

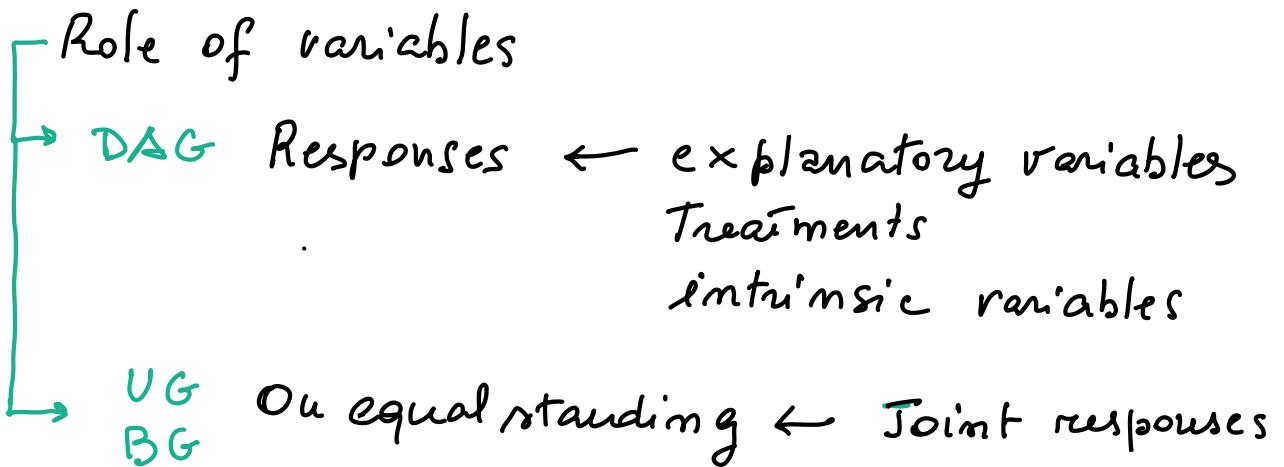
Graphical Markov Models

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

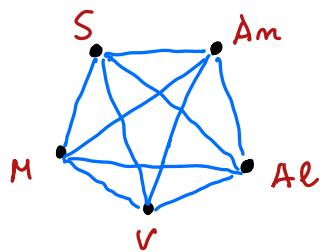
Undirected graphs



Example

Mathematics marks (88 students)

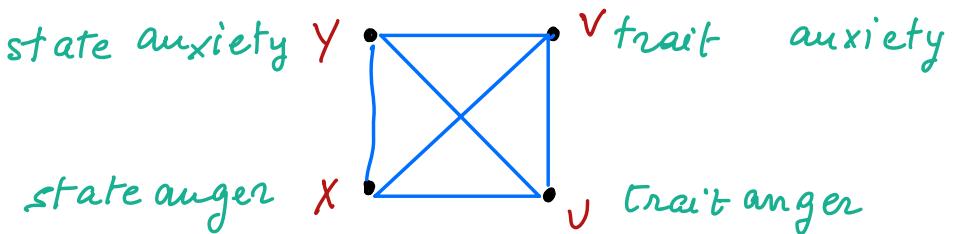
5 variables :



mechanics
vectors
algebra
analysis
statistics

Example for 684 fem. students

4 variables

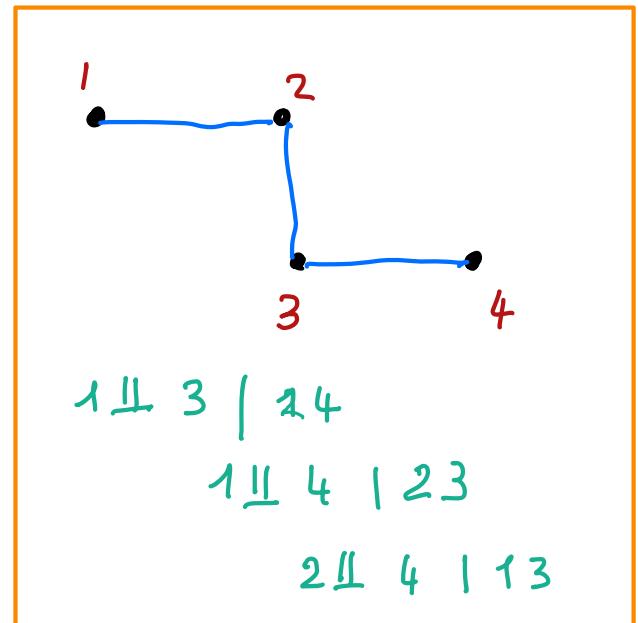


Trait : stable personality characteristic

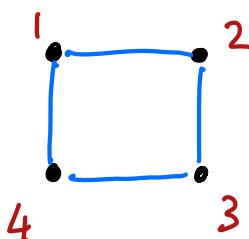
State : behaviour in specific situations

