Università degli Studi di Padova
Dipartimento di Scienze Statistiche

Scuola di Dottorato in
Scienze Statistiche
Ciclo XX

Combining information from Gaussian graphical models

Maria Sofia Massa

Direttore: Prof.ssa A. Salvan

Supervisore: Prof.ssa M. Chiogna

Co-supervisore: Prof. S.L. Lauritzen

31/01/2008

Riassunto

Data una distribuzione congiunta di probabilità, generalmente si possono determinare le sue componenti marginali. Il processo inverso non è però semplice e, in alcuni casi, non è nemmeno fattibile. In questa tesi studieremo il contesto generale sottostante alla composizione di famiglie di distribuzioni di probabilità, al fine di derivare una famiglia congiunta di distribuzioni. L'attenzione sarà rivolta specificatamente a distribuzioni di probabilità assolutamente continue rispetto ad un'opportuna misura prodotto.

Lo studio della compatibilità delle famiglie marginali sarà il punto di partenza del lavoro. Successivamente introdurremo due tipi di composizione di famiglie di distribuzioni di probabilità, corrispondenti all'informazione iniziale posseduta. La metodologia proposta permetterà poi di passare a studiare la composizione di famiglie normali multivariate che rispettano proprietà di indipendenza condizionale, ossia modelli grafici Gaussiani. Verranno forniti esempi di composizione di modelli grafici Gaussiani e proposti alcuni metodi per l'inferenza sui parametri del modello grafico congiunto. Infine, saranno presentati due studi di simulazione per verificare l'affidabilità della metodologia proposta.

Abstract

Given a joint probability distribution, one can generally find its marginal components. However, it is not straightforward, or even possible, the inverse procedure. In this thesis, we shall study the wide context of combining families of distributions. In particular, we shall consider absolutely continuous distributions with respect to a product measure.

The conditions for compatibility of the marginal families shall be the initial research problem to be investigated. Next, we shall classify two types of combination corresponding to the initial available information. The methodology previously introduced shall permit to lead the way to study the combination of families of multivariate normal distributions respecting some conditional independence relationships, i.e., Gaussian graphical models. Examples of combination of Gaussian graphical models and methods to estimate the parameters of the joint family of distributions shall be also provided. Eventually, we shall perform two simulation studies to assess the proposed methodology.

Acknowledgments

Foremost, I would like to thank my supervisor, Monica Chiogna, for her precious guidance and support during all these years. She, together with Chiara Romualdi and Duccio Cavalieri, suggested the research problem addressed in this thesis.

I would equally like to thank my co-supervisor, Steffen Lauritzen, for everything I learnt from him, and for having coordinated with Monica my stay in Oxford. While in Oxford, I worked on the core part of this thesis, and I am grateful to him for the interesting and motivating discussions we had.

I am also very grateful to everyone on the PhD School committee, for the organisation of the courses and the activities during these three years.

I would like to thank Alberto Roverato, for insightful comments and advices on topics related to this work. I am also very grateful to Luigi Pace, for his beneficial advices and suggestions. Marco Ferrante, Chiara Romualdi, and Erich Battistin are warmly acknowledged for their helpful opinions and encouragements.

I am grateful to my friends of the XIX and XX ciclo: Giuliana, Federico, Antonio, Daniela, Moreno, Simone, Francesca, Nedda, Marco and Aldo. I would also like to thank my friends in the Department of Statistics at Oxford: Ioana, Tim, Anjali, Klea, Kjell, Ludger, Dave, Oemetse, Liz, Rhoss, Amber, Alfred, Roland, Almund. I remember fondly the moments we spent together, both for study or lunches and coffee breaks at the University Clubs or in the colleges. Through them, I understood more about Romania, Canada, Australia, England, Botswana, US, Trinidad and Tobago, South Africa, and Germany.

I would like to thank my dear friends Giuseppina, Diego, Filippo, Francesca, Erica and Maria, for finding the time to meet and tell each other our stories. Roger also deserves a special thanks.

I am deeply grateful to my parents, Ottorino and Norma, my brother, Emanuele, and my sister, Caterina, for supporting and encouraging me.

Contents

Chapter 1

Introduction

1.1 Overview

Consider a number of independent studies that not necessarily address the same question, but involve some common variables. This may be the case, for example, when different laboratories perform independently studies which are only partially overlapping, or they perform the same study under different circumstances, or they analyse related phenomena using quite different methods. These situations are very common in some research fields, such as, for example, molecular biology. Recent progress in the field of genomics has made available a large amount of biological data from different sources and types of experiments, which provide considerable information on important biological cellular processes. In these cases, interest might lie in binding together information coming from these sources.

If a study is concerned with interactions among variables, the information is usually represented by a network. In molecular biology, for example, a network is used to represent metabolic pathways (chains of biochemistry reactions), and gene regulatory networks. They all have an evidence of experimental type and they are, usually, the results of independent experiments. A possible intent might be that of reconstructing a meaningful joint network, to improve the biological understanding. An example of this can be found in Goh et al. (2007), where a large number of studies of diseases and related genes are combined to form "the human diseasome" bipartite network. In Chapter 2, we shall describe in detail this example.

Such situations can be framed within the context of *meta-analysis*. Meta-analysis is usually meant to integrate and combine evidence from different studies that are welldesigned, address the same questions and use similar outcome measures (see Glass, 1976; Hedges and Olkin, 1985). Here, the concept of meta-analysis has to be extended so as to cover the more general process of producing a consistent overview of information from several independent studies, possibly only partially overlapping.

The aim of this thesis is to develop a general framework for meta-analysis of families of distributions, addressing in detail families which can be represented by graphical models, i.e., one possible statistical interpretation of the concept of network. Our focus is on a first development of formal concepts to underpin the ideas mentioned above. Throughout the work, we shall use the term *meta-analysis* in a general perspective, to embrace also studies that do not necessarily focus on the same research question. We shall consider the case of undirected Gaussian graphical models, referring in particular to the combination of two graphical Gaussian models, in the context of maximum likelihood estimation. However, the results that we have achieved can be generalised to more than two families.

The thesis is organised as follows. Chapter 2 describes in detail the statistical problem and presents some illustrative examples which motivate the following developments. The core of the thesis are Chapters 3, 4. Chapter 3 studies the conditions for compatibility of families of distributions and introduces a new methodology for their combination. Chapter 4 addresses the specific case of combining information from Gaussian graphical models. In Chapter 5, the emphasis is on maximum likelihood estimation of the joint set of distributions, among the classes of graphical and non-graphical combinations. In Chapter 6, we assess the proposed methodology by means of some studies with artificial data. We conclude with a general discussion and some possibilities for future work.

1.2 Contribution of this thesis

The research problem addressed in this thesis may be briefly described as meta-analysis of Gaussian graphical models. This appears to be a novel problem, which defines and opens several research issues: definition of compatibility and incompatibility for statistical models; identification of the joint graphical model along with rules for models combination; estimation of the parameters of a combination. It naturally leads to investigate the combination of families of distributions as a more general context.

Some work has been done on only partially touching topics. Fienberg and Kim (1999), and Kim (2006) address the problem of combining conditional graphical log-linear structures involving the same group of variables. They build a hyper model including the initial variables and a new one that takes into account the different structures. Then, they study the conditional independence relationships given by the hyper model. Kim (2001) proposes a method to combine the structures of marginal decomposable graphical models. He constructs a joint graph by analysing the structural properties and describing a *grafting process*, based on the minimal connectors of the marginal graphs. His idea is that of finding all the possible structures compatible with the initial ones, studying also the combination of graphs with empty intersection. Jiroušek and Vejnarova (2003) introduce a method to construct multidimensional distributions by integrating many low-dimensional ones using operators of composition. Their approach is an alternative construction and representation of multivariate distributions if compared with that of graphical models, and it is developed in both probabilistic and possibilistic frameworks. With respect to the current literature, the contributions of this thesis may be detailed as follows.

- A novel use of the definitions of meta-consistency and meta-Markov combination (Dawid and Lauritzen, 1993), to compose marginal families of distributions.
- Introduction of the concepts of *quasi-consistency* for families of distributions and *quasi-Markov combination* for the combination of two families of distributions.
- Study of the properties of the quasi-Markov combination with respect to preservation or reduction of the marginal families, and of its relations with the meta-Markov combination.
- Interpretation of the combination of Gaussian graphical models as a meta-Markov or a quasi-Markov combination, and definition of *graphical* and *non-graphical* combinations.
- Identification of the missing data perspective as an appropriate way to achieve the maximum likelihood estimation of the composed models.

Chapter 2

The problem

In this chapter, we shall present in more detail the considered research problem from both an applied point of view, and a statistical perspective.

2.1 A motivating example

Suppose that we have a large number of biological networks, each of them representing a known disease and the genes implicated in it. For example, imagine to have a network representing colon cancer by means of the genes whose mutation is related to this particular disease, one for breast cancer showing the relations between the corresponding genes, and so on. The intent might be that of finding a joint network showing the relations between all the diseases and genes. Goh et al. (2007) use network analysis to study a problem related to this. They begin with the lists of disorders, disease genes, and association between them documented in the Online Mendelian Inheritance in Man (OMIM). Their central idea is to condense in a single graph all the information about human genetic disorders and the corresponding disease genes, building the "diseasome" bipartite graph, as in Figure 2.1 (see also Figure 2.2 that shows a zoom on a smaller number of diseases and related genes). It can be seen as a prominent and large scale example of a graphical model meta-analysis, and its achievement is very important as it allows to derive biologically relevant marginal networks, such as the human disease network projection, and the disease gene network projection (see Figure 2.3 and Figure 2.4, respectively).

The same problem might be interpreted from a purely statistical point of view, as follows. Given some non-disjoint networks studied as graphical models, find and estimate a joint network (graphical model) by combining the initial ones. This is the problem that we shall tackle in this thesis.

Before going on, it is important to remark that, usually, the recent available data (for example, genomic, transcriptomic, proteomic data) suffer from many sources of experimental variability. For example, in a genomic context, variability arising throughout the measurement process of gene expression microarrays can obscure the biological signals of interest. Therefore, to make meaningful use of them, a number of preliminary actions (the so called *normalizations*) have to be taken to eliminate questionable measurements and correct low quality aspects. In this thesis, we shall assume that the available data has already been pre-processed (see Chiogna et al. (2007) for a discussion).

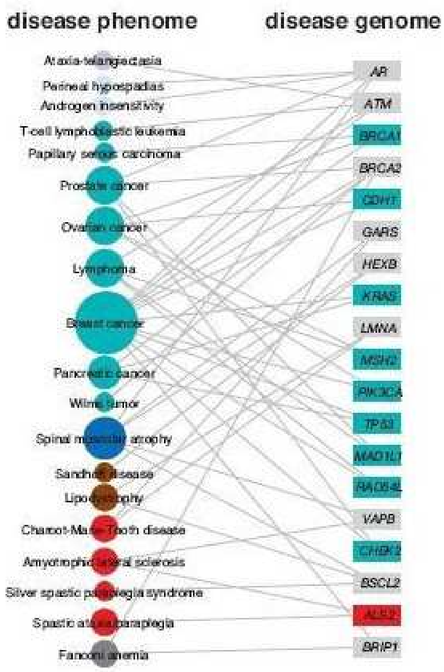


Figure 2.1: A small subset of the Diseasome bipartite network. On the left the **disorders**, on the right the **disease genes** (Goh et al., 2007).

2.2 Statistical aspects

Before developing a general approach for combination of Gaussian graphical models, in this section we provide some examples to introduce the ideas involved in the combination of graphical models. To begin with, we give the definition of a Gaussian graphical model, following Lauritzen (1996). We refer the reader also to the Appendix for a more detailed

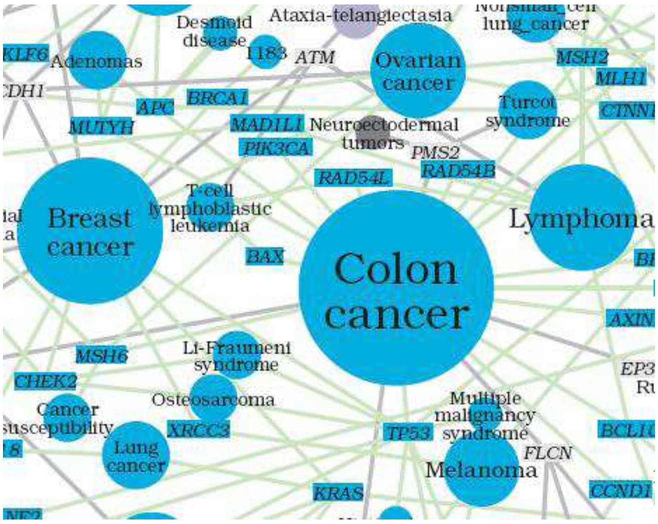


Figure 2.2: A zoom on the Diseasome bipartite network (Goh et al., 2007).

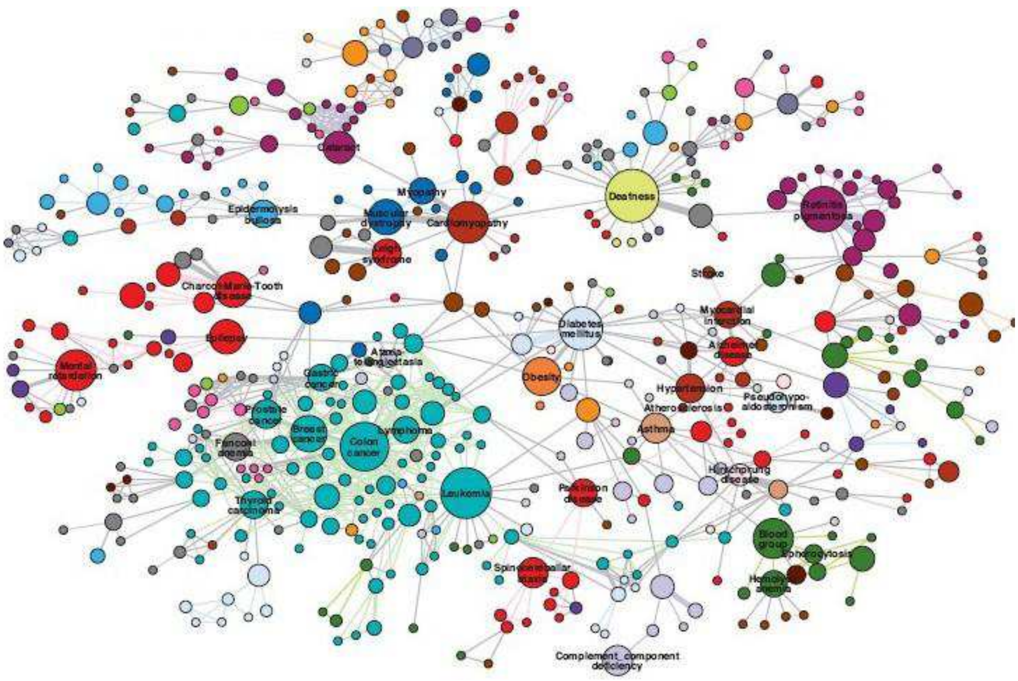


Figure 2.3: A subset of the Human Disease Network projection of the bipartite graph. Each node corresponds to a distinct **disorder** (Goh et al., 2007).

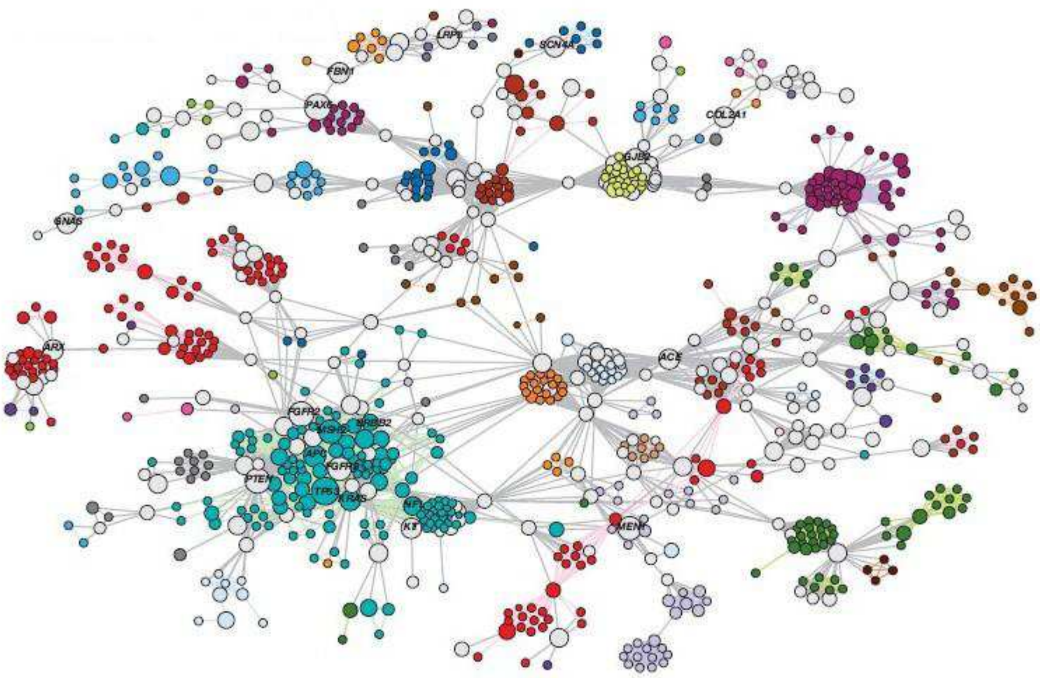


Figure 2.4: A subset of the Disease Gene Network projection of the bipartite graph. Each node is a **gene** (Goh et al., 2007).

exposition. Consider an undirected graph $G = (V, E)$, with $|V| = p$, where $|V|$ denotes the cardinality of V, that is the number of vertices of the graph G or, equivalently, the number of involved variables.

Definition 2.1. A Gaussian graphical model is the family of multivariate normal distributions $Y \sim N_p(\mu, \Sigma)$, where the mean μ is an arbitrary vector, the concentration matrix Σ^{-1} is assumed to be positive definite, and its elements are equal to zero whenever there is no edge between the corresponding elements in G.

For simplicity, we consider the case where the mean vector μ is set to zero, to focus the interest on issues concerning the covariance matrix. Thus, a Gaussian graphical model is represented by a set of multivariate normal distributions as $Y \sim N_p(0, \Sigma)$, where $\Sigma^{-1} \in S^+(G)$. $S^+(G)$ is the set of symmetric positive definite matrices, whose elements are equal to zero whenever there is no edge between the corresponding elements in G. Note that we use the term *graph* to indicate the conditional independence structure of the model, and the term *family* to indicate a graph and a set of distributions conforming with the conditional independence structure represented by the graph.

Example 2.1. The two leftmost graphs in Figure 2.5 represent two marginal Gaussian graphical models, $Y_A \sim N_3(0, \Sigma)$, $\Sigma^{-1} \in S^+(G_A)$, $Y_B \sim N_2(0, \Phi)$, $\Phi^{-1} \in S^+(G_B)$. We imagine that the graphical models represent information of two studies acquired from two laboratories. The studies have some, but not all, of the variables in common, and it is of interest to find and estimate the parameters of a model that combines the initial pieces of information in the best way. Each graphical model represents a family of probability

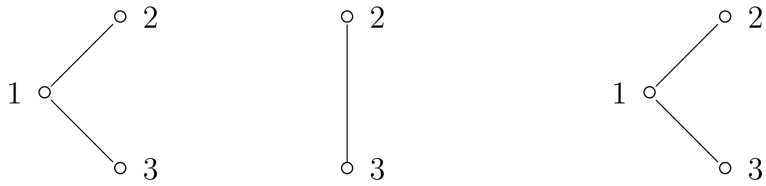


Figure 2.5: From left to right, graphs G_A and G_B and a joint graph G compatible with them.

distributions. For a simple combination of them to make sense, we have to ensure that at least one pair of distributions exist within the families which induce the same distribution over the common variables. It can be easily shown that, in this case, the second graphical model is the marginal family of the first one. Hence, we may simply take the joint family to be equal to the family determined by the leftmost graph of the figure, now represented by the rightmost graph in Figure 2.5. The next question that we would then ask is how to estimate the parameters of the joint family given the observations or the estimates of the parameters in each of the initial graphical models.

Example 2.2. The left-hand side of Figure 2.6 represents a graphical Gaussian model $Y_A \sim N_2(0, \Sigma)$, $\Sigma^{-1} \in S^+(G_A)$, in which the variables are independent. On the right-hand side, another graphical model is displayed in which the same variables are dependent. The two families are defined exactly on the same variables, and the family on the left is a subset of the family on the right, for the structure of the two concentration matrices. Again, the interest is in finding and estimating the parameters of a combination of the families.

Example 2.3. So far, we have considered examples in which there is a particular simple relation between the families, i.e., one family is a subset or the marginal of the other family. Figure 2.7 presents two Gaussian graphical models, $Y_A \sim N_3(0, \Sigma)$, $\Sigma^{-1} \in S^+(G_A)$ and $Y_B \sim N_3(0, \Phi)$, $\Phi^{-1} \in S^+(G_B)$ that are related only through the variables Y_2 and Y_3 and involve two different conditional independence relationships. Here, the definition of the joint graph is not obvious and it is not naturally given by the graph

Figure 2.6: From left to right, graphs G_A and G_B.

obtained as the union of the two graphs, as this would represent very different conditional independence restrictions.

