Statistical Modelling - Module I
Graphical models
Lecture 1

Federico Castelletti

Department of Statistical Sciences
Università Cattolica del Sacro Cuore
Milan

Introduction

Graphical models : multivariate statistical models based on a graph

nodes $\Longleftrightarrow$ variables

edges $\Longleftrightarrow$ dependence relations between variables

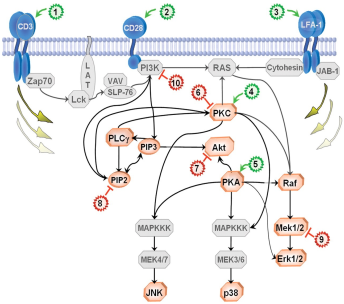
Given variables $X_1, \ldots, X_q$ and a graph $\mathcal{G}$
$\mathcal{G}$ implies a factorization of the joint density of $X_1, \ldots, X_q$
which makes computations with multivariate distributions easier

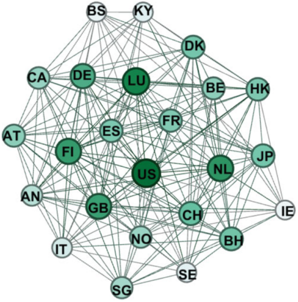
Directed Acyclic Graphs also used in causal inference
Be careful: the graph in itself represents dependence/association relationships
For causal statements we need further assumptions on the data generating mechanism

Literature on graphical models has grown increasingly in the last decades.
Applications are in many fields, e.g.

- computer science
- sociology
- biology

Example: genomics and finance





Left: a signaling network describing interactions between proteins from Sachs et al. (2005, *Science*)
Right: a financial network representing dependencies among liabilities of different countries; from Giudici and Spelta (2016, *Journal of Business and Economic Statistics*)

Graphical models: why?

Why needed? We already have Structural Equation Models (SEMs) and path analysis!

(1) Graph-based representations facilitate reading dependence relations between variables.

(2) The graph structure (path) is often *unknown* (*model uncertainty*) and must be inferred from the data.
The graph-model space (e.g. set of all directed/undirected graphs) is well characterized.
Search-algorithms built upon this space can be easily implemented.

Outline of the course

1. Conditional independence, graph theory and Markov properties
2. Some frequentist methods for graphical model selection: PC algorithm and Graphical Lasso
3. Bayesian graphical modelling
4. Bayesian networks and causal inference

Throughout the course, R code will be presented

Conditional independencies

Conditional independencies

$\boldsymbol{x} = \{X_1, \ldots, X_q\}$ set of variables with indexes $\{1, \ldots, q\} = V$
$\boldsymbol{x}_A$ subset of $\boldsymbol{x}$ with variables indexed by $A \subseteq V$
$p(\cdot)$ a continuous probability measure over $\boldsymbol{x}$

Given A, B, C disjoint subsets of V,
we say that $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ are (marginally) independent w.r.t. $p(\cdot)$ iff
$p(\boldsymbol{x}_A, \boldsymbol{x}_B) = p(\boldsymbol{x}_A)p(\boldsymbol{x}_B)$

Also, $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ are *conditionally* independent *given* $\boldsymbol{x}_C$ w.r.t. $p(\cdot)$ iff
$p(\boldsymbol{x}_A, \boldsymbol{x}_B \,|\, \boldsymbol{x}_C) = p(\boldsymbol{x}_A \,|\, \boldsymbol{x}_C)p(\boldsymbol{x}_B \,|\, \boldsymbol{x}_C)$
i.e. *once* the value of $\boldsymbol{x}_C$ is known, $\boldsymbol{x}_A$ and $\boldsymbol{x}_B$ does not "influence" each other

We then write: $\boldsymbol{x}_A \perp\!\!\!\perp \boldsymbol{x}_B \,|\, \boldsymbol{x}_C$

