GRAPHICAL MODELS
MODELE GRAFICZNE

Piotr GRACZYK(Université d'Angers)
graczyk@univ-angers.fr

Uniwersytet Wrocławski

MARZEC 2020

1. INTRODUCTION.

Simpson paradox

A university has 48 000 students
Half boys(24 000), half girls(24 000)

At the final exams: 10 000 boys and 14 000 girls fail
Feminist organizations threaten to close the university, girl students want to lynch the president!

However, the president of the university proves that the **results** R of the exams are **conditionally independent** of the **sex** S of a student, **knowing the department** D (notation $R \perp\!\!\!\perp S | D$)

3 departments

A(literature, history, languages),
B(law),
C(sciences)
16 000 students each

A	Succ.	Fail	**B** Succ.	Fail	**C** Succ.	Fail
Girls	3	9	4	4	3	1
Boys	1	3	4	4	9	3

Actually $\boxed{R \perp\!\!\!\perp S \mid D = d}$ for $d =$**A,B,C**

The notion of the conditional independence is necessary to understand the Simpson paradox.

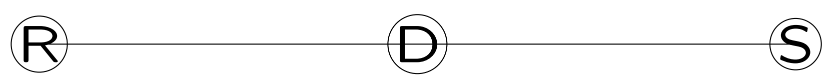
Graphical coding of conditional independence

Let $\mathcal{G}$ be a graph with vertices v_i. If 2 vertices v_i, v_j are connected by an (undirected) edge, we write $\boxed{v_i \sim v_j}$.

$R \perp\!\!\!\perp S | D$: $R \not\sim S$ **no edge between R and S**

Results depend on Department (knowing Sex): $R \sim D$

S depends on D (knowing Results): $S \sim D$



Remark. **D separates R from S**

(any path from R to S goes through D. Direct route from R to S impossible.)

GRAPHICAL MODELS IN STATISTICS

Consider an undirected graph $\mathcal{G} = (V, E)$ where:

- the set of nodes $V = \{1, \ldots, n\}$
- the set of edges $E \subset \{F \subset V |\ card(F) = 2\}$

Any set $\{i, j\} \in E$ will be called an edge. If $\{i, j\} \in E$, we write

$$i \sim j$$

Consider a system of random variables $X_1, \ldots, X_n$ on the same probability space $(\Omega, \mathcal{T}, P)$.

The information on **conditional independence** between the X_i's is schematized by an undirected graph $\mathcal{G} = (V, E)$ such that

$$X_l \perp\!\!\!\perp X_m \,|X_{V\setminus\{l,m\}} \iff l \not\sim m.$$

The graph $\mathcal{G}$ is called the *dependence graph* of the system of random variables $X_1, \ldots, X_n$.

Exercise 1. Alabama murderers. During 10 years, there were 4863 murders in Alabama. 2606 murderers were black, 60 were sentenced to death, others to prison. All the other 2257 murderers were white, 52 were sentenced to death, others to prison.
Question 1. *Does the proportion of death sentences depend on the race? Are Alabama judges racist?*

When the murdered victim was black, 2320 murderers were black, 12 were sentenced to death. 111 murderers were white, none was sentenced to death.
When the murdered victim was white, 286 murderers were black, 48 were sentenced to death. 2146 murderers were white, 52 were sentenced to death.
Question 2. *Knowing the victim race, does the proportion of death sentences depend on the race? Are Alabama judges racist?* Draw the dependence graph.

2. CONDITIONAL INDEPENDENCE.

In the preceding examples, we used intuitively the notion of CONDITIONAL INDEPENDENCE.

We will present it rigourously in this chapter.

Revisions on independence of r.v. X, Y (notation $X \perp Y$)

Def. $X \perp Y$ if $\forall A, B \in \mathcal{B}$

$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$

Equivalently, $\boxed{\mathcal{L}(X,Y) = \mathcal{L}(X) \otimes \mathcal{L}(Y)}$. In case with density, $\boxed{f_{X,Y}(x,y) = f_X(x) f_Y(y)}$.

