

SC6/SM9 Graphical Models

Michaelmas Term, 2017

Robin Evans

evans@stats.ox.ac.uk
Department of Statistics
University of Oxford

November 29, 2017

1

Course Information

There will be four problem sheets and four associated classes.

Part C students, your classes are weeks 3, 5, 7 and HT1. Sign up online for one of the two sessions.

Hand in work by Tuesday, 5pm.

MSc students, classes are at 2pm on Wednesdays, weeks 3, 5, 7, and Thursday week 8 in here (LG.01).

3

Course Website

The class site is at

<http://www.stats.ox.ac.uk/~evans/gms/>

You'll find

- lecture notes;
- slides;
- problem sheets;
- data sets.

2

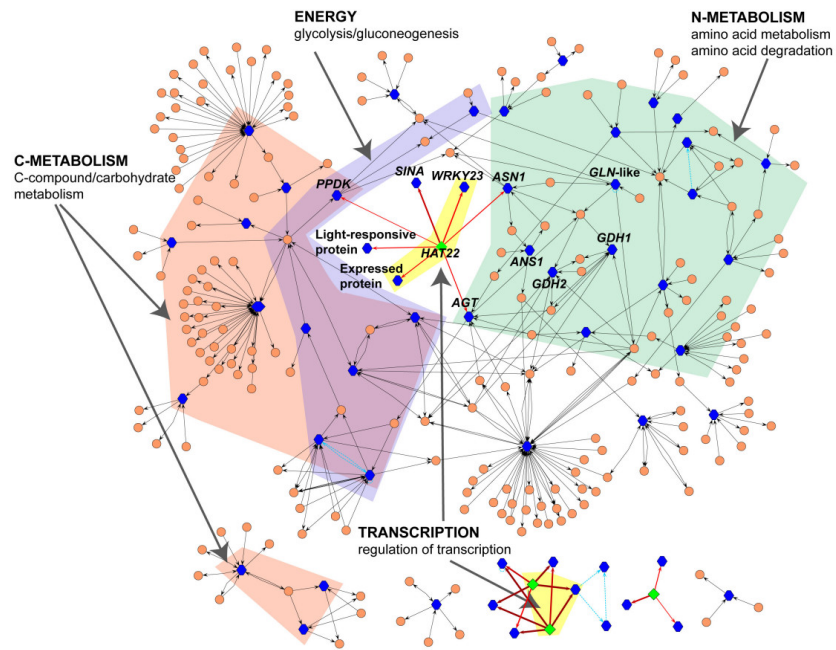
Books

These books might be useful.

- Lauritzen (1996). *Graphical Models*, OUP.
- Wainwright and Jordan (2008). *Graphical Models, Exponential Families, and Variational Inference*. (Available online).
- Pearl (2009). *Causality*, (3rd edition), Cambridge.
- Koller and Friedman (2009), *Probabilistic Graphical Models: Principles and Techniques*, MIT Press.

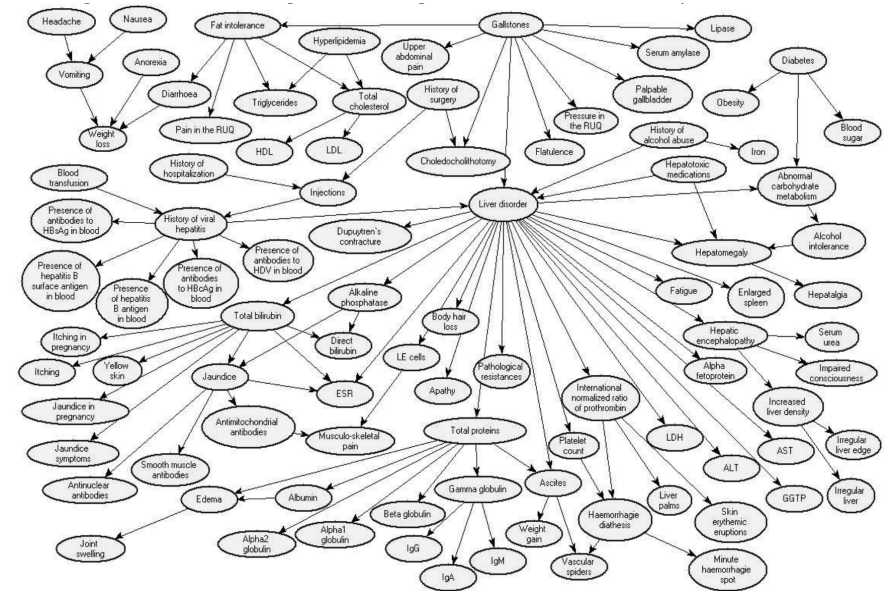
4

Gene Regulatory Networks



5

Medical Diagnosis



6

Main Issues

There are two main problems with large data sets that we will consider in this course:

- statistical;
 - we need to predict outcomes from scenarios that have never been observed (i.e., we need a model).
- computational:
 - we can't store probabilities for all combinations of variables;
 - even if we could, we can't sum/integrate them to find a marginal or conditional probability:

$$P(X = x) = \sum_{\mathbf{y}} P(X = x, \mathbf{Y} = \mathbf{y}).$$

Our solution will be to impose nonparametric structure, in the form of conditional independences.

