loew's Golem.* By the 19th century, ideas about artificial men and thinking machines were developed in fiction, as in *Mary Shelley's Frankenstein* or *Karel Čapek's R.U.R.* (*Rossum's*

Universal Robots). AI has continued to be an important element of science fiction into the present. Among other things, let us also mention the legendary *three laws of robotics* published by *Isaac Asimov* in 1950.

The seeds of modern AI were planted by classical philosophers who attempted to subscribe the process of human thinking as the mechanical manipulation of symbols. This work culminated in the invention of the programmable digital computer in the 1940s, a machine based on the abstract essence of mathematical reasoning. Since then, AI has been inherently tied to computer science.

In 1956, at *Dartmouth Conference*, the term *"Artificial Intelligence"* was coined by *John McCarthy*, following by an influential proposal for building automated reasoning systems. This proposal, depicted in Figure 1.1, calls for a system with two components: a knowledge base, which encodes what we know about the world, and a reasoner (inference engine), which acts on the knowledge base to answer queries of interest. For example, the knowledge base may encode what we know about the theory of sets in mathematics, and the reasoner may be used to prove various theorems in this domain.

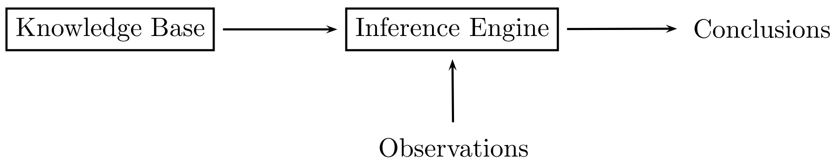


Figure 1.1: A reasoning system in which the knowledge base is separated from the reasoning process. The knowledge base is often called a *model*, giving rise to the term *model-based reasoning*

McCarthy's proposal was actually more specific than what was suggested by Figure 1.1, as he called for expressing the knowledge base using statements in suitable logic, and for using logical deduction in realizing the reasoning engine. This approach proved to be ineffective later, and that is why we state it here in its general form. While the knowledge base can be domain-specific, changing from one application to another, the reasoner is quite general and fixed. This aspect became the basis for a class of reasoning systems known as *knowledge-based* or *model-based systems*. We will also subscribe to this knowledge-based approach for reasoning, except that our knowledge bases will be *compositional models* and our reasoning engine will be based on *probability theory*.

Expert systems

When computers with large memories became available around 1970, AI researchers started to build applications. This led to the development and deployment of *expert systems* (introduced by *Edward Feigenbaum*), the first truly commercial, successful form of AI software. Expert systems are sometimes labeled as *knowledge-based systems* and they strictly follow the structure designed by McCurthy's proposal, shown in Figure 1.1. The expert knowledge is stored in the form of *if/then type statements* (rules) – in a *rulebase*. The principal distinction between expert systems and traditional problem-solving programs is the way in which the problem-related expertise is coded. In traditional applications, problem-related expertise is encoded in both program and data structures. In the expert system approach, the problem expertise is encoded mostly in data structures.

In an example related to tax advice, the traditional approach has data structures that describe the taxpayer and tax tables, and a program that contains rules (encoding expert knowledge) that relate information about the taxpayer to tax table choices. In contrast, using the expert system approach, the latter information is also encoded in data structures. The collective data structures are called the *knowledge base*. This approach has a benefit in the form of simple extending/altering rules. More about expert systems can be found in previously published literature, such as in [20]. The typical applications are clinical decision support systems, troubleshooting, or computer games. An example and good demonstration of limitations of an expert system is the *Windows operating system troubleshooting software* embedded in the Windows XP operating system. It is designed to provide solutions, advice, and suggestions to common errors encountered while using this operating system.

Qualification problem

Yet McCarthy's proposal is very elegant, and as the approach was being applied to more application areas, a key difficulty called for an alternative proposal. In particular, it was observed that although deductive logic is a natural framework for representing and reasoning about facts, it was not capable of dealing with assumptions that tend to be prevalent. This problem is known as the *qualification problem* and it was stated formally by McCarthy in the late 1970s. Let us illustrate this difficulty on the example originally published in [11]:

Consider the following statement:

Birds fly.

In this case, the deductive logic would be able to obtain the expected conclusion when it sees a bird. However, it would meet with an inconsistency if it encounters a bird that cannot fly. On the other hand, if we write

If a bird is normal, it flies,

deductive logic will not be able to reach the expected conclusion upon seeing a bird, as it would not know whether the bird is normal or not – contrary to what humans will do. Note that this is the consequence of *monotonicity* of deductive logic. In fact, deductive logic is monotonic in the sense that once we deduce something from a knowledge base (the bird flies), we can never invalidate the deduction by acquiring more knowledge (the bird has a broken wing). Intuitively, monotonicity indicates that learning a new piece of knowledge cannot reduce the set of what is known. This is however in conflict with how the human mind works.

Non-monotonic logics

This, together with another of McCarthy's influential proposals [38], which called for equipping logic with an ability to jump to certain conclusions, led to the rise of a new generation of logics, the so called *non-monotonic logics*, in the 1980s. These logics are equipped with mechanisms for managing assumptions, criterion for deciding which assumptions to assert and retract, and when to do so.

During studies of non-monotonic logic, several different problems have been addressed: *reasoning by default* (consequences may be derived only because of lack of evidence to the contrary), *abductive reasoning* (consequences are only deduced as most likely explanations) and some important approaches to *reasoning about knowledge* (the ignorance of a consequence must be retracted when the consequence becomes known), and similarly, *belief revision*. Regarding the scope of this text, note that belief revision [16] is the process of changing beliefs to accommodate a new belief that might be inconsistent with the old ones. On the assumption that the new belief is correct, some of the old ones have to be retracted in order to maintain consistency. The belief revision approach is an alternative to the so-called *paraconsistent logics*, which tolerate inconsistency rather than attempting to remove it.

Degree of belief

The discovery of the qualification problem, and the associated monotonicity problem of deductive logic, brought new evidence for alternative probabilistic approach to AI and gave numerical methods a second chance in AI. Among the most influential names in this trend are *Lauritzen, Pearl*, or *Shachter*. This alternative direction can be viewed as postulating the existence of a more fundamental notion, sometimes called a *degree of belief*, which, according to some treatments, can alleviate the need for assumptions altogether and, according to others, can be used as a basis for deciding which assumptions to make in the first place.

A degree of belief is a number that one assigns to a proposition in lieu of having to declare it as a fact (as in deductive logic) or an assumption (as in non-monotonic logic). For example, instead of assuming that a bird is normal unless observed otherwise – which leads us to tenuously believe that it also flies – we assign a degree of belief to the bird's normality, say, 99%, and then use this to derive a corresponding degree of belief in the bird's ability to fly.

Degrees of belief address the monotonicity problem by being revisable upward or downward, depending on what else is known. For example we may initially believe that a bird is normal with a belief of 99%, only to revise this to, say, 20% after learning that its wing is broken. One can argue that assigning a degree of belief is more committing than making an assumption. On the other hand, one can also argue to the contrary that working with degrees of belief is far less committing as they do not imply any particular truth of the underlying propositions [42].

The usage of numerical degrees of belief was proposed well before the qualification problem appeared. However, this approach was not warmly welcomed at first. AI hasits roots in symbolic logic, and for many years AI experts showed little interest in probability. There were several basic objections:

- Do *humans* use such degrees of belief in *their own reasoning*? (In those times, the correspondence with human cognition was highly valued)
- The *availability* of degrees of belief – where do the numbers come from?
- *Robustness* – what happens if I change this .90 to .95?
- *Application scale* – the size of probability distributions grows exponentially with the number of involved variables.

A number of different proposals have been published for interpreting degrees of belief including, for example, the notion of possibility on which fuzzy logic is based. Another very natural way (at least nowadays) is to interpret them as probabilities and manipulate them according to the laws of probability.

History of probability

The word *probable* comes to modern languages from Latin *probabilis*, and is generally applied to an opinion to mean *plausible* or *generally approved*. Two different aspects of probability can be considered:

- *Likelihood of hypotheses given the evidence for them.*
- *Behavior of stochastic processes such as throwing coins.*

The study of the former is historically older, and one can find it even in ancient *law of evidence* – developed for grading *degrees of proof, presumptions* and *half-proof* to deal with the uncertainties of evidence in court.

On the other hand, the mathematical methods of probability arose later – in the 1650s. In those days in France, gambling was popular and fashionable, not even restricted by law. As games became more complicated, and the stakes simultaneously became larger, a need arose for mathematical methods to compute chances. According to [44], a well-known gambler, *Chevalier De Mere* [15], consulted *Blaise Pascal* in Paris with questions about some games of chance. (He suffered severe financial losses for assessing incorrectly his chances of winning in certain games of dice.) This forced Pascal to begin correspondence with his friend *Pierre Fermat* about such problems. This correspondence between Pascal and Fermat represents the origin of the mathematical study of probability. The method they developed is now called the *classical approach* to computing probabilities. The method is as follows:

> Suppose a game has n equally likely outcomes, of which m outcomes correspond to winning. Then the probability of winning is m/n.

Note that this classical approach requires a game to be split into equally likely outcomes, which is not always possible to guarantee. It may even be unclear whether all possibilities are equally likely. Later, throughout the 18th century, the application of probability moved from games of chance to scientific problems like *theory of insurance* – the so-called *life tables*. Recall the book by *Pierre-Simon Laplace* from 1812 – *Theorie Analytique des Probabilities* – in which a mathematical theory of probability was presented with an emphasis on its scientific application. By 1850, many mathematicians found the classical method to be unrealistic for general use and were attempting to redefine probability in terms of the *frequency method*. This method consists of repeating a game a large number of times under the same conditions. The probability of winning is then approximately equal to the proportion of wins in the repeats. Note that this frequency method was used even by Pascal and Fermat to verify results obtained by the classical method. Moreover, *James Bernoulli* proved that the frequency method and the classical method are mutually consistent in his book *Ars Conjectandi* in 1713. In 1763, an influential theorem of *Thomas Bayes* for calculating *inverse probabilities* was posthumously published. It was later generalized by *Pierre-Simon Laplace* (1749-1827) to approach problems in medical statistics, reliability, and jurisprudence.

The power of probabilistic methods in dealing with uncertainty was shown by *Gauss's* determination of the orbit of *Ceres* (1801) [55]. However, the first rigorous approach to probability was developed by *Andrey Kolmogorov* in his 1933 monograph *Grundbegriffe der Wahrscheinlichkeitsrechnun*. He stated three fundamental axioms and built up probability from them in a way comparable with *Euclid's* treatment of geometry. Let us also mention Kolmogorov's contemporary, *B. de Finetti* [9], who rejected Kolmogorov's axiomatic approach and introduced a *subjectivist interpretation* to give particular insights as to the *meaning* of probability.

Probability reasoning

Considering the probability approach to AI, a group of researchers formed around the *UAI* conference (organized regularly every year), for the first time in Los Angeles, California in 1985 . Some of the most influential work from this group came from *Judea Pearl* – a key proponent of probabilistic reasoning. He was able not only to show that many problems and paradoxes in symbolic formalisms (such as monotonicity) simply do not surface in the probabilistic approach (see a summary in [42], Chapter 10), but he also created a representational and computational system that could compete with symbolic systems that were in commercial use at the time. This system is known as a *Bayesian network.* A very good summary of Pearl's work may be found in his influential book, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* published in 1988 [42]. On the computational side of Bayesian networks, we should mention *jointree algorithm* by Lauritzen and *Spiegelhalter* from 1988 [36]. Other notable works have been contributed by Czech researchers *Albert Perez, Petr Hájek* – [17], *Tomáš Havránek* - [17],[18], and *Radim Jiroušek* – [17, 21, 22, 24, 25, 27, 28]. One may say that the latter group of researchers did the parallel but independent research of probability reasoning using both probabilistic and logical/algebraic approaches.

Graphical models

In the 1970s and early 1980s, the probabilistic and especially decision-theoretic models were largely out of favor within AI (certainly within the dominant expert-systems paradigms of the day), due in part to

> "a perception that decision-theoretic approaches were hopelessly intractable and were inadequate for expressing the rich structure of human knowledge." (*Horovitz et al.* 1988)

Even as (rule based) expert systems fell out of favor, the logicist paradigm continued to dominate AI through the early 1990s. However, the introduction of graphical models to AI in the early 1980s, especially Bayesian networks [42] and *influence diagrams* [5] by *Howard* and *Matheson* [19], precipitated an enormous change. Influence diagrams demonstrated that the language of decision theory was not only rich enough to capture the intricacies of complex decision problems faced throughout AI but also made possible a suitable decomposition of a problem into the representational levels. Developments in sophisticated inference methods for Bayesian networks and solution techniques for influence diagrams also demonstrated the computational power of these approaches.

Many reasoning tasks arising in uncertainty reasoning in AI can be considerably simplified if a suitable concept of relevance or irrelevance of symptoms or variables is taken into consideration. The conditional irrelevance in probabilistic reasoning is modeled by means of the concept of probabilistic *conditional independence* (CI) [42]. Since every CI-statement can be interpreted as a certain qualitative relationship among involved variables, the dimensionality of the problem can be reduced and a more effective way of storing the knowledge base may be found. Note that the concept of CI has also been introduced and studied in non-probabilistic calculi for dealing with uncertainty in AI. We can also mention much earlier work by *A.A. Markov*, who used the notion of conditional independence around 1900 to formulate simple multivariate models, now called *Markov chains*. A Markov chain for n categorical variables can be represented by undirected graph (UG) – a sequence of nodes $i_1, i_2, \ldots, i_n$ just where each consecutive pair is connected by an edge.

Conditional Independence structures were at first described by means of graphs in literature. The central idea is that each variable is represented by a node in a graph. Any pair of nodes may be joined by an edge. For most types of graphs, a missing edge represents some form of independency between the pair of variables. Because the independency may be either marginal or conditional on some or all of the other variables, a variety of types of graphs are needed. Graphs whose nodes correspond to random variables have been a traditional tool for describing structures of multidimensional probability distributions.

Graphical models were first developed by *Darroch* [10] and *Wermuth* [37] as special subclasses of *log-linear models* [1] for contingency tables, and of multivariate Gaussian distributions which can be interpreted in terms of conditional independencies and represented by UGs. Let us mention *Tomáš Havránek* [18], a Czech scientist who studied ways to determine which log-linear models are best supported by the data. We must also mention his influential collaboration with *David Edwards* [13] and [14] where they present efficient model-search procedures for some classes of hierarchical log-linear models. Their approach was recently reused in the work of *Vladislav Bína* [2] in the area of compositional models.

Two types stand out: UGs and directed graphs without directed cycles that are usually called *acyclic directed graphs* (DAG) (probabilistic influence diagrams [45] [47] or Bayesian network [42]). The early usage of DAGs dates back to the 1930s, when geneticist *Sewall Wright* [60] introduced and used path diagrams for recursive linear relations. Nevertheless, Wright and his successors did not thoroughly analyze their directed graphs in terms of conditional independence.

In addition to DAGs and UGs, various types of hybrid graphs were used. For example, chain graphs were introduced by *Lauritzen* and *Wermuth* in the mid-1980s [37]. Comparison of all the above-mentioned approaches as well as several others can be found in [50]. Nevertheless, the graphical approach cannot describe all possible probabilistic CI-structures. A natural way to remove this disadvantage is to describe this structure by means of so-called *independency models*, that is, lists of CI-statements. But such an approach would be unnecessarily wide: owing to the above- mentioned formal properties of CI [12], many independency models cannot be models of probabilistic CI-structures. Therefore, *Pearl* and *Paz* [41] introduced the concept of *semigraphoid* (resp. graphoid) as an independency model closed under four (or five) concrete inference rules expressing the above-mentioned properties of CI. Note that the graphoid axioms were initially identified in [12] and [48] and then rediscovered and named by Pearl and Paz. The graphoid axioms are designed in a way that every model of a probabilistic CI-structure is a semigraphoid as well. Pearl [42] conjectured the converse statement:

> "Every semigraphoid is a CI-model" (or "every graphoid is a CI model induced by a strictly positive probability distribution" [41]).

This conjecture was later refuted by *Milan Studený* [49].

Separation criteria

We interpret a UG by saying that two families of variables U and V are independent given a third family Z whenever all paths between U and V go through Z. We interpret a DAG by saying that each variable u is independent, given its parents (variables with arrows to u) of its nondescendants (variables to which there is no directed path from u).

Graphical models are attractive for several reasons. Most importantly, perhaps, they facilitate the construction of probability models. The conditional independence assumptions represented by a graph are equivalent to the assumption that the joint probability density for the variables in the graph can be decomposed into factors involving only neighboring variables (cliques in UGs, each variable and its parents in DAGs), and thus the graph represents the first step in model construction. Graphical representation also serves to facilitate computations, especially updating after observation of some of the variables. In addition, powerful *separation criterion* permits one to read all CI statements induced by the model directly from the graph. Logically, we use the notion of graph separation to derive a graphical rule (separation criterion) for inferring conditional independence, especially in the case of undirected graphs.

A first path criterion for directed graphs was given by J. Pearl, who called it *d-separation* (for separation in directed graphs). Another equivalent criterion (the so-called *moralization criterion*) is due to S.L. Lauritzen.

D-separation is direct in the sense that it defines when a trail connecting two nodes is d-separated by a set of nodes Z and corresponding variables U are conditionally independent of V given Z if every trail from U to V is d-separated by Z. On the contrary, Lauritzen's moralization criterion is indirect in the sense that for every disjoint triple U, V, Z one takes a special induced subgraph that is converted into its *moral graph* H. Since the moral graph is undirected, one uses the classic separation criterion for UGs.

Equivalence problem

While noting the advantages of the graphical representation of CI, we should also note that the representation is imperfect. On the one hand, not all sets of CI relations that might be satisfied by a probability distribution can be represented by a graphical model. On the other hand, two or more graphical models often represent the same CI relations. When graphical models do represent the same independence relations, we say they are *independence equivalent* (sometimes it is called *Markov equivalence*.) Figure 1.2 gives an example.

$$\mathrm{U} \longrightarrow \mathrm{Z} \longrightarrow \mathrm{V} \qquad \mathrm{U} \longleftarrow \mathrm{Z} \longleftarrow \mathrm{V}$$

$$\mathrm{U} \longleftarrow \mathrm{Z} \longrightarrow \mathrm{V} \qquad \mathrm{U} \text{ --- } \mathrm{Z} \text{ --- } \mathrm{V}$$

Figure 1.2: These four graphs encode the same CI relation: U and V are independent given Z.

The so-called *equivalence problem* is how to recognize whether two given graphs are independence equivalent. It is also of special importance to have a simple rule that allows us to recognize this equivalence (solved in [56]), and an easy way to get from one graph to another in terms of some elementary operations on graphs (solved in [8]). Another very important aspect is the ability to generate all graphs equivalent to a given one.

A classic graphical characterization of equivalent graphs [56] by *Verma* and Pearl states that they are equivalent iff they have the same adjacencies and immoralities, which are special induced subgraphs. Representing a CI structure by any of the DAGs defining it leads to a non-unique description, causing later identification problems. It became apparent that each independence equivalence class can be represented by a unique chain graph, called an *essential graph.* This representation might be used to facilitate selection among models, where enumeration of essential graphs [29] and hierarchical clustering might also be useful.

Non-graphical/algebraic approach

The idea of an algebraic approach, introduced in [52], is to use an algebraic representative called a *standard imset*, which is, at the same time, a unique CI structure representative. It is a vector whose components are integers indexed by subsets of the set of variables (= nodes). The most important point is that from a geometric point of view, the set of standard imsets is the set of vertices (i.e., extremal points) of a certain polytope. Then the score-and-search method for learning a CI structure from data can be re-formulated as a classic *linear programming problem.*

Recently, in [54], they propose an even simpler algebraic representative called the *characteristic imset* It is a 0-1 vector obtained from the standard imset by an affine transformation.

State of the Art

The interest in computer-aided reasoning within computer science dates back to the very early days of Artificial Inteligence (AI), when much work had been initiated for developing computer programs to solve problems that require a high degree of intelligence.

The seeds of modern AI were planted by classical philosophers who attempted to subscribe the process of human thinking as the mechanical manipulation of symbols. This work culminated in the invention of the programmable digital computer in the 1940s, a machine based on the abstract essence of mathematical reasoning. Since then, AI has been inherently tied to computer science.

The ability to represent and process multidimensional probability distributions is necessary for application of probabilistic methods in Artificial Intelligence and computer-aided reasoning. Among the most popular approaches nowadays are the methods based on Graphical Markov Models, e.g., Bayesian Networks. An alternative approach to Graphical Markov Models is represented by the so-called compositional models, which seem to be more efficient than Bayesian networks (more efficient in computations, etc.). Nevertheless, many substantial problems remain to be solved. One of them, the so-called equivalence problem, is a primary focus of this text.

In 1956, at Dartmouth Conference, the term "Artificial Intelligence" was coined by John McCarthy, following by an influential proposal for building automated reasoning systems. This proposal, depicted in Figure 1.1, calls for a system with two components: a knowledge base, which encodes what we know about the world, and a reasoner (inference engine), which acts on the knowledge base to answer queries of interest.

While the knowledge base can be domain-specific, changing from one application to another, the reasoner is quite general and fixed. This aspect became the basis for a class of reasoning systems known as knowledge-based or model-based systems. We will also subscribe to this knowledge-based approach for reasoning, except that our knowledge bases will be compositional models and our reasoning engine will be based on probability theory.

Many reasoning tasks arising in uncertainty reasoning in AI can be considerably simplified if a suitable concept of relevance or irrelevance of symptoms or variables is taken into consideration. The conditional irrelevance in probabilistic reasoning is modeled by means of the concept of probabilistic conditional independence (CI) [42]. Since every CI-statement can be interpreted as a certain qualitative relationship among involved variables, the dimensionality of the problem can be reduced and a more effective way of storing the knowledge base may be found.

While noting the advantages of the graphical representation of CI, we should also note that the representation is imperfect. On the one hand, not all sets of CI relations that might be satisfied by a probability distribution can be represented by a graphical model. On the other hand, two or more graphical models often represent the same CI relations. When graphical models do represent the same independence relations, we say they are independence equivalent (sometimes it is called Markov equivalence.) Figure 1.2 gives an example. The so-called equivalence problem is how to recognize whether two given graphs are independence equivalent. It is also of special importance to have CHAPTER 1. INTRODUCTION 16 U Z V U Z V U Z V U Z V Figure 1.2: These four graphs encode the same CI relation: U and V are independent given Z. a simple rule that allows us to recognize this equivalence (solved in [56]), and an easy way to get from one graph to another in terms of some elementary operations on graphs (solved in [8]). Another very important aspect is the ability to generate all graphs equivalent to a given one. A classic graphical characterization of equivalent graphs [56] by Verma and Pearl states that they are equivalent iff they have the same adjacencies and immoralities, which are special induced subgraphs. Representing a CI structure by any of the DAGs defining it leads to a non-unique description, causing later identification problems. It became apparent that each independence equivalence class can be represented by a unique chain graph, called an essential graph. This representation might be used to facilitate selection among models, where enumeration of essential graphs [29] and hierarchical clustering might also be useful.

Compositional models are probabilistic models presenting an alternative to well-known Bayesian networks. But unlike the graphical models, the compositional models represent a purely algebraic approach based on directly assembling low-dimensional probability distributions with the aid of the *operator of composition*, without the necessity to employ graphs. Yet graphs (hypergraphs) can be used for the sake of visualization.

It can be shown that both approaches – Bayesian networks and compositional models – are equivalent in the sense that they can both represent the same class of probability distributions [23], but the compositional models appear to be less computationally demanding for the frequent task of computing marginal probability distributions [3]. Another advantage may be that, by redefining the operator of composition, three different frameworks for uncertainty description may be considered: probability and possibility theories, and Dempster-Shafer theory of belief functions. Special operators of composition are introduced within all three frameworks in [26]. This fact enables us, among other things, to define the concept of conditional independence meeting all the semigraphoid axioms. (It became apparent that while factorization and conditional independence coincide for probability and possibility theories, they differ from each other for belief functions.)

The basic properties of compositional models are described well in [24] and in the recent review paper [27]. But the approaches for understanding how CI is coded in a model structure are rather undeveloped. Similar to other probabilistic models, the structure of the model induces a system of CI relations. Moreover, two different structures may induce the same system. The complete solution of equivalence problems in the case of compositional model structures has not yet been published.

Generally, the solution can be inspired by the approaches used in other probability models, especially in DAGs mentioned above. Here we employ the fact that compositional models are equivalent with Bayesian networks in such a way that one can convert a compositional model into equivalent Bayesian network and vice versa [24].

In case of Bayesian networks and DAGs, the solution of an equivalence problem is usually split up into two main branches: direct and indirect characterization of independence equivalence. While the direct characterization contains invariable properties of independence equivalent structures, the indirect one covers elementary operations preserving the induced independence model. We use a similar partition in this text as well.

Goals of the Thesis

This thesis elaborates on two major topics in detail, namely, solution of the equivalence problem in theory of compositional models, and its usage in other areas like conditioning probability distributions represented by a compositional model. Concerning the former, we will focus on properties that are invariable for the class of independence equivalent structures as well as on elementary operations preserving the induced independence model of those structures. We will demonstrate that both of these approaches – partial solutions of an equivalence problem – may be connected by stating them as necessary and sufficient conditions for independence equivalence of two arbitrary structures. Next, we will derive additional conditions guaranteeing that a probability distribution represented by a compositional model is invariant with respect to the operations mentioned above.

The latter topic was motivated by an open problem posed by Radim Jiroušek in [24], i.e., generating sequence flexibility. Note that in the case of conditioning of a probability distribution represented by a compositional model, the problem is converted to that of respective generating sequence flexibility. The flexibility problem will be divided into two sub-problems, where the first one will be solved using the solution knowledge derived in the first part of this text – the equivalence problem solution. For the second sub-problem we will illustrate its complexity, using several examples.

Chapter 2

Notation

We introduce basic probability theory notions in this chapter as a tool for representing knowledge. Moreover, we present notation and conventions used thoroughout the text. We use the following classification of various important statements. Statements taken from literature are labeled as *Assertions* and they are stated without proofs. The original statements implied in this work are labeled as *Lemmata, Theorems* and *Corollaries* with the following difference: While a lemma comprehends an interesting and nontrivial statement, a theorem indicates a significant/essential statement. Both of them are always proven. In contrast, a corollary contains some trivial but important consequence of a lemma (or a combination of more than one) that is used later.

In addition, further explanatory text is provided within *examples* and *remarks.*

2.1 Basic notions of probability theory

In this text, we will deal with a non-empty finite set of finite-valued variables and the symbol N will denote such a set. The symbols U, V, W, Z will be used for subsets of N. $|U|$ will denote the number of elements in U, that is, its *cardinality.* Symbols u, v, w, x, y, z denote variables as well as singletons $\{u\}, \{v\}, \{w\}, \{x\}$, $\{y\}$, $\{z\}$. Two set inclusion symbols are used thoroughout the paper, namely $\subset$ and $\subseteq$. The symbol $U \subseteq V$ (also $V \supseteq U$) denotes that U is a *subset* of V (alternatively V is a *superset* of U) which involves the situation $U = V$. However, *strict inclusion* is denoted as follows: $U \subset V$ or $V \supset U$ means that $U \subseteq V$ but $U \neq V$. The *power set* of a non-empty set N is a class of all of its subsets $\{U; U \subseteq N\}$, denoted by 2^N. Each variable u from N is assumed to have a finite (non-empty) set of values $\mathbf{X}_u$. The set of all combinations of the considered values will be denoted $\mathbf{X}_N = \times_{u \in N} \mathbf{X}_u$. Analogously for $U \subset N$, $\mathbf{X}_U = \times_{u \in U} \mathbf{X}_u$.

All probability distributions of the considered variables will be denoted by Greek letters (π, κ, etc. with possible indices); thus for $U \subseteq N$, we consider a distribution $\pi(U)$ which is defined on variables U. Symbol $\pi(U)$ represents a $|U|$-dimensional probability distribution and $\pi(x)$ a value of probability distribution π for point $x \in \mathbf{X}_U$.

Having two distributions $\pi(U)$ and $\kappa(V)$, we say that κ dominates π (in symbols $\pi \ll \kappa$) if, for all $x \in \mathbf{X}_{U\cap V}$,

$$\kappa(x) = 0 \Rightarrow \pi(x) = 0.$$

2.2 Marginal distribution

For a probability distribution $\pi(U)$ and $V \subseteq U$ we denote by $\pi(V)$ or $\pi^{\downarrow V}$ the respective *marginal distribution*, which can, for all $x \in \mathbf{X}_V$, be computed as

$$\pi^{\downarrow V}(x) = \sum_{y \in \mathbf{X}_U : y_V = x} \pi(y)$$

where y_V denotes the *projection* of $y \in \mathbf{X}_U$ into $\mathbf{X}_V$. For computation of marginal distributions we need not exclude situations when $V = \emptyset$. In accordance with the formula introduced above, we get $\pi^{\downarrow\emptyset} = 1$. By $\pi(x_V)$ we denote the value of marginal probability distribution $\pi(V)$ for point $x_V \in \mathbf{X}_V$. That is, the marginal distributions can be viewed as the projection of the joint distribution onto a smaller set of variables V.

Example 2.1. *Consider a 3-dimensional distribution $\pi(u, v, w)$ given in Table 2.1. Its marginal distributions $\pi^{\downarrow\{u,v\}}$ and $\pi^{\downarrow\{v,w\}}$ are shown in Table 2.2*

π	$u = 0$		$u = 1$	
	$v = 0$	$v = 1$	$v = 0$	$v = 1$
$w = 0$	0.1	0.1	0.2	0.1
$w = 1$	0	0.1	0	0.1
$w = 2$	0.2	0	0	0.1

Table 2.1: 3-dimensional distribution

$\pi^{\downarrow\{u,v\}}$	$u = 0$	$u = 1$
$v = 0$	0.3	0.2
$v = 1$	0.2	0.3

$\pi^{\downarrow\{v,w\}}$	$v = 0$	$v = 1$
$w = 0$	0.3	0.2
$w = 1$	0	0.2
$w = 2$	0.2	0.1

Table 2.2: 2-dimensional marginal distributions

2.3 Extensions of distribution

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

$$\Pi^{(V)}(\pi) = \left\{\kappa \in \Pi^{(V)} : \kappa(U) = \pi(U)\right\}.$$

(Recall that $\kappa(U)$ is the marginal distribution of κ for variables U.) Having a system

$$\Xi = \{\pi_1(U_1), \pi_2(U_2), \ldots \pi_n(U_n)\},$$

of oligodimensional distributions ($U_1 \cup \ldots \cup U_n \subset V$), the symbol $\Pi^{(V)}(\Xi)$ denotes the system of all distributions that are extensions of all the distributions from Ξ:

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

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

2.4 Consistency

Definition 2.2. *We say that distributions $\pi(U)$ and $\kappa(V)$ are consistent if*

$$\pi^{\downarrow U \cap V} = \kappa^{\downarrow U \cap V}.$$

Remark 2.3. *Notice that if $U \cap V = \emptyset$, the distributions $\pi(U)$ and $\kappa(V)$ are always consistent.*

2.5 Conditional distribution

For a distribution $\pi(U)$ and two disjoint subsets $V, W \subseteq U$ we will often consider a *conditional distribution* $\pi(V|W)$, which is, for each fixed $x_W \in \mathbf{X}_W$, a $|V|$-dimensional probability distribution, for which

$$\pi(x_V|x_W)\pi(x_W) = \pi(x_{V \cup W})$$

for each $x \in \mathbf{X}_{V \cup W}$. It is important to realize that, if $\pi(x_W) = 0$ for some combination(s) of values $x_W \in \mathbf{X}_W$, then $\pi(x_{V \cup W}) = 0$ by the definition of a marginal distribution and the definition is ambiguous. Nevertheless, the advantage of this definition is that the conditional distribution is always defined. Observe that if $V = \emptyset$ then

$$\pi(x_V|x_W) = 1,$$

and if $W = \emptyset$ then

$$\pi(x_V|x_W) = \pi(x_V).$$

Example 2.4. *Since $\pi(w) > 0$ for all $w = 0, 1, 2$ in Table 2.1, the conditional distributions $\pi(u, v|w)$, $\pi(u|w)$, and $\pi(v|w)$ are uniquely defined by*

$$\begin{aligned} \pi(u, v|w) &= \frac{\pi(u, v, w)}{\pi(w)}, \\ \pi(u|w) &= \frac{\pi(u, w)}{\pi(w)}, \\ \pi(v|w) &= \frac{\pi(v, w)}{\pi(w)} \end{aligned}$$

respectively.

2.6 Conditional independence of variables

In this section, we introduce one of the most important notions of this text, the concept of *conditional independence*, which generalizes the well-known (unconditional) independence of variables. For more examples illustrating this notion, the reader is referred to basic textbooks like [43] or [52]. Here we introduce it along with its two most important properties.

Definition 2.5. *For a probability distribution $\pi(N)$ and three disjoint subsets $U, V, Z \subset N$ such that $U \neq \emptyset \neq V$, we say that sets of variables U and V are* conditionally independent *given Z in π (in symbol $U \perp\!\!\!\perp V|Z[\pi]$) if*

$$\pi^{\downarrow U\cup V\cup Z}(x) \cdot \pi^{\downarrow Z}(x) = \pi^{\downarrow U\cup Z}(x) \cdot \pi^{\downarrow V\cup Z}(x) \tag{2.6.1}$$

for all $x \in \mathbf{X}_{U\cup V\cup Z}$.

Observe that, if $Z = \emptyset$, then the conditional independence coincides with the well-known (unconditional) independence. Recall that the unconditional independence of variable sets U and V in π is denoted by $U \perp\!\!\!\perp V[\pi]$.

Another alternative definition of conditional independence is sometimes used. We use (2.6.1) to emphasize the *symmetry* of independence $U \perp\!\!\!\perp V|Z[\pi]$ in variables U and V:

Assertion 2.6. *For a probability distribution $\pi(N)$ and three disjoint subsets $U, V, Z \subset N$ such that $U \neq \emptyset \neq V$, it holds that*

$$U \perp\!\!\!\perp V | Z [\pi] \Leftrightarrow \forall x \in \boldsymbol{X}_{U \cup V \cup Z} : \pi(x_{U \cup Z}) > 0 \, (\pi(x_V | x_Z) = \pi(x_V | x_{U \cup Z})) \,. \quad (2.6.2)$$

Note that (2.6.2) is often used to explain the concept of conditional independence. It says that conditional probability of variables V given variables Z is the same as conditional probability of these variables given $U \cup Z$. In other words, if one knows the values of variables Z, the additional knowledge of values of U does not affect the conditional probability of V.

Assertion 2.7. ***(Factorization lemma:)*** *Let $U, V, Z \subset N$ be disjoint such that $U \neq \emptyset \neq V$. Then, for any probability distribution $\pi(U \cup V \cup Z)$, it holds that*

$$U \perp\!\!\!\perp V | Z [\pi]$$

if and only if there exists functions

$$\psi_1 : \boldsymbol{X}_{U \cup Z} \to [0, +\infty), \qquad \psi_2 : \boldsymbol{X}_{V \cup Z} \to [0, +\infty),$$

such that for all $x \in \boldsymbol{X}_{U \cup V \cup Z}$

$$\pi(x) = \psi_1(x_{U \cup Z}) \psi_2(x_{V \cup Z}).$$

Assertion 2.8. ***(Block independence lemma:)*** *Let $U, V, W, Z \subset N$ be disjoint and $U \neq \emptyset$, $V \neq \emptyset$, $W \neq \emptyset$. Then, for any probability distribution $\pi(U \cup V \cup W \cup Z)$, the following two statements are equivalent*

(a) $U \perp\!\!\!\perp V \cup W | Z [\pi]$,

(b) $U \perp\!\!\!\perp W | Z [\pi]$ *and* $U \perp\!\!\!\perp V | W \cup Z [\pi]$.

Note that this lemma is a summary of properties of conditional independence: while $(a) \Rightarrow (b)$ is sometimes called *decomposition* and *weak union* respectively, the opposite implication $(b) \Rightarrow (a)$ is usually denoted as *contraction.* These properties, combined with the property called *symmetry*, are known as the *semigraphoid axioms*. Symmetry simply states that $U \perp\!\!\!\perp V | Z [\pi] \Leftrightarrow V \perp\!\!\!\perp U | Z [\pi]$ logically.

Chapter 3

Compositional Models

A Bayesian network may be defined as a multidimensional distribution factorizable with respect to a DAG. Alternatively, it may be defined by its graph and an appropriate system of low-dimensional (oligodimensional) *conditional distributions*. Similarly, a compositional model is defined as a multidimensional distribution assembled from a sequence of oligodimensional *unconditional distributions*, with the help of operators of composition. The main advantage of both models lies in the fact that oligodimensional distributions could easily be stored in computer memory – the size of a joint probability distribution grows exponentially with the number of variables of interest. However, computing with a multidimensional distribution that is split into many pieces may be extremely complicated. The advantage of compositional models in comparison with Bayesian networks lies in the fact that compositional models explicitly express some marginals, whose computation in a Bayesian network may be demanding.

To be able to compose low-dimensional distributions to get a distribution of a higher dimension, we will introduce an operator of composition. We introduce it here in the form of a definition along with several of the, from our point of view, most important properties. They can be found in overview texts of compositional models, for example [24], or in a recent review paper, such as [27].

3.1 Operator of composition

The keystone of Compositional Models is an operator of composition $\triangleright$. It is used to compose low-dimensional distributions to get a distribution of a higher dimension. The composition is described in the following definition.

Definition 3.1. *For two arbitrary distributions $\pi(U)$ and $\kappa(V)$ for which $\pi \ll \kappa$,* *their composition is given by the Formula*

$$(\pi \triangleright \kappa)(x) = \frac{\pi(x_U)\kappa(x_V)}{\kappa(x_{U\cap V})} \tag{3.1.1}$$

for each $x \in \boldsymbol{X}_{U\cup V}$. *In a case where* $\pi \not\ll \kappa$, *the composition remains undefined. If, for any* $x \in \boldsymbol{X}_{U\cap V}$, $\kappa(x) = 0$, *then by dominance* $\pi \ll \kappa$ *and by definition of a marginal distribution, there are two zeros in the numerator and we take* $\frac{0\cdot 0}{0} = 0$

Example 3.2. *Let us illustrate Formula (3.1.1) by computing*

$$\pi(u,v) \triangleright \pi(v,w) = \frac{\pi(u,v)\pi(v,w)}{\pi(v)}$$

where the 2-dimensional distributions involved are the marginal distributions from Example 2.1, Table 2.2. This computation results in a distribution presented in Table 3.1. One can immediately see that $\pi(u,v,w) \neq \pi(u,v) \triangleright \pi(v,w)$. *Recall that the original 3-dimensional distribution corresponding to these marginals can be found in Table 2.1.*

	$u=0$		$u=1$	
	$v=0$	$v=1$	$v=0$	$v=1$
$w=0$	0.18	0.08	0.12	0.12
$w=1$	0	0.08	0	0.12
$w=2$	0.12	0.04	0.08	0.06

Table 3.1: Composed 3-dimensional distribution $\pi(u,v) \triangleright \pi(v,w)$

To make it clear from the very beginning, let us stress that this is just a generalization of the idea of computing the three-dimensional distribution from two two-dimensional ones introducing the conditional independence

$$\pi(u,v) \triangleright \kappa(v,w) = \frac{\pi(u,v)\kappa(v,w)}{\kappa(v)} = \pi(u,v)\kappa(w|v).$$

Hence, considering Factorization lemma (Assertion 2.7), application of the operator of composition introduces conditional independence of respective variables. The exact meaning of this statement can be seen from the following important assertion.

Assertion 3.3. *Let* $\kappa(U\cup V) = \pi_1(U) \triangleright \pi_2(V)$ *be defined, and* $U\setminus V \neq \emptyset \neq V\setminus U$. *Then*

$$(U \setminus V) \perp\!\!\!\perp (V \setminus U) | (U \cap V)[\kappa].$$

One can easily see from the definition of the operator of composition that it is neither commutative nor associative. However, the commutativity holds under special conditions:

	$u=0$	$u=1$
$w=0$	0.26	0.24
$w=1$	0.08	0.12
$w=2$	0.16	0.14

Table 3.2: Marginal distribution $(\pi^{\downarrow\{u,v\}} \rhd \pi^{\downarrow\{v,w\}})^{\downarrow\{u,w\}}$

Assertion 3.4. *Let $\pi \in \Pi^{(U)}$ and $\kappa \in \Pi^{(V)}$. If π and κ are consistent then*

$$\pi \rhd \kappa = \kappa \rhd \pi.$$

If either $\pi \ll \kappa$ or $\kappa \ll \pi$ (i.e., $\pi \rhd \kappa$ or $\kappa \rhd \pi$ is defined respectively) then the reverse implication also holds true:

$$\pi \rhd \kappa = \kappa \rhd \pi \Rightarrow \pi \text{ and } \kappa \text{ are consistent.}$$

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

Assertion 3.5. *Let π, κ be probability distributions from $\Pi^{(U)}, \Pi^{(V)}$ respectively. If $\pi \ll \kappa$ (i.e., if $\pi \rhd \kappa$ is defined) then $\pi \rhd \kappa$ is a probability distribution from $\Pi^{(U\cup V)}(\pi)$, i.e., it is a probability distribution and its marginal distribution for variables U equals π:*

$$(\pi \rhd \kappa)(x_U) = \pi(x_U)$$

for all $x_U \in \boldsymbol{X}_U$.

