Graphical Markov Models

3 Lectures

Giovanni M. Marchetti

2024

1 Conditional independence
Graphs (DAG, UG, BG, RCG)
Markov Properties

Conditional Indep. tests
Examples

2 Undirected graphs.
Markov properties
separation, decomposable graphs
Fitting by ML, Examples

3 Bidirected graphs
Regression chain graphs
Markov properties
Parametrizations
Fitting graphs to data
Examples

Undirected graphs

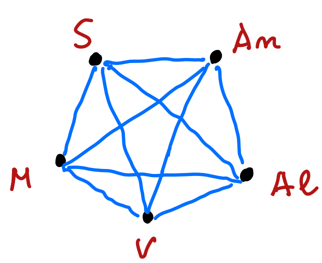
Role of variables

- DAG Responses ← explanatory variables
  Treatments
  intrinsic variables
- UG
  BG On equal standing ← Joint responses

Example

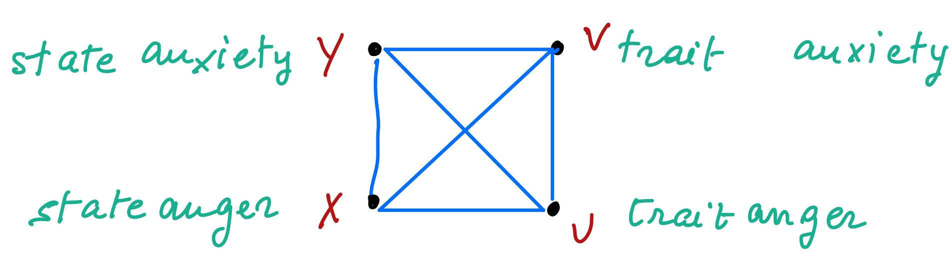
Mathematics marks (88 students)

5 variables : mechanics
vectors
algebra
analysis
statistics



Example for 684 fem. students

4 variables :



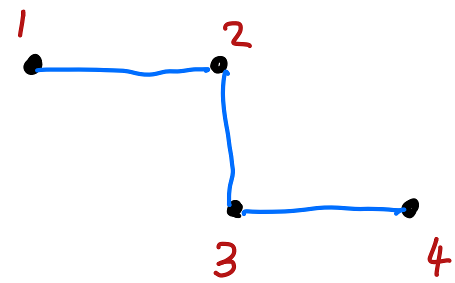
Trait : stable personality characteristic
State : behaviour in specific situations

Markov properties (1)

Pairwise MP

for any $i \neq j$

$X_i \perp\!\!\!\perp X_j \mid X_{V \setminus ij}$

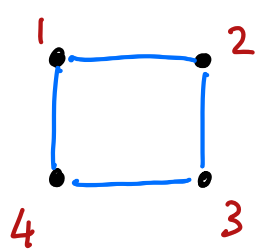


$1 \perp\!\!\!\perp 3 \mid 14$

$1 \perp\!\!\!\perp 4 \mid 23$

$2 \perp\!\!\!\perp 4 \mid 13$

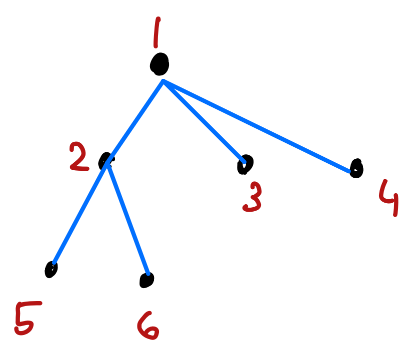
Examples



chordless 4-cycle

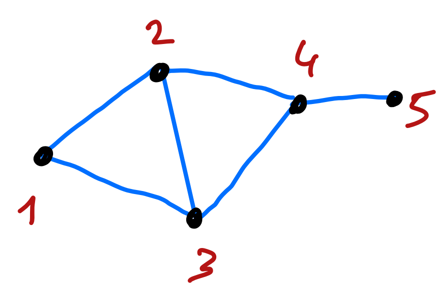
$1 \perp\!\!\!\perp 3 \mid 24$

$2 \perp\!\!\!\perp 4 \mid 13$



Tree

Exercise



Chordal graph

Exercise

list the CIs !