Markov properties (1)

Pairwise MP
 for any $i \neq j$
 $\rightarrow X_i \perp\!\!\!\perp X_j \mid X_{V \setminus ij}$



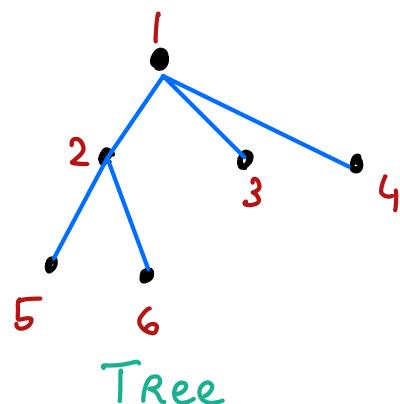
Examples



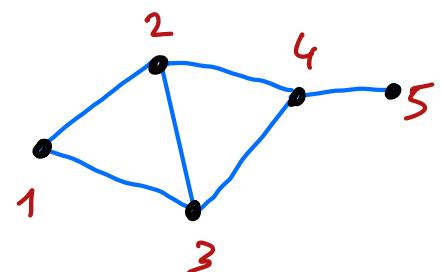
chordless 4-cycle

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

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



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

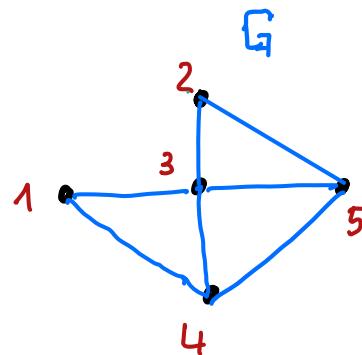
Example

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 disjoint 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 $\{3\}$

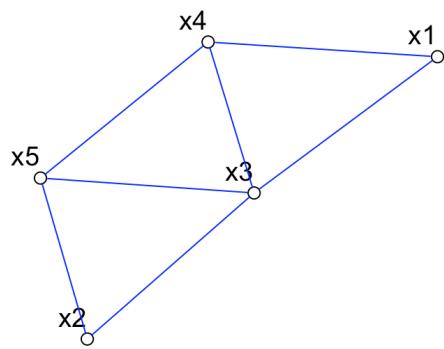
Compare separation and α -separation

graphs - R

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)

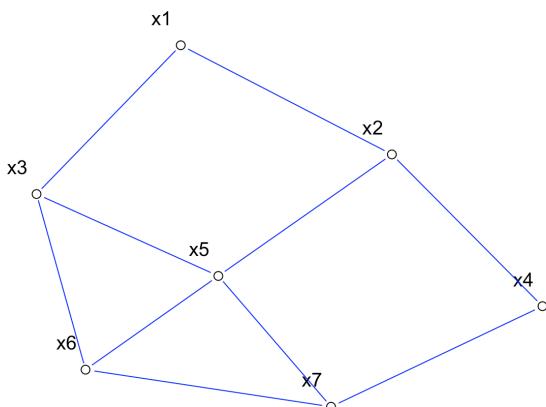
Global Markov property

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

If A and B are separated by C in G

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

Example



Theorem

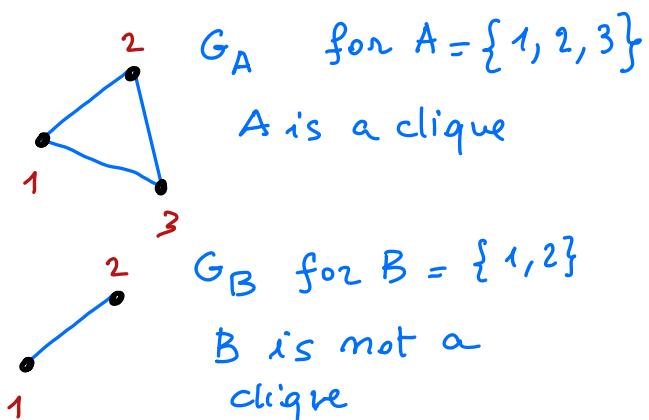
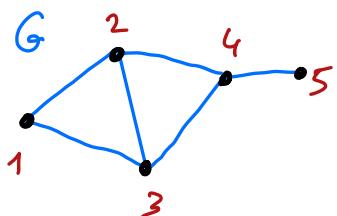
If G is an UG satisfying the pairwise MP and C separates A and B then $X_A \perp\!\!\!\perp X_B \mid X_C$.