Prop.1. In case with density, X, Y are independent iff the joint density factorizes $\boxed{f_{X,Y}(x,y) = g(x)h(y).}$

Proof. **Exercise 2.**

Equivalent def. of independence: X, Y independent if $\boxed{\mathcal{L}(X|Y) = \mathcal{L}(X)}$, i.e. $\forall A, B \in \mathcal{B}$, with $P(Y \in B) > 0$,

$$P(X \in A \mid Y \in B) = P(X \in A)$$

Exercise 3. Prove the equivalence of 2 definitions of independence.

Independence in a normal (Gaussian) vector: zeros in the covariance matrix Σ

Let $X = (X_1, \ldots, X_d)$ be a normal(Gaussian) vector on $\mathbb{R}^d$, with law $N(\xi, \Sigma)$.

Exercise 4. Let $i \neq j$.

1. What is the marginal law of (X_i, X_j)?
2. Show that $X_i \perp X_j \;\Leftrightarrow\; \Sigma_{ij} = 0$.

Covariance matrix Σ of a Gaussian vector contains:

- **marginal covariances**
- **information on independence:** $\boxed{X_i \perp X_j \;\Leftrightarrow\; \Sigma_{ij} = 0}$
- in practice, correlation $\rho_{ij} = \dfrac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}}\sqrt{\Sigma_{jj}}} \approx 0 \;\Rightarrow\; X_i \perp X_j$.

Remark. In statistics, the discrete case ($X(\Omega)$ finite or countable) may be also considered with **density**

$$f_X(x) = P(X = x)$$

w.r. to the counting measure $\nu = \sum_{x \in X(\Omega)} \delta_x$.

Law of X is $\mu_X = \sum_{x \in X(\Omega)} P(X = x)\delta_x$.

Functions on discrete prob. spaces are continuous.

So in the sequel we will write densities. The discrete case is included.

Conditional Independence of random variables.

Def. Conditional density

$$f_{U|V}(u|v) = f_{U|V=v}(u) = \frac{f_{U,V}(u,v)}{f_V(v)}.$$

The marginal density is supposed strictly positive:
$f_V(v) = \int f_{U,V}(u,v)du > 0$

Prop.2. The conditional density is a probability density.

Proof. $\int f_{U|V}(u|v)du = \dfrac{f_V(v)}{f_V(v)} = 1.$ $\square$

Exercise 5. Let $(X, Y) \sim N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$

Compute the conditional density $f_{X|Y=y}$.

Determine the conditional law $X|Y = y$.

Definition.

Let X, Y, Z **be three random variables on a probability space** $(\Omega, \mathcal{T}, P)$**.**

X **and** Y **are conditionally independent given** Z **if** $\boxed{\mathcal{L}(X|\ Y, Z) = \mathcal{L}(X|\ Z)}$**.**

A possible interpretation: **Knowing** Z **renders** Y **irrelevant for predicting** X**.**

Notation: $X \perp\!\!\!\perp Y\ |Z$**.**

(1) In case with density, $\boxed{X \perp\!\!\!\perp Y \;|Z \;\Leftrightarrow\; f_{X|Y,Z} = f_{X|Z}}$,
that is, $\forall x, y, z$

$$f_{X|\,(Y,Z)=(y,z)}(x,y) = \frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)} = \frac{f_{X,Z}(x,z)}{f_Z(z)} = f_{X|\,Z=z}(x)$$

(2) $X \perp\!\!\!\perp Y \;|Z$ iff $\boxed{\mathcal{L}(X,Y|\;Z) = \mathcal{L}(X|\;Z) \otimes \mathcal{L}(Y|\;Z)}$,
i.e. $\boxed{f_{X,Y|Z} = f_{X|Z} f_{Y|Z}}$.
(consequence of (1); CHECK IT!)

Prop.3. Factorization Property.
$X \perp\!\!\!\perp Y \;|Z$ **iff** $\boxed{f_{X,Y,Z}(x,y,z) = g(x,z)h(y,z)}$
Proof. **Exercise 6**.

Conditional Independence in a Gaussian vector: zeros in the precision matrix $K = \Sigma^{-1}$

Let $X = (X_1, \ldots, X_d)$ be a Gaussian vector on $\mathbb{R}^d$, with law $N(\xi, \Sigma)$ and invertible Σ.