In the proofs, we shall often compute a marginal distribution from a distribution defined as a composition model of two (or more) low-dimensional distributions. Therefore, it is important to realize that generally for $\pi_1(U_1), \pi_2(U_2)$ and $V \subseteq U_1 \cup U_2$

$$(\pi_1 \rhd \pi_2)^{\downarrow V} \neq \pi_1^{\downarrow U_1\cap V} \rhd \pi_2^{\downarrow U_2 \cap V}. \tag{3.1.2}$$

To illustrate the situation when equality in Formula (3.1.2) does not hold, consider $\pi^{\downarrow\{u,v\}} \rhd \pi^{\downarrow\{v,w\}}$ from Example 3.2 (see Table 3.1) and its marginal distribution $(\pi^{\downarrow\{u,v\}} \rhd \pi^{\downarrow\{v,w\}})^{\downarrow\{u,w\}}$, which is shown in Table 3.2. Examining this marginal distribution, we see that variables u and w are not independent. Therefore

$$\begin{aligned}(\pi(u,v) \rhd \pi(v,w))^{\downarrow\{u,w\}} &\neq (\pi(u,v))^{\downarrow u} \rhd (\pi(v,w))^{\downarrow w} \\ &= \pi(u) \rhd \pi(w) = \pi(u)\pi(v).\end{aligned}$$

The following simple assertion presents a sufficient condition under which the equality in (3.1.2) holds.

Assertion 3.6. *Let $U_1, U_2, V \subseteq N$. If $U_1 \cap U_2 \subseteq V \subseteq U_1 \cup U_2$, then for any probability distributions $\pi_1(U_1)$ and $\pi_2(U_2)$ it holds that*

$$(\pi_1 \rhd \pi_2)^{\downarrow V} = \pi_1^{\downarrow U_1\cap V} \rhd \pi_2^{\downarrow U_2 \cap V}.$$

3.2 Generating sequences

In this section we will start considering *multidimensional compositional models*, i.e., multidimensional probability distributions assembled from sequences of oligodimensional distributions using the operator of composition. The result of the composition (if defined) is a new distribution. We can iteratively repeat the process of composition to obtain a multidimensional distribution. That is why such a multidimensional distribution is called a *compositional model*.

To simplify the following considerations, let us present three important conventions. In this and the following chapters we will consider a system of n oligodimensional distributions $\pi_1(U_1)$, $\pi_2(U_2), \ldots,$ $\pi_n(U_n)$. Therefore, whenever distribution π_i is used, if not specified otherwise, the distribution will be assumed to be a distribution from $\Pi^{(U_k)}$, which means it will be a distribution $\pi(U_i)$. Thus, the formula $\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n$, if it is defined, will determine the distribution of variables $U_1 \cup U_2 \cup \ldots \cup U_n$. Our second convention concerns the fact that the operator $\triangleright$ is neither commutative nor associative. To avoid writing too many parentheses in the formulas, we will always apply the operator from left to right. Thus,

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

Therefore, in order to construct such a model it is sufficient to determine a sequence of oligodimensional distributions $\pi_1, \pi_2, \ldots, \pi_n$ – we call it a *generating sequence*. Note that there are situations in which the result of the composition is not defined. That is why, whenever we speak about a generating sequence in the following, we always assume that the respective compositional model is well defined. And this is our third convention.

Considering Assertion 3.5 and the fact that we apply the operator of composition from left to right, one can easily determine marginal distributions of a compositional model for variables first appearing in the arguments of the first distribution in the corresponding generating sequence. Note that this assertion will be crucial in the following text, particularly in Chapters 8 and 9.

Assertion 3.7. *Consider a compositional model π with a generating sequence $\pi_1(U_1)$, $\pi_2(U_2)$, $\ldots, \pi_n(U_n)$ ($\pi = \pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n$). Then, for all $i = 1, \ldots, n$,*

$$\pi^{\downarrow U_1 \cup \ldots \cup U_i} = \pi_1 \triangleright \ldots \triangleright \pi_i.$$

Among other things, it means that π_1 equals a marginal of $\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n$ for variables in arguments of π_1, which will mostly be used in subsequent proofs.

3.2.1 Perfect sequences

Not all generating sequences are equally efficient in representing multidimensional distributions. Among them, so-called perfect sequences hold an important position [28]. From the original definition, one can hardly see the importance of this class of generating sequences. Instead, for the purpose of this text, let us define it by another equivalent property which is more suitable for our needs.

Definition 3.8. *A sequence of probability distributions $\pi_1, \pi_2, \ldots, \pi_n$ is perfect iff $\pi = \pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n$ is defined and all the distributions from this sequence are marginals of the distribution π.*

A compositional model defined by a perfect sequence is called a *perfect sequence model.* Perfect sequences have many beneficial properties which are advantageous for representing multidimensional distributions. One of them – a direct consequence of the definition – is frequently used in other parts of this text.

Assertion 3.9. *If a sequence of distributions is perfect, then all distributions in this sequence are pairwise consistent.*

Remark 3.10. *Notice that when defining a perfect sequence, let alone a generating sequence, we do not impose any conditions on sets of variables for which the distributions are defined. For example, considering a generating sequence where one distribution is defined for a subset of variables of another distribution (i.e., $U_i \subset U_j$) is fully meaningful and may carry information about the distribution. If, for $\pi(u, v, w)$, we state $\pi^{\downarrow u}, \pi^{\downarrow v}, \pi$ as a perfect sequence, it is obvious that*

$$\pi^{\downarrow u} \triangleright \pi^{\downarrow v} \triangleright \pi = \pi$$

(because all the elements of a perfect sequence are marginals of the resulting distribution and therefore π must be marginal to $\pi^{\downarrow u} \triangleright \pi^{\downarrow v} \triangleright \pi$). Nevertheless, it can happen that for some reason it may be advantageous to work with model π. From the model one can immediately see that variables u and v are independent, which, not knowing the numbers defining the distribution, one cannot say about distribution π. (How to read all the conditional independence relations from the structure of a compositional model is shown in Section 3.4.)

3.3 Model structure

Consider a compositional model defined by a generating sequence $\pi_1(U_1), \pi_2(U_2), \ldots, \pi_n(U_n)$. Then the sequence of sets $U_1, U_2, \ldots, U_n$ is called its *model structure* and it is usually denoted by symbol $\mathcal{P}$. If not specified otherwise, $\mathcal{P} = U_1, \ldots, U_n$ where $(U_1 \cup \ldots \cup U_n) = N$, and we say that $\mathcal{P}$ is defined over N and $U_i \in \mathcal{P}$ for every $i \in \{1, \ldots, n\}$. Moreover, we recognize the auxiliary sets $K_i^{\mathcal{P}}$ which reflect the ordering in $\mathcal{P}$ – $K_i^{\mathcal{P}}$ is the i-th set from $\mathcal{P}$; e.g., for $\mathcal{P} = U_1, \ldots, U_n$ it holds that $K_i^{\mathcal{P}} = U_i$ for all $i = 1, \ldots, n$.

The reason for the double notation of the same set within this structure is as follows: Consider $\mathcal{P} = U_1, U_2, U_3, U_4$ and let $\mathcal{P}'$ be a permutation, for example, such that $\mathcal{P}' = U_3, U_1, U_4, U_2$. Then U_3 is the first set in sequence $\mathcal{P}'$ and U_1 is the second set in $\mathcal{P}'$. This can now be easily expressed as $U_3 \equiv K_1^{\mathcal{P}'}$ and $U_1 = K_2^{\mathcal{P}'}$.

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

$$\begin{aligned} R_1^{\mathcal{P}} &= K_1^{\mathcal{P}} \\ R_i^{\mathcal{P}} &= K_i^{\mathcal{P}} \backslash (K_1^{\mathcal{P}} \cup \ldots \cup K_{i-1}^{\mathcal{P}}) \ \forall i = \{2, \ldots, |\mathcal{P}|\} \end{aligned}$$

and

$$\begin{aligned} S_1^{\mathcal{P}} &= \emptyset \\ S_i^{\mathcal{P}} &= K_i^{\mathcal{P}} \cap (K_1^{\mathcal{P}} \cup \ldots \cup K_{i-1}^{\mathcal{P}}) \ \forall i = \{2, \ldots, |\mathcal{P}|\} \end{aligned}$$

It has the following meaning: $R_i^{\mathcal{P}}$ denotes the variables first occurring in the i-th set of the sequence $\mathcal{P}$ (taken from left to right). Conversely, $S_i^{\mathcal{P}}$ denotes variables from the i-th set of $\mathcal{P}$ which have already been used in a foregoing set. Observe that $K_i^{\mathcal{P}} = R_i^{\mathcal{P}} \cup S_i^{\mathcal{P}}$. The super index $\mathcal{P}$ may be omitted if the context is clear. $|\mathcal{P}|$ denotes the number of sets in the structure, i.e., $|\mathcal{P}| = n$ for $\mathcal{P} = U_1, \ldots, U_n$.

Example 3.11. *For a generating sequence* $\pi_1(u)$, $\pi_2(v, w)$, $\pi_3(u, v, x)$, $\pi_4(w, x, y)$, $\pi_5(x, y, z)$, *its structure is* $\mathcal{P} = \{u\}, \{v, w\}, \{u, v, x\}, \{w, x, y\}, \{w, y, z\}$ *and* $|\mathcal{P}| = 5$.

$$\begin{array}{lll} K_1^{\mathcal{P}} = u & R_1^{\mathcal{P}} = u & S_1^{\mathcal{P}} = \emptyset \\ K_2^{\mathcal{P}} = \{v, w\} & R_2^{\mathcal{P}} = \{v, w\} & S_2^{\mathcal{P}} = \emptyset \\ K_3^{\mathcal{P}} = \{u, v, x\} & R_3^{\mathcal{P}} = x & S_3^{\mathcal{P}} = \{u, v\} \\ K_4^{\mathcal{P}} = \{w, x, y\} & R_4^{\mathcal{P}} = y & S_4^{\mathcal{P}} = \{w, x\} \\ K_5^{\mathcal{P}} = \{w, y, z\} & R_5^{\mathcal{P}} = z & S_5^{\mathcal{P}} = \{x, y\} \end{array}$$

To be able to simply handle characteristic properties of the respective structures, we introduce a function

$$]\cdot[_{\mathcal{P}} \colon 2^N \to \{1, \ldots, |\mathcal{P}|\}$$

such that, for fixed structure $\mathcal{P}$, $U \subseteq N$, $]U[_{\mathcal{P}} = max_{u \in U}\{i : u \in R_i^{\mathcal{P}}\}$. Hence $]U[_{\mathcal{P}}$ equals the maximal index i such that $u \in U$ and $u \in R_i^{\mathcal{P}}$. Due to the previously established notation, it can be said that $K_{]u[_{\mathcal{P}}}^{\mathcal{P}}$ is that set $K_i^{\mathcal{P}}$ for which $u \in R_i^{\mathcal{P}}$, i.e., $]u[_{\mathcal{P}} = i : u \in R_i^{\mathcal{P}}$. The symbol $\mathcal{P}$ may be omitted in $]u[_{\mathcal{P}}$ if the context is clear – for example when dealing with only one structure.

Example 3.12. *Consider structure $\mathcal{P}$ from Example 3.11. One can read the following properties: $]u[= 1$, $]\{u,v\}[= 2$, $]\{u,w\}[= 2$, $]x[= 3$, $]y[= 4$, $]\{u,v,w,x,y,z\}[= 5$, and $]z[= 5$. Similar to Example 3.11*

$$\begin{array}{lll} K^{\mathcal{P}}_{]u[} = u & R^{\mathcal{P}}_{]u[} = u & S^{\mathcal{P}}_{]u[} = \emptyset \\ K^{\mathcal{P}}_{]v[} = \{v,w\} & R^{\mathcal{P}}_{]w[} = \{v,w\} & S^{\mathcal{P}}_{]\{v,w\}[} = \emptyset \end{array}$$

etc.

Definition 3.13. *For a structure $\mathcal{P}$ over N we introduce a binary relation $\preceq_{\mathcal{P}}$ such that, for two non-empty sets of variables $U, V \subseteq N$, the relation $U \preceq_{\mathcal{P}} V$ holds if and only if $]U[_{\mathcal{P}} \leq]V[_{\mathcal{P}}$. Moreover, we introduce its strict version $\prec_{\mathcal{P}}$. in which $U \prec_{\mathcal{P}} V$ if and only if $]U[_{\mathcal{P}} <]V[_{\mathcal{P}}$.*

The symbol $\mathcal{P}$ may be omitted in $\prec_{\mathcal{P}}$ and $\preceq_{\mathcal{P}}$ if the context is clear.

Example 3.14. *Consider the structure from Example 3.11 again. According to the former Definition, one can see that $u \prec v \preceq w \prec x \prec y \prec z$ in that structure. Similarly, for subsets $\{u,v,w\} \prec z$, $\{u,v\} \preceq w$, $\{u,x\} \prec y$, etc.*

3.3.1 Persegrams

To visualize the structure of a compositional model (and its generating sequence) we use a tool called a *persegram.* This visualization tool was designed during development of a technique to read conditional independence relations from the structure of a compositional model [25].

Definition 3.15. Persegram *of a structure $\mathcal{P} = U_1, U_2, \ldots, U_n$ is a table in which rows correspond to variables from $U_1 \cup U_2 \cup \ldots \cup U_n$ (in an arbitrary order) and columns to sets of variables U_i for all $i \in \{1, \ldots, n\}$; ordering of the columns corresponds to the structure order. A position in the table is marked if the respective set contains the corresponding variable. Markers for the first occurrence of each variable (i.e., the leftmost markers in rows) are* box-markers*, and for other occurrences there are* bullets*.*

Example 3.16. *Let $\mathcal{P} = U_1, \ldots, U_5$ be a structure of a compositional model such that $U_1 = \{u\}, U_2 = \{v,w\}, U_3 = \{u,v,x\}, U_4 = \{w,x,y\}, U_5 = \{x,y,z\}$. Note that this is the very same structure as the one in Example 3.11. Since the row ordering is not specified in Definition 3.15, the corresponding persegram can be visualized not only as in Figure 3.1a, but also in many other ways. See another such persegram in Figure 3.1b.*

Since markers in the i-th column of the persegram corresponding to structure $\mathcal{P}$ corresponds to variables $K^{\mathcal{P}}_i$, we usually call sets from a structure its *columns.*

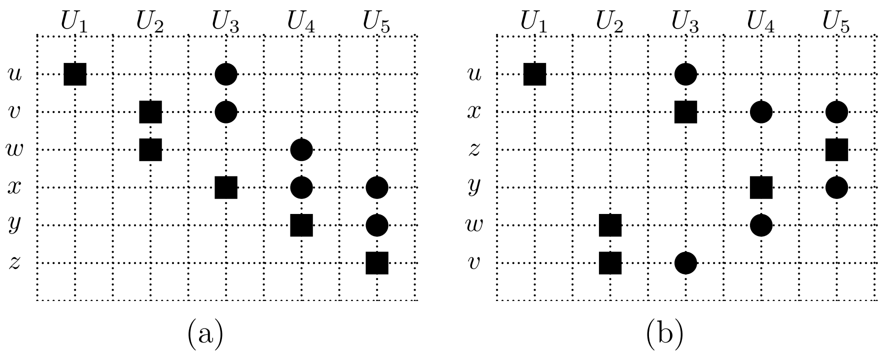


Figure 3.1: Different persegrams belonging to one model structure $\mathcal{P}$

Observe that bullets in the i-th column correspond to variables from $S_i^{\mathcal{P}}$ while box-markers to variables from $R_i^{\mathcal{P}}$. Compare results of Example 3.11 with Figure 3.1.

3.4 Induced independence relations

As stated in the introduction, while a model is put together, a system of (un)conditional independencies valid for the represented multidimensional probability distribution is simultaneously introduced by the structure of the generating sequence (see Assertion 3.3). Note that the induced independence relation mentioned in Assertion 3.3 is guaranteed solely by the structure of the generating sequence.

Example 3.17. *Let $\{u, v\} = N$, $u \neq v$. $\pi_1(u), \pi_2(v)$ be a generating sequence of a compositional model $\pi_1 \triangleright \pi_2$. Then $u \perp\!\!\!\perp v[\pi_1 \triangleright \pi_2]$. Indeed, by applying the operator of composition one gets*

$$\pi_1(u) \triangleright \pi_2(v) = \frac{\pi_1(u)\pi_2(v)}{\pi_2(\emptyset)} = \pi_1(u)\pi_2(v),$$

which corresponds to the definition of independence between variables u and v.

Similarly, assume that $\{u, v, w\} = N$ are three distinct variables $\pi_1(u, w)$, and $\pi_2(v, w)$ is a generating sequence of a compositional model $\pi_1 \triangleright \pi_2$. Observe that $u \perp\!\!\!\perp v|w[\pi_1 \triangleright \pi_2]$ by Assertion 3.3.

The more complex the model structure, the more difficult the seeking of induced independencies is. Note that independencies induced by a structure may not be all independencies valid for a multidimensional probability distribution represented by a compositional model with this structure. Some independencies are implied by properties of the respective low-dimensional distributions from the respective generating sequence. However, the set of independence relations induced by a structure is valid for any compositional model with this structure regardless of the generating distributions properties. That is the reason we speak about independence relations induced by a structure and why we connect it with structures instead of probability distributions in the discussion that follows.

Obviously, one should be able to read induced independencies directly from the structure. To increase the lucidity and readability of this text, we have decided to use a persegram as a visualization of a structure, and we present a procedure for reading the induced independencies using this tool. We demonstrate how to read the induced conditional independence relations from a persegram representing the structure of a compositional model in this section. Let us reveal that such independencies are indicated by the absence of a *trail connecting relevant markers and avoiding others*, which is defined below.

Definition 3.18. *A sequence of markers $m_0, \ldots, m_t$ of a persegram corresponding to structure $\mathcal{P}$ over N is called a Z-*avoiding trail *($Z \subseteq N$) that connects m_0 and m_t if it meets the following five conditions:*

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

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

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

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

4. a) vertical and horizontal connections regularly alternate with the following possible exception:

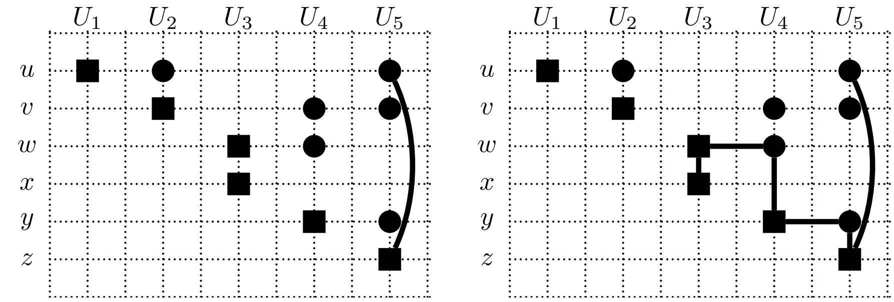
b) at most two vertical connections may be in direct succession if their common adjacent marker is a box-marker of a variable from Z;

If a Z-avoiding trail connects two markers corresponding to variables u and v, we say that these variables are connected by a Z-avoiding trail. *This situation will be denoted by $u \not\perp\!\!\!\perp v | Z[\mathcal{P}]$.*

By investigating Definition 3.18 further, the reader will find that no condition of this definition is dependent on the order of rows in the considered persegram. That would be inappropriate, because all persegrams representing the structure of a generating sequence are equivalent, regardless of the row ordering. (See the definition of persegram – Definition 3.15). Then the system of Z-avoiding trails induced by a persegram can be obtained by any other persegram of the considered structure. In the sense of the previous definition, all persegrams corresponding to $\mathcal{P}$ are equivalent.

Example 3.19. *Consider structure $\mathcal{P}$ shown in Figure 3.2. There are two different sequences of markers highlighted in the Figure, each in its respective part of the Figure. In order to illustrate vertical and horizontal connections and to highlight the ordering, each two consecutive markers are connected with a line.*

There is a sequence of markers $[U_5, u], [U_5, z]$ in Figure 3.2a. Considering $Z = \emptyset$, it forms a Z-avoiding trail connecting u and z. However, considering Definition 3.18 we can see that, this sequence avoids many other variables and Z may have a variable content. In fact, Z can be any subset of $\{v, w, x, y\}$. Hence, $u \not\!\perp\!\!\!\perp v | Z[\mathcal{P}]$ for any (even empty) $Z \subseteq \{v, w, x, y\}$.



(a): $\mathcal{P}$: $u \not\!\perp\!\!\!\perp z | \emptyset$, $u \not\!\perp\!\!\!\perp z | \{v, w, x, y\}$ (b): $\mathcal{P} : u \not\!\perp\!\!\!\perp x | \{z\}$, $u \not\!\perp\!\!\!\perp x | \{v, z\}$

Figure 3.2: Different trails connecting u with some other variables

Similarly, the sequence of markers $[U_5, u]$, $[U_5, z]$, $[U_5, y]$, $[U_4, y]$, $[U_4, w]$, $[U_3, w]$, $[U_3, x]$ in Figure 3.2b is a Z-avoiding trail where $Z = z$. Contrary to Figure 3.2a, one cannot replace z by any other variable since z has to always be part of Z in this case. Indeed, otherwise the Condition 4.b in Definition 3.18 would be corrupted. However, the trail from Figure 3.2b is $\{v, z\}$-avoiding too.

The *(un)conditional (in)dependencies induced by a structure* are introduced with the help of Z-avoiding trails.

Definition 3.20. *Consider a structure $\mathcal{P}$ over N and three disjoint subsets $U, V, Z \subset N$ such that $U \neq \emptyset \neq V$. The sets of variables U and V are* conditionally independent given Z in $\mathcal{P}$ *(in symbol $U \perp\!\!\!\perp V | Z[\mathcal{P}]$), if no $u \in U$ is connected with any $v \in V$ by a Z-avoiding trail in the corresponding persegram. Otherwise U and V are* conditionally dependent given by Z in $\mathcal{P}$, *written $U \not\!\perp\!\!\!\perp V | Z[\mathcal{P}]$.*

The induced independence model $\mathcal{I}(\mathcal{P})$ *and the* induced dependence model $\mathcal{D}(\mathcal{P})$ *of structure $\mathcal{P}$ are defined as follows:*

$$\mathcal{I}(\mathcal{P}) = \{\langle U, V | Z \rangle \in \mathcal{T}(N); U \perp\!\!\!\perp V | Z[\mathcal{P}]\}$$

$$\mathcal{D}(\mathcal{P}) = \{\langle U, V | Z \rangle \in \mathcal{T}(N); U \not\!\perp\!\!\!\perp V | Z[\mathcal{P}]\},$$

where the symbol $\mathcal{T}(N)$ denotes the class of all disjoint triplets over N:

$$\mathcal{T}(N) = \{\langle U, V | Z \rangle : U, V, Z \subseteq N, U \neq \emptyset \neq V, U \cap V = V \cap Z = Z \cap U = \emptyset\}$$

To avoid any misunderstanding, when talking about a triplet $\langle u, v|Z\rangle$ in the discussion that follows, we always consider a disjoint triplet.

Remark 3.21. *Realize that, for any structure $\mathcal{P}$ over N, its induced dependence model $\mathcal{D}(\mathcal{P})$ (and hence independence model $\mathcal{I}(\mathcal{P})$ as well) is uniquely determined by the set of "elementary relations" $\langle u, v|Z\rangle$ where u and v are singletons. Indeed, recall that there is a one-to-one correspondence between the fact $\langle u, v|Z\rangle \in \mathcal{D}(\mathcal{P})$ and the fact that there exists a Z-avoiding trail connecting u with v in the corresponding persegram ($u \not\perp\!\!\!\perp v|Z[\mathcal{P}]$). And these trails are just what determines whether $\langle U, V|Z\rangle \in \mathcal{D}(\mathcal{P})$ for any disjoint $U, V \subset N$ by the previous definition.*

The concept of induced *(in)dependencies* lives up to expectations that there is a parallel between this and independencies valid in any compositional model with the respective structure. The connection between independence read from a compositional model and from its structure (persegram) is elucidated by the following theorem. The proof of this assertion is rather technical and requires results proven in lemmata found in other research papers. The reader is referred to [25].

Theorem 3.22. *Consider a generating sequence $\pi_1, \ldots, \pi_n$ with structure $\mathcal{P}$ over N and three disjoint subsets $U, V, Z \subseteq N$ such that $U \neq \emptyset \neq V$. Then:*

$$U \perp\!\!\!\perp V|Z[\mathcal{P}] \Rightarrow U \perp\!\!\!\perp V|Z[\pi_1 \triangleright \ldots \triangleright \pi_n].$$

It is important to realize that (analogous to Bayesian networks or decomposable models) one can be sure about the validity of the indicated independence relations for any distribution which is represented by a compositional model with the given structure.

3.4.1 Other preliminaries

A trivial fact follows from Definition 3.18. It concerns variables appearing for the first time in the last column. Before we introduce this fact in the form of a lemma, let us illustrate it with the help of the following example.

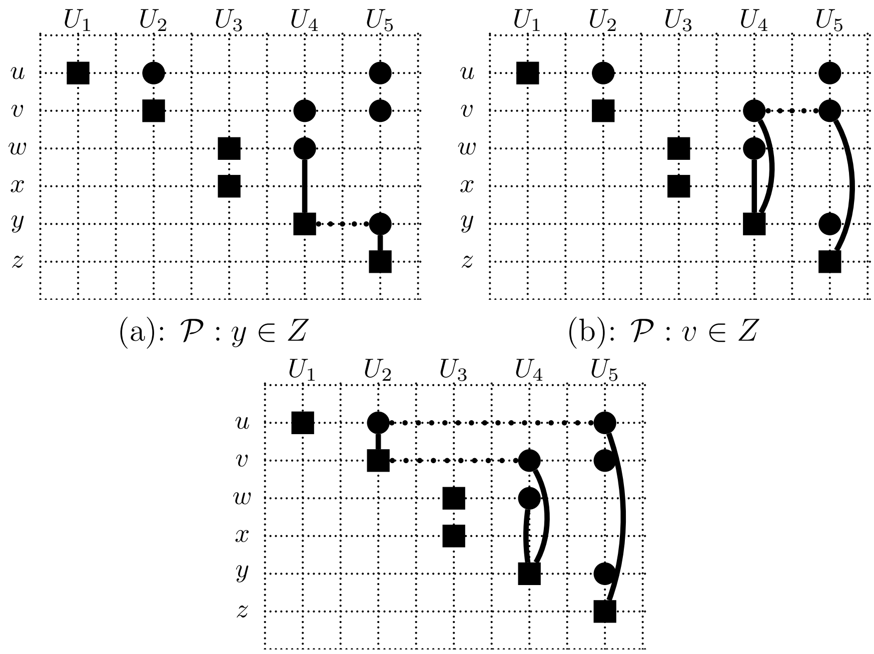
Example 3.23. *Consider structure $U_1, \ldots, U_5$ from Figure 3.3. Let us show that there is no S_5-avoiding trail connecting $z \in R_5$ (first appearing in the last column) with $w \notin U_5$ (not belonging to the last column). Let us try to construct such a sequence of markers forming an S_5-avoiding trail. Observe that $S_5 = \{u, v, y\}$.*

Three different sequences of markers are shown in Figure 3.3. Let us summarize requirements for Z that are necessary for these sequences to be Z-avoiding trails:

- *Consider the sequence of markers highlighted in Figure 3.3a: By Condition 3 of Definition 3.18 (i.e., no horizontal connection corresponds to a variable from Z), Z must not contain a variable y ($y \notin Z$).*

- *For the sequence of markers in Figure 3.3b: Similarly, $v \notin Z$ for the same reason and $y \in Z$ by Condition 4 of Definition 3.18 (i.e., two vertical connections may be in direct succession if their common adjacent marker is a box-marker corresponding to a variable from Z).*

- *Figure 3.3c: $u, v \notin Z$, $y \in Z$ for the same reasons.*



(a): $\mathcal{P} : y \in Z$ (b): $\mathcal{P} : v \in Z$

(c): $\mathcal{P} : u, v \in Z$

Figure 3.3: Different trails violating Condition 3. of Definition 3.18 for $Z = \{u, v, y\}$.

Combining all of those restrictions on Z together, one gets the following corollary: By choosing $Z = S_5 = \{u, v, y\}$, none of the above-discussed sequences forms an S_5-avoiding trail, since each of them contains a horizontal connection corresponding to a variable from S_5. These horizontal connections violating Condition 3 of Definition 3.18 are drawn by dotted lines. Since there is no other possible S_5-avoiding trail between w and z, $w \perp\!\!\!\perp z | S(K_5)$ by Definition 3.20.

Lemma 3.24. *Consider a structure $\mathcal{P}$ over N and two distinct variables $u, v \in N$ such that $u \in R^{\mathcal{P}}_{|\mathcal{P}|}$ and $v \notin K^{\mathcal{P}}_{|\mathcal{P}|}$. Then $u \perp\!\!\!\perp v | S^{\mathcal{P}}_{|\mathcal{P}|}[\mathcal{P}]$.*

Proof. Suppose that $|\mathcal{P}| = n$ and consider the respective persegram. Since u belongs to the last column of $\mathcal{P}$ ($u \in R_n \Rightarrow u \in K_n$), every trail from u has to begin with a vertical connection in K_n to a marker corresponding to a variable from S_n (otherwise, in a case where the vertical connection joins two variables from R_n, the horizontal and vertical connections could not regularly alternate). However, no S_n-avoiding trail may contain a horizontal connection corresponding to a variable from S_n, and such a trail must not contain any marker out of the last column. Since $u \notin K_n$, a trail representing $u \not\perp\!\!\!\perp v | S_n$ cannot exist; therefore $u \perp\!\!\!\perp v | S_n[\mathcal{P}]$ by Definition 3.20. $\square$

To simplify the following, we introduce the concept of the *substructure induced by a set of variables*. Unlike the subgraph which contains exactly those variables that induce it, the substructure is usually defined for a superset.

Definition 3.25. *A* substructure *of a structure $\mathcal{P}$ over N induced by a set $U \subseteq N$ is its minimal left part containing all variables from U, i.e., $\mathcal{P}[U] = K_1^{\mathcal{P}}, \ldots, K_{]U[}^{\mathcal{P}}$.*

A persegram of $\mathcal{P}[U]$ is created from a persegram of $\mathcal{P}$ by removing columns to the right of the one with the farthest right box-marker corresponding to a variable from U.

Example 3.26. *Consider a structure $\mathcal{P} = U_1, \ldots, U_5$ shown in Figure 3.4a. Put $U = \{w, x\}$. One can find the corresponding substructure $\mathcal{P}[U]$ in Figure 3.4b. Observe that $\mathcal{P}[U]$ is defined not only over $\{w, x\}$, but also over $\{u, v\}$.*

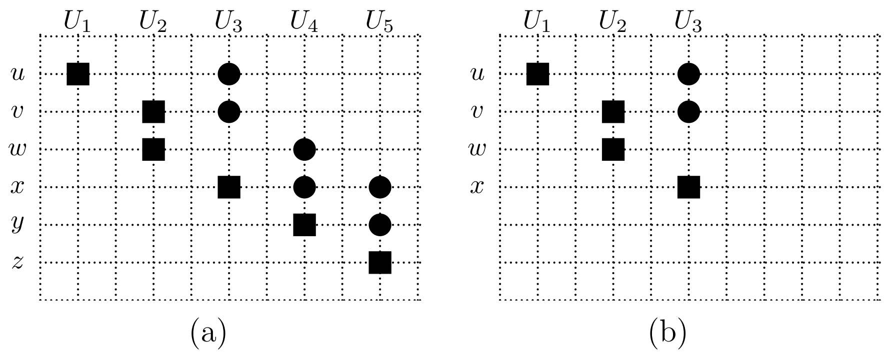


Figure 3.4: Structure $\mathcal{P}$ and its corresponding substructure $\mathcal{P}[U]$ for $U = \{w, x\}$

Remark 3.27. *Accepting the suggested concept that markers from persegram corresponding to substructure $\mathcal{P}[U]$ coincide with some markers from persegram of the respective structure $\mathcal{P}$, we can observe the following fact: For given $U \subseteq N$ and $Z \subset U$, any sequence of markers forming a Z-avoiding trail in a persegram of $\mathcal{P}[U]$ forms a Z-avoiding trail in a persegram of $\mathcal{P}$.*

The concept of an induced substructure brings one very important advantage. Searching for an area for Z-avoiding trails connecting u with v may be restricted to a persegram corresponding to the respective substructure induced by $\{u,v\}\cup Z$ only.

Lemma 3.28. *Consider a persegram of structure $\mathcal{P}$ over N, $u, v \in N$, and $Z \subseteq N \setminus \{u,v\}$. If $u \not\perp\!\!\!\perp v|Z[\mathcal{P}]$, then all Z-avoiding trails connecting u with v are present in the persegram corresponding to $\mathcal{P}[\{u,v\} \cup Z]$.*

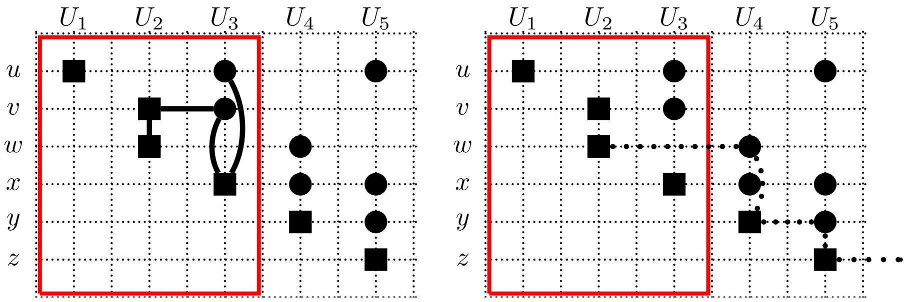
Proof. Assume by contradiction that there is a Z-avoiding trail representing $u \not\perp\!\!\!\perp v|Z[\mathcal{P}]$ containing markers out of the area defined by $\mathcal{P}[\{u,v\} \cup Z]$. We are going to show that if the Z-avoiding trail from u leaves the area defined by $\mathcal{P}[\{u,v\} \cup Z]$, then it cannot end up in v which contradicts the assumption. To understand our way of thinking, the reader should have a careful look at Figure 3.5b while considering the proof.

Assume that τ is a Z-avoiding trail representing $u \not\perp\!\!\!\perp v|Z[\mathcal{P}]$ with a marker out of $\mathcal{P}[\{u,v\} \cup Z]$, i.e., $\tau = m_0, \ldots, m_t$ is a sequence of markers where m_0 corresponds to u. Let m_i be the first marker in the sequence τ outside of the area corresponding to $\mathcal{P}[\{u,v\} \cup Z]$. Since it is the first marker outside of the area, a horizontal connection had to be used between m_{i-1} and m_i, and therefore m_i has to be a bullet. By Definition 3.18, the trail now has to continue with a vertical connection to a box-marker. Since this box-marker cannot correspond to any variable from $Z \cup v$ (it is outside of $\mathcal{P}[\{u,v\} \cup Z]$), one has to continue with a horizontal connection (to the right of the box-marker which is always the first marker in a row, and there is nothing on the left in the same row) to a bullet. Then a vertical connection to a box-marker has to follow again. It does not correspond to any variable from $Z \cup v$, etc. From such a trail τ, there is no return to v. Therefore, the respective Z-avoiding trail cannot exist, which contradicts the assumption. □

Example 3.29. *Let us illustrate the idea of proving Lemma 3.28. Consider the structure shown in Figure 3.5, its corresponding structure $\mathcal{P}$, and $Z = \{x\}$. First, we show that there is only one x-avoiding trail representing $u \not\perp\!\!\!\perp w|x[\mathcal{P}]$. One can find it in the highlighted area corresponding to $\mathcal{P}[\{u,w,x\}]$ in Figure 3.5a.*

Let us try to create an x-avoiding trail from w to u containing markers outside of the highlighted part corresponding to $\mathcal{P}[\{u,w,x\}]$. Such an experiment is depicted by the dotted line in Figure 3.5b.

Let us start with box-marker $[U_2, w]$ and continue to $[U_4, w]$ outside of $\mathcal{P}[\{u,w,x\}]$. To satisfy Definition 3.18 of a Z-avoiding trail, one has to continue with a vertical connection to a box-marker. (The only possible box-marker is $[U_4, y]$). Since $y \notin Z = x$, then by Condition 4 of Definition 3.18, one has to continue with a horizontal connection (to the right – there is nothing left of any box-marker)



(a): The shortest x-avoiding trail connecting u with w. It is located in the area induced by $\{u, w, x\}$

(b): An attempt to create a x-avoiding trail outside the area corresponding to the induced substructure.

Figure 3.5: Illustration of Lemma 3.28

to a bullet, etc. Since there is no box-marker corresponding to u, x *outside of* $\mathcal{P}[\{u, w, x\}]$*, the trail cannot end in* u*. Hence it cannot exist.*

Lemma 3.28 basically means that, if we are interested in relation $u \perp\!\!\!\perp v | Z [\mathcal{P}]$, we may focus only on the subpersegram $\mathcal{P}[\{u, v\} \cup Z]$. This observation is summarized in the following corollary – a trivial consequence of Lemma 3.28 and Remark 3.27.

Corollary 3.30. *Consider a structure* $\mathcal{P}$ *over* N*, two distinct variables* $u, v \in N$*, and* $Z \subseteq N \setminus \{u, v\}$*. Then*

$$u \perp\!\!\!\perp v | Z \ [\mathcal{P}[\{u, v\} \cup Z]] \Leftrightarrow u \perp\!\!\!\perp v | Z \ [\mathcal{P}].$$

Chapter 4

Equivalence problem

The equivalence problem is understood as a problem of how to recognize whether two given structures $\mathcal{P}$ and $\mathcal{P}'$ over the same set of variables N induce the same independence model ($\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}')$). A very readable overview of the solution to this problem in the area of structures represented by a DAG may be found in [30].

It is of special importance to have a simple rule to recognize that two structures are equivalent in this sense (the notion of a rule's simplicity may differ when considering whether people or a computer will use it), and an easy way to convert $\mathcal{P}$ into $\mathcal{P}'$ in terms of some elementary operations on structures. These issues are addressed in [32], [33] and [34]. Another very important aspect is the ability to generate all structures which are equivalent to a given structure.

This text covers the solution of all above-mentioned subproblems. In Chapter 5, we introduce and describe several properties of a model structure which are invariable in the class of equivalent structures. This means that they are necessary to guarantee the equivalence of different structures. They include the so-called *connection set* and *F-condition set*, *non-trivial sets, strong and weak core*, and *formal ratio* of the respective structure.

In Chapter 6 we introduce four elementary operations, (the so called *IE operations*) that allow us to convert any structure into another independence-equivalent one and (as shown in Chapter 7) to generate the complete class of structures that are equivalent with the given one – the so-called equivalence class.

We summarize it all in Chapter 7, where we show that some of the invariant properties from Chapter 5 (or their combinations) are not only necessary, but also sufficient to guarantee the equivalence of considered structures. In other words, these properties are real characteristics of equivalent structures.

Definition 4.1. *Structures $\mathcal{P}, \mathcal{P}'$ (over the same variable set N) are called* independence equivalent*, if they induce the same independence model $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}')$.*

Remark 4.2. *One may easily see that the above-mentioned definition could be formulated using a dependence model instead. Structures $\mathcal{P}, \mathcal{P}'$ (over the same variable set N) are independence equivalent iff $\mathcal{D}(\mathcal{P}) = \mathcal{D}(\mathcal{P}')$. This alternative is used in most of the proofs.*

Example 4.3. *1. Consider two simple structures $\mathcal{P}_1, \mathcal{P}_1'$ over $\{u, v\}$ shown in Figure 4.1. Since there is no possible vertical connection in both persegrams, there can be no Z-avoiding trail for any Z. Therefore $u \perp\!\!\!\perp v|\emptyset$ in both $\mathcal{P}_1$ and $\mathcal{P}_1'$. Hence*

$$\mathcal{I}(\mathcal{P}_1) = \mathcal{I}(\mathcal{P}_1') = \{\langle u, v|\emptyset\rangle\}.$$

The corresponding structures are independence equivalent.