Figure 2.7: Two marginal graphs.

Example 2.4. Yet another example is shown in Figure 2.8. The pair of graphs on the left-hand side represents two marginal graphical models $Y_A \sim N_3(0, \Sigma)$, $\Sigma^{-1} \in S^+(G_A)$, $Y_B \sim N_3(0, \Phi)$, $\Phi^{-1} \in S^+(G_B)$. The family shown on the right is not necessarily the joint

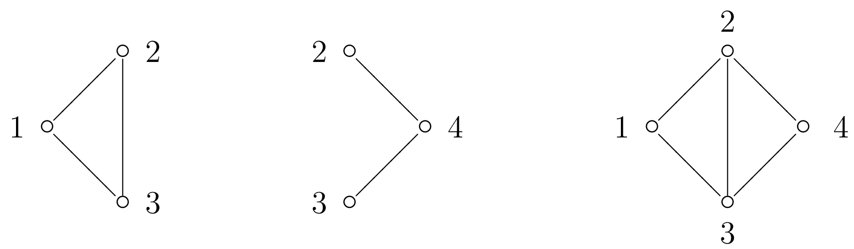


Figure 2.8: Two marginal graphs on the left, and a joint graph not compatible with them.

family, because the graph in the middle does not represent the marginal family of the graph on the right-hand side. As in the previous example, the determination of the joint graph is not straightforward. Note that not all types of combination are graphical. In some cases, it is not possible to represent the combination by a graphical model, unless we extend or reduce the initial models.

Example 2.5. In Figure 2.9, there are no conditional independence relationships expressed by the two graphs and one possible graph for the combined model is given by

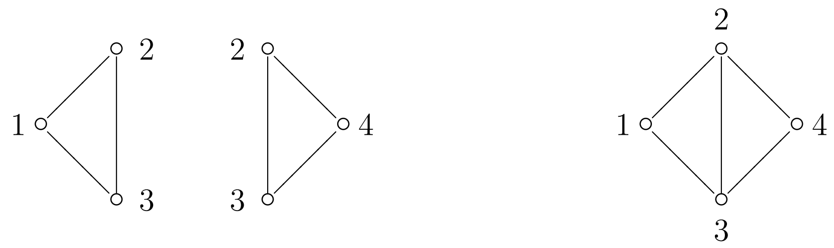


Figure 2.9: On the left, the marginal graphs. On the right, the joint graph.

the union of the two marginal graphs. Nevertheless, there are four different graphs which are compatible with the independence structure of the initial ones, all containing the cycle $(1, 2, 3, 4)$ but differing by the presence or absence of the edges $(2, 3)$ and $(1, 4)$. The idea of considering all possible graphical structures consistent with the given marginal ones is exploited in Kim (2001). We shall present an approach for combining graphical models that chooses the simplest graph which is compatible with the initial graphs.

We conclude this section by summarising the issues emerging from the above examples. Combining two graphical Gaussian models involves several steps. Firstly, it is important to query the establishment of the initial families. If they have been established on the basis of well-founded external knowledge (for example physical laws), it is generally assumed that they are both accurate and we should look for a joint model able to include both of them. On the other hand, if they have been empirically established, we consider the possibility of modifying them in order to make the combination possible. In general, we shall assume that the families are well established and we shall propose a modification of them only for the cases in which a joint graphical model does not exist, i.e., there is no graph corresponding to the joint family.

In order to combine two or more graphical models, we need to require some form of *consistency*, i.e. they must in some way specify the same probability law specified over the variables that they have in common.

The proper definition and visualization of the joint graph is also essential. In some cases, there are many joint graphs compatible with the given marginal ones, so that a method to choose between different options is needed. Our idea is to keep the joint graph as simple as possible, according to the initial information. We shall select the joint graph by introducing and classifying two types of combinations depending on the properties of the initial families.

Once the consistency assumptions and the joint graph are found, and a joint family of distributions is defined over it, the estimation of the parameters of the joint distribution has to be addressed. In a Gaussian context, this is equivalent to finding the estimate of the concentration matrix, coherently with the structure of the graph.

Chapter 3

A general framework

In this chapter, we shall describe a general framework for discussing compatibility of families of distributions with respect to the problem of combination of the families. Moreover, we shall present two types of family combination and we shall study their properties through several clarifying examples.

3.1 Consistency assumptions

Consider two families of distributions defined over different sets of variables. Note that we restrict our attention to distributions which are absolutely continuous w.r.t. a fixed product measure, and shall use the term distribution synonymous with the term density. We shall not consider cases in which the densities are not defined.

The notion of consistency is essential for combining two families and our intent is to verify that at least one distribution of a chosen family is compatible with at least one distribution of the other family, in the sense that they induce the same distribution over the variables in common. We distinguish two types of consistency, i.e. meta-consistency, as defined in Dawid and Lauritzen (1993) and quasi-consistency, to be defined below.
We start by giving the definition of consistency for two distributions defined over two sets, exploiting Dawid and Lauritzen (1993). They introduced it with the purpose of construction of a Markov distribution over a decomposable graph, when two distributions over two subsets of the same graph are given.

Consider two sets of variables, A and B. For a family $\mathcal{F} = \{f \,|\, f \text{ distribution over } A\}$, $\mathcal{F}^{\downarrow A\cap B}$ shall denote the family of marginal distributions induced by $\mathcal{F}$ over $A \cap B$. If $A \cap B = \emptyset$, $\mathcal{F}^{\downarrow A\cap B}$ is trivial and has only one constant element.

Definition 3.1 (Dawid and Lauritzen)**.** Two distributions f and g for random variables Y_A and Y_B, f over A and g over B, are said to be *consistent* if $f_{A\cap B} = g_{A\cap B}$.

This defininition is now extended to families as follows.

Definition 3.2 (Dawid and Lauritzen)**.** Two families of distributions, $\mathcal{F} = \{f |\, f \,\text{distribution over}\; A\}$ and $\mathcal{G} = \{g |\, g \,\text{distribution over}\; B\}$, for random variables Y_A and Y_B, are said to be *meta-consistent* if $\mathcal{F}^{\downarrow A\cap B} = \mathcal{G}^{\downarrow A\cap B}$.

The condition for meta-consistency can equivalently be expressed as:

$$\forall f \in \mathcal{F}, \exists\, g \in \mathcal{G}, \text{ such that } f_{A\cap B} = g_{A\cap B} \text{ and } \textit{vice versa.}$$

Note that Definition 3.1 and 3.2 can also be extended to more than two distributions or more than two families of distributions, and they do not require any Markov property.

The notion of meta-consistency of two families corresponds to the situation when the two marginal families representing different studies are identical, which, clearly, is very restrictive. Next, we introduce the less restrictive condition of quasi-consistency to study families that do not specify exactly the same set of distributions over the variables in the intersection, but have at least one distribution in common. Since a family of distributions represents the scientific assumption that the generating distribution is some member of the family, quasi-consistency reflects the situation where the assumptions associated with the two independent experiments are not in conflict, although they are not identical.

Definition 3.3. Two families of distributions $\mathcal{F} = \{f |\, f \,\text{distribution over}\; A\}$ and $\mathcal{G} = \{g |\, g \,\text{distribution over}\; B\}$ for random variables Y_A and Y_B are said to be *quasi-consistent* if $(\mathcal{F}^{\downarrow A\cap B}) \cap (\mathcal{G}^{\downarrow A\cap B}) \neq \emptyset$.

In other words, $\mathcal{F}$ and $\mathcal{G}$ are quasi-consistent if there are at least one consistent pair of elements (f, g) with $f \in \mathcal{F}$ and $g \in \mathcal{G}$ or, equivalently, if

$$\exists\, f \in \mathcal{F}, \exists\, g \in \mathcal{G}, \text{ such that } f_{A\cap B} = g_{A\cap B}.$$

Meta-consistency implies quasi-consistency but the converse is not generally true.

Proposition 3.1. *If $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent, then $\mathcal{F}$ and $\mathcal{G}$ are quasi-consistent.*

Proof. This follows directly from Definition 3.2 and Definition 3.3. □

Below, we give three simple examples of pairs of families that are meta-consistent, quasi-consistent, and neither of the two.

Example 3.1. Consider two families of distributions defined as follows.

$$\mathcal{F} = \left\{ \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \right), \quad \sigma_x^2 > 0,\, \sigma_y^2 > 0 \right\},$$

$$\mathcal{G} = \left\{ \begin{pmatrix} Y \\ Z \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda_y^2 & 0 \\ 0 & \lambda_z^2 \end{pmatrix} \right), \quad \lambda_y^2 > 0,\, \lambda_z^2 > 0 \right\}.$$

Here, $\mathcal{F}^{\downarrow Y} = \{Y \sim N(0, \sigma_y^2),\, \sigma_y^2 > 0\}$ and $\mathcal{G}^{\downarrow Y} = \{Y \sim N(0, \lambda_y^2),\, \lambda_y^2 > 0\}$. Therefore, $\mathcal{F}^{\downarrow Y} = \mathcal{G}^{\downarrow Y}$, i.e., $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent. For all $f \in \mathcal{F}$, there exists $g \in \mathcal{G}$ such that $f_Y = g_Y$. It is sufficient to take $\lambda_y^2 = \sigma_y^2$.

Example 3.2. Consider two families of distributions defined as follows.

$$\begin{aligned} \mathcal{F} &= \left\{ \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \right), \quad \sigma_x^2 > 0, \sigma_y^2 > 0 \right\}, \\ \mathcal{G} &= \left\{ \begin{pmatrix} Y \\ Z \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_z^2 \end{pmatrix} \right), \quad \sigma_z^2 > 0 \right\}. \end{aligned}$$

Here, $\mathcal{F}^{\downarrow Y} = \{Y \sim N(0, \sigma_y^2),\, \sigma_y^2 > 0\}$ and $\mathcal{G}^{\downarrow Y} = \{Y \sim N(0, 1)\}$. Therefore, $\mathcal{F}^{\downarrow Y} \cap \mathcal{G}^{\downarrow Y} \neq \emptyset$, i.e., $\mathcal{F}$ and $\mathcal{G}$ are quasi-consistent because we can take $\sigma_y^2 = 1$. However, they are not meta-consistent.

Example 3.3. Consider two families of distributions defined as follows.

$$\begin{aligned} \mathcal{F} &= \left\{ \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & 2 \end{pmatrix} \right), \quad \sigma_x^2 > 0 \right\}, \\ \mathcal{G} &= \left\{ \begin{pmatrix} Y \\ Z \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_z^2 \end{pmatrix} \right), \quad \sigma_z^2 > 0 \right\}. \end{aligned}$$

Here, $\mathcal{F}^{\downarrow Y} = \{Y \sim N(0, 2)\}$, $\mathcal{G}^{\downarrow Y} = \{Y \sim N(0, 1)\}$. Therefore, $\mathcal{F}$ and $\mathcal{G}$ are neither meta-consistent nor quasi-consistent.

In this work, we shall concentrate on families that are either meta-consistent or quasi-consistent.

3.2 Types of combination

The problem outlined in the introduction can be formalised as follows. Consider two sets A and B and two meta-consistent or quasi-consistent families of distributions, $\mathcal{F} =$

$\{f|\, f \,\text{distribution over}\, A\}$ and $\mathcal{G} = \{g|\, g \,\text{distribution over}\, B\}$ for Y_A and Y_B. We ideally search for a joint family of distributions, $\mathcal{H} = \{h|\, h \,\text{distribution over}\, A \cup B\}$ for $Y_{A\cup B}$, such that

$$\mathcal{H}^{\downarrow A} = \mathcal{F}, \quad \mathcal{H}^{\downarrow B} = \mathcal{G}, \tag{3.1}$$

where $\mathcal{H}^{\downarrow A}$ and $\mathcal{H}^{\downarrow B}$ are the families of induced marginal distributions of $\mathcal{H}$ over the sets A and B, respectively.

In Chapter 4, we shall study the problem (3.1) in the specific context of Gaussian graphical models; in what follows, we shall analyse the problem in the general context of meta-consistent and quasi-consistent families.

3.2.1 Meta-Markov combination

Different approaches have been proposed in the literature to tackle similar problems. Jiroušek and Vejnarova (2003), with the aim of representing a multivariate distribution, propose an iterative application of operators of composition called *generating sequences*, by which a multidimensional distribution is composed from a sequence of low dimensional ones. To this purpose, they define the operator of right composition as $f \triangleright g = \frac{f\cdot g}{g_{A\cap B}}$, and the operator of left composition as $f \triangleleft g = \frac{f\cdot g}{f_{A\cap B}}$. Furthermore, they show that under particular assumptions (if the sequence is perfect), the distribution defined by a generating sequence preserves all the given marginal distributions. Note that they do not consider families of distributions.

If $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent families of distributions, Dawid and Lauritzen (1993) define the *meta-Markov combination*, as follows.

Definition 3.4 (Dawid and Lauritzen)**.** The *meta-Markov combination* of $\mathcal{F}$ and $\mathcal{G}$ is defined as

$$\mathcal{F} \star \mathcal{G} = \{f \star g, f \in \mathcal{F}, g \in \mathcal{G}, f \,\text{and}\, g \,\text{consistent}\},$$

where $f \star g = \frac{f\cdot g}{g_{A\cap B}}$ is the *Markov combination* of distributions f and g.

Proposition 3.2. *The Markov combination $f \star g$ satisfies the following relations*

$$A \perp\!\!\!\perp B \,|\, (A \cap B), \quad (f \star g)^{\downarrow A} = f, \quad (f \star g)^{\downarrow B} = g. \tag{3.2}$$

Proof. This follows directly from Definition 3.4, see also Dawid and Lauritzen (1993). □

Note that we study families of distributions and, therefore, we need the Markov combination only to define the meta-Markov combination. However, it is worth noting that the Markov combination of f and g is the simplest possible, in the sense that it has maximal entropy among all distributions with the given marginals.

Theorem 3.1. *$H(f \star g) \geq H(h), \forall h \in \mathcal{Q}$, where* $\mathcal{Q} = \{h | h_A = f, h_B = g\}$.

Proof. This follows from the calculation below, which establishes and exploits the well-known fact that the entropy is a submodular function on the subsets of a finite set so that

$$H(A) + H(B) \geq H(A \cap B) + H(A \cup B),$$

with equality if and only if A and B are conditionally independent given $A \cap B$ (Studený, 2004).

Consider $h(x, y, z) = h(x, y)h(z|x, y)$, where $A = \{x, y\}$, $B = \{y, z\}$, $A \cap B = \{y\}$. Then,

$$\begin{aligned}
H(h(x, y, z)) &= E(-\log h(X, Y, Z)), \\
&= E\left(-\log h(X, Y)\right) + E\left(-\log h(Z|X, Y)\right), \\
&= H(h(x, y)) + H(h(z|x, y)), \\
&\leq H(h(x, y)) + H(h(z|y)), \\
&= H(f) + H(g) - H(g_{A \cap B}), \\
&= H(f \star g).
\end{aligned}$$

□

The meta-Markov combination is built by considering all the pairs of consistent distributions (f, g), $f \in \mathcal{F}$, $g \in \mathcal{G}$. If applied to meta-consistent families of distributions, the meta-Markov combination preserves the marginal families.

Proposition 3.3 (Dawid and Lauritzen)**.** *If $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent, $(\mathcal{F} \star \mathcal{G})^{\downarrow A} = \mathcal{F}$ and $(\mathcal{F} \star \mathcal{G})^{\downarrow B} = \mathcal{G}$.*

Proof. It follows from Definition 3.3 and from the construction of the Markov combination. □

It is important to stress that the operators of left and right composition for two distributions introduced by Jiroušek and Vejnarova (2003) are equivalent to the Markov combination when the two distributions are consistent.

Example 3.1 (continued)**.** In this case, the meta-Markov combination is given by

$$\mathcal{F} \star \mathcal{G} = \frac{\exp\left\{-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\lambda_z^2})\right\}}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\lambda_z},$$

for all $f \in \mathcal{F}$ and $g \in \mathcal{G}$ pairwise consistent, i.e. such that $\sigma_y^2 = \lambda_y^2$. It holds that $(\mathcal{F} \star \mathcal{G})^{\downarrow(X,Y)} = \mathcal{F}$ and $(\mathcal{F} \star \mathcal{G})^{\downarrow(Y,Z)} = \mathcal{G}$.

Notice that the meta-Markov combination can be used also when $\mathcal{F}$ and $\mathcal{G}$ are quasi-consistent families of distributions. In this case, it corresponds to the smallest set of meta-Markov combinations among the two families because, by definition, it associates only those $f \in \mathcal{F}$ and $g \in \mathcal{G}$ that are consistent. In general, when used for quasi-consistent families, the meta-Markov combination reduces the marginal families.

Proposition 3.4. *If $\mathcal{F}$ and $\mathcal{G}$ are quasi-consistent, $(\mathcal{F} \star \mathcal{G})^{\downarrow A} \subseteq \mathcal{F}$ and $(\mathcal{F} \star \mathcal{G})^{\downarrow B} \subseteq \mathcal{G}$. Further, if $\mathcal{F} \subset \mathcal{G}$ then $\mathcal{F} \star \mathcal{G} = \mathcal{F}$.*

Proof. This follows directly from (3.2). □

The following example uses the meta-Markov combination for quasi-consistent families of distributions.

Example 3.2 (continued)**.** In this case, the meta-Markov combination is given by

$$\mathcal{F} \star \mathcal{G} = \frac{\exp\left\{-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + y^2 + \frac{z^2}{\sigma_z^2})\right\}}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_z},$$

with $(\mathcal{F} \star \mathcal{G})^{\downarrow(X,Y)} \subset \mathcal{F}$ and $(\mathcal{F} \star \mathcal{G})^{\downarrow(Y,Z)} = \mathcal{G}$.

3.2.2 Quasi-Markov combination

As we have seen in the previous section, it is possible to use the meta-Markov combination for both meta-consistent and quasi-consistent families. In these cases, the marginals of the combined family are contained in the original marginal families. When the restrictions in the marginal families are less well established, it is relevant to consider a less restrictive form of combination obtained by preserving one marginal and modifying the other one. We introduce the notion of quasi-Markov combination to reflect this fact.

Consider two meta-consistent or quasi-consistent families of distributions $\mathcal{F}$ and $\mathcal{G}$, as described at the beginning of Section 3.2. The idea is to combine every marginal of f from $\mathcal{F}$ (f-marginal) with every $g_{B|A\cap B}$ for $g \in \mathcal{G}$ (g-conditional) or *vice versa.*

Definition 3.5. The *quasi-Markov combination* of $\mathcal{F}$ and $\mathcal{G}$ is defined as

$$\mathcal{F} \,\overline{\star}\, \mathcal{G} = \{f_{A|A\cap B} \cdot g, f \in \mathcal{F}, g \in \mathcal{G}\} \cup \{f \cdot g_{B|A\cap B}, f \in \mathcal{F}, g \in \mathcal{G}\}.$$

Note that f and g are not required to be consistent as it was for the meta-Markov combination (see Definition 3.4). For the cases in which this happens, the quasi-Markov combination is the same as the meta-Markov combination. In general, the quasi-Markov combination extends the marginal families.

Proposition 3.5. *The quasi-Markov combination implies that* $(\mathcal{F} \,\overline{\star}\, \mathcal{G})^{\downarrow A} \supseteq \mathcal{F}$ *and* $(\mathcal{F} \,\overline{\star}\, \mathcal{G})^{\downarrow B} \supseteq \mathcal{G}$. *Moreover, if* $\mathcal{F} \subset \mathcal{G}$ *then* $\mathcal{F} \,\overline{\star}\, \mathcal{G} = \mathcal{G}$.

Proof. From the definition, it follows that $(\mathcal{F} \,\overline{\star}\, \mathcal{G})^{\downarrow A} = \mathcal{F} \cup \{f_{A|A\cap B} \cdot g, f \in \mathcal{F}, g \in \mathcal{G}\}^{\downarrow A}$ and $(\mathcal{F} \,\overline{\star}\, \mathcal{G})^{\downarrow B} = \mathcal{G} \cup \{f \cdot g_{B|A\cap B}, f \in \mathcal{F}, g \in \mathcal{G}\}^{\downarrow B}$. □

In general, the meta-Markov combination defines a smaller family than the quasi-Markov combination. In fact, the meta-Markov combination is defined by associating only consistent distributions, whereas the quasi-Markov combination is defined by associating all distributions of the two families after having fixed one family.

Proposition 3.6. *It holds that* $\mathcal{F} \star \mathcal{G} \subseteq \mathcal{F} \,\overline{\star}\, \mathcal{G}$.