Some definitions

A path between two nodes i, j is a sequence of distinct nodes with consecutive edges

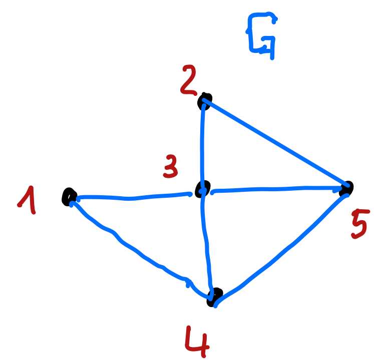
Example

A path from x_1 to x_5

(x_1, x_3, x_2, x_5)

another path

$(x_1, x_4, x_3, x_2, x_5)$



Separation

Let A, B, C 3 disjont subsets of V

A and B are separated by C in G if every path between A and B contains at least a node in C.

Exercise

$\{1, 4\}$ and $\{2\}$ are separated by $\{3, 5\}$

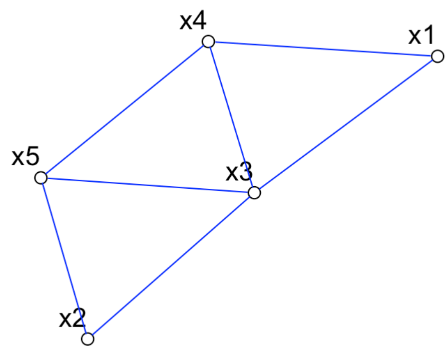
$\{1\}$ and $\{2\}$ are NOT separated by $\{3\}$

Compare separation and d-separation

Undirected graphs in R

Use package gRbase or ggm or igraph

```
G <- UG( ~ x1*x3*x4 + x2*x3*x5 + x3*x4*x5)

drawGraph(G, layout = layout_nicely)
```




```
g <- graph_from_adjacency_matrix(G, mode = "undirected")
separates("x5", "x1", "x4", g)

[1] FALSE

separates("x5", "x1", c("x4","x3"), g)

[1] TRUE
```

Markov properties (2)

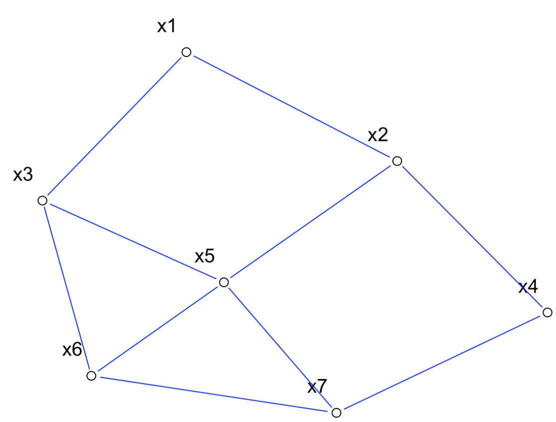
Global Markov property

For any disjoint $A, B, C \subset V$

If A and B are separated by C in G

then $\quad X_A \perp\!\!\!\perp X_B \mid X_C$

Example



Theorem If G is an UG satisfying the pairwise MP and C separates A and B then $X_A \perp\!\!\!\perp X_B \mid X_C$.

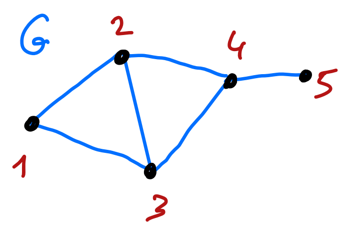
If Pairwise then Global

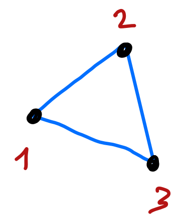
- Pairwise MP $\quad X_1 \perp\!\!\!\perp X_5 \mid X_2 X_3 X_4 X_6 X_7$
- Global MP $\quad X_1 \perp\!\!\!\perp X_5 \mid X_2 X_3$

more compact

Cliques of an UG

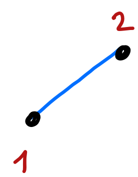
The cliques of G are all sets $A \subseteq V$ such that the subgraph G_A is maximally complete