The matrix $K = \Sigma^{-1}$ is called the precision(concentration) matrix of X.

This is the precision matrix $K = (\kappa_{ij})_{i,j \le d}$ *that appears in the density* $f(x) = (2\pi)^{-d/2} (\det K)^{1/2} e^{-(x-\xi)^T K (x-\xi)/2}$

Example.

$X \sim N(0, \Sigma)$ **in dim 3.** $K = (\kappa_{ij})_{i,j \le 3} = \Sigma^{-1}$

$$f(x_1, x_2, x_3) = \frac{(\det K)^{1/2}}{(2\pi)^{3/2}} \times$$
$$e^{-(\kappa_{11}x_1^2 + \kappa_{22}x_2^2 + \kappa_{33}x_3^2 + 2\kappa_{12}x_1x_2 + 2\kappa_{13}x_1x_3 + 2\kappa_{23}x_2x_3)/2}$$

Suppose $\boxed{X_1 \perp\!\!\!\perp X_2 | X_3}$ $\Leftrightarrow$ $f(x_1, x_2, x_3) = g(x_1, x_3) h(x_2, x_3) = C \times$
$$e^{-(\kappa_{11}x_1^2 + \kappa_{22}x_2^2 + \kappa_{33}x_3^2 + 2\kappa_{12}x_1x_2 + 2\kappa_{13}x_1x_3 + 2\kappa_{23}x_2x_3)/2}$$

Obligatorily $2\kappa_{12}x_1x_2 = 0 \;\Leftrightarrow\; \boxed{\kappa_{12} = 0}$.

Prop.4. Let X be a Gaussian vector in $\mathbb{R}^d$. Denote $V = \{1, \ldots, d\}$ the index set. Let $l, m \in V$ and $l \neq m$.

Then the marginals X_l, X_m are conditionally independent w.r. to all the other variables $X_{V \setminus \{l,m\}}$

$$X_l \perp\!\!\!\perp X_m \; | X_{V \setminus \{l,m\}} \iff \kappa_{lm} = 0$$

i.e. the lm-term of the precision matrix is equal 0.

Proof. The normal density $f(x)$ is a product of $e^{-\kappa_{ij} x_i x_j}$ and $e^{-\kappa_{ii} x_i^2 / 2}$. By Factorization Property in Prop. 3, $X_l \perp X_m \; | X_{V \setminus \{l,m\}}$ if and only if no "mixed" factor $e^{-\kappa_{lm} x_l x_m}$ appears (we apply ln and use the fact that $x_l x_m \neq a(x_l) + b(x_m)$, clear by taking $x_l = 0$ and next $x_m = 0$). MAKE SURE YOU CAN PROVE IT! $\square$

3. GAUSSIAN GRAPHICAL MODELS

Let $V = \{1, \ldots, n\}$ and let $\mathcal{G} = (V, E)$ be an undirected graph. Let $\mathcal{S}(\mathcal{G}) = \{Z \in Sym(n \times n)|\ i \not\sim j \ \Rightarrow\ Z_{ij} = 0\}$, the space of symmetric matrices with obligatory zero terms $Z_{ij} = 0$ for $i \not\sim j$.

Definition. The **GAUSSIAN GRAPHICAL MODEL governed by the graph** $\mathcal{G}$ is the set of all random Gaussian vectors $X = (X_v)_{v \in V} \sim N(\xi, \Sigma)$, with precision matrix $K = \Sigma^{-1} \in \mathcal{S}(\mathcal{G})$.

By Prop. 4, the **GAUSSIAN GRAPHICAL MODEL governed by the graph** $\mathcal{G}$ means the **constraint of conditional independence** $X_l \perp\!\!\!\perp X_m \,| X_{V \setminus \{l,m\}}$ for all graph nodes $l \not\sim m$ non-connected by an edge of $\mathcal{G}$.

Example 1. Let $\mathcal{G} : \overset{1}{\bullet} - \overset{2}{\bullet} - \cdots - \overset{n}{\bullet}$ be the graph corresponding to the Gaussian graphical model of **nearest neighbour interaction** in a Gaussian character. The graph $\mathcal{G}$ is called A_n.