Proof. The meta-Markov combination is the family of distributions $\mathcal{F} \star \mathcal{G} = \{f \cdot g_{B|A\cap B} = g \cdot f_{A|B\cap A}, f \in \mathcal{F}, g \in \mathcal{G}, f \text{ and } g \text{ consistent}\}$, where $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent or quasi-consistent. □

Proposition 3.7. *If* $\mathcal{F}$ *and* $\mathcal{G}$ *are meta-consistent, it further holds that* $\mathcal{F} \star \mathcal{G} = \mathcal{F} \,\overline{\star}\, \mathcal{G}$ *if and only if* $(\mathcal{F} \,\overline{\star}\, \mathcal{G})^{\downarrow A} = \mathcal{F}$ *and* $(\mathcal{F} \,\overline{\star}\, \mathcal{G})^{\downarrow B} = \mathcal{G}$.

Proof. It follows directly from Proposition 3.3 that, if $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent, $\mathcal{F} \star \mathcal{G} = \mathcal{F} \,\overline{\star}\, \mathcal{G}$ implies $(\mathcal{F} \,\overline{\star}\, \mathcal{G})^{\downarrow A} = \mathcal{F}$ and $(\mathcal{F} \,\overline{\star}\, \mathcal{G})^{\downarrow B} = \mathcal{G}$. If $(\mathcal{F} \,\overline{\star}\, \mathcal{G})^{\downarrow A} = \mathcal{F}$ then $\{f_{A\,|\,A\cap B} \cdot g, f \in \mathcal{F}, g \in \mathcal{G}\}^{\downarrow A} \subseteq \mathcal{F}$, in other words we must for all $f \in \mathcal{F}, g \in \mathcal{G}$ have that

$$\left(\frac{f \cdot g}{f_{A\cap B}}\right)^{\downarrow A} = \frac{f}{f_{A\cap B}} \cdot g_{A\cap B} = f^*$$

for some $f^* \in \mathcal{F}$. But then we have $f^*_{A\cap B} = g_{A\cap B}$ and thus $f_{A\,|\,A\cap B} \cdot g = f^* \star g$. Repeating the argument with f and g reversed yields that $\mathcal{F} \star \mathcal{G} = \mathcal{F} \,\overline{\star}\, \mathcal{G}$. □

In Proposition 3.7, we gave a necessary and sufficient condition for the equivalence of meta-Markov and quasi-Markov combination for meta-consistent families. We are in the same situation also when $Y_{A\cap B}$ is a cut.

Definition 3.6. $Y_{A\cap B}$ is a cut in $\mathcal{F}$ if $\mathcal{F} = \mathcal{F}^{\downarrow(A|A\cap B)} \times \mathcal{F}^{\downarrow A\cap B}$ or equivalently $A|A\cap B$ and $A\cap B$ are variation independent, $(A|A\cap B)\ddagger(A\cap B)[\mathcal{F}]$.

Proposition 3.8. *If $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent and $Y_{A\cap B}$ is a cut for both $\mathcal{F}$ and $\mathcal{G}$, $\mathcal{F} \star \mathcal{G} = \mathcal{F} \,\overline{\star}\, \mathcal{G}$.*

Proof. $Y_{A\cap B}$ is a cut for $\mathcal{F}$ implies that $\mathcal{F} = \mathcal{F}^{\downarrow(A|A\cap B)} \times \mathcal{F}^{\downarrow A\cap B}$. $Y_{A\cap B}$ is a cut for $\mathcal{G}$ implies that $\mathcal{G} = \mathcal{G}^{\downarrow(B|A\cap B)} \times \mathcal{G}^{\downarrow A\cap B}$. Now,

$$\begin{aligned}
\mathcal{F} \,\overline{\star}\, \mathcal{G} &= \{f \cdot g_{B|A\cap B}, f \in \mathcal{F}, g \in \mathcal{G}\} \cup \{f_{A|A\cap B} \cdot g, f \in \mathcal{F}, g \in \mathcal{G}\}, \\
&= \{f_{A|A\cap B} \cdot f_{A\cap B} \cdot g_{B|A\cap B}, f \in \mathcal{F}, g \in \mathcal{G}\} \cup \{f_{A|A\cap B} \cdot g_{A\cap B} \cdot g_{B|A\cap B}, f \in \mathcal{F}, g \in \mathcal{G}\}, \\
&= \{f_{A|A\cap B} \cdot f_{A\cap B} \cdot g_{B|A\cap B}, f \in \mathcal{F}, g \in \mathcal{G}\}, \\
&= \left\{\frac{f \cdot g}{g_{A\cap B}}, f \in \mathcal{F}, g \in \mathcal{G}\right\}, \\
&= \mathcal{F} \star \mathcal{G}.
\end{aligned}$$

The cut was used in the second line. Then, we exploited the fact that f and g are consistent under meta-consistency, that is to say, $f_{A\cap B} = g_{A\cap B}$. □

We conclude the section by presenting some examples of the quasi-Markov combination. In the first two examples, we want to stress that the quasi-Markov combination can be used for both meta-consistent and quasi-consistent families.

Example 3.1 (continued)**.** The families are meta-consistent and the quasi-Markov combination is given by

$$\mathcal{F} \,\overline{\star}\, \mathcal{G} = \left\{\frac{\exp\left\{-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\lambda_z^2})\right\}}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_y\lambda_y}\right\} \cup \left\{\frac{\exp\left\{-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\lambda_y^2} + \frac{z^2}{\lambda_z^2})\right\}}{(2\pi)^{\frac{3}{2}}\sigma_x\lambda_y\lambda_z}\right\}.$$

Example 3.2 (continued)**.** In this case, the families are quasi-consistent and the quasi-Markov combination is given by

$$\begin{aligned}
\mathcal{F} \,\overline{\star}\, \mathcal{G} &= \left\{\frac{\exp\left\{-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + y^2 + \frac{z^2}{\sigma_z^2})\right\}}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_z}\right\} \cup \left\{\frac{\exp\left\{-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2})\right\}}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_z\sigma_y}\right\}, \\
&= \frac{\exp\left\{-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2})\right\}}{(2\pi)^{\frac{3}{2}}\sigma_x\sigma_z\sigma_y}.
\end{aligned}$$

Example 3.4. Consider the case in which $\mathcal{F}$ and $\mathcal{G}$ consist of one element each, i.e., $\mathcal{F} = \{f^1\}$ and $\mathcal{G} = \{g^1\}$. This can be the case of $\mathcal{F}$ and $\mathcal{G}$ families that contain exactly one multivariate normal distribution defined over A and B with $A = \{X, Y, Z\}$, $B = \{Y, T\}$, respectively. If $f^1_{A\cap B} = g^1_{A\cap B}$, the families are meta-consistent and the meta-Markov combination is the singleton family

$$\mathcal{F} \star \mathcal{G} = \left\{f^1 \star g^1\right\} = \left\{\frac{f^1 \cdot g^1}{f^1_{A\cap B}}\right\}.$$

If $f^1_{A\cap B} \neq g^1_{A\cap B}$, the families are neither meta-consistent nor quasi-consistent because in the definition of quasi-consistency we require at least two distributions to be consistent (see Definition 3.3), so we do not define any type of combination, corresponding to the fact that the two models are mutually contradictory.

Example 3.5. Consider the case in which $\mathcal{F} = \{f^1, f^2\}$ and $\mathcal{G} = \{g\}$. If $f^1_{A\cap B} = f^2_{A\cap B} = g_{A\cap B}$, the families are meta-consistent. The meta-Markov combination is $\mathcal{F}\star\mathcal{G} = \left\{\frac{f^1\cdot g}{g_{A\cap B}}, \frac{f^2\cdot g}{g_{A\cap B}}\right\}$. If $f^1_{A\cap B} = g_{A\cap B}$, and $f^2_{A\cap B} \neq g_{A\cap B}$, $\mathcal{F}$ and $\mathcal{G}$ are quasi-consistent. The meta-Markov combination is given by $\mathcal{F} \star \mathcal{G} = \frac{f^1\cdot g}{g_{A\cap B}}$. The quasi-Markov combination is given by $\mathcal{F} \barwedge\!\!\star\, \mathcal{G} = \left\{f^1 \cdot g_{B|A\cap B}, f^2 \cdot g_{B|A\cap B}\right\} \cup \left\{f^1_{A|A\cap B} \cdot g, f^2_{A|A\cap B} \cdot g\right\}$.

Chapter 4

Combination of Gaussian graphical models

We now turn to the core of this work, namely, combining information from Gaussian graphical models. To this aim, we consider two undirected graphs $G_A = (A, E_A)$ and $G_B = (B, E_B)$ and two families of multivariate normal distributions $\mathcal{F}$ and $\mathcal{G}$, conforming to the structure of the graphs. G_A and G_B represent the graphical models $\mathcal{F}$ and $\mathcal{G}$, and both the families are defined as the set of Gaussian distributions which satisfy the conditional independence relations of the graphs. We search for a joint undirected graph $G = (V, E)$ with vertex set $V = A \cup B$, and a family of multivariate normal distributions $\mathcal{H}$ defined over V such that $\mathcal{H}^{\downarrow A} = \mathcal{F}$, $\mathcal{H}^{\downarrow B} = \mathcal{G}$ and all distributions in $\mathcal{H}$ conform with the independence properties of the marginal models.

In this framework, we consider the two types of combination already introduced in Chapter 3: the meta-Markov combination, $\mathcal{F} \star \mathcal{G}$, and the quasi-Markov combination, $\mathcal{F} \overline{\star} \mathcal{G}$, defined for both meta-consistent and quasi-consistent families. The combination is established by following the previously given results for families of distributions. Consequently, it is important to identify whether the families are quasi-consistent or meta-consistent, and to decide which combination is appropriate. When $Y_{A\cap B}$ is a cut in the original meta-consistent families, and this happens if and only if the corresponding graphs are collapsible onto $A \cap B$ (Frydenberg, 1990), from the general theory of Chapter 3 we see that the meta-Markov combination is the same as the quasi-Markov combination (see Proposition 3.8).

Proposition 4.1. *If two meta-consistent graphical Gaussian models $\mathcal{F}$ and $\mathcal{G}$ have both graphs collapsible onto $A \cap B$, then $\mathcal{F} \star \mathcal{G} = \mathcal{F} \overline{\star} \mathcal{G}$.*

Proof. It is the equivalent of Proposition 3.8 in the Gaussian case. □

Once we have defined the characteristics of the families and we have chosen the appropriate combination, the latter necessarily determines the joint family of distributions. Thus, the problem is how to determine the graph corresponding to the combination. Firstly, we note that the dependence graph of a family $\mathcal{F}$ is always well defined as the graph with no edge between α and β (written in the following as $\alpha \not\sim \beta$) if and only if $\alpha \perp\!\!\!\perp \beta \,|\, V \setminus \{\alpha, \beta\}$, for all $f \in \mathcal{F}$. It is the smallest graph such that all the distributions of the family are pairwise Markov with respect to it (see the Appendix for more details). Therefore, we can always define the dependence graph of a combination of two families. It is defined as the graph with $\alpha \not\sim \beta$ if and only if $\alpha \perp\!\!\!\perp \beta \,|\, V \setminus \{\alpha, \beta\}$, for all $h \in \mathcal{F} \star \mathcal{G}$ or $h \in \mathcal{F} \,\overline{\star}\, \mathcal{G}$.
We denote with $G(\mathcal{F} \star \mathcal{G})$ the dependence graph of the combined family $\mathcal{F} \star \mathcal{G}$, and with $G(\mathcal{F} \,\overline{\star}\, \mathcal{G})$ the dependence graph of the combined family $\mathcal{F} \,\overline{\star}\, \mathcal{G}$. Further, we say that a combination is graphical if the Gaussian family corresponding to the dependence graph of the combined family is equal to the combined family itself.

Definition 4.1. A meta-Markov combination $\mathcal{F} \star \mathcal{G}$ is said to be a *graphical meta-Markov combination*, if the Gaussian family corresponding to $G(\mathcal{F} \star \mathcal{G})$ is equal to the family $\mathcal{F} \star \mathcal{G}$.

Definition 4.2. A quasi-Markov combination $\mathcal{F} \,\overline{\star}\, \mathcal{G}$ is said to be a *graphical quasi-Markov combination*, if the Gaussian family corresponding to $G(\mathcal{F} \,\overline{\star}\, \mathcal{G})$ is equal to the family $\mathcal{F} \,\overline{\star}\, \mathcal{G}$.

If a combination is graphical, it defines the graphical structure of the joint graph and the multivariate normal distribution over it. The vertex set is the union of the vertex sets of the marginal graphs, and the edges set takes into account the conditional independence relationships and the graphical structure of the two families. In addition, the joint graph obtained by considering the two marginal graphs corresponds to the dependence graph derived from the combination. This leads to the definition of the meta-Markov combination of two marginal graphs. We define it as the dependence graph corresponding to the meta-Markov combination of the families.

Definition 4.3. Consider two graphs G_A and G_B, corresponding to the meta-consistent families $\mathcal{F}$ and $\mathcal{G}$, respectively. The dependence graph $G_A \star G_B = G(\mathcal{F} \star \mathcal{G})$ is said to be the *meta-Markov combination* of G_A and G_B.

Note that if the combination is graphical, the Gaussian family corresponding to $G_A \star G_B$ is equal to the family $\mathcal{F} \star \mathcal{G}$. A similar definition is given for quasi-consistent families.

Definition 4.4. Consider two graphs G_A and G_B, corresponding to the quasi-consistent families $\mathcal{F}$ and $\mathcal{G}$, respectively. The dependence graph $G_A \mathbin{\overline{\star}} G_B = G(\mathcal{F} \mathbin{\overline{\star}} \mathcal{G})$ is said to be the *quasi-Markov combination* of G_A and G_B.

As before, if the combination is graphical, the Gaussian family corresponding to $G_A \mathbin{\overline{\star}} G_B$ is equal to the family $\mathcal{F} \mathbin{\overline{\star}} \mathcal{G}$.

Another type of situation occurs when the joint set of distributions exists though it is not a graphical model, i.e., when the Gaussian family corresponding to $G(\mathcal{F} \star \mathcal{G})$ (or $G(\mathcal{F} \mathbin{\overline{\star}} \mathcal{G})$) is not equal to the family $\mathcal{F} \star \mathcal{G}$ (or $\mathcal{F} \mathbin{\overline{\star}} \mathcal{G}$). We call this combination a *non-graphical combination*. For example, the joint family might correspond to the union of two graphical models. We shall see that in some circumstances this type of combination may be studied with the tools of graphical models. In general, it might always be investigated from an algebraic point of view.

To avoid non-graphical combinations, one might wish to define a combination of two Gaussian graphical models as the Gaussian family with dependence graph $G(\mathcal{F} \star \mathcal{G})$ or $G(\mathcal{F} \mathbin{\overline{\star}} \mathcal{G})$. In this way, a graphical combination of two graphical models is always defined, but it is bigger than the meta-Markov or quasi-Markov combination of the families. It is equivalent to consider the smallest graphical model that is not contradicting the marginal families, and it is obtained by modifying the original families through addition or deletion of edges from one or both the corresponding graphs.

We prefer the first proposed approach with the distinction between graphical and non-graphical combinations. Now, we conclude the section providing some examples of these two types of combinations.

4.1 Graphical combinations

Example 4.1. We represent the two graphical models of Figure 4.1 as the families $\mathcal{F} = \{Y_A \sim N_3(0, \Sigma), \Sigma^{-1} \in S^+(G_A)\}$ and $\mathcal{G} = \{Y_B \sim N_2(0, \Phi), \Phi^{-1} \in S^+(G_B)\}$. $\mathcal{F}^{\downarrow(Y_2,Y_3)} = \{Y \sim N_2(0, \Sigma_{\{2,3\}}), \Sigma^{-1}_{\{2,3\}} \in S^+(G_B)\}$, $\mathcal{G}^{\downarrow(Y_2,Y_3)} = \mathcal{G}$. The families $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent because $\mathcal{F}^{\downarrow(Y_2,Y_3)} = \mathcal{G}^{\downarrow(Y_2,Y_3)}$. The meta-Markov combination is the family $\mathcal{F} \star \mathcal{G} = \left\{Y \sim N_3(0, \Sigma), \Sigma^{-1} \in S^+(G_A), \Sigma_{\{2,3\}} = \Phi_{\{2,3\}}\right\}$, and it is represented by the graph on the left. The quasi-Markov combination is given by

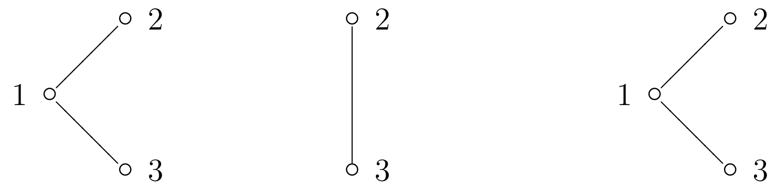


Figure 4.1: From left to right, graphs G_A and G_B and a joint graph G compatible with them.

$$\begin{aligned}\mathcal{F} \,\overline{\star}\, \mathcal{G} \quad \propto \quad & \exp\Big\{-\frac{1}{2}(y_1^2k_{11} + y_2^2h_{22} + y_3^2h_{33} + 2y_1y_2k_{12} + 2y_1y_3k_{13} + y_2y_3(h_{23} - k_{23}))\Big\} \cup \\ & \exp\Big\{-\frac{1}{2}(y_1^2k_{11} + y_2^2k_{22} + y_3^2k_{33} + 2y_1y_2k_{12} + 2y_1y_3k_{13})\Big\},\end{aligned}$$

where $\Sigma^{-1} = \{k_{ij}\}$, $\Phi^{-1} = \{h_{ij}\}$. The dependence graph associated with it is the complete graph. The quasi-Markov combination is a non-graphical combination because the Gaussian distribution associated with its dependence graph is not equal to $\mathcal{F} \,\overline{\star}\, \mathcal{G}$.

Example 4.2. We represent the two graphical models of Figure 4.2 as the families $\mathcal{F} = \{Y_A \sim N_2(0,\Sigma), \Sigma^{-1} \in S^+(G_A)\}$ and $\mathcal{G} = \{Y_B \sim N_2(0,\Phi), \Phi^{-1} \in S^+(G_B)\}$. $\mathcal{F}^{\downarrow(Y_2,Y_3)} = \mathcal{F}$, $\mathcal{G}^{\downarrow(Y_2,Y_3)} = \mathcal{G}$. The families $\mathcal{F}$ and $\mathcal{G}$ are quasi-consistent because $\mathcal{F}^{\downarrow(Y_2,Y_3)} \subset \mathcal{G}^{\downarrow(Y_2,Y_3)}$. The meta-Markov combination is built by considering all the consistent pairs $f \in \mathcal{F}$ and $g \in \mathcal{G}$. Therefore, it is the family $\mathcal{F} \star \mathcal{G} = \{Y \sim N_2(0,\Sigma), \Sigma^{-1} \in S^+(G_A)\}$, represented by the graph on the left. The quasi-Markov combination is the family $\mathcal{F} \,\overline{\star}\, \mathcal{G} = \{Y \sim N_2(0,\Sigma), \Sigma^{-1} \in S^+(G_B)\}$, represented by the graph on the right.

Figure 4.2: From left to right, graphs G_A and G_B.

Example 4.3. We represent the two graphical models of Figure 4.3 as the families $\mathcal{F} = \{Y_A \sim N_3(0,\Sigma), \Sigma > 0\}$ and $\mathcal{G} = \{Y_B \sim N_3(0,\Phi), \Phi > 0\}$. The notation $\Sigma > 0$ stands for Σ positive definite. We have that $\mathcal{F}^{\downarrow(Y_2,Y_3)} = \big\{Y \sim N_2(0,\Sigma_{\{2,3\}}), \Sigma_{\{2,3\}} > 0\big\}$, and $\mathcal{G}^{\downarrow(Y_2,Y_3)} = \big\{Y \sim N_2(0,\Phi_{\{2,3\}}), \Phi_{\{2,3\}} > 0\big\}$. $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent because $\mathcal{F}^{\downarrow(Y_2,Y_3)} = \mathcal{G}^{\downarrow(Y_2,Y_3)}$. The meta-Markov combination is the family of distributions represented by the graphical model obtained as the union of the two. It is equivalent to the quasi-Markov combination because the two graphs are both collapsible onto $\{2,3\}$.

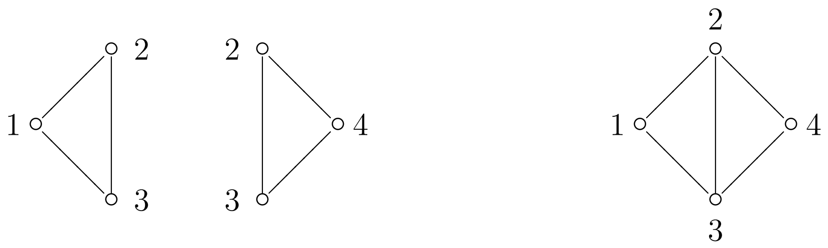


Figure 4.3: On the left, the marginal graphs. On the right, the joint graph.

Example 4.4. We represent the three graphical models of Figure 4.4 as the families $\mathcal{F} = \{Y_A \sim N_2(0, \Gamma), \Gamma^{-1} \in S^+(G_A)\}$, $\mathcal{G} = \{Y_B \sim N_3(0, \Omega), \Omega^{-1} \in S^+(G_B)\}$ and $\mathcal{H} = \{Y_A \sim N_2(0, \Phi), \Phi^{-1} \in S^+(G_C)\}$.