→ If Pairwise Then Global

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

Cliques of an UG

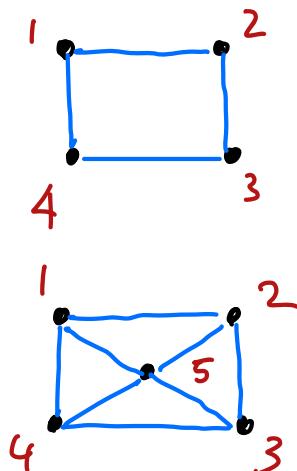
The cliques of G are all sets $A \subseteq V$ such that the subgraph G_A is maximally complete



⇒ the set of all cliques $\mathcal{P} = \{123, 243, 45\}$

```
> get_cliques(g)
[[1]]
[1] "x1" "x2" "x3"
[[2]]
[1] "x2" "x4" "x5"
[[3]]
[1] "x1" "x2"
```

Exercise
Find the cliques



→ R graphs_R

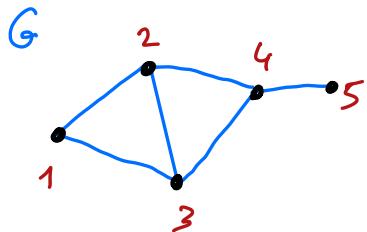
Factorization for UGs

If $X_v = (X_1, \dots, X_d)$ with density p and G is UG and exist functions such that $p(x_1, \dots, x_d) = \prod_{c \in C} \Psi_c(x_c)$

we say that P_v factorizes according to G

Example a distribution that factorizes

according to



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

Theorem

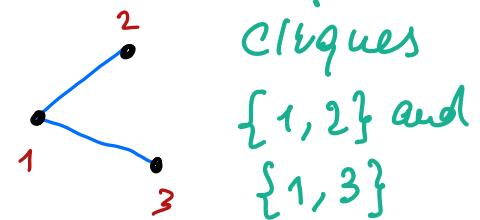
If $P(x_1, \dots, x_d)$ factorizes according to G then the global MP holds

R graph-loglin

Example with binary variables

X_1, X_2, X_3 have levels

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



log-linear parameters

$$\underline{\lambda} \quad \leftarrow \quad \begin{matrix} 8 \times 1 \\ \text{variation independent} \end{matrix}$$

	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

2^3 contingency Table

vector
of joint
probabilities

$$\underline{\pi} = (\pi_{ijk}) \quad 8 \times 1$$

with a constraint

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

Log-linear parameterization

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

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 parameterization defines a factorization according to G ?

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

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

The constraints are

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

Because

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

So that we get the factorization

$$\begin{aligned} \pi_{ijk} &= \exp \{ \lambda_0 + \lambda_1 i + \lambda_2 j + \lambda_{12} ij \} \times \\ &\quad \exp \{ \lambda_3 k + \lambda_{13} ik \} \\ &= \psi_{12}(i, j) \psi_{13}(i, k) \end{aligned}$$

that correspond to the cliques of G

Gaussian undirected graph models -

$$(x_1, \dots, x_d) \sim N_d(\mu, \Sigma)$$

where Σ is the concentration matrix

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

It can be shown that

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

is a bivariate normal with

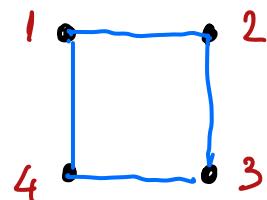
correlation matrix $\frac{1}{\sqrt{K_{11}K_{22}-K_{12}^2}} \begin{pmatrix} K_{22} - K_{12} \\ - & K_{11} \end{pmatrix}$

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

$$K_{12} = 0 \Leftrightarrow P_{12|rest} = 0 \Leftrightarrow X_1 \perp\!\!\!\perp X_2 | X_{V \setminus \{1,2\}}$$

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

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



Factorization in the Gaussian

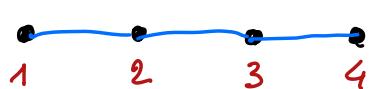
A multivariate normal density can be written as