Conditional independencies

Equivalently:

$$
\begin{aligned}
p(\boldsymbol{x}_A, \boldsymbol{x}_B \mid \boldsymbol{x}_C) &= p(\boldsymbol{x}_A \mid \boldsymbol{x}_C)\, p(\boldsymbol{x}_B \mid \boldsymbol{x}_C) \\
\frac{p(\boldsymbol{x}_A, \boldsymbol{x}_B, \boldsymbol{x}_C)}{p(\boldsymbol{x}_C)} &= \frac{p(\boldsymbol{x}_A, \boldsymbol{x}_C)}{p(\boldsymbol{x}_C)} \frac{p(\boldsymbol{x}_B, \boldsymbol{x}_C)}{p(\boldsymbol{x}_C)} \\
p(\boldsymbol{x}_A, \boldsymbol{x}_B, \boldsymbol{x}_C) &= \frac{p(\boldsymbol{x}_A, \boldsymbol{x}_C)\, p(\boldsymbol{x}_B, \boldsymbol{x}_C)}{p(\boldsymbol{x}_C)}
\end{aligned}
$$

Factorizes the joint density of $\boldsymbol{x}_A, \boldsymbol{x}_B, \boldsymbol{x}_C$ in order to reflect the conditional independence relation $\boldsymbol{x}_A \perp\!\!\!\perp \boldsymbol{x}_B \mid \boldsymbol{x}_C$.

Graphically (details later on):

$$A \;\text{———}\; C \;\text{———}\; B$$

Conditional independencies

Example 1 (categorical data)

$\boldsymbol{x} = \{X_1, X_2, X_3\}$

$X_1 \in \{a, b\} = \mathcal{X}_1$
$X_2 \in \{0, 1\} = \mathcal{X}_2$
$X_3 \in \{M, F\} = \mathcal{X}_3$

$\mathcal{X}_j$ set of levels of variable $X_j,\ j = 1, 2, 3$

Then $\boldsymbol{x}$ takes values in the space $\mathcal{X}_V = \bigtimes_{j=1}^{3} \mathcal{X}_j = \big\{\{a, 0, M\}, \ldots, \{b, 1, F\}\big\}$

$\mathcal{X}_V$ can be represented as a 3-dimensional contingency table
Each cell corresponds to an element of $\mathcal{X}_V$ (level of $\boldsymbol{x}$)

Conditional independencies

Example 1 (categorical data)

For example, we can represent 2-dim conditional contingency tables (slices of the 3-dim contingency table) such as:

$(X_1, X_2) \mid X_3 = m$

$X_1 \backslash X_2$	0	1	
a	$\{a, 0, m\}$	$\{a, 1, m\}$	
b	$\{b, 0, m\}$	$\{b, 1, m\}$	

$(X_1, X_2) \mid X_3 = f$

$X_1 \backslash X_2$	0	1	
a	$\{a, 0, f\}$	$\{a, 1, f\}$	
b	$\{b, 0, f\}$	$\{b, 1, f\}$	

Conditional independencies

Example 1 (categorical data)

For each cell (level) $x = (x_1, x_2, x_3) \in \mathcal{X}_V$ we can define the (joint) probability

$$
\begin{aligned}
p(x) &= p(x_1, x_2, x_3) \\
&= p\left(X_1 = x_1, X_2 = x_2, X_3 = x_3\right) = p_{x_1 x_2\, x_3}
\end{aligned}
$$

All (other) marginal (and conditional) probabilities can be obtained from the previous, e.g. by aggregation:

$$
\begin{aligned}
p(x_1, x_3) &= \sum_{x_2 \in \mathcal{X}_2} p(x_1, x_2, x_3) \\
p(x_3) &= \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} p(x_1, x_2, x_3) \\
p(x_1 \mid x_3) &= \frac{p(x_1, x_3)}{p(x_3)}
\end{aligned}
$$

Conditional independencies

Example 1 (categorical data)

If we want $X_1 \perp\!\!\!\perp X_2 \mid X_3$ then we must guarantee

$$p(x_1, x_2 \mid x_3) = p(x_1 \mid x_3)\, p(x_2 \mid x_3) \qquad \text{for each } x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, x_3 \in \mathcal{X}_3$$

or equivalently

$$p(x_1, x_2, x_3) = \frac{p(x_1, x_3)\, p(x_2, x_3)}{p(x_3)} \qquad \text{for each } x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, x_3 \in \mathcal{X}_3$$

