Contemporary Mathematics

Combining Statistical Models

M. Sofia Massa and Steffen L. Lauritzen

Abstract. This paper develops a general framework to support the combination of information from independent but related experiments, by introducing a formal way of combining statistical models represented by families of distributions. A typical example is the combination of multivariate Gaussian families respecting conditional independence constraints, i.e. Gaussian graphical models. Combining information from such models, represented by their dependence graphs, yields a formal basis for what could suitably be termed structural meta analysis. We consider issues of combination of pairs of distributions, extending the concept of meta-Markov combination introduced by Dawid and Lauritzen. The proposed theory is then applied to the special case of graphical models.

1. Introduction

Consider a number of independent statistical investigations that not necessarily address the same question but have some aspects in common, in particular the variables under study. This may be the case, for example, when different laboratories analyze related phenomena using different methods or when different sources of information relate to different parts of the same experiment. It may happen when small experiments are conducted with the purpose to reconstruct and make inference about a system where a global investigation involving all relevant variables for some reason is not feasible. Or it may happen simply because studies are performed independently and under different circumstances.

In a biological context, we may think of studies involving regulatory networks or signaling pathways with common elements (genes, proteins, etc.). The networks might be the results of independent experiments and the interest is in constructing a meaningful joint network to encapsulate the biological understanding. We mention the study of [**9**], where a large number of studies of diseases and related genes are combined to form "the human diseasome" bipartite network which can be seen as a simple but prominent and large scale example of the type of structural meta-analysis which we have in mind. Meta-analysis is usually meant to integrate and

2000 *Mathematics Subject Classification.* Primary 62H99; Secondary 68R99.

Key words and phrases. collapsibility; consistent distributions; cut; lower Markov combination; meta-consistent families; meta-Markov combination; quasi-consistent families; structural meta-analysis, super Markov combination; upper Markov combination.

The first author was supported by University of Padua grant CPDR070805.

combine evidence from different studies that are well-designed, address the same questions and use similar outcome measures (see [**8, 10**]). In our setup, we extend the concept of meta-analysis so that it covers the more general process of creating a consistent overview of information in several independent and related studies.

Some literature on similar topics already exists. [**2**] focus on constructing prior distributions for decomposable graphical models from prior distributions on its cliques, and considers in this connection the notion of meta-Markov combination of the models. This paper extends and generalizes the latter. [**6**] 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. [**14**] proposes a method to combine the structures of marginal decomposable graphical models. He focuses on the construction of a joint graph, analyzing the structural properties and describing a grafting process, based on the minimal connectors of the marginal graphs. Note that his idea is that of finding all the possible structures compatible with the initial ones. [**3**] yield a general axiomatic theory for combination of compatible objects in so-called conditional products and [**11**] introduce a method for constructing discrete multidimensional distributions by combining many low-dimensional ones using non-symmetric composition operations in the case where distributions are not necessarily compatible. [**21**] propose an algorithm to integrate partial ancestral graphs with overlapping variables without conflictual information.

In contrast, our attention is focused on constructing combinations which in some sense are as simple as possible. In our general setup, we consider families of distributions defined over subsets of variables with respect to a product measure which respect some form of compatibility and we are interested in constructing a joint family of distributions over all the variables of interest. We develop a general framework for combination of families of distributions and a first development of formal concepts to underpin the ideas mentioned above. Although we in the present paper focus on the general aspects, the prime application we have in mind is the combination of families of distributions respecting conditional independence constraints with respect to a graph G, i.e. graphical models.

We are conscious that it is also important to describe a procedure for combining the inferences from each of the models in the combination when data are available in some form. However, in this paper we investigate the combination of the models and combination of inferences will be discussed elsewhere.

The paper is organized as follows. In Section 2 we introduce a few motivating examples. Section 3 studies conditions for compatibility for distributions and families of distributions. Section 4 deals with combination of distributions and families of distributions and studies the properties of such combinations in some detail. Section 5 applies the ideas of the previous sections to the combination of Gaussian graphical models and gives some examples of combinations. We conclude with a general discussion and some possibilities for future work.

2. Motivating examples

Before developing the general concepts, we provide some simple examples to introduce the ideas behind the concepts. To begin with, we give the definition of a

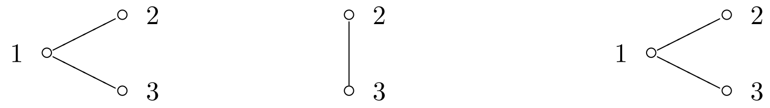

FIGURE 2.1. From left to right, graphs G_A and G_B and a possible combination of them.

Gaussian graphical model with undirected graph $G = (V, E)$ with $|V| = p$, following [**15**].

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

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 $Y_V \sim N_p(0, \Sigma)$ where $\Sigma^{-1} \in S^+(G)$, and $S^+(G)$ is the set of symmetric positive definite matrices whose elements are equal to zero whenever there is no edge between the corresponding elements of V. Note that we use *graph* to indicate the conditional independence structure of the model and *family* to indicate both the graph and the set of distributions conforming with the conditional independence structure.

EXAMPLE 2.1. The two leftmost graphs in Figure 2.1 represent two 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 (Y_2 and Y_3) but not all variables in common, and it is of interest to construct a model that combines the initial pieces of information in the best way. Each graphical model represents a family of probability distributions. For a simple combination of them to make sense, we have to ensure that at least one pair of distributions exist within the models which induce the same distribution over the variables in common. It can be easily shown in this case that the second graphical model is indeed the marginal family of the first one, hence we may simply take a combination with the Y_{23}-marginal determined by the leftmost graph of the figure, the joint family being represented by the rightmost graph in Figure 2.1. □

EXAMPLE 2.2. In the first example there is a particular simple relation between the families, i.e. one family is the marginal of the other family. Figure 2.2 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, there is no obvious way of defining the combination; in particular, it is not naturally given by the graph obtained as the union of the two graphs as this would represent very different conditional independence restrictions. □

EXAMPLE 2.3. In Figure 2.3, there are no conditional independence relationships expressed by the two graphs and one possible graph for the combined model is

FIGURE 2.2. From left to right, graphs G_A and G_B. It is not straightforward to define a combination of them.

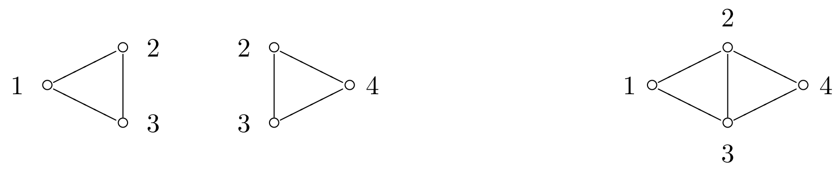

FIGURE 2.3. On the left, two complete graphs. On the right, a graph representing a possible combination of them.

the union of the two graphs. Nevertheless, there are several different graphs which are compatible with the independence structure of the initial ones, for example the four graphs containing the cycle (1243) but differing by the presence or absence of the edges $(2, 3)$ and $(1, 4)$. Our approach for combining models chooses the simplest model which is compatible with the initial graphical models, in this case represented by the graph to the right in Figure 2.3 having the edge $(1, 4)$ absent. $\square$

3. Consistency issues

We begin by describing some notation that we use throughout the paper. Let V be a set of variables. We let Y_v denote a random variable 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 $\mathcal{Y}_A = \times_{v \in A} \mathcal{Y}_v$. By a probability distribution over $A \subseteq V$, we indicate a joint distribution for Y_A over $\mathcal{Y}_A$. If f is a 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 $Y_{B \setminus A}$ given $Y_A = y_A$. For a family of distributions $\mathcal{F} = \{f |\, f \text{ distribution over } A\}$ and $C \subseteq A$, $\mathcal{F}^{\downarrow C}$ denotes the family of marginal distributions induced by $\mathcal{F}$ over C. If $C = \emptyset$, $\mathcal{F}^{\downarrow C}$ is trivial containing only the constant equal to one.

The general formulation of the problem addressed in this paper is as follows. We consider two sets of variables A and B which are possibly different subsets of a finite set V of variables, and two families $\mathcal{F}$ and $\mathcal{G}$ of distributions for Y_A and Y_B. We ideally search for a joint family of distributions $\mathcal{H}$ for $Y_{A \cup B}$, such that

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

We restrict attention to distributions which are absolutely continuous w.r.t. product measures $\otimes_{v \in A} \mu_v$ and $\otimes_{v \in B} \mu_v$, where the measures $\mu_v, v \in V$ are fixed. We can therefore use the term distribution synonymous with the term density.

When combining such families we must ensure that they in some way induce the same distributions over the variables in common. Following the works of [**12, 13, 22, 23**], we firstly define consistency for distributions.