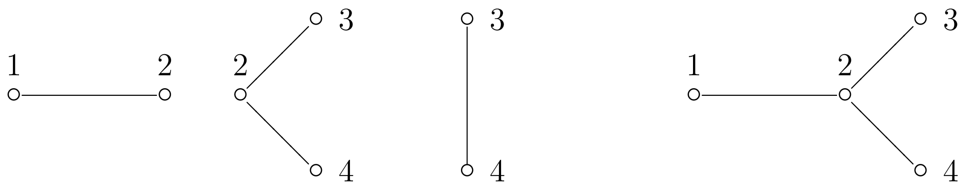


Figure 4.4: From left to right, graphs G_A, G_B, G_C, and a joint graph $G_{A\cup B\cup C}$ compatible with them.

On the one hand, we can combine families $\mathcal{F}$ and $\mathcal{G}$ and then combine the resulting family with $\mathcal{H}$. In this regard, $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent because $\mathcal{F}^{\downarrow Y_2} = \mathcal{G}^{\downarrow Y_2}$. Thus, the meta-Markov combination is $\mathcal{F} \star \mathcal{G} = \{Y \sim N_4(0, \Delta), \Delta^{-1} \in S^+(G_{A\cup B}), \Sigma_2 = \Phi_2\}$, represented by the graph $G_{A\cup B}$ obtained by merging the graphs G_A and G_B. Then $(\mathcal{F} \star \mathcal{G})$ and $\mathcal{H}$ are meta-consistent. The meta-Markov combination is

$$(\mathcal{F} \star \mathcal{G}) \star \mathcal{H} = \left\{Y \sim N_4(0, \Sigma), \Sigma^{-1} \in S^+(G_{A\cup B\cup C}), \Sigma_2 = \Phi_2, \Phi_{\{2,3\}} = \Omega_{\{2,3\}} \right\}.$$

On the other hand, we can combine families $\mathcal{G}$ and $\mathcal{H}$ and then combine the resulting family with $\mathcal{F}$. In this regard, $\mathcal{G}$ and $\mathcal{H}$ are meta-consistent (see Example 4.1). Thus, the meta-Markov combination is $\mathcal{G} \star \mathcal{H} = \left\{Y \sim N_3(0, \Omega), \Omega^{-1} \in S^+(G_B), \Phi_{\{2,3\}} = \Omega_{\{2,3\}}\right\}$, represented by the graph G_B. Then, $(\mathcal{G} \star \mathcal{H})$ and $\mathcal{F}$ are meta-consistent. The meta-Markov combination is

$$\mathcal{F} \star (\mathcal{G} \star \mathcal{H}) = \left\{Y \sim N_4(0, \Gamma), \quad \Gamma^{-1} \in S^+(G_{A\cup B\cup C}), \Sigma_2 = \Phi_2, \Phi_{\{2,3\}} = \Omega_{\{2,3\}}\right\},$$

represented by the graph $G_{A\cup B\cup C}$ obtained as the union of the three graphs. Note that, in this case, $(\mathcal{F} \star \mathcal{G}) \star \mathcal{H} = \mathcal{F} \star (\mathcal{G} \star \mathcal{H})$ and the described procedure permits to combine only one family at time.

4.2 Non-graphical combinations

Example 4.5. We represent the two graphical models of Figure 4.5 as the families $\mathcal{F} = \{Y_A \sim N_3(0,\Sigma), \Sigma^{-1} \in S^+(G_A)\}$ and $\mathcal{G} = \{Y_B \sim N_3(0,\Phi), \Phi^{-1} \in S^+(G_B)\}$. $\mathcal{F}^{\downarrow(Y_2,Y_3)} = \Big\{Y \sim N_2(0,\Sigma_{\{2,3\}}), \Sigma^{-1}_{\{2,3\}} \in S^+(G)\Big\}$, G is the graph with vertex set $\{2,3\}$ and edge $(2,3)$. $\mathcal{G}^{\downarrow(Y_2,Y_3)} = \Big\{Y \sim N_2(0,\Phi_{\{2,3\}}), \Phi^{-1}_{\{2,3\}} \in S^+(G)\Big\}$. $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent because $\mathcal{F}^{\downarrow(Y_2,Y_3)} = \mathcal{G}^{\downarrow(Y_2,Y_3)}$. In this case, the meta-Markov combination is not a graphical model and it is given by

$$\mathcal{F} \star \mathcal{G} = \Big\{Y \sim N_4(0,\Sigma), \Sigma_{\{2,3\}} = \Phi_{\{2,3\}},\, \sigma_{24}\sigma_{34}(\sigma_{44})^{-1} = \sigma_{13}\sigma_{12}(\sigma_{11})^{-1}\Big\},$$

where σ_{ij} are the elements of the covariance matrix Σ.

The quasi-Markov combination is given by:

$$\begin{aligned}\mathcal{F} \,\bar{\star}\, \mathcal{G} \;\propto\; & \exp\Big\{-\frac{1}{2}(y_1^2k_{11} + y_2^2h_{22} + y_3^2h_{33} + y_4^2h_{44} + 2y_1y_2k_{12} + 2y_1y_3k_{13} + \\ & 2y_2y_4h_{24} + 2y_3y_4h_{34} - 2y_2y_3k_{23})\Big\} \cup \\ & \exp\Big\{-\frac{1}{2}(y_1^2k_{11} + y_2^2k_{22} + y_3^2k_{33} + y_4^2h_{44} + 2y_1y_2k_{12} + 2y_1y_3k_{13} + \\ & 2y_2y_4h_{24} + 2y_3y_4h_{34} - 2y_2y_3h_{23})\Big\},\end{aligned}$$

where $\Sigma^{-1} = \{k_{ij}\}$, $\Phi^{-1} = \{h_{ij}\}$. It is a non-graphical combination and the dependence graph associated with it is the graph with vertex set $\{1,2,3,4\}$ and edges set $\{(1,2),(1,3),(2,3),(2,4),(3,4)\}$.

Figure 4.5: Two marginal graphs.

Example 4.6. We represent the two graphical models of Figure 4.6 as the families $\mathcal{F} = \{Y_A \sim N_3(0,\Sigma), \Sigma^{-1} \in S^+(G_A)\}$ and $\mathcal{G} = \{Y_B \sim N_2(0,\Phi), \Phi^{-1} \in S^+(G_B)\}$. $\mathcal{F}^{\downarrow(Y_2,Y_3)} = \Big\{Y \sim N_2(0,\Sigma_{\{2,3\}}), \Sigma^{-1}_{\{2,3\}} \in S^+(G)\Big\}$, and G is the graph with vertex set $\{2,3\}$ and edge $(2,3)$. $\mathcal{G}^{\downarrow(Y_2,Y_3)} = \mathcal{G}$. The families are quasi-consistent because $\mathcal{G}^{\downarrow(Y_2,Y_3)} \subset \mathcal{F}^{\downarrow(Y_2,Y_3)}$. The meta-Markov combination combines only consistent pairs but it is not a graphical model. It is given by $\mathcal{F} \star \mathcal{G} = \Big\{Y \sim N_3(0,\Sigma), \Sigma_{\{2,3\}} = \Phi_{\{2,3\}},\, \rho_{23} = \rho_{12}\rho_{13}, \rho_{23} = 0\Big\}$, where ρ_{ij}

are the elements of the correlation matrix of the joint family. It is the union of the graphical model with vertex set $\{1, 2, 3\}$ and edge $(1, 2)$ and the graphical model with vertex set $\{1, 2, 3\}$ and edge $(1, 3)$. The quasi-Markov combination combines all pairs and becomes the graphical model with the graph on the left in Figure 4.6.

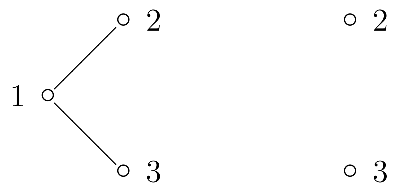


Figure 4.6: From left to right, graphs G_A and G_B.

Chapter 5

Maximum likelihood estimation

In the earlier chapters, we have defined the meta-Markov and quasi-Markov combinations and introduced the distinction between graphical and non-graphical combinations. Here, the emphasis is on the estimation of the parameters in each type of combination, graphical and non-graphical. To do this, we shall consider the problem from a missing data perspective.

5.1 Graphical combinations

Given two marginal families, a graphical combination has been defined in Chapter 4 as a joint family of distributions equal to the Gaussian family corresponding to the dependence graph of the combination. In this regard, either a meta-Markov combination or a quasi-Markov combination define the joint graphical Gaussian model assigning a multivariate normal distribution with concentration matrix having zero entries where the edges of the graph are missing. Thus, the next step is the estimation of the concentration matrix or its inverse, assuming that observations and (or) corresponding estimates are available for each of the marginal models. Where possible, we propose an approach that exploits all the available information and converts the problem in a missing data one, as it is explained in the remainder of this section.

The observed variables define a missing data pattern for the joint graph. In general, without conditional independence relationships (in a saturated model), a monotone missing data pattern would permit a factorisation of the likelihood and its estimate in separate pieces of complete-data likelihoods. For a non-monotone data pattern, iterative methods, such as the Expectation–Maximization (EM) algorithm, must generally be used (Little and Rubin, 2002). In large non-saturated models with general incomplete patterns, there are no straightforward factorisations that lead to a simplification of the estimation problem. Usually, one looks for a *lossless decomposition* of the graph or, if this is not the case, imputes missing data partially (Geng et al., 2000).

The examples investigated in this section involve small graphs and, therefore, we exploit the standard EM algorithm as detailed for mixed graphical models by Didelez and Pigeot (1998). However, the partial imputation EM algorithm (Geng et al., 2000) would be more efficient when dealing with high dimensional graphs and multiple combinations.

In our problem, the hypothetical complete data log-likelihood corresponds to an exponential family. Hence, the E-step computes the expected values of the complete data sufficient statistics conditioned on the observed data, and the M-step replaces the adjusted values found in the E-step in the likelihood equations for complete data. For decomposable joint graphs, the M-step is explicitly solved, while in the non-decomposable case it requires iterative methods like, for example, the Iterative Proportional Scaling (IPS) algorithm (Speed and Kiiveri, 1986). To avoid redundant iteration, only one cycle of the IPS algorithm needs to be performed in every M-step, as each iteration increases the likelihood (Lauritzen, 1996, pag. 135).

We describe our method in detail for both meta-consistent and quasi-consistent families. The examples involve decomposable and non-decomposable joint graphs and we show also cases in which the missing data perspective is not required.

5.1.1 Meta-consistent families

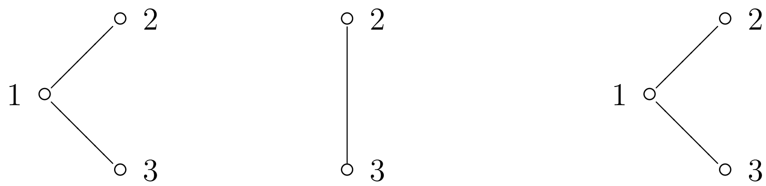


Figure 5.1: From left to right, graphs G_A and G_B and a joint graph G compatible with them.

Example 5.1. Consider two graphical Gaussian models $\mathcal{F}$ and $\mathcal{G}$ as in Figure 5.1, where $Y_A \sim N_3(0, \Sigma)$, $K = \Sigma^{-1} \in S^+(G_A)$, and $Y_B \sim N_2(0, \Phi)$, $\Phi^{-1} \in S^+(G_B)$. $y_A = (y_j^i)$ with $j = 1, 2, 3$ and $i = 1, \cdots, n_A$ are observations from family $\mathcal{F}$ and $y_B = (y_j^i)$ with $j = 2, 3$ and $i = 1, \cdots, n_B$ are observations from family $\mathcal{G}$, $n = n_A + n_B$. As we have seen in

Section 4.1, the families are meta-consistent, the meta-Markov combination of the two is $\mathcal{F} \star \mathcal{G} = \{Y \sim N_3(0, \Sigma),\ \Sigma^{-1} \in S^+(G_A)\}$, and its graph is shown on the right of Figure 5.1. The interest lies in estimating the parameter Σ, with inverse given by

$$K = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & 0 \\ k_{31} & 0 & k_{33} \end{pmatrix}.$$

We interpret this problem as a missing data problem with the missing pattern shown in Table 5.1. Here, variables Y_3 and Y_2 are more frequently observed than Y_1.

	Y_1	Y_2	Y_3
A	n_A	n_A	n_A
B		n_B	n_B

Table 5.1: Missing pattern for the problem considered.

It can be seen that the hypothetical complete data model has the form of a regular exponential family with unknown canonical parameter $\theta = K$; therefore we apply the version of the EM algorithm as it specialises for exponential families. Let W be the matrix of sufficient statistics for the complete data model, i.e., the matrix

$$W = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{12} & w_{22} & 0 \\ w_{13} & 0 & w_{33} \end{pmatrix},$$

where the sufficient statistics, w_{kj}, $k, j = 1, 2, 3$, are

$$w_{11} = \sum_{i=1}^{n} (y_1^i)^2,\ w_{22} = \sum_{i=1}^{n} (y_2^i)^2,\ w_{33} = \sum_{i=1}^{n} (y_3^i)^2,\ w_{12} = \sum_{i=1}^{n} y_1^i y_2^i,\ w_{13} = \sum_{i=1}^{n} y_1^i y_3^i.$$

In the following, $W_{[h,l]}$ will denote the submatrix $\begin{pmatrix} w_{ll} & w_{lh} \\ w_{hl} & w_{hh} \end{pmatrix}$.

As the graph is decomposable, if $n > 2$ the maximum likelihood estimate, $\hat{K}$, of K for the complete data case is given by

$$\hat{K} = n \begin{pmatrix} w_{[1,2]}^{11} & w_{[1,2]}^{12} & 0 \\ w_{[1,2]}^{21} & w_{[1,2]}^{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} + n \begin{pmatrix} w_{[1,3]}^{11} & 0 & w_{[1,3]}^{13} \\ 0 & 0 & 0 \\ w_{[1,3]}^{31} & 0 & w_{[1,3]}^{33} \end{pmatrix} - \begin{pmatrix} \frac{n}{w_{11}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{5.1}$$

where $w^{rs}_{[h,l]}$ is the rsth element in $W^{-1}_{[h,l]}$.

We apply the EM algorithm recalling that the restrictions on the parameter, induced by the graph, enter in the M-step only. The algorithm starts by specifying an initial value $K^{(0)}$ for K. The E-step computes the expected values of the complete data sufficient statistics, conditional on the observed data, using the current estimate of the parameter. At iteration (t), denote the current estimate of the parameter as $\theta^{(t)} = K^{(t)}$. Exploiting the fact that $\sum_{i=n_A+1}^{n} \mathrm{Cov}\left(Y_1^i, Y_j^i \middle| Y_{obs}, \theta^{(t)}\right) = 0$, $j = 2, 3$, we have

$$
\begin{aligned}
w_{1j}^{(t)} &= E\left(\sum_{i=1}^{n} Y_1^i Y_j^i \middle| Y_{obs}, \theta^{(t)}\right), \\
&= E\left(\sum_{i=1}^{n_A} Y_1^i Y_j^i + \sum_{i=n_A+1}^{n} Y_1^i Y_j^i \middle| Y_{obs}, \theta^{(t)}\right), \\
&= E\left(\sum_{i=1}^{n_A} Y_1^i Y_j^i \middle| Y_{obs}, \theta^{(t)}\right) + E\left(\sum_{i=n_A+1}^{n} Y_1^i Y_j^i \middle| Y_{obs}, \theta^{(t)}\right), \\
&= \sum_{i=1}^{n_A} y_1^i y_j^i + \sum_{i=n_A+1}^{n} E\left(Y_1^i Y_j^i \middle| Y_{obs}, \theta^{(t)}\right), \\
&= \sum_{i=1}^{n_A} y_1^i y_j^i + \sum_{i=n_A+1}^{n} \mathrm{Cov}\left(Y_1^i, Y_j^i \middle| Y_{obs}, \theta^{(t)}\right) + \sum_{i=n_A+1}^{n} E\left(Y_1^i \middle| Y_{obs}, \theta^{(t)}\right) E\left(Y_j^i \middle| Y_{obs}, \theta^{(t)}\right), \\
&= \sum_{i=1}^{n_A} y_1^i y_j^i + \sum_{i=n_A+1}^{n} E\left(Y_1^i \middle| Y_2^i = y_2^i, Y_3^i = y_3^i\right) y_j^i, \quad j = 2, 3, \\
&= \sum_{i=1}^{n_A} y_1^i y_j^i + \sum_{i=n_A+1}^{n} \left(y_1^i\right)^* y_j^i, \quad j = 2, 3,
\end{aligned}
$$

where

$$
\begin{aligned}
\left(y_1^i\right)^* &= E\left(Y_1^i \middle| Y_2^i = y_2^i, Y_3^i = y_3^i\right), \\
&= -k_{11}^{-1} \begin{pmatrix} k_{12} & k_{13} \end{pmatrix} \begin{pmatrix} y_2^i \\ y_3^i \end{pmatrix}, \\
&= -\frac{k_{12}}{k_{11}} y_2^i - \frac{k_{13}}{k_{11}} y_3^i.
\end{aligned}
$$

In a similar way,

$$\begin{aligned} w_{11}^{(t)} &= E\left(\sum_{i=1}^{n}(Y_1^i)^2 \middle| Y_{obs}, \theta^{(t)}\right), \\ &= \sum_{i=1}^{n_A}(y_1^i)^2 + \sum_{i=n_A+1}^{n} \mathrm{Var}\left(Y_1^i \middle| Y_2^i = y_2^i, Y_3^i = y_3^i\right) + \sum_{i=n_A+1}^{n}\left[E\left(Y_1^i | Y_2^i = y_2^i, Y_3^i = y_3^i\right)\right]^2, \\ &= \sum_{i=1}^{n_A}(y_1^i)^2 + (n-n_A)(k_{11})^{-1} + \sum_{i=n_A+1}^{n}\left[(y_1^i)^*\right]^2. \end{aligned}$$

Variables Y_2 and Y_3 are completely observed and, therefore, the conditional expectations of their sums of squares are equal to their observed values, i.e.,

$$w_{jj} = w_{jj}^{(t)} = E\left(\sum_{i=1}^{n}(Y_j^i)^2 | Y_{obs}, \theta^{(t)}\right) = \sum_{i=1}^{n}(y_j^i)^2, \quad j = 2, 3.$$

The M-step of the algorithm computes $\hat{K}^{(t+1)}$, exploiting (5.1) and updating the relevant quantities with the values obtained in the E-step:

$$\hat{K}^{(t+1)} = n \begin{pmatrix} {w_{[1,2]}^{11}}^{(t)} & {w_{[1,2]}^{12}}^{(t)} & 0 \\ {w_{[1,2]}^{21}}^{(t)} & w_{[1,2]}^{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} + n \begin{pmatrix} {w_{[1,3]}^{11}}^{(t)} & 0 & {w_{[1,3]}^{13}}^{(t)} \\ 0 & 0 & 0 \\ {w_{[1,3]}^{31}}^{(t)} & 0 & w_{[1,3]}^{33} \end{pmatrix} - \begin{pmatrix} \frac{n}{{w_{11}}^{(t)}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The algorithm iterates the two steps until convergence.

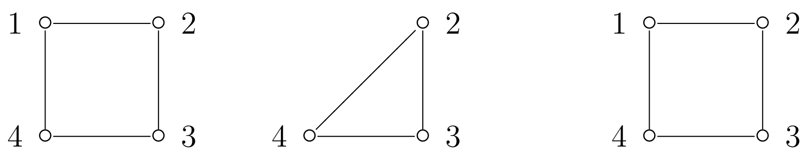


Figure 5.2: On the left, two marginal graphs G_A and G_B. On the right, the graph corresponding to the meta-Markov combination of them.

Example 5.2. Consider $Y_A \sim N_4(0, \Sigma)$, $\Sigma^{-1} = K \in S^+(G_A)$ and $Y_B \sim N_3(0, \Phi)$, $\Phi^{-1} \in S^+(G_B)$, corresponding to the families $\mathcal{F}$ and $\mathcal{G}$. The graphs are represented in Figure 5.2. As in the previous example, $y_A = (y_j^i)$ with $j = 1, 2, 3$ and $i = 1, \cdots, n_A$ and $y_B = (y_j^i)$ with $j = 2, 3$ and $i = 1, \cdots, n_B$, $n = n_A + n_B$, are observations from family $\mathcal{F}$ and $\mathcal{G}$, respectively. The two families are meta-consistent and their meta-Markov combination is $\mathcal{F} \star \mathcal{G} = \{Y \sim N_4(0, \Sigma),\ \Sigma^{-1} = K \in S^+(G_A)\}$, where

$$K = \begin{pmatrix} k_{11} & k_{12} & 0 & k_{14} \\ k_{21} & k_{22} & k_{23} & 0 \\ 0 & k_{32} & k_{33} & k_{34} \\ k_{41} & 0 & k_{43} & k_{44} \end{pmatrix}.$$

The corresponding graph is shown on the right of Figure 5.2. Here, the missing pattern is shown in Table 5.2.

	Y_1	Y_2	Y_3	Y_4
A	n_A	n_A	n_A	n_A
B		n_B	n_B	n_B

Table 5.2: Missing pattern for the problem considered.