$(X_1, X_2) \mid X_3 = m$

$X_1 \backslash X_2$	0	1	
a	$p_{a0\mid m}$	$p_{a1\mid m}$	$p_{a.\mid m}$
b	$p_{b0\mid m}$	$p_{b1\mid m}$	$p_{b.\mid m}$
	$p_{.0\mid m}$	$p_{.1\mid m}$	1

$(X_1, X_2) \mid X_3 = f$

$X_1 \backslash X_2$	0	1	
a	$p_{a0\mid f}$	$p_{a1\mid f}$	$p_{a.\mid f}$
b	$p_{b0\mid f}$	$p_{b1\mid f}$	$p_{b.\mid f}$
	$p_{.0\mid f}$	$p_{.1\mid f}$	1

Conditional independencies

Example 1 (categorical data)

Some observations:

(1) Conditional independencies imply *constraints* on *parameters* (cell probabilities)

(2) These constraints can be used as hypothesis to test (using some available data) in order to check for conditional independencies relations

This idea is at the basis of *constraint-based* methods for graphical model selection

(3) In a Bayesian framework, such constraints imply "restrictions“ on the priors assigned to (graph-dependent) model parameters

Conditional independencies

Example 2 (Gaussian data)

Consider the (Structural Equation) Model:

$$
\begin{aligned}
X_1 &= \varepsilon_1 & \varepsilon_1 &\sim \mathcal{N}(0, \sigma_1^2) & & \\
X_2 &= \beta_{21} X_1 + \varepsilon_2 & \varepsilon_2 &\sim \mathcal{N}(0, \sigma_2^2) & & \varepsilon_j \perp\!\!\!\perp \varepsilon_k \text{ for each } j \neq k \\
X_3 &= \beta_{32} X_2 + \varepsilon_3 & \varepsilon_3 &\sim \mathcal{N}(0, \sigma_3^2) & &
\end{aligned}
$$

In matrix notation:

$$
\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ 0 & \beta_{32} & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \qquad \boldsymbol{x} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{\varepsilon} \qquad \boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\varepsilon}
$$

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{x}) &= \boldsymbol{0} \\
\mathbb{V}\text{ar}(\boldsymbol{x}) &= \boldsymbol{\Sigma} = (\boldsymbol{I} - \boldsymbol{B})^{-1} \boldsymbol{D} (\boldsymbol{I} - \boldsymbol{B})^{-\top} \qquad \boldsymbol{D} = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2) \\
\boldsymbol{x} &\sim \mathcal{N}_q(\boldsymbol{0}, \boldsymbol{\Sigma})
\end{aligned}
$$

Conditional independencies

Example 2 (Gaussian data)

After some calculations we recover the variance and inverse-covariance matrices:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \cdots & \cdots \\ \beta_{21}\sigma_1^2 & \beta_{21}^2\sigma_2^2 & \cdots \\ \beta_{21}\beta_{32}\sigma_1^2 & \beta_{21}^2\beta_{32}\sigma_1^2 + \beta_{32}\sigma_2^2 & \beta_{21}^2\beta_{32}^2\sigma_1^2 + \beta_{32}^2\sigma_2^2 + \sigma_3^2 \end{bmatrix}$$

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} + \frac{\beta_{21}^2}{\sigma_2^2} & \cdots & \cdots \\ -\frac{\beta_{21}}{\sigma_2^2} & \frac{1}{\sigma_2^2} + \frac{\beta_{32}^2}{\sigma_3^2} & \cdots \\ 0 & -\frac{\beta_{32}}{\sigma_3^2} & \frac{1}{\sigma_3^2} \end{bmatrix}$$

Conditional independencies

Example 2 (Gaussian data)

Graphically, this Structural Equation model can be represented as

$$X_1 \longrightarrow X_2 \longrightarrow X_3$$

Homework. Prove that:

(1) $f(x_1, x_3) \neq f(x_1)f(x_3)$

(2) $f(x_1, x_3 \mid x_2) = f(x_1 \mid x_2)f(x_3 \mid x_2)$ that is $X_1 \perp\!\!\!\perp X_3 \mid X_2$

[using properties of multivariate Normal distribution]

Graph theory

Graph theory

Some definitions

$\mathcal{G} = (V, E)$ a graph
$V = \{1, \ldots, q\}$ a set of vertices (or nodes)
$E \subseteq V \times V$ a set of edges