7

Conditional Independence

8

Marginal Table

Victim's Race	Death Penalty?	Defendant's Race	
		White	Black
White	Yes	53	11
	No	414	37
Black	Yes	0	4
	No	16	139

If we sum out the Victim's race...

Death Penalty?	Defendant's Race	
	White	Black
Yes	53	15
No	430	176

13

Contingency Tables

The death penalty data is on the class website.

```
> deathpen <- read.table("deathpen.txt", header=TRUE)
> deathpen

  DeathPen Defendant Victim freq
1      Yes      White  White   53
2      No      White  White  414
3      Yes      Black  White   11
4      No      Black  White   37
5      Yes      White  Black    0
6      No      White  Black   16
7      Yes      Black  Black    4
8      No      Black  Black  139
```

14

Contingency Tables

We can fit models on it in R:

```
> summary(glm(freq ~ Victim*Defendant + Victim*DeathPen,
+             family=poisson, data=deathpen))
```

Coefficients:

	Estimate	Std. Error
(Intercept)	4.93737	0.08459
VictimWhite	-1.19886	0.16812
DefendantWhite	-2.19026	0.26362
DeathPenYes	-3.65713	0.50641
VictimWhite:DefendantWhite	4.46538	0.30408
VictimWhite:DeathPenYes	1.70455	0.52373

Residual deviance: 5.394 on 2 degrees of freedom

(So $p \approx 0.07$ in hypothesis test of model fit.)

15

Contingency Tables

If we fit the marginal table over the races of Victim and Defendant, the parameters involving 'Defendant' are the same.

```
> summary(glm(freq ~ Victim*Defendant,
+             family=poisson, data=deathpen))
```

Coefficients:

	Estimate	Std. Error
(Intercept)	4.26970	0.08362
VictimWhite	-1.09164	0.16681
DefendantWhite	-2.19026	0.26360
VictimWhite:DefendantWhite	4.46538	0.30407

16

Undirected Graphical Models

17

Multivariate Data

```
> library(ggm)
> data(marks)
> dim(marks)

[1] 88  5

> head(marks, 8)

  mechanics vectors algebra analysis statistics
1         77      82     67         67         81
2         63      78     80         70         81
3         75      73     71         66         81
4         55      72     63         70         68
5         63      63     65         70         63
6         53      61     72         64         73
7         51      67     65         65         68
8         59      70     68         62         56
```

18

Multivariate Data

```
> sapply(marks, mean)

 mechanics    vectors    algebra    analysis    statistics
    39.0         50.6         50.6         46.7         42.3

> cor(marks)

           mechanics vectors algebra analysis statistics
mechanics  1.000    0.553   0.546   0.410    0.389
vectors    0.553    1.000   0.610   0.485    0.436
algebra    0.546    0.610   1.000   0.711    0.665
analysis   0.410    0.485   0.711   1.000    0.607
statistics 0.389    0.436   0.665   0.607    1.000
```

19

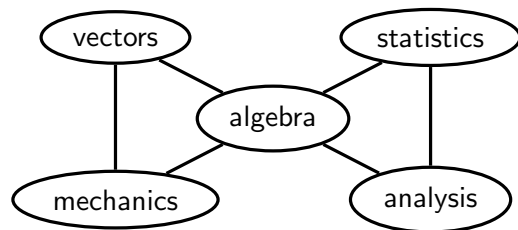
Multivariate Data

```
> conc <- solve(cov(marks)) # concentration matrix
> round(1000*conc, 2)

           mechanics vectors algebra analysis statistics
mechanics    5.24   -2.43   -2.72    0.01   -0.15
vectors     -2.43   10.42   -4.72   -0.79   -0.16
algebra     -2.72   -4.72   26.94   -7.05   -4.70
analysis     0.01   -0.79   -7.05    9.88   -2.02
statistics  -0.15  -0.16   -4.70   -2.02    6.45
```

20

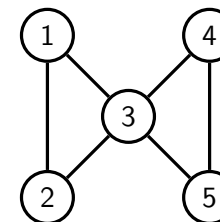
Undirected Graphs



	mech	vecs	alg	anlys	stats
mechanics	5.24	-2.43	-2.72	0.01	-0.15
vectors	-2.43	10.42	-4.72	-0.79	-0.16
algebra	-2.72	-4.72	26.94	-7.05	-4.70
analysis	0.01	-0.79	-7.05	9.88	-2.02
statistics	-0.15	-0.16	-4.70	-2.02	6.45

21

Undirected Graphs

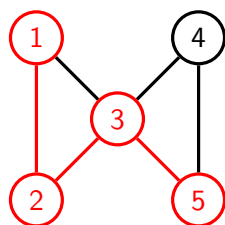


$$V = \{1, 2, 3, 4, 5\}$$

$$E = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{3, 4\}, \{3, 5\}, \{4, 5\}\}.$$

22

Paths



Paths:

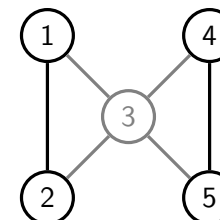
$$\pi_1 : 1 - 2 - 3 - 5$$

$$\pi_2 : 3$$

Note that paths may consist of one vertex and no edges.