G_A for $A = \{1, 2, 3\}$

A is a clique



G_B for $B = \{1, 2\}$

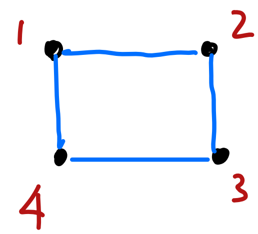
B is not a clique

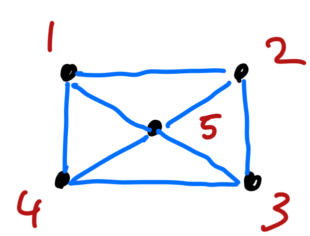
$\Rightarrow$ the set of all cliques $\mathcal{C} = \{123, 245, 45\}$

```
> get_cliques(g)
[[1]]
[1] "x1" "x2" "x3"

[[2]]
[1] "x2" "x4" "x5"

[[3]]
[1] "x1" "x2"
```

Exercise

Find the cliques





$\rightarrow$ R graphs_R

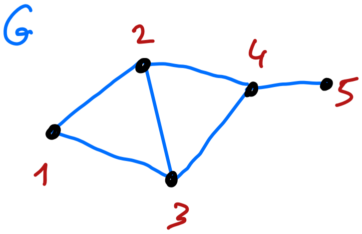
Factorization for UGs

If $X_V = (X_1, \dots, X_d)$ with density p and $\mathcal{G}$ is UG and exist functions ψ_C such that

$$p(x_1, \dots, x_d) = \prod_{C \in \mathcal{C}} \psi_C(\underline{x}_C)$$

we say that P_V factorizes according to G

Example a distribution that factorizes according to G



$$p(x_1, x_2, \dots x_5) = \psi_{123}(x_1, x_2, x_3)\, \psi_{234}(x_2, x_3, x_4)\, \psi_{45}(x_4, x_5)$$

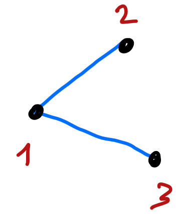
$$\boxed{\begin{array}{l} \textbf{Theorem} \quad \text{If } p(x_1, \dots, x_d) \text{ factorizes according to } G \text{ then} \\ \text{the global MP holds} \end{array}}$$

R graph_loglin

Example with binary variables

$X_1\ X_2\ X_3$ have levels

$i = 0,1 \quad j = 0,1 \quad k = 0,1$



cliques $\{1,2\}$ and $\{1,3\}$

log-linear parameters

2^3 contingency Table

	X_1	X_2	X_3
λ_0	0	0	0
λ_1	1	0	0
λ_2	0	1	0
λ_{12}	1	1	0
λ_3	0	0	1
λ_{13}	1	0	1
λ_{23}	0	1	1
λ_{123}	1	1	1

$\underline{\lambda}$ ← (table) →

8×1

variation independent

vector of joint probabilities

$\underset{\sim}{\pi} = (\pi_{ijk})$

8×1

with a constraint

$\sum_{i,j,k} \pi_{ijk} = 1$

Log-linear parameterization

$$\underset{\sim}{\pi} = \exp(M \underline{\lambda})$$

where
$$M = \underbrace{\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}}_{3 \text{ variables}}$$

8×8 matrix

When the loglinear parametrization defines a factorization according to G ?

CI constraints for $X_2 \perp\!\!\!\perp X_3 \mid X_1$

$$\boxed{\pi_{ijk} = \exp\left\{ \lambda_0 + \lambda_1 i + \lambda_2 j + \lambda_3 k + \lambda_{12} ij + \lambda_{13} ik + \lambda_{23} jk + \lambda_{123} ijk \right\}}$$

the constraints are

$$\lambda_{23} = 0 \quad \& \quad \lambda_{123} = 0$$

because

$$\boxed{\pi_{ijk} = \exp\left\{ \lambda_0 + \lambda_1 i + \lambda_2 j + \lambda_3 k + \lambda_{12} ij + \lambda_{13} ik + \cancel{\lambda_{23} jk} + \cancel{\lambda_{123} ijk} \right\}}$$

so that we get the factorization

$$\pi_{ijk} = \exp\left\{ \lambda_0 + \lambda_1 i + \lambda_2 j + \lambda_{12} ij \right\} \times \exp\left\{ \lambda_3 k + \lambda_{13} ik \right\}$$
$$= \psi_{12}(i,j)\, \psi_{13}(i,k)$$

that correspond to the cliques of G

Gaussian undirected graph models.