The hypothetical complete data model has the form of a regular exponential family with unknown canonical parameter $\theta = K$, and the sufficient statistics, $w_{kj},\ k, j = 1, \cdots, 4$, are given by

$$w_{11} = \sum_{i=1}^{n} (y_1^i)^2, \quad w_{22} = \sum_{i=1}^{n} (y_2^i)^2, \quad w_{33} = \sum_{i=1}^{n} (y_3^i)^2, \quad w_{44} = \sum_{i=1}^{n} (y_4^i)^2,$$

$$w_{12} = \sum_{i=1}^{n} y_1^i y_2^i, \quad w_{23} = \sum_{i=1}^{n} y_2^i y_3^i, \quad w_{34} = \sum_{i=1}^{n} y_3^i y_4^i, \quad w_{41} = \sum_{i=1}^{n} y_4^i y_1^i.$$

If $n \geq 3$ (Buhl, 1993), the maximum likelihood estimate, $\hat{K}$, of K for the complete data case is $\hat{K} = \hat{\Sigma}^{-1}$, where the likelihood equations are

$$\begin{aligned} n\hat{\Sigma}_{\{1,2\}} &= w_{\{1,2\}}, \\ n\hat{\Sigma}_{\{2,3\}} &= w_{\{2,3\}}, \\ n\hat{\Sigma}_{\{3,4\}} &= w_{\{3,4\}}, \\ n\hat{\Sigma}_{\{4,1\}} &= w_{\{4,1\}}, \end{aligned}$$

where w_{kj} are the elements of $W = \sum_{i=1}^{n} y^i (y^i)^T$, i.e., the matrix of sum of squares and products. These equations can be solved by Iterative Proportional Scaling (Speed and Kiiveri, 1986). In this case, $\hat{K} = \lim_{r\to\infty} K_r$ where, given $K_0 = I$, $K_{r+1} = (T_{12}T_{23}T_{34}T_{41})K_r$. $T_c K$ is the operation of adjusting the marginal. For example, $T_{12}K =$

$\begin{pmatrix} C & D \\ E & F \end{pmatrix}$, where

$$C = n \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}^{-1} + \begin{pmatrix} 0 & k_{14} \\ k_{23} & 0 \end{pmatrix} \cdot \begin{pmatrix} k_{33} & k_{34} \\ k_{43} & k_{44} \end{pmatrix}^{-1} \cdot \begin{pmatrix} 0 & k_{34} \\ k_{41} & 0 \end{pmatrix},$$

$$D = \begin{pmatrix} 0 & k_{14} \\ k_{32} & 0 \end{pmatrix}, \quad E = \begin{pmatrix} 0 & k_{32} \\ k_{41} & 0 \end{pmatrix}, \quad F = \begin{pmatrix} k_{33} & k_{34} \\ k_{43} & k_{44} \end{pmatrix}.$$

In the E-step, as in the previous example, we compute the value of the sufficient statistics, $w_{kj}^{(t)}$, with $\theta^{(t)} = K^{(t)}$ being the current estimate of the parameter. It is

$$\begin{aligned} w_{1j}^{(t)} &= E\left(\sum_{i=1}^{n} Y_1^i Y_j^i \middle| Y_{obs}, \theta^{(t)} \right), \\ &= E\left(\sum_{i=1}^{n_A} Y_1^i Y_j^i + \sum_{i=n_A+1}^{n} Y_1^i Y_j^i \middle| Y_{obs}, \theta^{(t)} \right), \\ &= \sum_{i=1}^{n_A} y_1^i y_j^i + \sum_{i=n_A+1}^{n} E\left(Y_1^i \middle| Y_2^i = y_2^i, Y_3^i = y_3^i, Y_4^i = y_4^i \right) y_j^i, \quad j = 2, 4, \\ &= \sum_{i=1}^{n_A} y_1^i y_j^i + \sum_{i=n_A+1}^{n} (y_1^i)^* y_j^i, \quad j = 2, 4, \end{aligned}$$

where

$$\begin{aligned} (y_1^i)^* &= E\left(Y_1^i \middle| Y_2^i = y_2^i, Y_3^i = y_3^i, Y_4^i = y_4^i \right), \\ &= -k_{11}^{-1} \begin{pmatrix} k_{12} & 0 & k_{14} \end{pmatrix} \begin{pmatrix} y_2^i \\ y_3^i \\ y_4^i \end{pmatrix}, \\ &= -\frac{k_{12}}{k_{11}} y_2^i - \frac{k_{14}}{k_{11}} y_4^i. \end{aligned}$$

Again, results derive by exploiting that $\sum_{i=n_A+1}^{n} \mathrm{Cov}\left(Y_1^i, Y_j^i \middle| Y_{obs}, \theta^{(t)}\right) = 0$, $j = 2, 4$.

$$
\begin{aligned}
w_{11}^{(t)} &= E\left(\sum_{i=1}^{n} (Y_1^i)^2 \middle| Y_{obs}, \theta^{(t)}\right), \\
&= \sum_{i=1}^{n_A} (y_1^i)^2 + \sum_{i=n_A+1}^{n} \mathrm{Var}\left(Y_1^i \middle| Y_2^i = y_2^i, Y_3^i = y_3^i, Y_4^i = y_4^i\right) \\
&\quad + \sum_{i=n_A+1}^{n} \left[E\left(Y_1^i | Y_2^i = y_2^i, Y_3^i = y_3^i, Y_4^i = y_4^i\right)\right]^2, \\
&= \sum_{i=1}^{n_A} (y_1^i)^2 + (n - n_A)(k_{11})^{-1} + \sum_{i=n_A+1}^{n} \left[(y_1^i)^*\right]^2.
\end{aligned}
$$

Variables Y_2, Y_3 and Y_4 are completely observed and, hence, the conditional expectations of the last sufficient statistics are equal to their observed values,

$$
\begin{aligned}
w_{jj}^{(t)} &= E\left(\sum_{i=1}^{n} (Y_j^i)^2 | Y_{obs}, \theta^{(t)}\right) = \sum_{i=1}^{n} (y_j^i)^2 = w_{jj}, \quad j = 2, 3, 4. \\
w_{3j}^{(t)} &= E\left(\sum_{i=1}^{n} Y_3^i Y_j^i | Y_{obs}, \theta^{(t)}\right) = \sum_{i=1}^{n} y_3^i y_j^i = w_{3j}, \quad j = 2, 4.
\end{aligned}
$$

At the (t)-th iteration, the M-step performs the IPS algorithm, with $w_{cc}^{(t)}$, $c \in \mathcal{C}$ found in the E-step. In particular, $w_{22}^{(t)} = w_{22}$, $w_{33}^{(t)} = w_{33}$, $w_{23}^{(t)} = w_{23}$ and $w_{34}^{(t)} = w_{34}$, while the values of $w_{11}^{(t)}$, $w_{12}^{(t)}$ and $w_{41}^{(t)}$ are adjusted. $T_c K$ is given by

$$
T_c K^{(t+1)} = \begin{pmatrix} n\left(w_{cc}^{(t)}\right)^{-1} + {K_{ca}}^{(t)} \cdot \left({K_{aa}}^{(t)}\right)^{-1} \cdot {K_{ac}}^{(t)} & {K_{ca}}^{(t)} \\ {K_{ac}}^{(t)} & {K_{aa}}^{(t)} \end{pmatrix}.
$$

In every M-step, the IPS algorithm leads to the value of $\hat{K}^{(t+1)}$ which is the input to the next E-step and the new starting value for the IPS algorithm of the following M-step.

Now we describe an example in which the EM algorithm is not necessary.

Example 5.3. We consider Figure 5.3, where $\mathcal{F} = \{Y_A \sim N_5(0, \Omega),\ \Omega^{-1} \in S^+(G_A)\}$, $\mathcal{G} = \{Y_B \sim N_3(0, \Phi),\ \Phi^{-1} \in S^+(G_B)\}$ and we indicate the densities of the families as $f \in \mathcal{F}$ and $g \in \mathcal{G}$, respectively. The families are meta-consistent and the meta-Markov combination is represented by the graph on the right of Figure 5.3 as $Y \sim N_6(0, \Sigma)$ and

$$\Sigma^{-1} = \begin{pmatrix} k_{11} & 0 & k_{13} & 0 & 0 & 0 \\ 0 & k_{22} & k_{23} & 0 & 0 & 0 \\ k_{31} & k_{32} & k_{33} & k_{34} & k_{35} & 0 \\ 0 & 0 & k_{43} & k_{44} & k_{45} & k_{46} \\ 0 & 0 & k_{53} & k_{54} & k_{55} & k_{56} \\ 0 & 0 & 0 & k_{64} & k_{65} & k_{66} \end{pmatrix}.$$

Consider two samples of observations, $y_A = (y_j^i)$, with $j = 1, \dots, 5$ and $i = 1, \dots, n_A$,

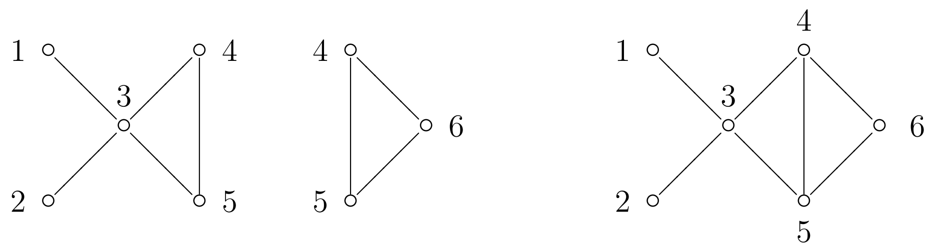


Figure 5.3: On the left, two marginal graphs G_A and G_B. On the right, the graph corresponding to their meta-Markov combination.

and $y_B = (y_j^i)$, with $j = 4, 5, 6$ and $i = 1, \dots, n_B$, $n = n_A + n_B$, for model $\mathcal{F}$ and $\mathcal{G}$, respectively. The meta-Markov combination is given by

$$\mathcal{F} \star \mathcal{G} = \left\{ \frac{f \cdot g}{g_S}, f \text{ and } g \text{ consistent} \right\}, \tag{5.2}$$

where $S = \{4, 5\}$ and g_S is the marginal density of a multivariate normal distribution and, therefore, it is itself a multivariate normal distribution, $Y_S \sim N_2(0, \Omega_S)$. The sample for the variables in S is built as $y_S = (y_A^\star \cup y_B^\star) = (y_j^i)$, $i = 1, \dots, n$, $j = 4, 5$, where $y_A^\star$ and $y_B^\star$ are the observations for S acquired from y_A and y_B. We might exploit the method introduced in the previous examples, but in this case the joint family can be directly estimated, as follows. If both $\hat{f}$ and $\hat{g}$ exist, the maximum likelihood estimate is found by placing $\hat{f}$ and $\hat{g}$ in (5.2). This result follows from Lauritzen (1996), Proposition 6.17, and it is equivalent to find the estimate, $\hat{\Sigma}^{-1}$, of Σ^{-1} as

$$\hat{\Sigma}^{-1} = [\hat{\Omega}^{-1}]^6 + [\hat{\Phi}^{-1}]^6 - [w_{\{4,5\}}^{-1}]^6, \tag{5.3}$$

where $w_{\{4,5\}} = \begin{pmatrix} w_{44} & w_{45} \\ w_{54} & w_{55} \end{pmatrix}$, $w_{\alpha\beta} = \sum_{i=1}^n y_\alpha^i y_\beta^i$, (see Lauritzen, 1996, Proposition 5.9), and $[A]^h$ denotes the matrix obtained from A by filling up with zero entries to obtain

full dimension $h \times h$.
Hence, we need to compute $\hat{f}$, $\hat{g}$ and $\hat{g}_S$, which is equivalent to calculate $\hat{\Phi}$ and $\hat{\Omega}$.
If $n_B > 1$, the estimate, $\hat{\Phi}$, of Φ is

$$\hat{\Phi} = \frac{1}{n_B}\begin{pmatrix} \hat{\phi}_{44} & \hat{\phi}_{45} & \hat{\phi}_{46} \\ \hat{\phi}_{54} & \hat{\phi}_{55} & \hat{\phi}_{56} \\ \hat{\phi}_{64} & \hat{\phi}_{65} & \hat{\phi}_{66} \end{pmatrix},$$

because the graph is complete and $\hat{\phi}_{\alpha\beta} = \sum_{i=1}^{n-n_A} y_\alpha^i y_\beta^i$. Once we have estimated Φ, we get also the estimate, $\hat{\Phi}_S$, for Φ_S as

$$\hat{\Phi}_S = \hat{\Omega}_S = \begin{pmatrix} \hat{\phi}_{44} & \hat{\phi}_{45} \\ \hat{\phi}_{54} & \hat{\phi}_{55} \end{pmatrix},$$

exploiting the fact that Y_S is the marginal of a multivariate normal distribution. Assuming $n_A > 3$, and with the usual notation, we calculate $\hat{\Omega}^{-1}$ as

$$\begin{aligned} \hat{\Omega}^{-1} \;&=\; n_A \begin{pmatrix} \omega^{11}_{[1,3]} & 0 & \omega^{13}_{[1,3]} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \omega^{31}_{[1,3]} & 0 & \omega^{33}_{[1,3]} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + n_A \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \omega^{22}_{[2,3]} & \omega^{23}_{[2,3]} & 0 & 0 \\ 0 & \omega^{32}_{[2,3]} & \omega^{33}_{[2,3]} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\ &+ n_A \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \omega^{33}_{[3,4,5]} & \omega^{34}_{[3,4,5]} & \omega^{35}_{[3,4,5]} \\ 0 & 0 & \omega^{43}_{[3,4,5]} & \omega^{44}_{[3,4,5]} & \omega^{45}_{[3,4,5]} \\ 0 & 0 & \omega^{53}_{[3,4,5]} & \omega^{54}_{[3,4,5]} & \omega^{55}_{[3,4,5]} \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2n_A}{\omega_{33}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \end{aligned}$$

where ω_V^{rs} is the rsth element in $W_{[V]}^{-1}$ and $W = \sum_{i=1}^{n_A} y^i (y^i)^T$, i.e., matrix of sum of squares and products. Then, we get the estimate of Σ using (5.3). Note that to estimate Φ and Ω we used the complete observations.

This is an example where it would make no difference whether we had the maximum likelihood estimates from each of the experiments or the raw data. This is because $Y_{\{4,5\}}$ is a cut, so each of the models is collapsible onto the intersection, which again is equivalent to the meta-Markov and quasi-Markov combinations being identical.

Example 5.4. Consider three Gaussian graphical models $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{H}$ with
$\mathcal{F} = \{Y_A \sim N_2(0, \Gamma),\ \Gamma^{-1} \in S^+(G_A)\}$, $\mathcal{G} = \{Y_B \sim N_3(0, \Omega),\ \Omega^{-1} \in S^+(G_B)\}$, and $\mathcal{H} =$

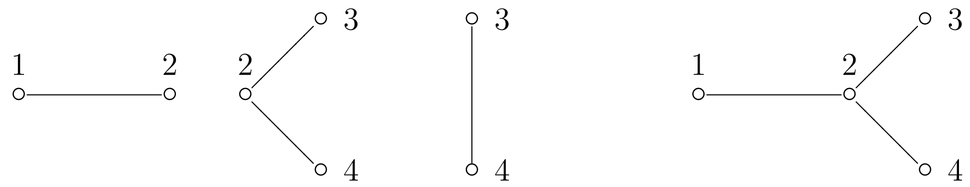


Figure 5.4: From left to right, graphs G_A, G_B and G_C and a joint graph compatible with them.

$\{Y_C \sim N_2(0, \Phi),\ \Phi^{-1} \in S^+(G_C)\}$, as in Figure 5.4. Here, $y_A = (y_j^i)$ with $j = 1, 2$, and $i = 1, \dots, n_A$, are observations from family $\mathcal{F}$, $y_B = (y_j^i)$ with $j = 2, 3, 4$, and $i = 1, \dots, n_B$, are observations from family $\mathcal{G}$, and $y_C = (y_j^i)$ with $j = 3, 4$, and $i = 1, \dots, n_C$, are observations from family $\mathcal{H}$, where $n = n_A + n_B + n_C$. The missing pattern is shown in Table 5.3.

Now, we give the details of the estimation of the parameters of the meta-Markov combination $\mathcal{F} \star (\mathcal{G} \star \mathcal{H})$ (see Example 4.4). The same procedure is appropriate also for $(\mathcal{F} \star \mathcal{G}) \star \mathcal{H}$.

	Y_1	Y_2	Y_3	Y_4
A	n_A	n_A		
B		n_B	n_B	n_B
C			n_C	n_C

Table 5.3: Missing pattern for the problem considered.

If $n > 2$, the maximum likelihood estimate, $\hat{K}$, of K for the complete data case is

$$\begin{aligned}\hat{K} \;=\; & n \begin{pmatrix} w^{11}_{[1,2]} & w^{12}_{[1,2]} & 0 & 0 \\ w^{21}_{[1,2]} & w^{22}_{[1,2]} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + n \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & w^{22}_{[2,3]} & w^{23}_{[2,3]} & 0 \\ 0 & w^{32}_{[2,3]} & w^{33}_{[2,3]} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + n \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & w^{22}_{[2,4]} & 0 & w^{24}_{[2,4]} \\ 0 & 0 & 0 & 0 \\ 0 & w^{42}_{[2,4]} & 0 & w^{44}_{[2,4]} \end{pmatrix} \\ & - \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{2n}{w_{22}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\end{aligned}$$

since the graph is decomposable. The sufficient statistics, w_{kj}, $k, j = 1, \cdots, 4$, are

$$w_{2j} = \sum_{i=1}^{n} y_2^i y_j^i, \quad j = 1, 3, 4, \quad w_{jj} = \sum_{i=1}^{n} (y_j^i)^2, \quad j = 1, \ldots, 4.$$

Denote the current estimate of the parameter as $\theta^{(t)} = K^{(t)}$.

As in the previous examples, we use $E\left(Y_j^i Y_k^i \middle| Y_{obs}\right) = \mathrm{Cov}(Y_j^i, Y_k^i | Y_{obs}) + E(Y_j^i | Y_{obs}) E(Y_j^i | Y_{obs})$, and $\mathrm{Cov}(Y_j^i, Y_k^i | Y_{obs}) = 0$, if Y_j^i or Y_k^i is observed. In the E-step we compute the following quantities:

$$\begin{aligned}
w_{12}^{(t)} &= E\left(\sum_{i=1}^{n} Y_1^i Y_2^i \middle| Y_{obs}, \theta^{(t)}\right), \\
&= E\left(\sum_{i=1}^{n_A} Y_1^i Y_2^i \middle| Y_1^i, Y_2^i\right) + E\left(\sum_{i=n_A+1}^{n_A+n_B} Y_1^i Y_2^i \middle| Y_2^i, Y_3^i, Y_4^i\right) + E\left(\sum_{i=n_A+n_B+1}^{n} Y_1^i Y_2^i \middle| Y_3^i, Y_4^i\right), \\
&= \sum_{i=1}^{n_A} y_1^i y_2^i + \sum_{i=n_A+1}^{n_A+n_B} E(Y_1^i | Y_2^i, Y_3^i, Y_4^i) y_2^i + \sum_{i=n_A+n_B+1}^{n} E\left(Y_1^i \middle| Y_3^i, Y_4^i\right) E\left(Y_2^i \middle| Y_3^i, Y_4^i\right) \\
&\quad + \sum_{i=n_A+n_B+1}^{n} \mathrm{Cov}(Y_1^i, Y_2^i \middle| Y_3^i, Y_4^i), \\
&= \sum_{i=1}^{n_A} y_1^i y_2^i + \sum_{i=n_A+1}^{n_A+n_B} (y_1^i)^* y_2^i + \sum_{i=n_A+n_B+1}^{n} z_1^i z_2^i + \sum_{i=n_A+n_B+1}^{n} u_{12}^i,
\end{aligned}$$

where

$$u_{12}^i = \mathrm{Cov}(Y_1^i, Y_2^i \middle| Y_3^i, Y_4^i) = -\frac{k_{21}}{k_{11}k_{22} - k_{12}^2},$$

$$\begin{aligned}
(y_1^i)^* &= E\left(Y_1^i \middle| Y_2^i, Y_3^i, Y_4^i\right), \\
&= -k_{11}^{-1} \begin{pmatrix} k_{12} & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} y_2^i \\ y_3^i \\ y_4^i \end{pmatrix}, \\
&= -\frac{k_{12}}{k_{11}} y_2^i,
\end{aligned}$$

and $z_1^i = E\left(Y_1^i \middle| Y_3^i, Y_4^i\right)$, $z_2^i = E\left(Y_2^i \middle| Y_3^i, Y_4^i\right)$ are computed as the marginal components of $E\left(Y_1^i, Y_2^i \middle| Y_3^i, Y_4^i\right)$. Therefore,