In the Gaussian character $(X_1, X_2, \ldots, X_n)$, **non-neighbours** X_i, X_j**,** $|i - j| > 1$ **are conditionally independent** with respect to other variables. Only neighbours interact.
$K \in \mathcal{S}(\mathcal{G})$ is equivalent to $K \in Sym^+(n \times n)$ **tridiagonal**.

Example 2. The **complete graph** $\mathcal{G}$ (i.e. $\mathcal{G}$ containing all possible edges) defines Gaussian graphical model containing all Gaussian vectors supported by $\mathbb{R}^n$, with no constraint. Such model is called **saturated**.

Example 3. The **totally disconnected graph** $\mathcal{G}$ has no edges. What is the corresponding Gaussian graphical model?

MORE on CONDITIONAL LAWS IN GAUSSIAN CASE

Let X be a Gaussian $N(\xi, \Sigma)$ vector in $\mathbb{R}^d$.
Partition $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ into $X_1 \in \mathbb{R}^r$ and $X_2 \in \mathbb{R}^s$, with $r + s = d$, and, similarly, $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$.

Partition mean vector, covariance and precision matrix accordingly in bloc $\begin{pmatrix} r \times r & r \times s \\ s \times r & s \times s \end{pmatrix}$ matrices as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \; K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$$

with $\Sigma_{21} = \Sigma_{12}^T$ and $K_{21} = K_{12}^T$. Suppose Σ invertible.

Prop.5. The conditional law $X_1 | \, X_2 = x_2 \sim N_r(\xi_{1|2}, \Sigma_{1|2})$ where

$\xi_{1|2} = \xi_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \xi_2)$ and $\Sigma_{1|2} = K_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Prop. 5 implies

Cor.6 (i) The precision matrix of $X_1|X_2$ equals K_{11}.

(ii) $X_l \perp\!\!\!\perp X_m \; | X_{V \setminus \{l,m\}} \iff \kappa_{lm} = 0$

(iii) X_1 and X_2 are independent if and only if the bloc $\Sigma_{12} = 0$ if and only if the bloc $K_{12} = 0$.

Exercise 7. Prove Corollary 6.

We will use the symbol $\propto$ "is proportional to", e.g. if ϕ is the $N(0,1)$ density, then $\phi(x) \propto e^{-x^2/2}$.

Proof of Prop.5.

Note that the bloc multiplication gives the real number $(x-\xi)^T K(x-\xi) = (x_1-\xi_1)^T K_{11}(x_1-\xi_1) + 2(x_1-\xi_1)^T K_{12}(x_2-\xi_2) + (x_2-\xi_2)^T K_{22}(x_2-\xi_2)$.

In the following, we point out the argument x_1 of the density $f_{X_1|X_2}$ and we push x_2 to the constant factor:

$f_{X_1|X_2}(x_1|x_2) \propto f_{N(\xi,\Sigma)}(x_1,x_2) \propto$
$\exp\{-(x_1-\xi_1)^T K_{11}(x_1-\xi_1)/2-(x_1-\xi_1)^T K_{12}(x_2-\xi_2)\} \propto$
$\exp\{x_1^T[K_{11}\xi_1 - K_{12}(x_2-\xi_2)] - x_1^T K_{11}x_1/2\} =$
$\exp\{x_1^T K_{11}[\xi_1 - K_{11}^{-1}K_{12}(x_2-\xi_2)] - x_1^T K_{11}x_1/2\} \propto$
$\exp\{-(x_1-m)^T K_{11}(x_1-m)/2\}$

with $m = \xi_1 - K_{11}^{-1}K_{12}(x_2-\xi_2)$. Next we use formulas $K_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and $K_{11}^{-1}K_{12} = -\Sigma_{12}\Sigma_{22}^{-1}$ following from the Lemma. $\square$

Lemma 7. Let the invertible matrix $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.
Then
$$K = \Sigma^{-1} = \begin{pmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1} & S \\ S^T & (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{pmatrix}$$
where $S = -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} = -(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\Sigma_{12}\Sigma_{22}^{-1}$.