DEFINITION 3.1. 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 is now extended to families following [**2**]:

DEFINITION 3.2. Two families of distributions $\mathcal{F}$ and $\mathcal{G}$ 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 notion of meta-consistency is very restrictive. We shall therefore consider combination of families also in cases where this does not hold. Therefore we introduce a weaker form of compatibility by only requiring the existence of comparable distributions. Let

$$\mathcal{F}^{\mathcal{G}} = \{f \in \mathcal{F} \mid \exists\, g \in \mathcal{G} : f << g\}, \tag{3.1}$$

where $f << g$ is set to mean that the densities satisfy

$$g_{A\cap B}(x_{A\cap B}) = 0 \implies f_{A\cap B}(x_{A\cap B}) = 0,$$

i.e. that $f_{A\cap B}$ is dominated by $g_{A\cap B}$. We define

DEFINITION 3.3. Two families $\mathcal{F}$ of distributions over A and $\mathcal{G}$ of distributions over B are said to be quasi-consistent if $\mathcal{F}^{\mathcal{G}} = \mathcal{F}$, and $\mathcal{G}^{\mathcal{F}} = \mathcal{G}$.

Clearly, if two families are meta-consistent, they are also quasi-consistent.

4. Markov combinations

4.1. Combination of distributions. [**2**] introduce the Markov combination of two consistent distributions as

DEFINITION 4.1. The Markov combination of a pair f and g of consistent distributions is defined as

$$f \star g = f \cdot g / g_{A\cap B}.$$

If $A \cap B = \emptyset$, $f_{A\cap B} = g_{A\cap B} = f_{\emptyset} = 1$ and $f \star g = f \cdot g$. It is worth noting that the Markov combination of f and g preserves the marginal distributions and it is the simplest possible with that property, in the sense that it has maximal entropy among all distributions with the given marginals, as detailed below.

PROPOSITION 4.1. *The Markov combination $f \star g$ preserves the marginal distributions*

$$(f \star g)^{\downarrow A} = f, \quad (f \star g)^{\downarrow B} = g.$$

PROOF. This is immediate, see also [**2**]. □

PROPOSITION 4.2. *Let $H_{\mathcal{F}}(A)$ be the entropy of a family of distributions $\mathcal{F}$ over A. Then, $H_{f\star g}(A \cup B) \geq H_h(A \cup B), \forall h \in \mathcal{Q}$, where $\mathcal{Q} = \{h : h_A = f, h_B = g\}$.*

PROOF. This follows from the calculation below which establishes and exploits that the entropy is a submodular function on the subsets of a finite set (see [**19**]) so it holds that

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

with equality if and only if A and B are conditionally independent given $A \cap B$. Then, since $A \perp\!\!\!\perp B \,|\, C$ with respect to $f \star g$ we have

$$\begin{aligned} H_{f\star g}(A \cup B) &= H_{f\star g}(A) + H_{f\star g}(B) - H_{f\star g}(A \cap B), \\ &= H_f(A) + H_g(B) - H_f(A \cap B), \\ &= H_h(A) + H_h(B) - H_h(A \cap B) \geq H_h(A \cup B) \end{aligned}$$

for any h which has marginals equal to f and g respectively. □

The Markov combination is commutative from the way it is defined. Now, consider three pairwise consistent distributions f, g and h defined over A, B, C, respectively. If $A \cap B = A \cap C = B \cap C = A \cap B \cap C$, the Markov combination also satisfies

$$f \star (g \star h) = (f \star g) \star h = (f \star h) \star g,$$

but it is not in general associative. If $A \cap C \subset B$ there is a limited associativity in the sense that

$$f \star (g \star h) = (f \star g) \star h \neq (f \star h) \star g. \tag{4.1}$$

This is seen by applying the definition of Markov combination twice; we get

$$f \star (g \star h) = \frac{f \cdot g \cdot h}{f_{A\cap(B\cup C)} \cdot g_{B\cap C}} = \frac{f \cdot g \cdot h}{f_{A\cap B} \cdot g_{B\cap C}},$$

where we have also used that $A \cap C \subset B$, and similarly

$$(f \star g) \star h = \frac{f \cdot g \cdot h}{f_{A\cap B} \cdot h_{(A\cup B)\cap C}} = \frac{f \cdot g \cdot h}{f_{A\cap B} \cdot h_{B\cap C}}.$$

The consistency condition $h_{B\cap C} = g_{B\cap C}$ gives the first equality, whereas

$$(f \star h) \star g = \frac{f \cdot h \cdot g}{f_{A\cap C} \cdot h_{(A\cup C)\cap B}},$$

which clearly is different in general.

If two distributions f and g are not consistent, there is no single obvious way of combining them, so we follow [**11**] and define

Definition 4.2. The operator of right composition of f and g is defined as

$$f \triangleright g = \begin{cases} f \cdot \frac{g}{g_{A\cap B}} & \text{if } f << g \\ \text{undefined} & \text{otherwise} \end{cases}.$$

Definition 4.3. The operator of left composition of f and g is defined as

$$f \triangleleft g := g \triangleright f.$$

Note that the expressions make sense when we define $0/0 = 0$ since $f_{A\cap B}(x_{A\cap B}) = 0$ implies $f(x) = 0$ almost everywhere.

The two operators of composition were originally introduced for discrete distributions and they are equivalent to the Markov combination when the two distributions f and g are consistent. Hence, if f and g are two consistent distributions, $f \triangleright g = f \triangleleft g = f \star g$.

[**11**] say that two distributions f and g are a perfect pair if $f \triangleright g = f \triangleleft g$. Clearly, a pair of distributions is perfect if and only if they are consistent because $f_{A\cap B} = g_{A\cap B}$ implies $f \triangleright g = f \triangleleft g$ and *vice versa*.

The operators of left and right composition preserve only one marginal distribution. In particular, $(f \triangleright g)^{\downarrow A} = f$ and $(f \triangleleft g)^{\downarrow B} = g$, provided the expressions are well-defined. The operators $f \triangleright g$ and $f \triangleleft g$ are neither commutative nor associative in general. However, the combination satisfies a fundamental conditional independence relation and can as such be considered a type of Markov combination:

PROPOSITION 4.3. *Consider two distributions f defined over A, and g defined over B. It then holds that*

$$A \perp\!\!\!\perp B \,|\, (A \cap B), \tag{4.2}$$

for any of the distributions $f \star g$, $f \triangleleft g$, and $f \triangleright g$.

PROOF. This follows directly from the definition of conditional independence and Definitions 4.1, 4.2, 4.3. □

4.2. Combination of families. This section is concerned with lifting the notions of combinations for distributions to families of distributions.

We have two different specific situations in mind. In the first situation we imagine the two families $\mathcal{F}$ for Y_A and $\mathcal{G}$ for Y_B having been established in their respective laboratories with a high degree of certainty so that restrictions in each of the families necessarily must be respected and taken at their face value. Below we define a lower Markov combination which reflects this by reducing the families to satisfy this demand. In the second situations the families may only reflect conjectures about the issue under study and we would like to combine the two studies in a way that allows either of the laboratories to be correct, but not necessarily both. We therefore also define an upper Markov combination which extends the families to incorporate both of the original ones.

The lower Markov combination combines only pairs of consistent distributions (f, g), $f \in \mathcal{F}$, $g \in \mathcal{G}$:

DEFINITION 4.4. The lower Markov combination of $\mathcal{F}$ and $\mathcal{G}$ is defined as

$$\mathcal{F} \,\underline{\star}\, \mathcal{G} = \{f \star g : f \in \mathcal{F}, g \in \mathcal{G}, \},$$

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

Note that only consistent pairs of distributions are combined. In the case where $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent, this specializes to the meta-Markov combination $\mathcal{F} \star \mathcal{G}$ of [**2**] which again is a special instance of a conditional product (see [**3**]).

As for the Markov combination of distributions, the lower Markov combination of families is commutative but has only limited associativity so that for $A \cap C \subset B$ it holds that

$$\mathcal{F} \,\underline{\star}\, (\mathcal{G} \,\underline{\star}\, \mathcal{H}) = (\mathcal{F} \,\underline{\star}\, \mathcal{G}) \,\underline{\star}\, \mathcal{H} \neq (\mathcal{F} \,\underline{\star}\, \mathcal{H}) \,\underline{\star}\, \mathcal{G}. \tag{4.3}$$

Note also that if $\mathcal{F}$ and $\mathcal{G}$ have no consistent pairs, we have $\mathcal{F} \,\underline{\star}\, \mathcal{G} = \emptyset$.

Now we provide two examples with combination of families of multivariate normal distributions.