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

where α is a normalization constant

$$\underline{\beta} \text{ is } = K \underline{\mu}$$

Assume that you have the graph



Exercise:

$$Q: \text{ Is } 1 \perp\!\!\!\perp 3 | 2 ?$$

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

This implies the factorization

$$\Psi_{12}(x_1, x_2) \quad \Psi_{23}(x_2, x_3) \quad \Psi_{34}(x_3, x_4)$$

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

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

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 $N_d(\mu; \Sigma)$
 $\hookrightarrow \underline{x}^i = (x_1^i, \dots, x_d^i)^T$

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

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

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

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

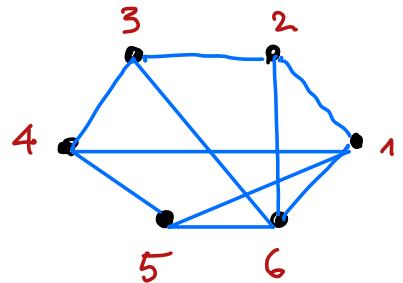
- In any GGM the MLE of μ is $\bar{\underline{x}}$
- The MLE for K and S are found by solving
 $\hat{K}_{ij} = 0 \quad \text{if } i \neq j$
and $\hat{\Sigma}_{cc} = S_{cc} \quad \text{if } c \in \mathcal{C}$

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	0	-1.188
3 *		-0.051	0	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

$$\text{Deviance} = 0.037$$

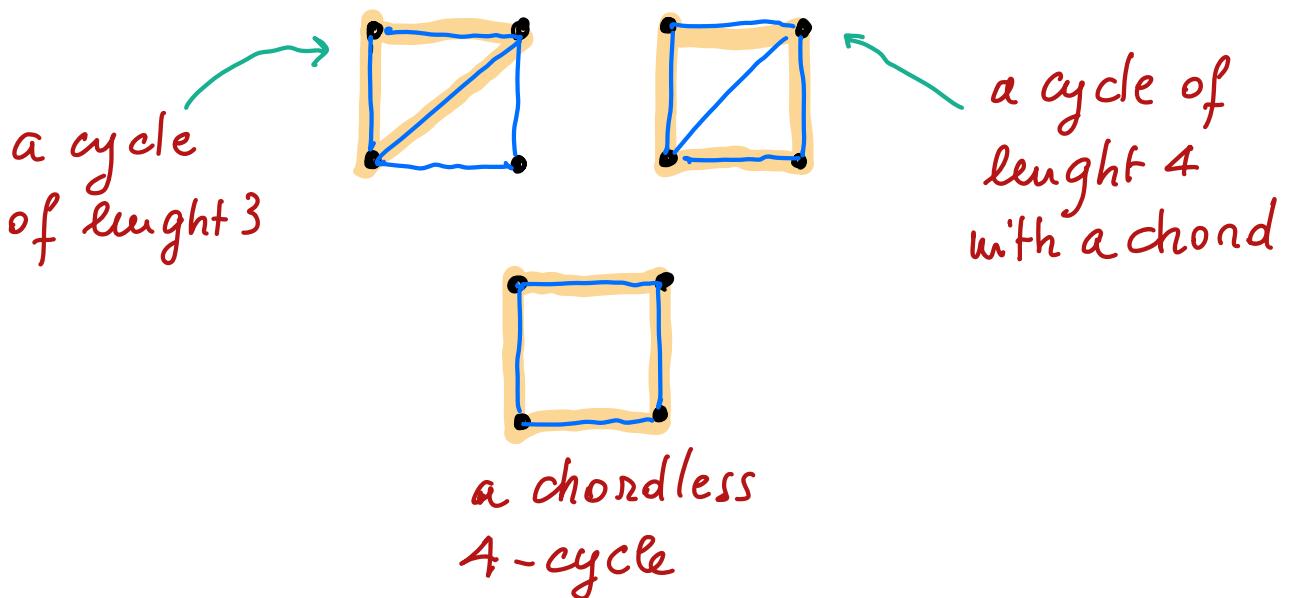
$$df = 5$$

4 iterations

Decomposable graphs

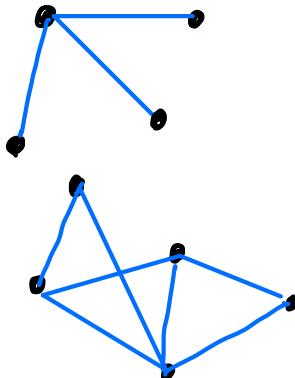
- play a special role in MLE
- Are important theoretically

An UG is decomposable (or chordal) iff
every undirected cycle
of length 3 or more has a chord