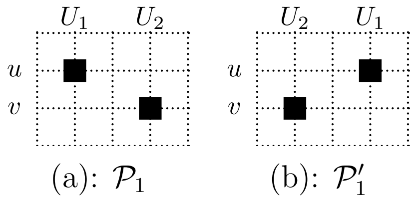


Figure 4.1: Two equivalent structures

2. On the other hand, structures with the same sets ($U \in \mathcal{P} \Leftrightarrow U \in \mathcal{P}'$) in a different ordering need not be equivalent. Let $N = \{u, v, w\}$ and consider the following two structures $\mathcal{P}_2, \mathcal{P}_2'$ from Figure 4.2. Observe that $u \perp\!\!\!\perp v|\emptyset[\mathcal{P}_2]$ but $u \not\perp\!\!\!\perp v|\emptyset[\mathcal{P}_2']$. On the contrary, $u \not\perp\!\!\!\perp v|w[\mathcal{P}_2]$ and $u \not\perp\!\!\!\perp v|w[\mathcal{P}_2']$. Hence

$$\mathcal{I}(\mathcal{P}_2) \neq \mathcal{I}(\mathcal{P}_2').$$

This example highlights that set-ordering in the structure is important.

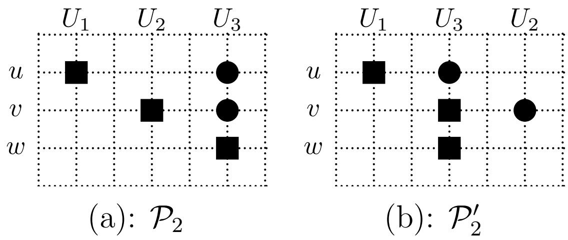


Figure 4.2: Two non-equivalent structures

Recall that each Z-avoiding trail contains one or several vertical connections. However, contrary to the persegram in Figure 4.2b, there is no possible vertical connection between markers corresponding to variables u, v in the persegram in Figure 4.2a. That is why these structures are not equivalent. One of the characteristic properties of equivalence classes is based on this observation.

Chapter 5

Invariants of independence equivalent structures

This chapter deals with some of those attributes of compositional model structures which are invariable with respect to independence equivalence.

5.1 DAG based properties

Now, step by step, we deduce two structural properties necessary for independence equivalence of the respective structures: the *connection set* and the so-called *F-condition set* necessary for description of independence equivalence of the underlying structures. We show later that these properties are sufficient to guarantee the independence equivalence as well.

Note that these properties were inspired by the equivalence problem solution in the theory of Bayesian networks. Recall that the structure of a Bayesian network is represented by a DAG. The reason for this inspiration is very simple: As it is shown in [24], one can transform any perfect compositional model into an equivalent Bayesian network and vice versa. Moreover, in [31] it was shown that any compositional model structure may be converted into an equivalent Bayesian network structure – DAG, i.e., they both induce the same system of (un)conditional independence relations, in other words, the same induced independence model.

To increase the similarity as well as to enlighten this for Bayesian network experts, we decided to use similar notation in this section.

5.1.1 Connection set

While an induced independence relation is highlighted by the fact that there is no respective Z-avoiding trail in the corresponding persegram, each dependence relation is represented by at least one such Z-avoiding trail. Two structures are equivalent if and only if they induce the same dependence models. Thus, in the case of two equivalent structures, one should be able to create the same set of Z-avoiding trails, including the elementary ones that are composed of only two markers – one vertical connection.

It turns out that the set of vertical connections in the respective persegrams is a property invariable with respect to a class of equivalence. Hence, in the following, there is nothing else to show than proper redefinition of a vertical connection in the area of structures.

Definition 5.1. *Consider a structure $\mathcal{P}$ over N and two distinct variables $u, v \in N$. We say that u, v are* connected *in $\mathcal{P}$ ($u \leftrightarrow_{\mathcal{P}} v$) if $u \in K^{\mathcal{P}}_{]v[}$ or $v \in K^{\mathcal{P}}_{]u[}$. Otherwise u, v are not connected in $\mathcal{P}$ ($u \not\leftrightarrow_{\mathcal{P}} v$). The set of all pairs $\mathcal{E}(\mathcal{P}) = \{\{u, v\} : u, v \in N, u \leftrightarrow_{\mathcal{P}} v\}$ is called a* connection set *of $\mathcal{P}$.*

Remark 5.2. *The previous definition basically means that u, v are* connected *in $\mathcal{P}$ iff there is a column in its persegram containing markers of both variables and at least one of them is a box-marker. Hence, connection $u \leftrightarrow_{\mathcal{P}} v$ corresponds to a vertical connection in Definition 3.18.*

The following convention will be used throughout the paper: Given a variable $v \in N$, $U \subseteq N \setminus v$ and a structure $\mathcal{P}$ over N, the term $U \leftrightarrow_{\mathcal{P}} v$ denotes that $u \leftrightarrow_{\mathcal{P}} v$ for every $u \in U$. The symbol $\mathcal{P}$ may be omitted in $u \leftrightarrow_{\mathcal{P}} v$ if the context is clear.

For purposes of the following text, one should realize that when $u \leftrightarrow v$, there is an obvious parallel between ordering of variables u, v and the content of respective columns $K_{]u[}, K_{]v[}$. It is summarized in the following remark.

Remark 5.3. *Let $\mathcal{P}$ be a structure over N and $u, v \in N$ be two distinct variables. By Definition 5.1, $u \leftrightarrow_{\mathcal{P}} v \Leftrightarrow u \in K^{\mathcal{P}}_{]v[} \vee v \in K^{\mathcal{P}}_{]u[}$. If $u \prec_{\mathcal{P}} v$ then $v \notin K^{\mathcal{P}}_{]u[}$. Hence*

$$u \preceq_{\mathcal{P}} v \wedge u \leftrightarrow_{\mathcal{P}} v \Leftrightarrow u \in K^{\mathcal{P}}_{]v[}. \tag{5.1.1}$$

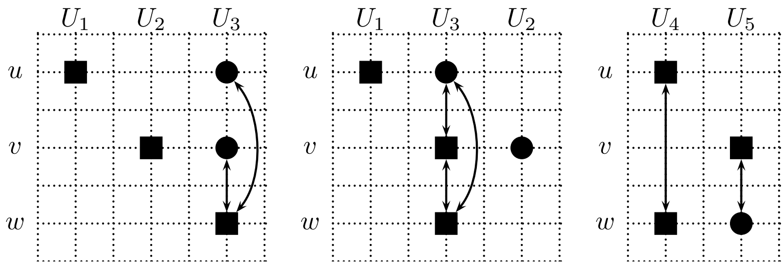
Observe that $u \in S^{\mathcal{P}}_{]v[}$ in (5.1.1) in a strict version of $u \prec_{\mathcal{P}} v$.

Example 5.4. *Consider non-equivalent structures $\mathcal{P}_2$ and $\mathcal{P}'_2$ from Example 4.3 and structure $\mathcal{P}_3$ depicted in Figure 5.1c. Connections as well as connection sets of all considered structures are highlighted in Figure 5.1.*

With the help of Figure 5.1 one can read the following relations:

$\mathcal{P}_2:$	$\{u, v\} \leftrightarrow_{\mathcal{P}_2} w$	$\mathcal{E}(\mathcal{P}_2) = \{\{u, w\}, \{v, w\}\}$
$\mathcal{P}'_2:$	$u \leftrightarrow_{\mathcal{P}'_2} w$, $v \leftrightarrow_{\mathcal{P}'_2} w$, $u \leftrightarrow_{\mathcal{P}'_2} v$	$\mathcal{E}(\mathcal{P}'_2) = \{\{u, w\}, \{v, w\}, \{u, v\}\}$
$\mathcal{P}_3:$	$u \leftrightarrow_{\mathcal{P}_3} w$, $v \leftrightarrow_{\mathcal{P}_3} w$	$\mathcal{E}(\mathcal{P}_3) = \{\{u, w\}, \{v, w\}\} = \mathcal{E}(\mathcal{P}_2)$

As previously stated, the connection $u \leftrightarrow v$ corresponds to the possibility of an existence of a vertical connection between markers corresponding to u and v.



(a): $\mathcal{P}_2$ (b): $\mathcal{P}_2'$ (c): $\mathcal{P}_3$

Figure 5.1: Connections in different structures

Therefore, if there is a connection between two variables, then there is a simple trail connecting the corresponding variables. Since the trail contains no other markers, it is Z-avoiding for any Z such that $Z \subseteq N \setminus \{u, v\}$.

Let us introduce the following specific notation, which allows us to express more than one dependence statement by a single term. Given a structure $\mathcal{P}$ over N, distinct variables $u, v \in N$ and a subset $U \subseteq N \setminus \{u, v\}$, the symbol $u \not\perp\!\!\!\perp v | +U[\mathcal{P}]$ will be interpreted as follows:

$$u \not\perp\!\!\!\perp v | +U[\mathcal{P}] \equiv \forall W \text{ such that } U \subseteq W \subseteq N \setminus \{u, v\} \text{ one has } u \not\perp\!\!\!\perp v | W[\mathcal{P}].$$

In words, u and v are (conditionally) dependent in $\mathcal{P}$ given any superset of U. If U is empty, we write $*$ instead of $+\emptyset$.

$$u \not\perp\!\!\!\perp v | *[\mathcal{P}] \equiv \forall W \text{ such that } W \subseteq N \setminus \{u, v\} \; u \not\perp\!\!\!\perp v | W[\mathcal{P}].$$

Lemma 5.5. *Consider a structure $\mathcal{P}$ over N and two distinct variables $u, v \in N$. Then*

$$u \leftrightarrow_{\mathcal{P}} v \Rightarrow u \not\perp\!\!\!\perp v | *[\mathcal{P}].$$

Proof. Without affecting the generality, suppose $u \preceq_{\mathcal{P}} v$. Then by Remark 5.3, $u \in K_{]v[}$. The sequence of markers $[K_{]v[}, u]$, $[K_{]v[}, v]$ is a W-avoiding trail for any $W \subseteq N \setminus \{u, v\}$. Hence $u \not\perp\!\!\!\perp v | *[\mathcal{P}]$. □

As shown below, one can prove that the *connection set* is one of the properties common to all equivalent structures using this lemma. That is, $\mathcal{E}(\mathcal{P})$ is an invariable property for every class of equivalence in the sense that a connection set of every structure independence that is equivalent with $\mathcal{P}$ coincides with $\mathcal{E}(\mathcal{P})$.

Lemma 5.6. *Let $\mathcal{P}$ be a structure over N. Then for any two distinct variables $u, v \in N$ such that $u \preceq_{\mathcal{P}} v$ it holds that*

$$u \perp\!\!\!\perp v | S^{\mathcal{P}}_{]v[}[\mathcal{P}] \Leftrightarrow u \not\leftrightarrow_{\mathcal{P}} v.$$

Proof. $\Rightarrow$ Suppose $u \perp\!\!\!\perp v | S_{]v[}[\mathcal{P}]$ and $u \leftrightarrow v$. This, however, contradicts Lemma 5.5 based on $u \leftrightarrow v$, which asserts that $u \not\perp\!\!\!\perp v | *[\mathcal{P}]$ and therefore $u \not\perp\!\!\!\perp v | S_{]v[}[\mathcal{P}]$ as well.

$\Leftarrow$ Suppose $u \not\leftrightarrow v$. Using assumption $u \preceq v$, this excludes $u \in K_{]v[}$ by Remark 5.3 and therefore even $u \notin S_{]v[}$. Thus, assume $u \not\perp\!\!\!\perp v | S_{]v[}[\mathcal{P}]$. Since $u \preceq v$, and $S_{]v[} \prec v$, then according to Lemma 3.24 $u \perp\!\!\!\perp v | S_{]v[}[\mathcal{P}[v]]$. By Corollary 3.30, we get $u \perp\!\!\!\perp v | S_{]v[}[\mathcal{P}]$, which contradicts our assumption.

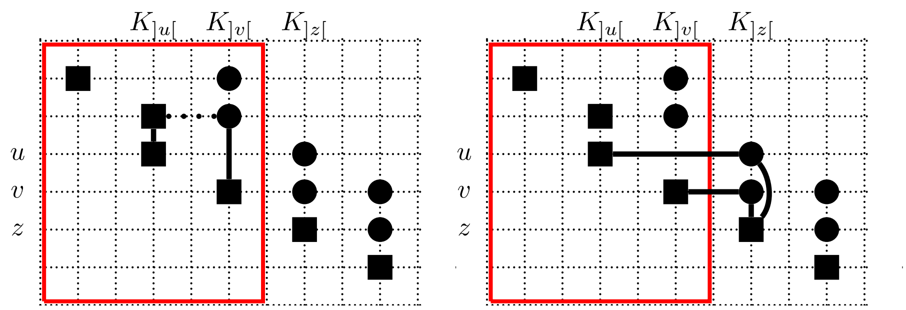
□

Interestingly, under assumption $u \preceq_{\mathcal{P}} v$ notice that while a more general implication of Lemma 5.6

$$u \perp\!\!\!\perp v | +S^{\mathcal{P}}_{]v[} \setminus \{u, v\}[\mathcal{P}] \Rightarrow u \not\leftrightarrow_{\mathcal{P}} v$$

holds, the opposite one does not. Note that the term $+S^{\mathcal{P}}_{]v[} \setminus \{u, v\}$ represents any superset of $S^{\mathcal{P}}_{]v[}$ not containing u or v. One can find a counterexample to the validity of the opposite implication in the following example:

Example 5.7. *Let $\mathcal{P}$ be a structure shown in Figure 5.2. (Recall that $u \prec v$ together with $u \not\leftrightarrow v$ imply that $u \perp\!\!\!\perp v | S_{]v[}[\mathcal{P}]$ by Lemma 5.6). Let us check whether there is an $S_{]v[}$-avoiding trail connecting u with v in Figure 5.2a. Due to Corollary 3.30, we may restrict the searching area to induced substructure $\mathcal{P}[\{u, v\} \cup S_{]v[}] \equiv \mathcal{P}[v]$. The area corresponding to this substructure is highlighted. Since the only sequence of markers regularly alternating horizontal and vertical connections joining u with v contains a horizontal connection corresponding to a variable from $S_{]v[}$, it cannot be an $S_{]v[}$-avoiding trail by Definition 3.18 in $\mathcal{P}[v]$. Thus, $u \perp\!\!\!\perp v | S_{]v[}[\mathcal{P}]$ by Corollary 3.30.*



(a): $u \perp\!\!\!\perp v | S_{]v[}$ (b): $u \not\perp\!\!\!\perp v | +S_{]v[}$

Figure 5.2: A counterexample that Lemma 5.6 cannot be generalized.

One can easily find an example that $u \not\leftrightarrow v \not\Rightarrow u \perp\!\!\!\perp v | +S_{]v[}[\mathcal{P}]$ in Figure 5.2b. It is enough to realize that $z \cup S_{]v[}$ is just a special case of $+S_{]v[}$.

With the help of the previous lemma, one can prove the following important assertion.

Lemma 5.8. *Let $\mathcal{P}$ be a structure over N and $u, v \in N$ be two distinct variables. Then*

$$u \leftrightarrow_{\mathcal{P}} v \Leftrightarrow u \not\perp\!\!\!\perp v | * [\mathcal{P}].$$

Proof. By Lemma 5.5, it is enough to prove the necessity ($\Leftarrow$). Suppose for contradiction that $u \not\perp\!\!\!\perp v | * [\mathcal{P}]$ and $u \not\leftrightarrow v$; one can assume without loss of generality that $u \preceq v$. Then Lemma 5.6 leads to contradiction, since $u \perp\!\!\!\perp v | S_{]v[}[\mathcal{P}]$. $\square$

For a structure, the previous lemma puts an induced connection set on the one hand and a special subset of induced independence relations on the other hand, on the same level. Hence, for two independence equivalent structures, its connection sets coincide. The connection set is an invariable property with respect to a class of independence equivalence:

Corollary 5.9. *Let $\mathcal{P}, \mathcal{P}'$ be two structures over N.*

$$\text{If } \mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}') \text{ then } \mathcal{E}(\mathcal{P}) = \mathcal{E}(\mathcal{P}').$$

Remark 5.10. *A compositional model is basically a multidimensional probability distribution and, as such, it can be represented by a Bayesian network as well. If one uses the conversion algorithm from [24], then the structure of a created Bayesian network $G(N, E)$ – DAG – induces the same independence model as the input compositional model structure $\mathcal{P}$ [31]. Moreover, every* connection $\leftrightarrow_{\mathcal{P}}$ *defined above corresponds precisely to an edge of the DAG corresponding to this conversion algorithm, i.e., $u \leftrightarrow_{\mathcal{P}} v \Leftrightarrow u \to v$ in G or $u \leftarrow v$ in G. This gives us a certain confirmation that our conclusions are correct. Indeed, the set of connections $\mathcal{E}(\mathcal{P})$ (sometimes named as a* skeleton*) is a characteristic property of all DAGs equivalent with G by [56].*

Example 5.11. *The equivalence of different structures was discussed in Example 4.3. While the first two structures $\mathcal{P}_1, \mathcal{P}_1'$ were equivalent, the second two ($\mathcal{P}_2, \mathcal{P}_2'$) were not. Let us look at that example again in light of the previous corollary.*

1. *Let $\mathcal{P}_1, \mathcal{P}_1'$ be two simple structures from Figure 4.1. One can easily see that $\mathcal{E}(\mathcal{P}_1) = \mathcal{E}(\mathcal{P}_1') = \emptyset$. The equality*
$$\mathcal{I}(\mathcal{P}_1) = \mathcal{I}(\mathcal{P}_1') = \{\langle u, v | \emptyset \rangle\}$$
was shown in Example 4.3.

2. *On the other hand, consider structures $\mathcal{P}_2, \mathcal{P}_2'$ depicted in Figure 4.2. Notice that the corresponding connections are highlighted by arrows in Figures 5.1a and 5.1b. Due to Example 4.3, the reader knows that $\mathcal{I}(\mathcal{P}_2) \neq \mathcal{I}(\mathcal{P}_2')$. Since $\mathcal{E}(\mathcal{P}_2') = \mathcal{E}(\mathcal{P}_2) \cup \{\langle u, v \rangle\}$, the reason for non-equivalence is obvious now.*

3. *Consider structure $\mathcal{P}_2$ depicted in Figure 5.1a again. Is there any structure over the same variables not equivalent with $\mathcal{P}_2$ but inducing the same connection set? Indeed, for an example see structure $\mathcal{P}_3$ depicted in Figure 5.1c. Observe that $u \not\perp\!\!\!\perp v | w [\mathcal{P}_2]$ but $u \perp\!\!\!\perp v | w [\mathcal{P}_3]$. Hence, $\mathcal{I}(\mathcal{P}_3) \neq \mathcal{I}(\mathcal{P}_2)$ while $\mathcal{E}(\mathcal{P}_3) = \mathcal{E}(\mathcal{P}_2)$.*

The 3rd part of Example 5.11 illustrates the fact that the same- connection-set condition is necessary but not sufficient to guarantee the equivalence of respective structures. Therefore it is necessary to find an additional property invariant through a class of equivalent structures.

5.1.2 F-condition set

We know that structures $\mathcal{P}_2, \mathcal{P}_3$ from the third part of Example 5.11 are not independence equivalent, despite the fact that $\mathcal{E}(\mathcal{P}_2) = \mathcal{E}(\mathcal{P}_3)$. Considering relation $\preceq_{\mathcal{P}}$, every structure induces a partial ordering of variables. One can easily verify that $u \prec_{\mathcal{P}_2} v \prec_{\mathcal{P}_2} w$ while $u \preceq_{\mathcal{P}_3} w \prec_{\mathcal{P}_3} v$. The induced variable ordering is different for non-equivalent structures. May the ordering of variables be some kind of a characteristic property? Definitely not in this simple way: See Example 4.3 and Figure 4.1, where $\mathcal{I}(\mathcal{P}_1) = \mathcal{I}(\mathcal{P}'_1)$ while $u \prec_{\mathcal{P}_1} v$ and $u \succ_{\mathcal{P}'_1} v$.

It follows that two structures may induce different orderings of variables despite being equivalent. However, if we are interested in the ordering of only groups of specially connected variables, we obtain another property invariable with respect to a class of equivalent structures. This property is based on the so-called *F conditions* defined below.

Definition 5.12. *Consider a structure $\mathcal{P}$ over N and three disjoint variables $u, v, w \in N$. We say that the triplet $\langle u, v|w\rangle$ forms* F-condition *if*

$$\{u, v\} \prec_{\mathcal{P}} w \wedge \{u, v\} \leftrightarrow_{\mathcal{P}} w \wedge u \not\leftrightarrow_{\mathcal{P}} v.$$

The set of triplets $\mathcal{F}(\mathcal{P}) = \{\langle u, v|w\rangle : \{u, v\} \prec_{\mathcal{P}} w \wedge \{u, v\} \leftrightarrow_{\mathcal{P}} w \wedge u \not\leftrightarrow_{\mathcal{P}} v\}$ is called F-condition set *induced by $\mathcal{P}$.*

The reason for calling the above-defined condition "F-condition" is very prosaic. Consider, for example, the structure $\mathcal{P}$ from Figure 5.3. The reader can easily verify that $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$. Observe that the shortest w-avoiding trail connecting box-markers of u and v evokes a mirror image of letter F. An example of F-condition can be found in $\mathcal{P}_2$ depicted in Figure 5.1a, where $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P}_2)$. There is no F-condition in $\mathcal{P}_{2'}$ (Figure 5.1b) and $\mathcal{P}_3$ (Figure 5.1c).

Remark 5.13. *Remark 5.3 states that the fact of $u \prec_{\mathcal{P}} w$ altogether with $u \leftrightarrow_{\mathcal{P}} w$ is equivalent to $u \in S^{\mathcal{P}}_{]w[}$. Regarding this, the previous definition may be reformulated in the following way:* Let $\mathcal{P}$ be a structure over N and $u, v, w \in N$ be three distinct variables. F-condition $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$ is a disjoint triplet of variables such that $u, v \in S^{\mathcal{P}}_{]w[}$ and $u \not\leftrightarrow_{\mathcal{P}} v$.

We have already shown that possessing the same *connection sets* is a necessary condition for equivalence of given structures. Therefore, when comparing two equivalent structures, the *connection set may be considered as fixed.* In addition, we show that the F-condition set is another characteristic property of a class of equivalent structures.

Lemma 5.14. *For structure $\mathcal{P}$ over N and any triplet of distinct variables $u, v, w \in N$ such that $\{u, v\} \leftrightarrow_{\mathcal{P}} w$ and $u \not\leftrightarrow_{\mathcal{P}} v$,*

$$\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P}) \Leftrightarrow u \not\perp\!\!\!\perp v|{+}w[\mathcal{P}].$$

Proof. $\Rightarrow$ Suppose $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$. Then $u, v \in S_{]w[}$ by Remark 5.13. As one can see in Figure 5.3, the sequence of markers $[K_{]u[}, u]$, $[K_{]w[}, u]$, $[K_{]w[}, w]$, $[K_{]w[}, v]$, $[K_{]v[}, v]$ is a W-avoiding trail for every $W \subseteq N \setminus \{u, v\}$ such that $w \in W$. Hence, $u \not\perp\!\!\!\perp v|W[\mathcal{P}]$ for every such W, which can be written as $u \not\perp\!\!\!\perp v|{+}w[\mathcal{P}]$.

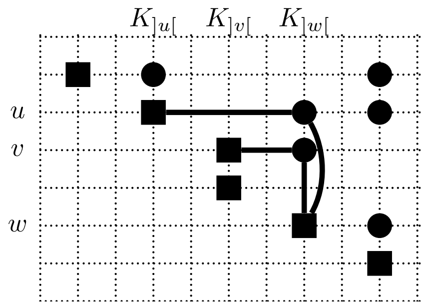


Figure 5.3: $u \not\perp\!\!\!\perp v| + w$

$\Leftarrow$ To prove sufficiency by contradiction, assume that $u \not\perp\!\!\!\perp v|{+}w[\mathcal{P}]$ and not $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$; since the respective connections are fixed by the lemma assumptions, either $u \succeq w$ or $v \succeq w$. Assume without loss of generality $v \succ u$ (equality may be omitted: if $u \not\leftrightarrow_{\mathcal{P}} v$) then $v \succeq w$. If $]v[=]w[$ then $u \prec w$ and $u \in S_{]w[}$ by Remark 5.3 as well as $v \in R_{]v[} \equiv R_{]w[}$. Then $u \leftrightarrow v$, which contradicts with the lemma assumptions. Hence $v \succ w$, which implies that $w \in S_{]v[}$ by assumption of $w \leftrightarrow v$ and Remark 5.3. Using the fact that $u \not\leftrightarrow v$, it follows that $u \perp\!\!\!\perp v|S_{]v[}[\mathcal{P}]$ by Lemma 5.6. This, however, contradicts with $u \not\perp\!\!\!\perp v| + w[\mathcal{P}]$ because $w \in S_{]v[}$.

$\square$

Remark 5.15. *Regarding assumptions of fixed connections $\{u, v\} \leftrightarrow_{\mathcal{P}}$ and $u \not\leftrightarrow_{\mathcal{P}} v$ in the previous lemma: While in the left-to-right implication they are also duplicated by $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$, and therefore they may seem to be useless, they are necessary in the opposite – right-to-left – implication.*

Using Corollary 5.9, one can easily see that

Corollary 5.16. *Let $\mathcal{P}, \mathcal{P}'$ be two structures over N. If $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}')$ then $\mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}')$.*

Remark 5.17. *It was mentioned in Remark 5.10 that there is an algorithm in [24] that enables us to create a DAG G that induces the same independence model as the structure $\mathcal{P}$ – i.e., $\mathcal{I}(\mathcal{P}) = \mathcal{I}(G)$. Moreover, each edge in G corresponds to a connection within $\mathcal{P}$. Note that there is an edge $u \leftrightarrow v$ in G if and only if $u \to v$ or $u \leftarrow v$ in G. Since arrow orientation is given by relation $\prec_{\mathcal{P}}$ (if $u \prec_{\mathcal{P}} v$ and $u \leftrightarrow_{\mathcal{P}} v$ then $u \to v$ in G) in the conversion algorithm, each F-condition defined above implies an* immorality *(vee-triple) in the respective DAG G. Recall that we say that distinct nodes u, v, w form an immorality in a DAG $G = (N, E)$ if $u \to w$ in G, $v \to w$ in G, and $u \not\leftrightarrow_G v$.*

We have derived two properties necessary for independence equivalence of given structures: *same connection, and F-condition sets*. However, are these properties also sufficient to guarantee the equivalence of respective structures? Let us reveal that the answer is positive. However, we will not be able to show it before the end of this chapter.

5.2 Non-trivial sets

One may disclose a possible non-equivalence of given structures with the help of the previously introduced structure properties that are invariable with respect to independence equivalence. A problem arises when the considered structures are more complex. Then the rule resulting from Corollary 5.9 (If two structures were equivalent, then their connection sets would have to coincide.) becomes less easily verifiable. It would be of special importance to have a rule concerning sets defining the structure instead of connections, i.e., sets $\{U_i\}_{i=1\ldots n}$ for structure $U_1, U_2, \ldots, U_n$

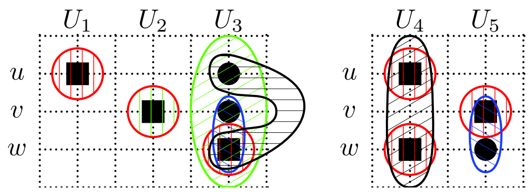
Is there such a rule? To cope with this question, we need the following concept of *non-trivial sets*. While connection $u \leftrightarrow v$ is generally a set of cardinality two, a non-trivial set represents its generalization – a set of an arbitrary cardinality:

Definition 5.18. *Let $\mathcal{P}$ be a structure over N. We say that a non-empty $U \subseteq N$ is* non-trivial in $\mathcal{P}$ *if $U \subseteq K^{\mathcal{P}}_{]U[}$. Otherwise the set U is* trivial *in $\mathcal{P}$. The set of all non-trivial sets in $\mathcal{P}$ is denoted by $\mathcal{N}(\mathcal{P})$.*

Remark 5.19. *Observe a relationship between Definition 5.1 and Definition 5.18. Basically, $\mathcal{N}(\mathcal{P})$ is a generalization of $\mathcal{E}(\mathcal{P})$. Observe that $\mathcal{E}(\mathcal{P}) \subseteq \mathcal{N}(\mathcal{P})$. For $U \in \mathcal{N}(\mathcal{P})$ it holds that $U \in \mathcal{E}(\mathcal{P})$ if and only if $|U| = 2$. However, note that $\mathcal{E}(\mathcal{P})$ does not aim at representing anything in the sense of a "minimal skeleton" for $\mathcal{N}(\mathcal{P})$. As one can see in the following example, there are structures that have the same connection sets but different non-trivial sets, i.e., $\mathcal{E}(\mathcal{P}) = \mathcal{E}(\mathcal{P}') \not\Rightarrow \mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$.*

Example 5.20. *Consider two structures $\mathcal{P}_2$ and $\mathcal{P}_3$ from Example 5.4. In that example, it was shown that both of these structures induce identical connection sets: $\mathcal{E}(\mathcal{P}_2) = \mathcal{E}(\mathcal{P}_3)$. To illustrate the assertion discussed above, $\mathcal{E}(\mathcal{P}) = \mathcal{E}(\mathcal{P}') \nRightarrow \mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$, see Figure 5.4 where structure $\mathcal{P}_2$ and $\mathcal{P}_3$ are again depicted with highlighted non-trivial sets. In the following summary table, one can compare non-trivial sets corresponding to the respective structures split by their cardinalities.*

cardinality	1	2	3
$\mathcal{N}(\mathcal{P}_2)$	u,v,w	{u,w},{v,w}	{u,v,w}
$\mathcal{N}(\mathcal{P}_3)$	u,v,w	{u,w},{v,w}	-



(a): $\mathcal{P}_2$ (b): $\mathcal{P}_3$

Figure 5.4: Non-trivial sets in different structures

For a structure $\mathcal{P}$, $\mathcal{N}(\mathcal{P})$ generally consists of sets with different cardinalities. Considering two independence equivalent structures, it is obvious that their classes of non-trivial sets contain the same singletons as well as sets with cardinality two. Indeed, while it is guaranteed for singletons by the fact that independence equivalent structures are defined over the same set, for sets of cardinality two it is guaranteed by Corollary 5.9 (a set of cardinality two represents a connection). Whether this interesting observation also holds for sets of higher cardinality is a question answered in Corollary 5.26

The following lemma might serve as an alternative definition of the set $\mathcal{N}(\mathcal{P})$. This lemma will later be used in the proof of equivalence of $\mathcal{N}(\mathcal{P})$ with other structure properties important for the equivalence problem solution, specifically in Chapter 7.

Lemma 5.21. *Let $\mathcal{P}$ be a structure over N, $U \subseteq N$, and $I = \{i : \ U \subseteq K_i^{\mathcal{P}}\}$.*

$$U \notin \mathcal{N}(\mathcal{P}) \Leftrightarrow (I = \emptyset) \vee (\forall i \in I \text{ it holds that } U \subseteq S_i^{\mathcal{P}})$$

Proof. Suppose $U \notin \mathcal{N}(\mathcal{P})$ – i.e., $U \nsubseteq K^{\mathcal{P}}_{]U[}$; to prove the necessity, suppose by contradiction that $I \neq \emptyset$ and $\exists i \in I$ such that $U \nsubseteq S_i^{\mathcal{P}}$. Since $U \subseteq K_i^{\mathcal{P}}$ by definition of I then $U \cap R_i^{\mathcal{P}} \neq \emptyset$. Hence $]U[_{\mathcal{P}} = i$ which leads to contradiction. The converse implication follows directly from Definition 5.18. $\square$

Remark 5.22. *Observe that if $U \subseteq S^{\mathcal{P}}_{]v[}$ then $U \cup \{v\}$ is non-trivial in $\mathcal{P}$. Indeed, note that the relation $U \subseteq S^{\mathcal{P}}_{]v[}$ implies that $U \prec_{\mathcal{P}} v$ and therefore $]v[_{\mathcal{P}}=]U \cup v[_{\mathcal{P}}$. Since $v \in K^{\mathcal{P}}_{]v[}$ by definition of $]\cdot[$,$U \cup v \subseteq K^{\mathcal{P}}_{]v[} \equiv K^{\mathcal{P}}_{]U\cup v[}$, which corresponds to the definition of non-triviality, Definition 5.18.*

Example 5.23. *Consider structure $\mathcal{P}$ in Figure 5.5. Observe that while $U_i \in \mathcal{P}$ as well as $U_i \in \mathcal{N}(\mathcal{P})$ for all $i = 1, \ldots, 5$, the sets $\{u, x\}$ and $\{v, x\}$ are necessarily non-trivial ($\{u, x\} \in \mathcal{N}(\mathcal{P})$ but $\{u, x\} \notin \mathcal{P}$ etc.).*

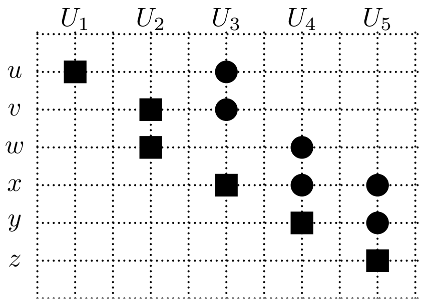


Figure 5.5: Structure $\mathcal{P}$

The following lemma states the importance of sets of mutually connected variables. Notice that for sets with cardinality two it coincides with Remark 5.3. Hence, one could say that the following lemma is a generalization of Remark 5.3 in the same way as a non-trivial set (Definition 5.18) represents a generalization of a connection defined in Definition 5.1.

Lemma 5.24. *Let U be a non-empty set of mutually connected variables in $\mathcal{P}$ ($u \leftrightarrow_{\mathcal{P}} u'$ for all disjoint $u, u' \in U$). Then U is non-trivial in $\mathcal{P}$, i.e., $U \in \mathcal{N}(\mathcal{P})$.*

Proof. Choose $u \in U$ such that $u \succeq_{\mathcal{P}} u'$ for all other $u' \in U$. This choice is always possible and ensures that $]U[_{\mathcal{P}}=]u[_{\mathcal{P}}$ by definition of $]\cdot[_{\mathcal{P}}$. Moreover $U \subseteq K^{\mathcal{P}}_{]u[}$. Indeed, realize that $u \leftrightarrow_{\mathcal{P}} u'$ by assumption. Hence $u' \in K^{\mathcal{P}}_{]u[}$ by Remark 5.3 for all $u' \in U \setminus u$, which implies that $U \subseteq K^{\mathcal{P}}_{]U[}$. Hence $U \in \mathcal{N}(\mathcal{P})$ by Definition 5.18. □

With the help of the previous lemma, one can prove the following interesting assertion relating non-trivial sets to other structure properties ($\mathcal{E}()$ and $\mathcal{F}()$) that are invariable with respect to equivalence classes. We show that $\mathcal{N}(\mathcal{P})$ is just another invariable property of equivalent structures.

Lemma 5.25. *For any two structures $\mathcal{P}, \mathcal{P}'$ over N it holds that*

$$\text{if } \mathcal{E}(\mathcal{P}) = \mathcal{E}(\mathcal{P}') \text{ and } \mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}') \text{ then } \mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}').$$

Proof. Since the connection sets of both $\mathcal{P}, \mathcal{P}'$ are identical, we take the connection set as fixed and we do not use the structure as the index in $\leftrightarrow_{\mathcal{P}}, \leftrightarrow_{\mathcal{P}'}$ to distinguish between the corresponding structures. In addition, the roles of $\mathcal{P}$ and $\mathcal{P}'$ are interchangeable and therefore it is enough to prove that $\mathcal{N}(\mathcal{P}) \subseteq \mathcal{N}(\mathcal{P}')$.

Consider an arbitrary $U \neq \emptyset$ such that $U \in \mathcal{N}(\mathcal{P})$ – i.e., $U \subseteq K^{\mathcal{P}}_{]U[}$ by the definition of a non-trivial set. Choose $u \in U$ such that $u \succeq_{\mathcal{P}} u'$ for all other $u' \in U$. This choice is always possible ($U \neq \emptyset$) and ensures that $]U[_{\mathcal{P}}=]u[_{\mathcal{P}}$. Then $U \subseteq K^{\mathcal{P}}_{]u[}$, which together with Remark 5.3, implies that

$$u \leftrightarrow U \setminus u. \tag{5.2.1}$$

Let $M \subseteq U$ be the maximal subset of mutually connected variables in $\mathcal{P}$ such that both $R^{\mathcal{P}}_{]U[} \cap U \subseteq M$ and $M \leftrightarrow u'$ for all $u' \in U \setminus M$. Then put $V = U \setminus M$. Observe that not only $M \neq \emptyset$ ($u \in M$ by (5.2.1)) but also $V \prec_{\mathcal{P}} u$. Indeed, assume for contradiction that $\exists v \in V$ such that $v \succeq_{\mathcal{P}} u$. Then $]v[_{\mathcal{P}}=]u[_{\mathcal{P}}=]U[_{\mathcal{P}}$ by definition of u and the fact that $V \subset U$. Then $v \in R^{\mathcal{P}}_{]U[}$ which contradicts the choice of M and the fact that $v \notin M$.

Since M is a set of mutually connected variables in $\mathcal{P}'$ by $\mathcal{E}(\mathcal{P}) = \mathcal{E}(\mathcal{P}')$, M is non-trivial in $\mathcal{P}'$ by Lemma 5.24 and $M \subseteq K^{\mathcal{P}'}_{]M[}$. One can distinguish two cases: $V = \emptyset$ and $V \neq \emptyset$.

If $V = \emptyset$, then $U \equiv M$ and the lemma is proven. Suppose that $V \neq \emptyset$. The next step is to prove that $V \subset K^{\mathcal{P}'}_{]M[}$ as well. Assume for contradiction that $\exists v \in V$ such that $v \notin K^{\mathcal{P}'}_{]M[}$. Then there exists $v' \in V \setminus v$ such that $v' \not\leftrightarrow v$ (otherwise $v \in M$). Considering the fact $V \prec_{\mathcal{P}} u$ and $\{v, v'\} \leftrightarrow u$ by (5.2.1), we get $\langle v, v'|u\rangle \in \mathcal{F}(\mathcal{P})$ and $\langle v, v'|u\rangle \in \mathcal{F}(\mathcal{P}')$ by $\mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}')$. The fact that $u \in M \subseteq K^{\mathcal{P}'}_{]M[}$ implies that $u \preceq_{\mathcal{P}'} M$, which together with $v \prec_{\mathcal{P}'} u$ (because of $\langle v, v'|u\rangle \in \mathcal{F}(\mathcal{P}')$) can be composed into $v \prec_{\mathcal{P}'} M$. Moreover, $v \leftrightarrow M$ by definition of M, and therefore $v \in K^{\mathcal{P}'}_{]M[}$ by Remark 5.3, which contradicts the assumption.

Hence, $V \cup M \subseteq K^{\mathcal{P}'}_{]M[}$ necessarily and since $]U[_{\mathcal{P}'} = max(]V[_{\mathcal{P}'},]M[_{\mathcal{P}'}) =]M[_{\mathcal{P}'}$, $U = V \cup M \subseteq K^{\mathcal{P}'}_{]M[} \equiv K^{\mathcal{P}'}_{]U[}$ and $U \in \mathcal{N}(\mathcal{P}')$ by Definition 5.18. □

With respect to the previous lemma and Corollaries 5.9, and 5.16 (i.e., $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}') \Rightarrow \mathcal{E}(\mathcal{P}) = \mathcal{E}(\mathcal{P}') \wedge \mathcal{F}(\mathcal{P}) = \mathcal{F}(\mathcal{P}')$) we can easily declare that the set $\mathcal{N}(\mathcal{P})$ is another invariable property of independence equivalent structures:

Corollary 5.26. *If two structures over N are independence equivalent then they induce the same set of non-trivial sets. ($\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}') \Rightarrow \mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$)*

Remark 5.27. *It became apparent that there is a close connection between the system of non-trivial sets and* characteristic imsets *introduced by Studený, Hemmecke, and Lindner [54] as a unique algebraic representative of a Bayesian network structure. It is a 0-1 vector indexed by subsets of the set of variables. It* *appears that in the case of Bayesian network structure – DAG – inducing the same independence model as a given compositional model structure, the characteristic imset of the DAG takes value 1 on components corresponding to non-trivial sets only in the respective compositional model structure.*

Note that the basic idea for introducing such an algebraic representative lies in possibility of using classical linear programming methods for learning the Bayesian network structure. For example, we refer to [54] for solution in the case of undirected forests.