EXAMPLE 4.1. [Combining meta-consistent families] Consider two families of bivariate normal distributions defined as

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

Here, $\mathcal{F}^{\downarrow Y} = \{Y \sim N(0, \sigma_y^2)\}$ and $\mathcal{G}^{\downarrow Y} = \{Y \sim N(0, \lambda_y^2)\}$. 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 $\lambda_y^2 = \sigma_y^2$. In this case, the lower (and meta-) Markov combination is given by

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

obtained by combining all $f \in \mathcal{F}$ and $g \in \mathcal{G}$ which are pairwise consistent, i.e. with $\sigma_y^2 = \lambda_y^2$. It holds that $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow(X,Y)} = \mathcal{F}$ and $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow(Y,Z)} = \mathcal{G}$ and all distributions in the resulting family are products of univariate Gaussian distributions. □

In the next example, the families are not meta-consistent so the lower Markov combination represents a proper restriction of the families.

Example 4.2. [Combining non-meta-consistent families] Consider two families of distributions defined as follows: $\mathcal{F}$ as in the above example and

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

Here, $\mathcal{F}^{\downarrow Y} = \{Y \sim N(0, \sigma_y^2)\}$ and $\mathcal{G}^{\downarrow Y} = \{Y \sim N(0, 1)\}$. Therefore the families are not meta-consistent. The lower Markov combination is given by

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

Note that here $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow(X,Y)} \subset \mathcal{F}$ and $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow(Y,Z)} = \mathcal{G}$ so the marginal families are genuinely reduced. □

The upper Markov combination, as defined below, is freely combining marginals of f from $\mathcal{F}$ with conditionals $g_{B|A\cap B}$ for $g \in \mathcal{G}$ and *vice versa*, corresponding to the set of left and right compositions of elements of $\mathcal{F}$ with elements of $\mathcal{G}$.

Definition 4.5. The upper Markov combination of $\mathcal{F}$ and $\mathcal{G}$ is

$$\mathcal{F} \overline{\star} \mathcal{G} = \{f \triangleright g \mid f \in \mathcal{F}, g \in \mathcal{G}, f << g\} \cup \{f \triangleleft g \mid f \in \mathcal{F}, g \in \mathcal{G}, g << f\}.$$

The lower Markov combination defines a smaller family than the upper Markov combination since it associates consistent pairs only, whereas the upper Markov combination associates more pairs.

Proposition 4.4. *It holds that* $\mathcal{F} \underline{\star} \mathcal{G} \subseteq \mathcal{F} \overline{\star} \mathcal{G}$.

Proof. Immediate from the definition. □

If the families $\mathcal{F}$ and $\mathcal{G}$ are not quasi-consistent, their upper Markov combination contains the upper Markov combination of $\mathcal{F}^{\mathcal{G}}$ and $\mathcal{G}^{\mathcal{F}}$ where these are defined in (3.1). Note that $\mathcal{F}^{\mathcal{G}} \overline{\star} \mathcal{G}^{\mathcal{F}} \subseteq \mathcal{F} \overline{\star} \mathcal{G}$, but the inclusion is strict in general.

All of the combinations imply conditional independence of the marginals given the intersection, thus deserving to be called Markov combinations.

Proposition 4.5. *For all distributions h in the upper Markov combination $\mathcal{F} \overline{\star} \mathcal{G}$, and hence also in the meta or lower Markov combination, it holds that $A \perp\!\!\!\perp B \mid A \cap B$.*

Proof. This follows directly from Proposition 4.3. □

Now we continue the two examples on combination of families of multivariate normal distributions.

Example 4.1 (continued). [Upper Markov combination for meta-consistent families] The families are meta-consistent and hence they define equivalent measures, that is to say $f << g$ and $g << f$ for each pair (f, g) with $f \in \mathcal{F}$ and $g \in \mathcal{G}$. Thus the upper Markov combination is given by

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

Indeed this is the same family as for the lower Markov combination so we have in this particular case

$$\mathcal{F} \,\underline{\star}\, \mathcal{G} = \mathcal{F} \star \mathcal{G} = \mathcal{F} \,\overline{\star}\, \mathcal{G}.$$

Proposition 4.9 below yields conditions under which this identity holds. □

Example 4.2 (continued). [Upper Markov combination for non-meta-consistent families] The families are quasi-consistent and the upper Markov combination is

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

□

As for the lower Markov combination, the upper Markov combination is commutative but not associative unless we impose additional strong conditions such as $A \cap B = A \cap C = B \cap C = A \cap B \cap C$, cf. the examples below.

Example 4.3. Consider $A = \{X, W\}$, $B = \{Y, W\}$, $C = \{Z, W\}$ and let all families consist of single elements: $\mathcal{F} = \{f\}$, $\mathcal{G} = \{g\}$, $\mathcal{H} = \{h\}$ defined over A, B, C, and assume, for simplicity, that all state spaces are discrete. A typical element of $(\mathcal{F} \,\overline{\star}\, \mathcal{G}) \,\overline{\star}\, \mathcal{H}$ is

$$(f \triangleright g) \triangleleft h = \frac{f(x, w) \cdot g(y, w) \cdot h(z, w)}{g(w) \cdot \tilde{f}(w)},$$

where $\tilde{f}(w)$ is the marginal given as

$$\tilde{f}(w) = \sum_{x,y} \frac{f(x, w) \cdot g(y, w)}{g(w)} = f(w),$$

so in this case $(f \triangleright g) \triangleleft h = f \triangleleft (g \triangleleft h)$, and the latter element is also in $\mathcal{F} \,\overline{\star}\, (\mathcal{G} \,\overline{\star}\, \mathcal{H})$. Similarly with other combinations. □

If we only have $A \cap C \subseteq B$, the associativity fails, as shown in the next example.

EXAMPLE 4.4. Consider $A = \{X, Y\}$, $B = \{Y, Z\}$, $C = \{Z, W\}$ and let all families consist of single elements: $\mathcal{F} = \{f\}$, $\mathcal{G} = \{g\}$, $\mathcal{H} = \{h\}$ defined over A, B, C and assume that all state spaces are discrete. One element of $(\mathcal{F} \overline{\star} \mathcal{G}) \overline{\star} \mathcal{H}$ is

$$(f \triangleright g) \triangleleft h = \frac{f(x, y) \cdot g(y, z) \cdot h(z, w)}{g(y) \cdot \tilde{g}(z)},$$

and now $\tilde{g}(z)$ is

$$\tilde{g}(z) = \sum_{x,y} \frac{f(x, y) \cdot g(y, z)}{g(y)} = \sum_{y} \frac{f(y) \cdot g(y, z)}{g(y)},$$

where this marginal does not simplify further so $(f \triangleright g) \triangleleft h$ is not contained in $\mathcal{F} \overline{\star} (\mathcal{G} \overline{\star} \mathcal{H})$. □

In general, the lower Markov combination reduces or preserves the marginal families, whereas the upper Markov combination preserves or extends them:

PROPOSITION 4.6. *$(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow A} \subseteq \mathcal{F}$ and $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow B} \subseteq \mathcal{G}$. Further, if $A = B$ then $\mathcal{F} \underline{\star} \mathcal{G} = \mathcal{F} \cap \mathcal{G}$.*

PROOF. This follows directly from Proposition 4.1 and Proposition 4.3. □

PROPOSITION 4.7. *$(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow A} \supseteq \mathcal{F}^{\mathcal{G}}$ and $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow B} \supseteq \mathcal{G}^{\mathcal{F}}$. Moreover, if $A = B$ then $\mathcal{F} \overline{\star} \mathcal{G} = \mathcal{F}^{\mathcal{G}} \cup \mathcal{G}^{\mathcal{F}}$.*

PROOF. From the definition it follows that $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow A} = \mathcal{F}^{\mathcal{G}} \cup \{f \triangleleft g, f \in \mathcal{F}, g \in \mathcal{G}, g << f\}^{\downarrow A}$ and $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow B} = \mathcal{G}^{\mathcal{F}} \cup \{f \triangleright g, f \in \mathcal{F}, g \in \mathcal{G}, f << g\}^{\downarrow B}$. □

If applied to meta-consistent families, the lower Markov combination preserves the marginal families, as also noted by Dawid and Lauritzen dawid:lauritzen:93.

PROPOSITION 4.8. *If $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent, $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow A} = \mathcal{F}$ and $(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow B} = \mathcal{G}$, and $\mathcal{F} \underline{\star} \mathcal{G} = \mathcal{F} \star \mathcal{G}$.*

PROOF. The first part follows from Definition 4.4 and the construction of the Markov combination. The final statement follows from [**2**] who show that for a meta-consistent pair, the lower (meta-) Markov combination is the unique Markov combination which preserves the marginal families. □

For the upper Markov combination the result is the reverse of that in Proposition 4.8 and it will be shown in Proposition 4.9.

The converse to Proposition 4.8 does not hold in general, but as summarized in Proposition 4.9 below, it does so when marginalizations to the variables in common form cuts. We recall the definition of a cut from [**1**] and [**2**].

