Estimating Graphical Models Combinations

Sofia Massa, University of Oxford

Workshop on Graphical Models: Mathematics, Statistics and Computer Sciences

Fields Institute, Toronto, 16-18 April 2012

Outline

Outline

1. Meta-analysis vs structural meta-analysis

Outline

1. Meta-analysis vs structural meta-analysis
2. Motivating examples

Outline

1. Meta-analysis vs structural meta-analysis
2. Motivating examples
3. Statistical models combination
    - Combination of distributions
    - Combination of families

Outline

1. Meta-analysis vs structural meta-analysis
2. Motivating examples
3. Statistical models combination
    - Combination of distributions
    - Combination of families
4. Graphical models combination
    - Examples
    - Graphical and non-graphical combinations
    - Estimation

Meta-analysis vs Structural Meta-Analysis

Meta-analysis vs Structural Meta-Analysis

- **Meta-analysis** is combining evidence about **important parameters** from experiments or studies performed independently under partially comparable circumstances;

Meta-analysis vs Structural Meta-Analysis

- **Meta-analysis** is combining evidence about **important parameters** from experiments or studies performed independently under partially comparable circumstances;
- **Structural meta-analysis** (Massa and Lauritzen, 2010), is the combination of evidence about **relationships** between variables from studies or experiments carried out independently under partially comparable conditions.

Meta-analysis vs Structural Meta-Analysis

- **Meta-analysis** is combining evidence about **important parameters** from experiments or studies performed independently under partially comparable circumstances;
- **Structural meta-analysis** (Massa and Lauritzen, 2010), is the combination of evidence about **relationships** between variables from studies or experiments carried out independently under partially comparable conditions.
- By relationships between variables we mean **conditional independence relations**, as defined by Dawid (1979).

Example: Two surveys

Gilula and McCullogh (2011)

Suppose data are available from two surveys A and B.

Example: Two surveys

Gilula and McCullogh (2011)

Suppose data are available from two surveys A and B.
Suppose there is no single investigation that contains all variables of interest.

Example: Two surveys

Gilula and McCullogh (2011)

Suppose data are available from two surveys A and B.
Suppose there is no single investigation that contains all variables of interest.

For example:

- Survey A investigates: age, income, gender, smoking habits, . . .;
- Survey B investigates: age, income, opinion about banning smoking in public, ...

Example: Cancer Studies

DuMouchel and Harris (1983)

Rows: 5 studies.

Columns: 10 variables investigated.

A filled circle means data available.

	ROOF TAR	COKE OVEN	GAS ENGINE	BaP	CIG.	DIESEL A	DIESEL B	DIESEL C	DIESEL D	DIESEL E
LUNG CANCER	●	●			●					●
SKIN TUMOR INIT.	●	●	●	●	●	●	●		●	
VIRAL TRANSFORM.	●	●	●	●	●	●	●	●	●	
MUTAGENESIS −MA	●	●	●		●	●	●	●	●	
MUTAGENESIS +MA	●	●	●		●	●	●	●	●	
	1	2	3	4	5	6	7	8	9	10

Example: Teen Drinking

Dee and Evans (2003)

- Data on teen drinking;
- Data on educational attainment;
- No dataset that contains both information on teen drinking and educational attainment.
- Study the effect of teen drinking on educational attainment.

Example: Diseasome

Goh et al. (2007)

- Example of a **graphical models meta-analysis**.
- A large set of diseases and relevant genes is combined to form "the human diseasome" bipartite network.

Example: Diseasome

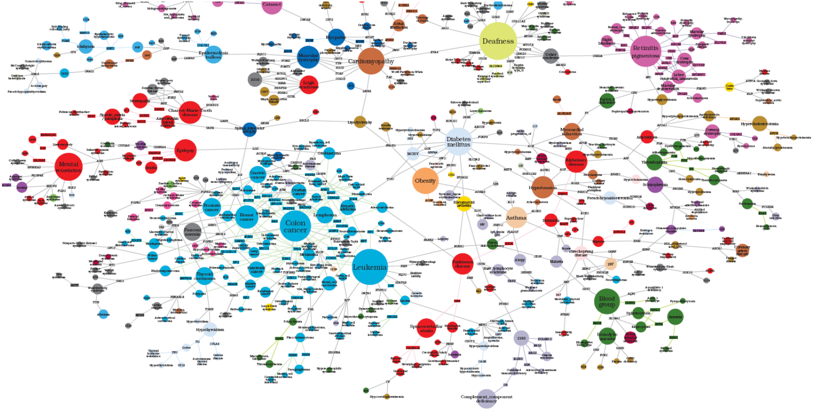
Goh et al. (2007)

- Example of a **graphical models meta-analysis**.
- A large set of diseases and relevant genes is combined to form "the human diseasome" bipartite network.
- The authors also generate two biologically relevant networks, the human disease network projection and the disease gene network projection.