5.3 Column approach

Just as the Connection set $\mathcal{E}()$ and F-conditions $\mathcal{F}()$, neither nontrivial sets $\mathcal{N}()$ represent an appropriate tool for the decision of the non-equivalence of two compositional model structures. For this reason we continued in the study of equivalence classes and have discovered other invariants of equivalent structures based on columns.

The following idea is very simple: if we take into account the maximal non-trivial sets in the sense of inclusion we can say that, if two structures induce the same non-trivial set, then they naturally induce the same maximal non-trivial sets as well. Moreover, as shown below, every maximal non-trivial set corresponds to a column. This is also why we call this approach a *column approach.*

Definition 5.28. *Let $\mathcal{P} = U_1, \ldots, U_n$ be a structure. A set U_i is a* non-trivial column *of $\mathcal{P}$ if $U_i = K_{]U_i[}$. Otherwise it is a* trivial column *of $\mathcal{P}$. The symbol $ntriv(\mathcal{P})$ denotes the set of all non-trivial columns in $\mathcal{P}$.*

Remark 5.29. *Observe that $ntriv(\mathcal{P}) = \{K_i^{\mathcal{P}} \in \mathcal{P} : R_i^{\mathcal{P}} \neq \emptyset\}$.*

Remark 5.30. *Recall that $R_1^{\mathcal{P}}, \ldots, R_{|\mathcal{P}|}^{\mathcal{P}}$ is a disjoint partition of $K_1^{\mathcal{P}} \cup \ldots \cup K_{|\mathcal{P}|}^{\mathcal{P}} = N$ for every structure $\mathcal{P}$ over N. Since $R_i^{\mathcal{P}} \neq \emptyset$ for nontrivial column $K_i^{\mathcal{P}}$, it is obvious that $|ntriv(\mathcal{P})| \leq |N|$.*

Observe that there is a close relationship between non-trivial columns and non-trivial sets of variables. In fact, an arbitrary non-trivial column of a structure $\mathcal{P}$ is a non-trivial set in $\mathcal{P}$ as well, i.e., $ntriv(\mathcal{P}) \subseteq \mathcal{N}(\mathcal{P})$. Conversely, if a set U is non-trivial in $\mathcal{P}$ then U is a subset of a non-trivial column from $\mathcal{P}$.

Lemma 5.31. *Having fixed structure $\mathcal{P}$, the maximal non-trivial sets (with respect to inclusion) in $\mathcal{P}$ coincide with maximal sets in $ntriv(\mathcal{P})$ (with respect to inclusion), that is, maximal columns with at least one box-marker.*

Proof. To prove this assertion it is enough to realize that every non-trivial column $U_i \in ntriv(\mathcal{P})$ represents a non-trivial set of variables $U = K_{]U[}^{\mathcal{P}}$ at the same time. Similarly, the existence of a non-trivial set U implies that $U \subseteq K_{]U[}$ by Definition 5.18. □

5.3.1 Strong core

By Corollary 5.26 and Lemma 5.25 non-trivial sets are the same for a class of equivalent structures. Hence maximal non-trivial sets, that is, maximal non-trivial columns (by Lemma 5.31) in them, coincide as well. In other words, Lemma 5.25 induces another characteristic property of a class of equivalent structures. This property is known as the *strong core* of a structure.

Definition 5.32. *For a structure $\mathcal{P}$ over N, its* strong core $\mathcal{C}^*(\mathcal{P})$ *is the set of its maximal non-trivial columns with respect to inclusion.* ($\mathcal{C}^*(\mathcal{P}) = \{U \in ntriv(\mathcal{P}) : \nexists V \in ntriv(\mathcal{P})$ *such that* $U \subset V\}$ *where* $U, V \subseteq N$.)

Example 5.33. *Consider structure $\mathcal{P} = U_1, U_2, \ldots, U_7$ in Figure 5.6. Note that $\mathcal{C}^*(\mathcal{P}) = \{U_4, U_5, U_7\}$. Indeed, observe that $U_1 \subset U_4, U_2 \subset U_4, U_3 \subset U_4$ where $U_4 \in ntriv(\mathcal{P})$. Since U_6 is a trivial column in $\mathcal{P}$, the inclusion $U_5 \subset U_6$ has no impact.*

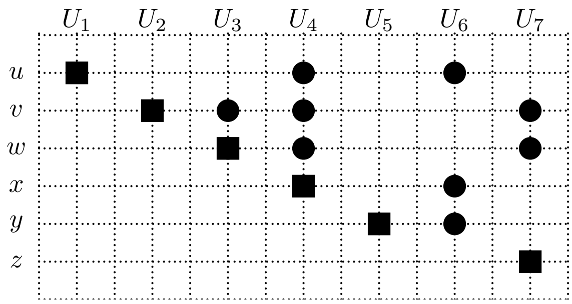


Figure 5.6: Structure $\mathcal{P}$

Recall that $\mathcal{C}^*(\mathcal{P}) \subseteq \mathcal{N}(\mathcal{P})$. Since non-trivial sets of two equivalent structures coincide by Corollary 5.26, the maximal non-trivial sets coincide as well. Thus, maximal non-trivial columns (by Lemma 5.31) coincide for independence equivalent structures.

Corollary 5.34. *Let $\mathcal{P}$ be a structure over N. Then $\mathcal{C}^*(\mathcal{P}) = \mathcal{C}^*(\mathcal{P}')$ for every equivalent structure $\mathcal{P}'$.*

Example 5.35. *Consider three structures $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$ over the same variable set $\{u, v, w, x\}$ shown in Figure 5.7. Observe that these structures are composed from eight different sets – columns: $U_1, \ldots U_8$. Notice that the* strong core *is the same for all three structures: $\mathcal{C}^*(\mathcal{P}_1) = \mathcal{C}^*(\mathcal{P}_2) = \mathcal{C}^*(\mathcal{P}_3) = \{U_3, U_4\} = \{\{u, v, w\}, \{v, w, x\}\}$. Thus, considering the necessary condition from Corollary 5.34 only, these structures could be equivalent. But they are not, as we shall see in Example 5.38.*

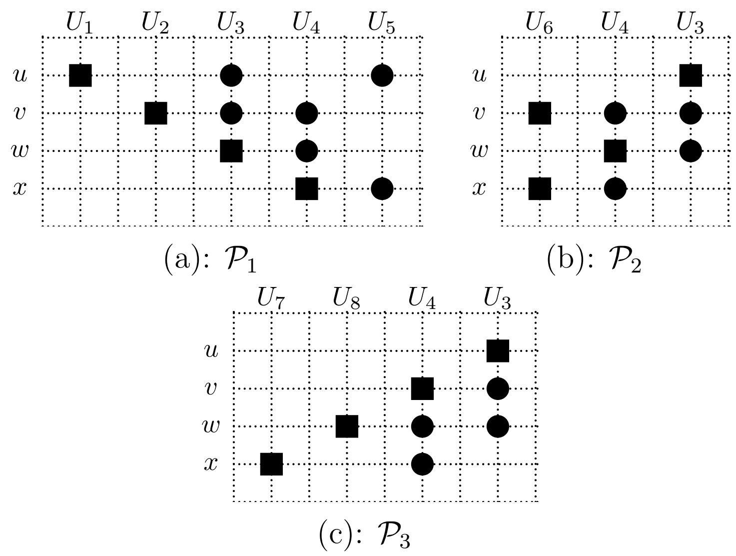


Figure 5.7: Structures with the same strong core

Remark 5.36. *Unlike the invariants previously discussed in Remarks 5.10 and 5.17, this characteristic property does not correspond to any standard characteristics of equivalent DAGs. Still, given the definition,* strong structure core *could correspond to a set of maximal families induced by the corresponding* $G = (N, E)$. *Note that family* $fam(u)$ *means the set* $u \cup pa(u)$ *where* $pa(u) = \{v \in N : (v \to u) \in E\}$.

Recall we can generate all DAGs which are equivalent to a given one with the help of the so-called legal arrow reversal. *By a* legal arrow reversal *we understand the change of DAG* G *into DAG* G' *by replacement of an arrow* $u \to v$ *(in* G*) with* $u \leftarrow v$ *(in* G'*) under the condition that* $pa_G(u) \cup u = pa_G(v)$. *If* $fam_G(v) = V = pa_G(u) \cup \{u, v\}$ *belongs to maximal families (with respect to inclusion), then it belongs to maximal families in* G' *as well. Indeed, since* $pa_{G'}(u) = pa_G(u) \cup v$ *then* $fam_{G'}(u) = V$. *Since there are no other arrow changes, all other families remain the same and* V *belongs to maximal families in* G' *and in all other equivalent DAGs.*

5.3.2 Weak core

Consider two equivalent structures, one of which has a non-trivial column U which does not exist in the other. What can be said about it? Not much so far. According to Corollary 5.34, there has to be another non-trivial column V such that $U \subset V$. Does this *superset column* V have any special relationship to U? The reader can immediately see that $]U[<]V[$. Indeed, U would otherwise be trivial. Another very interesting relationship may be discovered by multiple applications of Lemma 5.25. It appears that it is useful to extend *strong structure* *core* to obtain the so-called *weak structure core* – a more complex set of columns that are common to all structures equivalent with the given one.

Definition 5.37. *For a structure $\mathcal{P}$, its* weak core *$\mathcal{C}(\mathcal{P})$ is the set of non-trivial columns that are not equivalent to the S-part of any other non-trivial column. ($\mathcal{C}(\mathcal{P}) = \{K_i^{\mathcal{P}} \in ntriv(\mathcal{P}) : K_i^{\mathcal{P}} \neq S_j^{\mathcal{P}}$ for all other $K_j^{\mathcal{P}} \in ntriv(\mathcal{P})\}$.)*

Note that, for any arbitrary structure, the extended *weak core* includes not only the already defined *strong core* but also all non-trivial columns that are not, let us say, *S-subsets* of another non-trivial column. (By the term S-subset we have in mind the situation when $K_i^{\mathcal{P}} = S_j^{\mathcal{P}}$. Here $K_i^{\mathcal{P}}$ is an S-subset of $K_j^{\mathcal{P}}$ in $\mathcal{P}$.)

Example 5.38. *Consider structure $\mathcal{P}$ from Figure 5.6 once more. One can see that $\mathcal{C}(\mathcal{P}) = \{U_1, U_4, U_5, U_7\}$. Observe that $\mathcal{C}^*(\mathcal{P}) \subset \mathcal{C}(\mathcal{P})$ in this structure.*

Observe that a *weak core* contains almost all non-trivial columns of the structure. In Example 5.38, $\mathcal{C}(\mathcal{P})$ contains all non-trivial columns except U_2, U_3. Now we will show that a *weak core* is another invariant property of a class of equivalent structures. As a result, we have a very simple and powerful tool in our hands. Especially in comparison with Corollaries 5.9, 5.16, and 5.34.

First, we have to prove three essential auxiliary assertions. They enable us to prove the desired invariance of the weak core with respect to a class of equivalent structures.

Lemma 5.39. *Consider a structure $\mathcal{P}$ over N, $U \subset N$, and $v \in N \setminus U$ such that $(U \cup v) \in \mathcal{N}(\mathcal{P})$. If $v \notin K_{]U[}^{\mathcal{P}}$ then $U \subseteq S_{]v[}^{\mathcal{P}}$.*

Proof. First prove that $U \prec v$. Assume for contradiction that $U \succeq v$ while $v \notin K_{]U[}$. Then $]U \cup v[=]U[$ and $(U \cup v) \subseteq K_{]U[}$ by non-triviality of $(U \cup v)$ in $\mathcal{P}$. This, however, contradicts the assumption of $v \notin K_{]U[}$. Hence, $U \prec v$.

It follows that $]U \cup v[=]v[$ and $U \subset K_{]v[}$ by non-triviality of $(U \cup v)$ in $\mathcal{P}$. Combining it with relation $U \prec v$, one gets that $U \subseteq S_{]v[}^{\mathcal{P}}$. □

Corollary 5.40. *If, for a given structure $\mathcal{P}$, it holds that $\{u, v, w\} \in \mathcal{N}(\mathcal{P})$ and $\{u, v\} \prec_{\mathcal{P}} w$, then $\{u, v\} \subseteq S_{]w[}^{\mathcal{P}}$.*

The following lemma is purely technical. It covers a very interesting part of the reasoning process that was consistent in several previous proofs. Therefore, for the sake of clarity, it is stated separately and we refer to it in respective proofs.

Note that the most important part is this: *If a non-trivial set U is a subset of S-part of a non-trivial column then, under certain circumstances, there is another subsequent non-trivial column that contains U in its S-part as well.*

Lemma 5.41. *Consider two structures $\mathcal{P}, \mathcal{P}'$ over N such that $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$, $U \subset N$, $v \in N \setminus U$, $U \in \mathcal{N}(\mathcal{P})$, $]U[_{\mathcal{P}} <]v[_{\mathcal{P}}$, and $K^{\mathcal{P}}_{]U[} \cap K^{\mathcal{P}}_{i} \subseteq U$ for all $]U[_{\mathcal{P}} < i \leq]v[_{\mathcal{P}}$. If*

$$U \subset S^{\mathcal{P}}_{]v[} \quad \textit{and} \quad S^{\mathcal{P}}_{]v[} \cap K^{\mathcal{P}'}_{]U \cup v[} \subseteq U,$$

then $\exists v' \in N :\ U \prec_{\mathcal{P}} v' \prec_{\mathcal{P}} v$ such that

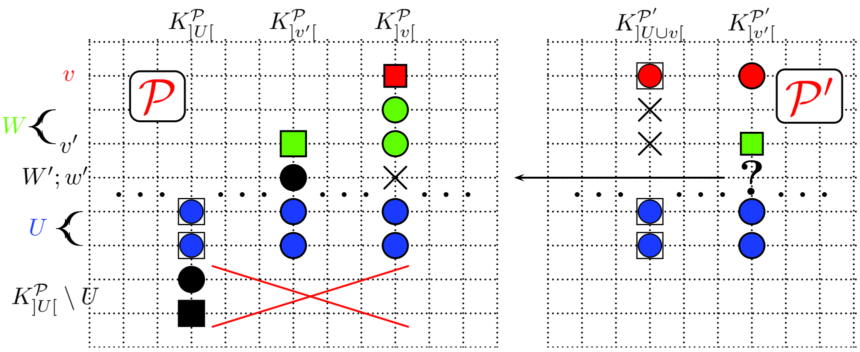
$$U \subseteq S^{\mathcal{P}}_{]v'[} \quad \textit{and} \quad S^{\mathcal{P}}_{]v'[} \cap K^{\mathcal{P}'}_{]U \cup v'[} \subseteq U.$$

Proof. Put $W = S^{\mathcal{P}}_{]v[} \setminus U$. Note that $W \neq \emptyset$ by assumption. Choose and fix $v' \in W$ such that $v' \preceq_{\mathcal{P}} w$ for all other $w \in W$. Realize that

$$v' \notin K^{\mathcal{P}}_{]U[} \tag{5.3.1}$$

by the fact that $v' \notin U$ by its definition and assumption of $K^{\mathcal{P}}_{]U[} \cap K^{\mathcal{P}}_{]v[} \subseteq U$.

Since $U \subset S^{\mathcal{P}}_{]v[}$ by assumption and $v' \in S^{\mathcal{P}}_{]v[}$ by its definition, $(U \cup \{v, v'\}) \in \mathcal{N}(\mathcal{P})$ by Remark 5.22. (See Figure 5.8 for illustration where, however, columns out of focus are omitted for the sake of lucidity. Note that by a set of bullets in a box in one column we denote the situation in which we are not sure about the markers' shapes, but at least one of them is a box-marker. A situation in which there is definitely no marker in a certain position is marked as a cross at that position.)



(a): part of $\mathcal{P}$ (b): part of $\mathcal{P}'$

Figure 5.8: Proof of Lemma 5.41

Since we assume that $v' \notin K^{\mathcal{P}'}_{]U \cup v[}$ (see the assumption $S^{\mathcal{P}}_{]v[} \cap K^{\mathcal{P}'}_{]U \cup v[} \subseteq U$),

$$U \cup v \subseteq S^{\mathcal{P}'}_{]v'[} \text{ and } U \prec_{\mathcal{P}'} v' \tag{5.3.2}$$

by Lemma 5.39. That is why $U \cup v' \in \mathcal{N}(\mathcal{P}')$ by Remark 5.22. Then, however,

$$U \subseteq S^{\mathcal{P}}_{]v'[} \tag{5.3.3}$$

by Lemma 5.39, (5.3.1) and the fact that $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$. It also implies that

$$U \prec_{\mathcal{P}} v' \prec_{\mathcal{P}} v. \tag{5.3.4}$$

Put $W' = S^{\mathcal{P}}_{]v'[} \setminus U$. Then the choice of v' guarantees that $W \cap W' = \emptyset$. Let us prove that $S^{\mathcal{P}}_{]v'[} \cap K^{\mathcal{P}'}_{]U\cup v'[} \subseteq U$. Assume for contradiction that for at least one $w' \in W'$ it holds that $w' \in K^{\mathcal{P}'}_{]U\cup v'[}$. Observe that $K^{\mathcal{P}'}_{]U\cup v'[} \equiv K^{\mathcal{P}'}_{]v'[}$ by (5.3.2) and moreover, $U \cup v \subseteq S^{\mathcal{P}'}_{]v'[}$. Then $\{v, v', w'\}$ is a non-trivial set in $\mathcal{P}'$. Since $w' \prec_{\mathcal{P}} v' \prec_{\mathcal{P}} v$ by definition of W' and W then $\{v', w'\} \subseteq S^{\mathcal{P}}_{]v[}$ by Lemma 5.40. Hence, $w' \in W$ which contradicts with $W \cap W' = \emptyset$. Therefore, $S^{\mathcal{P}}_{]v'[} \cap K^{\mathcal{P}'}_{]U\cup v'[} \subseteq U$. This, together with (5.3.3) and (5.3.4), proves the lemma.

□

If we properly investigate the previous lemma, we find out that it might be applied recursively on its own. Since, moreover, $U \prec_{\mathcal{P}} v' \prec_{\mathcal{P}} v$ and the number of variables is finite, it is clear that there must exist v'' such that $U = S^{\mathcal{P}}_{]v''[}$:

Corollary 5.42. *Consider two structures $\mathcal{P}, \mathcal{P}'$ over N such that $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$. Let $U \in \mathcal{N}(\mathcal{P})$ and $v \in N \setminus U$ be such that $K^{\mathcal{P}}_{]U[} \cap K^{\mathcal{P}}_i \subseteq U$ for all $i \in \mathbb{N}$ for which $]U[_{\mathcal{P}} < i \leq]v[_{\mathcal{P}}$. If*

$$U \subset S^{\mathcal{P}}_{]v[} \quad \textit{and} \quad S^{\mathcal{P}}_{]v[} \cap K^{\mathcal{P}'}_{]U\cup v[} \subseteq U$$

then $\exists w \in N$ such that $U = S^{\mathcal{P}}_{]w[}$.

In a case where U is a non-trivial column, this fact has a very interesting consequence concerning weak core $\mathcal{C}(\mathcal{P})$, stated in the following lemma.

Lemma 5.43. *Consider two structures $\mathcal{P}, \mathcal{P}'$ over N with $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$, and $U \in ntriv(\mathcal{P})$ such that $U \subset S^{\mathcal{P}}_{]v[}$ for some $v \in N \setminus U$. If $S^{\mathcal{P}}_{]v[} \cap K^{\mathcal{P}'}_{]U\cup v[} \subseteq U$ then $U \notin \mathcal{C}(\mathcal{P})$.*

Proof. Since $U \in ntriv(\mathcal{P})$, $U = K^{\mathcal{P}}_{]U[}$ and $K^{\mathcal{P}}_{]U[} \cap K^{\mathcal{P}}_i \subseteq U$ for all $i = 1, \ldots, |\mathcal{P}|$ and therefore even for $]U[_{\mathcal{P}} < i \leq]v[_{\mathcal{P}}$. Using Corollary 5.42 we get $w \in N$ such that $U = S^{\mathcal{P}}_{]w[}$. Then, however, $U \notin \mathcal{C}(\mathcal{P})$ by Definition 5.37. □

Now we can prove the professed claim that the weak core is another invariant property over a class of equivalent structures.

Theorem 5.44. *For two structures $\mathcal{P}, \mathcal{P}'$ such that $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$ it holds that $\mathcal{C}(\mathcal{P}) = \mathcal{C}(\mathcal{P}')$.*

Proof. Since the roles of $\mathcal{P}, \mathcal{P}'$ are interchangeable, it suffices to verify $\mathcal{C}(\mathcal{P}) \subseteq \mathcal{C}(\mathcal{P}')$. Suppose for contradiction that $\exists U \in ntriv(\mathcal{P})$ such that $U \in \mathcal{C}(\mathcal{P})$ and $U \notin \mathcal{C}(\mathcal{P}')$. We can distinguish two possible cases of $U \notin \mathcal{C}(\mathcal{P}')$:

1) $U \in ntriv(\mathcal{P}')$

2) $U \notin ntriv(\mathcal{P}')$

First, we will show that in both cases there exists a column $K_i^{\mathcal{P}'}$ such that $U \subset K_i^{\mathcal{P}'}$ and, moreover, for all $v \in K_i^{\mathcal{P}'} \setminus U$ it holds that $U \cup v$ is a non-trivial set in both $\mathcal{P}, \mathcal{P}'$.

1) Assume that $U \in ntriv(\mathcal{P}')$. Since $U \notin \mathcal{C}(\mathcal{P}')$, $\exists i \in \mathbb{N}$, such that $U = S_i^{\mathcal{P}'}$ where $K_i^{\mathcal{P}'} \in ntriv(\mathcal{P}')$. Put $V = K_i^{\mathcal{P}'} \setminus U \equiv R_i^{\mathcal{P}'}$ (note that $V \neq \emptyset$) and observe that $U \cup v \in \mathcal{N}(\mathcal{P}') = \mathcal{N}(\mathcal{P})$ for all $v \in V$ by Remark 5.22.

2) Similarly, assume $U \notin ntriv(\mathcal{P}')$. Nevertheless, U is a non-trivial set in $\mathcal{P}'$ by $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$ and the fact that $U \in \mathcal{C}(\mathcal{P})$. Therefore there exists a column such that $U \subset K_{]U[}^{\mathcal{P}'}$. Put $V = K_{]U[}^{\mathcal{P}'} \setminus U$. Now $U \cup \{v\}$ is non-trivial for all $v \in V$. Let us point out that $i =]U[_{\mathcal{P}'}$ in this case.

Observe that
$$i =]U \cup v[_{\mathcal{P}'} \tag{5.3.5}$$
holds for all $v \in V$ in both cases $U \in ntriv(\mathcal{P}')$ and $U \notin ntriv(\mathcal{P}')$.

Since U, V are disjoint and $\mathcal{N}(\mathcal{P}') = \mathcal{N}(\mathcal{P})$, $U \subseteq S_{]v[}^{\mathcal{P}}$ for all $v \in V$ by Lemma 5.39, where the inclusion may be made strict using the fact that $U \in \mathcal{C}(\mathcal{P})$. Hence
$$U \subset S_{]v[}^{\mathcal{P}}. \tag{5.3.6}$$
for all $v \in V$. Indeed, otherwise $U = S_{]v[}^{\mathcal{P}}$ for some $v \in V$ and $U \notin \mathcal{C}(\mathcal{P})$ by definition of weak core. Choose and fix $v \in V$ and the corresponding $K_{]v[}^{\mathcal{P}}$ such that $v \preceq_{\mathcal{P}} v'$ for all other $v' \in V$. This choice is always possible and ensures that $S_{]v[}^{\mathcal{P}} \cap V = \emptyset$. Given that $K_i^{\mathcal{P}'} = U \cup V$, it implies that

$$S_{]v[}^{\mathcal{P}} \cap K_i^{\mathcal{P}'} \subseteq U. \tag{5.3.7}$$

Using (5.3.5), expression (5.3.7) may be rewritten into

$$S_{]v[}^{\mathcal{P}} \cap K_{]U \cup v[}^{\mathcal{P}'} \subseteq U. \tag{5.3.8}$$

Consider our assumption of $U \in ntriv(\mathcal{P})$, (5.3.6), and (5.3.8). Then Lemma 5.43 may be applied and one gets $U \notin \mathcal{C}(\mathcal{P})$, which contradicts with the assumption. □

By combining both Corollary 5.26 and the previous lemma, we easily come to the conclusion that the *weak core* is another invariant property of a class of equivalent compositional model structures:

Corollary 5.45. *Let $\mathcal{P}, \mathcal{P}'$ be two structures over N. If $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}')$ then $\mathcal{C}(\mathcal{P}) = \mathcal{C}(\mathcal{P}')$.*

Example 5.46. *Consider structures* $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$ *from Figure 5.7 again. By Example 5.35, the* strong cores *of the respective structures coincide.* $(\mathcal{C}^*(\mathcal{P}_1) = \mathcal{C}^*(\mathcal{P}_2) = \mathcal{C}^*(\mathcal{P}_3) = \{U_3, U_4\})$*. Thus, considering only Corollary 5.34, these structures could be equivalent.*

Conversely, weak cores of these structures are completely different:

- $\mathcal{C}(\mathcal{P}_1)$*:* $\{U_1, U_2, U_3, U_4\}$
- $\mathcal{C}(\mathcal{P}_2)$*:* $\{U_3, U_4\}$
- $\mathcal{C}(\mathcal{P}_3)$*:* $\{U_3, U_4, U_7, U_8\}$

Therefore these structures cannot be equivalent by Corollary 5.45.

The conclusion of the previous Example 5.46 could also be obtained using another (previously discovered) property invariable for equivalent structures, such as the F-condition set. However, we believe that this solution is more elegant and easier to use.

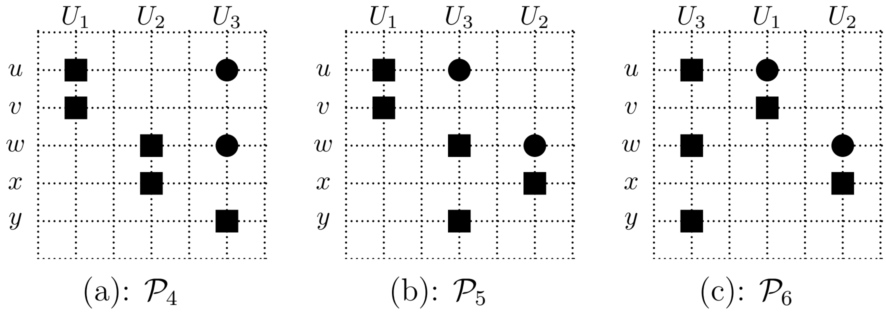


Figure 5.9: Non-equivalent structures with the same weak core.

Nevertheless, even the *weak core* characteristic is not sufficient to guarantee the equivalence of respective structures. For example, consider structures $\mathcal{P}_4, \mathcal{P}_5, \mathcal{P}_6$ in Figure 5.9. Observe that they all induce the same weak structure core $\mathcal{C}(\mathcal{P}_4) = \mathcal{C}(\mathcal{P}_5) = \mathcal{C}(\mathcal{P}_6) = \{U_1, U_2, U_3\}$ and, considering only Corollary 5.45, these structures could be equivalent. Nevertheless, they do not induce equal F-conditions: $\mathcal{F}(\mathcal{P}_4) = \{\langle u, w|y\rangle\}$ while $\mathcal{F}(\mathcal{P}_5) = \mathcal{F}(\mathcal{P}_6) = \emptyset$. Hence $\mathcal{P}_4, \mathcal{P}_5$ and $\mathcal{P}_4, \mathcal{P}_6$ cannot be equivalent by Corollary 5.16. (Note that $\mathcal{P}_5, \mathcal{P}_6$ are independence equivalent.)

5.4 Reduced structure

Another natural reasoning arises in connection with Corollary 5.45. Consider a structure $\mathcal{P}$ such that all its columns belong to its *weak core* $\mathcal{C}(\mathcal{P})$. Thus, considering Corollary 5.45, every equivalent structure $\mathcal{P}'$ contains all columns from $\mathcal{P}$. Above that, these columns are not only in its *weak core* $\mathcal{C}(\mathcal{P}')$, but they literally represent its complete *weak core* (i.e., if a column from $\mathcal{P}'$ does not belong to $\mathcal{P}$ it also lies outside of $\mathcal{C}(\mathcal{P}')$).

This gives us quite a good notion of possible equivalent structures. Such a structure in which where all its columns lie in its weak core as well, is literally a "core" of a class of independence equivalent structures. Beyond that, it explains why we call such a set of columns a *core*. It is literally the core of a class of equivalent structures. We denote such a structure by the epithet *reduced*.

Definition 5.47. *By a* reduced structure *we understand a structure $\mathcal{P}$ in which $K_i^{\mathcal{P}} \in \mathcal{C}(\mathcal{P})$ for all $i = 1, \ldots, |\mathcal{P}|$.*

Example 5.48. *Observe that structure $\mathcal{P}_1$ depicted in Figure 5.7a is not reduced since $K_5^{\mathcal{P}_1} \equiv U_5$ is trivial column, and as such it cannot be a part of the weak core. Similarly, $K_1^{\mathcal{P}_2} = S_2^{\mathcal{P}_2}$ in structure $\mathcal{P}_2$ from Figure 5.7b and therefore $K_1^{\mathcal{P}_2} \notin \mathcal{C}(\mathcal{P}_2)$ and $\mathcal{P}_2$ is not reduced. On the contrary, structure $\mathcal{P}_3$ in Figure 5.7c is reduced.*

Remark 5.49. *The epithet* "reduced" *comes from a parallel with fractions and their reduction (cancelation). When one writes out operators of composition applied on a generating sequence $\pi_1(K_1)$, $\pi_2(K_2)$, $\ldots$, $\pi_n(K_n)$, the following fraction appears:*

$$\frac{\pi_1(K_1) \cdot \pi_2(K_2) \cdot \ldots \cdot \pi_n(K_n)}{\pi_2(S_2) \cdot \ldots \cdot \pi_n(S_n)} \tag{5.4.1}$$

In the case of structures, we deal only with sets of variables $K_1, \ldots, K_n$. One can imagine the fraction (5.4.1) as

$$\frac{K_1 \cdot K_2 \cdot \ldots \cdot K_n}{S_2 \cdot \ldots \cdot S_n} \tag{5.4.2}$$

in that case. Considering the fact that there are sets instead of numbers in the respective fractions, we use the expression cancelation *for "reduction" of respective fractions. If, for example, $K_2 = S_n$, one can then perform the cancelation of K_2 and S_n in (5.4.2). The corresponding structure is not* reduced.

For every non-trivial column $K_i^{\mathcal{P}} \in ntriv(\mathcal{P})$, it holds that $R_i^{\mathcal{P}} \neq \emptyset$. Hence such a column introduces at least one new variable into the sequence $\mathcal{P} = K_1^{\mathcal{P}}, \ldots, K_n^{\mathcal{P}}$. Note that $\{R_i^{\mathcal{P}}\}_{i=1,\ldots,n}$ is a partition of $K_1^{\mathcal{P}} \cup \ldots \cup K_n^{\mathcal{P}}$. There is no trivial column in a reduced structure. Since the number of variables is limited, the number of columns has to be limited too. Moreover, it is limited by the number of variables $|N|$ over which the structure is defined.

Remark 5.50. *Let $\mathcal{P}$ be a reduced structure over N. Then $|\mathcal{P}| \leq |N|$.*

Corollary 5.51. *Let $\mathcal{P}$ be a reduced structure over N. Then every equivalent structure $\mathcal{P}'$ contains all columns from $\mathcal{P}$.*

Corollary 5.52. *Let $\mathcal{P}, \mathcal{P}'$ be two equivalent and reduced structures over N. Then $\mathcal{P}, \mathcal{P}'$ consist of the same columns (possibly in different ordering – they are each other's "permutations").*

5.5 Formal ratio

The concept of a *reduced structure* – or more precisely, Formula (5.4.2) – brought us to the idea of the so-called *formal ratio.* Formal ratio is a fraction with sets of variables in both numerator and denominator, assigned to every structure. Two formal ratios coincide if their numerators consist of the same sets and, at the same time, their denominators consist of the same sets.

Definition 5.53. *One writes a* formal ratio *for every structure $\mathcal{P}$ as follows: in the numerator one lists sets $K_i^{\mathcal{P}}$ for $i = 1, \ldots, |\mathcal{P}|$, while in the denominator one lists sets $S_i^{\mathcal{P}}$ for $i = 2, \ldots, |\mathcal{P}|$. Then "cancelation" is performed: one occurrence of a set $U \subseteq N$ in the denominator is canceled against one occurrence of U in the numerator.*

Example 5.54. *For example the structure $\mathcal{P}_2$ in Figure 5.7b induces the following "ratio".*

$$\frac{\{v,x\} \cdot \{v,w,x\} \cdot \{u,v,w\}}{\{v,x\} \cdot \{v,w\}} \tag{5.5.1}$$

Since the set $\{v,x\}$ is in both the numerator and the denominator, one can performs its "cancelation" to produce the formal ratio of $\mathcal{P}_2$

$$\frac{\{v,w,x\} \cdot \{u,v,w\}}{\{v,w\}}. \tag{5.5.2}$$

Compare this definition with Definition 5.37, the definition of a *weak core.* You will find out that these two definitions are symmetrical in a way. Let $\mathcal{P}$ be a structure. In a case where $\mathcal{P}$ is reduced, there can be no cancelation during the formal ratio creation process, and the formal ratio corresponds precisely to $\mathcal{P}$. This means:

Remark 5.55. *In the case of reduced structure $\mathcal{P}$, the numerator of the respective formal ratio corresponds to columns ($\{K_i^{\mathcal{P}}\}_{i=1,\ldots,|\mathcal{P}|}$) while the denominator consists of $\{S_i^{\mathcal{P}}\}_{i=2,\ldots,|\mathcal{P}|}$.*

In the case of non-reduced structure $\mathcal{P}$, there are at least two indices $i, j \in \mathbb{N}$ such that $K_i^{\mathcal{P}} = S_j^{\mathcal{P}}$. Hence a cancelation may be performed and these two sets are removed. It follows that:

Corollary 5.56. *The numerator of a formal ratio corresponds to the weak core of the corresponding structure.*

Example 5.57. *Consider structures $\mathcal{P}_4$, $\mathcal{P}_5$, and $\mathcal{P}_6$ from Figure 5.9. Note that all of these structures are reduced. Hence, their formal ratios are as follows:*

$$\mathcal{P}_4: \ \frac{\{u,v\}\cdot\{w,x\}\cdot\{u,w,y\}}{\{u,w\}}$$

$$\mathcal{P}_5: \ \frac{\{u,v\}\cdot\{w,x\}\cdot\{u,w,y\}}{u\cdot w}$$

$$\mathcal{P}_6: \ \frac{\{u,v\}\cdot\{w,x\}\cdot\{u,w,y\}}{u\cdot w}$$

Observe that formal ratios corresponding to $\mathcal{P}_5$ and $\mathcal{P}_6$ coincide.

The *formal ratio* is a very important feature of structures. As is shown later, it is one of the invariable properties of equivalent structures. In addition, it is a sufficient property through which one can decide on equivalence of the corresponding structures. It is another direct characterization of equivalent structures.

Considering Remark 5.55 and Lemma 5.21, one can prove a very interesting lemma that relates a formal ratio to non-trivial sets.

Lemma 5.58. *If formal ratios of two reduced structures $\mathcal{P},\mathcal{P}'$ coincide, then*

$$\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$$

.

Proof. Assume that both $\mathcal{P}$ and $\mathcal{P}'$ are defined over the set of variables N. We show that $(2^N \setminus \mathcal{N}(\mathcal{P})) = (2^N \setminus \mathcal{N}(\mathcal{P}))$, i.e., $U \notin \mathcal{N}(\mathcal{P}) \Leftrightarrow U \notin \mathcal{N}(\mathcal{P}')$ for $U \subseteq N$. Since the role of $\mathcal{P}$ and $\mathcal{P}'$ is interchangeable, it suffices to verify $(2^N \setminus \mathcal{N}(\mathcal{P})) \subseteq (2^N \setminus \mathcal{N}(\mathcal{P}))$. Assume that formal ratios of both structures coincide and they are equal to

$$\frac{U_1\cdot U_2\cdot\ldots\cdot U_n}{S_2\cdot\ldots\cdot S_n} \tag{5.5.3}$$

Put $I = \{i \in \mathbb{N} : \ U \subseteq K_i^{\mathcal{P}}\}$. Suppose $U \notin \mathcal{N}(\mathcal{P})$, then one can distinguish two cases by Lemma 5.21

I. $I = \emptyset$

II. $I \neq \emptyset$ and $\forall i \in I$ holds that $U \subseteq S_i^{\mathcal{P}}$

If $I = \emptyset$ then $U \nsubseteq K_k^{\mathcal{P}}$ for all $k = 1,\ldots,n$. Realizing the fact that the numerator of the respective formal ratio (5.5.3) corresponds to columns $\{K_k^{\mathcal{P}}\}_{k=1,\ldots,n}$ as well as $\{K_l^{\mathcal{P}'}\}_{l=1,\ldots,n}$ by Remark 5.55, it follows that $U \nsubseteq K_l^{\mathcal{P}'}$ for all $l = 1,\ldots,n$, which indicates that $U \notin \mathcal{N}(\mathcal{P}')$ by Lemma 5.21.

If $|I| \geq 1$ then the number of supersets of U in the numerator of (5.5.3) equals to $|I|$ as well as the number of supersets of U in its denominator. Since (5.5.3) is the same for $\mathcal{P}'$ and $\mathcal{P}'$ is reduced, then there are $|I|$ columns from $\mathcal{P}'$ such that $U \subseteq K_k^{\mathcal{P}'}$ for $k \in \mathbb{N}$, and $|I|$ columns where $U \subseteq S_l^{\mathcal{P}'}$ for $l \in \mathbb{N}$ by Remark 5.55. The fact ($U \subseteq S_i^{\mathcal{P}'} \Rightarrow U \subseteq K_i^{\mathcal{P}'}$) implies that $\forall i \in \{1, \ldots, n\}$ such that $U \subseteq K_i^{\mathcal{P}'}$ holds that $U \subseteq S_i^{\mathcal{P}'}$. This concludes $U \notin \mathcal{N}(\mathcal{P}')$ by Lemma 5.21. $\square$

Chapter 6

Indirect characterization

Recall that the previously formulated *Equivalence problem* includes the problem of how to recognize whether two given compositional model structures $\mathcal{P}, \mathcal{P}'$ over N induce the same independence model. This includes both:

(a) an easy rule for recognizing that two structures are equivalent in this sense. (Solution of this subproblem is usually named as *direct characterization* and it was partially solved in previous sections.)

(b) an easy way to get from $\mathcal{P}$ to $\mathcal{P}'$ in terms of some elementary operations. This is usually denoted as *indirect characterization*.

Note that another very important aspect of the equivalence problem is the ability to generate all structures that are equivalent to a given structure.

A reasonable solution to this problem would be a group of elementary operations. We consider those that change structure $\mathcal{P}$ while the *induced independence model* $\mathcal{I}(\mathcal{P})$ remains untouched. Concatenation of these operations would gradually generate a class (preferably the whole class) of equivalent structures.

What type of such operations might be expected? For simplicity, we must first restrict our attention to *reduced structures* defined at the end of the previous section. Recall that, according to Corollary 5.52, equivalent reduced structures consist of the same columns. Therefore they may differ only in column ordering. Thus, reduced equivalent structures are their own "permutations". In contrast, non-reduced structures may generally contain some extra columns. However, they are limited to those that do not affect the *weak core* $\mathcal{C}(\mathcal{P})$; trivial columns ($K_i = S_i$) and "strict" subsets of weak core columns ($K_i = S_j$ for some $K_j \in \mathcal{C}(\mathcal{P})$).

This means we may basically have only two types of operations: *permutations* and *adding/removing columns*. Considering the definition of a *weak core*, no other operations make sense.

By experimenting with equivalent structures, four different elementary operations on a structure-preserving independence model were discovered. We call them *IE (Independence Equivalent) operations.* Let us note that when these operations are applied to a structure, the respective generating sequence of the corresponding compositional model is accordingly modified. This issue is addressed in Chapter 8. We recognize the following operations:

IE operations

- *constant transposition*
- *box transposition*
- *simple reduction*
- *simple extension*

Furthermore, we distinguish several complex operations. As it will be proven later, each of these operations may be replaced by a sequence of IE operations. These complex operations are more than useful in proofs of various assertions.