For two nodes $u, v \in V$ $(u \neq v)$
if $(u, v) \in E$ and $(v, u) \notin E$ then $\mathcal{G}$ contains the directed edge $u \rightarrow v$
if $(u, v) \in E$ and $(v, u) \in E$ then $\mathcal{G}$ contains the undirected edge $u - v$
u and v are adjacent if they connected by an edge (directed or undirected)

If $u - v$ in $\mathcal{G}$ then u is a *neighbor* of v in $\mathcal{G}$
$\mathrm{ne}_{\mathcal{G}}(v)$ set of neighbors of v in $\mathcal{G}$
If $u \rightarrow v$ in $\mathcal{G}$ then u is a *parent* of v and v is a *son* of u
$\mathrm{pa}_{\mathcal{G}}(u)$ set of all parents of u in $\mathcal{G}$

For $A \subseteq V$, induced sub-graph of $\mathcal{G} = (V, E)$ is
$\mathcal{G}_A = (A, E_A)$ with $E_A = \{(u, v) \in E \,|\, u \in A, v \in A\}$

Graph theory

Some definitions: Example

$$\begin{tikzcd}
1 \arrow[r, dash] & 2 \arrow[r, dash] \arrow[d, dash] & 3 \arrow[d] & \\
 & 4 \arrow[r] & 5 \arrow[r] & 6
\end{tikzcd}$$

$V = \{1, 2, 3, 4, 5, 6\}$
$E = \{(1, 2), (2, 1), \dots, (5, 6)\}$

$\mathrm{ne}_{\mathcal{G}}(2) = \{1, 3, 4\}$
$\mathrm{ne}_{\mathcal{G}}(4) = \{2\}$

$\mathrm{pa}_{\mathcal{G}}(2) = \emptyset$
$\mathrm{pa}_{\mathcal{G}}(5) = \{3, 4\}$

Graph theory

Some definitions

If $\mathcal{G}$ contains only directed edges we call it a *directed graph*
If $\mathcal{G}$ contains only undirected edges we call it a *undirected graph* (UG)

The sequence of distinct vertices $\{v_0, v_1, \dots, v_k\}$ in $\mathcal{G}$ is a *path* from v_0 to v_k
if $\mathcal{G}$ contains $v_{j-1} - v_j$ or $v_{j-1} \rightarrow v_j$ for all $j = 1, \dots, k$.
A path is *directed* (*undirected*) if all edges are directed (undirected)
A path is *partially directed* if it contains at least one directed edge

If there exists a path from u to v we say that v is a *descendant* of u

A sequence of nodes $\{v_0, v_1, \dots, v_k\}$ with $v_0 = v_k$
and $v_{j-1} - v_j$ or $v_{j-1} \rightarrow v_j$ for all $j = 1, \dots, k$ is called a *cycle*
A cycle is directed (undirected) if it contains only directed (undirected) edges

Graph theory

Some definitions: Example

$$
\begin{tikzcd}
1 \arrow[r, dash] & 2 \arrow[r, dash] \arrow[d, dash] & 3 \arrow[d] & \\
 & 4 \arrow[r] & 5 \arrow[r] & 6
\end{tikzcd}
$$

$1 - 2 - 4$ is an undirected path

$3 \to 5 \to 6$ is a directed path

$1 - 2 - 4 \to 5 \to 6$ is a partially directed path

5 is a descendant of 1

3 is not a descendant of 5

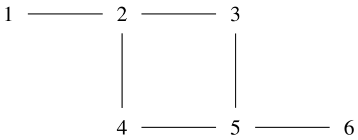
$\mathcal{G}$ has no (directed/undirected) cycles

Graph theory

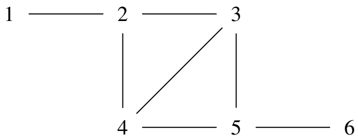
Some definitions: decomposable graphs

$\mathcal{G}$ UG is *decomposable* (or *chordal* or *triangulated*)
if every cycle of length $l \geq 4$ has a *chord* (two non-consecutive adjacent vertices)