Diseasome

Goh et al. (2007)

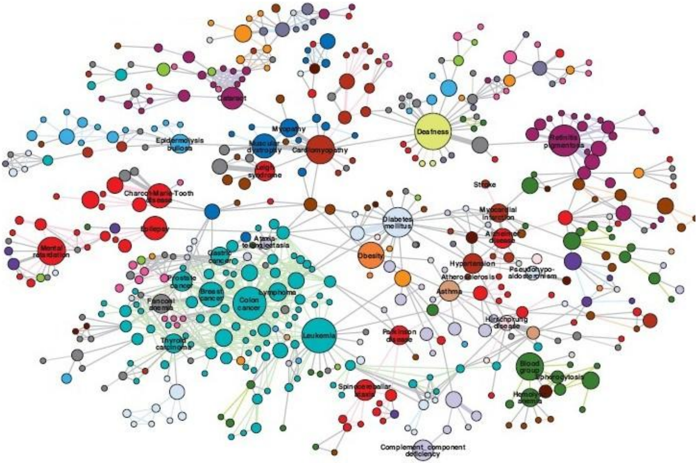
A subset of the diseasome bipartite network. Circles are **disorders**, rectangles are **disease genes**.



The Human Disease Network

Goh et al. (2007)

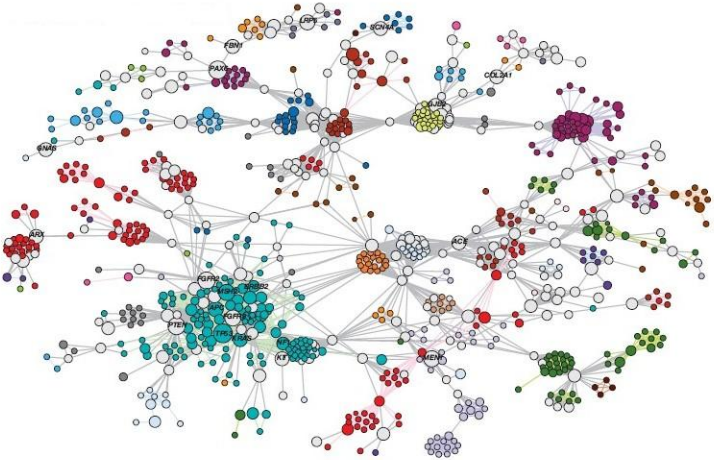
A subset of the **human disease network projection** of the bipartite graph. Each node corresponds to a distinct **disorder**.



The Disease Gene Network

Goh et al. (2007)

A subset of the **disease gene network projection** of the bipartite graph.
Each node is a **gene**.



Examples

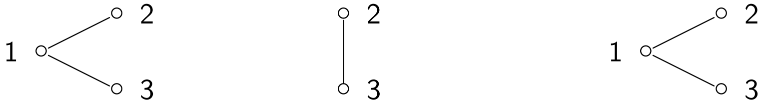


Figure: The rightmost graph represents the combination since both models imply the same constraints for the joint distribution of (Y_2, Y_3).

Examples

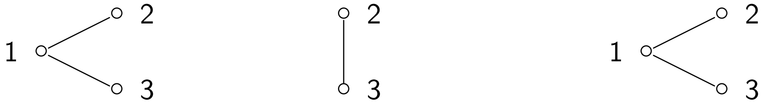


Figure: The rightmost graph represents the combination since both models imply the same constraints for the joint distribution of (Y_2, Y_3).

Figure: Here it is less obvious to define the combination and to represent the combination graphically.

Example

Figure: There are no conditional independence relationships expressed by the two graphs on the left. The graph on the right represents their combination.

General Setting

We can think of the previous examples in different contexts:

General Setting

We can think of the previous examples in different contexts:

- Different types of variables: discrete nodes, continuous and discrete nodes;

General Setting

We can think of the previous examples in different contexts:

- Different types of variables: discrete nodes, continuous and discrete nodes;
- Different types of graphs: for example DAGs.