Composed operations

- *left cycle permutation*
- *right cycle permutation*
- *box cycle permutation*
- *reduction*

The next subsection deals with permutations using the definitions of elementary permutations – the so-called *transpositions* – in which two adjacent columns exchange their position in a structure. We prove that an *induced independence model* is invariant with respect to these transpositions. Similarly, the induced independence model is invariant to permutations that can be implemented as sequences of these transpositions.

Operations that add or remove columns are introduced in the last part of this section. We similarly prove that the *induced independence model* of the structure is invariant with respect to these operations. In addition, we present an algorithm which converts any structure to its equivalent reduced form.

6.1 Transpositions and permutations

The question of whether two structures are equivalent – that is, whether they induce the same independence models – can be converted to the question of whether they induce the same dependence models. Recall that induced dependence model $\mathcal{D}(\mathcal{P})$ is a set of triples $\langle U, V|Z\rangle$ such that $U\not\perp\!\!\!\perp V|Z[\mathcal{P}]$. Note that we can confine our efforts to cases where U and V are singletons. Any other complex dependence relation can be replaced by a set of these elementary ones. This nontrivial consequence was already discussed in Remark 3.21.

By Definition 3.20, a triple $\langle u, v|Z\rangle \in \mathcal{D}(\mathcal{P})$ if and only if there exists a Z-avoiding trail between markers of u and v ($u\not\perp\!\!\!\perp v|Z[\mathcal{P}]$) in a corresponding persegram. If a structure and its permutation are independence equivalent, then one should be able to find "the same" Z-avoiding trails in both corresponding persegrams (and in all other equivalent permutations). By the same Z-avoiding trails we understand arbitrary Z-avoiding trails connecting the same couple of variables. Recall that a Z-avoiding trail is a sequence of markers; every marker is defined by its coordinates – the column and the variable.

In the case of permutation, we consider the following convention illustrated by Figure 6.1. We take permutation as a function and hence we can say that both the structure and its permutation consist of the same columns. Similarly, in case of persegrams, they consist of the same columns even if the shapes of several markers may differ. Observe that for structure U_1, U_2, U_3 in Figure 6.1a, its permutation U_3, U_1, U_2, and variable $v \in U_2$, while marker $[U_2, v]$ is a box-marker in U_1, U_2, U_3 (Figure 6.1a), it is a bullet in U_3, U_1, U_2 (Figure 6.1b). Hence, we suppose that markers are tied to columns and they move during permutations as well.

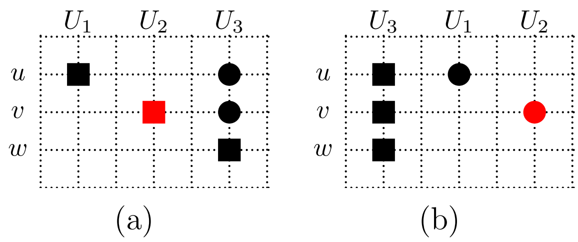


Figure 6.1: Marker $[U_2, v]$ moving and changing during permutation

Let $\mathcal{P}$ be a structure such that $u\not\perp\!\!\!\perp v|Z[\mathcal{P}]$. Then there exists at least one Z-avoiding trail connecting u with v in the respective persegram. Choose one of the trails of this type and denote it as τ.

Since columns are moving during a permutation, we move the sequence of markers τ in a corresponding way during the permutation as well. (Recall that τ is a sequence of markers in the permutated structure as well – possibly some bullets have changed into box-markers and vice versa). However, the new sequence of markers τ may not meet all conditions required from a Z-avoiding trail (see

Definition 3.18) in the permutated structure. Luckily, several properties from the definition are invariant with respect to a permutation. We state this observation in the form of a lemma since we refer to it in several following proofs.

Lemma 6.1. *Consider a structure and a sequence of markers τ such that τ is a Z-avoiding trail in the structure. Then τ meets conditions 0., 1., 3., and 4a. of Definition 3.18 in every permutation of the structure.*

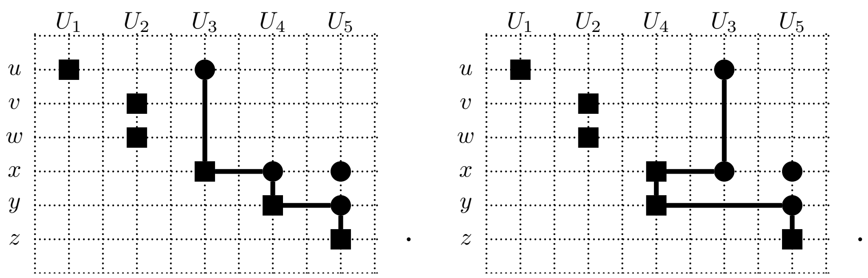
Proof. It is quite obvious: vertical and horizontal connections remain vertical and horizontal, and if they regularly alternate in $\mathcal{P}$, they do in $\mathcal{P}\sigma$ as well. Similarly, no horizontal connection starts to correspond to any variable from Z if it didn't before the permutation. □

So, the only possibility where τ may not meet the conditions of Definition 3.18 are the following:

- *Each vertical connection must be adjacent to a box-marker (one of the markers is a box-marker).*
- *Two vertical connections may be in direct succession if their common adjacent marker is a box-marker of a variable from Z.* (Notice that this condition is closely linked to the previous one.)

Remark 6.2. *Considering the previous, the only problem that can occur during a permutation of a structure is when a box-marker of a Z-avoiding trail τ turns into a bullet. Then τ may not be a Z-avoiding trail in the permuted structure.*

One can see the typical feature of Definition 3.18 during a permutation in the following example:

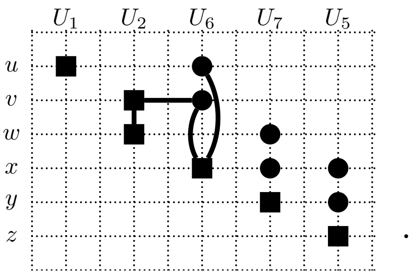


(a): $\{v, w\}$-avoiding trail (b): a sequence of markers

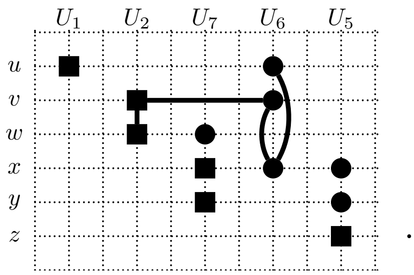
Figure 6.2: Feature of Condition 2. from Definition 3.18 during a structure permutation

Example 6.3. *Let $U_1, \ldots, U_5$ be a structure in Figure 6.2a. Observe the highlighted sequence of markers $[U_3, u], [U_3, x]$, $[U_4, x], [U_4, y]$, $[U_5, y], [U_5, z]$ which fulfils the conditions of Definition 3.18 and which forms a Z-avoiding trial representing $u \not\perp\!\!\!\perp z | Z[U_1, \ldots, U_5]$ for $Z \subseteq \{v, w\}$. One can see another structure*

U_1, U_2, U_4, U_3, U_5 depicted in Figure 6.2b – a permutation of $U_1, \ldots, U_5$. The sequence of the same markers $[U_3, u]$, $[U_3, x]$, $[U_4, x]$, $[U_4, y]$, $[U_5, y]$, $[U_5, z]$ is highlighted there as well. However, this one does not fulfill Condition 2 of Definition 3.18 – each vertical connection must be adjacent to a box-marker. *The sequence of markers in Figure 6.2b does not represent a Z-avoiding trail (for any such Z). Moreover $u \perp\!\!\!\perp z | \{v, w\}[U_1, U_2, U_4, U_3, U_5]$ and therefore $\mathcal{I}(U_1, \ldots, U_5) \neq \mathcal{I}(U_1, U_2, U_4, U_3, U_5)$.*



(a): $\{x\}$-avoiding trail



(b): a sequence of markers

Figure 6.3: Permutation causing a breakthrough of Condition 4b. from Definition 3.18

Similarly, Condition 4b from Definition 3.18 (Two vertical connections may be in direct succession if their common adjacent marker is a box-marker of a variable from Z) *may be broken. See Figure 6.3 where, in addition, Condition 2 is broken as well.*

Considering Lemma 6.1 and especially Remark 6.2, one can find out that the only way to cancel a Z-avoiding trail for a sequence of markers is when a box-marker changes into a bullet during the permutation. Hence, when no box-marker changes its shape during a permutation, one can conclude the following interesting assertion:

Corollary 6.4. *If τ is a Z-avoiding trail in a structure and none of its box-markers is changed into a bullet during its permutation, then τ is a Z-avoiding trail in the permuted structure as well.*

6.1.1 Basic notions of permutations

To increase the clarity of the text, the standard concept of permutation (including the notation) is used. Check [59] or [4] for more details. In the following, the basic notion of permutation is recalled, as well as examples for understanding permutations in the area of compositional models.

In mathematics, the notion of permutation is used with several slightly different meanings, all related to the act of permuting (rearranging in an ordered fashion) objects or values. Informally, a permutation of a set of values is an arrangement of those values into a particular order. Thus there are six permutations of the set $\{1,2,3\}$, namely $[1,2,3]$, $[1,3,2]$, $[2,1,3]$, $[2,3,1]$, $[3,1,2]$, and $[3,2,1]$.

Definition 6.5. *Let X be a set, then a* permutation *of X is a bijection $\sigma : X \to X$.*

There are several ways to write permutations. We use the *product of disjoint cycles* method, since it is well known that one can express every permutation as a product of disjoint cycles. Let $i_1, i_2, \ldots, i_r$ be distinct elements of X. The *r-cycle* $(i_1\ i_2\ \ldots\ i_r)$ is the permutation which maps $i_1 \mapsto i_2$, $i_2 \mapsto i_3$, $\ldots$, $i_{r-1} \mapsto i_r$, $i_r \mapsto i_1$ and fixes all other points in X. For a permutation of n symbols, the collection of all permutations of this set is denoted by T_n.

Note that the cycle $(1\ 2\ 4\ 5)$ can also be written as $(2\ 4\ 5\ 1)$, $(4\ 5\ 1\ 2)$, or $(5\ 1\ 2\ 4)$ since all these expressions contain the same information. It could not have been written as $(5\ 4\ 2\ 1)$.

If σ is a permutation of the set X, we shall write $i\sigma$ for the image of the element $i \in X$ under σ (rather then $\sigma(i)$). The principal reason for doing this is that it makes composition of permutations much easier: $\sigma_1\sigma_2$ will mean we apply σ_1 first and then apply σ_2, rather than the other way around.

In case of a compositional model structure (where elements are not integers), we should first find an association between each element (column) and an integer – the column index will be used for this purpose. Hence by σ permutation of structure $\mathcal{P} = K_1^{\mathcal{P}}, \ldots, K_n^{\mathcal{P}}$ we think of its permutation $\mathcal{P}\sigma$ where $K_{i\sigma}^{\mathcal{P}\sigma} = K_i^{\mathcal{P}}$.

Example 6.6. *Let $\mathcal{P} = U_1, U_2, \ldots, U_8$ be a compositional model structure and $\sigma \in T_5$ be a permutation. In a cycle such as $\sigma = (1\ 2\ 4\ 5)$ we mean that the permutation maps 1 to 2, 2 to 4, 4 to 5 and 5 to 1.*

Since $K_i^{\mathcal{P}} = K_{i\sigma}^{\mathcal{P}\sigma}$ by definition, then

$$\begin{aligned}
U_1 = K_1^{\mathcal{P}} &= K_{1\sigma}^{\mathcal{P}\sigma} \equiv K_2^{\mathcal{P}\sigma} \\
U_2 = K_2^{\mathcal{P}} &= K_{2\sigma}^{\mathcal{P}\sigma} \equiv K_4^{\mathcal{P}\sigma} \\
U_3 = K_3^{\mathcal{P}} &= K_{3\sigma}^{\mathcal{P}\sigma} \equiv K_3^{\mathcal{P}\sigma} \\
U_4 = K_4^{\mathcal{P}} &= K_{4\sigma}^{\mathcal{P}\sigma} \equiv K_5^{\mathcal{P}\sigma} \\
U_5 = K_5^{\mathcal{P}} &= K_{5\sigma}^{\mathcal{P}\sigma} \equiv K_1^{\mathcal{P}\sigma} \\
U_6 = K_6^{\mathcal{P}} &= K_{6\sigma}^{\mathcal{P}\sigma} \equiv K_6^{\mathcal{P}\sigma} \\
U_7 = K_7^{\mathcal{P}} &= K_{7\sigma}^{\mathcal{P}\sigma} \equiv K_7^{\mathcal{P}\sigma} \\
U_8 = K_8^{\mathcal{P}} &= K_{8\sigma}^{\mathcal{P}\sigma} \equiv K_8^{\mathcal{P}\sigma}
\end{aligned}$$

by definition of σ. Hence $\mathcal{P}\sigma = U_5, U_1, U_3, U_2, U_4, U_6, U_7, U_8$.

The "composition" of two permutations σ_1 and σ_2 is the function obtained by applying σ_1 first and then applying σ_2. Since we are writing maps on the right, we denote this by $\sigma_1\sigma_2$. Note that in general $\sigma_1\sigma_2 \neq \sigma_2\sigma_1$.

We put a special emphasis on the following very short cycles – *transpositions*: A *transposition* is a cycle of length two (that is, with two elements) – the so-called 2-cycle. Thus a transposition is a permutation $(i\ j)$ which simply swaps around the two elements i and j. Transpositions are useful for the following reason: Is is well known that *every permutation can be expressed as a product of transpositions.*

For example,

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

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

However, note that a decomposition into a product of transpositions is not unique nor is the number of transpositions unique. For example, the cycle (1 2 3) may be written as (1 2)(1 3), or (1 3)(2 3)(1 2)(1 2).

The final thing we need to do with permutations is *invert* them: Since each permutation is a bijection, it has an inverse which also is a bijection. The inverse transposition σ^{-1} of σ is the permutation that undoes the effect of applying σ. Thus if $\sigma : i \mapsto j$, then $\sigma^{-1} : j \mapsto i$.)

To find the inverse of a permutation that is a cycle, all we have to do is write elements of the cycle in reverse order. Thus, the inverse of (1 2 3 4) is (4 3 2 1). Since a cycle can be written with any of its elements as the first term, we can also write the inverse as (1 4 3 2). This provides an alternative way to write down the inverse of a cycle. Fix the first element in the cycle and write the remaining element in reverse order. Thus, the inverse of (1 2 3 4 5) is (1 5 4 3 2).

In a case where the permutation is a product of cycles, we must reverse the order of cycles as well as invert each cycle. For our need it is enough to find the inverse of a permutation that is a product of transpositions. Since for every transposition $\sigma = (i\ j) = (j\ i) = \sigma^{-1}$ then for the permutation written as the product of *transpositions* we need only reverse the order of the transpositions. Thus if $\sigma = \sigma_1\sigma_2 \dots \sigma_r$ where σ_i is a transposition for all $i = 1, \dots, r$ then $\sigma^{-1} = \sigma_r\sigma_{r-1} \dots \sigma_1$, i.e., for $\sigma = (1\,2)(1\,3)(1\,4)(1\,5)$ holds that $\sigma^{-1} = (1\,5)(1\,4)(1\,3)(1\,2)$.

A *permutation* of a given compositional model structure is a permutation of the columns it contains. As an example, if $\mathcal{P} = U_1, \dots, U_5$ and $\sigma = (1\ 2\ 4\ 5)$, then $\mathcal{P}\sigma = U_5, U_1, U_3, U_2, U_4$.

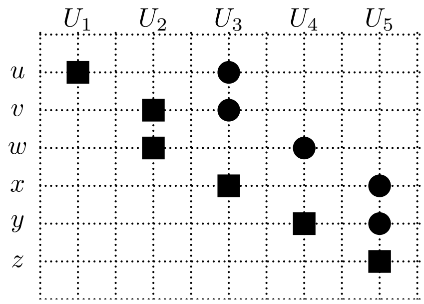
6.1.2 Constant transposition

Definition 6.7. *For $\mathcal{P}$ with $|\mathcal{P}| \geq 2$ and $k \in \{2, \dots, |\mathcal{P}|\}$ a transposition $\sigma = (k{-}1\ k) \equiv (k\ k{-}1)$ is said to be* constant *in $\mathcal{P}$ if $R^{\mathcal{P}}_{k-1} \cap K^{\mathcal{P}}_k = \emptyset$. We say that $\mathcal{P}\sigma$ is a* constant transposition *of $\mathcal{P}$.*

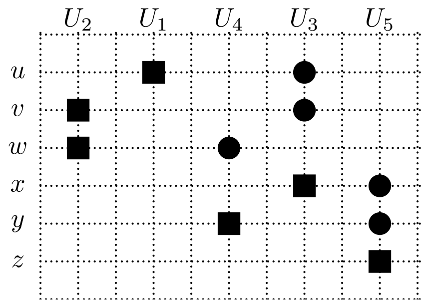
Remark 6.8. *Notice that, for $\sigma = (k{-}1\ k)$ and $\mathcal{P} = U_1, \dots, U_n$, it holds that $\mathcal{P}\sigma = U_1, \dots, U_{k-2}, U_k, U_{k-1}, U_{k+1}, \dots, U_n$. This means that $\mathcal{P}\sigma$ is created from $\mathcal{P}$ by swapping the positions of two adjacent columns U_{k-1}, U_k.*

Example 6.9. *Let $\mathcal{P}$ be the structure in Figure 6.4a. Since $R_1^{\mathcal{P}} \cap K_2^{\mathcal{P}} = \{u\} \cap \{v, w\} = \emptyset$, then by Definition 6.7 $\sigma_1 = (1\ 2)$ is a constant transposition in $\mathcal{P}$.*

Since $R_3^{\mathcal{P}\sigma_1} \cap K_4^{\mathcal{P}\sigma_1} = \{x\} \cap \{y, w\} = \emptyset$ then similarly $\sigma_2 = (3\ 4)$ is a constant transposition in $\mathcal{P}\sigma_1$. Put $\sigma = \sigma_1\sigma_2$, i.e., $\mathcal{P}\sigma$ – a permutation of $\mathcal{P}$ – was created from $\mathcal{P}$ by two constant transpositions. One can find $\mathcal{P}\sigma$ in Figure 6.4b.



(a): $\mathcal{P} = U_1, U_2, U_3, U_4, U_5$



(b): $\mathcal{P}\sigma = U_2, U_1, U_4, U_3, U_5$

Figure 6.4: A structure and its permutation caused by a double application of constant transposition

Lemma 6.10. *Consider structure $\mathcal{P}$ and a transposition σ which is* constant *in $\mathcal{P}$. Then $R_i^{\mathcal{P}} = R_{i\sigma}^{\mathcal{P}\sigma}$ and $S_i^{\mathcal{P}} = S_{i\sigma}^{\mathcal{P}\sigma}$ for all $i \in \{1, \dots, |\mathcal{P}|\}$.*

Proof. Since $K_i^{\mathcal{P}} = K_{i\sigma}^{\mathcal{P}\sigma}$ by definition of σ and $S_i = K_i \setminus R_i$, it is enough to prove that $R_i^{\mathcal{P}} = R_{i\sigma}^{\mathcal{P}\sigma}$ for all $i \in \{1, \dots, |\mathcal{P}|\}$.

Let $\sigma = (k{-}1\ k)$. Since $j = j\sigma$ for all $j \in \{1, \dots, |\mathcal{P}|\} \setminus \{k{-}1, k\}$, then $K_j^{\mathcal{P}} = K_j^{\mathcal{P}\sigma}$ and therefore $R_{j\sigma}^{\mathcal{P}\sigma} = R_j^{\mathcal{P}\sigma} = K_j^{\mathcal{P}\sigma} \setminus (K_1^{\mathcal{P}\sigma} \cup \dots \cup K_{j-1}^{\mathcal{P}\sigma}) = R_j^{\mathcal{P}}$ for the respective j. (Note that in the case of $j > k$, the formula $K_{(k-1)\sigma}^{\mathcal{P}\sigma} \cup K_{k\sigma}^{\mathcal{P}\sigma} = K_k^{\mathcal{P}} \cup K_{k-1}^{\mathcal{P}}$ is applied, the validity of which is guaranteed by the nature of transposition – a transposition only changes the order of columns but not their contents)

For index $k - 1$:

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

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

The part where $R^{\mathcal{P}\sigma}_{k\sigma} = R^{\mathcal{P}}_k$ is a direct consequence of the fact that $R^{\mathcal{P}\sigma}_1, \ldots, R^{\mathcal{P}\sigma}_n$ is a partition of $K^{\mathcal{P}\sigma}_1 \cup \ldots \cup K^{\mathcal{P}\sigma}_n = K^{\mathcal{P}}_1 \cup \ldots \cup K^{\mathcal{P}}_n$. □

Lemma 6.11. *Let σ be a constant transposition in $\mathcal{P}$. Then the formal ratios of both $\mathcal{P}$ and $\mathcal{P}\sigma$ coincide.*

Proof. Observe that both numerators and denominators of non-canceled ratios corresponding to $\mathcal{P}, \mathcal{P}\sigma$ are the same by the nature of permutation ($K^{\mathcal{P}}_i = K^{\mathcal{P}\sigma}_{i\sigma}$) and Lemma 6.10. Then the canceled ratios have to be the same as well, which proves the lemma. □

In the case of a persegram the set $R^{\mathcal{P}}_i$, or $S^{\mathcal{P}}_i$, corresponds to box-markers or to bullets, respectively, in the corresponding column. If the reader realizes this link, the adjective "constant" makes sense. It is summarized in the following corollary:

Corollary 6.12. *No persegram marker changes its shape during repositioning caused by* constant transposition *in the respective structure.*

Check the validity of Corollary 6.12 in Example 6.9 which deals with *constant transposition.* Observe that no box-marker turns into a bullet and vice versa in persegrams in Figure 6.4. To illustrate, see the following table listing all box markers of structure $\mathcal{P}$ and its transposition $\mathcal{P}\sigma$:

Box markers:

column	$\mathcal{P}$	$\mathcal{P}\sigma$
1st	$[U_1, u]$	$[U_2, v][U_2, w]$
2nd	$[U_2, v][U_2, w]$	$[U_1, u]$
3rd	$[U_3, x]$	$[U_4, y]$
4th	$[U_4, y]$	$[U_3, x]$
5th	$[U_5, z]$	$[U_5, z]$

Similarly, one can check all bullets in both structures $\mathcal{P}$ and $\mathcal{P}\sigma$:

Bullets:

column	$\mathcal{P}$	$\mathcal{P}\sigma$
1st	$\emptyset$	$\emptyset$
2nd	$\emptyset$	$\emptyset$
3rd	$[U_3, u][U_3, v]$	$[U_4, w]$
4th	$[U_4, w]$	$[U_3, u][U_3, v]$
5th	$[U_5, x][U_5, y]$	$[U_5, x][U_5, y]$

Remark 6.13. *Notice that if $\mathcal{P}\sigma$ is a constant transposition of $\mathcal{P}$, then σ is a constant transposition in $\mathcal{P}\sigma$ as well. Indeed, considering $\sigma = (k{-}1\ k)$ – a constant transposition in $\mathcal{P}$ – observe that $R^{\mathcal{P}\sigma}_{k-1} \cap K^{\mathcal{P}\sigma}_{k} = R^{\mathcal{P}}_{k} \cap K^{\mathcal{P}}_{k-1} = \emptyset$ by definition of σ and $R^{\mathcal{P}}_{k}$. This also means that that the role of $\mathcal{P}$ and $\mathcal{P}\sigma$ is interchangeable in a way that if $\mathcal{P}' = \mathcal{P}\sigma$ then $\mathcal{P} = \mathcal{P}'\sigma$.*

Notice that since each transposition is its own inversion $\sigma = \sigma^{-1}$, then $\sigma\sigma$ is an identical permutation and that is why $\mathcal{P}\sigma\sigma = \mathcal{P}$.

Lemma 6.14. *For a structure $\mathcal{P}$ and its constant transposition $\mathcal{P}\sigma$ it holds that*

$$\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}\sigma).$$

Proof. We show that $\mathcal{D}_{\mathcal{P}} = \mathcal{D}_{\mathcal{P}\sigma}$. Since the roles of $\mathcal{P}$ and $\mathcal{P}\sigma$ are interchangeable by Remark 6.13, it suffices to verify $\mathcal{D}_{\mathcal{P}} \subseteq \mathcal{D}_{\mathcal{P}\sigma}$.

Choose an arbitrary $u \not\perp\!\!\!\perp v|Z[\mathcal{P}]$. Then there exists a Z-avoiding trail τ between markers corresponding to u and v in $\mathcal{P}$. Label each marker of τ uniquely so that it can be identified in $\mathcal{P}\sigma$. Since no marker changes its shape by Corollary 6.12 and τ is a Z-avoiding trail in $\mathcal{P}$, then τ is a Z-avoiding trail in $\mathcal{P}\sigma$ by Corollary 6.4 as well. Hence $u \not\perp\!\!\!\perp v|Z[\mathcal{P}\sigma]$. $\mathcal{D}_{\mathcal{P}} \subseteq \mathcal{D}_{\mathcal{P}\sigma}$ holds true and the proof is complete. ☐

Observe that the proof of the previous lemma would be valid even for a more general permutation than a *constant transposition* defined in Definition 6.7. One might easily omit the restriction of two consecutive columns and consider an arbitrary permutation of all columns such that no box-marker from the whole persegram turns into a bullet and vice versa. The restriction of *two consecutive columns* was implemented due to Chapter 8 where the impact of these operations on the respective generating sequence and compositional model is investigated.

To simplify some of the following proofs, we introduce two special permutations that can be created by iterative applications of constant transpositions.

6.1.2.1 Left cycle permutation

Definition 6.15. *Consider the structure $\mathcal{P}$, $i \in \{1,\ldots,|\mathcal{P}|-2\}$, $k \in \{2,\ldots,|\mathcal{P}|-i\}$ such that $K^{\mathcal{P}}_{i} \supseteq S^{\mathcal{P}}_{i+k}$. Then we call a cycle $\sigma_L = (i{+}1\ i{+}2\ \ldots\ i{+}k)$ a* left cycle permutation *in $\mathcal{P}$. We say that $\mathcal{P}\sigma_L$ is a left cycle permutation of $\mathcal{P}$.*

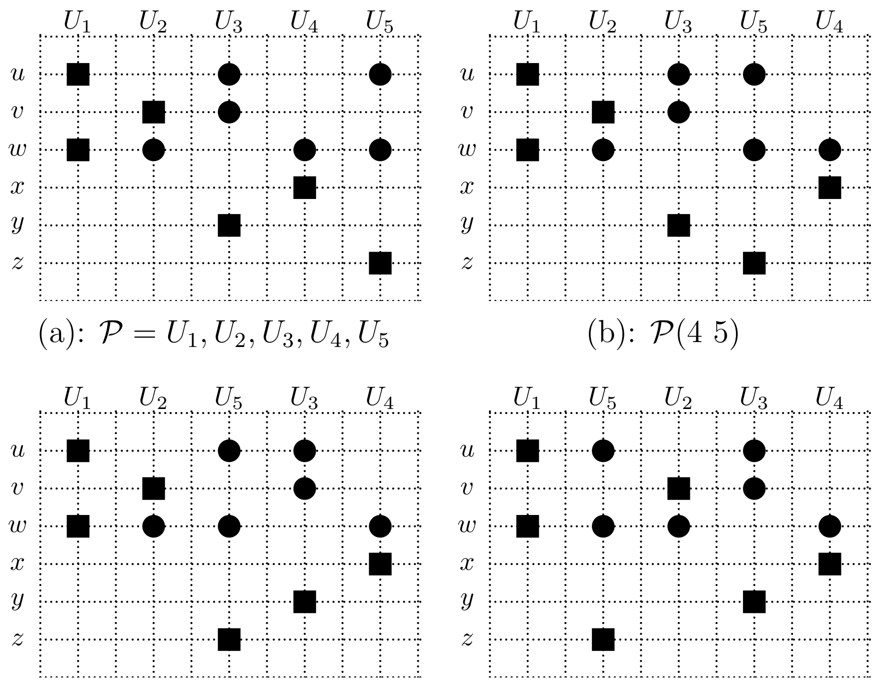
Example 6.16. *As an example of a left cycle permutation, take the structure $\mathcal{P} = U_1,\ldots,U_5$ in Figure 6.5a and consider the permutation $\sigma = (2\ 3\ 4\ 5)$. One can immediately see that σ is a left cycle permutation in this structure, since $K^{\mathcal{P}}_{1} = \{u,w\} \subseteq S^{\mathcal{P}}_{5} = \{u.w\}$. See permutation $\mathcal{P}\sigma$ in Figure 6.5d.*

Lemma 6.17. *If $\mathcal{P}\sigma_L$ is a* left cycle permutation *of $\mathcal{P}$ then one may obtain $\mathcal{P}\sigma_L$ from $\mathcal{P}$ by iterative applications of constant transpositions.*

Proof. Let $\sigma_L = (i{+}1 \ i{+}2 \ \dots \ i{+}k)$, then $K_i^{\mathcal{P}} \supseteq S_{i+k}^{\mathcal{P}}$ by Definition 6.15. Since every cycle of length k can be rewritten into a product of $k-1$ transpositions, we can express the considered permutation σ_L as $\sigma_L = \sigma_1\sigma_2\dots\sigma_{k-1} = (i{+}k \ i{+}k{-}1)(i{+}k{-}1 \ i{+}k{-}2)\dots(i{+}2 \ i{+}1)$.

Observe that $\forall j : i < j < i{+}k$ holds $S_{i+k}^{\mathcal{P}} \cap R_j^{\mathcal{P}} = \emptyset$ which, using Lemma 6.10, finishes the proof. □

Example 6.18. *Consider a structure $\mathcal{P}$ in Figure 6.5a. According to the previous proof, one can easily see that the permutation $\sigma = (2\ 3\ 4\ 5)$ (note that σ is a left cycle permutation in $\mathcal{P}$, cf. Example 6.16) may be replaced by a product of transpositions – $\sigma = (4\ 5)(3\ 4)(2\ 3)$ – which are all constant transpositions. While structures $\mathcal{P}(4\ 5)$ and $\mathcal{P}(4\ 5)(3\ 4)$ are on Figures 6.5b and 6.5c, respectively, structure $\mathcal{P}(4\ 5)(3\ 4)(2\ 3) \equiv \mathcal{P}(2\ 3\ 4\ 5)$ is in Figure 6.5d.*



(a): $\mathcal{P} = U_1, U_2, U_3, U_4, U_5$ (b): $\mathcal{P}(4\ 5)$

(c): $\mathcal{P}(4\ 5)(3\ 4)$ (d): $\mathcal{P}(4\ 5)(3\ 4)(2\ 3) = \mathcal{P}(2\ 3\ 4\ 5)$

Figure 6.5: Left cycle permutation as a product of constant transpositions

6.1.2.2 Right cycle permutation

The following special permutation can be similarly replaced by a sequence of *constant transpositions.* Nevertheless, it is obviously a simple generalization of the *constant transposition* as it is defined in Definition 6.7.

Definition 6.19. *Let $\mathcal{P}$ be a structure with $|\mathcal{P}| \geq 2$ and $i \in \{1, \ldots, |\mathcal{P}| - 1\}$, $k \in \{1, \ldots, |\mathcal{P}| - i\}$. We call a cycle $\sigma_R = (i{+}k \;\; i{+}k{-}1 \;\; \ldots \;\; i)$ a* right cycle permutation *in $\mathcal{P}$ if $R_i^{\mathcal{P}} \cap (K_{i+1}^{\mathcal{P}} \cup \ldots \cup K_{i+k}^{\mathcal{P}}) = \emptyset$. We say that $\mathcal{P}\sigma_R$ is a right cycle permutation of $\mathcal{P}$.*

Lemma 6.20. *Right cycle permutation $\mathcal{P}\sigma_R$ of $\mathcal{P}$ can by obtained from $\mathcal{P}$ by iterative applications of constant transpositions.*

Proof. The proof is conducted in the same style as the proof of the previous lemma – Lemma 6.17. Assume $\sigma_R = (i{+}k \quad i{+}k{-}1 \quad \ldots \quad i)$, then $R_i^{\mathcal{P}} \cap (K_{i+1}^{\mathcal{P}} \cup \ldots \cup K_{i+k}^{\mathcal{P}}) = \emptyset$ by Definition 6.19. Observe that $\sigma_R = \sigma_1 \ldots \sigma_k$ where $\sigma_x = (i{+}x{-}1 \quad i{+}x)$ for every $x \in \{1, \ldots, k\}$. Put $\mathcal{P}^0 = \mathcal{P}$ and $\mathcal{P}^x = \mathcal{P}^{x-1}\sigma_x$ for all $x = 1, \ldots, k$.

Let us prove by induction on x that σ_x is a *constant transposition* in $\mathcal{P}^{x-1}$. It is obvious for $x = 1$. Assume that the induction hypothesis holds for $x - 1$ and $\mathcal{P}^{x-1}$ is obtained from $\mathcal{P}$ by a sequence of *constant transpositions* $\sigma_1, \ldots, \sigma_{x-1}$. To prove that σ_x is a local transposition in $\mathcal{P}^{x-1}$, it is sufficient to show that $R_{i+x-1}^{\mathcal{P}^{x-1}} \cap K_{i+x}^{\mathcal{P}^{x-1}} = \emptyset$. Note that to make it clear, there is an index number in brackets, while parentheses contain transpositions in what follows. Since

$$\begin{aligned}
[i + x - 1](\sigma_1 \ldots \sigma_{x-1})^{-1} &= [i + x - 1]\sigma_{x-1}^{-1} \ldots \sigma_1^{-1} \\
&= [i + x - 1]\sigma_{x-1} \ldots \sigma_1 \\
&= [i + x - 1](i{+}x{-}1 \quad i{+}x{-}2)(i{+}x{-}2 \quad i{+}x{-}3) \ldots (i{+}1 \quad i) \\
&= i
\end{aligned}$$

and

$$\begin{aligned}
[i + x](\sigma_1 \ldots \sigma_{x-1})^{-1} &= [i + x]\sigma_{x-1}^{-1} \ldots \sigma_1^{-1} \\
&= [i + x]\sigma_{x-1} \ldots \sigma_1 \\
&= [i + x](i{+}x{-}1 \quad i{+}x{-}2)(i{+}x{-}2 \quad i{+}x{-}3) \ldots (i{+}1 \quad i) \\
&= i + x
\end{aligned}$$

then $R_{i+x-1}^{\mathcal{P}^{x-1}} \cap K_{i+x}^{\mathcal{P}^{x-1}} = R_i^{\mathcal{P}} \cap K_{i+x}^{\mathcal{P}}$ by induction assumption and Lemma 6.10 where $R_i^{\mathcal{P}} \cap K_{i+x}^{\mathcal{P}} \subseteq R_i^{\mathcal{P}} \cap (K_{i+1}^{\mathcal{P}} \cup \ldots \cup K_{i+k}^{\mathcal{P}}) = \emptyset$ by the assumption and the fact that $1 \leq x \leq k$. □

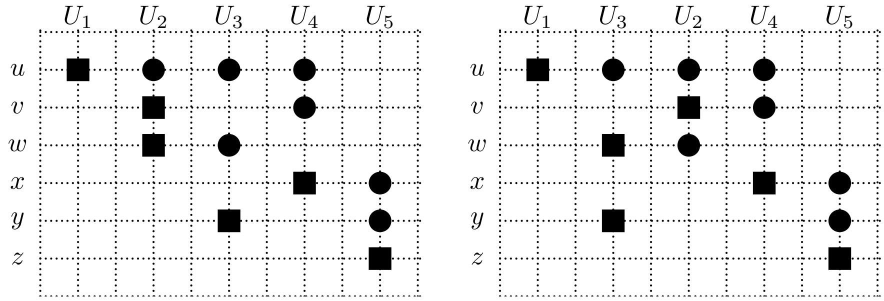
Since an *induced independence model* is invariable with respect to a *constant transposition* by Lemma 6.14, we get that an *induced independence model* is invariant with respect to *left and right cycle permutations* as well, by Lemmata 6.17 and 6.20. This observation is summarized in the following corollary:

Corollary 6.21. *If σ is a left or right cycle permutation in $\mathcal{P}$, then $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}\sigma)$.*

6.1.3 Box transposition

Definition 6.22. *For $\mathcal{P}$ with $|\mathcal{P}| \geq 2$ and $k \in \{2, \ldots, |\mathcal{P}|\}$ we call transposition $\sigma = (k{-}1\ k) \equiv (k\ k{-}1)$ a* box transposition *in $\mathcal{P}$ if $S^{\mathcal{P}}_{k-1} \subseteq S^{\mathcal{P}}_{k} \subseteq K^{\mathcal{P}}_{k-1}$. We say that $\mathcal{P}\sigma$ is* box transposition *of $\mathcal{P}$.*

Example 6.23. *Let $\mathcal{P} = U_1, U_2, U_3, U_4, U_5$ be a structure from Figure 6.6a. Observe that for adjacent columns $U_2 \equiv K^{\mathcal{P}}_2$ and $U_3 \equiv K^{\mathcal{P}}_3$ holds: $S^{\mathcal{P}}_2 \subset S^{\mathcal{P}}_3 \subset K^{\mathcal{P}}_2$. By Definition 6.22 the transposition $\sigma = (2\ 3)$ is a box transposition in $\mathcal{P}$: $\mathcal{P}\sigma = U_1, U_3, U_2, U_4, U_5$ was created from $\mathcal{P}$ by box-transposition. See both, structure $\mathcal{P}$ and its box transposition $\mathcal{P}\sigma$, in Figure 6.6.*



(a): $\mathcal{P}$: original structure (b): $\mathcal{P}\sigma$ – box transposition of $\mathcal{P}$

Figure 6.6: Box transposition

If one examines the two structures in Figure 6.6 more closely, one finds out that, regardless of the permutation, all bullets seem to hold their positions. The only thing that has changed is the location of some box-markers. This observation holds for any box-transposition and that is also the reason why we call this transposition a *box transposition*. This observation is precisely formulated in the following assertion.

Lemma 6.24. *If σ is a box transposition in $\mathcal{P}$ then $S^{\mathcal{P}}_i = S^{\mathcal{P}\sigma}_i$ for all $i = 1, \ldots, |\mathcal{P}|$.*

Proof. Suppose $\sigma = (k{-}1\ k)$. First, we will show that

$$(K^{\mathcal{P}}_i \cap K^{\mathcal{P}}_{k-1}) = (K^{\mathcal{P}}_i \cap K^{\mathcal{P}}_k) \text{ for all } i \in \{1, \ldots, k{-}2\}. \tag{6.1.1}$$

To do so, realize that since σ is a box transposition in $\mathcal{P}$ ($S^{\mathcal{P}}_{k-1} \subseteq S^{\mathcal{P}}_{k} \subseteq K^{\mathcal{P}}_{k-1}$) and $i < k{-}1$ then

$$\begin{aligned} K^{\mathcal{P}}_i \cap K^{\mathcal{P}}_{k-1} &= (K^{\mathcal{P}}_i \cap S^{\mathcal{P}}_{k-1}) \cup (K^{\mathcal{P}}_i \cap R^{\mathcal{P}}_{k-1}) \\ &= K^{\mathcal{P}}_i \cap S^{\mathcal{P}}_{k-1} \\ &\subseteq K^{\mathcal{P}}_i \cap S^{\mathcal{P}}_{k} \\ &\subseteq K^{\mathcal{P}}_i \cap K^{\mathcal{P}}_{k-1}. \end{aligned}$$

As the first and last member of the previous inclusion sequence are the same, then all elements are equal and therefore $K_i^{\mathcal{P}} \cap K_{k-1}^{\mathcal{P}} = K_i^{\mathcal{P}} \cap S_k^{\mathcal{P}} = K_i^{\mathcal{P}} \cap K_k^{\mathcal{P}}$ since $K_i^{\mathcal{P}} \cap R_k^{\mathcal{P}} = \emptyset$ by the fact that $i < k$. So we have proven that (6.1.1) holds.

Now, since $j = j\sigma$ for all $j \in \{1, \ldots, |\mathcal{P}|\} \setminus \{k-1, k\}$ then $K_j^{\mathcal{P}} = K_j^{\mathcal{P}\sigma}$ and therefore $S_j^{\mathcal{P}\sigma} = K_j^{\mathcal{P}\sigma} \cap (K_1^{\mathcal{P}\sigma} \cup \ldots \cup K_{j-1}^{\mathcal{P}\sigma}) = S_j^{\mathcal{P}}$ (note that in case of $j > k$ the knowledge of $K_{k-1}^{\mathcal{P}\sigma} \cup K_k^{\mathcal{P}\sigma} = K_k^{\mathcal{P}} \cup K_{k-1}^{\mathcal{P}}$ is successfully applied, which follows from the nature of permutations).