Proof. One has the decomposition (easy to check) $\Sigma =$
$$\begin{pmatrix} I_r & 0 \\ \Sigma_{21}\Sigma_{11}^{-1} & I_s \end{pmatrix} \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix} \begin{pmatrix} I_r & \Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I_s \end{pmatrix}.$$
Use this and $\begin{pmatrix} I & A \\ 0 & I \end{pmatrix}^{-1} = \begin{pmatrix} I & -A \\ 0 & I \end{pmatrix}$ to express Σ^{-1} as a product of 3 matrices. From this, get K_{22} and S. By symmetry of indices 1 and 2, get K_{11}. DO IT YOURSELF! □

Precision matrix $K = \Sigma^{-1}$ of a Gaussian vector contains:

- **conditional precision matrices ($K_{X_1|X_2} = K_{11}$)**

- **information on conditional independence:**

$$\boxed{X_l \perp\!\!\!\perp X_m \,|X_{V\setminus\{l,m\}} \iff \kappa_{lm} = 0} \qquad (l \neq m)$$

- **In practice, we use conditional correlation** $(l \neq m)$

$$\rho_{lm|V\setminus\{l,m\}} = \frac{Cov(X_l, X_m | X_{V\setminus\{l,m\}})}{Var(X_l | X_{V\setminus\{l,m\}})^{\frac{1}{2}} Var(X_m | X_{V\setminus\{l,m\}})^{\frac{1}{2}}} = -\tilde{\kappa}_{lm}$$

where $\tilde{\kappa}_{lm} \overset{\text{df}}{=} \frac{\kappa_{lm}}{\sqrt{\kappa_{ll}}\sqrt{\kappa_{mm}}}$ **is an element of the so-called** *scaled precision matrix* $\tilde{K}$**.**

The formula $\boxed{\rho_{lm|V\setminus\{l,m\}} = -\tilde{\kappa}_{lm}}$ is justified by:

$$K_{lm|V\setminus\{l,m\}} = \begin{pmatrix} \kappa_{ll} & \kappa_{lm} \\ \kappa_{lm} & \kappa_{mm} \end{pmatrix}$$

$$\Sigma_{lm|V\setminus\{l,m\}} = K^{-1}_{lm|V\setminus\{l,m\}} = \frac{1}{k} \begin{pmatrix} \kappa_{mm} & -\kappa_{lm} \\ -\kappa_{lm} & \kappa_{ll} \end{pmatrix},$$

where $k = \det K_{lm|V\setminus\{l,m\}}$.

END THE PROOF YOURSELF

If $\rho_{lm|V\setminus\{l,m\}} = -\tilde{\kappa}_{lm} \approx 0 \;\Rightarrow$

we naturally conjecture that

$$X_l \perp\!\!\!\perp X_m \;|X_{V\setminus\{l,m\}}$$

To be serious:

we should prove it

we should test it statistically

Exercise 8. Let $X \sim N_3(0, \Sigma)$ with covariance matrix $\Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$.

0. Are there X_i independent?

1. Find the precision matrix K. Are there X_i conditionally independent? Draw the dependence graph.

2. Find the marginal law of (X_1, X_2).

3. Find the conditional law of $(X_1, X_2)|X_3 = x_3$.

4. Compute the scaled precision matrix $\tilde{K}$.

5. Find the conditional correlations $\rho_{X_1,X_2|X_3=x}$, $\rho_{X_1,X_3|X_2=x}$ and $\rho_{X_2,X_3|X_1=x}$.

Example. Marks of 88 students in 5 exams: $N_5(\xi, \Sigma)$

Table 1.1.1: Marks in five mathematics exams for 88 students. From Mardia, Kent and Bibby (1979). [WHITTAKER]