General Setting

We can think of the previous examples in different contexts:

- Different types of variables: discrete nodes, continuous and discrete nodes;
- Different types of graphs: for example DAGs.
- In this talk, for simplicity the focus is only on **Gaussian variables**.

Basic Notation

- V is a set of variables.

Basic Notation

- V is a set of variables.
- Y_V, $v \in V$ random variable taking values in $\mathcal{Y}_v$.

Basic Notation

- V is a set of variables.
- Y_V, $v \in V$ random variable taking values in $\mathcal{Y}_v$.
- For $A \subseteq V$, $Y_A = (Y_v)_{v \in A}$ with values in $\mathcal{Y}_A = \times_{v \in A} \mathcal{Y}_v$.

Basic Notation

- V is a set of variables.
- Y_V, $v \in V$ random variable taking values in $\mathcal{Y}_v$.
- For $A \subseteq V$, $Y_A = (Y_v)_{v \in A}$ with values in $\mathcal{Y}_A = \times_{v \in A} \mathcal{Y}_v$.
- For f distribution over V and $A, B \subset V$, f_A denotes the marginal distribution of Y_A and $f_{B|A}$ the conditional distribution of $Y_{B \setminus A}$ given $Y_A = y_A$.

Basic Notation

- V is a set of variables.
- Y_V, $v \in V$ random variable taking values in $\mathcal{Y}_v$.
- For $A \subseteq V$, $Y_A = (Y_v)_{v \in A}$ with values in $\mathcal{Y}_A = \times_{v \in A} \mathcal{Y}_v$.
- For f distribution over V and $A, B \subset V$, f_A denotes the marginal distribution of Y_A and $f_{B|A}$ the conditional distribution of $Y_{B \setminus A}$ given $Y_A = y_A$.
- $\mathcal{F} = \{f\}$ is the family of distributions over $A \subseteq V$.

Basic Notation

- V is a set of variables.
- Y_V, $v \in V$ random variable taking values in $\mathcal{Y}_v$.
- For $A \subseteq V$, $Y_A = (Y_v)_{v \in A}$ with values in $\mathcal{Y}_A = \times_{v \in A} \mathcal{Y}_v$.
- For f distribution over V and $A, B \subset V$, f_A denotes the marginal distribution of Y_A and $f_{B|A}$ the conditional distribution of $Y_{B \setminus A}$ given $Y_A = y_A$.
- $\mathcal{F} = \{f\}$ is the family of distributions over $A \subseteq V$.
- $\mathcal{F}^{\downarrow C}$ are the induced marginal distributions over $C \subseteq A$.

Set-up

Consider two sets of variables A and B, and two families $\mathcal{F}$ and $\mathcal{G}$ of distributions for Y_A and Y_B, where A and $B \subseteq V$.

Set-up

Consider two sets of variables A and B, and two families $\mathcal{F}$ and $\mathcal{G}$ of distributions for Y_A and Y_B, where A and $B \subseteq V$.

Ideally search for a joint family of distributions, $\mathcal{H}$ for $Y_{A \cup B}$, such that

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

Combining Distributions

Combining Distributions

If f and g are **consistent**, i.e., $f_{A \cap B} = g_{A \cap B}$, their **Markov combination** (Dawid and Lauritzen, 1993) is

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

Combining Distributions

If f and g are **consistent**, i.e., $f_{A \cap B} = g_{A \cap B}$, their **Markov combination** (Dawid and Lauritzen, 1993) is

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

If f and g are **not consistent**, but $f_{A \cap B} << g_{A \cap B}$, their **right composition** (Jiroušek and Vejnarová, 2003) is

$$f \triangleright g = f \cdot \frac{g}{g_{A \cap B}}.$$

Combining Distributions

If f and g are **consistent**, i.e., $f_{A \cap B} = g_{A \cap B}$, their **Markov combination** (Dawid and Lauritzen, 1993) is