Using (6.1.1) let us show for index k:

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

Similarly one can see that for index $k-1$:

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

$\square$

Example 6.25. *Consider again structures $\mathcal{P}, \mathcal{P}\sigma$ from Figure 6.6 where $\sigma = (2\ 3)$ is a box transposition in $\mathcal{P}$. Observe that $S_2^{\mathcal{P}} = u = S_2^{\mathcal{P}\sigma}$, $S_3^{\mathcal{P}} = \{u, w\} = S_3^{\mathcal{P}\sigma}$, etc.*

While box-marker $[K_2, w]_{\mathcal{P}}$ remains at its position – i.e., $[K_2, w]_{\mathcal{P}\sigma}$ is also box marker, the box-marker corresponding to y moves from the third column to the second one ($[K_3, y]_{\mathcal{P}} \equiv [K_2, y]_{\mathcal{P}\sigma}$) and box-marker $[K_2, v]_{\mathcal{P}}$ moves to the third column on position $[K_3, v]_{\mathcal{P}\sigma}$.

Remark 6.26. *Observe that if σ is a box transposition in $\mathcal{P}$, then it is a box transposition in $\mathcal{P}\sigma$ as well. To prove this, check $\sigma = (k-1\ k)$ for the validity of $S_{k-1}^{\mathcal{P}\sigma} \subseteq S_k^{\mathcal{P}\sigma} \subseteq K_{k-1}^{\mathcal{P}\sigma}$. It corresponds to $S_{k-1}^{\mathcal{P}} \subseteq S_k^{\mathcal{P}} \subseteq K_k^{\mathcal{P}}$ by definition of σ and Lemma 6.24. While the first inclusion is guaranteed by the fact that σ is a box transposition in $\mathcal{P}$, the second follows from the definition of $S_k^{\mathcal{P}}$. Moreover, since any transposition is its own inversion: $\sigma = \sigma^{-1}$ then $\mathcal{P}\sigma\sigma = \mathcal{P}$. Hence the roles of $\mathcal{P}$ and $\mathcal{P}\sigma$ are interchangeable with respect to σ. ($\mathcal{P}' = \mathcal{P}\sigma$ if and only if $\mathcal{P} = \mathcal{P}'\sigma$)*

We have shown that no marker changes its shape during constant transposition. The question, "How do markers behave during box transposition?" naturally arises. Note that a box transposition affects only two adjacent columns. All other columns (as well as shape of the corresponding markers) remain untouched.

Lemma 6.27. *Let $\sigma = (k-1 \quad k)$ be a box-transposition in structure $\mathcal{P} = U_1, \ldots, U_n$. Then no box-marker turns into a bullet during box-transposition σ in all columns except for U_{k-1}.*

Proof. Recall that $i\sigma = i$ for all $i \in \{1, \ldots, n\} \setminus \{k–1, k\}$ by definition of σ. Since $S_i^{\mathcal{P}} = S_i^{\mathcal{P}\sigma}$ by Lemma 6.24 then the lemma is proven for all $i \in \{1, \ldots, n\} \setminus \{k–1, k\}$ and corresponding columns U_i.

Check the lemma also in the case of U_k. Observe that $U_k \equiv K_k^{\mathcal{P}} = K_{k-1}^{\mathcal{P}\sigma}$ and $S_{k-1}^{\mathcal{P}\sigma} = S_{k-1}^{\mathcal{P}} \subseteq S_k^{\mathcal{P}}$ by Lemma 6.24 and Definition 6.22. This has the following meaning: All bullets from the column corresponding to U_k in $\mathcal{P}\sigma$ (i.e., $S_{k-1}^{\mathcal{P}\sigma}$) are bullets in the column corresponding to U_k in $\mathcal{P}$ as well (i.e., $S_{k-1}^{\mathcal{P}\sigma} \subseteq S_k^{\mathcal{P}}$). Hence, no box-marker could change into a bullet in the column corresponding to U_k during the transposition σ from $\mathcal{P}$ to $\mathcal{P}\sigma$. ∎

Example 6.28. *See Figure 6.6 to illustrate Lemma 6.27. We list box markers corresponding to all sets from the structure. Note that structure $\mathcal{P}\sigma = U_1, U_3, U_2, U_4, U_5$ from 6.6b was obtained from $\mathcal{P} = U_1, U_2, U_3, U_4, U_5$ (depicted in Figure 6.6a) by box transposition $\sigma = (2\ 3)$. Observe that, as Lemma 6.27 claims, no box-marker, except for those from U_2, changes into a bullet during the transposition.*

Box markers:

column	$\mathcal{P}$	$\mathcal{P}\sigma$
U_1	$[U_1, u]$	$[U_1, u]$
U_2	$[U_2, v]$ $[U_2, w]$	$[U_2, v]$
U_3	$[U_3, y]$	$[U_3, w]$ $[U_3, y]$
U_4	$[U_4, x]$	$[U_4, x]$
U_5	$[U_5, z]$	$[U_5, z]$

Observe that while the emphasized box-marker corresponding to variable w appears for structure $\mathcal{P}$ in U_2, for $\mathcal{P}(2\ 3)$ it is in U_3.

Lemma 6.29. *If σ is a box transposition in $\mathcal{P}$ then the formal ratio is the same for both $\mathcal{P}$ and $\mathcal{P}\sigma$.*

Proof. Observe that both numerators and denominators of non-canceled ratios corresponding to $\mathcal{P}, \mathcal{P}\sigma$ are the same by the nature of permutation ($K_i^{\mathcal{P}} = K_{i\sigma}^{\mathcal{P}\sigma}$) and Lemma 6.24 ($S_i^{\mathcal{P}} = S_i^{\mathcal{P}\sigma}$). Then the formal ratios have to be the same as well, which proves the lemma. ∎

The fact that a formal ratio is an invariable attribute with respect to box transpositions implies that a *weak structure core* is invariable in relation to box transpositions as well. Realizing the fact that a weak core is invariable with respect to independence equivalence, another natural question arises in connection with Lemma 6.29: is the induced independence model $\mathcal{D}(\mathcal{P})$ invariant with respect to box transpositions? The answer is yes:

Lemma 6.30. *If σ is a box transposition in $\mathcal{P}$ then $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}\sigma)$.*

Proof. We show that $\mathcal{D}(\mathcal{P}) = \mathcal{D}(\mathcal{P}\sigma)$. Since the roles of $\mathcal{P}$ and $\mathcal{P}\sigma$ are interchangeable by Remark 6.26 it suffices to verify $\mathcal{D}(\mathcal{P}) \subseteq \mathcal{D}(\mathcal{P}\sigma)$. Let $\mathcal{P} = U_1, \ldots, U_n$ and $\sigma = (k{-}1\ k)$ where $k \leq n$.

Assume for contradiction $u \not\perp\!\!\!\perp v | Z[\mathcal{P}]$ while $u \perp\!\!\!\perp v | Z[\mathcal{P}\sigma]$. Without loss of generality, consider a trail $\tau = [U_{i_0}, u_{j_0}], [U_{i_1}, u_{j_1}], \ldots [U_{i_m}, u_{j_m}]$ representing $u \not\perp\!\!\!\perp v | Z[\mathcal{P}]$ such that τ involves the minimal number of markers among all trails of this type. (Observe that $u_{j_0} = u$ and $u_{j_m} = v$.) Based on Lemma 6.1 and Corollary 6.4, one can distinguish two cases:

I) No marker of τ turns from a box-marker into a bullet during the transposition σ.

II) At least one box-marker from τ turns into a bullet during the transposition σ.

In a case where no box marker turns into a bullet during the transposition, τ is a Z-avoiding trail by Corollary 6.4 in $\mathcal{P}\sigma$ as well, and the proof is complete.

Assume that there is at least one marker of τ which turns from a box-marker in $\mathcal{P}$ into a bullet in $\mathcal{P}\sigma$. Every Z-avoiding trail consists of horizontal and vertical connections, where each vertical connection must be adjacent to a box-marker by Definition 3.18. Assume the existence of at least one marker such that the respective vertical connection is adjacent only to bullets. (Otherwise τ fulfills Definition 3.18 and the proof is complete.) Note that by Lemma 6.27 the only opportunity for this to happen is when the marker corresponds to a variable from U_{k-1}.

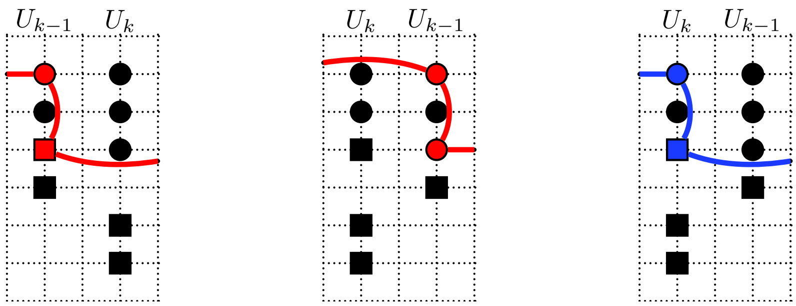
Hence $\tau = [U_{i_0}, u], \ldots, [U_{k-1}, u_{x-1}], [U_{k-1}, u_x], \ldots, [U_{i_m}, v]$ where

$$\{u_{x-1}, u_x\} \subseteq S_k^{\mathcal{P}\sigma} \tag{6.1.2}$$

while $\{u_{x-1}, u_x\} \not\subseteq S_{k-1}^{\mathcal{P}}$ by our assumptions (indeed, recall that τ is a Z-avoiding trail in $\mathcal{P}$, and each vertical connection is adjacent to a box-marker; box markers of $K_{k-1}^{\mathcal{P}}$ correspond to $R_{k-1}^{\mathcal{P}}$). Without loss of generality, let us assume that

$$u_x \in R_{k-1}^{\mathcal{P}},\ (i.e.\ u_x \notin S_{k-1}^{\mathcal{P}}). \tag{6.1.3}$$

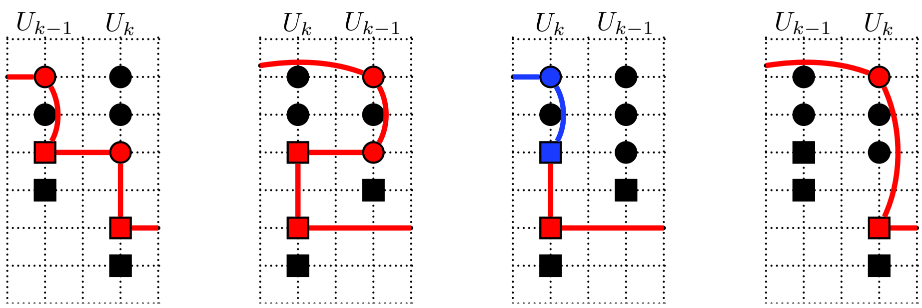
Then, however, $\{u_{x-1}, u_x\} \subseteq S_k^{\mathcal{P}} \subseteq U_k$ by (6.1.2) and Lemma 6.24. One can move the respective vertical connection to the columns corresponding to U_k. Then $u_x \in R_{k-1}^{\mathcal{P}\sigma}$ by the fact that $K_{k-1}^{\mathcal{P}\sigma} = U_k$, Lemma 6.24, and (6.1.3), the new vertical connection is adjacent to a box marker. Repeat the process for all "broken" vertical connections and then denote the new sequence of markers by τ' – note that each of its vertical connections is adjacent to a box-marker now. See Figure 6.7 for illustration – however, columns out of focus are omitted in this Figure for the sake of lucidity.



(a): $\mathcal{P} : \tau$ (b): $\mathcal{P}\sigma : \tau$ (c): $\mathcal{P}\sigma : \tau'$

Figure 6.7: Correction of a trail damaged by box transposition $\sigma = (k-1, k)$

A careful reader may object that Condition 4.a in Definition 3.18 may be corrupted during execution of the previous correcting algorithm (*vertical and horizontal connections may not regularly alternate in* τ'); see Figure 6.8 for illustration:



(a): $\mathcal{P}:\ \tau$ (b): $\mathcal{P}\sigma:\ \tau$ (c): $\mathcal{P}\sigma:\ \tau'$ (d): $\mathcal{P}$: trail shorter than τ

Figure 6.8: First trail, which leads to two verticals, contradicts with the trail on fourth figure which is shorter.

Considering Lemma 6.1 and the fact that τ is a Z-avoiding trail in $\mathcal{P}$, it may happen that by moving an invalid vertical connection from U_{k-1} to U_k, two adjacent vertical connections appear in U_k. (See Figure 6.8c.)

In that case, however, τ is not the shortest Z-avoiding trail in $\mathcal{P}$, which contradicts the assumption (Compare Figures 6.8a and 6.8d). □

6.1.3.1 Box cycle permutation

Definition 6.31. *Consider a structure $\mathcal{P}$ with $|\mathcal{P}| \geq 2$, $i \in \{1, \ldots, |\mathcal{P}|-1\}$ and $k \in \{1, \ldots, |\mathcal{P}|-i\}$. We call a cycle $\sigma_B = (i\ i{+}1\ \ldots\ i{+}k)$ a* box cycle permutation *in $\mathcal{P}$ if $S_i^{\mathcal{P}} \subseteq S_{i+k}^{\mathcal{P}} \subseteq K_i^{\mathcal{P}}$. We say that $\mathcal{P}\sigma_B$ is a* box cycle permutation *of $\mathcal{P}$.*

Lemma 6.32. *If $\mathcal{P}\sigma_B$ is a* box cycle permutation *of $\mathcal{P}$, then one may obtain $\mathcal{P}\sigma_B$ from $\mathcal{P}$ by a sequence of constant and box transpositions.*

Proof. Let $\sigma_B = (i\ i{+}1\ \ldots\ i{+}k)$ for some $i \in \{1, \ldots, |\mathcal{P}|{-}1\}$ and $k \in \{1, \ldots, |\mathcal{P}|-i\}$. Since σ_B is a *box cycle permutation* in $\mathcal{P}$,

$$S_i^{\mathcal{P}} \subseteq S_{i+k}^{\mathcal{P}} \subseteq K_i^{\mathcal{P}} \tag{6.1.4}$$

by Definition 6.31. If $k = 1$, then σ_B coincides with *box transposition* $(i\ i{+}1)$ and the lemma is proven.

Suppose that $k \geq 2$. In that case σ_B may be expressed as a convolution $\sigma_B = \sigma_L\sigma_b$ where $\sigma_L = (i{+}1\ \ldots\ i{+}k)$ and $\sigma_b = (i\ i{+}1)$. Observe that σ_L is a *left cycle permutation* in $\mathcal{P}$ by (6.1.4). Recall that any *left cycle permutation* may be replaced by a sequence of constant transpositions due to Lemma 6.17, and that is also why

$$S_i^{\mathcal{P}\sigma_L} \subseteq S_{i+1}^{\mathcal{P}\sigma_L} \subseteq K_i^{\mathcal{P}\sigma_L} \tag{6.1.5}$$

holds by Lemma 6.10 and the fact that $(i + k)\sigma_L = i + 1$, $i\sigma_L = i$, and (6.1.4). Hence, however, σ_b is a *box transposition* in $\mathcal{P}\sigma_L$ by (6.1.5). And the lemma is proved. □

Remark 6.33. *Following the proof of Lemma 6.32, one may notice that, since $\sigma_L = (i{+}1\ \ldots\ i{+}k)$ is a left cycle in $\mathcal{P}$, then $S_{i+k}^{\mathcal{P}} = S_{i+1}^{\mathcal{P}\sigma_L}$ by Lemma 6.17 and Lemma 6.10. Considering $\sigma_b = (i\ i{+}1)$ – a box transposition in $\mathcal{P}\sigma_L$ – Lemma 6.24 implies that $S_{i+1}^{\mathcal{P}\sigma_L} = S_{i+1}^{\mathcal{P}\sigma_L\sigma_b}$. Then, $\sigma_B = \sigma_L\sigma_b$ implies*

$$S_{i+1}^{\mathcal{P}\sigma_B} = S_{i+k}^{\mathcal{P}}. \tag{6.1.6}$$

Note that

$$K_{i+1}^{\mathcal{P}\sigma_B} = K_i^{\mathcal{P}} \tag{6.1.7}$$

by definition of cycle σ_B (recall that σ_B maps i to $i{+}1$).

The observations (6.1.6) and (6.1.7) discussed in the previous remark have a very important and interesting impact on non-reduced structures. Recall that a structure is not reduced if at least one of its columns does not simultaneously belong to its weak core.

Let $\mathcal{P}$ be a non-reduced structure with a non-trivial column $K_i^{\mathcal{P}}$ such that $K_i^{\mathcal{P}} \notin \mathcal{C}(\mathcal{P})$. Then, by the definition of a weak core – Definition 5.37 – there exists $k \geq 0$ such that

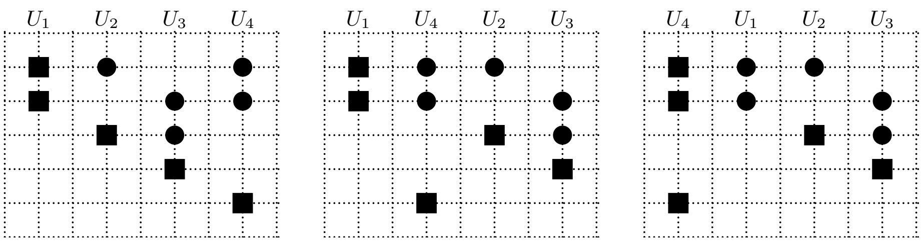
$$K_i^{\mathcal{P}} = S_{i+k}^{\mathcal{P}}. \tag{6.1.8}$$

Observe that

$$\boldsymbol{S_i^{\mathcal{P}}} \subseteq K_i^{\mathcal{P}} = \boldsymbol{S_{i+k}^{\mathcal{P}} = K_i^{\mathcal{P}}} \tag{6.1.9}$$

by our assumptions where the first inclusion is guaranteed by the general relationship between K_i and S_i. Then, however, considering the emphasized parts in (6.1.9), one gets that $\sigma_B = (i\ i{+}1\ \dots\ i{+}k)$ is a *box cycle* permutation in $\mathcal{P}$. If we use the consequences following from the previous remark, then $S_{i+1}^{\mathcal{P}\sigma_B} = S_{i+k}^{\mathcal{P}} = K_i^{\mathcal{P}} = K_{i+1}^{\mathcal{P}\sigma_B}$. Note that the respective equations correspond to (6.1.6), (6.1.8), and (6.1.7), in this order. Hence, $K_{i+1}^{\mathcal{P}\sigma_B} \notin ntriv(\mathcal{P})$ and $\mathcal{P}\sigma$ is a little bit closer to be a reduced structure while $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}\sigma_B)$ by Lemma 6.32 and Lemmata 6.14 and 6.30.

This is consistent with the concept of a *weak structure core*: the column not belonging to the corresponding *weak structure core* need not be included in all equivalent structures. (Recall that the trivial column brings no additional information with respect to the *induced independence model* $\mathcal{I}(\mathcal{P})$).



(a): $\mathcal{P}$ (b): $\mathcal{P}\sigma_L$ (c): $\mathcal{P}\sigma_L\sigma_b$

Figure 6.9: Sequence of independence equivalent structures

Example 6.34. *Consider structure $\mathcal{P} = U_1, \dots, U_4$ in Figure 6.9a. Observe that $K_1^{\mathcal{P}} = S_4^{\mathcal{P}}$. Hence, following the proof of Lemma 6.32, one can easily apply the left cycle permutation $\sigma_L = (2\ 3\ 4)$ where $\mathcal{P}\sigma_L$ can be found in Figure 6.9b. Observe that $\sigma_b = (1\ 2)$ is a box transposition in $\mathcal{P}\sigma_L$. $\mathcal{P}\sigma_L\sigma_b$ is shown in Figure 6.9c. Note that $\sigma_L\sigma_b = (1\ 2\ 3\ 4)$ is a box cycle permutation in $\mathcal{P}$.*

Example 6.35. *Consider structure U_1, U_2, U_3 for $U_1 = \{u, v, w\}, U_2 = \{v, w, x\}$, and $U_3 = \{w, x, y\}$ in Figure 6.10a. Recall that there are six permutations of the set $\{1, 2, 3\}$, namely $[1, 2, 3]$, $[1, 3, 2]$, $[2, 1, 3]$, $[2, 3, 1]$, $[3, 1, 2]$, and $[3, 2, 1]$. Similarly, there are six permutations of U_1, U_2, U_3 including the identical one. Note that only four of them are independence equivalent with U_1, U_2, U_3 in this case. See the overview in the following table, and the explanation below:*

permutation	*equivalence with* U_1, U_2, U_3
U_1, U_2, U_3	✓
U_1, U_3, U_2	$\times$
U_2, U_1, U_3	✓
U_2, U_3, U_1	✓
U_3, U_1, U_2	$\times$
U_3, U_2, U_1	✓

Regardless of the fact that we do not have any necessary and simultaneously sufficient conditions for independence equivalence between two given structures yet, verification of the table from above is simple. Using Figures 6.11a and 6.11b, observe that $\mathcal{C}(U_1, U_2, U_3) = \{U_1, U_2, U_3, \}$ *while* $\mathcal{C}(U_1, U_3, U_2) = \{U_1, U_3\} = \mathcal{C}(U_3, U_1, U_2)$*. Hence* $\mathcal{I}(U_1, U_3, U_2) \neq \mathcal{I}(U_1, U_2, U_3) \neq \mathcal{I}(U_3, U_1, U_2)$ *by Corollary 5.45, which claims the equality of the corresponding weak cores to be a condition necessary for independence equivalence.*

The independence equivalence of the other permutations is given by the fact that one can easily organize these structures into a sequence such that:

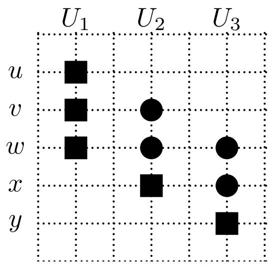
- U_1, U_2, U_3 *is the first structure in the sequence.*
- *Any other structure in the sequence may be obtained from the previous one by a constant or box-transposition.*

Then by Lemmata 6.14 and 6.30 these structures are independence equivalent. In detail: realize that U_2, U_1, U_3 *(see Figure 6.10b) is a box transposition of* U_1, U_2, U_3*. Similarly,* U_2, U_3, U_1 *is a constant transposition of* U_2, U_1, U_3 *and finally* U_3, U_2, U_1 *is a box transposition of* U_2, U_3, U_1*. See the sequence in Figure 6.10.*

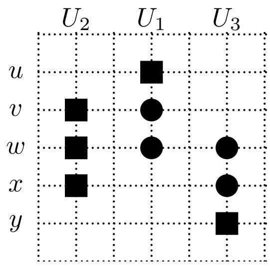
6.2 Extensions and Reductions

At the beginning of Chapter 6 we mentioned the operations that can add or remove a column to/from the structure. Since any two equivalent structures must have the same *weak core* due to Corollary 5.45, it is not so surprising that subsequent operations may only add or remove columns in a way not affecting the corresponding *weak core*. In other words: *The following operations add and remove only those columns that do not belong to the weak core of the structure.* We call them *extensions* and *reductions*.

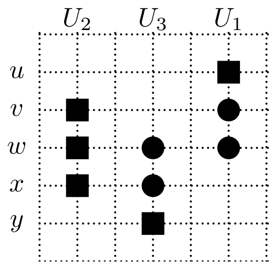
After thorough investigation of the *weak core* definition (Definition 5.37), one realizes that there are generally two types of columns not belonging to the *weak core* in a structure. They are:



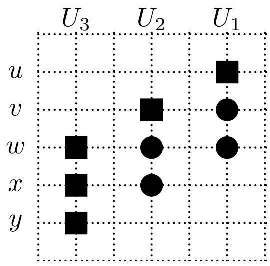
(a): $\mathcal{P} = U_1, U_2, U_3$



(b): $\mathcal{P}\sigma_1$
$\sigma_1 = (1\ 2)$

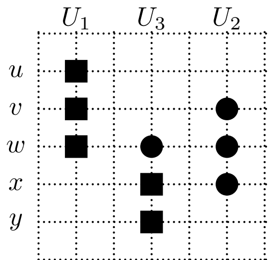


(c): $\mathcal{P}\sigma_1\sigma_2$
$\sigma_2 = (2\ 3)$

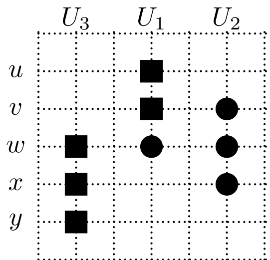


(d): $\mathcal{P}\sigma_1\sigma_2\sigma_3$
$\sigma_3 = (1\ 2)$

Figure 6.10: Another sequence of independence equivalent structures



(a)



(b)

Figure 6.11: Non-equivalent permutations of U_1, U_2, U_3

(a) trivial columns,

(b) non-trivial columns that equal S-parts of another non-trivial column.

Having this structure in mind, we introduce an operation that removes trivial columns first. Then, using Remark 6.33 and its resulting considerations, we introduce an operation able to remove any non-trivial columns not belonging to the *weak core* of the corresponding structure.

Recall that the so-called *reduced structure* was introduced in Section 5.4. Summarizing, these two operations make reduction of any structure possible. In other words: *One can convert any structure into its equivalent and reduced form with the help of the following operations.*

6.2.1 Simple reduction/extension

Definition 6.36. Simple reduction *means a change of structure* $\mathcal{P}$ *into structure* $\mathcal{P}'$ *by removing a trivial column. The structure formed by iterative applications of simple reductions from* $\mathcal{P}$ *such that all trivial columns are removed is denoted by* $red(\mathcal{P})$. *(*$\mathcal{P}' = red(\mathcal{P})$*).*

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

Remark 6.37. *Observe that* $ntriv(\mathcal{P}) = ntriv(\mathcal{P}') = ntriv(\mathcal{P}'')$ *for all* $\mathcal{P}, \mathcal{P}', \mathcal{P}''$ *mentioned in the previous definition.*

Example 6.38. *Using an example of* simple reduction *we can bring removal of the next to last column from the structure in Figure 6.11a, or similarly from the structure in Figure 6.11b. Observe that the resulting structures are box-transpositions of each other.*

Observe that adding/removing of a trivial column corresponds to the cancelation during the creation process of a *formal ratio*. Hence one can conclude with the following assertion. We state it here in the form of a labeled corollary, since we refer to it later:

Corollary 6.39. *Consider two structures* $\mathcal{P}$ *and* $\mathcal{P}'$ *such that* $\mathcal{P}'$ *can be obtained from* $\mathcal{P}$ *by a simple reduction/extension. Then these structures have the same formal ratio.*

Lemma 6.40. *Let* $\mathcal{P}$ *and* $\mathcal{P}'$ *be two structures over* N *such that* $\mathcal{P}'$ *is obtained from* $\mathcal{P}$ *by a* simple reduction/extension. *Then* $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}')$.

Proof. Suppose that a column was added to (removed from) the considered persegram in such a way that all its markers are only bullets, i.e., every added variable has to appear in some preceding column. According to Definition 3.18 of a Z-avoiding trail, no vertical connection can pass through the column without a box-marker. Therefore, the addition (removal) of such a column will bring no change in its Z-avoiding trail system, and subsequently it causes no change in its dependence or independence model. □

6.2.2 Reduction

In Remark 6.33 and the following considerations we showed a very interesting impact of a *box cycle permutation* on a structure column not belonging to the *weak core* of the structure. To recall, a column $K_i^{\mathcal{P}}$ does not belong to a structure's weak core $\mathcal{C}(\mathcal{P})$ if it is the so called S-subset of a superset column $K_{i+k}^{\mathcal{P}}$, i.e., $K_i^{\mathcal{P}} = S_{i+k}^{\mathcal{P}}$. An application of the box-cycle permutation σ between such a $K_i^{\mathcal{P}}$

and the respective $K^{\mathcal{P}}_{i+k}$ ($\sigma = (i\ i{+}1\ \ldots\ i{+}k)$) on structure $\mathcal{P}$ results in a structure whose originally non-trivial column $K^{\mathcal{P}}_i$ became trivial $K^{\mathcal{P}\sigma}_{i+1} = S^{\mathcal{P}\sigma}_{i+1}$.

Note that the following operation is not an elementary one. It can be substituted by a sequence of *constant and box transpositions* and one *simple reduction.* That is the reason why we do not classify it as one of the *IE operations* although it has a special importance in the process of creating a reduced equivalent for every structure. We call it *reduction.*

Definition 6.41. *Consider a structure $\mathcal{P}$ without trivial columns, and two indices $i \in \{1, \ldots, |\mathcal{P}|\}$ and $k \in \{0, \ldots, |\mathcal{P}| - i\}$.* Reduction *means a change of $\mathcal{P}$ into red($\mathcal{P}\sigma$) such that $\sigma = (i\ i{+}1\ \ldots\ i{+}k)$ and $K^{\mathcal{P}}_i = S^{\mathcal{P}}_{i+k}$. We say that structure red($\mathcal{P}\sigma$) is a reduction of $\mathcal{P}$.*

Lemma 6.42. *If red($\mathcal{P}\sigma$) is a reduction of $\mathcal{P}$, then one may obtain red($\mathcal{P}\sigma$) from $\mathcal{P}$ by iterative applications of IE operations.*

Proof. Since *red*() is one of the IE operations, it is enough to prove that $\mathcal{P}\sigma$ may be obtained from $\mathcal{P}$ by iterative applications of IE operations. Assume $\sigma = (i\ i{+}1\ \ldots\ i{+}k)$. Then by Definition 6.41:

$$K^{\mathcal{P}}_i = S^{\mathcal{P}}_{i+k} \tag{6.2.1}$$

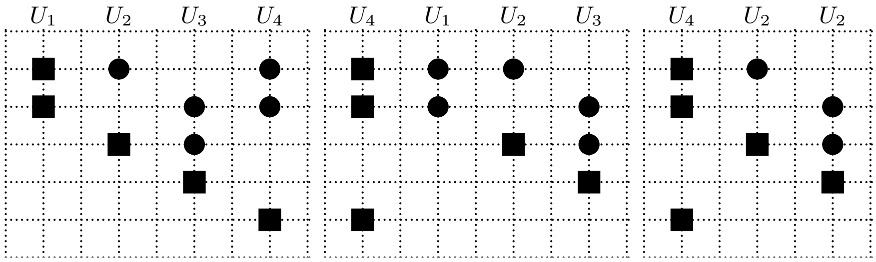
If $k = 0$, then σ is a trivial transposition, $\sigma = (i\ i)$, and the above-defined *reduction* coincides with the *simple reduction* from Definition 6.36.

Consider $k \geq 1$. Note that in that case $S^{\mathcal{P}}_i \subseteq K^{\mathcal{P}}_i = S^{\mathcal{P}}_{i+k} = K^{\mathcal{P}}_i$ by definition of S_i and (6.2.1), which guarantees that σ is a *box cycle permutation* in $\mathcal{P}$ by Definition 6.31. Recall that by Lemma 6.32, any *box cycle permutation* may be replaced by a sequence of *IE operations.* Hence the proof is finished. ☐

Regarding the fact that each of the *IE operations* preserves the *induced independence model* (see Lemata 6.14, 6.24 and 6.42), Lemma 6.42 has a very natural consequence concerning reduction and the induced independence model.

Corollary 6.43. *Let $\mathcal{P}, \mathcal{P}'$ be two structures over N such that $\mathcal{P}'$ is a reduction of $\mathcal{P}$. Then $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}')$.*

Example 6.44. *Consider structure $\mathcal{P} = U_1, \ldots, U_4$ in Figure 6.12a. Since $K^{\mathcal{P}}_1 = S^{\mathcal{P}}_4$, one can apply a reduction. Put $\sigma = (1\ 2\ 3\ 4)$ where $\mathcal{P}\sigma$ can be found in Figure 6.12b. Observe that $K^{\mathcal{P}\sigma}_2 \notin ntriv(\mathcal{P}\sigma)$ and one can apply a simple reduction and remove this column. The final reduced structure is shown in Figure 6.12c.*



(a): $\mathcal{P}$ (b): $\mathcal{P}\sigma$ (c): $red(\mathcal{P}\sigma)$

Figure 6.12: Process of a structure reduction

6.2.2.1 Reduction algorithm

Considering assertions of the previous subsection, one can remove any column not belonging to the weak core of the corresponding structure using a *reduction* operation. So if we apply reductions for long enough to a certain structure, in the end we get a reduced structure equivalent with the original one. Recall that reduction may by replaced by a sequence of *IE operations*. Hence, every structure can be converted into a reduced form that is independence equivalent with the respective structure. See a simple algorithm to do this:

Lemma 6.45. *Let $\mathcal{P}$ be a compositional model structure. Then $\mathcal{P}$ can be transformed into its reduced form $\mathcal{P}'$ by iterative applications of IE operations.*

Proof. Suppose that $\mathcal{P}$ is a non-reduced structure. Hence there exists a pair $i, k \in \mathbb{N}$ such that $K_i^{\mathcal{P}} = S_{i+k}^{\mathcal{P}}$. Use the following algorithm to convert $\mathcal{P}$ into its equivalent and reduced form :

$\mathcal{P} = red(\mathcal{P})$; {simple reduction – removing of all trivial columns}
while $\mathcal{P}$ is not reduced **do**
  Find i, k such that $K_i^{\mathcal{P}} = S_{i+k}^{\mathcal{P}}$; {such a pair exists since $\mathcal{P}$ is not reduced}
  $\sigma = (i\ i{+}1\ \ldots\ i{+}k)$;
  $\mathcal{P} = red(\mathcal{P}\sigma)$; {reduction of $\mathcal{P}$}
end while
$\mathcal{P}' = \mathcal{P}$;
return $\mathcal{P}'$;

Since *reduction* may be replaced by a sequence of *IE operations* by Lemma 6.42, the lemma is proven. □

Example 6.46. *Following the algorithm presented in the previous lemma and structure $\mathcal{P}$ shown in Figure 6.13a, one can create the following sequence of equivalent structures (see Figure 6.13) where the last one is reduced: $U_1, \ldots, U_7 \Rightarrow U_1, U_2, U_3, U_4, U_5, U_7 \Rightarrow U_1, U_3, U_4, U_5, U_7 \Rightarrow U_1, U_7, U_4, U_5 = \mathcal{P}'$. Observe that $\mathcal{P}'$ contains columns only from $\mathcal{C}(\mathcal{P}') = \mathcal{C}(\mathcal{P})$.*

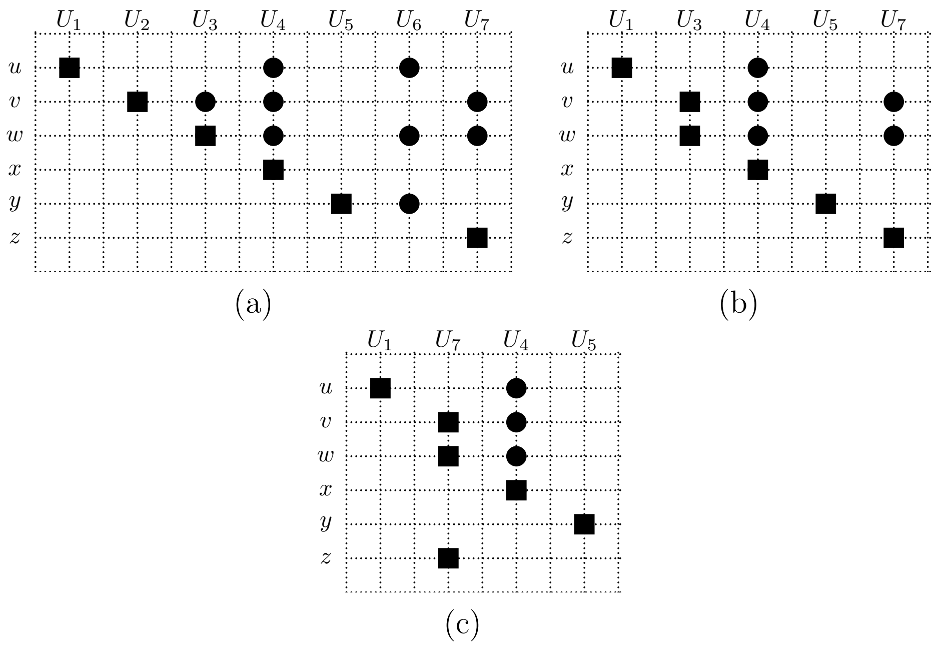


Figure 6.13: Complete reduction by reduction algorithm

6.3 IE operations

At this moment we have defined and explored all of the IE operations mentioned in the beginning of this chapter. Based on Lemmata 6.14, 6.30, and 6.40, we can make the following conclusion:

Corollary 6.47. *Each of the IE operations preserves the induced independence model* $\mathcal{I}(\mathcal{P})$*.*

Similarly, using Lemmata 6.11 and 6.29, and Corollary 6.39 it follows that:

Corollary 6.48. *Each of the IE operations preserves the induced formal ratio.*

Remark 6.49. *If an operation belongs to IE operations in an induced substructure* $\mathcal{P}[U]$ *of structure* $\mathcal{P}$*, then it is an IE operation in the structure* $\mathcal{P}$ *itself as well. Indeed, observe that whether the operation belongs to* IE operations *depends on the involved and foregoing columns – and that is exactly the definition of the* induced substructure*. See Definition 3.25.*

Chapter 7

Solution to the Equivalence problem

We have already found several operations on structures that preserve the *induced independence model*. The most elementary of them are called *IE operations*. With the help of these operations, we can iteratively change any structure and simultaneously create a class of independence equivalent structures (the so called *class of equivalence*). Since we want to find a simple characterization of equivalence, we should restrict our attention to a representative or a special subclass to simplify the following lemmata and clarify the whole theory. To do so, we will use reduced structures.

Lemma 7.1. *Let $\mathcal{P}, \mathcal{P}'$ be two reduced structures such that $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}')$. Then one can transform $\mathcal{P}'$ to have the same last column as $\mathcal{P}$ with the help of box and constant transpositions (including the division into S_i and R_i).*

Proof. Observe that $\mathcal{C}(\mathcal{P}) = \mathcal{C}(\mathcal{P}')$ by Theorem 5.44. This, together with the fact that $\mathcal{P}, \mathcal{P}'$ are reduced, implies that $|\mathcal{P}| = n = |\mathcal{P}'|$ and $\mathcal{P}' = \mathcal{P}\sigma$ for some permutation σ. Then, however, $K_n^{\mathcal{P}} = K_{n\sigma}^{\mathcal{P}\sigma}$. Recall that since both $\mathcal{P}$ and $\mathcal{P}\sigma$ are reduced, $R_n^{\mathcal{P}} \neq \emptyset$ and $K_{n\sigma}^{\mathcal{P}\sigma}$ is the only column from $\mathcal{P}\sigma$ containing variables from $R_n^{\mathcal{P}}$:

$$R_n^{\mathcal{P}} \cap K_i^{\mathcal{P}\sigma} = \emptyset \textit{ for all } i \neq n\sigma. \tag{7.0.1}$$

Therefore, $R_n^{\mathcal{P}} \subseteq R_{n\sigma}^{\mathcal{P}\sigma}$. Put $R = R_{n\sigma}^{\mathcal{P}\sigma} \setminus R_n^{\mathcal{P}}$ and $U = S_n^{\mathcal{P}}$. Observe that $R \subseteq U$. We can distinguish two cases:

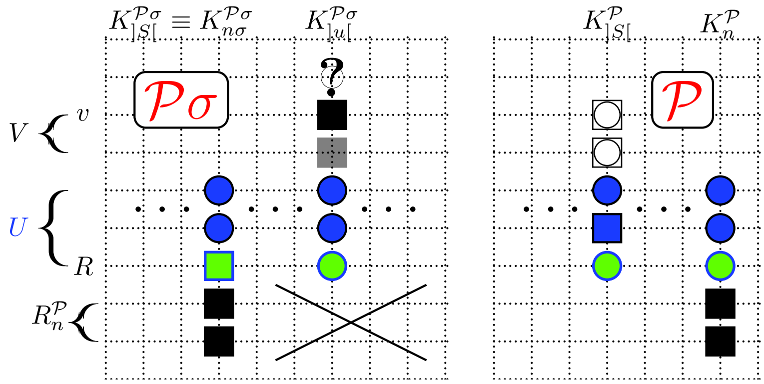
1. $R = \emptyset$
2. $R \neq \emptyset$, which means that

$$]U[_{\mathcal{P}\sigma} = n\sigma \textit{ and therefore } U \subseteq K_{]U[}^{\mathcal{P}\sigma} \tag{7.0.2}$$

If $R = \emptyset$ then $R^{\mathcal{P}\sigma}_{n\sigma} = R^{\mathcal{P}}_n$ and $R^{\mathcal{P}\sigma}_{n\sigma} \cap (K^{\mathcal{P}\sigma}_{n\sigma+1} \cup \ldots \cup K^{\mathcal{P}\sigma}_n) = \emptyset$ by (7.0.1). Put $\sigma_R = (n\ n{-}1\ \ldots\ n\sigma)$. Then σ_R is a right cycle permutation in $\mathcal{P}\sigma$ by Definition 6.19. Hence, $K^{\mathcal{P}\sigma\sigma_R}_n = K^{\mathcal{P}\sigma}_{n\sigma} = K^{\mathcal{P}}_n$ and the proof is finished using Lemma 6.20.