DEFINITION 4.6. $Y_{A \cap B}$ is a cut in $\mathcal{F}$ if $\mathcal{F} = \mathcal{F}^{\downarrow (A|A \cap B)} \times \mathcal{F}^{\downarrow A \cap B}$ i.e. if $A|A \cap B$ and $A \cap B$ are variation independent: $(A|A \cap B) \ddagger (A \cap B)[\mathcal{F}]$.

It is equivalent to say that any product $f^1_{A \cap B} \cdot f^2_{A|A \cap B}$, with $f^1_{A \cap B} \in \mathcal{F}^{\downarrow A \cap B}$ and $f^2_{A|A \cap B} \in \mathcal{F}^{\downarrow (A|A \cap B)}$ defines a distribution f in $\mathcal{F}$.

The following proposition summarizes the interplay between cuts and combinations for meta-consistent families of distributions.

PROPOSITION 4.9. *Let $\mathcal{F}$ and $\mathcal{G}$ be two families of distributions for random variables Y_A and Y_B. The following are equivalent:*

(i) $\mathcal{F}$ *and* $\mathcal{G}$ *are quasi-consistent and* $\mathcal{F} \underline{\star} \mathcal{G} = \mathcal{F} \overline{\star} \mathcal{G}$.
(ii) $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow A} = \mathcal{F}$ *and* $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow B} = \mathcal{G}$.
(iii) $\mathcal{F}$ *and* $\mathcal{G}$ *are meta-consistent and* $Y_{A \cap B}$ *is a cut for* $\mathcal{F}$ *and* $\mathcal{G}$.
(iv) $\mathcal{F}$ *and* $\mathcal{G}$ *are meta-consistent and* $Y_{A \cap B}$ *is a cut for* $\mathcal{F} \overline{\star} \mathcal{G}$.

PROOF. (i) $\Rightarrow$ (ii): From Propositions 4.6 and 4.7 it follows that

$$(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow A} \subseteq \mathcal{F}^{\mathcal{G}} \text{ and } (\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow A} \supseteq \mathcal{F}^{\mathcal{G}},$$

and hence, as the combinations are equal we must have

$$(\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow A} = (\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow A} = \mathcal{F}^{\mathcal{G}} = \mathcal{F},$$

where the last equality follows from quasi-consistency. The case for the B-marginal is analogous.

(ii) $\Rightarrow$ (iii): Recall that we have $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow A} = \mathcal{F}^{\mathcal{G}} \cup \{f \triangleleft g, f \in \mathcal{F}, g \in \mathcal{G}, g << f\}^{\downarrow A}$. If

$$\{f \triangleleft g, f \in \mathcal{F}, g \in \mathcal{G}, g << f\}^{\downarrow A} \subseteq \mathcal{F},$$

it must also hold that $\{f \triangleleft g, f \in \mathcal{F}, g \in \mathcal{G}, g << f\}^{\downarrow A} \subseteq \mathcal{F}^{\mathcal{G}}$ as $(f \triangleleft g)^{\downarrow A} << g$. Hence then also $\mathcal{F}^{\mathcal{G}} = \mathcal{F}$ whereby $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow A \cap B} = \mathcal{F}^{\downarrow A \cap B} = \mathcal{G}^{\downarrow A \cap B}$, i.e. the families are meta-consistent. An arbitrary element of $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow A}$ is, for example, of the form

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

Since the families are meta-consistent, $g_{A \cap B} = f^*_{A \cap B}$ for some $f^* \in \mathcal{F}$, and hence

$$(f \triangleleft g)^{\downarrow A} = f_{A|A \cap B} \cdot f^*_{A \cap B},$$

showing that $Y_{A \cap B}$ forms a cut in $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow A} = \mathcal{F}$ and similarly in $\mathcal{G}$.

(iii) $\Rightarrow$ (i): Since $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent they are also quasi-consistent. As $\mathcal{F} \underline{\star} \mathcal{G} \subseteq \mathcal{F} \overline{\star} \mathcal{G}$ follows from Proposition 4.4, we must show the reverse inclusion. Assume $Y_{A \cap B}$ is a cut for $\mathcal{F}$ and $\mathcal{G}$, and consider an arbitrary element of $\mathcal{F} \overline{\star} \mathcal{G}$ of the form $f \triangleleft g = f_{A|A \cap B} \cdot g$. We must show it is also in $\mathcal{F} \underline{\star} \mathcal{G}$, i.e. find a consistent pair which gives the same combination. Since $\mathcal{F}$ and $\mathcal{G}$ are assumed meta-consistent, we can find f^* with $f^*_{A \cap B} = g_{A \cap B}$. As $Y_{A \cap B}$ is a cut in $\mathcal{F}$ it holds that

$$\tilde{f} = f_{A|A \cap B} \cdot f^*_{A \cap B} \in \mathcal{F},$$

and since $\tilde{f}_{A|A \cap B} = f_{A|A \cap B}$ and $\tilde{f}_{A \cap B} = f^*_{A \cap B} = g_{A \cap B}$, $\tilde{f}$ and g are consistent and $\tilde{f} \star g = f \triangleleft g$, hence the latter is in $\mathcal{F} \underline{\star} \mathcal{G}$. The argument is analogous for an element of the form $f \triangleright g$.

(iii) $\Leftrightarrow$ (iv): Proposition 4.5 yields the conditional independence $A \perp\!\!\!\perp B \mid A \cap B$ which implies that $Y_{A \cap B}$ is a cut for $\mathcal{F}$ and $\mathcal{G}$ if and only if it is a cut for $\mathcal{F} \overline{\star} \mathcal{G}$. $\square$

COROLLARY 4.1. *For* $C \subseteq A$, Y_C *is a cut for* $\mathcal{F}$ *if and only if* $\mathcal{F} \overline{\star}\ \mathcal{F}^{\downarrow C} = \mathcal{F}$.

PROOF. This follows from Proposition 4.9 by letting $\mathcal{G} = \mathcal{F}^{\downarrow C}$. $\square$

In all of the Markov combinations it holds that the pair of marginal families in the combined families are necessarily meta-consistent, by construction. If we let

$$\mathcal{F}_\star = (\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow A}, \quad \mathcal{F}^\star = (\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow A}, \quad \mathcal{G}_\star = (\mathcal{F} \underline{\star} \mathcal{G})^{\downarrow B}, \quad \mathcal{G}^\star = (\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow B},$$

we have

Proposition 4.10. *The pairs of marginal families* $(\mathcal{F}_\star, \mathcal{G}_\star)$ *and* $(\mathcal{F}^\star, \mathcal{G}^\star)$ *are meta-consistent. Each of the combinations* $\mathcal{F} \underline{\star}\, \mathcal{G}$ *and* $\mathcal{F} \overline{\star}\, \mathcal{G}$ *is the meta-Markov combination of its constituent pair of marginals. Further, it holds that*

$$\mathcal{F} \underline{\star}\, \mathcal{G} = \mathcal{F}_\star \underline{\star}\, \mathcal{G}_\star = \mathcal{F}_\star \star \mathcal{G}_\star \subseteq \mathcal{F}_\star \overline{\star}\, \mathcal{G}_\star, \tag{4.4}$$

and

$$\mathcal{F} \overline{\star}\, \mathcal{G} = \mathcal{F}^\star \underline{\star}\, \mathcal{G}^\star = \mathcal{F}^\star \star \mathcal{G}^\star \subseteq \mathcal{F}^\star \overline{\star}\, \mathcal{G}^\star. \tag{4.5}$$

Proof. The pair of marginal families for any family of joint distributions is necessarily meta-consistent, which gives the first statement. This automatically implies the equalities in (4.4) and (4.5), and the last inequality in these follows from Proposition 4.4. □

This proposition highlights that the Markov combinations can be seen as first reducing or extending the given marginal families as appropriate and subsequently forming their meta-Markov combination. Note that the notation is slightly imprecise as $\mathcal{F}_\star$ and $\mathcal{F}^\star$ depend on $\mathcal{G}$ as well, so $\mathcal{F}_\star(\mathcal{G})$ etc. would be more appropriate. However, for simplicity of notation we shall leave this fact as implicit. This is clear also from the following, where we give an alternative expression for $\mathcal{F}_\star$ and $\mathcal{G}_\star$.

Proposition 4.11. $\mathcal{F}_\star = \{f \in \mathcal{F} \mid f_{A\cap B} = g_{A\cap B} \text{ for some } g \in \mathcal{G}\}$ *and* $\mathcal{G}_\star = \{g \in \mathcal{G} \mid f_{A\cap B} = g_{A\cap B} \text{ for some } f \in \mathcal{F}\}$.