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

If f and g are **not consistent**, but $f_{A \cap B} << g_{A \cap B}$, their **right composition** (Jiroušek and Vejnarová, 2003) is

$$f \triangleright g = f \cdot \frac{g}{g_{A \cap B}}.$$

If $g_{A \cap B} << f_{A \cap B}$, their **left composition** (Jiroušek and Vejnarová, 2003) is

$$f \triangleleft g = \frac{f}{f_{A \cap B}} \cdot g.$$

Restrictive Type of Combination

Restrictive Type of Combination

The **lower Markov combination** of $\mathcal{F}$ and $\mathcal{G}$ is

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

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

Restrictive Type of Combination

The **lower Markov combination** of $\mathcal{F}$ and $\mathcal{G}$ is

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

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

The combination respects $(\mathcal{F} \,\underline{\star}\, \mathcal{G})^{\downarrow A} \subseteq \mathcal{F}, \quad (\mathcal{F} \,\underline{\star}\, \mathcal{G})^{\downarrow B} \subseteq \mathcal{G}$.

Restrictive Type of Combination

The **lower Markov combination** of $\mathcal{F}$ and $\mathcal{G}$ is

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

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

The combination respects $(\mathcal{F} \,\underline{\star}\, \mathcal{G})^{\downarrow A} \subseteq \mathcal{F}, \quad (\mathcal{F} \,\underline{\star}\, \mathcal{G})^{\downarrow B} \subseteq \mathcal{G}$.

If $\mathcal{F}$ and $\mathcal{G}$ are **meta-consistent**, i.e., $\mathcal{F}^{\downarrow A \cap B} = \mathcal{G}^{\downarrow A \cap B}$, this is the **meta-Markov combination** $\mathcal{F} \star \mathcal{G}$ of Dawid and Lauritzen (1993).

Less Restrictive Type of Combination

Less Restrictive Type of Combination

The **upper Markov combination** of $\mathcal{F}$ and $\mathcal{G}$ is

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

Less Restrictive Type of Combination

The **upper Markov combination** of $\mathcal{F}$ and $\mathcal{G}$ is

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

All marginal distributions of the two families are represented also in the combined family.
Here we have

$$(\mathcal{F} \,\bar{\star}\, \mathcal{G})^{\downarrow A} \supseteq \mathcal{F}^{\mathcal{G}}, \quad (\mathcal{F} \,\bar{\star}\, \mathcal{G})^{\downarrow B} \supseteq \mathcal{G}^{\mathcal{F}},$$

where $\mathcal{F}^{\mathcal{G}} = \{f \in \mathcal{F} \mid \exists g : f_{A\cap B} << g_{A\cap B}\}$.

Less Restrictive Type of Combination

The **upper Markov combination** of $\mathcal{F}$ and $\mathcal{G}$ is

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

All marginal distributions of the two families are represented also in the combined family.
Here we have

$$(\mathcal{F} \,\bar{\star}\, \mathcal{G})^{\downarrow A} \supseteq \mathcal{F}^{\mathcal{G}}, \quad (\mathcal{F} \,\bar{\star}\, \mathcal{G})^{\downarrow B} \supseteq \mathcal{G}^{\mathcal{F}},$$

where $\mathcal{F}^{\mathcal{G}} = \{f \in \mathcal{F} \mid \exists g : f_{A \cap B} << g_{A \cap B}\}$.

If $\mathcal{F}^{\mathcal{G}} = \mathcal{F}, \mathcal{G}^{\mathcal{F}} = \mathcal{G}$ we say that the families are **quasi-consistent**.

Less Restrictive Type of Combination

The **upper Markov combination** of $\mathcal{F}$ and $\mathcal{G}$ is

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

All marginal distributions of the two families are represented also in the combined family.
Here we have

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

where $\mathcal{F}^{\mathcal{G}} = \{f \in \mathcal{F} \mid \exists g : f_{A\cap B} << g_{A\cap B}\}$.

If $\mathcal{F}^{\mathcal{G}} = \mathcal{F}, \mathcal{G}^{\mathcal{F}} = \mathcal{G}$ we say that the families are **quasi-consistent**.

It holds that $\mathcal{F} \,\underline{\star}\, \mathcal{G} \subseteq \mathcal{F} \,\overline{\star}\, \mathcal{G}$.

Cuts and Equivalence of Combinations

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

Cuts and Equivalence of Combinations

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 $(\mathcal{F} \overline{\star} \mathcal{G})^{\downarrow B} = \mathcal{G}$.

Cuts and Equivalence of Combinations

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 $\mathcal{G}$.

Cuts and Equivalence of Combinations

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 $\mathcal{F} \overline{\star} \mathcal{G}$.