23

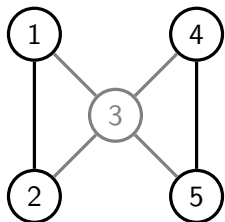
Induced Subgraph



The **induced subgraph** $\mathcal{G}_{\{1,2,4,5\}}$ drops any edges that involve $\{3\}$.

24

Separation

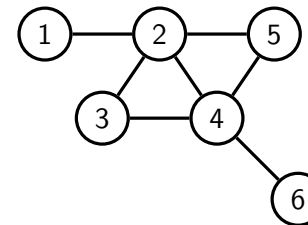


All paths between $\{1, 2\}$ and $\{5\}$ pass through $\{3\}$.

Hence $\{1, 2\}$ and $\{5\}$ are **separated** by $\{3\}$.

25

Cliques and Running Intersection



Cliques:

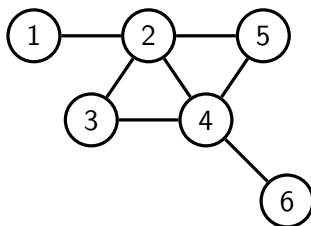
$\{1, 2\}$ $\{2, 3, 4\}$ $\{2, 4, 5\}$ $\{4, 6\}$.

Separator sets:

\emptyset $\{2\}$ $\{2, 4\}$ $\{4\}$.

26

Cliques and Running Intersection



A different ordering of the cliques:

$\{2, 3, 4\}$ $\{2, 4, 5\}$ $\{4, 6\}$ $\{1, 2\}$.

Separator sets:

\emptyset $\{2, 4\}$ $\{4\}$ $\{2\}$.

Any ordering works in this case as long $\{1, 2\}$ and $\{4, 6\}$ aren't the first two entries.

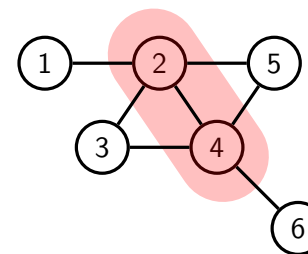
27

Estimation

Given a decomposition of the graph, we have an associated conditional independence: e.g. $(\{1, 3\}, \{2, 4\}, \{5, 6\})$ suggests

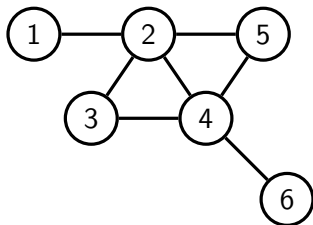
$$X_1, X_3 \perp\!\!\!\perp X_5, X_6 \mid X_2, X_4$$

$$p(x_{123456}) \cdot p(x_{24}) = p(x_{1234}) \cdot p(x_{2456}).$$



And $p(x_{1234})$ and $p(x_{2456})$ are Markov with respect to \mathcal{G}_{1234} and \mathcal{G}_{2456} respectively.

28

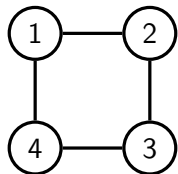


Repeating this process on each subgraph we obtain:

$$p(x_{123456}) \cdot p(x_{24}) \cdot p(x_2) \cdot p(x_4) = p(x_{12}) \cdot p(x_{234}) \cdot p(x_{245}) \cdot p(x_{46}).$$

i.e.

$$p(x_{123456}) = \frac{p(x_{12}) \cdot p(x_{234}) \cdot p(x_{245}) \cdot p(x_{46})}{p(x_{24}) \cdot p(x_2) \cdot p(x_4)}.$$

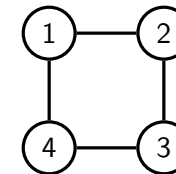


If we 'decompose' without a complete separator set then we introduce constraints between the separate terms:

$$p(x_{1234}) = p(x_1 | x_2, x_4) \cdot p(x_3 | x_2, x_4),$$

but how to ensure that $X_2 \perp\!\!\!\perp X_4 | X_1, X_3$?

But can't we do this for any factorization?



No! Although

$$p(x_{1234}) = \psi_{12}(x_{12}) \cdot \psi_{23}(x_{23}) \cdot \psi_{34}(x_{34}) \cdot \psi_{14}(x_{14}),$$

the ψ s are constrained by the requirement that

$$\sum_{x_{1234}} p(x_{1234}) = 1.$$

These is no nice representation of the ψ_C s in terms of p .

Iterative Proportional Fitting

The Iterative Proportional Fitting Algorithm

```

function IPF(collection of margins  $q(x_{C_i})$ )
  set  $p(x_V)$  to uniform distribution;
  while  $\max_i \max_{x_{C_i}} |p(x_{C_i}) - q(x_{C_i})| > \text{tol}$  do
    for  $i$  in  $1, \dots, k$  do
      update  $p(x_V)$  to  $p(x_{V \setminus C_i} | x_{C_i}) \cdot q(x_{C_i})$ ;
    end for
  end while
  return distribution  $p$  with margins  $p(x_{C_i}) = q(x_{C_i})$ .
end function

```

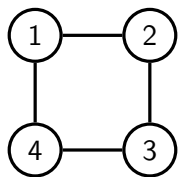
If any distribution satisfying $p(x_{C_i}) = q(x_{C_i})$ for each $i = 1, \dots, k$ exists, then the algorithm converges to the **unique distribution** with those margins and which is Markov with respect to the graph with cliques C_1, \dots, C_k .