$$\begin{aligned}
E\left(Y_1^i, Y_2^i \middle| Y_3^i, Y_4^i\right) &= -\begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}^{-1} \cdot \begin{pmatrix} 0 & 0 \\ k_{23} & k_{24} \end{pmatrix} \cdot \begin{pmatrix} y_3^i \\ y_4^i \end{pmatrix}, \\
&= -\frac{1}{k_{11}k_{22} - k_{12}^2} \begin{pmatrix} k_{12}k_{23}y_3^i + k_{12}k_{24}y_4^i \\ k_{11}k_{23}y_3^i + k_{11}k_{24}y_4^i \end{pmatrix},
\end{aligned}$$

yields

$$\begin{aligned} z_1^i &= E\left(Y_1^i|Y_3^i, Y_4^i\right) = -\frac{k_{12}k_{23}y_3^i + k_{12}k_{24}y_4^i}{k_{11}k_{22} - k_{12}^2}, \\ z_2^i &= E\left(Y_2^i|Y_3^i, Y_4^i\right) = -\frac{k_{11}k_{23}y_3^i + k_{11}k_{24}y_4^i}{k_{11}k_{22} - k_{12}^2}. \end{aligned}$$

$$\begin{aligned} w_{23}^{(t)} &= E\left(\sum_{i=1}^{n_A} Y_2^i Y_3^i \middle| Y_1^i, Y_2^i\right) + E\left(\sum_{i=n_A+1}^{n_A+n_B} Y_2^i Y_3^i \middle| Y_2^i, Y_3^i, Y_4^i\right) + E\left(\sum_{i=n_A+n_B+1}^{n} Y_2^i Y_3^i \middle| Y_3^i, Y_4^i\right), \\ &= \sum_{i=1}^{n_A} E(Y_3^i|Y_1^i, Y_2^i) y_2^i + \sum_{i=n_A+1}^{n_A+n_B} y_2^i y_3^i + \sum_{i=n_A+n_B+1}^{n} E(Y_2^i|Y_3^i, Y_4^i) y_3^i, \\ &= \sum_{i=1}^{n_A} z_3^i y_2^i + \sum_{i=n_A+1}^{n_A+n_B} y_2^i y_3^i + \sum_{i=n_A+n_B+1}^{n} z_2^i y_3^i, \end{aligned}$$

where

$$E\left(Y_3^i, Y_4^i \middle| Y_1^i, Y_2^i\right) = -\begin{pmatrix} k_{33} & 0 \\ 0 & k_{44} \end{pmatrix}^{-1} \cdot \begin{pmatrix} 0 & k_{32} \\ 0 & k_{42} \end{pmatrix} \cdot \begin{pmatrix} y_1^i \\ y_2^i \end{pmatrix} = \begin{pmatrix} -\frac{k_{32}}{k_{33}} y_2^i \\ -\frac{k_{42}}{k_{44}} y_2^i \end{pmatrix} \quad \text{and}$$

$$\begin{aligned} z_3^i = E\left(Y_3^i \middle| Y_1^i, Y_2^i\right) &= -\frac{k_{32}}{k_{33}} y_2^i, \\ z_4^i = E\left(Y_4^i \middle| Y_1^i, Y_2^i\right) &= -\frac{k_{42}}{k_{44}} y_2^i. \end{aligned}$$

$$\begin{aligned} w_{24}^{(t)} &= E\left(\sum_{i=1}^{n_A} Y_2^i Y_4^i \middle| Y_1^i, Y_2^i\right) + E\left(\sum_{i=n_A+1}^{n_A+n_B} Y_2^i Y_4^i \middle| Y_2^i, Y_3^i, Y_4^i\right) + E\left(\sum_{i=n_A+n_B+1}^{n} Y_2^i Y_4^i \middle| Y_3^i, Y_4^i\right), \\ &= \sum_{i=1}^{n_A} E(Y_4^i|Y_1^i, Y_2^i) y_2^i + \sum_{i=n_A+1}^{n_A+n_B} y_2^i y_4^i + \sum_{i=n_A+n_B+1}^{n} E(Y_2^i|Y_3^i, Y_4^i) y_4^i, \\ &= \sum_{i=1}^{n_A} z_4^i y_2^i + \sum_{i=n_A+1}^{n_A+n_B} y_2^i y_4^i + \sum_{i=n_A+n_B+1}^{n} z_2^i y_4^i. \end{aligned}$$

$$\begin{aligned}
w_{11}^{(t)} &= E\left(\sum_{i=1}^{n}(Y_1^i)^2\Big|Y_{obs},\theta^{(t)}\right),\\
&= \sum_{i=1}^{n_A}(y_1^i)^2+\sum_{i=n_A+1}^{n_A+n_B}\mathrm{Var}\left(Y_1^i\big|Y_2^i,Y_3^i,Y_4^i\right)+\sum_{i=n_A+1}^{n_A+n_B}(E\left(Y_1^i|Y_2^i,Y_3^i,Y_4^i\right))^2\\
&\quad+\sum_{i=n_A+n_B+1}^{n}\mathrm{Var}\left(Y_1^i\big|Y_3^i,Y_4^i\right)+\sum_{i=n_A+n_B+1}^{n}(E\left(Y_1^i|Y_3^i,Y_4^i\right))^2,\\
&= \sum_{i=1}^{n_A}(y_1^i)^2+n_Bk_{11}^{-1}+\sum_{i=n_A+1}^{n_A+n_B}\left[(y_1^i)^*\right]^2+n_Cv_1+\sum_{i=n_A+n_B+1}^{n}(z_1^i)^2,
\end{aligned}$$

where

$$v_1 = \mathrm{Var}\left(Y_1^i|Y_3^i,Y_4^i\right)=\frac{k_{22}}{k_{11}k_{22}-k_{12}^2}.$$

$$\begin{aligned}
w_{22}^{(t)} &= E\left(\sum_{i=1}^{n}(Y_2^i)^2\Big|Y_{obs},\theta^{(t)}\right),\\
&= \sum_{i=1}^{n_A}(y_2^i)^2+\sum_{i=n_A+1}^{n_A+n_B}\mathrm{Var}\left(Y_2^i\big|Y_1^i,Y_3^i,Y_4^i\right)+\sum_{i=n_A+1}^{n_A+n_B}(E\left(Y_2^i|Y_1^i,Y_3^i,Y_4^i\right))^2\\
&\quad+\sum_{i=n_A+n_B+1}^{n}\mathrm{Var}\left(Y_2^i\big|Y_3^i,Y_4^i\right)+\sum_{i=n_A+n_B+1}^{n}(E\left(Y_2^i|Y_3^i,Y_4^i\right))^2,\\
&= \sum_{i=1}^{n_A}(y_2^i)^2+n_Bk_{22}^{-1}+\sum_{i=n_A+1}^{n_A+n_B}\left[(y_2^i)^*\right]^2+n_Cv_2+\sum_{i=n_A+n_B+1}^{n}(z_2^i)^2,
\end{aligned}$$

where

$$v_2 = \mathrm{Var}\left(Y_2^i|Y_3^i,Y_4^i\right)=\frac{k_{11}}{k_{11}k_{22}-k_{12}^2},$$

$$\begin{aligned}
(y_2^i)^* &= E\left(Y_2^i|Y_1^i,Y_3^i,Y_4^i\right),\\
&= -\frac{1}{k_{22}}\cdot\begin{pmatrix}k_{21} & k_{23} & k_{24}\end{pmatrix}\cdot\begin{pmatrix}y_1^i\\ y_3^i\\ y_4^i\end{pmatrix},\\
&= -\frac{k_{21}y_1^i+k_{23}y_3^i+k_{24}y_4^i}{k_{22}}.
\end{aligned}$$

$$\begin{aligned}
w_{33}^{(t)} &= E\left(\sum_{i=1}^{n}(Y_3^i)^2 \middle| Y_{obs}, \theta^{(t)}\right), \\
&= \sum_{i=n_A+1}^{n}(y_3^i)^2 + \sum_{i=1}^{n_A}\mathrm{Var}\left(Y_3^i|Y_1^i, Y_2^i\right) + \sum_{i=1}^{n_A}(E\left(Y_3^i|Y_1^i, Y_2^i\right))^2, \\
&= \sum_{i=n_A+1}^{n}(y_3^i)^2 + n_A k_{33}^{-1} + \sum_{i=1}^{n_A}\left(z_3^i\right)^2.
\end{aligned}$$

$$\begin{aligned}
w_{44}^{(t)} &= E\left(\sum_{i=1}^{n}(Y_4^i)^2 \middle| Y_{obs}, \theta^{(t)}\right), \\
&= \sum_{i=n_A+1}^{n}(y_4^i)^2 + \sum_{i=1}^{n_A}\mathrm{Var}\left(Y_4^i|Y_1^i, Y_2^i\right) + \sum_{i=1}^{n_A}(E\left(Y_4^i|Y_1^i, Y_2^i\right))^2, \\
&= \sum_{i=n_A+1}^{n}(y_4^i)^2 + n_A k_{44}^{-1} + \sum_{i=1}^{n_A}\left(z_4^i\right)^2.
\end{aligned}$$

To compute $\hat{K}^{(t+1)}$, the M-step plugs in the values found in the E-step in the estimate $\hat{K}^{(t)}$ for the complete data case. Since the graph is decomposable, we have

$$\begin{aligned}
\hat{K}^{(t+1)} &= n\begin{pmatrix} {w_{[1,2]}^{11}}^{(t)} & {w_{[1,2]}^{12}}^{(t)} & 0 & 0 \\ {w_{[1,2]}^{21}}^{(t)} & {w_{[1,2]}^{22}}^{(t)} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + n\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & {w_{[2,3]}^{22}}^{(t)} & {w_{[2,3]}^{23}}^{(t)} & 0 \\ 0 & {w_{[2,3]}^{32}}^{(t)} & {w_{[2,3]}^{33}}^{(t)} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
&+ n\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & {w_{[2,4]}^{22}}^{(t)} & 0 & {w_{[2,4]}^{24}}^{(t)} \\ 0 & 0 & 0 & 0 \\ 0 & {w_{[2,4]}^{42}}^{(t)} & 0 & {w_{[2,4]}^{44}}^{(t)} \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{2n}{{w_{22}}^{(t)}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.
\end{aligned}$$

We notice that, in comparison with the previous cases, all the sufficient statistics are updated after the E-step, and this is due to the particular missing pattern studied.

5.1.2 Quasi-consistent families

In this section, an example of combination of quasi-consistent families is presented. The focus is on the meta-Markov and the quasi-Markov combination. Moreover, we assume that the available data are the sufficient statistics, rather than the observations themselves, showing the difficulties that may arise under this different circumstance.

Figure 5.5: From left to right, graphs G_A and G_B.

Example 5.5. Figure 5.5 displays two graphical Gaussian models, $Y_A \sim N_2(0,\Sigma)$, with $\Sigma^{-1} = \begin{pmatrix} k_{11} & 0 \\ 0 & k_{22} \end{pmatrix}$, corresponding to family $\mathcal{F}$, and $Y_B \sim N_2(0,\Phi)$, with $\Phi^{-1} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}$, corresponding to family $\mathcal{G}$. As we have seen in Section 4.1, $\mathcal{F}$ and $\mathcal{G}$ are quasi-consistent.

Consider $y_A = (y_j^i)$, $j = 1,2$, $i = 1,\ldots,n_A$, and $y_B = (y_j^i)$, $j = 1,2$, and $i = 1,\ldots,n_B$, observations for the families $\mathcal{F}$ and $\mathcal{G}$. Further, suppose that the sufficient statistics $\sum_{i=1}^{n_A}(y_1^i)^2$, $\sum_{i=1}^{n_A}(y_2^i)^2$ for family $\mathcal{F}$, and $\sum_{i=n_A+1}^{n} y_1^i y_2^i$, $\sum_{i=n_A+1}^{n}(y_1^i)^2$, $\sum_{i=n_A+1}^{n}(y_2^i)^2$ for family $\mathcal{G}$ are the only available information. In this case, the observations under analysis are not the raw data but derived quantities, i.e. sufficient statistics or, equivalently, the maximum likelihood estimates in the marginal models. For the sake of simplicity, we study only this example with such particular initial data. Note that also the previous cases might have been studied under this initial assumption.

The meta-Markov combination, which associates only consistent distributions, is represented by the graph on the left of Figure 5.5 as $\mathcal{F}\star\mathcal{G} = \{Y \sim N_2(0,\Sigma), \Sigma^{-1} \in S^+(G_A)\}$. The quasi-Markov combination is represented by the graph on the right of Figure 5.5 as $\mathcal{F} \,\overline{\star}\, \mathcal{G} = \{Y \sim N_2(0,\Phi), \Phi^{-1} \in S^+(G_B)\}$, and it associates all distributions from the two families.

In the meta-Markov combination, we do not need the EM algorithm. The estimates, $\hat{\sigma}_{11}$ and $\hat{\sigma}_{22}$, of the elements σ_{11} and σ_{22} of Σ are obtained as follows:

$$\begin{aligned} \hat{\sigma}_{11} &= \frac{\frac{1}{n_A}\sum_{i=1}^{n_A}(y_1^i)^2 + \frac{1}{n-n_A}\sum_{i=n_A+1}^{n}(y_1^i)^2}{n}, \\ \hat{\sigma}_{22} &= \frac{\frac{1}{n_A}\sum_{i=1}^{n_A}(y_2^i)^2 + \frac{1}{n-n_A}\sum_{i=n_A+1}^{n}(y_2^i)^2}{n}. \end{aligned}$$

In the quasi-Markov combination, it is of interest to estimate Φ and the main difficulty is that there are no data for the edge $(1,2)$ acquired from the family $\mathcal{F}$. Hence, we apply the EM algorithm with the sufficient statistics for the complete data case given by

$$w_{jj} = \sum_{i=1}^{n} (y_j^i)^2 \quad j = 1, 2 \quad \text{and} \quad w_{12} = \sum_{i=1}^{n} y_1^i y_2^i,$$

and the maximum likelihood estimate, $\hat{\Phi}$, of Φ given by:

$$\hat{\Phi} = \frac{1}{n} \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}.$$

Denoting the current estimate of the canonical parameter as $\theta^{(t)} = K^{(t)}$, in the E-step we compute

$$E\left(\sum_{i=1}^{n} (Y_j^i)^2 \middle| Y_{obs}, \theta^{(t)} \right) = w_{jj}^{(t)}, \quad j = 1, 2,$$

exploiting the fact that Y_1 and Y_2 are completely observed. The edge $(1, 2)$ is not completely observed, thus, we have

$$\begin{aligned} w_{12}^{(t)} = E\left(\sum_{i=1}^{n} Y_1^i Y_2^i \middle| Y_{obs}, \theta^{(t)} \right) &= E\left(\sum_{i=1}^{n_A} Y_1^i Y_2^i | Y_{obs}, \theta^{(t)} \right) + E\left(\sum_{i=n_A+1}^{n} Y_1^i Y_2^i \middle| Y_{obs}, \theta^{(t)} \right), \\ &= E\left(S_{12} | S_{11} = s_{11}, S_{22} = s_2 \right) + \sum_{i=n_A+1}^{n} y_1^i y_2^i, \end{aligned}$$

where $S_{12} = \sum_{i=1}^{n_A} Y_1^i Y_2^i$, $S_{11} = \sum_{i=1}^{n_A} (Y_1^i)^2$, $s_{11} = \sum_{i=1}^{n_A} (y_1^i)^2$, $S_{22} = \sum_{i=1}^{n_A} (Y_2^i)^2$, $s_{22} = \sum_{i=1}^{n_A} (y_2^i)^2$. To calculate $E\left(S_{12} | S_{11} = s_{11}, S_{22} = s_2 \right)$, we need to find the distribution of $(S_{12} | S_{11} = s_{11}, S_{22} = s_2)$. We know that

$$f\left(S_{12} | S_{11}, S_{22} \right) = \frac{f\left(S_{12}, S_{11}, S_{22} \right)}{f\left(S_{11}, S_{22} \right)},$$

where $\mathcal{L}\left(S_{12}, S_{11}, S_{22} \right) = \mathcal{W}_2(n_A, \Phi)$ and $\mathcal{L}\left(S_{11}, S_{22} \right)$ is found by marginalising $f(S_{12}, S_{11}, S_{22})$ over S_{12}. Utilising these facts, we have

$$\begin{aligned} f\left(S_{12}, S_{11}, S_{22} \right) &= c(2, n_A)^{-1} (\det \Phi)^{-n_A/2} (\det S)^{(n_A-3)/2} \exp\left\{ -\frac{1}{2} \mathrm{tr}(\Phi^{-1} S) \right\}, \\ &= \mathrm{const} \cdot (S_{11} S_{22} - S_{12}^2)^{(n_A-3)/2} \exp\{ -h_{12} S_{12} \}, \end{aligned}$$

where $\mathrm{const} = c(2, n_A)^{-1} (\det \Phi)^{-n_A/2} e^{-1/2(h_{11} S_{11} + h_{22} S_{22})}$, $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$, and

$c(2, n_A) = (2^{n_A})(\pi^{1/2}) \prod_{i=1}^{2} \Gamma\left(\frac{n+1-i}{2}\right)$. The denominator is given by

$$\begin{aligned} f(S_{11}, S_{22}) &= \int f(S_{11}, S_{22}, S_{12})\, dS_{12}, \\ &= \text{const} \int_{-\sqrt{m}}^{+\sqrt{m}} (m - S_{12}^2)^{(n_A-3)/2} \exp\{-h_{12}S_{12}\} dS_{12}, \\ &= \text{const} \int_{-\sqrt{m}}^{+\sqrt{m}} \left(m - x^2\right)^{\alpha} e^{-bx} dx, \end{aligned}$$

where $m = S_{11}S_{22}$, $\alpha = (n_A - 3)/2$, $b = h_{12}$ and $x = S_{12}$. To derive the results, we used the fact that $-1 \leq \frac{S_{12}}{\sqrt{S_{11}S_{22}}} \leq 1$ and so $-\sqrt{m} \leq S_{12} \leq +\sqrt{m}$. The conditional density is

$$f(S_{12}|S_{11}, S_{22}) = \frac{(m - S_{12}^2)^{(n_A-3)/2} \exp\{-h_{12}S_{12}\}}{\int_{-\sqrt{m}}^{+\sqrt{m}} (m - S_{12}^2)^{(n_A-3)/2} \exp\{-h_{12}S_{12}\} dS_{12}},$$

and its expected value is

$$\begin{aligned} E(S_{12}|S_{11}, S_{22}) &= \frac{\int_{-\sqrt{m}}^{+\sqrt{m}} S_{12}(m - S_{12}^2)^{(n_A-3)/2} \exp\{-h_{12}S_{12}\} dS_{12}}{\int_{-\sqrt{m}}^{+\sqrt{m}} (m - S_{12}^2)^{(n_A-3)/2} \exp\{-h_{12}S_{12}\} dS_{12}}, \\ &= \frac{\int_{-\sqrt{m}}^{+\sqrt{m}} x(m - x^2)^{\alpha} e^{-bx} dx}{\int_{-\sqrt{m}}^{+\sqrt{m}} (m - x^2)^{\alpha} e^{-bx} dx}. \end{aligned}$$

Now,

$$\begin{aligned} \int_{-\sqrt{m}}^{+\sqrt{m}} x(m - x^2)^{\alpha} e^{-bx} dx &= \int_{-\sqrt{m}}^{+0} x(m - x^2)^{\alpha} e^{-bx} dx + \int_{0}^{+\sqrt{m}} x(m - x^2)^{\alpha} e^{-bx} dx, \\ &= -\int_{0}^{+\sqrt{m}} x(m - x^2)^{\alpha} e^{bx} dx + \int_{0}^{+\sqrt{m}} x(m - x^2)^{\alpha} e^{-bx} dx, \\ &= -\frac{\sqrt{\pi}}{2} \left(\frac{\mu}{2}\right)^{\frac{1}{2}-\nu} u^{\nu+\frac{1}{2}} \Gamma(\nu) [I_{\nu+\frac{1}{2}}(\mu u) + L_{\nu+\frac{1}{2}}(\mu u) + I_{\nu+\frac{1}{2}}(-\mu u) \\ &\quad + L_{\nu+\frac{1}{2}}(-\mu u)], \end{aligned}$$

where $\mu = b$, $u = \sqrt{m}$, $\nu = \alpha + 1$.

$$\begin{aligned} \int_{-\sqrt{m}}^{+\sqrt{m}} (m - x^2)^{\alpha} e^{-bx} dx &= \int_{-\sqrt{m}}^{+0} (m - x^2)^{\alpha} e^{-bx} dx + \int_{0}^{+\sqrt{m}} (m - x^2)^{\alpha} e^{-bx} dx, \\ &= -\int_{0}^{+\sqrt{m}} (m - x^2)^{\alpha} e^{bx} dx + \int_{0}^{+\sqrt{m}} (m - x^2)^{\alpha} e^{-bx} dx, \\ &= -\frac{\sqrt{\pi}}{2} \left(\frac{2u}{\mu}\right)^{\nu-\frac{1}{2}} \Gamma(\nu) [I_{\nu-\frac{1}{2}}(\mu u) + L_{\nu-\frac{1}{2}}(\mu u) - I_{\nu-\frac{1}{2}}(-\mu u) \\ &\quad - L_{\nu-\frac{1}{2}}(-\mu u)], \end{aligned}$$

where $\mu = b$, $u = \sqrt{m}$, $\nu = \alpha + 1$. $I_\nu(z)$ is the Bessel function of an imaginary argument (Gradshteyn and Ryzhik, 2000),i.e.,

$$\begin{aligned} I_\nu(z) &= \frac{\left(\frac{z}{2}\right)^\nu}{\Gamma(\nu + \frac{1}{2})\Gamma(\frac{1}{2})} \int_{-1}^{1} (1-t^2)^{\nu - \frac{1}{2}} e^{\pm zt} dt, \\ &= \frac{1}{2\pi i} \oint e^{\frac{z}{2}(t+1/t)} t^{-n-1} dt. \end{aligned}$$

$L_\nu(z)$ is the modified Struve function, (Gradshteyn and Ryzhik, 2000), given by

$$\begin{aligned} L_\nu(z) &= \left(\frac{1}{2}z\right)^{\nu+1} \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}z\right)^{2k}}{\Gamma(k+\frac{3}{2})\Gamma(k+\nu+\frac{3}{2})}, \\ &= \frac{2\left(\frac{z}{2}\right)^\nu}{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})} \int_0^{\frac{\pi}{2}} \mathrm{sh}(z\cos\theta)(\sin\theta) d\theta. \end{aligned}$$

Once we have done this, the M-step easily calculates the maximum likelihood estimate, $\hat{\Phi}^{(t+1)}$, for Φ as

$$\hat{\Phi}^{(t+1)} = \frac{1}{n} \begin{pmatrix} w_{11} & w_{12}^{(t)} \\ w_{21}^{(t)} & w_{22} \end{pmatrix}.$$

5.2 Non-graphical combinations

A non-graphical combination has been described in Chapter 4 as the joint family of distributions that does not correspond to the Gaussian family related to the dependence graph of the combination. Now, we present some examples of non-graphical combinations and their estimation. Even if the joint families of distribution are not graphical models, in some cases they can still be studied applying the tools of graphical models.