Recall that $Y_{A\cap B}$ is a cut in $\mathcal{F}$ if $\mathcal{F} \sim \mathcal{F}^{\downarrow A|(A\cap B)} \times \mathcal{F}^{\downarrow A\cap B}$, (Barndorff-Nielsen, 1978).

Maximal Extension of the Families

Maximal Extension of the Families

If we let

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

and

Maximal Extension of the Families

If we let

$$\mathcal{F}^\star = (\mathcal{F} \mathbin{\bar{\star}} \mathcal{G})^{\downarrow A}, \quad \mathcal{G}^\star = (\mathcal{F} \mathbin{\bar{\star}} \mathcal{G})^{\downarrow B},$$

and

$$\mathcal{F}^{\star\star} = (\mathcal{F}^\star \mathbin{\bar{\star}} \mathcal{G}^\star)^{\downarrow A}, \quad \mathcal{G}^{\star\star} = (\mathcal{F}^\star \mathbin{\bar{\star}} \mathcal{G}^\star)^{\downarrow B},$$

Maximal Extension of the Families

If we let

$$\mathcal{F}^\star = (\mathcal{F} \mathbin{\bar{\star}} \mathcal{G})^{\downarrow A}, \quad \mathcal{G}^\star = (\mathcal{F} \mathbin{\bar{\star}} \mathcal{G})^{\downarrow B},$$

and

$$\mathcal{F}^{\star\star} = (\mathcal{F}^\star \mathbin{\bar{\star}} \mathcal{G}^\star)^{\downarrow A}, \quad \mathcal{G}^{\star\star} = (\mathcal{F}^\star \mathbin{\bar{\star}} \mathcal{G}^\star)^{\downarrow B},$$

we have that

$$\mathcal{F} \mathbin{\underline{\star}} \mathcal{G} \subseteq \mathcal{F} \mathbin{\bar{\star}} \mathcal{G} \subseteq \mathcal{F}^{\star\star} \star \mathcal{G}^{\star\star},$$

Maximal Extension of the Families

If we let

$$\mathcal{F}^{\star} = (\mathcal{F} \mathbin{\bar{\star}} \mathcal{G})^{\downarrow A}, \quad \mathcal{G}^{\star} = (\mathcal{F} \mathbin{\bar{\star}} \mathcal{G})^{\downarrow B},$$

and

$$\mathcal{F}^{\star\star} = (\mathcal{F}^{\star} \mathbin{\bar{\star}} \mathcal{G}^{\star})^{\downarrow A}, \quad \mathcal{G}^{\star\star} = (\mathcal{F}^{\star} \mathbin{\bar{\star}} \mathcal{G}^{\star})^{\downarrow B},$$

we have that

$$\mathcal{F} \mathbin{\underline{\star}} \mathcal{G} \subseteq \mathcal{F} \mathbin{\bar{\star}} \mathcal{G} \subseteq \mathcal{F}^{\star\star} \star \mathcal{G}^{\star\star},$$

Super Markov combination $\mathcal{F} \otimes \mathcal{G}$ of $\mathcal{F}$ and $\mathcal{G}$ as the meta-Markov combination of the maximally extended families:

Maximal Extension of the Families

If we let

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

and

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

we have that

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

Super Markov combination $\mathcal{F} \otimes \mathcal{G}$ of $\mathcal{F}$ and $\mathcal{G}$ as the meta-Markov combination of the maximally extended families:

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

Summarising

- **Lower** Markov combination (restrictive case)

$$\mathcal{F} \underline{\star} \mathcal{G} = \left\{ \frac{f \cdot g}{f_{A \cap B}}, f \in \mathcal{F}, g \in \mathcal{G}, f \text{ and } g \text{ consistent} \right\}.$$

Summarising

- **Lower** Markov combination (restrictive case)

$$\mathcal{F} \underline{\star} \mathcal{G} = \left\{ \frac{f \cdot g}{f_{A \cap B}}, f \in \mathcal{F}, g \in \mathcal{G}, f \text{ and } g \text{ consistent} \right\}.$$

- **Upper** Markov combination (less restrictive case)

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

Summarising

- **Lower** Markov combination (restrictive case)

$$\mathcal{F} \underline{\star} \mathcal{G} = \left\{ \frac{f \cdot g}{f_{A \cap B}}, f \in \mathcal{F}, g \in \mathcal{G}, f \text{ and } g \text{ consistent} \right\}.$$

- **Upper** Markov combination (less restrictive case)

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

- **Super** Markov combination (maximal extension of the families)
$\mathcal{F} \otimes \mathcal{G} = \mathcal{F}^{\star\star} \star \mathcal{G}^{\star\star}$ also written as

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

Conditional Independence Assumption

- All combinations use the conditional independence assumption $A \perp\!\!\!\perp B | (A \cap B)$.
- If $A \perp\!\!\!\perp B | (A \cap B)$ does not hold, then the separate analyses of A and B can potentially be very misleading as the missing data in each case typically will induce spurious correlations.
- It may make sense to use this assumption and then consider the distortions that this may induce.