If $R \neq \emptyset$ then $U \in \mathcal{N}(\mathcal{P}\sigma)$ by (7.0.2). Since $\mathcal{N}(\mathcal{P}) = \mathcal{N}(\mathcal{P}\sigma)$ by assumption, then

$$U \subseteq K^{\mathcal{P}}_{]U[} \tag{7.0.3}$$



(a): part of $\mathcal{P}\sigma$ (b): part of $\mathcal{P}$

Figure 7.1: Proof of Lemma 7.1 – illustration of where $R \neq \emptyset$

Observe that $K^{\mathcal{P}}_{]U[} = U \cup V$ where $V \neq \emptyset$. (Indeed, otherwise $K^{\mathcal{P}}_{]U[} = U = S^{\mathcal{P}}_n$ would hold and $\mathcal{P}$ would not be reduced, which would contradict the assumption.) Moreover, $v \cup U \in \mathcal{N}(\mathcal{P})$ for all $v \in V$. (See Figure 7.1 where, however, columns out of focus are omitted for the sake of clarity. We use the same notation as in the proof of Lemma 5.41, where bullets in a box in one column denote the situation when we are not sure about the markers' shapes, but at least one of them is a box-marker.) Since $v \notin K^{\mathcal{P}\sigma}_{n\sigma} \equiv K^{\mathcal{P}\sigma}_{]U[}$ by definition of V, then

$$U \subseteq S^{\mathcal{P}\sigma}_{]v[} \tag{7.0.4}$$

by Lemma 5.39 applied to $\mathcal{P}\sigma$ for all $v \in V$. Choose and fix $v \in V$ such that $v \preceq_{\mathcal{P}\sigma} v'$ for all other $v' \in V$. This choice is always possible ($V \neq \emptyset$) and it guarantees that $S^{\mathcal{P}\sigma}_{]v[} \cap V = \emptyset$. It follows that

$$\begin{aligned} S^{\mathcal{P}\sigma}_{]v[} \cap K^{\mathcal{P}}_{]U[} &= S^{\mathcal{P}\sigma}_{]v[} \cap (U \cup V) \\ &= (S^{\mathcal{P}\sigma}_{]v[} \cap U) \cup (S^{\mathcal{P}\sigma}_{]v[} \cap V) \\ &= U \end{aligned} \tag{7.0.5}$$

by (7.0.4) and the choice of v. Furthermore, since $K^{\mathcal{P}\sigma}_{]U[} \setminus U = R^{\mathcal{P}}_n$ by the fact that $K^{\mathcal{P}\sigma}_{]U[} \equiv K^{\mathcal{P}\sigma}_{n\sigma}$, and we obtain that

$$(K^{\mathcal{P}\sigma}_{]U[} \setminus U) \cap K^{\mathcal{P}\sigma}_i = \emptyset \textit{ for all } i \neq]U[_{\mathcal{P}\sigma} \tag{7.0.6}$$

in particular $K^{\mathcal{P}\sigma}_{]U[} \cap K^{\mathcal{P}\sigma}_{i} \subseteq U$ for $i >]U[_{\mathcal{P}\sigma}= n\sigma$ using (7.0.1).

Considering (7.0.4) one can distinguish two cases:

1. $U = S^{\mathcal{P}\sigma}_{]v[}$

2. $U \subset S^{\mathcal{P}\sigma}_{]v[}$

However in the latter case we can apply Corollary 5.42 (using both (7.0.5) and (7.0.6)). The conclusion is that $\exists w \in N\ ]w[_{\mathcal{P}\sigma}>]U[_{\mathcal{P}\sigma}$ such that $U = S^{\mathcal{P}\sigma}_{]w[}$. Put $v = w$ in that case. Hence $U = S^{\mathcal{P}\sigma}_{]v[}$ in both cases. Realizing that $S^{\mathcal{P}\sigma}_{]U[} \subseteq U$, it follows that

$$S^{\mathcal{P}\sigma}_{]U[} \subseteq U = S^{\mathcal{P}\sigma}_{]v[} \subseteq K^{\mathcal{P}\sigma}_{]U[}. \tag{7.0.7}$$

Put $i =]U[_{\mathcal{P}\sigma}$ and $k = (]v[_{\mathcal{P}\sigma}-i)$ for easier orientation in the problem. Then $\sigma_B = (i\ i{+}1\ \ldots\ i{+}k)$ is a *box-cycle permutation* in $\mathcal{P}\sigma$ by (7.0.7), which means that

$$\begin{aligned} K^{\mathcal{P}\sigma\sigma_B}_{i+1} &= K^{\mathcal{P}\sigma}_{i} = K^{\mathcal{P}}_{n} \\ S^{\mathcal{P}\sigma\sigma_B}_{i+1} &= S^{\mathcal{P}\sigma}_{i+k} = U = S^{\mathcal{P}}_{n} \end{aligned}$$

by Remark 6.33 and the definition of i and k.

Then $R^{\mathcal{P}\sigma\sigma_B}_{i+1} \cap (K^{\mathcal{P}\sigma\sigma_B}_{i+2} \cup \ldots \cup K^{\mathcal{P}\sigma\sigma_B}_{n}) = \emptyset$ by the fact that $K^{\mathcal{P}}_{n}$ is the only column containing $R^{\mathcal{P}}_{n} = K^{\mathcal{P}}_{n} \setminus S^{\mathcal{P}}_{n}$. Therefore $\sigma_R = (n\ n{-}1\ \ldots\ i{+}1)$ is a *right cycle permutation* in $\mathcal{P}\sigma\sigma_B$. Observe that $K^{\mathcal{P}}_{n} = K^{\mathcal{P}\sigma\sigma_B\sigma_R}_{n}$.

Since both σ_B and σ_R may be internally replaced by *IE operations* through use of Lemmata 6.32 and 6.20, the proof is finished. □

The following theorem provides a complete solution of the Equivalence problem as it was defined in the beginning of this section.

Theorem 7.2. *Supposing $\mathcal{P}_A$ and $\mathcal{P}_B$ are structures over N, the following five conditions are mutually equivalent:*

(1) $\mathcal{I}(\mathcal{P}_A) = \mathcal{I}(\mathcal{P}_B)$

(2) $\mathcal{E}(\mathcal{P}_A) = \mathcal{E}(\mathcal{P}_B)$ *and* $\mathcal{F}(\mathcal{P}_A) = \mathcal{F}(\mathcal{P}_B)$

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

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

(5) formal ratios of $\mathcal{P}_A$ and $\mathcal{P}_B$ coincide

Proof. We show (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (1) and (4) $\Rightarrow$ (5) $\Rightarrow$ (3). While the implication (1) $\Rightarrow$ (2) is a combination of Corollaries 5.9 and 5.16, the implication (2) $\Rightarrow$ (3) is asserted in Corollary 5.26.

The proof of (3) $\Rightarrow$ (4) is carried out first for reduced structures by induction on $|\mathcal{P}_A|$, where $|\mathcal{P}_A| = |\mathcal{P}_B|$ by Corollary 5.52. The induction statement for $n \geq 1$ is that (3) $\Rightarrow$ (4) holds for any pair of reduced structures $\mathcal{P}_A, \mathcal{P}_B$ over N with $|\mathcal{P}_A| = |\mathcal{P}_B| \leq n$. The implication is evident for $n = 1$. Assume $n = |\mathcal{C}(\mathcal{P}_A)| \geq 2$ and that the implication holds for reduced structures $\mathcal{P}$ over N with $|\mathcal{C}(\mathcal{P})| < n$.

Observe that $\mathcal{N}(\mathcal{P}_A) = \mathcal{N}(\mathcal{P}_B)$ implies the existence of a sequence of structures $\mathcal{P}_A = \mathcal{P}_1, \ldots, \mathcal{P}_k$ such that $K_n^{\mathcal{P}_k} = K_n^{\mathcal{P}_B}$ and $\mathcal{P}_{i+1}$ is obtained from $\mathcal{P}_i$ by box or constant transposition according to Lemma 7.1 for all $i = 2, \ldots k$. Then introduce $\mathcal{P}'_k$ or $\mathcal{P}'_B$ as the induced substructures of $\mathcal{P}_k$ or $\mathcal{P}_B$, respectively, over $N \setminus R_n^{\mathcal{P}_k}$. Observe that $\mathcal{N}(\mathcal{P}'_k) = \mathcal{N}(\mathcal{P}'_B)$ by Corollaries 6.47, 5.26 and the fact that removed the same set of non-trivial set from both $\mathcal{N}(\mathcal{P}_k) = \mathcal{N}(\mathcal{P}_B)$ induced by common last column with the same division to R_i and S_i. Moreover $|\mathcal{P}'_k| = n - 1 = |\mathcal{P}'_B|$. By the induction hypothesis there exists a desired sequence of $\mathcal{P}'_k, \ldots, \mathcal{P}'_{k+m} = \mathcal{P}'_B$ where $m \geq 1$,. Introduce $\mathcal{P}_{k+i}$ as a structure over N obtained from $\mathcal{P}'_{k+i}$ by adding a column $K_n^{\mathcal{P}_k}$ at the last position for $i = 1, \ldots, m$. Since $\mathcal{P}'_{k+i}$ is a substructure of $\mathcal{P}_{k+i}$, $\mathcal{P}_{k+i+1}$ is obtained from $\mathcal{P}_{k+i}$ by *box and constant transpositions* for $i = 1, \ldots, m-1$ according to Remark 6.49.

If $\mathcal{P}_A$ or $\mathcal{P}_B$ is not reduced, then one may easily create sequences of structures $\mathcal{P}_A = \mathcal{P}_{A_1}, \ldots, \mathcal{P}_{A_l}$ and $\mathcal{P}_B = \mathcal{P}_{B_1}, \ldots, \mathcal{P}_{B_n}$ where both $\mathcal{P}_{A_l}$ and $\mathcal{P}_{B_n}$ are reduced and $\mathcal{P}_{A_{i+1}}$ or $\mathcal{P}_{B_{j+1}}$ is obtained from $\mathcal{P}_{A_i}$ or $\mathcal{P}_{B_j}$, respectively, by an IE operation for $i = 1, \ldots, l-1$ and $j = 1, \ldots, n-1$ by Lemma 6.42. Then Case I occurs for the pair $(\mathcal{P}_{A_l}, \mathcal{P}_{B_n})$, which has already been dealt with. This concludes the induction step.

The proof of (4) $\Rightarrow$ (1) follows from repetitive applications of Corollary 6.47, then (4) $\Leftrightarrow$ (3); the proof of (4) $\Rightarrow$ (5) follows from repetitive applications of Corollary 6.48.

To finish the proof of the theorem, let us show the validity of the last non-proven implication (5) $\Rightarrow$ (3). Recall that by Lemma 6.42, any structure may be transformed into its equivalent and reduced form by repetitive applications of IE operations. Using this, denote by $\mathcal{P}'_A$ or $\mathcal{P}'_B$ an equivalent and reduced form of $\mathcal{P}_A$ or $\mathcal{P}_B$, respectively. Since each of the IE operations preserves the corresponding formal ratio by Corollary 6.48, the formal ratios of both $\mathcal{P}'_A$ and $\mathcal{P}'_B$ coincide. Now we can apply Lemma 5.58 which asserts that for two reduced structures with equal formal ratio, their non-trivial sets coincide ($\mathcal{N}(\mathcal{P}'_A) = \mathcal{N}(\mathcal{P}'_B)$), which finishes the proof. $\square$

It has already been mentioned that an arbitrary permutation may be expressed as a product of transpositions [4]. Imagine a slightly modified algorithm from

Lemma 6.45 where columns out of respective weak core are not removed by simple reduction, but they are moved at the end of the structure using right cycle permutation. Then, we can transform any structure $\mathcal{P}$ into the form where first $|\mathcal{C}(\mathcal{P})|$ columns represent its reduced form and the rest are trivial columns. Following the induction in the (3) $\Rightarrow$ (4) part of the previous proof, (as well as the notion of the previous theorem itself), one can conclude:

Corollary 7.3. *A structure and its permutation are independence equivalent if and only if one can be obtained from the other by repetitive applications of box and constant transpositions.*

Chapter 8

Equivalent structures and generating sequences

Considering a generating sequence, its structure is inherently tied to the represented multidimensional probability distribution – the compositional model. Since the *operator of composition* is neither commutative nor associative, an arbitrary change of the generating sequence may result in a change of the represented compositional model.

Let us generalize IE operations in a way that instead of changing structure, we change (permute, extend/reduce) the respective generating sequence, i.e., operations originally introduced for columns in the compositional model structures will also be used for distributions in the generating sequences.

Example 8.1. *For example, let π_2, π_1 be a sequence that arose from a generating sequence π_1, π_2 by transposition* (1 2)*, which is constant in the structure of π_1, π_2. Then we will say that π_2, π_1 is a constant transposition of sequence π_1, π_2.*

It has been proven at the end of the previous chapter that two structures are equivalent if (and only if) there exists a sequence of IE operations such that one can transform one structure to the other one using these operations. Now, if we consider generating sequences instead of structures, the obvious question arises: If there are two arbitrary generating sequences such that one is obtained from the other one by iterative applications of IE operations, are they equivalent? (i.e., do they represent the same multidimensional probability distribution?)

The answer for this question is generally negative:

Remark 8.2. *Note that, generally, box transposition π_2, π_1 of a generating sequence π_1, π_2 does not have to be a* generating *sequence (i.e., $\pi_2 \rhd \pi_1$ may be undefined even if $\pi_1 \rhd \pi_2$ is defined).*

IE operations applied on a generating sequence consequently change the entire compositional model – the represented multidimensional probability distribution.

So it is appropriate to examine how the resulting model changes, and under which additional conditions the resulting model is the same.

In order to simplify the following lemmata, we will work with the model whose generating sequence consists only of three distributions $\pi_1(U_1), \pi_2(U_2), \pi_3(U_3)$, and we will apply IE operations on π_2 and π_3. This simplification is not in any way at the expense of generality. Indeed, realize that $\pi_1(U_1)$ can be internally composed from several distributions and $\pi_1(U_1), \pi_2(U_2), \pi_3(U_3)$ can represent the beginning of a much longer sequence, and as such it coincides with its corresponding marginal distribution by Assertion 3.7.

8.1 Constant transposition

Lemma 8.3. *Consider three distributions $\pi_1(U_1), \pi_2(U_2)$, and $\pi_3(U_3)$. If $(2\ 3)$ is a constant transposition in structure U_1, U_2, U_3, then*

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_1 \triangleright \pi_3 \triangleright \pi_2. \tag{8.1.1}$$

The proof of a similar assertion can be found in [24] with the difference that a constant transposition is not defined there due to different approach. However, for the sake of this text's completeness , we present the proof here as well.

Proof. First, let us show that the left side expression in (8.1.1) is not defined *iff* the right side of the formula is not defined. Put $\mathcal{P} = U_1, U_2, U_3$ and $\sigma = (2\ 3)$. From the definition of operator $\triangleright$ we know that $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is not defined *iff*

$$\pi_1^{\downarrow S_2^{\mathcal{P}}} \not\ll \pi_2^{\downarrow S_2^{\mathcal{P}}}$$

or

$$(\pi_1 \triangleright \pi_2)^{\downarrow S_3^{\mathcal{P}}} \not\ll \pi_3^{\downarrow S_3^{\mathcal{P}}}.$$

Analogously, considering the fact that structure of π_1, π_3, π_2 coincides with $\mathcal{P}\sigma$, $\pi_1 \triangleright \pi_3 \triangleright \pi_2$ is not defined *iff*

$$\pi_1^{\downarrow S_2^{\mathcal{P}\sigma}} \not\ll \pi_3^{\downarrow S_2^{\mathcal{P}\sigma}}$$

or

$$(\pi_1 \triangleright \pi_3)^{\downarrow S_3^{\mathcal{P}\sigma}} \not\ll \pi_2^{\downarrow S_3^{\mathcal{P}\sigma}}.$$

Under the given assumption that $(2\ 3)$ is a *constant transposition* in $\mathcal{P} = U_1, U_2, U_3$, these two conditions coincide because:

$$S_2^{\mathcal{P}} = S_3^{\mathcal{P}\sigma} \text{ and } S_3^{\mathcal{P}} = S_2^{\mathcal{P}\sigma} \tag{8.1.2}$$

by Lemma 6.10 and

$$
\begin{aligned}
(\pi_1 \rhd \pi_2)^{\downarrow S_3^{\mathcal{P}}} = (\pi_1 \rhd \pi_2)^{\downarrow S_2^{\mathcal{P}\sigma}} &= \pi_1^{\downarrow S_2^{\mathcal{P}\sigma}} \\
(\pi_1 \rhd \pi_3)^{\downarrow S_3^{\mathcal{P}\sigma}} = (\pi_1 \rhd \pi_3)^{\downarrow S_2^{\mathcal{P}}} &= \pi_1^{\downarrow S_2^{\mathcal{P}}}
\end{aligned}
$$

where the last equation in each row is guaranteed by Assertion 3.7 and the fact that both $S_2^{\mathcal{P}}$ and $S_2^{\mathcal{P}\sigma}$ are subsets of $K_1^{\mathcal{P}} = U_1$. Indeed, realize that in the case of $S_2^{\mathcal{P}\sigma}$ it holds that $S_2^{\mathcal{P}\sigma} \subseteq K_1^{\mathcal{P}\sigma} = K_1^{\mathcal{P}}$ by the nature of transposition $\sigma = (2\ 3)$.

Now, let us assume that both expressions in (8.1.1) are defined. Because of (8.1.2), the expressions

$$
\begin{aligned}
\pi_1 \rhd \pi_2 \rhd \pi_3 &= \frac{\pi_1\pi_2\pi_3}{\pi_2^{\downarrow S_2^{\mathcal{P}}}\pi_3^{\downarrow S_3^{\mathcal{P}}}} \\
\pi_1 \rhd \pi_3 \rhd \pi_2 &= \frac{\pi_1\pi_2\pi_3}{\pi_3^{\downarrow S_2^{\mathcal{P}\sigma}}\pi_2^{\downarrow S_3^{\mathcal{P}\sigma}}}
\end{aligned}
$$

are equivalent to each other, which finishes the proof.

□

8.2 Box transposition

Lemma 8.4. *Consider three distributions $\pi_1(U_1)$, $\pi_2(U_2)$ and $\pi_3(U_3)$ such that π_2 and π_3 are consistent. If $(2\ 3)$ is a box transposition in structure U_1, U_2, U_3, then*

$$\pi_1 \rhd \pi_2 \rhd \pi_3 = \pi_1 \rhd \pi_3 \rhd \pi_2. \tag{8.2.1}$$

Proof. Let us start, again, by showing that, under the given assumption, $\pi_1 \rhd \pi_2 \rhd \pi_3$ is undefined *iff* $\pi_1 \rhd \pi_3 \rhd \pi_2$ is undefined. Put $\mathcal{P} = U_1, U_2, U_3$ and $\sigma = (2\ 3)$. Similarly as in the proof of Lemma 8.3 $\pi_1 \rhd \pi_2 \rhd \pi_3$ is not defined *iff*

$$\pi_1^{\downarrow S_2^{\mathcal{P}}} \not\ll \pi_2^{\downarrow S_2^{\mathcal{P}}} \tag{8.2.2}$$

or

$$(\pi_1 \rhd \pi_2)^{\downarrow S_3^{\mathcal{P}}} \not\ll \pi_3^{\downarrow S_3^{\mathcal{P}}}. \tag{8.2.3}$$

Analogously, (realizing the fact that $\mathcal{P}\sigma$ represents the structure of π_1, π_3, π_2) $\pi_1 \rhd \pi_3 \rhd \pi_2$ is not defined *iff*

$$\pi_1^{\downarrow S_2^{\mathcal{P}\sigma}} \not\ll \pi_3^{\downarrow S_2^{\mathcal{P}\sigma}} \tag{8.2.4}$$

or

$$(\pi_1 \rhd \pi_3)^{\downarrow S_3^{\mathcal{P}\sigma}} \not\ll \pi_2^{\downarrow S_3^{\mathcal{P}\sigma}}. \tag{8.2.5}$$

Moreover, since (2 3) is a *box transposition* in $\mathcal{P} = U_1, U_2, U_3$ then

$$S_2^{\mathcal{P}} = S_2^{\mathcal{P}\sigma} \text{ and } S_3^{\mathcal{P}} = S_3^{\mathcal{P}\sigma}. \tag{8.2.6}$$

Now, since π_2 and π_3 are consistent by our assumption, (8.2.2) and (8.2.4) coincide. To prove the same for (8.2.3) and (8.2.5), put $S = S_3^{\mathcal{P}}$. Definition 6.22 of a box transposition induces that $S_2^{\mathcal{P}} \subseteq S_3^{\mathcal{P}} \subseteq U_2$. Considering the fact $S_2^{\mathcal{P}} = (U_1 \cap U_2)$, this implies not only

$$(U_1 \cap U_2) \subseteq S \subseteq (U_1 \cup U_2) \tag{8.2.7}$$

(where the second inclusion is guaranteed by the fact that $U_2 \subseteq (U_1 \cup U_2)$) but also

$$(S \cap U_2) = (S \cap U_3) = S \tag{8.2.8}$$

(where the second equation is guaranteed by the choice of S).

Then

$$\begin{aligned} (\pi_1 \rhd \pi_2)^{\downarrow S} &= \pi_1^{\downarrow S\cap U_1} \rhd \pi_2^{\downarrow S\cap U_2} \\ &= \pi_1^{\downarrow S\cap U_1} \rhd \pi_3^{\downarrow S\cap U_3} \\ &= (\pi_1 \rhd \pi_3)^{\downarrow S} \end{aligned}$$

where the first equation is guaranteed by (8.2.7) and Assertion 3.6, the second one by (8.2.8) and the assumption of consistency on π_2 and π_3, and finally the last one by Assertion 3.6 again. Thus we got that (8.2.3) is equivalent to (8.2.5) and both conditions coincide.

Let us now assume that both expressions in Formula (8.2.1) are defined. Because of (8.2.6) and the fact that π_2 and π_3 are consistent, the expressions

$$\pi_1 \rhd \pi_2 \rhd \pi_3 = \frac{\pi_1\pi_2\pi_3}{\pi_2^{\downarrow S_2^{\mathcal{P}}}\,\pi_3^{\downarrow S_3^{\mathcal{P}}}},$$
$$\pi_1 \rhd \pi_3 \rhd \pi_2 = \frac{\pi_1\pi_3\pi_2}{\pi_3^{\downarrow S_2^{\mathcal{P}\sigma}}\,\pi_2^{\downarrow S_2^{\mathcal{P}\sigma}}}$$

are mutually equivalent, which finishes the proof. □

8.3 Simple extension/reduction

Lemma 8.5. *Consider three distributions $\pi_1(U_1), \pi_2(U_2)$, and $\pi_3(U_3)$ such that $\pi_1 \rhd \pi_2$ is defined. If U_2 is a trivial column in structure U_1, U_2, U_3 then*

$$\pi_1 \rhd \pi_2 \rhd \pi_3 = \pi_1 \rhd \pi_3. \tag{8.3.1}$$

Proof. First, we start by showing that, under the given assumptions, $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is undefined *iff* $\pi_1 \triangleright \pi_3$ is undefined. Considering the fact that $\pi_1 \triangleright \pi_2$ is defined by our assumption, $\pi_1 \triangleright \pi_2 \triangleright \pi_3$ is not defined *iff*

$$(\pi_1 \triangleright \pi_2)^{\downarrow (U_1 \cup U_2) \cap U_3} \not\ll \pi_3^{\downarrow (U_1 \cup U_2) \cap U_3}.$$

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

$$\pi_1^{\downarrow U_1 \cap U_3} \not\ll \pi_3^{\downarrow U_1 \cap U_3}.$$

Under the given assumption that U_2 is a trivial column in U_1, U_2, U_3, these two conditions coincide because $U_2 \subseteq U_1$ by definition of non-triviality – this induces that

$$(U_1 \cup U_2) \cap U_3 = U_1 \cap U_3 \tag{8.3.2}$$

and the fact that

$$(\pi_1 \triangleright \pi_2)^{\downarrow U_1 \cap U_3} = \pi_1^{\downarrow U_1 \cap U_3} \tag{8.3.3}$$

which is illustrated by Assertion 3.7 and its statement of $(\pi_1 \triangleright \pi_2)^{\downarrow U_1} = \pi_1^{\downarrow U_1}$.

Now, assume that both expressions in Formula (8.3.1) are defined. Because of (8.3.2) and the fact that $U_2 \cap U_1 = U_2$ (i.e., $\pi_2^{\downarrow U_1 \cap U_2} = \pi_2$), the expressions

$$\begin{aligned}
\pi_1 \triangleright \pi_2 \triangleright \pi_3 &= \frac{\pi_1 \pi_2 \pi_3}{\pi_2^{\downarrow U_1 \cap U_2} \pi_3^{\downarrow (U_1 \cup U_2) \cap U_3}}, \\
\pi_1 \triangleright \pi_3 &= \frac{\pi_1 \pi_3}{\pi_3^{\downarrow U_1 \cap U_3}}
\end{aligned}$$

are mutually equivalent, which finishes the proof. □

8.4 Other properties

Consider a generating sequence, which represents a multidimensional probability distribution – π – the so-called compositional model. Denote by $\mathcal{P}$ the structure of the corresponding generating sequence of compositional model π. It was proven in Chapter 7 that every structure $\mathcal{P}'$ equivalent with $\mathcal{P}$ can be obtained from $\mathcal{P}$ by iterative applications of IE operations. In the previous three lemmata we added a link between IE operations and their respective compositional models – generating sequences.

If one focuses on the relationships between compositional models that differ by one of the IE operations (i.e., one arose from the other one by applying one of IE operations on its generating sequence), the following conclusions can be drawn about their equality or inequality:

Two sequences of probability distributions represent the same multidimensional probability distributions if one is a generating one and:

a) the other one is its *constant transposition*, or

b) the other one is its *box transposition* and the respective (the permuted) probability distributions are consistent, or

c) the other one arose from the original one by a *simple reduction*. (The assumption of Lemma 8.5 – i.e., $\pi_1^{\downarrow U_2} \ll \pi_2$ in that case – is guaranteed by the fact that we consider a *generating sequence*, i.e., all the operators of composition are defined when applied on the corresponding generating sequence.)

Remark 8.6. *One may conclude with similar assertions for composed operations (*left and right cycle permutation, box cycle permutation, and reduction *as well). It has been proven that one may replace any of these composed operations by a sequence of IE operations. It is therefore sufficient to answer the question, "What are these sequences composed of?" Recall that the answer to this question can be found in Lemmata 6.17, 6.20,6.32, and 6.42 (or in their proofs in the cases of the box cycle permutation and reduction).*

Two sequences of probability distributions represent the same multidimensional probability distribution if one is a generating sequence and

i. the other one is its left or right cycle permutation*, or*

ii. the other one is its box cycle permutation *with $\sigma = (i\ i{+}1\ \dots\ i{+}k)$, and the i-th and $(i{+}k)$-th distributions are consistent.*

Since a reduction can be replaced by a box-cycle transposition and simple reduction, one can conclude:

iii. A generating sequence represents the same multidimensional distribution as its reduction*, where the i-th distribution was removed and the $(i{+}k)$-th one was moved into its position (i-th) simultaneously, if the respective distributions were consistent.*

Now, focus only on permutations of generating sequences.

Definition 8.7. *For a generating sequence, we understand its* equivalent permutation *to be a generating sequence composed from the same distributions (in a different order) representing the same multidimensional probability distribution.*

On the other hand, we understand an IE permutation *to be a sequence of the same distributions inducing independence equivalent structures.*

Observe that while for a generating sequence, its equivalent permutation is a generating sequence as well, its IE permutation is a common sequence of distributions (not a generating sequence, generally). So for a sequence $\pi_1, \pi_2, \dots, \pi_n$ and its permutation $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_n}$ such that the respective structures are independence equivalent, the expression $\pi_{i_1} \triangleright \pi_{i_2} \triangleright \ldots \triangleright \pi_{i_n}$ does not have to be defined. Thus a sequence $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_n}$ (an IE permutation of $\pi_1, \pi_2, \ldots, \pi_n$) does not have to be a generating sequence.

Recall that an arbitrary permutation can be expressed as a product of transpositions [4]. In the case of IE permutations, they can be expressed as a product of constant and box transpositions (see Corollary 7.3). Regarding equality of the represented multidimensional distributions, while a constant transposition needs no additional requirement by Lemma 8.3, a box transposition demands consistency of the involved distributions.

Hence, in the case of IE permutations, they represent the same multidimensional distribution as the original one if certain pairs of low-dimensional distributions in the sequence are consistent. Thus, considering pairwise consistency, one can easily prove the following simple assertion:

Lemma 8.8. *Let $\pi_1, \pi_2, \ldots, \pi_n$ be a sequence of pairwise consistent distributions. Then for its arbitrary IE permutation $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_n}$ holds that*

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

Proof. Denote by $\mathcal{P}$ the structure of $\pi_1, \pi_2, \ldots, \pi_n$. Let $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_n}$ arose from $\pi_1, \pi_2, \ldots, \pi_n$ by permutation σ. Observe that then the permutation $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_n}$ has a structure $\mathcal{P}\sigma$ and that both $\mathcal{P}$ and $\mathcal{P}\sigma$ are independence equivalent by the lemma assumption. Then σ can be expressed as a product of *constant* and *box transpositions* by Corollary 7.3. The proof follows from repetitive applications of Lemmata 8.3 and 8.4.

□

Remark 8.9. *Note that pairwise consistency is an unnecessarily strong condition in the previous lemma. Recall that the only IE operation that requires the consistency of involved distributions is a box transposition (Lemma 8.4). Hence it would be sufficient to demand consistency of those pairs of distributions that are transposed just by a box transposition. More about this problem can be found in Sections 9.3.1 and 9.3.2.*

Considering the fact that distributions in a perfect sequence are pairwise consistent (see Assertion 3.9), one can apply the previous lemma and conclude that any IE permutation of a perfect sequence is its equivalent permutation as well. Moreover, every distribution in the sequence is its marginal by Definition 3.8, and any of its IE permutations is perfect as well:

Corollary 8.10. *Let $\pi_1, \pi_2, \ldots, \pi_n$ be a perfect sequence. Then its IE permutation for permutation $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_n}$ is perfect and represents the same multidimensional distribution (i.e., it is its equivalent permutation).*

Chapter 9

Conditioning

Let us introduce the problem of conditioning a distribution that is represented in the form of a compositional model. Generally, the conditioning process can be viewed as a transformation of one probability distribution into another. When representing a distribution in the form of a compositional model, we understand conditioning as a transformation of its generating sequence into another one – preferably with the smallest number of *local changes* (inspired by Lauritzen-Spiegelhalter's local computations [36]). By a local change we understand either a change of just one distribution in the corresponding generating sequence (its recalculation), or a permutation of the generating sequence.

The conditioning problem was briefly discussed in [24]. In the same publication, the following example is also given. It illustrates conditioning in a simple distribution $\pi(u, v, w)$ represented by a compositional model with a generating sequence $\pi_1(u, v), \pi_2(v, w)$. A theorem was also stated for how to deal with the case when a conditioning variable appears in the argument of the first distribution of the corresponding generating sequence – Assertion 9.1 in here – as well as the concept of *flexible sequences*. We further investigate flexible sequences in this text – primarily using new evidence about IE permutations of generating sequences.

9.1 Illustrating example

To illustrate the problems with computation of a conditional distribution given by one variable, we repeat the following example from [24]. Consider a generating sequence of two distributions $\pi_1(u, v), \pi_2(v, w)$ and compute three conditional distributions for some $x_u \in \mathbf{X}_u, x_v \in \mathbf{X}_v, x_w \in \mathbf{X}_w$:

$$(\pi_1 \triangleright \pi_2)(v, w|u = x_u),$$
$$(\pi_1 \triangleright \pi_2)(u, w|v = x_v),$$
$$(\pi_1 \triangleright \pi_2)(u, v|w = x_w).$$

By definition of a conditional distribution, let us express

$$\begin{aligned}(\pi_1 \triangleright \pi_2)(u = x_u, v, w) &= \pi_1(u = x_u, v)\pi_2(w|v)\\ &= \pi_1(u = x_u)\pi_1(v|u = x_u)\pi_2(w|v)\\ &= \pi_1(u = x_u)(\pi_1(v|u = x_u) \triangleright \pi_2(v, w)),\end{aligned}$$

which implies (using the fact that π_1 is a marginal of $\pi_1 \triangleright \pi_2$ by Assertion 3.7) that the respective conditional distribution may be expressed as a compositional model with a generating sequence similar to the original one. The involved distribution was slightly changed:

$$(\pi_1 \triangleright \pi_2)(v, w|u = x_u) = \pi_1(v|u = x_u) \triangleright \pi_2(v, w).$$

Analogously, we can express

$$\begin{aligned}(\pi_1 \triangleright \pi_2)(u, v = x_v, w) &= \pi_1(u, v = x_v)\pi_2(w|v = x_v)\\ &= \pi_1(v = x_v)\pi_1(u|v = x_v)\pi_2(w|v = x_v)\\ &= \pi_1(v = x_v)(\pi_1(u|v = x_v) \triangleright \pi_2(w|v = x_v)),\end{aligned}$$

where the last equation is guaranteed by the fact that distributions $\pi_1(u|v = x_v)$ and $\pi_2(w|v = x_v)$ are one-dimensional distributions defined on different variables. That is why the operator of composition degenerates to a product of the respective distributions. (Note that this corresponds to the fact $u \perp\!\!\!\perp w|v[\mathcal{P}]$ where $\mathcal{P} = \{u, v\}, \{v, w\}$ is the structure of the respective compositional model.) However, (using the fact that π_1 is marginal of $\pi_1 \triangleright \pi_2$ by Assertion 3.7 as well), the second considered conditional distribution may be expressed as a compositional model that is transformed from $\pi_1 \triangleright \pi_2$ by "local operations" only.

$$(\pi_1 \triangleright \pi_2)(u, w|v = x_v) = \pi_1(u|v = x_v) \triangleright \pi_2(w|v = x_v).$$

In contrast to the previous two cases, the conditional distribution $(\pi_1 \triangleright \pi_2)(u, v|w = x_w)$ cannot be expressed as a compositional model of $\pi_1(u, v)$ in the first position and some other distribution. In this case, namely

$$(\pi_1 \triangleright \pi_2)(u, v, w = x_w) = \pi_1(u, v)\pi_2(w = x_w|v).$$

Note that $\pi_2(w = x_w|v)$ is a function of variable v and as such it is generally not a probability distribution. Assume the consistency of π_1 and π_2 for now. Moreover, realizing the fact that the transposition of the first two distributions is always either a box or constant transposition (the specific situation depends on whether the intersection of the sets of their arguments is nonempty – i.e., (1 2) is a box transposition in this case), then

$$\pi_1 \triangleright \pi_2 = \pi_2 \triangleright \pi_1 \tag{9.1.1}$$

by either Lemmata 8.3 or 8.4 and hence $(\pi_1 \triangleright \pi_2)^{\downarrow\{v,w\}} = \pi_2$ by Assertion 3.7. It implies that

$$(\pi_1 \triangleright \pi_2)(u, v|w = x_w) = \pi_1(u, v)\frac{\pi_2(w = x_w|v)}{\pi_2(w = x_w)}.$$

Observe that the ratio $\pi_2(w = x_w|v)/\pi_2(w = x_w)$ (as a function of v) can achieve values greater than one. Therefore $(\pi_1 \triangleright \pi_2)(u, v|w = x_w)$ cannot be expressed as a composition of $\pi_1(u, v)$ and another distribution.

Nevertheless, one can easily see that (9.1.1) holds if π_1, π_2 are consistent, and one can show it similarly for $(\pi_1 \triangleright \pi_2)(v, w|u = x_u)$ that

$$(\pi_1 \triangleright \pi_2)(u, v|w = x_w) = \pi_2(v|w = x_w) \triangleright \pi_1(u, v).$$

This example illustrates a general fact: *computation of a conditional distribution $\pi(\cdot|u = x_u)$ (for $\pi = \pi_1 \triangleright \ldots \triangleright \pi_n$) is easy only if the conditioning variable u appears among the arguments of the first distribution π_1.* In the following assertion, originally published in [24], one can see the process of such a conditioning.

Assertion 9.1. *Let $\pi_1, \pi_2, \ldots, \pi_n$ be a generating sequence with structure $\mathcal{P}$ over N and $u \in K_1^{\mathcal{P}}$. Then, for any value x_u of the variable u for which $\pi_1(u = x_u) > 0$, the following formula holds*

$$(\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n)\big((K_1^{\mathcal{P}} \cup \ldots \cup K_n^{\mathcal{P}}) \setminus u|u = x_u\big) = \kappa_1 \triangleright \kappa_2 \triangleright \ldots \triangleright \kappa_n,$$

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

$$\kappa_i(K_i^{\mathcal{P}} \setminus u) = \begin{cases} \pi_i(K_i^{\mathcal{P}}) & \text{if } u \notin K_i^{\mathcal{P}}, \\ \pi_i(K_i^{\mathcal{P}}|u = x_u) & \text{if } u \in K_i^{\mathcal{P}}. \end{cases}$$

9.2 Flexible sequence

Assume a generating sequence $\pi_1, \pi_2, \ldots, \pi_6$ with structure $\mathcal{P} = U_1, U_2, \ldots, U_6$ and permutation $\sigma \in T_6$. Let, for example, $\sigma = (1\ 2\ 4\ 5)$. Then $\mathcal{P}\sigma =$

$U_5, U_1, U_3, U_2, U_4, U_6$. Similarly, in case of the sequence $\pi_1, \pi_2, \ldots, \pi_6$ we logically assume that its σ permutation coincides with $\pi_5, \pi_1, \pi_3, \pi_2, \pi_4, \pi_6$. Observe that the interpretation of the permutation σ is that the 1st distribution moves to the 2nd position, the 2nd one to the 4th position, etc. Then, when one writes the final permutated sequence, one asks: Which distributions moved to the 1st position, to the 2nd one, etc.? The answer is, the $1\sigma^{-1}$th one, the $2\sigma^{-1}$th one, etc. Hence the σ permutation of $\pi_1, \pi_2, \ldots, \pi_6$ can be written also as $\pi_{1\sigma^{-1}}, \pi_{2\sigma^{-1}}, \ldots, \pi_{6\sigma^{-1}}$. This notation will be used several times in the following.

In light of Assertion 9.1, it seems reasonable to study the question it raises: "In which case may we reorder the generating sequence in such a way that the desired variable appears among the arguments of its first probability distribution?" To specify the problem, we define the so-called *flexible sequences.* Note that the concept of generating sequence *flexibility* was originally introduced in [24] during studies of a stronger property – the so-called *decomposability.*

Definition 9.2. *A generating sequence $\pi_1, \pi_2, \ldots, \pi_n$ with structure $\mathcal{P}$ is called* flexible *if, for all $u \in K_1^{\mathcal{P}} \cup \ldots \cup K_n^{\mathcal{P}}$, there exists a permutation $\sigma \in T_n$ such that $u \in K_1^{\mathcal{P}\sigma}$ and*

$$\pi_{1\sigma^{-1}} \triangleright \pi_{2\sigma^{-1}} \triangleright \ldots \triangleright \pi_{n\sigma^{-1}} = \pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n.$$

In other words, flexible sequences are those which can be reordered in many ways so that each variable can appear among the arguments of the first distribution. However, it does not mean that each distribution appears at the beginning of the generating sequence. If this were the case, a flexible sequence would be a subclass of perfect sequences (since each distribution would be a marginal of the composed distribution by Assertion 3.7).

Observe that the problem of conditioning by a variable turns into a problem of *flexibility* in light of Assertion 9.1. We avoid the study of other possible conditioning algorithms and we require that the conditional variable is among the arguments of the first distribution in the generating sequence.

Recall that we have already dealt with permutations in this text. By Corollary 5.52: *Two reduced and equivalent structures are permutations of each other.* Since each independence equivalent permutation of a structure may be equivalently obtained by iterative applications of IE operations (specifically using only box and constant transpositions by Corollary 7.3), we may use the knowledge from Chapter 8, where a connection between a permutation of a structure and its corresponding generating sequence was introduced. Consistency of some corresponding pairs of distributions in a generating sequence ensures that both generating sequences represent the same multidimensional probability distribution.