5.2.1 Quasi-consistent families

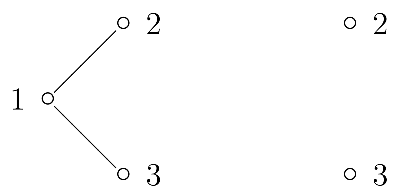


Figure 5.6: From left to right, graphs G_A and G_B.

Example 5.6. Consider two graphical Gaussian models $Y_A \sim N_3(0, \Omega)$, $\Omega^{-1} \in S^+(G_A)$ and $Y_B \sim N_2(0, \Phi)$, $\Phi^{-1} \in S^+(G_B)$, corresponding to the families $\mathcal{F}$ and $\mathcal{G}$, as in Figure 5.6. As we have seen in Section 4.2, the two families are quasi-consistent. The quasi-Markov combination is represented by the graph on the left in Figure 5.6, and it can be estimated following the details of Example 5.1. The meta-Markov combination is given by considering the distributions that satisfy $\rho_{23} = \rho_{12}\rho_{13}$ and $\rho_{23} = 0$, where ρ_{ij} is the usual correlation matrix of the joint family. It is equivalent to $\{\rho_{23} = 0 \wedge \rho_{12} = 0\} \cup \{\rho_{23} = 0 \wedge \rho_{13} = 0\}$ and can be represented by the union of two graphs, as shown in Figure 5.7. Even if this type of combination can be seen as the union of two graphs,

Figure 5.7: Meta-Markov combination of $\mathcal{F}$ and $\mathcal{G}$.

it is not a graphical model.

Now, we give the details of the estimation of the parameters of the joint family. Suppose $y_A = (y_j^i)$, $j = 1, 2, 3$, $i = 1, \ldots, n_A$ and $y_B = (y_j^i)$, $j = 2, 3$, $i = 1, \ldots, n_B$, are the observations from the original families of Figure 5.6. The missing pattern is the same as in Table 5.1. In order to estimate the meta-Markov combination as the union of the two families of Figure 5.7, we need the maximum likelihood estimates of them. We provide the details for the family on the left, since a similar approach can be applied to the other one.

We assume a multivariate normal distribution $Y \sim N_3(0, \Sigma)$, with $\Sigma^{-1} = K_1 \in S^+(G)$, where G is the graph on the left of Figure 5.7. Since the graph is collapsible onto $\{2, 3\}$, $Y_{\{2,3\}}$ is a cut and the unknown density is factorised as $f(y_1, y_2, y_3) = f(y_1|y_2, y_3)f(y_2, y_3)$. Therefore, the parameters involved in the factorization are distinct and the estimates can be obtained by maximising separately the likelihood corresponding to these two factors. In particular, $\begin{pmatrix} Y_2 \\ Y_3 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} \end{pmatrix}\right)$, and

$(Y_1|Y_2 = y_2, Y_3 = y_3) \sim N(\mu_{1|23}, \sigma_{1|23})$, where

$$\begin{aligned}
\mu_{1|23} &= \frac{(\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23})y_2 + (\sigma_{13}\sigma_{22} - \sigma_{12}\sigma_{32})y_3}{\sigma_{22}\sigma_{33} - (\sigma_{23})^2}, \\
\sigma_{1|23} &= \frac{\sigma_{11}\sigma_{22}\sigma_{33} - \sigma_{23}(\sigma_{23}\sigma_{11} - \sigma_{13}\sigma_{12}) - (\sigma_{13})^2\sigma_{22}}{\sigma_{22}\sigma_{33} - (\sigma_{23})^2}.
\end{aligned}$$

From the conditional independence relationships we have $\sigma_{21}\sigma_{33} - \sigma_{31}\sigma_{23} = 0$ and $\sigma_{11}\sigma_{23} - \sigma_{21}\sigma_{13} = 0$. Hence,

$$\begin{aligned}
\mu_{1|23} &= \frac{\sigma_{13}\sigma_{22} - \sigma_{12}\sigma_{32}}{\sigma_{22}\sigma_{33} - (\sigma_{23})^2} y_3, \\
&= \beta_{1|23} y_3. \\
\sigma_{1|23} &= \sigma_{22}\frac{\sigma_{11}\sigma_{33} - (\sigma_{13})^2}{\sigma_{22}\sigma_{33} - (\sigma_{23})^2}.
\end{aligned}$$

Maximising the likelihood corresponding to $f(y_2, y_3)$ is equivalent to estimating the parameters σ_{33}, σ_{22} and σ_{23} as the sample variances and covariances, considering all the observations, yielding

$$\hat{\sigma}_{jj} = \frac{1}{n}\sum_{i=1}^{n}(y_j^i)^2, \quad j = 2, 3, \quad \hat{\sigma}_{23} = \frac{1}{n}\sum_{i=1}^{n}(y_2^i)(y_3^i).$$

Maximising the likelihood corresponding to $f(y_1|y_2, y_3)$ is equivalent to estimating $\beta_{1|23}$ and $\sigma_{1|23}$. They are estimated considering only n_A observations, yielding

$$\hat{\beta}_{1|23} = \frac{s_{13}s_{22} - s_{12}s_{32}}{s_{22}s_{33} - (s_{23})^2}, \quad \hat{\sigma}_{1|23} = s_{22}\frac{s_{11}s_{33} - (s_{13})^2}{s_{22}s_{33} - (s_{23})^2},$$

with

$$s_{jk} = \frac{1}{n_A}\sum_{i=1}^{n_A}(y_k^i)(y_j^i), \quad j, k = 1, 2, 3.$$

Now, we estimate the elements of the covariance matrix Σ. In particular $\hat{\sigma}_{11}$, $\hat{\sigma}_{13}$ and $\hat{\sigma}_{12}$ can be found solving the system of equations

$$\begin{cases}
\sigma_{13} = \frac{\sigma_{11}\hat{\sigma}_{23}}{\sigma_{21}} \\
\hat{\beta}_{1|23} = \frac{\sigma_{13}\hat{\sigma}_{22} - \sigma_{12}\hat{\sigma}_{32}}{\hat{\sigma}_{22}\hat{\sigma}_{33} - (\hat{\sigma}_{23})^2} \\
\hat{\sigma}_{1|23} = \hat{\sigma}_{22}\frac{\sigma_{11}\hat{\sigma}_{33} - (\sigma_{13})^2}{\hat{\sigma}_{22}\hat{\sigma}_{33} - (\hat{\sigma}_{23})^2}
\end{cases}.$$

Alternatively, we might apply the sweep operator (Schafer, 1997; Little and Rubin, 2002). After having repeated a similar procedure also for the second model, we obtain two estimates, $\hat{K}_1$ and $\hat{K}_2$, of the two covariance matrices K_1 and K_2.

Figure 5.8: Two marginal graphs G_A and G_B.

5.2.2 Meta-consistent families

Example 5.7. Consider Figure 5.8, where $Y_A \sim N_3(0,\Sigma)$, $\Sigma^{-1} \in S^+(G_A)$ and $Y_B \sim N_3(0,\Phi)$, $\Phi^{-1} \in S^+(G_B)$, corresponding to families $\mathcal{F}$ and $\mathcal{G}$, respectively. As we have seen in Section 4.2, the families are meta-consistent. The conditional independence relationships of the single models are $Y_2 \perp\!\!\!\perp Y_3|Y_1$ and $Y_2 \perp\!\!\!\perp Y_3|Y_4$. Hence, the meta-Markov combination is given by the family of distributions satisfying both $\rho_{23} = \rho_{12}\rho_{13}$ as well as $\rho_{23} = \rho_{24}\rho_{34}$, where ρ is the correlation coefficient for the joint family. If we put $\rho_{12} = a$, $\rho_{13} = b$, $\rho_{14} = c$, $\rho_{24} = d$, the correlation matrix R corresponding to the constraints is

$$R = \begin{pmatrix} 1 & a & b & c \\ a & 1 & ab & d \\ b & ab & 1 & \frac{ab}{d} \\ c & d & \frac{ab}{d} & 1 \end{pmatrix}.$$

We look for a joint family with no edge $(1,4)$. Therefore, if K is the concentration matrix, it should be $k_{14} = 0$. It is equivalent to

$$\det \begin{pmatrix} a & 1 & ab \\ b & ab & 1 \\ c & d & \frac{ab}{d} \end{pmatrix} = 0.$$

We have $a^3b^2d^{-1} + c + ab^2d - a^2b^2c - ad - ab^2d^{-1} = 0$ that can be written also as $cd(1-a^2b^2) = a(d^2(1-b^2)+b^2(1-a^2))$. There is no graphical combination corresponding to this equation and it might be studied from an algebraic point of view, giving the details of the algebraic surface obtained.

As we have stated in Section 4.2, since the families are meta-consistent we can also consider the graphical meta-Markov combination, using the Gaussian family corresponding to $G(\mathcal{F}\star\mathcal{G})$. It is represented by the graph in Figure 5.9 and it is equivalent to extend the marginal graphs by adding the edge $(2,3)$. In this case, the estimation of the parameter of the joint graphical model requires the EM algorithm, since there are no observed data for the edge $(2,3)$. We do not provide the details of the estimation since it is very similar to the previous examples.

Figure 5.9: Extension of the families and a joint family associated with them.

Chapter 6

Simulation studies

This chapter is devoted to assess the methodology that we have proposed in the previous chapters through some simulation studies. Firstly, we describe some procedures to generate data from a given Gaussian graphical model and then, we provide two detailed simulation studies.

6.1 Simulation of data from Gaussian graphical models

Data simulated from a given Gaussian graphical model are defined by a multivariate normal distribution with mean zero and a positive definite covariance matrix, such that its inverse has zeros whenever there is a missing edge on the corresponding graph. If the considered graph is relatively small, we can proceed as follows. We create a matrix with 1's along the main diagonal and a constant value ρ, say, for all the other elements, where ρ is chosen such that the matrix is positive definite. Then, we multiply the matrix by a positive value σ, say, and update the elements of the matrix either running the IPS algorithm in case of non-decomposable or large graphs, or for small decomposable graphs, imposing the constraints given by the structure of the graph. The IPS algorithm ensures that the concentration matrix is coherent with the conditional independence relationships given by the graph.

For example, consider graph G_1 and suppose we want to simulate data from a Gaussian graphical model corresponding to it. In this case, $Y \sim N_3(0, \Sigma)$, $\Sigma^{-1} \in S^+(G_1)$,

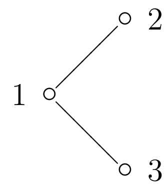


Figure 6.1: Graph G_1.

that is to say, $\sigma_{23} = \sigma_{12}\sigma_{13}\sigma_{11}^{-1}$, where $\Sigma = \{\sigma_{ij}\}$ and

$$\Sigma^{-1} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & 0 \\ h_{31} & 0 & h_{33} \end{pmatrix}.$$

If we consider a covariance matrix such as

$$\Sigma = \sigma \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix},$$

where σ is a positive real number and $\rho \in]-0.5, 1[$, we ensure that Σ is positive definite. Now, we can either run the IPS algorithm to find the exact covariance matrix corresponding to the graph, or we can directly use the constraints on the elements of the covariance matrix taking

$$\Sigma = \sigma \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho^2 \\ \rho & \rho^2 & 1 \end{pmatrix},$$

with $\rho \in]-1, 1[$ and $\sigma > 0$.

In general, it is convenient to run the IPS algorithm for large or non-decomposable graphs.

Another possibility is to generate a random correlation matrix, multiply it by a coefficient $\sigma > 0$, say, and then run the IPS algorithm to adjust the matrix to the structure of the graph. Various algorithms have been given in the statistics literature for generating random correlation matrices. The R package *ggm* (Marchetti and Drton, 2006) provides a method to generate a sample from a random correlation matrix exploiting Marsaglia and Olkin (1984) by forming AA^T from an $m \times n$ matrix A whose rows are random points on the unit m-sphere.

6.2 First simulation

The aim of this simulation is to investigate whether the proposal of estimating the joint Gaussian graphical model using a missing data perspective is reliable. We choose a Gaussian graphical model, say $\mathcal{G}$, and two marginal families whose combination is $\mathcal{G}$. We simulate a sample from $\mathcal{G}$ and deduce a sample also for the two marginal families, using the same data. We combine the two marginal families following the methods described in Chapters 4 and 5 and we estimate the parameters of the joint family, i.e., the concentration matrix. Then, we check whether the estimates obtained in this way are in accordance with the estimated concentration matrix of the distribution from which the data were generated. In order to do so, we compare the estimated elements of the two concentration matrices. We repeat the same procedure for two Gaussian graphical models $\mathcal{G}_1$ and $\mathcal{G}_2$ corresponding to the graphs G_1 and G_2. For each graph we select two marginal graphs whose combination is exactly G_1 and G_2, as in Figure 6.2 and Figure 6.3. We simulate samples of size $n = 25, 50, 100$ from the distributions $Y_A \sim N_3(0, \Sigma)$, $\Sigma^{-1} = \{\sigma^{ij}\} \in S^+(G_1)$, and $Y_B \sim N_4(0, \Phi)$, $\Phi^{-1} = \{\phi^{ij}\} \in S^+(G_2)$, respectively. Σ is chosen following the procedure described in the previous section with $\rho = 0.6$ and $\sigma = 2$ whereas Φ is chosen with $\rho = 0.7$ and $\sigma = 2$. Since the graph of the joint family is already defined, we estimate the parameters of the concentration matrix using the missing data perspective proposed in Chapter 5 (see Table 6.1 and Table 6.2). We call the estimated matrices as $\hat{\Sigma}_{comb}$, $\hat{\Sigma}^{-1}_{comb} = \{\hat{\sigma}^{ij}_{comb}\} \in S^+(G_1)$ and $\hat{\Phi}_{comb}$, $\hat{\Phi}^{-1}_{comb} = \{\hat{\phi}^{ij}_{comb}\} \in S^+(G_2)$.

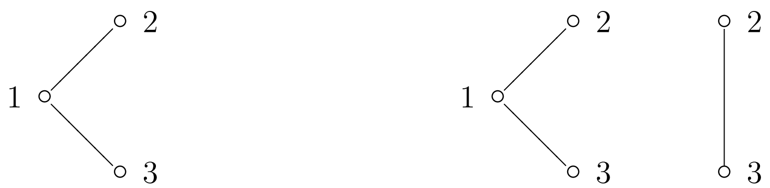


Figure 6.2: From left to right, graph G_1 and its marginal graphs.

	Y_1	Y_2	Y_3
A	n	n	n
B		n	n

Table 6.1: Missing pattern for the first joint family considered.

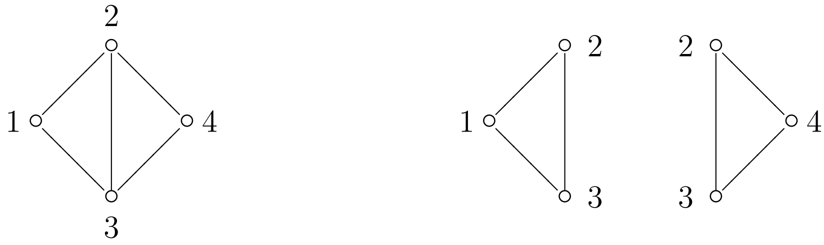
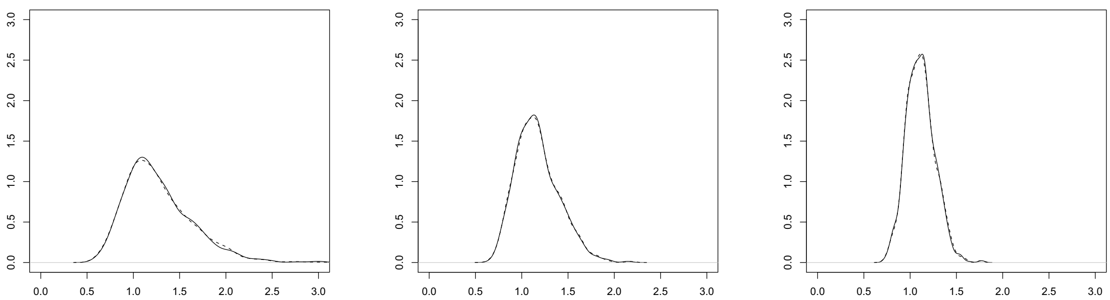


Figure 6.3: From left to right, graph G_2 and its marginal graphs.

	Y_1	Y_2	Y_3	Y_4
A	n	n	n	
B		n	n	n

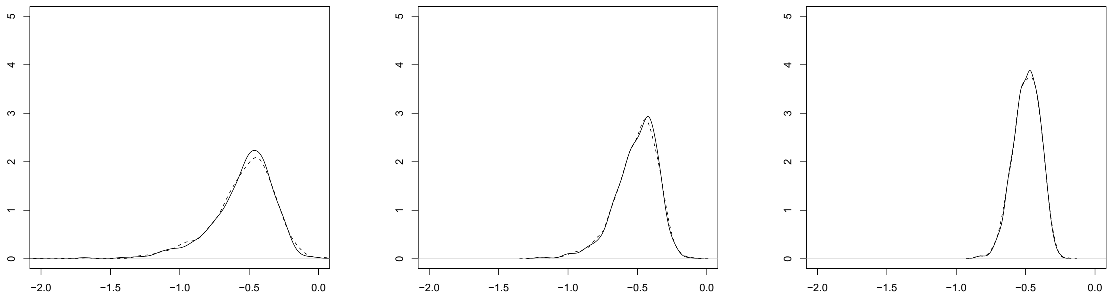
Table 6.2: Missing pattern for the second joint family considered.

Finally, the estimates of the elements of the two concentration matrices are compared. The simulations are carried out using R 2.6.0 and the package mimR (Højsgaard, 2004, 2007) that is an interface between MIM (Edwards, 2000) and R. In Figures 6.4-6.8 the results for the first Gaussian graphical model $\mathcal{G}_1$ are shown. We note that the two estimates are almost identical.



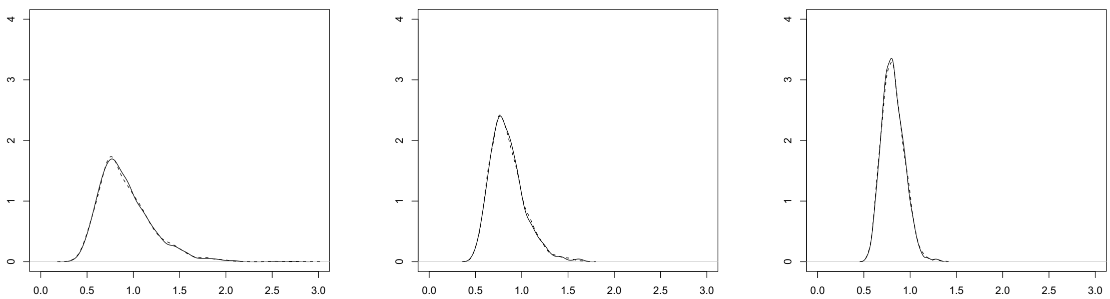
(a) n=25 (b) n=50 (c) n=100

Figure 6.4: $\hat{\sigma}^{11}$ (straight line) and $\hat{\sigma}^{11}_{comb}$ (dotted line) with (a) 25 observations, (b) 50 observations (c) 100 observations.