Combinations of Graphical Models

Combinations of Graphical Models

- Can we relate **meta-consistency** to a property of the graphs?

Combinations of Graphical Models

- Can we relate **meta-consistency** to a property of the graphs?
- Is there any context when the problem becomes very simple?

Combinations of Graphical Models

- Can we relate **meta-consistency** to a property of the graphs?
- Is there any context when the problem becomes very simple?
- Difference between **graphical** and **non-graphical** combinations.

Meta-consistency of Graphical Models

Meta-consistency of Graphical Models

- If two Gaussian graphical models are meta-consistent, the **marginal graphs** over the variables in the intersection must be identical. The converse is not generally true.

Meta-consistency of Graphical Models

- If two Gaussian graphical models are meta-consistent, the **marginal graphs** over the variables in the intersection must be identical. The converse is not generally true.
- If the two **dependence graphs** are isomorphic then meta-consistency is guaranteed.

Meta-consistency of Graphical Models

- If two Gaussian graphical models are meta-consistent, the **marginal graphs** over the variables in the intersection must be identical. The converse is not generally true.
- If the two **dependence graphs** are isomorphic then meta-consistency is guaranteed.
- Meta-consistency is related to but different from **collapsibility** of the two dependence graphs onto $A \cap B$.

Meta-consistency of Graphical Models

- If two Gaussian graphical models are meta-consistent, the **marginal graphs** over the variables in the intersection must be identical. The converse is not generally true.

- If the two **dependence graphs** are isomorphic then meta-consistency is guaranteed.

- Meta-consistency is related to but different from **collapsibility** of the two dependence graphs onto $A \cap B$.

- In general, all constraints on the common variables must be investigated.

Examples

Examples

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

Examples

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

Figure: Two meta-consistent graphical Gaussian models (the two graphs are isomorphic and therefore induce the same restrictions on the common variables).

Conditions for Equivalence of Combinations

If the graphs are collapsible onto $A \cap B$ and the induced subgraphs on the common variables are the same, then

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

Conditions for Equivalence of Combinations

If the graphs are collapsible onto $A \cap B$ and the induced subgraphs on the common variables are the same, then

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

Here the combination is **graphical** and its dependence graph is given by

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

Combinations of Graphical Models

Combinations of Graphical Models

- The **lower Markov combination** is the family of distributions that satisfies all induced constraints. These can be polynomial equality relations (conditional independence), tetrad and pentad constraints, and also inequality constraints (Drton et al., 2007).

Combinations of Graphical Models

- The **lower Markov combination** is the family of distributions that satisfies all induced constraints. These can be polynomial equality relations (conditional independence), tetrad and pentad constraints, and also inequality constraints (Drton et al., 2007).

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

Combinations of Graphical Models

- The **lower Markov combination** is the family of distributions that satisfies all induced constraints. These can be polynomial equality relations (conditional independence), tetrad and pentad constraints, and also inequality constraints (Drton et al., 2007).

- 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.

Combinations of Graphical Models

- The **lower Markov combination** is the family of distributions that satisfies all induced constraints. These can be polynomial equality relations (conditional independence), tetrad and pentad constraints, and also inequality constraints (Drton et al., 2007).

- 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.

- We need to distinguish between **graphical** and **non-graphical** combinations.

Example - Graphical Combination

Example - Graphical Combination

The **lower Markov combination** is
$\mathcal{F} \underline{\star} \mathcal{G} = \mathcal{F} \star \mathcal{G} = \{ Y \sim N_3(0, \Sigma), \Sigma^{-1} \in S^+(G_A), \Sigma_{\{2,3\}} = \Phi_{\{2,3\}} \}$.

The **upper** and **super Markov** combination are identical and the corresponding graph is a complete graph.