Some Data

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	5	10	18	1
	1	0	3	4	0
$X_4 = 1$	0	24	0	9	3
	1	1	2	2	7

Margins

Suppose we want to fit the 4-cycle model:



The relevant margins are:

$n(x_{12})$	$X_2 = 0$	1
$X_1 = 0$	30	33
1	15	11

$n(x_{23})$	$X_3 = 0$	1
$X_2 = 0$	39	6
1	31	13

$n(x_{34})$	$X_4 = 0$	1
$X_3 = 0$	34	36
1	7	12

$n(x_{14})$	$X_4 = 0$	1
$X_1 = 0$	27	36
1	14	12

Start with a Uniform Table

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	5.56	5.56	5.56	5.56
	1	5.56	5.56	5.56	5.56
$X_4 = 1$	0	5.56	5.56	5.56	5.56
	1	5.56	5.56	5.56	5.56

Set Margin X_1, X_2 to Correct Value

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	7.5	3.75	8.25	2.75
	1	7.5	3.75	8.25	2.75
$X_4 = 1$	0	7.5	3.75	8.25	2.75
	1	7.5	3.75	8.25	2.75

Replace

$$p^{(i+1)}(x_1, x_2, x_3, x_4) = p^{(i)}(x_1, x_2, x_3, x_4) \cdot \frac{n(x_1, x_2)}{p^{(i)}(x_1, x_2)}$$

37

Set Margin X_2, X_3 to Correct Value

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	13	6.5	11.62	3.88
	1	2	1	4.88	1.62
$X_4 = 1$	0	13	6.5	11.62	3.88
	1	2	1	4.88	1.62

Replace

$$p^{(i+1)}(x_1, x_2, x_3, x_4) = p^{(i)}(x_1, x_2, x_3, x_4) \cdot \frac{n(x_2, x_3)}{p^{(i)}(x_2, x_3)}$$

38

Set Margin X_3, X_4 to Correct Value

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	12.63	6.31	11.29	3.76
	1	1.47	0.74	3.59	1.2
$X_4 = 1$	0	13.37	6.69	11.96	3.99
	1	2.53	1.26	6.16	2.05

Replace

$$p^{(i+1)}(x_1, x_2, x_3, x_4) = p^{(i)}(x_1, x_2, x_3, x_4) \cdot \frac{n(x_3, x_4)}{p^{(i)}(x_3, x_4)}$$

39

Set Margin X_1, X_4 to Correct Value

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	11.76	7.36	10.52	4.39
	1	1.37	0.86	3.35	1.4
$X_4 = 1$	0	14.15	5.74	12.66	3.42
	1	2.67	1.08	6.52	1.76

Replace

$$p^{(i+1)}(x_1, x_2, x_3, x_4) = p^{(i)}(x_1, x_2, x_3, x_4) \cdot \frac{n(x_1, x_4)}{p^{(i)}(x_1, x_4)}$$

Notice that sum of first column is now 29.96.

40

Set Margin X_1, X_2 to Correct Value

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	11.78	7.37	10.53	4.39
	1	1.37	0.86	3.34	1.39
$X_4 = 1$	0	14.14	5.73	12.64	3.42
	1	2.68	1.09	6.54	1.77

Eventually:

Waiting for this process to converge leads to the MLE:

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	11.76	7.33	10.5	4.4
	1	1.38	0.86	3.35	1.4
$X_4 = 1$	0	14.18	5.72	12.66	3.44
	1	2.68	1.08	6.48	1.76

41

42

Gaussian Graphical Models

The Multivariate Gaussian Distribution

Let $X_V \sim N_p(0, \Sigma)$, where $\Sigma \in \mathbb{R}^{p \times p}$ is a symmetric positive definite matrix.

$$\log p(x_V; \Sigma) = -\frac{1}{2} \log |\Sigma| - \frac{1}{2} x_V^T \Sigma^{-1} x_V + \text{const.}$$

The log-likelihood for Σ is

$$l(\Sigma) = -\frac{n}{2} \log |\Sigma| - \frac{n}{2} \text{tr}(S \Sigma^{-1})$$

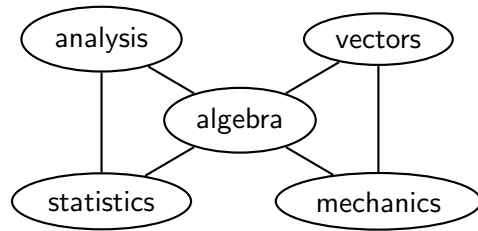
where S is the sample covariance matrix, and this is maximized by choosing $\hat{\Sigma} = S$.