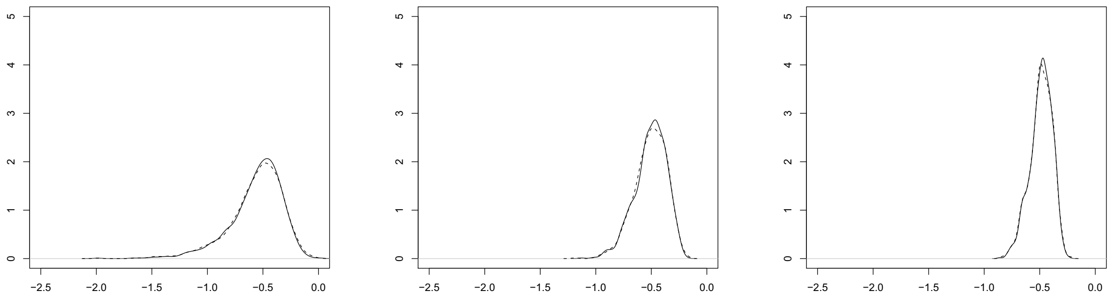
(a) n=25 (b) n=50 (c) n=100

Figure 6.5: : $\hat{\sigma}^{12}$ (straight line) and $\hat{\sigma}^{12}_{comb}$ (dotted line) with (a) 25 observations, (b) 50 observations (c) 100 observations.



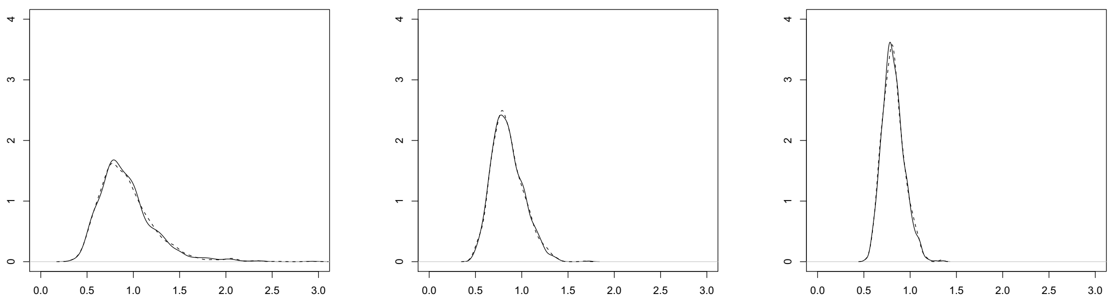
(a) n=25 (b) n=50 (c) n=100

Figure 6.6: $\hat{\sigma}^{13}$ (straight line) and $\hat{\sigma}^{13}_{comb}$ (dotted line) with (a) 25 observations, (b) 50 observations (c) 100 observations.



(a) n=25 (b) n=50 (c) n=100

Figure 6.7: $\hat{\sigma}^{22}$ (straight line) and $\hat{\sigma}^{22}_{comb}$ (dotted line) with (a) 25 observations, (b) 50 observations (c) 100 observations.



(a) n=25 (b) n=50 (c) n=100

Figure 6.8: $\hat{\sigma}^{23}$ (straight line) and $\hat{\sigma}^{23}_{comb}$ (dotted line) with (a) 25 observations, (b) 50 observations (c) 100 observations.

6.3 Second simulation

In this simulation, we use the methods proposed in the previous chapters to examine the possibility of conducting several small studies instead of a bigger one. To do this, we consider a Gaussian graphical model, say $\mathcal{G}$, and two marginal models, say $\mathcal{G}_A$ and $\mathcal{G}_B$ whose combination is $\mathcal{G}$, corresponding to the graphs in Figure 6.3. We simulate a sample of size n for $\mathcal{G}$ from the distribution $Y \sim N_4(0, \Sigma)$, $\Sigma^{-1} \in S^+(G_2)$. From the same dataset, we randomly select two samples of size $n_A = n_B = n/2$ for the variables involved in the families $\mathcal{G}_A$ and $\mathcal{G}_B$, respectively. Afterwards, we select through an unrestricted stepwise backward selection (Edwards, 2000) the Gaussian graphical model for the three samples. We repeat the procedure 10000 times and we count (1) the number of times the stepwise procedure finds the initial family $\mathcal{G}$, (*cont1*); (2) the number of times the stepwise procedure finds the marginal family $\mathcal{G}_A$, (*cont2*); (3) the number of times the stepwise procedure finds the marginal family $\mathcal{G}_B$, (*cont3*); (4) the number of times the stepwise procedure finds both the marginal families $\mathcal{G}_A$ and $\mathcal{G}_B$, (*cont4*). Note that, if the stepwise procedure finds both the marginal families $\mathcal{G}_A$ and $\mathcal{G}_B$, it also finds the initial one, because the combination of $\mathcal{G}_A$ and $\mathcal{G}_B$ is exactly $\mathcal{G}$. This means that we are interested in comparing $cont1$ with $cont4$, only when they are obtained with the same number of observations.

The stepwise procedure uses n observations to select the initial family $\mathcal{G}$, whereas only half of them are used for the marginal families. Therefore, to compare *cont1* with *cont4* we have to compare the values obtained with $n/2$ and n observations, respectively. So, for example, *cont4* = 1479 is compared with *cont1* = 337 (see Table 6.3). In this case, they are both computed with 40 observations. Moreover, even if $cont2$ and $cont3$ alone are not of direct interest to us, they are both greater than $cont1$. For example, with 50 observations $cont2 = 6032$, $cont3 = 6020$ and $cont1 = 1097$.
Table 6.3 shows that, with the same number of observations, the combination of two subfamilies finds more often the true joint family. This can be motivated by noticing that the subgraphs consist of three variables, whereas the original graph consist of six variables, but the structure of both the original graph and the subgraphs is estimated with the same number of observations.

This result appears to be quite interesting for the problem considered in this thesis. It can be useful in the case of composing many graphical models in order to obtain a larger model, especially in the $n < p$ paradigm, i.e., when the number of observations is less than the number of variables under analysis. This happens very often with microarray data, where the sample size is usually less than the number of data (genes). Hence, instead of estimating a model including all the variables, one might compose smaller models in such a way that a more appropriate number of observations is used, if compared with the number of variables involved. For example, suppose we have a graph representing 100 variables. If we have only a small number of observation, 25 say, instead of estimating the initial graphical model, we could estimate smaller models where the number of variables is less than 25. However, it is not always feasible, because it depends on the structure of the initial graph and on the possibility of finding all the subgraphs where n is greater than the number of the vertices of the subgraph. Nevertheless, it might be achieved in a recursive way. If a submodel cannot be estimated because the number of variables of the submodel n is not greater than p, then the submodel itself might be composed and estimated by combining smaller models, and so on.

n	n_A	n_B	cont1	cont2	cont3	cont4
40	20	20	337	42	46	0
50	25	25	1097	445	455	27
60	30	30	2194	1305	1315	160
70	35	35	3367	2519	2587	620
80	40	40	4230	3852	3779	1479
90	45	45	5019	5027	4962	2476
100	50	50	5712	6032	6020	3601
120	60	60	6540	7606	7654	5809
140	70	70	7251	8552	8576	7314
160	80	80	7808	9189	9199	8462

Table 6.3: Results of a stepwise procedure to find the structure of a joint family by combining two subfamilies. Here, n is the number of observation used for the initial family, whereas n_A and n_B are the number of observations used for the subfamilies.

Chapter 7

Conclusions

This thesis introduced a general context for combination of families of distributions, focussing on undirected Gaussian graphical models. Throughout the thesis we have given several examples, which provide beneficial insights for the more general problem of combination of large or many models. A motivating example for the choice of this research problem regarded the possibility of combining relevant biological networks, when they can be interpreted as Gaussian graphical models.

The central idea was to consider the wider perspective of combining families of distributions. Hence, the starting point of this work were two marginal families defined over two sets of variables. We supposed that they represented two marginal families of an unknown higher dimensional family of distributions. Although it is not straightforward or even always possible to find a joint family of distributions, we specifically restricted our attention to the feasible cases. To this aim, the compatibility of the families was studied by introducing the concepts of meta-consistent families, as families whose induced distributions over the common variables is the same, and quasi-consistent families, as families whose induced distributions over the common variables is only partially the same. The joint family was established by an appropriate combination of the two initial families. Two types of combination were presented: the meta-Markov combination and the quasi-Markov combination. The first combines only the consistent distributions, while the second combines all the distributions of the two families. Further, we showed that these two combinations can be used for composing both meta-consistent and quasi-consistent families. The choice between meta-Markov combination and quasi-Markov combination is taken in the light of the information available on the establishment of the initial families. We studied also conditions under which the two combinations are equivalent and, in general, the properties related to each type of combination. Overall, we showed that the meta-Markov combination preserves the marginal families, property that is not always satisfied by the quasi-Markov combination.

With this particular framework in mind, we then studied the combination of two Gaussian graphical models, which can be seen as families of distributions respecting conditional independence constraints. We needed to distinguish between graphical and non-graphical combinations. In the first case, the multivariate normal distribution associated with the dependence graph of the meta-Markov (or quasi-Markov) combination corresponds to the meta-Markov (or quasi-Markov) combination. This is not the case for combinations of non-graphical type. We provided also one example of a combination of three models, showing an idea of how might be faced the composition of a larger number of models.

The following part of the work was concerned with the estimation of the joint family of distributions. In the majority of the cases, we achieved this by proposing an estimation process as if we were handling missing data problems. The standard EM algorithm was exploited, in the version given for mixed graphical models. However, when the missing data context was not necessary, we also presented other direct ways of estimation. We attempted to give an exhaustive view of the possible situations arising when estimating a combination of Gaussian graphical models. To this purpose, we analysed the estimation of joint families corresponding to graphical and non graphical combinations, decomposable and non-decomposable joint graphs, marginal graphs collapsible onto the common variables. We further noted that the type of available data for the original families determines the complexity of the estimation procedure.

The last chapter considered two simulation studies. On the one hand, we wanted to show the reliability of the missing data perspective previously introduced, on the other hand, we wanted to suggest a possible way to study and estimate larger families by composing smaller ones. This procedure might be interesting especially when the sample size is less than the number of variables involved, and the standard methods of inference for graphical models cannot be used.

Finally, we highlight some research directions for future work both in the theoretic and applied framework. When composing more than two families of distributions, it becomes interesting to study the associative property for the two types of combination, i.e., check whether the order of composition changes or not the final combination. In fact, this condition does not generally hold for both types of combination.

Another future orientation is to conduct an application with some relevant regulatory networks or biological pathways, so that our framework can be specialised for an applied context.

An interesting point that deserves our attention is the Bayesian framework. For example, consider the case in which we are given two independent studies and for each study there is available some information provided by an expert. When possible, it is of interest to combine the information of the experts and then also the information acquired from the studies.

The theory of algebraic statistics (Pistone et al., 2001; Pachter and Sturmfels, 2005) might be another approach to study the combination of models and it might in particular throw interesting light on the non-graphical combinations.

Appendix A

Gaussian graphical models

Here, we concisely review some relevant concepts about graphs and the theory of graphical models. We will follow the exposition of Lauritzen (1996), to which we refer for further details. Other reference books on graphical models are Whittaker (1990) and Edwards (2000).
Let V a finite set with $|V| = p$.

Definition A.1. An undirected graph is a pair $G = (V, E)$ where V is a set of vertices, and $E \subseteq V \times V$ is a finite set of edges between these vertices.

The edges are indicated as $(\alpha, \beta) \in E$ or equivalently $(\beta, \alpha) \in E$. By definition, an undirected graph has no multiple edges and no loops. Vertices that are joined by an edge are said to be neighbours, i.e. $\alpha \sim \beta$. If this is not the case, we use the notation $\alpha \not\sim \beta$. The boundary of a subset A of V, bd(A), is the set of vertices in $V \setminus A$ that are neighbours of vertices in A. The closure of A, cl(A), is cl(A)= $A \cup$ bd(A).

If $A \subset V$ is a subset of the vertex set of a graph G, it induces a subgraph $G_A = (A, E_A)$ where $E_A = E \cap (A \times A)$. The set of edges E_A is obtained from G by keeping edges with both endpoints in A. A graph or a subgraph is complete if all its vertices are joined by an edge. A complete subgraph that is not contained within another complete subgraph is called a clique.

A sequence of n different vertices, $\alpha_1 = \alpha, \cdots, \alpha_n = \beta$, is called a path of length $n-1$ from α to β, if $\alpha_i \sim \alpha_{i+1}$ for $i = 1, \cdots, n-1$. For subsets A, B, S of V, S separates A from B in G if every path from A to B intersects S. A connected component of a graph is a subset of vertices $A \subset V$, such that any vertex in the subgraph G_A is connected by some path to any other vertex. An undirected graph G is said to be collapsible onto

$A \subset V$, if every connected component of $V \setminus A$ has complete boundary in G (Asmussen and Edwards, 1983).

A partitioning of V into a triple (A, B, S) of subsets of V forms a decomposition of an undirected graph G if (i) $V = A \cup B \cup S$, (ii) S separates A from B, (iii) S is complete. The decomposition is proper if both A and B are non-empty sets and the components of G are the induced subgraphs $G_{A\cup S}$ and $G_{B\cup S}$. A graph is said to be prime if no proper decomposition exists. It can be shown that any finite undirected graph can be recursively decomposed into its uniquely defined prime components. A central notion, strictly linked with that of a decomposition, is that of a decomposable graph.

Definition A.2. An undirected graph is decomposable if its prime components are cliques.

So far, we have listed some important notions of graph theory. To introduce a statistical setting that is related to a graph, with each vertex $v \in V$ we associate a random variable Y_v, taking values in a sample space $\mathcal{Y}_v$. For $A \subseteq V$, we let $Y_A = (Y_v)_{v\in A}$ with values in the product space $\mathcal{Y}_A = \times_{v\in A}\mathcal{Y}_v$. Along all the work, we consider only distributions that are absolutely continuous w.r.t. a product measure, thus, by a probability distribution over $A \subseteq V$, we indicate a joint distribution for Y_A over $\mathcal{Y}_A$. If f is the distribution over V and $A, B \subset V$, then f_A denotes the marginal distribution of Y_A, and $f_{B|A}$ the conditional distribution of Y_B given $Y_A = y_A$.

Now, we see the connection between the specific constraints implied by an undirected graph G and the distributions defined over the variables associated with G. Firstly, we need the notion of conditional independence, (Dawid, 1979, 1980).

Definition A.3. Random variables X and Y are conditionally independent given random variable Z (or shortly $X \perp\!\!\!\perp Y|Z$), if X and Y are independent in their joint distribution given $Z = z$, for any value of z.

The above definition can also be reformulated using equivalent expressions in term of probability distributions (see Lauritzen, 1996, pag. 29 for details). Then, the Markov properties provide the link between the graph and the distributions defined over it.

Definition A.4. A distribution f over V is said to obey the pairwise Markov property with respect to G if $\alpha \not\sim \beta$ implies $\alpha \perp\!\!\!\perp \beta|V \setminus \{\alpha, \beta\}$.

Definition A.5. A distribution f over V is said to obey the local Markov property with respect to G if $\alpha \perp\!\!\!\perp V\setminus \mathrm{cl}(\alpha) \mid \mathrm{bd}(\alpha)$, for all α in V.

Definition A.6. A distribution f over V is said to obey the global Markov property with respect to G if, for any decomposition (A, B, S) of G, $Y_A \perp\!\!\!\perp Y_B | Y_S$ (or shortly $A \perp\!\!\!\perp B | S$).

In general, the global Markov property implies the local Markov property, which, in turn implies the pairwise Markov property. If the distribution is strictly positive, the three Markov properties all are equivalent (Pearl and Paz, 1987). At this point, we are ready to give the definition of a graphical model and a Gaussian graphical model.

Definition A.7. A graphical model is a family of probability distributions such that all the distributions of the family obey the pairwise Markov property with respect to G.

The graph G is called dependence graph. In general, a dependence graph of a joint family of distributions $\mathcal{F}$ is a graph with a missing edge between two vertices α and β, if and only if $\alpha \perp\!\!\!\perp \beta \,|\, V \setminus \{\alpha, \beta\}$, for all $f \in \mathcal{F}$. It is the smallest graph such that all the distributions are pairwise Markov with respect to it.

Definition A.8. A Gaussian graphical model, also known as a covariance selection model (Dempster, 1972), is a family of multivariate normal distributions, $Y \sim N_p(\mu, \Sigma)$, with Σ positive definite and such that all the distributions of the family obey the pairwise Markov property with respect to G.

In this case, the local and global Markov property are equivalent to the pairwise Markov property. If we take $K = \Sigma^{-1}$, the latter is equivalent to say that if two vertices α and β are not neighbours, then it holds that $k_{\alpha\beta} = 0$ and conversely (see Lauritzen, 1996, Proposition 5.2).

Bibliography

Asmussen, S. and Edwards, D. (1983) Collapsibility and response variables in contingency tables. *Biometrika*, **70**, 567–578.

Buhl, S. L. (1993) On the existence of maximum likelihood estimators for graphical Gaussian models. *Scandinavian Journal of Statistics*, **20**, 263–270.

Chiogna, M., Massa, M. S. and Romualdi, C. (2007) Effect of normalization on detecting differentially expressed genes with cDNA microarray experiments. Submitted.

Dawid, A. P. (1979) Conditional independence in statistical theory. *Journal of the Royal Statistical Society Series B*, **41**, 1–31.

— (1980) Conditional independence for statistical operations. *Annals of Statistics*, **8**, 598–617.

Dawid, A. P. and Lauritzen, S. L. (1993) Hyper–Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics*, **21**, 1272–1317.

Dempster, A. (1972) Covariance selection. *Biometrics*, **28**, 157–175.

Didelez, V. and Pigeot, I. (1998) Maximum likelihood estimation in graphical models with missing values. *Biometrika*, **85**, 960–966.

Edwards, D. (2000) *Introduction to graphical modelling*. New York: Springer-Verlag.

Fienberg, S. E. and Kim, S.-H. (1999) Combining conditional log-linear structures. *Journal of the American Statistical Association*, **94**, 229–239.

Frydenberg, M. (1990) The chain graph Markov property. *Scandinavian Journal of Statistics*, **17**, 333–353.

Geng, Z., Wan, K. and Tao, F. (2000) Mixed graphical models with missing data and the partial imputation EM algorithm. *Scandinavian Journal of Statistics*, **27**, 433–444.

Glass, G. V. (1976) Primary, secondary, and meta-analysis of research. *Educational Researcher*, **5**, 3–8.

Goh, K.-I., Cusick, M., Valle, D., Childs, B., Vidal, M. and Barabasi, A.-L. (2007) The human disease network. *Proceedings of the National Academy of Sciences*, **104**, 8685–8690.

Gradshteyn, I. S. and Ryzhik, I. M. (2000) *Table of integrals, series, and products*. San Diego, CA: Academic Press Inc., sixth edn.

Hedges, L. V. and Olkin, I. (1985) *Statistical methods for meta-analysis*. Orlando, FL: Academic Press Inc.

Højsgaard, S. (2004) The mimR package for graphical modelling in R. *Journal of Statistical Software*, **11**, (6).

— (2007) Addendum to the paper: The mimR package for graphical modelling in R. *The Comprehensive R Archive Network*.

Jiroušek, R. and Vejnarova, J. (2003) Construction of multidimensional models by operator of composition: current state of art. *Soft Computing*, **7**, 328–335.

Kim, S.-H. (2001) Towards a statistical foundation in combining structures of decomposable graphical models. *Tech. Rep. 1-2*, Division of Applied Mathematics, KAIST, Daejeon, 305-701, South Korea.

— (2006) Conditional log-linear structures for log-linear modelling. *Computational Statistics and Data Analysis*, **50**, 2044–2064.

Lauritzen, S. L. (1996) *Graphical models*. Clarendon Press, Oxford.

Little, R. J. A. and Rubin, D. B. (2002) *Statistical analysis with missing data*. Hoboken, NJ: Wiley-Interscience, second edn.

Marchetti, G. M. and Drton, M. (2006) ggm: an R package for Gaussian graphical models. *The Comprehensive R Archive Network. http://cran.r-project.org/*.

Marsaglia, G. and Olkin, I. (1984) Generating correlation matrices. *SIAM Journal on Scientific and Statistical Computing*, **5**, 470–475.

Pachter, L. and Sturmfels, B. (2005) *Algebraic Statistics for Computational Biology*. New York: Cambridge Univ. Press.

Pearl, J. and Paz, A. (1987) Graphoids: A graph based logic for reasoning about relevancy relations. In *Advances in Artificial Intelligence – II* (eds. B. D. Boulay, D. Hogg and L. Steel), 357–363. Amsterdam, The Netherlands.

Pistone, G., Riccomagno, E. and Wynn, H. P. (2001) *Algebraic Statistics. Computational Commutative Algebra in Statistics.* Boca Raton, FL: Chapman & Hall/CRC.

Schafer, J. L. (1997) *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Speed, T. P. and Kiiveri, H. (1986) Gaussian Markov distributions over finite graphs. *Annals of Statistics*, **14**, 138–150.

Studený, M. (2004) *Probabilistic Conditional Independence Structures.* New York: Springer-Verlag.

Whittaker, J. (1990) *Graphical models in applied multivariate statistics.* Chichester: John Wiley & Sons Ltd.