Example - Non-Graphical Combination

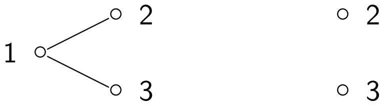
Example - Non-Graphical Combination

The **lower Markov combination** is

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

where $\Omega = \{\omega_{ij}\}$.

Example - Non-Graphical Combination



The **lower Markov combination** is

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

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 $(1, 3)$.

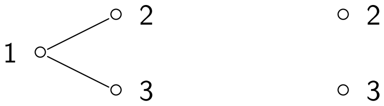
Example (cont.)

Example (cont.)

The **upper Markov combination** is

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

Example (cont.)



The **upper Markov combination** is

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

The **super Markov combination** is

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

and they are both **graphical** combinations.

Estimation

- Present interest only in **graphical** cases.

Estimation

- Present interest only in **graphical** cases.
- Consider an approach that exploits all the available information and converts the problem in a missing data one.

Estimation

- Present interest only in **graphical** cases.
- Consider an approach that exploits all the available information and converts the problem in a missing data one.
- The available initial data define the complexity of the estimation process (**raw data** vs **derived quantities**).

Missing Data Approach

An example

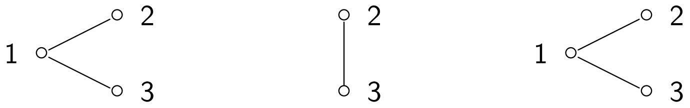


Figure 1: From left to right, families $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{F} \star \mathcal{G}$.

- $y_A = (y_j^i)$ with $j = 1, 2, 3$ and $i = 1, \cdots, n_A$ observations from $\mathcal{F}$.

Missing Data Approach

An example

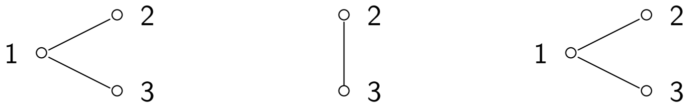


Figure 1: From left to right, families $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{F} \star \mathcal{G}$.

- $y_A = (y_j^i)$ with $j = 1, 2, 3$ and $i = 1, \cdots, n_A$ observations from $\mathcal{F}$.
- $y_B = (y_j^i)$ with $j = 2, 3$ and $i = 1, \cdots, n_B$ observations from $\mathcal{G}$, $n = n_A + n_B$.

Missing Data Approach

An example

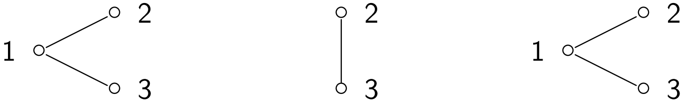


Figure 1: From left to right, families $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{F} \star \mathcal{G}$.

- $y_A = (y_j^i)$ with $j = 1, 2, 3$ and $i = 1, \cdots, n_A$ observations from $\mathcal{F}$.
- $y_B = (y_j^i)$ with $j = 2, 3$ and $i = 1, \cdots, n_B$ observations from $\mathcal{G}$, $n = n_A + n_B$.

	1	2	3
$\mathcal{F}$	n_A	n_A	n_A
$\mathcal{G}$		n_B	n_B

Table: Missing pattern for the problem considered.

EM Algorithm

The hypothetical complete data model has the form of a **regular exponential family** with unknown canonical parameter K, concentration matrix of the combination.

EM Algorithm

The hypothetical complete data model has the form of a **regular exponential family** with unknown canonical parameter K, concentration matrix of the combination.

Apply standard EM algorithm as detailed for mixed graphical models by Didelez and Pigeot (1998).

EM Algorithm

The hypothetical complete data model has the form of a **regular exponential family** with unknown canonical parameter K, concentration matrix of the combination.

Apply standard EM algorithm as detailed for mixed graphical models by Didelez and Pigeot (1998).

The partial imputation EM algorithm (Geng et al., 2000) would be more efficient when dealing with **high dimensional graphs** and **multiple combinations**.