$\mathcal{G}_1$



$\mathcal{G}_2$



$\mathcal{G}_2$ is decomposable since cycle $2-4-5-3-2$ has the chord $3-4$
$\mathcal{G}_1$ is not decomposable

A complete subset that is maximal with respect to inclusion is called a *clique*

Graph theory

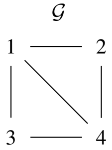
Some definitions: decomposable graphs

Let $\mathcal{C} = \{C_1, \dots, C_K\}$ sequence of cliques of $\mathcal{G}$
We can define, for $k = 2, \dots, K$, the three types of sets

$$H_k = C_1 \cup \cdots \cup C_k \quad S_k = C_k \cap H_{k-1} \quad R_k = C_k \setminus H_{k-1}$$

called *history*, *separators* and *residuals*
and with $S_1 = \emptyset, H_1 = C_1$ by assumption

A decomposable UG can be uniquely represented through its sets of cliques and separators



A decomposable graph $\mathcal{G}$ on the set of vertices $V = \{1, 2, 3, 4\}$; the cycle $\{1, 2, 4, 3\}$ of length $l = 4$ contains the chord $1 - 4$. $\mathcal{G}$ has the perfect sequence of cliques $\{C_1, C_2\}$, with $C_1 = \{1, 2, 4\}$, $C_2 = \{1, 3, 4\}$, and then $H_2 = V, S_2 = \{1, 4\}, R_2 = \{3\}$.

Markov properties

Markov properties

A graph $\mathcal{G}$ (directed or undirected) encodes conditional independence relations between nodes

The set of *all* conditional independencies encoded in $\mathcal{G}$ defines its *Markov property*

If we model the joint density of $X_1, \dots, X_q$, say $f(\cdot)$, according to $\mathcal{G}$
we want the cond. indep. encoded by $\mathcal{G}$ to be reflected by $f(\cdot)$

This is often (but not always) guaranteed and depends on the family of $f(\cdot)$

Yet, there are different types of Markov properties that $f(\cdot)$ can satisfy

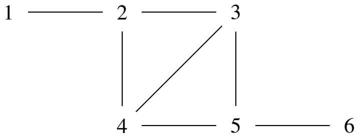
Markov properties

Undirected graphs: separation

A, B, C disjoint subsets of V

Def. A and B are separated by C in $\mathcal{G}$ if every walk from A to B contains a node in C (equivalently, after removing C, A and B are separated)

If A and B are separated by C in $\mathcal{G}$, then $A \perp\!\!\!\perp B \mid C$ in $\mathcal{G}$



$A = \{2\}$ and $B = \{5\}$ are separated by $C = \{3, 4\}$

Hence, $2 \perp\!\!\!\perp 5 \mid \{3, 4\}$

Also: $1 \perp\!\!\!\perp 5 \mid \{2, 3, 4\}$ but simply $1 \perp\!\!\!\perp 5 \mid 2$ too

Markov properties

Undirected graphs: Global Markov property

Let $\mathcal{M}_{\mathcal{G}}$ be the set of (all) conditional independencies encoded by $\mathcal{G}$

Consider a probability measure $p(\cdot)$ over V (nodes/variables)

Let $\mathcal{M}_p$ be the set of conditional independencies in $p(\cdot)$

If $\mathcal{M}_{\mathcal{G}} \subseteq \mathcal{M}_p$ then $p(\cdot)$ satisfies the *global* Markov property of $\mathcal{G}$:
A and B separated by C in $\mathcal{G} \implies A \perp\!\!\!\perp B \mid C$ in $p(\cdot)$
and so $p(\boldsymbol{x}_A, \boldsymbol{x}_B \mid \boldsymbol{x}_C) = p(\boldsymbol{x}_A \mid \boldsymbol{x}_C)\, p(\boldsymbol{x}_B \mid \boldsymbol{x}_C)$

If also $\mathcal{M}_{\mathcal{G}} = \mathcal{M}_p$ then $p(\cdot)$ is perfectly Markovian to $\mathcal{G}$

For some parametric families (e.g. Gaussian, categorical/multinomial),
the existence of a perfectly Markovian distribution for each UG $\mathcal{G}$ has been proved

Markov properties

Undirected graphs: Local and pairwise Markov properties

Global Markov property is difficult to establish
(many conditional indep. statements to check; see also next example)