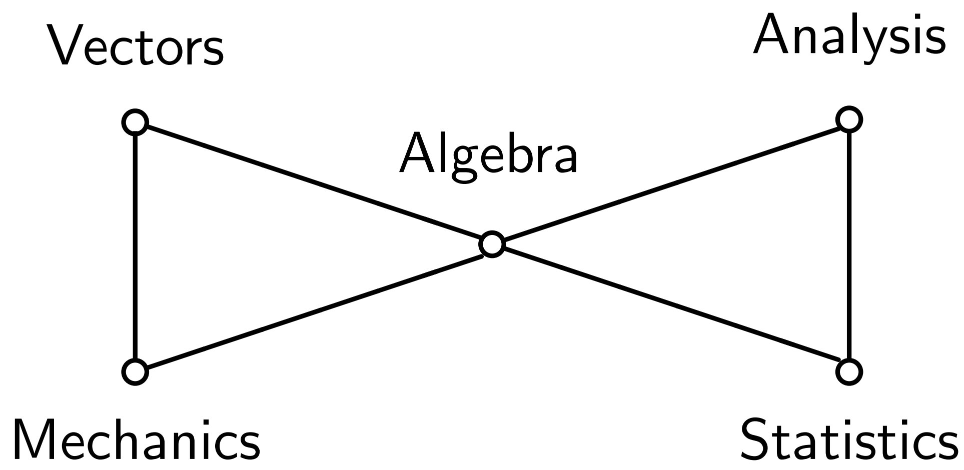
me	ve	al	an	st	me	ve	al	an	st	me	ve	al	an	st
77	82	67	67	81	30	69	50	52	45	62	44	36	22	42
63	78	80	70	81	46	49	53	59	37	48	38	41	44	33
75	73	71	66	81	40	27	54	61	61	34	42	50	47	29
55	72	63	70	68	31	42	48	54	68	18	51	40	56	30
63	63	65	70	63	36	59	51	45	51	35	36	46	48	29
53	61	72	64	73	56	40	56	54	35	59	53	37	22	19
51	67	65	65	68	46	56	57	49	32	41	41	43	30	33
59	70	68	62	56	45	42	55	56	40	31	52	37	27	40
62	60	58	62	70	42	60	54	49	33	17	51	52	35	31
64	72	60	62	45	40	63	53	54	25	34	30	50	47	36
52	64	60	63	54	23	55	59	53	44	46	40	47	29	17
55	67	59	62	44	48	48	49	51	37	10	46	36	47	39
50	50	64	55	63	41	63	49	46	34	46	37	45	15	30
65	63	58	56	37	46	52	53	41	40	30	34	43	46	18
31	55	60	57	73	46	61	46	38	41	13	51	50	25	31
60	64	56	54	40	40	57	51	52	31	49	50	38	23	9
44	69	53	53	53	49	49	45	48	39	18	32	31	45	40
42	69	61	55	45	22	58	53	56	41	8	42	48	26	40
62	46	61	57	45	35	60	47	54	33	23	38	36	48	15
31	49	62	63	62	48	56	49	42	32	30	24	43	33	25
44	61	52	62	46	31	57	50	54	34	3	9	51	47	40
49	41	61	49	64	17	53	57	43	51	7	51	43	17	22
12	58	61	63	67	49	57	47	39	26	15	40	43	23	18
49	53	49	62	47	59	50	47	15	46	15	38	39	28	17
54	49	56	47	53	37	56	49	28	45	5	30	44	36	18
54	53	46	59	44	40	43	48	21	61	12	30	32	35	21
44	56	55	61	36	35	35	41	51	50	5	26	15	20	20
18	44	50	57	81	38	44	54	47	24	0	40	21	9	14
46	52	65	50	35	43	43	38	34	49					
32	45	49	57	64	39	46	46	32	43					

Mathematics marks

Examination marks of 88 students in 5 different mathematical subjects. The empirical concentrations (on or above diagonal) and partial correlations (below diagonal) are

	Mechanics	Vectors	Algebra	Analysis	Statistics
Mechanics	5.24	−2.44	−2.74	0.01	−0.14
Vectors	0.33	10.43	−4.71	−0.79	−0.17
Algebra	0.23	0.28	26.95	−7.05	−4.70
Analysis	−0.00	0.08	0.43	9.88	−2.02
Statistics	0.02	0.02	0.36	0.25	6.45

Graphical model for mathmarks



This analysis is from Whittaker (1990).
We have An, Stats $\perp\!\!\!\perp$ Mech, Vec | Alg.

Gaussian Graphical Models

Steffen Lauritzen
University of Oxford

CIMPA Summerschool, Hammamet 2011, Tunisia

September 8, 2011