$$(X_1, \dots, X_d) \sim N_d(\underline{\mu}, \Sigma)$$

where K is the concentration matrix

$$K = \Sigma^{-1} = (K_{ij}) \text{ symmetric}$$

It can be shown that

$$(X_1, X_2 \mid X_3, \dots, X_d)$$

is a bivariate normal with

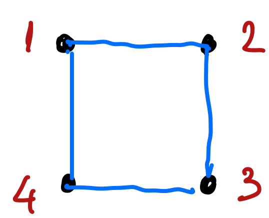
covariance matrix $\dfrac{1}{K_{11}K_{22} - K_{12}^2} \begin{pmatrix} K_{22} & -K_{12} \\ \cdot & K_{11} \end{pmatrix}$

The correlation coeff. from ↑ is the partial correlation coefficient $\rho_{12|rest}$

$$\boxed{K_{12} = 0 \iff \rho_{12|rest} = 0 \iff X_1 \perp\!\!\!\perp X_2 \mid X_{V \setminus \{1,2\}}}$$

Graphical Gaussian models are defined by setting specified concentrations K_{ij} to zero

$$\begin{pmatrix} K_{11} & K_{12} & 0 & K_{14} \\ \cdot & K_{22} & K_{23} & 0 \\ \cdot & \cdot & K_{33} & K_{34} \\ \cdot & \cdot & \cdot & K_{44} \end{pmatrix}$$



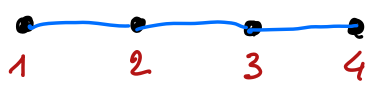
Factorization in the Gaussian

A multivariate normal density can be written as

$$p(\underline{x}) = \exp\left(\alpha + \underline{\beta}^T \underline{x} - \frac{1}{2}\underline{x}^T K \underline{x}\right)$$

where α is a normalization constant
$\underline{\beta}$ is $= K\underline{\mu}$

Assume that you have the graph



Exercise:
Q: $1 \perp\!\!\!\perp 3 \mid 2$?

$$K = \begin{pmatrix} K_{11} & K_{12} & 0 & 0 \\ \cdot & K_{22} & K_{23} & 0 \\ \cdot & \cdot & K_{33} & K_{34} \\ \cdot & \cdot & \cdot & K_{44} \end{pmatrix}$$

take $\underline{\mu} = 0$

This implies the factorization

$$\boxed{\psi_{12}(x_1, x_2)\, \psi_{23}(x_2, x_3)\, \psi_{34}(x_3, x_4)}$$

cliques $= \{1,2\}\{2,3\}\{3,4\}$

$$\boxed{\exp\left(\alpha - \frac{1}{2}K_{11}^2 x_1^2 - \frac{1}{2}K_{22}^2 x_2^2 - K_{12}x_1x_2\right)}$$

Thus, the global MP holds $\Rightarrow$ A: YES!

MLE for a Gaussian Graphical Model

- Assume that you know the graph G
- Assume that the data are a sample $(\underline{X}^1, \dots, \underline{X}^N)$ from $N_d(\mu; \Sigma)$

  $\underline{X}^i = (X_1^i, \dots X_d^i)^T$

log profile likelihood $\hat{\mu} = \bar{\underset{\sim}{x}}$

$$l(K, \hat{\underline{\mu}}) = \frac{N}{2}\{\log|K| - \mathrm{tr}(KS)\}$$

where $S = \frac{1}{N} \sum (\underline{x}^i - \bar{\underline{x}})(\underline{x}^i - \bar{\underset{\sim}{x}})^T$

Let G an undirected graph with cliques $\mathcal{C}$.