Proof. Follows by direct calculation of $(\mathcal{F} \underline{\star}\, \mathcal{G})^{\downarrow A}$ and $(\mathcal{F} \underline{\star}\, \mathcal{G})^{\downarrow B}$. □

In general, the inclusions on the right-hand sides of (4.4) and (4.5) are strict. For example, if we combine an element of the form $(f \lhd g)^{\downarrow A} \in \mathcal{F}^\star$ with one of the form $(\tilde{f} \rhd \tilde{g})^{\downarrow B} \in \mathcal{G}^\star$, we get

$$(f \lhd g)^{\downarrow A} = \left(\frac{f \cdot g}{f_{A\cap B}}\right)^{\downarrow A} = \frac{f \cdot g_{A\cap B}}{f_{A\cap B}}, \quad (\tilde{f} \rhd \tilde{g})^{\downarrow B} = \left(\frac{\tilde{f} \cdot \tilde{g}}{\tilde{g}_{A\cap B}}\right)^{\downarrow B} = \frac{\tilde{f}_{A\cap B} \cdot \tilde{g}}{\tilde{g}_{A\cap B}},$$

and further

$$(f \lhd g)^{\downarrow A\cap B} = g_{A\cap B}, \quad (\tilde{f} \rhd \tilde{g})^{\downarrow A\cap B} = \tilde{f}_{A\cap B},$$

so that, for example,

$$(f \lhd g)^{\downarrow A} \lhd (\tilde{f} \rhd \tilde{g})^{\downarrow B} = \frac{\frac{f \cdot g_{A\cap B}}{f_{A\cap B}} \cdot \frac{\tilde{f}_{A\cap B} \cdot \tilde{g}}{\tilde{g}_{A\cap B}}}{g_{A\cap B}} = \frac{f \cdot \frac{\tilde{f}_{A\cap B}}{f_{A\cap B}} \cdot \tilde{g}}{\tilde{g}_{A\cap B}}, \tag{4.6}$$

which is not necessarily an element of the upper Markov combination. If we let

$$\mathcal{F}_{\star\star} = (\mathcal{F}_\star \underline{\star}\, \mathcal{G}_\star)^{\downarrow A}, \quad \mathcal{G}_{\star\star} = (\mathcal{F}_\star \underline{\star}\, \mathcal{G}_\star)^{\downarrow B}, \quad \mathcal{F}^{\star\star} = (\mathcal{F}^\star \overline{\star}\, \mathcal{G}^\star)^{\downarrow A}, \quad \mathcal{G}^{\star\star} = (\mathcal{F}^\star \overline{\star}\, \mathcal{G}^\star)^{\downarrow B},$$

and so on, we have

Proposition 4.12. *The iterated extensions* $\mathcal{F}^{\star\star}$ *and* $\mathcal{G}^{\star\star}$ *are given as*

$$\begin{aligned} \mathcal{F}^{\star\star} &= \{f_1 \rhd f_2 : f_1 \in \mathcal{F}^{\downarrow A\cap B}, f_2 \in \mathcal{F}\} \cup \{f \lhd g : f \in \mathcal{F}, g \in \mathcal{G}\}^{\downarrow A}, \\ &= \left\{f \cdot \frac{h_{A\cap B}}{f_{A\cap B}}, f \in \mathcal{F}, h \in \mathcal{F} \cup \mathcal{G}\right\}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{G}^{\star\star} &= \{g_1 \rhd g_2 : g_1 \in \mathcal{G}^{\downarrow A\cap B}, g_2 \in \mathcal{G}\} \cup \{f \rhd g : f \in \mathcal{F}, g \in \mathcal{G}\}^{\downarrow B}, \\ &= \left\{\frac{h_{A\cap B}}{g_{A\cap B}} \cdot g, h \in \mathcal{F} \cup \mathcal{G}, g \in \mathcal{G}\right\}. \end{aligned}$$

Proof. Follows by direct calculation of $(\mathcal{F}^\star \,\overline{\star}\, \mathcal{G}^\star)^{\downarrow A}$ and $(\mathcal{F}^\star \,\overline{\star}\, \mathcal{G}^\star)^{\downarrow B}$. $\square$

In words, the iterated extension $\mathcal{F}^{\star\star}$ is obtained by combining all $A \cap B$ marginals from $\mathcal{F}$ and $\mathcal{G}$ with all conditionals of $B \setminus A$ given $A \cap B$ from $\mathcal{F}$, thus constructing a cut to $A \cap B$. $\mathcal{G}^{\star\star}$ is derived in an equivalent manner.

If we modify the families repeatedly by combining them, the families remain unchanged after a finite number of steps.

Proposition 4.13. $\mathcal{F}_{\star\star} = \mathcal{F}_\star, \quad \mathcal{G}_{\star\star} = \mathcal{G}_\star, \quad \mathcal{F}^{\star\star\star} = \mathcal{F}^{\star\star}, \quad \mathcal{G}^{\star\star\star} = \mathcal{G}^{\star\star}.$

Proof. The first two identities follow directly from Proposition 4.10. Proposition 4.12 shows that $\mathcal{F}^{\star\star}$ and $\mathcal{G}^{\star\star}$ are meta-consistent and also that $Y_{A\cap B}$ is a cut for $\mathcal{F}^{\star\star}$ and $\mathcal{G}^{\star\star}$. Therefore $\mathcal{F}^{\star\star} \,\overline{\star}\, \mathcal{G}^{\star\star} = \mathcal{F}^{\star\star} \star \mathcal{G}^{\star\star}$ and $\mathcal{F}^{\star\star\star} = (\mathcal{F}^{\star\star} \star \mathcal{G}^{\star\star})^{\downarrow A} = \mathcal{F}^{\star\star}$ because the meta-Markov combination preserves the marginal families (see Proposition 4.8). The last identity just involves interchanging $\mathcal{F}$ with $\mathcal{G}$. $\square$

Note that, once we have reduced the families with the lower Markov combination to obtain $\mathcal{F}_\star$ and $\mathcal{G}_\star$, they will not be reduced any further by repeated combination, whereas with the upper Markov combination we may extend them one more step by forming their upper combination. Combining this with Proposition 4.10 and (ii) $\Rightarrow$ (i) of Proposition 4.9 yields

$$\mathcal{F} \,\overline{\star}\, \mathcal{G} = \mathcal{F}^\star \,\underline{\star}\, \mathcal{G}^\star = \mathcal{F}^\star \star \mathcal{G}^\star \subseteq \mathcal{F}^\star \,\overline{\star}\, \mathcal{G}^\star = \mathcal{F}^{\star\star} \,\underline{\star}\, \mathcal{G}^{\star\star} = \mathcal{F}^{\star\star} \star \mathcal{G}^{\star\star} = \mathcal{F}^{\star\star} \,\overline{\star}\, \mathcal{G}^{\star\star}, \tag{4.7}$$

i.e. when the families have been sufficiently modified, they combine in the obvious way. Note in particular that

$$\mathcal{F} \,\underline{\star}\, \mathcal{G} \subseteq \mathcal{F} \,\overline{\star}\, \mathcal{G} \subseteq \mathcal{F}^{\star\star} \star \mathcal{G}^{\star\star}.$$

This motivates the introduction of a super Markov combination as the meta-Markov combination of the maximally extended families:

Definition 4.7. The super Markov combination $\mathcal{F} \otimes \mathcal{G}$ of $\mathcal{F}$ and $\mathcal{G}$ is given as

$$\mathcal{F} \otimes \mathcal{G} = \mathcal{F}^{\star\star} \star \mathcal{G}^{\star\star}.$$

Note that $Y_{A\cap B}$ is a cut in $\mathcal{F} \otimes \mathcal{G}$ as well as in the families $\mathcal{F}^{\star\star}$ and $\mathcal{G}^{\star\star}$. This implies in particular that the super Markov combination is a strong meta-Markov combination of its constituent marginal families $\mathcal{F}^{\star\star}$ and $\mathcal{G}^{\star\star}$, as is also the case for the lower Markov combination when the conditions of Proposition 4.9 are fulfilled as then $\mathcal{F} = \mathcal{F}^{\star\star}$. We also have, from (4.7), that

$$\mathcal{F} \otimes \mathcal{G} = \mathcal{F}^\star \,\overline{\star}\, \mathcal{G}^\star,$$

so the super Markov combination can also be obtained by double upper combination of the families. The super Markov combination has no associativity properties.

Proposition 4.14. *The super Markov combination $\mathcal{F} \otimes \mathcal{G}$ can also be written as*

$$\mathcal{F} \otimes \mathcal{G} = \left\{ f_{A|A\cap B} \cdot h_{A\cap B} \cdot g_{B|A\cap B},\ f \in \mathcal{F}, h \in \mathcal{F} \cup \mathcal{G}, g \in \mathcal{G} \right\}.$$

Proof. This follows from Definition 4.7. $\square$

In words, the super Markov combination is the combination of all conditional distributions from the two families with any marginal distributions.