43

44

Gaussian Graphical Models

We have $X_a \perp\!\!\!\perp X_b \mid X_{V \setminus \{a,b\}}$ if and only if $k_{ab} = 0$.



	mechanics	vectors	algebra	analysis	statistics
mechanics	k_{11}	k_{12}	k_{13}	0	0
vectors		k_{22}	k_{23}	0	0
algebra			k_{33}	k_{34}	k_{35}
analysis				k_{44}	k_{45}
statistics					k_{55}

45

Likelihood

From Lemma 4.23, we have

$$\log p(x_V) + \log p(x_S) = \log p(x_A, x_S) + \log p(x_B, x_S).$$

This becomes

$$x_V^T \Sigma^{-1} x_V + x_S^T (\Sigma_{SS})^{-1} x_S - x_{AS}^T (\Sigma_{AS,AS})^{-1} x_{AS} - x_{SB}^T (\Sigma_{SB,SB})^{-1} x_{SB} = 0$$

But can rewrite each term in the form $x_V^T M x_V$, e.g.:

$$x_{AS}^T (\Sigma_{AS,AS})^{-1} x_{AS} = x_V^T \begin{pmatrix} (\Sigma_{AS,AS})^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} x_V$$

Equating terms gives:

$$\Sigma^{-1} = \begin{pmatrix} (\Sigma_{AS,AS})^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & (\Sigma_{SB,SB})^{-1} \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & (\Sigma_{SS})^{-1} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

46

Maximum Likelihood Estimation

Iterating this process with a decomposable graph shows that:

$$\Sigma^{-1} = \sum_{i=1}^k \{(\Sigma_{C_i, C_i})^{-1}\}_{C_i, C_i} - \sum_{i=1}^k \{(\Sigma_{S_i, S_i})^{-1}\}_{S_i, S_i}.$$

For maximum likelihood estimation, using Lemma 4.23 we have

$$\begin{aligned} \hat{\Sigma}^{-1} &= \sum_{i=1}^k \{(\hat{\Sigma}_{C_i, C_i})^{-1}\}_{C_i, C_i} - \sum_{i=1}^k \{(\hat{\Sigma}_{S_i, S_i})^{-1}\}_{S_i, S_i} \\ &= \sum_{i=1}^k \{(W_{C_i, C_i})^{-1}\}_{C_i, C_i} - \sum_{i=1}^k \{(W_{S_i, S_i})^{-1}\}_{S_i, S_i} \end{aligned}$$

where $W_{CC} = \frac{1}{n} \sum_i X_C^{(i)} X_C^{(i)T}$ is the sample covariance matrix.

47

Example

```

> true_inv          # true concentration matrix

      [,1] [,2] [,3] [,4]
[1,]  1.0  0.3  0.2  0.0
[2,]  0.3  1.0 -0.1  0.0
[3,]  0.2 -0.1  1.0  0.3
[4,]  0.0  0.0  0.3  1.0

> solve(true_inv)  # Sigma

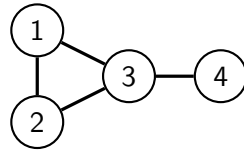
      [,1] [,2] [,3] [,4]
[1,]  1.17 -0.382 -0.30  0.090
[2,] -0.38  1.136  0.21 -0.063
[3,] -0.30  0.209  1.19 -0.356
[4,]  0.09 -0.063 -0.36  1.107

> # rmvnorm is in the mvtnorm package
> dat <- rmvnorm(1000, mean=rep(0,4), sigma = solve(true_inv))
> W <- cov(dat)      # sample covariance
  
```

48

Example

Fit the model with decomposition
({1, 2}, {3}, {4}):



```
> K_hat = matrix(0, 4, 4)
> K_hat[1:3, 1:3] = solve(W[1:3, 1:3])
> K_hat[3:4, 3:4] = K_hat[3:4, 3:4] + solve(W[3:4, 3:4])
> K_hat[3, 3] = K_hat[3, 3] - 1/W[3, 3]
> K_hat
```

```
      [,1] [,2] [,3] [,4]
[1,] 1.079 0.2851 0.2298 0.000
[2,] 0.285 0.9930 -0.0687 0.000
[3,] 0.230 -0.0687 1.0017 0.296
[4,] 0.000 0.0000 0.2958 1.082
```

Note this is close to the true concentration matrix.