To illustrate this consequence, recall Lemma 8.8: *Let $\pi_1, \pi_2, \ldots, \pi_n$ be a sequence of pairwise consistent distributions. Then, for its IE σ-permutation holds*

$$\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n = \pi_{1\sigma^{-1}} \triangleright \pi_{2\sigma^{-1}} \triangleright \ldots \triangleright \pi_{n\sigma^{-1}}.$$

In the light of this knowledge, we distinguish between two different approaches to flexible sequences:

I. In our first approach, we restrict ourselves only to those flexible sequences whose all appropriate permutations induce structures that are mutually independence equivalent.

II. In our second approach, we focus on permutations with non-equivalent structures as well.

9.3 Flexible structure

To explore our first approach to flexible sequences from above (where we restrict ourselves to those flexible sequences that simultaneously induce independence equivalent structures only), we will define the so-called *flexible structure*. But as we shall see, structure flexibility coincides with another well-known property called *running intersection property* sometimes abbreviated as *RIP*.

Definition 9.3. *Let $\mathcal{P}$ be a structure over N. $\mathcal{P}$ is* flexible *if $\forall u \in N$ exists a permutation σ such that $u \in K_1^{\mathcal{P}\sigma}$ and $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}\sigma)$.*

The problem of independence equivalent structures was well examined in Chapter 7. Recall that several characterizations of independence equivalence exist. In this context let us highlight the so-called *F-condition set* – one of independence equivalence invariants (Corollary 5.16). Besides, they form, along with the *Connection set*, $\mathcal{E}(\mathcal{P})$ one of direct characterizations of independence equivalence. See Theorem 7.2.

Non-emptiness of the induced F-condition set has a major impact on the flexibility of the respective structure. Judge for yourself: recall that that the disjoint triplet $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$ if all of the following conditions hold:

(a) $\{u, v\} \prec_{\mathcal{P}} w$,

(b) $\{u, v\} \leftrightarrow_{\mathcal{P}} w$,

(c) $u \not\leftrightarrow_{\mathcal{P}} v$.

Hence, if $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$ then by (a) and the fact that $\mathcal{F}(\mathcal{P})$ is an independence class invariant, it follows that there is no structure $\mathcal{P}'$ equivalent with $\mathcal{P}$ such that $w \in K_1^{\mathcal{P}'}$. Indeed, there always have to be some foregoing columns introducing u and v first in $\mathcal{P}'$. Moreover, considering the condition (c), it implies that if $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$ then variable w cannot appear earlier than in the third column of any independence equivalent structure for the first time. Therefore the emptiness of $\mathcal{F}(\mathcal{P})$ is a necessary condition for the flexibility of $\mathcal{P}$:

Corollary 9.4. *If a structure $\mathcal{P}$ is flexible, then $\mathcal{F}(\mathcal{P}) = \emptyset$.*

9.3.1 Column covering

In this section we shall see that the emptiness of $\mathcal{F}(\mathcal{P})$ is not only necessary but also a sufficient condition for the flexibility of the respective structure $\mathcal{P}$. To prove this, we need to employ the specific consequence of $\mathcal{F}(\mathcal{P}) = \emptyset$. This consequence deals with specific structure's behavior regarding any non-trivial column – the so called *column covering.*

Definition 9.5. *Consider structure $\mathcal{P}$ over N and variable $u \in N$. Column $K^{\mathcal{P}}_{]u[}$ is* covered *in structure $\mathcal{P}$ if either $]u[_{\mathcal{P}}= 1$ or if there exists variable $v \in N$: $v \prec_{\mathcal{P}} u$ such that $S^{\mathcal{P}}_{]u[} \subseteq K^{\mathcal{P}}_{]v[}$. We say that $K^{\mathcal{P}}_{]v[}$ is a* covering column *of $K^{\mathcal{P}}_{]u[}$.*

The essential feature of $\mathcal{F}(\mathcal{P}) = \emptyset$ lies in the fact formulated in the following lemma:

Lemma 9.6. *Let $\mathcal{P}$ be a structure such that $\mathcal{F}(\mathcal{P}) = \emptyset$. Then all of its non-trivial columns are covered in it.*

Proof. While the first column of structure $\mathcal{P}$ is covered by definition, it holds that $S^{\mathcal{P}}_2 = K^{\mathcal{P}}_1 \cap K^{\mathcal{P}}_2 \subseteq K^{\mathcal{P}}_1$. So, if $K^{\mathcal{P}}_2$ is non-trivial, it is covered as well.

Choose an arbitrary $w \in N$ such that $]w[_{\mathcal{P}}\geq 3$. One can distinguish two cases:

I. $|S^{\mathcal{P}}_{]w[}| \leq 1$

II. $|S^{\mathcal{P}}_{]w[}| \geq 2$

In the case of $S^{\mathcal{P}}_{]w[} \leq 1$, either $S^{\mathcal{P}}_{]w[} = \emptyset$, and then $K^{\mathcal{P}}_1$ can be its covering column, or $|S^{\mathcal{P}}_{]w[}| = 1$. Put $u = S^{\mathcal{P}}_{]w[}$ and observe that $K^{\mathcal{P}}_{]u[}$ is its covering column by Definition 9.5.

Assume now that $|S^{\mathcal{P}}_{]w[}| \geq 2$. Choose and fix $v \in S^{\mathcal{P}}_{]w[}$ such that $v \succeq_{\mathcal{P}} v'$ for all other $v' \in S^{\mathcal{P}}_{]w[}$. Now, let us show that $S^{\mathcal{P}}_{]w[} \subseteq K^{\mathcal{P}}_{]v[}$ by considering the opposite for a contradiction – i.e., let $\exists u \in S^{\mathcal{P}}_{]w[}$ such that $u \notin K^{\mathcal{P}}_{]v[}$. Then $u \prec_{\mathcal{P}} v$ by the choice of v and $u \not\leftrightarrow_{\mathcal{P}} v$ by Definition 5.1. It implies that $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$ by Remark 5.13, which contradicts the lemma supposition $\mathcal{F}(\mathcal{P}) = \emptyset$. Hence, $S^{\mathcal{P}}_{]w[} \subseteq K^{\mathcal{P}}_{]v[}$ and $K^{\mathcal{P}}_{]w[}$ is covered by $K^{\mathcal{P}}_{]v[}$ in this case, which finishes the proof. $\square$

The notion of a covering column has a close connection to the so-called *left cycle permutation.* More precisely, in the view of Definition 6.15, any covered column may be moved just behind its covering column using a left cycle permutation. Indeed, let $K^{\mathcal{P}}_i, K^{\mathcal{P}}_{i+k}$ be a couple of covering and covered column,i.e., $K^{\mathcal{P}}_i \supseteq S^{\mathcal{P}}_{i+k}$ by Definition 9.5. Then $(i{+}1 \quad i{+}2 \quad \ldots \quad i{+}k)$ is a left cycle permutation in

$\mathcal{P}$. Using this and the fact that (1 2) is either a box transposition or a constant one in any structure $\mathcal{P}$, one can easily prove that in the case of $\mathcal{F}(\mathcal{P}) = \emptyset$, any non-trivial column may be moved to the first position in the structure using IE operations only:

Lemma 9.7. *For structure $\mathcal{P}$ over N with $\mathcal{F}(\mathcal{P}) = \emptyset$ and every variable $u \in N$, there exists a permutation $\sigma \in T_{|\mathcal{P}|}$ such that $(]u[_{\mathcal{P}})\sigma = 1$ ($K^{\mathcal{P}}_{]u[} \equiv K^{\mathcal{P}\sigma}_{1}$) and $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}\sigma)$.*

Proof. We will prove a little bit stronger version of this assertion by induction on N (we organize the set of variables using the relation $\preceq_{\mathcal{P}}$): The induction hypothesis for $u \in N$ such that $]u[_{\mathcal{P}} \geq 1$ is that there exists a permutation σ such that $(]u[_{\mathcal{P}})\sigma = 1$, $(]v[_{\mathcal{P}})\sigma =]v[_{\mathcal{P}}$ for all $v \succ_{\mathcal{P}} u$, and $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}\sigma)$. It is evident for $u \in N$ such that $]u[_{\mathcal{P}} = 1$, consider *identical permutation* in this case.

Assume $u \in N$ such that $]u[_{\mathcal{P}} \geq 2$ and that the implication holds for every $v \in N$ such that $v \prec_{\mathcal{P}} u$. Then $K^{\mathcal{P}}_{]u[}$ is covered by Lemma 9.6, i.e., $\exists K^{\mathcal{P}}_{]v[} \in \mathcal{P}$ such that

$$S^{\mathcal{P}}_{]u[} \subseteq K^{\mathcal{P}}_{]v[}. \tag{9.3.1}$$

Since $v \prec_{\mathcal{P}} u$ then by induction hypothesis, there exists $\sigma \in T_{|\mathcal{P}|}$ such that

$$(]v[_{\mathcal{P}})\sigma = 1 \tag{9.3.2}$$

and $(]w[_{\mathcal{P}})\sigma =]w[_{\mathcal{P}}$ for all $w \in N$ such that $w \succ_{\mathcal{P}} v$.

Observe, however, that then

$$S^{\mathcal{P}}_{]w[} = S^{\mathcal{P}\sigma}_{]w[} \textit{ for all such } w \in N, w \succ_{\mathcal{P}} v \tag{9.3.3}$$

by definition of $S^{\cdot}_{]w[}$. Combining all of the expressions (9.3.1), (9.3.2), and (9.3.3) with the fact that $u \succ_{\mathcal{P}} v$, we can easily obtain the relationship $S^{\mathcal{P}\sigma}_{]u[} \subseteq K^{\mathcal{P}\sigma}_{1}$ guaranteeing that $\sigma_L = (2\ 3\ \ldots\ ]u[_{\mathcal{P}})$ is a left cycle permutation in $\mathcal{P}\sigma$. Put $\sigma_{cb} = (1\ 2)$. Then $\sigma' = \sigma\sigma_L\sigma_{cb}$ guarantees that $(]u[_{\mathcal{P}})\sigma' = 1$, $(]v[_{\mathcal{P}})\sigma =]v[_{\mathcal{P}}$ for all $v \succ_{\mathcal{P}} u$ by their definition. Moreover, since $\sigma_{cb} = (1\ 2)$ is either a constant or a box transposition in every structure $|\mathcal{P}| \geq 2$, then $\mathcal{I}(\mathcal{P}) = \mathcal{I}(\mathcal{P}\sigma) = \mathcal{I}(\mathcal{P}\sigma\sigma_L\sigma_{cb})$ by induction hypothesis and Theorem 7.2 respectively. ☐

Realizing the fact that a set of non-trivial columns of a structure over N contains all variables from N, it follows that $\mathcal{F}(\mathcal{P}) = \emptyset$ is a sufficient condition for flexibility of the respective structure $\mathcal{P}$. Hence, using Corollary 9.4, emptiness of $\mathcal{F}(\mathcal{P})$ is not only necessary but also a sufficient condition for flexibility of structure $\mathcal{P}$:

Corollary 9.8. *Structure $\mathcal{P}$ is flexible $\Leftrightarrow \mathcal{F}(\mathcal{P}) = \emptyset$.*

Notice that the condition of $\mathcal{F}(\mathcal{P}) = \emptyset$ means the existence of an equivalent permutation of $\mathcal{P}$ for every variable so that the selected variable appears in the first column of this permutation. But perhaps there could be non-trivial columns in $\mathcal{P}$ that do not appear at the beginning of any such an equivalent structure. However, $\mathcal{F}(\mathcal{P}) = \emptyset$ guarantees the existence of an equivalent permutation for every non-trivial column such that this column appears at the beginning of the equivalent permutation.

In other words, it allows us to move any non-trivial column to the beginning of the sequence using an IE permutation. See the following example to illustrate the difference:

Example 9.9. *Consider $U_1 = \{u, v\}, U_2 = \{v, w\}$, and $U_3 = \{w, x\}$. Observe that for structures U_1, U_2, U_3 and U_3, U_2, U_1, the corresponding formal ratios coincide:*

$$U_1, U_2, U_3: \ \frac{\{u,v\} \cdot \{v,w\} \cdot \{w,x\}}{v \cdot w},$$

$$U_3, U_2, U_1: \ \frac{\{u,v\} \cdot \{v,w\} \cdot \{w,x\}}{v \cdot w}.$$

Then structures U_1, U_2, U_3 and U_3, U_2, U_1 are independence equivalent by Theorem 7.2. Moreover, if any of u, v, w, and x appears in the first column of one of those structures, then this fact is sufficient for the flexibility of U_1, U_2, U_3.

Note that since $\mathcal{F}(U_1, U_2, U_3) = \emptyset$, there also exists an equivalent permutation with U_2 at the first position by Lemma 9.7 and Corollary 6.47 (specifically U_2, U_1, U_3). However, its existence is not sufficient for the flexibility of U_1, U_2, U_3 in this case.

9.3.2 Flexible structures versus flexible sequences

For a generating sequence and its IE permutation (the corresponding structures are independence equivalent), the pairwise consistency of the considered distributions guarantees that both sequences are equivalent simultaneously (they represent identical multidimensional distributions) by Lemma 8.8. Considering the definition of structure flexibility, flexible structures are closely connected with independence equivalence and thus with IE operations as well. The impact of IE operations on an arbitrary distribution represented by a compositional model was described in Chapter 8. Recall that every equivalent structure permutation may be obtained by iterative applications of constant and box transpositions. To guarantee the same compositional model, consistency of those pairs of distributions that were affected by a box-transposition is required. Then, in the case of pairwise consistency, one can derive the following deduction:

Lemma 9.10. *If $\pi_1(U_1), \pi_2(U_2), \ldots, \pi_n(U_n)$ is a sequence of pairwise consistent probability distributions with flexible structure $U_1, U_2, \ldots, U_n$, then this sequence is flexible.*

Proof. This is a simple consequence of Corollary 9.8, Lemma 9.7, and iterative applications of Lemmata 8.3 and 8.4. □

Remark 9.11. *One may object that the previous lemma – Lemma 9.10 – represents literally "reinvention of the wheel". Indeed, note that the condition $\mathcal{F}(\mathcal{P}) = \emptyset$ corresponds to the so-called* running intersection property (RIP). *Its definition is the following: Let $U_1, U_2, \ldots, U_n$ be a sequence of sets. Then this sequence meets RIP if*

$$\forall i = 2, \ldots, n \quad \exists j : (1 \leq j < i) \left(U_i \cap (\bigcup_{k=0}^{i-1} U_k) \subseteq U_j \right).$$

This definition can easily be rewritten when the sequence of sets represents a structure $\mathcal{P}$: Structure $\mathcal{P}$ meets RIP if

$$\forall i = 2, \ldots, |\mathcal{P}| \quad \exists j : (1 \leq j < i) \left(S_i^{\mathcal{P}} \subseteq K_j^{\mathcal{P}} \right).$$

For example, in the case of RIP there is a covering column for each column (including trivial ones). In cases when we do not consider trivial columns, the condition of $\mathcal{F}(\mathcal{P}) = \emptyset$ coincides with RIP and one can find the following lemma in [24]: If $\pi_1(U_1), \pi_2(U_2), \ldots, \pi_n(U_n)$ is a sequence of pairwise consistent probability distributions such that $U_1, U_2, \ldots, U_n$ meets RIP then this sequence is flexible.

Considering the proof of Lemma 9.7 we can see that Lemma 9.10 may be slightly modified. Use knowledge of Chapter 8 to check the demands of IE transpositions pertaining to the respective distributions. Recall that while a box transposition requires the consistency of respective distributions to guarantee the equality of the permuted compositional model, a constant transposition has no additional claims on the corresponding distributions. Then, given the proof of Lemma 9.7, we do not need to require the pairwise consistency, we only need the consistency of those pairs of distributions that correspond to the covering/covered pairs of columns in the respective structure:

Corollary 9.12. *If $\pi_1(U_1), \pi_2(U_2), \ldots, \pi_n(U_n)$ is a sequence of probability distributions with flexible structure $U_1, U_2, \ldots, U_n$ such that the pairs of distributions which correspond to the covering/covered pairs of columns in the structure are consistent, then this sequence is flexible.*

Note that this could become handy for automatic checking of flexibility. *In the case of a generating sequence of length n, we restrict the number of consistencies to verify from*

$$\frac{n(n-1)}{2}$$

to

$$n-1.$$

Note that the consistency of the "covered $\times$ covering" distributions ensures pairwise consistency in the case of structure flexibility,. It is then (in the case of structure consistency) a sort of minimum spanning for pairwise consistency.

In cases of a structure, flexibility guarantees that for every non-trivial column there exists an independence equivalent permutation with this column at the first position. Then, by generalizing it to a generating sequence (with a structure without non-trivial columns), it holds that for a distribution in the sequence there exists an equivalent permutation with the distribution at the beginning, i.e., every distribution represents a marginal of the respective compositional model by Assertion 3.7. Hence:

Corollary 9.13. *If a generating sequence has flexible structures that do not contain any trivial column, and the distributions corresponding to the covering/covered column pairs in the structure are consistent, then this sequence is perfect.*

9.4 Inflexible structure

It has been shown that the question of flexibility of a generating sequence is simple for sequences having flexible structures. In this section we will focus on generating sequences whose structures induce at least one F-condition, i.e., sequences with *inflexible structures.*

To illustrate the problem, choose a simple generating sequence with an inflexible structure. Put $\mathcal{P} = u, v, \{u, v, w\}$. Observe that $\langle u, v|w\rangle \in \mathcal{F}(\mathcal{P})$. See its visualization in Figure 9.1. Assume the flexibility of the corresponding generating sequence $\pi_1(u)$, $\pi_2(v)$, $\pi_3(u, v, w)$. and put $\pi = \pi_1 \rhd \pi_2 \rhd \pi_3$.

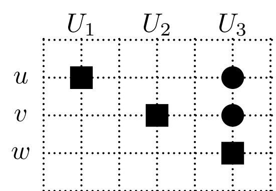


Figure 9.1: $\mathcal{P}$: inflexible structure

Since π_3 is the only distribution having w in its argument, due to the flexibility there has to exist an equivalent permutation of π_1, π_2, π_3 starting with π_3. Let π_3, π_1, π_2 be such a permutation. (Both π_3, π_1, π_2 and π_3, π_2, π_1 are symmetric in u and v, and therefore a particular choice makes no difference.) Then by the flexibility

$$\frac{\pi_1(u)\pi_2(v)\pi_3(u,v,w)}{\pi_3(u,v)} = \frac{\pi_3(u,v,w)\pi_1(u)\pi_2(v)}{\pi_1(u)\pi_2(v)}. \tag{9.4.1}$$

Observe that (1 2) is a constant transposition in $u, v, \{u, v, w\}$. Therefore both π_1 and π_2 represent marginals of π according to Assertion 3.7. Moreover, the flexibility ($\pi_1 \rhd \pi_2 \rhd \pi_3 = \pi_3 \rhd \pi_1 \rhd \pi_2$) and Assertion 3.7 guarantee that π_3 is a marginal of π too – i.e., π_1, π_2, π_3 is a perfect sequence. Then (9.4.1) may be rewritten as

$$\frac{\pi(u)\pi(v)\pi(u,v,w)}{\pi(u,v)} = \frac{\pi(u,v,w)\pi(u)\pi(v)}{\pi(u)\pi(v)}, \tag{9.4.2}$$

which holds if and only if $u \perp\!\!\!\perp v[\pi]$. This is, however, guaranteed by structure $\mathcal{P} = u, v, \{u, v, w\}$ of π_1, π_2, π_3.

> *Hence, perfectness is a necessary condition for flexibility in the case of an arbitrary generating sequence $\pi_1(u), \pi_2(v), \pi_3(u, v, w)$.*

One may object that this generating sequence – as an example – is irrelevant because it can be equivalently replaced by only distribution π_3. Let us slightly extend this model so that this problem is avoided.

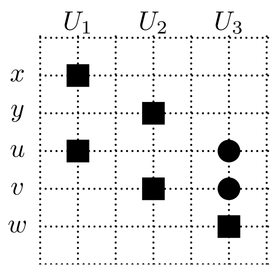


Figure 9.2: $\mathcal{P}$: another inflexible structure

Let $\mathcal{P}' = \{u, x\}, \{v, y\}, \{u, v, w\}$ be given. See its visualization in Figure 9.2, and assume flexibility of the corresponding generating sequence $\pi_1(u, x), \pi_2(v, y), \pi_3(u, v, w)$. Moreover, put $\pi' = \pi_1 \rhd \pi_2 \rhd \pi_3$. Observe that $\sigma = (1\ 2)$ is a constant transposition in $\mathcal{P}'$ and that therefore both π_1 and π_2 represent marginals of π' by Assertion 3.7. Since π_3 is the only distribution having w among its arguments again, the flexibility induces the existence of an equivalent permutation of π_1, π_2, π_3 starting with π_3. Let π_3, π_1, π_2 be such a permutation. (Note that the other possible permutation – π_3, π_2, π_1 – is a constant transposition of π_3, π_1, π_2, i.e., they are equivalent in this case). Then

$$\pi' = \frac{\pi_1(u,x)\pi_2(v,y)\pi_3(u,v,w)}{\pi_3(u,v)} = \frac{\pi_3(u,v,w)\pi_1(u,x)\pi_2(v,y)}{\pi_1(u)\pi_2(v)} \tag{9.4.3}$$

and π_3 is a marginal distribution of π' due to Assertion 3.7. Hence, π_1, π_2, π_3 is a perfect sequence as well. One can rewrite (9.4.3) as

$$\frac{\pi'(u,x)\pi'(v,y)\pi'(u,v,w)}{\pi'(u,v)} = \frac{\pi'(u,v,w)\pi'(u,x)\pi'(v,y)}{\pi'(u)\pi'(v)} \tag{9.4.4}$$

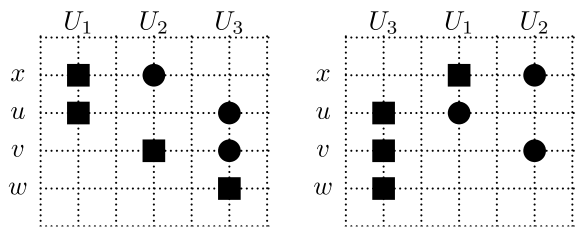
which holds true if and only if $\pi'(u,v) = \pi'(u)\pi'(v)$, i.e. $u \perp\!\!\!\perp v[\pi']$. This is, however, a guaranteed structure $\mathcal{P}'$ of this compositional model representing π'.

> *Hence perfectness is a necessary condition of flexibility for an arbitrary generating sequence* $\pi_1(u,x), \pi_2(v,y), \pi_3(u,v,w)$.

These two foregoing examples would lead us to a hypothesis that *perfectness of a generating sequence is a sufficient condition for its flexibility.* Unfortunately, this hypothesis can easily be refuted by the following counter-example:

Let $\pi_1(u,x), \pi_2(v,x), \pi_3(u,v,w)$ be a perfect generating sequence. (See its structure on Figure 9.3a.) Put $\pi'' = \pi_1 \triangleright \pi_2 \triangleright \pi_3$. Since $\pi_1(u,x), \pi_2(v,x), \pi_3(u,v,w)$ is perfect, then the definition of perfectness implies that all of π_1, π_2, π_3 represent marginals of π'' and

$$\pi'' = \frac{\pi_1(u,x)\pi_2(v,x)\pi_3(u,v,w)}{\pi_2(x)\pi_3(u,v)} = \frac{\pi''(u,x)\pi''(v,x)\pi''(u,v,w)}{\pi''(x)\pi''(u,v)} \tag{9.4.5}$$



(a): U_1, U_2, U_3 (b): U_3, U_1, U_2

Figure 9.3: A contra-example of an inflexible structure

π_3 is the only distribution which has w among its arguments. If π_1, π_2, π_3 were flexible, then there should exist an equivalent permutation of π_1, π_2, π_3 starting with π_3. Since both possible permutations π_3, π_1, π_2 and π_3, π_2, π_1 are interchangeable from this point of view, let π_3, π_1, π_2 be the expected permutation. Then, however, by definition of flexibility

$$\frac{\pi''(u,x)\pi''(v,x)\pi''(u,v,w)}{\pi''(x)\pi''(u,v)} = \frac{\pi''(u,v,w)\pi''(u,x)\pi''(v,x)}{\pi''(u)\pi''(v,x)} \tag{9.4.6}$$

which holds true if their numerators are equal – i.e.,

$$\pi''(x)\pi''(u,v) = \pi''(u)\pi''(v,x). \tag{9.4.7}$$

Considering the definition of conditional distribution, one can transform (9.4.7) into

$$\pi''(x)\pi''(v|u)\pi''(u) = \pi''(u)\pi''(v|x)\pi''(x). \tag{9.4.8}$$

which can be reduced for all $x \in \mathbf{X}_{u\cup v\cup x}$ such that $\pi''(x_u) > 0$ and $\pi''(x_x) > 0$ cancel into the form

$$\pi''(v|u) = \pi''(v|x). \tag{9.4.9}$$

Since structure $\{u,x\},\{v,x\},\{u,v,w\}$ guarantees that $u \perp\!\!\!\perp v|x[\pi'']$ as well as the alternative definition (2.6.2) of conditional independence of variables, one can replace the right hand side of (9.4.9) by

$$\pi''(v|x) = \pi''(v|u,x). \tag{9.4.10}$$

Consequently,

$$\pi''(v|u,x) = \pi''(v|u). \tag{9.4.11}$$

That is nothing else than the definition of conditional independence $v \perp\!\!\!\perp x|u[\pi''] \equiv v \perp\!\!\!\perp x|u[\pi_1 \rhd \pi_2 \rhd \pi_3]$, which represents a necessary condition for the flexibility of perfect sequence structure $\pi_1(u,x), \pi_2(v,x), \pi_3(u,v,w)$.

The relationship $v \perp\!\!\!\perp x|u[\pi_1 \rhd \pi_2 \rhd \pi_3]$ is not guaranteed by its structure and it has to be primarily encoded in the respective distributions. Observe that this independence is nothing else than the independence encoded by a structure of permutation $\pi_3(u,v,w), \pi_1(u,x), \pi_2(v,x)$, cf. Figure 9.3b. Non-equivalent permutations of the structure have different properties and therefore induce other independence relationships. Logically, it is necessary to have identified all of these independence relations. It does not matter whether they are induced by the structure or by properties of low-dimensional distributions of the respective generating sequence.

Remark 9.14. *Regarding the choice of permutation π_3, π_1, π_2, we could use the other permutation in the beginning of this counter-example – π_3, π_2, π_1. We would finish with condition $u \perp\!\!\!\perp x|v[\pi'']$ in that case. That is why we can conclude with the following:*

> Consider a perfect sequence $\pi_1(u,x), \pi_2(v,x), \pi_3(u,v,w)$. This sequence is flexible if either $v \perp\!\!\!\perp x|u[\pi_1 \rhd \pi_2 \rhd \pi_3]$ or $u \perp\!\!\!\perp x|v[\pi_1 \rhd \pi_2 \rhd \pi_3]$.

From the previous discussion, one can see that the requirement for generating sequence's flexibility is very strong for a non-flexible structure. Nevertheless, we think that the concept of flexible sequences is worth further study. But it goes beyond the scope of this work. One of the questions that would be good to answer is: "If you have a flexible sequence, is the conditional generating sequence flexible again?" Consider a flexible generating sequence representing probability distribution $\pi(U)$, variable $u \in U$ and its specific value $x_u \in \mathbf{X}_u$. Transform the generating sequence into a sequence representing the conditional probability distribution $\pi(U \setminus u | u = x_u)$ using local changes described at the beginning of this chapter. Is the new generating sequence flexible again?

Chapter 10

Conclusions

This thesis deals with the properties of compositional models (Chapter 3) as a subclass of probabilistic models and analyzes them from the perspective of conditional independence assertions induced by their structures. The basic goal of the thesis is the design and description of properties characteristic of the structures inducing the same system of independence assertions – such structures are denoted as *independence equivalent*.

The introductory Chapter 1 summarizes not only the current state of the art in probabilistic models but also brings significant historic reference to certain areas of Artificial intelligence, probabilistic models, and graphical models. Chapter 2 then gathers fundamental notions and assertions from probability theory useful in the context of this text. In Chapter 3 we start to focus on compositional models. Except for fundamental definitions (*operator of composition*, *generating sequence*,...) this chapter also introduces compositional model's *structure* and related notions of the *induced independence relations*. At the end of the chapter one can also find several original and important observations (like an induced substructure) necessary for further development.

Chapter 4 provides an introduction to the so-called *equivalence problem* whose solution occupies a large part of this text. Note that among other related subproblems, this problem generally covers another question: when do two different structures of two compositional models induce the same set of conditional independence assertions?. Chapters 5 and 6 contain partial solutions of the equivalence problem from the perspective of properties invariant within a class of independence equivalent structures as well as elementary operations on structures preserving an induced independence model, and Chapter 7 puts all those partial solutions together by stating necessary and sufficient conditions for independence equivalence of two arbitrary structures – Theorem 7.2.

We have presented three properties characterizing independence equivalence of the respective compositional model structures in Chapter 5: *connection set* *combined with F-condition set, non-trivial sets*, and *formal ratio*. The first property was inspired by Verma and Pearl's characterization of equivalent acyclic directed graphs [56], but this property does not appear to be very suitable for compositional model structures. Therefore, the other two characterizations were derived. We consider the derivation of the *weak core* of a structure as an important achievement. Let us realize that the structure of a compositional model is basically a sequence of sets of variables (we refer to such sets as *columns* to distinguish them from common sets of variables). From this point of view, a weak core – consisting of several columns within the structure – denotes those columns common to all structures equivalent with the given one (Corollary 5.45). Hence, the weak core comparison represents a simple rule to recognize possible non-equivalence and, moreover, it represents a natural and logical transition to the indirect characterization of structure equivalence using some elementary operations on structures. In our case, these operations are permutations in particular. In connection with permutations, let us also mention the so called *reduced structure* as a non-unique representative of a class of equivalent structures. However, all reduced structures in one class of equivalence are permutations of each other. We may view a structure's formal ratio as a unique representative of an independence equivalence class even though a formal ratio is not a structure.

The second basic part was devoted to the study of applying an equivalence problem solution to other open problems. In Chapter 8 we investigate the impact of generalized elementary operations, originally introduced only for structures in Chapter 6, to the generating sequences and corresponding represented probability distributions. We identified other necessary conditions guaranteeing that probability distributions represented by respective compositional models are identical. Finally, Chapter 9 closes the thesis with a partial solution for determining *flexibility* and connected the *conditioning problem* of a generating sequence, using a solution of the equivalence problem from Chapter 7.

Bibliography

[1] Y.M.M. Bishop, S.E. Fienberg, P.W. Holland: *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, Massachusetts: MIT Press (1975)

[2] V. Bína: *Exhaustive Search among Compositional Models of Decomposable Type*, Proceedings of 13th Czech-Japan Seminar on Data Analysis and Decision Making in Service Science, Otaru, Japan, (2010), pp. 103-108.

[3] V. Bína, R. Jiroušek: *Marginalization in Multidimensional Compositional Models*, Kybernetika 42, (2006), pp. 405-422.

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

[5] C. Boutilier: *The Influence of Influence Diagrams on Artificial Intelligence*, Decision Analysis, Vol.2 No. 4, (2005), pp. 229-231.

[6] B. G. Buchanan: *A (Very) Brief History of Artificial Intelligence.* [online] AI Magazine 26(4). 2005, [cited 2011-07-04]. Available from: ⟨http://www.aaai.org/AITopics/assets/PDF/AIMag26-04-016.pdf⟩.

[7] B.G. Buchanan: *Timeline: A Brief History of Artificial Intelligence* [online]. AITopics:The Association for the Advancement of Artificial Intelligence; 2011 [cited 2011-07-07]. Available from: ⟨http://www.aaai.org/AITopics/pmwiki/pmwiki.php/AITopics/BriefHistory⟩.

[8] M.D. Chickering: *A transformational characterization of equivalent Bayesian networks*, Uncertainty in Artificial Intelligence 11, Morgan Kaufmann (1995), pp. 87-98.

[9] D.M. Cifarelli, E. Regazzini: *De Finetti's Contribution to Probability and Statistics*, Statistical Science, Vol. 11, Issue 4, (1996), pp. 253-282.

[10] J.N. Darroch, S.L. Lauritzen and T.P. Speed, *Markov field theory and loglinear interaction models for contingency tables*, Annals of Statistic, 8, (1980), pp. 522-539.

[11] A. Darwiche: *Modeling and reasoning with Bayesian networks*, Cambridge University Press, (2009)

[12] A.P. Dawid: *Conditional independence in statistical theory*, Journal of the Royal Statistical Society, Series B,41 (1979) pp. 1-31.

[13] D. Edwards, T. Havránek: *A Fast Procedure for Model Search in Multidimensional Contingency Tables*, Biometrika, 72, pp. 339- 351. (1985)

[14] D. Edwards, T. Havránek: *A Fast Model Selection Procedure for Large Families of Models*, Journal of the American Statistical Association Vol. 82, No. 397 (1987), pp. 205-213

[15] *Felynx Cougati Project* [online]. 1999 [cited 2011-07-07]. *Gambling and the Chevalier De Mere*. Available from: ⟨http://www.ualberta.ca/MATH/gauss/fcm/BscIdeas/probability/DeMere.htm⟩.

[16] P. Gärdenfors: *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA, (1988).

[17] P. Hájek, T. Havránek, R. Jiroušek: *Uncertain Information Processing in Expert Systems*, CRC Press, (1992).

[18] T. Havránek: *A Procedure for Model Search in Multidimensional Contingency Tables*, Biometrics, 40,(1984) pp. 95-100.

[19] R.A. Howard, J.E. Matheson: *Readings on the principles and Applications of Decision Analysis*, Strategic Decisions Group, Menlo Park, Ca., (1984).

[20] J. Ignizio: *Introduction to Expert Systems*, (2001). ISBN 0-07-909785-5.

[21] R. Jiroušek: *A survey of methods used in probabilistic expert systems for knowledge integration*, Knowledge-Based Systems vol.3, 1 (1990), pp. 7-12.

[22] R. Jiroušek: *Solution of the Marginal Problem and Decomposable Distributions*, Kybernetika vol.27, 5 (1991), pp. 403-412

[23] R. Jiroušek: *What is the difference between Bayesian networks and compositional models?*, Proceedings of the 7th Czech–Japan Seminar on Data Analysis and Decision Making under Uncertainty, Osaka, Japan, (2004), pp. 191-196.

[24] R. Jiroušek: *Multidimensional Compositional Models*. Preprint DAR - ÚTIA 2006/4, ÚTIA AV ČR, Prague, (2006).

[25] R. Jiroušek: *Persegrams of Compositional Models Revisited: conditional independence.* In Proceedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, Malaga, (2008).

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

[27] R. Jiroušek: *Foundations of compositional model theory*, International Journal of General Systems - Volume 40, Issue 6, (2011), pp. 623-678.

[28] G.D. Kleiter, R.Jiroušek: *Perfect sequences: a contribution to structuring conditional independence models* , Proceedings of the 8th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty, (2005), pp. 65-75.

[29] G.D. Kleiter: *Ordering and counting essential graphs*, 7th Workshop on Uncertainty Processing, Mikulov, (2006), pp. 112-125.

[30] T. Kočka, R. R. Bouckaert, M. Studený: *On the Inclusion Problem.* Research report 2001, ÚTIA AV ČR, Prague (2001).

[31] V. Kratochvíl: *Dissertation thesis analysis: research report*, Prague, (2008).

[32] V. Kratochvíl: *Equivalence Problem in Compositional Models*, WUPES'09, Eds: Kroupa T., Vejnarová J., WUPES'09, Liblice (2009).

[33] V. Kratochvíl: *Motivatio for different characterization of Equivalent Persegrams*, Proceedings of the 12th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty, Eds: Novák V., Pavliska V., Štěpnička M., Czech-Japan Seminar on Data Analysis and Decision-making under Uncertainty /12./, Litomyšl (2009).

[34] V. Kratochvíl: *Different Approaches of Study Direct Equivalence Characterization*, Doktorandské dny 2009, sborník workshopu doktorandů FJFI oboru Matematické inženýrství, Eds: Ambrož P., Masáková Z., ČVUT Praha (2009), pp. 101-110. ISBN 978-80-01-04436-0

[35] V. Kratochvíl: *Characteristic Properties of Equivalent Structures in Compositional Models*, International Journal of Approximate Reasoning vol. 52,5 (2011), pp. 599-612.

[36] S. Lauritzen, D.J. Spiegelhalter: *Local computations with probabilities on graphical structures and their application to expert systems.* Journal of Royal Statistics Society, Series B, 50(2), (1988), pp. 157-224.

[37] S.L. Laurizent and N. Wermuth: *Graphical models with associations between variables, some of which are quantitative and some qualitative*, Annals of Statistics, 17 (1989), pp. 31-57.

[38] J. McCarthy: *Epistemological problems of artificial intelligence.* In Proceedings of the Fifth International Joint Conference on Artificial Intelligence, (1977).

[39] R. Merris: *Graph Theory.* Wiley Interscience, New York (2001).

[40] J. Pearl: *Probabilistic reasoning using graphs* In Proceedings of IPMU'1986, (1986), pp. 200-202.

[41] J. Pearl and A. Paz: *Graphoids: a graph-based logic for reasoning about relevance relations*, Advances in Artificial Intelligence II, eds. B.Du Boulay, D. Hogg and L. Steels, (1987). pp. 357-363

[42] J. Pearl: *Probabilistic Reasoning in Intelligent systems: Networks of Plausible Inference*, Margan Kaufmann, San Mateo, CA, (1988).

[43] J. Pearl: *Causality: Models, Reasoning, and Inference*, Cambridge University Press, (2000).

[44] A.M. Polansky: Northern Illinois University [online] Division of Statistics, 2002 [cited 2011-07-07]. *A short history of Probability*, Division of statistics, Northern Illinois University, Available from:
⟨http://staff.ustc.edu.cn/ zwp/teach/Prob-Stat/A short history of probability.pdf⟩

[45] R.D. Schater: *Evaluating influence diagrams*, Annals of Operations Research,34 (1986), pp.871-882.

[46] G. Shafer: *Advances in the Understanding and Use of Conditional Independence*, Annals of Mathematics and Artificial Intelligence, Vol. 21, Issue 1, (1996), pp. 1-11.

[47] J.Q. Smith: *Influence diagrams for statistical modeling*, Annals of Statistics, 17 (1989), pp. 654-672.

[48] W. Spohn: *Stochastic independence, causal independence, and shieldability*, Journal of Philosophical Logic, 9 (1980), pp. 73-99.

[49] M. Studený: *Conditional independence relations have no finite complete characterization. In Information Theory, Statistical Decision Functions and Random Processes*. Transactions of the 11th Prague Conference vol. B (S. Kubik, J.A. Visek eds.) Kluwer, Dordrecht - Boston - London (also Academia, Prague) (1992), pp. 377-396.

[50] M. Studený: *Comparison of graphical approaches to description of conditional independence structures*. In Proceedings of WUPES‘97, Prague, Czech Republic, (1997), pp. 156-172

[51] M. Studený: *Semigraphoids and structures of probabilistic conditional independence,* Baltzer Journals, (1997).

[52] M. Studený: *Probabilistic Conditional Independence Structures*, Springer, London, (series Information Science and Statistics), (2005).

[53] M. Studený: *O strukturách podmíněné nezávislosti.* Rukopis série přednášek. Prague (2008).

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

[55] D. Teets, K. Whitehead: *The Discovery of Ceres: How Gauss Became Famous* Mathematics Magazine Vol. 72, No. 2 (1999), pp. 83-93.

[56] T.S. Verma, J. Pearl: *Equivalence and synthesis of Causal models*, Uncertainty in Artificial Inteligence 6, (1991).

[57] N. Wermuth, D.R. Cox: *Graphical Models: Overview.* International Encyclopedia of the Social & Behavioral Sciences, Elsevier Science Ltd., (2001), pp. 6379-6386

[58] Wikipedia contributors: *History of artificial intelligence* [online]. Wikipedia, The Free Encyclopedia; 2011-06-23 [cited 2011-07-07]. Available from: ⟨http://en.wikipedia.org/wiki/History_of_artificial_intelligence⟩.

[59] Wikipedia contributors: *Permutation* [online]. Wikipedia, The Free Encyclopedia; 2011-06-24 [cited 2011-07-07]. Available from: ⟨http://en.wikipedia.org/wiki/Permutation⟩.

[60] S. Wright: *The method of path coefficients.* Annals of Mathematical Statistics, 5 (1934), pp. 161-212.

[61] *Leonardo Da Vinci's Lion Robot for the King of France, Year-1515* [online]. YouTube; 2008 [cited 2011-07-04]. Available from: ⟨http://www.youtube.com/watch?v=7jBkwCWxaic⟩.