EXAMPLE 4.5. [General example with all combinations and properties of cuts] Let $\mathcal{F} = \{f^1, f^2\}$ and $\mathcal{G} = \{g^1, g^2\}$. If $f^1_{A\cap B} = g^1_{A\cap B}$, and $f^2_{A\cap B} = g^2_{A\cap B}$, the families are meta-consistent. The lower Markov combination is then equal to the meta-Markov combination

$$\mathcal{F} \underline{\star} \mathcal{G} = \mathcal{F} \star \mathcal{G} = \left\{ f^1 \star g^1, f^2 \star g^2 \right\}.$$

The upper Markov combination in the general case is equal to

$$\mathcal{F} \overline{\star} \mathcal{G} = \left\{ f^1 \triangleleft g^1, f^1 \triangleright g^1, f^1 \triangleleft g^2, f^1 \triangleright g^2, f^2 \triangleleft g^1, f^2 \triangleright g^1, f^2 \triangleleft g^2, f^2 \triangleright g^2 \right\},$$

yielding

$$\mathcal{F}^\star = \mathcal{F} \cup \left\{ \frac{f^1 \cdot g^1_{A\cap B}}{f^1_{A\cap B}}, \frac{f^1 \cdot g^2_{A\cap B}}{f^1_{A\cap B}}, \frac{f^2 \cdot g^1_{A\cap B}}{f^2_{A\cap B}}, \frac{f^2 \cdot g^2_{A\cap B}}{f^2_{A\cap B}} \right\},$$

$$\mathcal{G}^\star = \mathcal{G} \cup \left\{ \frac{f^1_{A\cap B} \cdot g^1}{g^1_{A\cap B}}, \frac{f^1_{A\cap B} \cdot g^2}{g^2_{A\cap B}}, \frac{f^2_{A\cap B} \cdot g^1}{g^1_{A\cap B}}, \frac{f^2_{A\cap B} \cdot g^2}{g^2_{A\cap B}} \right\},$$

whereas $\mathcal{F}^{\star\star}$ and $\mathcal{G}^{\star\star}$ have 8 elements each because they combine all 4 marginals to $A \cap B$ with 2 conditionals from $\mathcal{F}$ or 2 conditionals from $\mathcal{G}$, respectively, see Proposition 4.12. The super Markov combination $\mathcal{F} \otimes \mathcal{G}$ has $2 \times 4 \times 2 = 16$ elements, reflecting the fact that it combines all 4 marginals to $A \cap B$ with both conditionals from $\mathcal{F}$ and both conditionals from $\mathcal{G}$ (see Proposition 4.14), i.e.

$$\mathcal{F} \otimes \mathcal{G} \quad = \quad \left\{ \frac{f^1 \cdot g^1 \cdot g^2_{A\cap B}}{f^1_{A\cap B} \cdot g^1_{A\cap B}}, \frac{f^1 \cdot g^1 \cdot f^2_{A\cap B}}{f^1_{A\cap B} \cdot g^1_{A\cap B}}, \frac{f^1 \cdot g^2 \cdot g^1_{A\cap B}}{f^1_{A\cap B} \cdot g^2_{A\cap B}}, \dots \right\}.$$

If $Y_{A\cap B}$ is a cut for $\mathcal{F}$, it holds that $f^1_{A|A\cap B} \cdot f^2_{A\cap B} = f^i$ for some $i \in \{1,2\}$, $f^2_{A|A\cap B} \cdot f^1_{A\cap B} = f^i$ for some $i \in \{1,2\}$. Therefore if $f^1_{A|A\cap B} \cdot f^2_{A\cap B} = f^1$, then we must have $f^1_{A\cap B} = f^2_{A\cap B}$, otherwise $f^1_{A|A\cap B} = f^2_{A|A\cap B}$. Therefore there are two conditions for $Y_{A\cap B}$ to be a cut for $\mathcal{F}$: $f^1_{A\cap B} = f^2_{A\cap B}$ or $f^1_{A|A\cap B} = f^2_{A|A\cap B}$. If $Y_{A\cap B}$ is a cut for $\mathcal{G}$, there are two other conditions, i.e. $g^1_{A\cap B} = g^2_{A\cap B}$ or $g^1_{B|A\cap B} = g^2_{B|A\cap B}$.

Therefore, let us suppose that $Y_{A\cap B}$ is a cut for both $\mathcal{F}$ and $\mathcal{G}$. There are four possible cases to which the meta-consistency conditions ($f^1_{A\cap B} = g^1_{A\cap B}$, $f^2_{A\cap B} = g^2_{A\cap B}$) are added :

(1) $f^1_{A\cap B} = f^2_{A\cap B}$ and $g^1_{A\cap B} = g^2_{A\cap B}$,
(2) $f^1_{A\cap B} = f^2_{A\cap B}$ and $g^1_{B|A\cap B} = g^2_{B|A\cap B}$,
(3) $f^1_{A|A\cap B} = f^2_{A|A\cap B}$ and $g^1_{A\cap B} = g^2_{A\cap B}$,
(4) $f^1_{A|A\cap B} = f^2_{A|A\cap B}$ and $g^1_{B|A\cap B} = g^2_{B|A\cap B}$.

Under meta-consistency and the hypotheses of each case we get

$$\mathcal{F} \underline{\star} \mathcal{G} = \mathcal{F} \star \mathcal{G} = \mathcal{F} \overline{\star} \mathcal{G} = \mathcal{F} \otimes \mathcal{G};$$

see also Proposition 4.9. □

5. Combining Gaussian graphical models

We first recall the definitions of dependence graph, marginal graph and graphical collapsibility.

DEFINITION 5.1. The dependence graph $G(f)$ of a distribution f is defined by

$$\alpha \not\sim \beta \iff \alpha \perp\!\!\!\perp_f \beta \,|\, V \setminus \{\alpha, \beta\}.$$

The dependence graph $G(\mathcal{F})$ of a family $\mathcal{F}$ is the union of the graphs $G(f)$ for $f \in \mathcal{F}$.

Thus $G(\mathcal{F})$ is the smallest graph G such that all $f \in \mathcal{F}$ are pairwise Markov with respect to G.

Definition 5.2. For $C \subseteq V$ the marginal graph $G_{[C]}$ is defined by $\alpha \sim \beta$ unless $C \setminus \{\alpha, \beta\}$ separates α and β in the original graph.

The marginal graph $G(\mathcal{F})_{[C]}$ is the dependence graph of the marginal family $\mathcal{F}^{\downarrow C}$. It is in general different from the induced subgraph $G(\mathcal{F})_C$.

Definition 5.3. An undirected graph G is collapsible onto A if the boundary of every connected component of $V \setminus A$ is complete.

Recall that G is collapsible onto A if and only if $G_{[A]} = G_A$ which again is true if and only if the marginal $\mathcal{F}^{\downarrow A}$ of the graphical Gaussian model $\mathcal{F}$ with graph G is exactly equal to the graphical Gaussian model determined by G_A (see [**7**]).

We now consider two families of multivariate Gaussian distributions $\mathcal{F}$ and $\mathcal{G}$ determined as the graphical Gaussian models with graphs (A, E_A) and (B, E_B) so $G(\mathcal{F}) = (A, E_A)$ and $G(\mathcal{G}) = (B, E_B)$. We now wish to consider combinations of these and would ideally search for an undirected graph $G = (V, E)$ with vertex set $V = A \cup B$ so that the combined family is the graphical Gaussian model with graph G.

If two Gaussian graphical models are meta-consistent, the marginal graphs over the variables in the intersection must be identical. However, the converse is not generally true as there may be other constraints, such as vanishing tetrad differences [**16, 17**]; see also the example below, taken from [**18**, pp. 149–150].

Example 5.1. [Identical marginal graphs but not meta-consistent]
Consider the two Gaussian graphical models as displayed in Figure 5.1. The marginal graphs over the vertices $\{3, 4, 5, 6\}$ are the same but the families of distributions induced over the corresponding variables are different.
The graph on the left-hand side implies the constraint (among others)

$$\sigma_{35}\sigma_{46} = \sigma_{36}\sigma_{45}, \tag{5.1}$$

whereas the graph on the right-hand side implies in addition

$$\sigma_{34}\sigma_{56} = \sigma_{36}\sigma_{45} \text{ and } \sigma_{35}\sigma_{46} = \sigma_{34}\sigma_{56}. \tag{5.2}$$

Since the structure of the concentration matrix of the graph on the right-hand side gives

$$\sigma_{34}\sigma_{77} = \sigma_{47}\sigma_{37}, \quad \sigma_{35}\sigma_{77} = \sigma_{57}\sigma_{37}, \quad \sigma_{45}\sigma_{77} = \sigma_{57}\sigma_{47},$$

there are further constraints induced because, for example, if $\sigma_{34} > 0$ and $\sigma_{35} > 0$ then also $\sigma_{45} > 0$. Moreover we have also

$$\frac{\sigma_{47}\sigma_{37}}{\sigma_{34}} = \frac{\sigma_{57}\sigma_{37}}{\sigma_{35}} = \frac{\sigma_{57}\sigma_{47}}{\sigma_{45}}.$$

As the families $\mathcal{F}$ and $\mathcal{G}$ do not induce the same constraints on the common variables they are not meta-consistent. □

If the two graphs are isomorphic then the meta-consistency of the families is guaranteed, as in example 5.2 below. Otherwise, all the constraints on the common variables must be investigated.