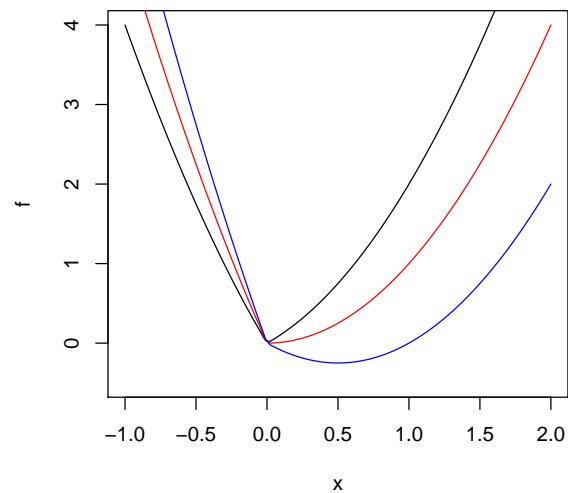
49

The Lasso

50

Cusps

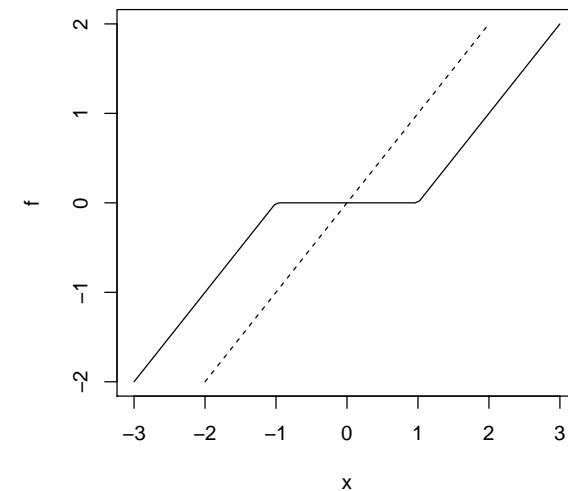
$$f(x) = x^2 - ax + 2|x|, \quad a = 1, a = 2, a = 3.$$



51

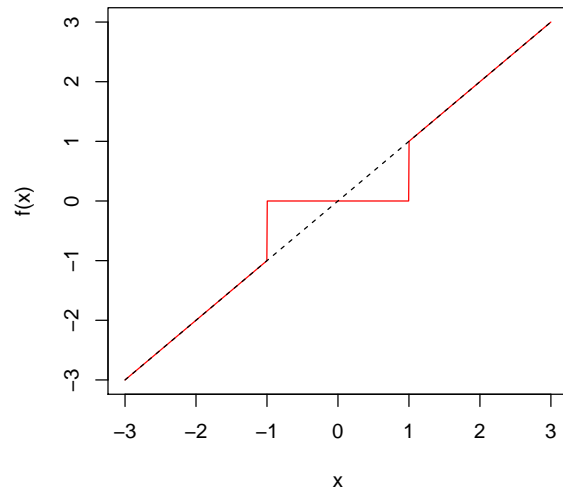
Soft Thresholding

$$f(x) = \text{sign}(x)(|x| - \lambda)_+ \text{ for } x \in (-5, 5) \text{ and } \lambda = 1.$$



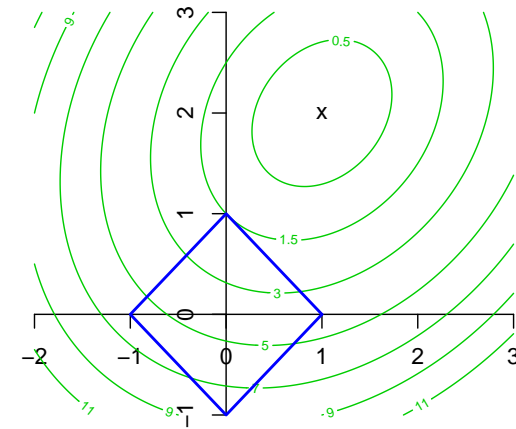
52

Hard Thresholding (Significance Tests)



53

Sparsity



54

Riboflavin Data

The Riboflavin production dataset is available in R's `hdi` package; it consists of $p = 4088$ gene expression measurements in 71 cells, and also measures Riboflavin production.

Here is some code to load it.

```
> library(hdi)
> data("riboflavin")
> Y <- riboflavin[,1] # response variable
> X <- as.matrix(riboflavin[,2]) # some preprocessing!
> class(X) <- "numeric"
> X <- as.data.frame(X)
> dim(X) # n=71, p=4088

[1] 71 4088
```

55

Riboflavin Data

We would like to find a small subset of genes that explain the Riboflavin production as well as possible, without overfitting.

```
> lm0 <- lm(Y ~ 1, data=X) # fit a null model
> # make a formula for the maximal model with all variables
> form <- paste("~ ", paste(names(X), collapse=" + "),
+               sep="")
> substr(form, 1, 50)

[1] "~ AADK_at + AAPA_at + ABFA_at + ABH_at + ABNA_at +"
> # stepAIC(lm0, scope=form) # doesn't work!
```

The `stepAIC()` function in R seems unable to deal with systems of this size (in principle it ought to work though).

56

Riboflavin Data

On the other hand, we can use the lasso to select variables and fit the model all at once.

The lars package gives an algorithm (called LARS) that can fit the entire lasso solution path (i.e. for all λ).

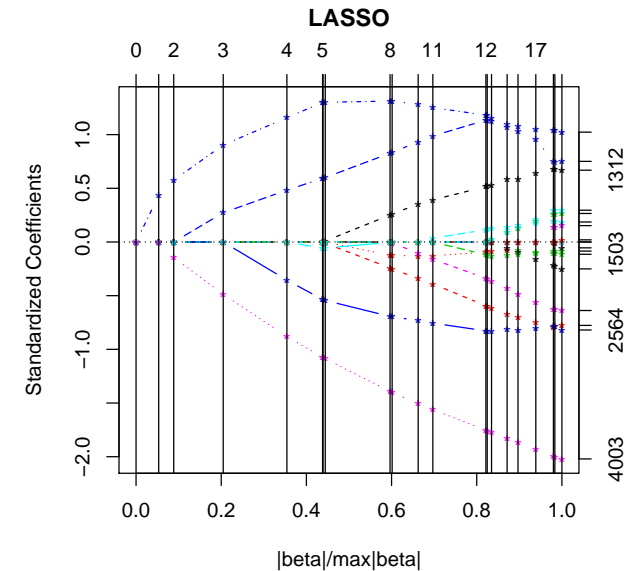
```
> library(lars)
> out <- lars(x=as.matrix(X), y=Y, max.steps = 20)
```