Def. (Local Markov property)
$p(\cdot)$ satisfies the *local* Markov property of $\mathcal{G}$ if
$u \perp\!\!\!\perp V \setminus \{u \cup \mathrm{ne}_{\mathcal{G}}(u)\} \mid \mathrm{ne}_{\mathcal{G}}(u)$ in $p(\cdot)$

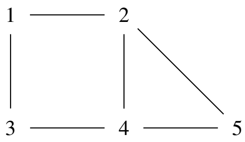
Def. (Pairwise Markov property)
$p(\cdot)$ satisfies the *pairwise* Markov property of $\mathcal{G}$ if
$u \perp\!\!\!\perp v \mid V \setminus \{u, v\}$ in $p(\cdot)$

It is possible to show that $(G) \implies (L) \implies (P)$
By converse, $(P) \implies (L) \implies (G)$ holds only if $p(\cdot)$ has a strictly positive density

The conditional independence statements relative to a graph $\mathcal{G}$ are of different types (G, L, P);
this means that if $p(\cdot)$ has a strictly positive density then we can limit to check
conditional independencies of type P to prove that $p(\cdot)$ satisfies the Global Markov property of $\mathcal{G}$

Markov properties

Example



$1 \perp\!\!\!\perp 4 \mid \{2,3\}$			G
$1 \perp\!\!\!\perp 4 \mid \{2,3,5\}$	P	L	G
$2 \perp\!\!\!\perp 3 \mid \{1,4\}$			G
$2 \perp\!\!\!\perp 3 \mid \{1,4,5\}$	P	L	G
$1 \perp\!\!\!\perp 5 \mid \{2,4\}$			G
$1 \perp\!\!\!\perp 5 \mid \{2,3\}$			G
$1 \perp\!\!\!\perp 5 \mid \{2,3,4\}$	P		G
$3 \perp\!\!\!\perp 5 \mid \{2,4\}$			G
$3 \perp\!\!\!\perp 5 \mid \{1,4\}$			G
$3 \perp\!\!\!\perp 5 \mid \{1,2,4\}$	P		G
$\{1,3\} \perp\!\!\!\perp 5 \mid \{2,4\}$		L	G
$\{2,5\} \perp\!\!\!\perp 3 \mid \{1,4\}$		L	G
$\{4,5\} \perp\!\!\!\perp 1 \mid \{2,3\}$		L	G

Markov properties

Undirected graphs: Factorization property

We say that $p(\cdot)$ factorize according to $\mathcal{G}$ if $p(\cdot)$ can be written as

$$p(x_1, \ldots, x_q) \propto \prod_{C \in \mathcal{C}} \phi(\boldsymbol{x}_C)$$

$\phi(\cdot)$ are called potentials
$\mathcal{C}$ is the set of cliques of $\mathcal{G}$

This condition is known as *factorization* property (F)
$(F) \implies (G)$ always!

So, if we can factorize $p(\cdot)$ according to $\mathcal{G}$ as above,

then the (global) Markov property of $\mathcal{G}$ is translated to $p(\cdot)$

Markov properties

Directed Acyclic Graphs

Several criteria available to read conditional independencies from a DAG $\mathcal{D}$

- d-separation (Pearl)
- moralization criterion (Lauritzen) [...]

They are all equivalent. We focus on the *moralization* criterion.
Moral graph of $\mathcal{D}$, $\mathcal{D}^m$, is the undirected graph with same vertex set of $\mathcal{D}$ and $u - v$ if and only if one of the following

- $u \to v$ or $u \leftarrow v$ are in $\mathcal{D}$
- u and v are involved in a v-structure $u \to z \leftarrow v$

$\mathcal{D}$

$$\begin{tikzcd} 1 \arrow[r] \arrow[d] & 2 \arrow[d] \\ 3 \arrow[r] & 4 \end{tikzcd}$$

$\mathcal{D}^m$

$$\begin{tikzcd} 1 \arrow[r, dash] \arrow[d, dash] & 2 \arrow[d, dash] \arrow[dl, dash] \\ 3 \arrow[r, dash] & 4 \end{tikzcd}$$