EM Algorithm

- The sufficient statistics are given by

$$w_{jj} = \sum_{i=1}^{n} (y_j^i)^2, \; j = 1, 2, 3 \quad w_{1j} = \sum_{i=1}^{n} y_1^i y_j^i, \; j = 2, 3.$$

EM Algorithm

- The sufficient statistics are given by

$$w_{jj} = \sum_{i=1}^{n} (y_j^i)^2,\ j = 1, 2, 3 \quad w_{1j} = \sum_{i=1}^{n} y_1^i y_j^i,\ j = 2, 3.$$

- The maximum likelihood estimate for the complete data case is

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

$w_{[A]}^{ij}$ is the ijth element in $\hat{W}_{[A]}^{-1}$, where $W = \sum_{i=1}^{n} y^i (y^i)^T$.

E-Step

Compute the expected values of the complete data sufficient statistics, conditional on the observed data, using the current estimate of the parameter.

E-Step

Compute 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 $\theta^{(t)} = K^{(t)}$. The E-step computes

$$
\begin{aligned}
w_{1j}^{(t)} &= E\left(\sum_{i=1}^{n} Y_1^i Y_j^i \middle| Y_{obs}, \theta^{(t)}\right), \\
w_{11}^{(t)} &= E\left(\sum_{i=1}^{n} (Y_1^i)^2 \middle| Y_{obs}, \theta^{(t)}\right), \\
w_{jj} &= w_{jj}^{(t)} = E\left(\sum_{i=1}^{n} (Y_j^i)^2 | Y_{obs}, \theta^{(t)}\right).
\end{aligned}
$$

M-Step

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

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

The algorithm performs the two steps until convergence, after having specified an initial value K_0 for K.

Direct Estimation

- If the original graphs are both collapsible onto $A \cap B$ and the induced subgraphs on the common variables are the same all the combinations are identical.

Direct Estimation

- If the original graphs are both collapsible onto $A \cap B$ and the induced subgraphs on the common variables are the same all the combinations are identical.

- It would make no difference whether we have the **maximum likelihood estimates** from each of the experiments or the **raw data**.

Direct Estimation

- If the original graphs are both collapsible onto $A \cap B$ and the induced subgraphs on the common variables are the same all the combinations are identical.
- It would make no difference whether we have the **maximum likelihood estimates** from each of the experiments or the **raw data**.
- We can directly combine the estimates of the single models and a missing data approach is **not** required.

Direct Estimation

Figure: All combinations are identical to the combination on the right.

All the combinations are equivalent to

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

The estimation of the combination is given by the combination of the separate estimates, i.e., $\hat{f}$ and $\hat{g}$.

Bayesian Approach

- This part of the work is under development.

Bayesian Approach

- This part of the work is under development.
- Specification of the prior distribution depends on the type of combination. In general, hyper-Markov laws are needed.

Bayesian Approach

- This part of the work is under development.
- Specification of the prior distribution depends on the type of combination. In general, hyper-Markov laws are needed.
- MCMC is necessary for the estimation of the combined graphical model.

Bayesian Approach

- This part of the work is under development.
- Specification of the prior distribution depends on the type of combination. In general, hyper-Markov laws are needed.
- MCMC is necessary for the estimation of the combined graphical model.
- Variational methods can be investigated in this case and also for the previous context.

Bayesian Approach

- This part of the work is under development.
- Specification of the prior distribution depends on the type of combination. In general, hyper-Markov laws are needed.
- MCMC is necessary for the estimation of the combined graphical model.
- Variational methods can be investigated in this case and also for the previous context.
- A comparative evaluation of the two approaches is also interesting.

Some References

- Dawid, A. P., Lauritzen, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models, Ann. Statist., **21**, 1272–1317.
- Drton, M., Sturmfel, B., Sullivant, S. (2007). Algebraic factor analysis: tetrads, pentads and beyond, Probab. Theory Relat. Fields, **138**, 463 –493.
- Geng, Z., Wan, K., Tao, F. (2000). Mixed graphical models with missing data and the partial imputation EM algorithm. Scandin. J. of Statist., **27**, 433–444.
- Goh, K-Il., Cusick, M. E., Valle, D., Childs, B., Vidal,M., Barabasi, A-L. (2007). The human disease network, Proc. Nat. Acad. Sci. **104**, 8685–8690.
- Jiroušek, R., Vejnarová, J. (2003). Construction of multidimensional models by operator of composition: current state of art. Soft Computing, **7**, 328–335.
- Massa, M.S., Lauritzen, S.L. (2010). Combining statistical models. In Algebraic methods in statistics and probability II. AMS Contemporary Mathematics Series, 239–259.
- Massa, M.S., Lauritzen, S.L. (2012). Combining information from graphical models. Working paper.