There are more than 500 variables and $n < m$;
You may wish to restart and set `use.Gram=FALSE`

```
> ## plot(out)
```

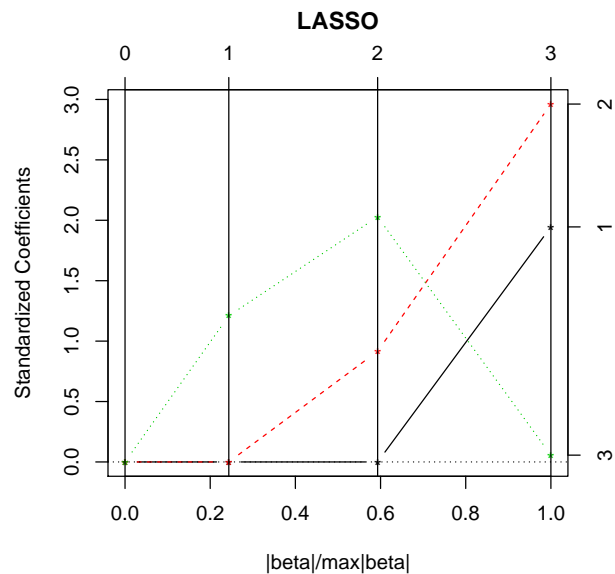
57

Lasso Solution Path



58

Irrepresentability



59

Sachs et al. Data

Sachs et al. measured the quantity of 11 signalling proteins in 7,466 cells.

```
> dat <- read.table("sachs_et_al.txt", header=TRUE)
> head(dat)
```

	PIP3	Plcg	PIP2	PKC	PKA	Raf	Mek	Erk	P38	Jnk	Akt
1	58.8	8.8	18.3	17.0	414	26	13.2	6.6	45	40	17
2	8.1	12.3	16.8	3.4	352	36	16.5	18.6	16	62	32
3	13.0	14.6	10.2	11.4	403	59	44.1	14.9	32	20	32
4	1.3	23.1	13.5	13.7	528	73	82.8	5.8	29	23	12
5	24.8	5.2	9.7	4.7	305	34	19.8	21.1	26	81	46
6	10.9	17.6	22.1	13.7	610	19	3.8	11.9	49	58	26

```
> S <- cov(dat)
```

60

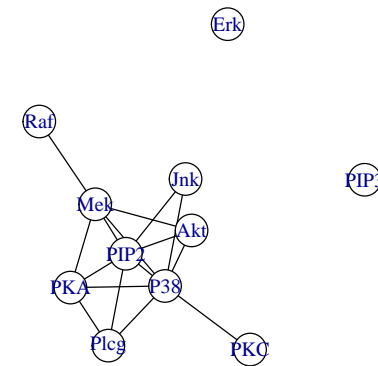
The glasso package runs the coordinate descent algorithm.

```
> library(glasso)
> out <- glassopath(S,
+   rholist = 10^seq(from=3, to=5, length=21),
+   penalize.diagonal = FALSE, trace=0)
```

61

Directed Graphical Models

Here is the graph for $\lambda = 10^4$.



62

Directed Graphs

We have so far used **undirected graphs**.

Directed graphs give each edge an orientation.

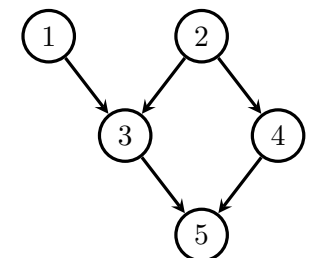
A directed graph \mathcal{G} is a pair (V, D) , where

- V is a set of vertices;
- D is a set of ordered pairs of vertices (i, j) such that $i, j \in V$ and $i \neq j$.

If $(i, j) \in D$ we write $i \rightarrow j$.

$$V = \{1, 2, 3, 4, 5\}$$

$$D = \{(1, 3), (2, 3), (2, 4), (3, 5), (4, 5)\}$$



63

64

Acyclicity

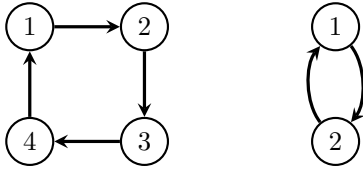
Paths are sequences of adjacent vertices, without repetition:

$$1 \rightarrow 3 \leftarrow 2 \rightarrow 4 \rightarrow 5 \qquad 1 \rightarrow 3 \rightarrow 5.$$

The path is **directed** if all the arrows point away from the start.

(A path of length 0 is just a single vertex.)

A **directed cycle** is a directed path from i to $j \neq i$, together with $j \rightarrow i$.



Graphs that contain no directed cycles are called **acyclic**. or more specifically, **directed acyclic graphs** (DAGs).

All the directed graphs we consider are acyclic.