Example 5.2. [Meta-consistency] The graphical Gaussian models displayed in Figure 5.2 are meta-consistent because the two graphs are isomorphic and therefore induce the same restrictions on the common variables. □

FIGURE 5.1. Two graphical Gaussian models with the same marginal graphs over vertices $\{3, 4, 5, 6\}$ which are not meta-consistent.

FIGURE 5.2. Two graphical Gaussian models which are meta-consistent.

FIGURE 5.3. Two graphical Gaussian models whose graphs are collapsible onto $\{1, 2, 3\}$ but the families are not meta-consistent.

Meta-consistency is related to but different from collapsibility of the two graphs onto $A \cap B$. In example 5.2, the two Gaussian graphical models are meta-consistent, however, none of the graphs are collapsible onto $\{1, 2, 3, 4\}$. This is also true for the graphs displayed in Figure 2.2, although there in a trivial way as the common marginals are unrestricted. In Example 5.1, the families are not meta-consistent and the graphs are not collapsible onto $\{3, 4, 5, 6\}$. In example 5.3, the graphs are collapsible onto $\{1, 2, 3\}$ but the families are not meta-consistent.

EXAMPLE 5.3. [Collapsibility without meta-consistency] In Figure 5.3, both graphs are collapsible onto $\{1, 2, 3\}$ but the Gaussian graphical models are not meta-consistent because the two marginal graphs over $\{1, 2, 3\}$ are not identical. □

If the graphs are collapsible onto $A \cap B$ and the induced subgraphs on the common variables are the same, then the families are meta-consistent and the lower, meta-, upper and super Markov combination are identical.

PROPOSITION 5.1. *If two graphical Gaussian models $\mathcal{F}$ and $\mathcal{G}$ have both graphs collapsible onto $A \cap B$ and $G(\mathcal{F})_{A\cap B} = G(\mathcal{G})_{A\cap B}$, then*

$$\mathcal{F} \underline{\star} \mathcal{G} = \mathcal{F} \star \mathcal{G} = \mathcal{F} \overline{\star} \mathcal{G} = \mathcal{F} \otimes \mathcal{G}.$$

PROOF. It follows from the fact that the graphical collapsibility ensures that $Y_{A\cap B}$ is a cut in the original families [**7**] and then by the results of Proposition 4.9. □

Now we describe in detail the combinations already studied in Section 4.2 when the two initial families are two Gaussian graphical models. The lower Markov combination is given by the family of distributions that satisfy all induced constraints.

These can be polynomial equality relations, i.e. conditional independence, tetrad and pentad constraints, and also inequality constraints, see [**5**]. In general, the equality constraints can all be identified by algebraic methods, see [**20**]. For example, the tetrad representation theorem in the version of [**20**] permits to compute all the tetrad constraints for a directed acyclic graph. In contrast, no method for computation of all inequality constraints is known to us. In example 5.1, the equality constraints are the conditional independence constraints of each graph converted into vanishing minors (see Lemma 2.1 of [**4**]) and the tetrad constraints (shown in 5.1 and 5.2). In Example 5.2, the families are meta-consistent, the lower Markov combination is equal to the meta-Markov combination which is seen by considering all induced constraints.

The upper Markov combination combines all the marginal distributions from one family with all the conditional distributions from the other family.

The super Markov combination is the combination of all conditional distributions from the two families with any marginal distributions on the common variables. In Example 5.2, the sets of marginal distributions of $Y_{A\cap B}$ from the two families are identical, $Y_{A\cap B}$ becomes a cut in each extended marginal family $\mathcal{F}^\star$ and $\mathcal{G}^\star$. Hence, from Propositions 4.9 and 4.10 we have

$$\mathcal{F} \,\overline{\star}\, \mathcal{G} = \mathcal{F}^\star \,\underline{\star}\, \mathcal{G}^\star = \mathcal{F}^\star \star \mathcal{G}^\star = \mathcal{F}^\star \,\overline{\star}\, \mathcal{G}^\star = \mathcal{F} \otimes \mathcal{G},$$

and also $\mathcal{F}^\star = \mathcal{F}^{\star\star}$ and $\mathcal{G}^\star = \mathcal{G}^{\star\star}$.

The chosen combination determines the joint family of distributions. The next step is to find the dependence graph $G(\mathcal{F} * \mathcal{G})$ of the combined family $\mathcal{F} * \mathcal{G}$, where the symbol $*$ is one of $\{\star, \underline{\star}, \overline{\star}, \otimes\}$, and its relation with the two initial graphs. As we saw in Definition 5.1, the dependence graph $G(\mathcal{F} * \mathcal{G})$ is always defined but the Gaussian graphical model with graph $G(\mathcal{F} * \mathcal{G})$ may be different from $\mathcal{F} * \mathcal{G}$. Moreover, $G(\mathcal{F} * \mathcal{G})$ may not be the graph union of the two original graphs. We therefore introduce the concept of a graphical combination as a combination in which $\mathcal{F} * \mathcal{G}$ is the Gaussian graphical model with graph $G(\mathcal{F} * \mathcal{G})$. Thus, a graphical combination takes into account only conditional independence constraints. So for example, a graphical lower Markov combination $\mathcal{F} \,\underline{\star}\, \mathcal{G}$ is the Gaussian graphical model with graph $G(\mathcal{F} \,\underline{\star}\, \mathcal{G})$.

If the graphs of the original families are collapsible onto $A \cap B$ and the families are meta-consistent, as in Proposition 5.1, the combination of the families is graphical and

$$G(\mathcal{F} * \mathcal{G}) = G(\mathcal{F}) \cup G(\mathcal{G}),$$

for all combinations.

Another type of situation occurs when the joint family of distributions is not a graphical model, so that $\mathcal{F} * \mathcal{G}$ has dependence graph $G(\mathcal{F} * \mathcal{G})$, but it is not a Gaussian graphical model with graph $G(\mathcal{F} * \mathcal{G})$. This combination is said to be a non-graphical combination. In such cases, the dependence graph of the combination does not fully describe the features of the combined family. Nevertheless, a graphical combination can be identified by the Gaussian family corresponding to $G(\mathcal{F} * \mathcal{G})$.

In Example 5.1, the upper Markov combination and the super Markov combination are both non-graphical combinations. A graphical combination is the Gaussian graphical model with graph $G(\mathcal{F} \,\underline{\star}\, \mathcal{G})$ which is the graph obtained by merging the two graphs after having added all the edges that render the induced subgraph over $\{3, 4, 5, 6\}$ a complete graph. In Example 5.2, the upper Markov combination (equal

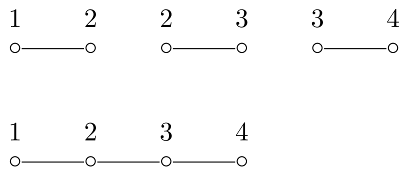

FIGURE 5.4. The three upper graphs correspond to families $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$, respectively. The lower graph (a line) represents any joint combination of the three families.

to the super Markov combination) is a non-graphical combination. We conclude the section providing further examples of such combinations.