Markov properties

Directed Acyclic Graphs: separation

Lemma (Lauritzen, 1996)
Let $\mathcal{D} = (V, E)$ be a DAG and $A, B, S \subseteq V$ three disjoint subsets.
Then $A \perp\!\!\!\perp B \,|\, S$ whenever A and B are separated by S in $(\mathcal{D}_{\mathrm{An}(A \cup B \cup S)})^m$,
the moral graph of the smallest ancestral set containing $A \cup B \cup S$.

v *ancestor* of u in $\mathcal{D}$ if there is a (directed) path from v to u
(by convention u is ancestor of itself)

Equivalently, u is an *descendant* of v

$\mathrm{An}(v)$ is the set of all ancestors of u

$\mathcal{D}$

$$\begin{tikzcd} 1 \arrow[r] \arrow[d] & 2 \arrow[d] \\ 3 \arrow[r] & 4 \end{tikzcd}$$

$\mathcal{D}^m$

$$\begin{tikzcd} 1 \arrow[r, dash] \arrow[d, dash] & 2 \arrow[d, dash] \arrow[dl, dash] \\ 3 \arrow[r, dash] & 4 \end{tikzcd}$$

$(\mathcal{D}_{\{1,2,3\}})^m$

$$\begin{tikzcd} 1 \arrow[r, dash] \arrow[d, dash] & 2 \\ 3 & \end{tikzcd}$$

Check if $2 \perp\!\!\!\perp 3 \,|\, 1$ and $2 \perp\!\!\!\perp 3 \,|\, 4$ in $\mathcal{D}$

Markov properties

Directed Acyclic Graphs: Global Markov property

Let $\mathcal{M}_{\mathcal{D}}$ be the set of (all) conditional independencies encoded by $\mathcal{D}$

Consider a probability measure $p(\cdot)$ over V (nodes/variables)

Let $\mathcal{M}_p$ be the set of conditional independencies in $p(\cdot)$

If $\mathcal{M}_{\mathcal{D}} \subseteq \mathcal{M}_p$ then $p(\cdot)$ satisfies the *global* Markov property of $\mathcal{G}$:
$A \perp\!\!\!\perp B \,|\, C$ in $\mathcal{D} \implies A \perp\!\!\!\perp B \,|\, C$ in $p(\cdot)$
and so $p(\boldsymbol{x}_A, \boldsymbol{x}_B \,|\, \boldsymbol{x}_C) = p(\boldsymbol{x}_A \,|\, \boldsymbol{x}_C)\, p(\boldsymbol{x}_B \,|\, \boldsymbol{x}_C)$

If also $\mathcal{M}_{\mathcal{D}} = \mathcal{M}_p$ then $p(\cdot)$ is perfectly Markovian to $\mathcal{D}$

As for UGs, the existence of a perfectly Markovian distribution for each DAG $\mathcal{D}$
has been proved for some parametric families (e.g. Gaussian, categorical/multinomial)

Markov properties

Directed Acyclic Graphs: Factorization property

We say that $p(\cdot)$ factorize according to $\mathcal{D}$ if $p(\cdot)$ can be written as

$$p(x_1, \dots, x_q) = \prod_{j=1}^{q} f(x_j \,|\, \boldsymbol{x}_{\mathrm{pa}_{\mathcal{D}}(j)})$$

$\mathrm{pa}_{\mathcal{D}}(j)$ set of parents of node j

$f(x_j \,|\, \boldsymbol{x}_{\mathrm{pa}_{\mathcal{D}}(j)})$ conditional distribution of X_j

This condition is known as *factorization* property (F) of a DAG
$(F) \implies (G)$ always

So, as for the UG case, if we can factorize $p(\cdot)$ according to $\mathcal{D}$ as above,

then the (global) Markov property of $\mathcal{D}$ is translated to $p(\cdot)$

Markov properties

Directed Acyclic Graphs: Markov equivalence

Different DAGs can encode the same conditional independencies

For instance:

$$\mathcal{D}_1 \qquad \begin{tikzcd} 1 \arrow[r] \arrow[d] & 2 \arrow[d] \\ 3 \arrow[r] & 4 \end{tikzcd} \qquad\qquad \mathcal{D}_2 \qquad \begin{tikzcd} 1 \arrow[d] & 2 \arrow[l] \arrow[d] \\ 3 \arrow[r] & 4 \end{tikzcd}$$

If so, they are called *Markov equivalent*

Markov equivalent DAGs represent the *same* statistical model

Important in *score-based* model selection methods:

Markov equivalent DAGs should be "scored" equally

Markov equivalence has been also characterized graphically

Markov properties

Directed Acyclic Graphs: Markov equivalence

Theorem (Verma and Pearl, 1990)
Two DAGs $\mathcal{D}_1$ and $\mathcal{D}_2$ are Markov equivalent
if and only if they have the same skeleton and the same *v*-structures.