- In any GGM the MLE of $\underline{\mu}$ is $\bar{\underline{x}}$
- The MLE for K and S are found by solving $\hat{K}_{ij} = 0$ if $i \nsim j$

  and $\hat{\Sigma}_{cc} = S_{cc}$ if $c \in \mathcal{C}$

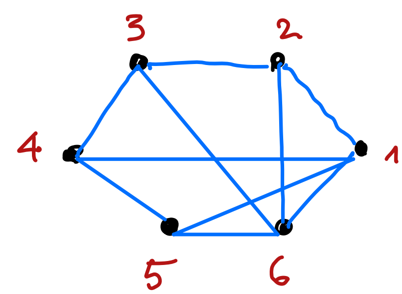
$\rightarrow R$

200 × 6 data matrix

Example (simdatUG.)

Partial correlation matrix

1	0.132	−0.046	0.347	0.205	−0.573
	1	−0.365	−0.018	−0.037	0.141
		1	0.074	−0.044	0.485
			1	−0.057	−0.005
				1	0.257
					1



Fitted concentration matrix

	1	2	3	4	5	6
1	*	−0.220	0	−0.380	−0.202	0.476
2		*	0.424	0	0	−1.188
3			*	−0.051	0	−0.287
4				*	0.066	0
5					*	−0.188
6						*

Sample covariance matrix

2.485	0.291	−1.630	0.697	0.003	−2.415
	0.488	−0.593	0.053	−0.023	−0.337
		4.040	−0.320	0.173	2.813
			1.009	−0.047	−0.647
				1.094	0.418
					4.569

Fitted covariance matrix

2.485	0.291	−1.536	0.697	0.003	−2.415
	0.488	−0.593	0.065	−0.009	−0.337
		4.040	−0.320	0.240	2.813
			1.009	−0.047	−0.645
				1.094	0.418
					4.569

Deviance = 0.037   4 iterations

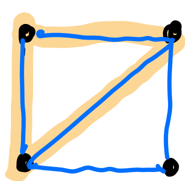
df = 5

Decomposable graphs

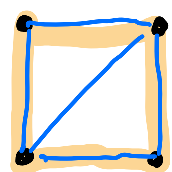
- play a special role in MLE
- Are important theoretically

An UG is decomposable (or chordal) iff

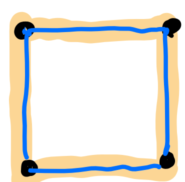
every undirected cycle of lenght 3 or more has a chord



a cycle of lenght 3



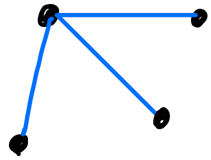
a cycle of lenght 4 with a chord

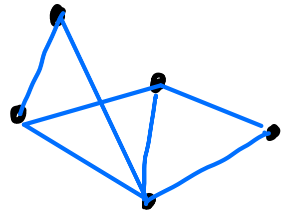


a chordless 4-cycle

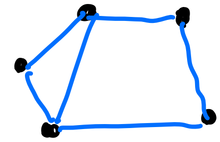
Examples

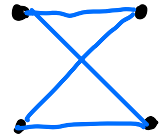
Decomposable

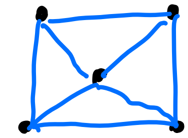




Not decomposable







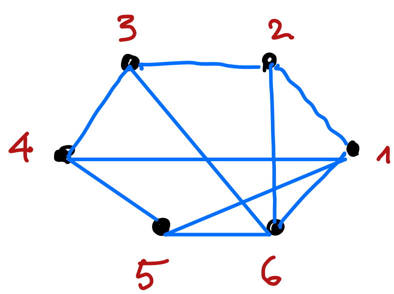
Explicit solutions

- MLE of an UG model requires in general an iterative method.
- If G is decomposable an explicit solution exist.

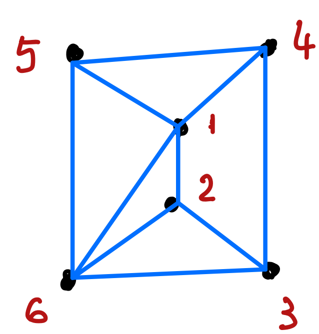
Thus it is relevant to check decomposability

Not so easy to identify by eye.

Example



is it decomposable?