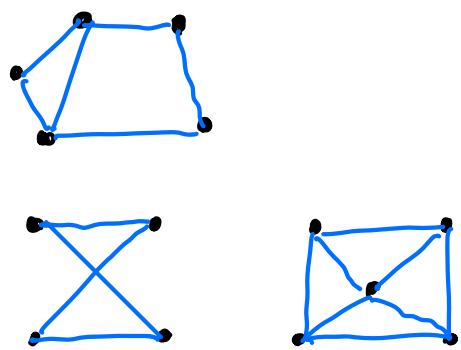


Examples

Decomposable



Not decomposable



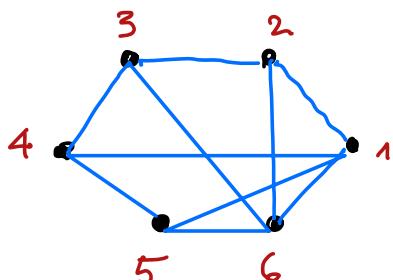
Explicit solutions

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

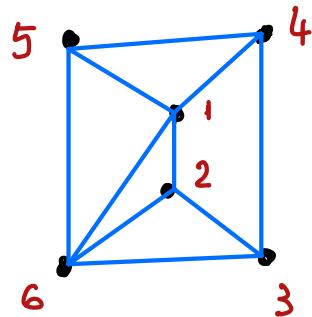
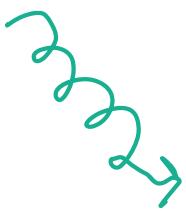
Thus it is relevant to check decomposability

Not so easy to identify by eye.

Example



is it decomposable?



NOT
Decomposable

(required 4 iterations)

```
is.triangulated(G)
```

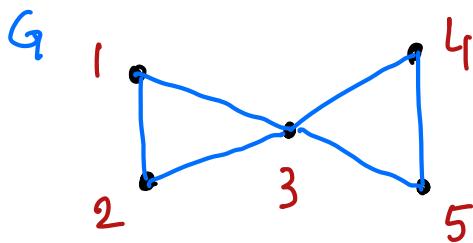
```
[1] FALSE
```

→ simdat UG

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 $1 \ 2 \perp\!\!\!\perp 4 \ 5 \mid 3$

and

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

MLE

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

the ML estimate of Σ constrained by G is

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

Esercizio!

Test for conditional independence (1)

Deviance for a model M

$$D = 2(\hat{L}_{\text{sat}} - \hat{L}_M)$$

if M is true, $D \rightarrow \chi^2_K$

where $K = \text{difference in dim. of}$
the saturated model and 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

$$\hat{L} = \sum_{\underline{x}} n(\underline{x}) \log \hat{P}(\underline{x})$$

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

↑ counts
↓ fitted counts

→ R reimis

Test of conditional independence (2)

- For a Gaussian model

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

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

If the reduced model specifies a single conditional independence

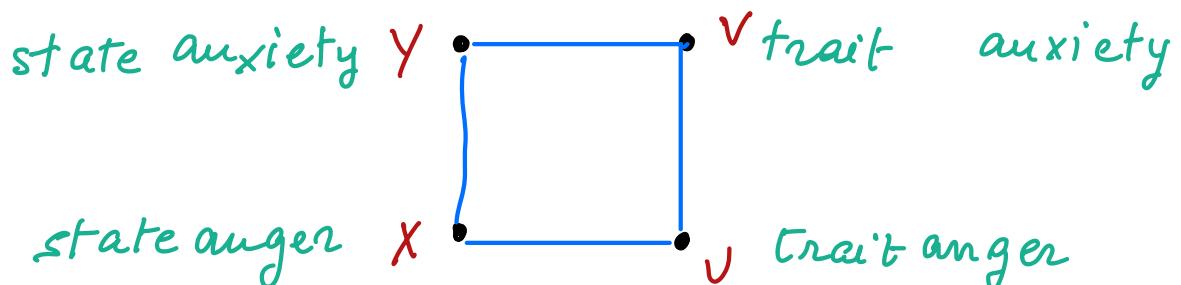
$$X_i \perp\!\!\!\perp X_j \mid X_{V \setminus ij}$$

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

called the edge exclusion deviance

Example

Anger Anxiety state/trait (684 fem. stud.)



edge $Y+U$ dev = 1.22 (1 df)

edge $X+V$ dev = 0.33 (1 df)

deviance for the graph = 2.10 (2 df)

7 iterations

$\rightarrow R$ anger.