EXAMPLE 5.4. [Graphical combinations] We represent the two graphical models of Figure 2.1 as the families
$\mathcal{F} = \left\{Y_A \sim N_3(0,\Sigma), \Sigma^{-1} \in S^+(G_A)\right\}$ and $\mathcal{G} = \left\{Y_B \sim N_2(0,\Phi), \Phi^{-1} \in S^+(G_B)\right\}$.
Here, $\mathcal{F}^{\downarrow\{2,3\}} = \{Y \sim N_2(0,\Sigma_{\{2,3\}}), \Sigma^{-1}_{\{2,3\}} \in S^+(G_B)\}$, and $\mathcal{G}^{\downarrow\{2,3\}} = \mathcal{G}$. The families $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent because $\mathcal{F}^{\downarrow\{2,3\}} = \mathcal{G}^{\downarrow\{2,3\}}$. The lower Markov combination is the family

$$\mathcal{F} \underline{\star} \mathcal{G} = \left\{Y \sim N_3(0,\Sigma), \Sigma^{-1} \in S^+(G_A), \Sigma_{\{2,3\}} = \Phi_{\{2,3\}}\right\},$$

and it is represented by the graph on the left.
The upper Markov combination (equal to the super Markov combination) is a graphical combination and its dependence graph is the complete graph. □

EXAMPLE 5.5. [Combinations and cuts] We represent the two graphical models of Figure 2.3 as $\mathcal{F} = \{Y_A \sim N_3(0,\Sigma)\}$ and $\mathcal{G} = \{Y_B \sim N_3(0,\Phi)\}$. We have that $\mathcal{F}^{\downarrow\{2,3\}} = \left\{Y \sim N_2(0,\Sigma_{\{2,3\}})\right\}$, and $\mathcal{G}^{\downarrow\{2,3\}} = \left\{Y \sim N_2(0,\Phi_{\{2,3\}})\right\}$. The families $\mathcal{F}$ and $\mathcal{G}$ are meta-consistent because $\mathcal{F}^{\downarrow\{2,3\}} = \mathcal{G}^{\downarrow\{2,3\}}$. The lower Markov combination is the graphical model obtained as the union of the two. It is equivalent to the meta-, upper and super Markov combinations because the two graphs are both collapsible onto $\{2,3\}$ and the families are meta-consistent. □

EXAMPLE 5.6. [Associativity and commutativity] Suppose $\mathcal{F} = \{Y_A \sim N_2(0,\Gamma)\}$, $\mathcal{G} = \{Y_B \sim N_2(0,\Omega)\}$, and $\mathcal{H} = \{Y_C \sim N_2(0,\Phi)\}$ represent three complete Gaussian graphical models as in Figure 5.4. Here all the combinations satisfy $(\mathcal{F}*\mathcal{G})*\mathcal{H} = \mathcal{F}*(\mathcal{G}*\mathcal{H}) = \mathcal{F}*(\mathcal{H}*\mathcal{G})$, and are the Gaussian families corresponding to the union of the three graphs (a line). It follows from the associativity property of the meta-Markov combination and the fact that the families are pairwise meta-consistent and form a cut over the common variables so that all the combinations are identical (see Proposition 5.1). If we change the order of combination, $(\mathcal{F} \underline{\star} \mathcal{H}) \underline{\star} \mathcal{G}$ represents the independence model $(1,2) \perp\!\!\!\perp (3,4)$ and $(\mathcal{F} \overline{\star} \mathcal{H}) \overline{\star} \mathcal{G}$ is again the lower graph in Figure 5.4. Suppose now to add another complete graph with variables 4 and 1. The lower Markov combination of all four graphs is the graphical model with $1 \perp\!\!\!\perp 4 \mid (2,3)$. The upper Markov combination is the cordless four cycle because it combines all marginal distributions over $\{1,4\}$ with all conditional distributions over $\{1,2,3,4\}$ and all marginal distributions over $\{1,2,3,4\}$ with all conditional distributions over $\{1,4\}$. □

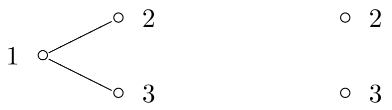

Figure 5.5. From left to right, graphs G_A and G_B.

Example 5.7. [Non-graphical and graphical combinations] We represent the two graphical models of Figure 5.5 as the families $\mathcal{F} = \{Y_A \sim N_3(0, \Sigma), \Sigma^{-1} \in S^+(G_A)\}$ and $\mathcal{G} = \{Y_B \sim N_2(0, \Phi), \Phi^{-1} \in S^+(G_B)\}$.
Moreover $\mathcal{F}^{\downarrow\{2,3\}} = \Big\{Y \sim N_2(0, \Sigma_{\{2,3\}}), \Sigma^{-1}_{\{2,3\}} \in S^+(G)\Big\}$, and G is the graph with vertex set $\{2, 3\}$ and edge $(2, 3)$. Here, $\mathcal{G}^{\downarrow\{2,3\}} = \mathcal{G}$ and the families are not meta-consistent because $\mathcal{G}^{\downarrow\{2,3\}} \subset \mathcal{F}^{\downarrow\{2,3\}}$. The lower Markov combination combines only consistent pairs but it is not a graphical combination. It is given by

$$\mathcal{F} \underline{\star}\, \mathcal{G} = \{Y \sim N_3(0, \Omega),\, \omega_{23}\omega_{11} = \omega_{12}\omega_{13},\, \omega_{23} = 0\}\,,$$

where $\Omega = \{\omega_{ij}\}$. 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 upper Markov combination is

$$\mathcal{F} \,\overline{\star}\, \mathcal{G} = \{Y \sim N_3(0, \Omega),\, \omega_{23} = 0\}\,,$$

whereas the super Markov combination becomes

$$\mathcal{F} \otimes \mathcal{G} = \left\{ \frac{f_{123} \cdot g_{23}}{f_{23}}, f_{123} \right\},$$

and thus a graphical combination. □

6. Conclusion

In this paper, we developed a general framework for combination of statistical models which was then specialized to Gaussian graphical models. After having described the combination of consistent and non-consistent distributions and introduced the concept of meta-consistent and quasi-consistent families of distributions, we presented three ways of combination for families of distributions: the lower Markov combination, the upper Markov combination and the super Markov combination. The first combines only consistent pairs of distributions and will in general reduce the original marginal families; it is equivalent to the meta-Markov combination of [**2**] when the original families are meta-consistent. The second combines all marginal distributions from the first family with all the conditional distributions of the second family and *vice versa*. In general, it extends the original marginal families. The third one permits to extend the marginal families in such a way that they become meta-consistent and can be combined by using the meta-Markov combination. It corresponds to combine any marginal distribution with any conditional distribution. Within this framework we then studied the combination of Gaussian graphical models.

The goal for the combination of models is also to combine inferences from the associated statistical analyses based on available data or for the individual graphical models. We shall study this aspect in detail in the future, both concerning combination of maximum likelihood estimates and Bayesian inferences.

References

[1] O. Barndorff-Nielsen, *Information and exponential families in statistical theory*, John Wiley & Sons Ltd., Chichester, 1978, Wiley Series in Probability and Mathematical Statistics.

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

[3] A. P. Dawid and M. Studený, *Conditional products: An alternative approach to conditional independence*, Artificial Intelligence and Statistics 99 (David Heckerman and Joe Whittaker, eds.), 1999, pp. 32–40.

[4] M. Drton, H. Massam, and I. Olkin, *Moments of minors of Wishart matrices*, Annals of Statistics **36** (2008), 2261–2283.

[5] M. Drton, B. Sturmfels, and S. Sullivant, *Algebraic factor analysis: Tetrads, pentads and beyond*, Probability Theory and Related Fields **138** (2007), 463–493.

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

[7] M. Frydenberg, *Marginalization and collapsibility in graphical interaction models*, Annals of Statistics **18** (1990), 790–805.

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

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

[10] L. V. Hedges and I. Olkin, *Statistical methods for meta-analysis*, Academic Press Inc., Orlando, FL, 1985.

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

[12] H. G. Kellerer, *Maßtheoretische Marginalprobleme*, Mathematische Annalen **153** (1964), 168–198.

[13] ———, *Verteilungsfunktionen mit gegebenen Marginalverteilungen*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **3** (1964), 247–270.

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

[15] S. L. Lauritzen, *Graphical models*, Clarendon Press, Oxford, 1996.

[16] C. Spearman, *"General intelligence", objectively determined and measured*, The American Journal of Psychology **15** (1904), 201–292.

[17] ———, *Pearson's contribution to the theory of two factors*, British Journal of Psychology **19** (1928), 95–101.

[18] P. Spirtes, C. Glymour, and R. Scheines, *Causation, prediction and search*, Springer-Verlag, New York, 1993.

[19] M. Studený, *Probabilistic conditional independence structures*, New York, 2005.

[20] S. Sullivant, *Algebraic geometry of Gaussian Bayesian networks*, Advances in Applied Mathematics **40** (2008), 482–513.

[21] R. E. Tillman, D. Danks, and C. Glymour, *Integrating locally learned causal structures with overlapping variables*, Advances in Neural Information Processing Systems 21, Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems, Vancouver, December 8-11, 2008 (Daphne Koller, Dale Schuurmans, Yoshua Bengio, and Léon Bottou, eds.), MIT Press, 2009, pp. 1665–1672.

[22] N. N. Vorob'ev, *Consistent families of measures and their extensions*, Theory of Probability and its Application **7** (1962), 147–163.

[23] ———, *Markov measures and Markov extensions*, Theory of Probability and its Application **8** (1963), 420–429.

Department of Statistical Sciences, University of Padua, Via Cesare Battisti 241, 35121 Padova, Italy

E-mail address: `massa@stat.unipd.it`

Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, OX1 3TG, United Kingdom

E-mail address: `steffen@stats.ox.ac.uk`