NOT Decomposable

(required 4 iterations)

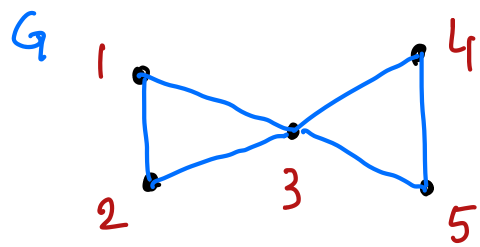
```
is.triangulated(G)

[1] FALSE
```

→ simdatUG

decomposable
triangulated
chordal

are synonym

Explicit solutions



This graph is decomposable

The cliques are $\{1,2,3\}$ and $\{3,4,5\}$

$c = \{3\}$ separates $a = \{1,2\}$ and $b = \{4,5\}$

Thus $\quad 1\,2 \perp\!\!\!\perp 4\,5 \mid 3$

and

$\Sigma_{ab|c} = 0 \quad$ i.e. $\Sigma_{ab} - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{cb} = 0$

MLE

$$\Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} & \Sigma_{ac} \\ & \Sigma_{bb} & \Sigma_{bc} \\ & & \Sigma_{cc} \end{bmatrix}$$ is estimated by $$S = \begin{bmatrix} S_{aa} & S_{ab} & S_{ac} \\ & S_{bb} & S_{bc} \\ & & S_{cc} \end{bmatrix}$$

the ML estimate of Σ constrained by G is

$$\begin{bmatrix} \Sigma_{aa} & S_{ac}S_{cc}^{-1}S_{cb} & S_{ac} \\ \cdot & S_{bb} & S_{bc} \\ \cdot & \cdot & S_{cc} \end{bmatrix}$$

Esercizio!

Test for conditional independence (1)

Deviance for a model $\mathcal{M}$

$$D = 2\left(\hat{\ell}_{sat} - \hat{\ell}_{\mathcal{M}}\right)$$

if $\mathcal{M}$ is true, $D \to \chi^2_k$

where k = difference in dim. of the saturated model and $\mathcal{M}$

Is equivalent to log likelihood ratio test for comparing a reduced model with the "saturated" model.

- For a discrete model (UG = log-linear model)

log-lik of the multinomial

$$\hat{\ell} = \sum_{\underline{x}} n(\underline{x}) \log \hat{p}(\underline{x})$$

counts

$$\text{deviance} = D = G^2 = 2\sum_{\underline{x}} n(\underline{x}) \log \frac{n(\underline{x})}{\hat{m}(\underline{x})}$$

fitted counts

→ R reimi's

Test of conditional independence (2)

- For a Gaussian model

$$\hat{\ell}(K, \hat{\mu}) = \frac{N}{2}\left(\log|K| - \mathrm{tr}(KS)\right)$$

$$D = N \log |S\hat{K}|$$

If the reduced model specifies a single conditional independence

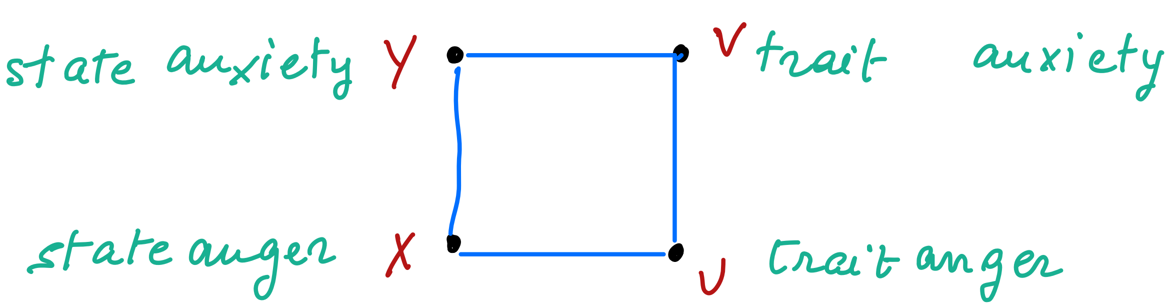
$$X_i \perp\!\!\!\perp X_j \mid X_{V \setminus ij}$$

$$D = -N\left(1 - \hat{\rho}^2_{ij|rest}\right)$$

called the edge exclusion deviance

Example

Anger Anxiety state/trait (684 fem. stud)



edge $Y \nsim U$ dev = 1.22 (1 df)

edge $X \nsim V$ dev = 0.33 (1 df)

deviance for the graph = 2.10 (2 df)

7 iterations

→ R anger.