For instance, $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ are Markov equivalent
Also, there are not other DAGs that are Markov equivalent to them
In other terms, they represent a Markov equivalence *class*

$\mathcal{D}_1$

$$\begin{tikzcd} 1 \arrow[r] \arrow[d] \arrow[dr] & 2 \arrow[d] \\ 3 \arrow[r] & 4 \end{tikzcd}$$

$\mathcal{D}_2$

$$\begin{tikzcd} 1 \arrow[r] \arrow[dr] & 2 \arrow[d] \\ 3 \arrow[r] \arrow[u] & 4 \end{tikzcd}$$

$\mathcal{D}_3$

$$\begin{tikzcd} 1 \arrow[d] \arrow[dr] & 2 \arrow[l] \arrow[d] \\ 3 \arrow[r] & 4 \end{tikzcd}$$

Markov properties

Directed Acyclic Graphs: Markov equivalence

The space of DAGs can be partitioned into Markov equivalence classes

Each class can be represented by a *partially directed* graph called *Essential Graph* (EG) or *Completed Partially Directed Acyclic Graph* (CPDAG)

$\mathcal{D}_1$

$$\begin{tikzcd} 1 \arrow[r] \arrow[d] \arrow[dr] & 2 \arrow[d] \\ 3 \arrow[r] & 4 \end{tikzcd}$$

$\mathcal{D}_2$

$$\begin{tikzcd} 1 \arrow[r] \arrow[dr] & 2 \arrow[d] \\ 3 \arrow[u] \arrow[r] & 4 \end{tikzcd}$$

$\mathcal{D}_3$

$$\begin{tikzcd} 1 \arrow[d] \arrow[dr] & 2 \arrow[l] \arrow[d] \\ 3 \arrow[r] & 4 \end{tikzcd}$$

$\mathcal{G}(\mathcal{D}_1)$

$$\begin{tikzcd} 1 \arrow[r, dash] \arrow[d, dash] \arrow[dr] & 2 \arrow[d] \\ 3 \arrow[r] & 4 \end{tikzcd}$$

Definition (Essential Graph)

Let $\mathcal{D}$ be a DAG and $[\mathcal{D}]$ its Markov equivalence class.

The essential graph of $\mathcal{D}$ is defined as $\mathcal{G}(\mathcal{D}) := \bigcup_{\mathcal{D}^* \in [\mathcal{D}]} \mathcal{D}^*$.

Markov properties

Directed Acyclic Graphs: Markov equivalence

Essential Graphs are also characterized graphically (Andersson et al., 1997) as chain graphs with *decomposable* chain components

Obs. We cannot estimate a DAG in general, but only its equivalence class (or representative EG)

Markov equivalent DAGs can be however *different* from a *causal* perspective

[More on this in Lecture 4]

References

SACHS, K., PEREZ, O., PE'ER, D., PE'ER, D., LAUFFENBURGER, D.A. & NOLAN, G.P. (2005).
Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data.
Science **308**, 523-529.

GIUDICI, P., SPELTA, A. (2016).
Graphical Network Models for International Financial Flows.
Journal of Business and Economic Statistics **34**, 128-138.

MAATHUIS, M., DRTON, M., LAURITZEN, S., WAINWRIGHT, M (Eds.) (2019).
Handbook of Graphical Models.
CRC Press.

LAURITZEN, S.L. (1996).
Graphical Models.
Oxford University Press.

ANDERSSON, S.A., MADIGAN, D. & PERLMAN, M.D. (1997).
A characterization of Markov equivalence classes for acyclic digraphs.
The Annals of Statistics **25(2)**, 505-541.