65

Happy Families

$$i \rightarrow j \quad \begin{cases} i \in \text{pa}_G(j) & i \text{ is a parent of } j \\ j \in \text{ch}_G(i) & j \text{ is a child of } i \end{cases}$$

$$a \rightarrow \dots \rightarrow b \quad \begin{cases} a \in \text{ang}(b) & a \text{ is an ancestor of } b \\ b \in \text{deg}(a) & b \text{ is a descendant of } a \end{cases}$$

or $a = b$

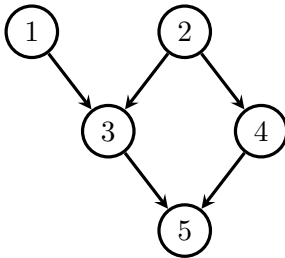
If $w \notin \text{deg}(v)$ then w is a **non-descendant** of v :

$$\text{nd}_G(v) = V \setminus \text{deg}(v).$$

(Notice that no v is a non-descendant of itself).

66

Examples



$$\text{pa}_G(3) = \{1, 2\}$$

$$\text{ch}_G(5) = \emptyset$$

$$\text{ang}(4) = \{2, 4\}$$

$$\text{deg}(1) = \{1, 3, 5\}$$

$$\text{nd}_G(1) = \{2, 4\}.$$

67

Topological Orderings

If the graph is acyclic, we can find a **topological ordering**: i.e. one in which no vertex comes before any of its parents. (Proof: induction)

Topological orderings:

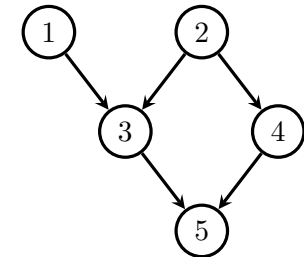
$$1, 2, 3, 4, 5$$

$$1, 2, 4, 3, 5$$

$$2, 1, 3, 4, 5$$

$$2, 1, 4, 3, 5$$

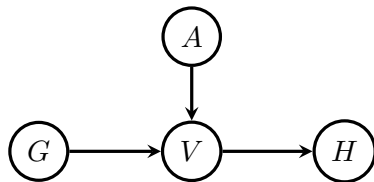
$$2, 4, 1, 3, 5$$



68

Parameter Estimation

- G : group assigned to patient;
- A : patient's age in years;
- V : whether patient received flu vaccine;
- H : patient hospitalized with respiratory problems;

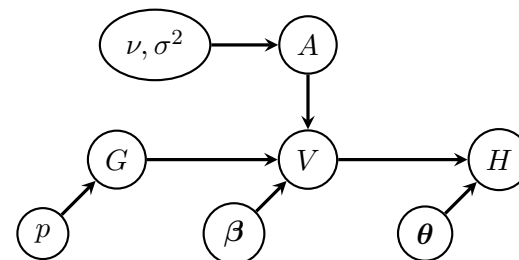


Parameter Estimation

We can model the data (G_i, A_i, V_i, H_i) as

- group : $G_i \sim \text{Bernoulli}(p)$;
- age : $A_i \sim N(\nu, \sigma^2)$;
- vaccine : $V_i | A_i, G_i \sim \text{Bernoulli}(\mu_i)$ where
 $\text{logit } \mu_i = \beta_0 + \beta_1 A_i + \beta_2 G_i$.
- hospital : $H_i | V_i \sim \text{Bernoulli}(\text{expit}(\theta + \theta_1 V_i))$.

Assuming independent priors:

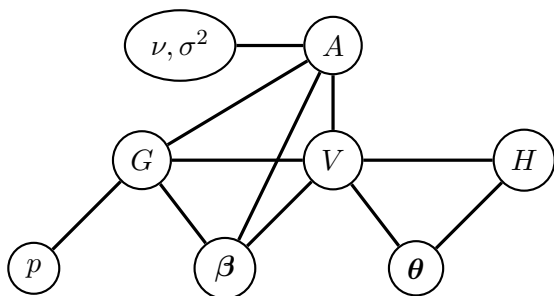


Bayesian Inference

From our argument, we have

$$\pi(\beta | G, A, V, H) = \pi(\beta | G, A, V) \propto p(V | A, G, \beta) \cdot \pi(\beta).$$

Looking at the moral graph we see

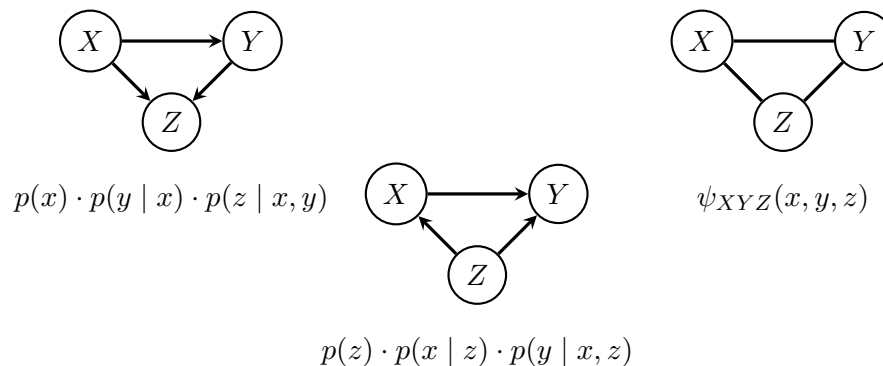


Markov Equivalence

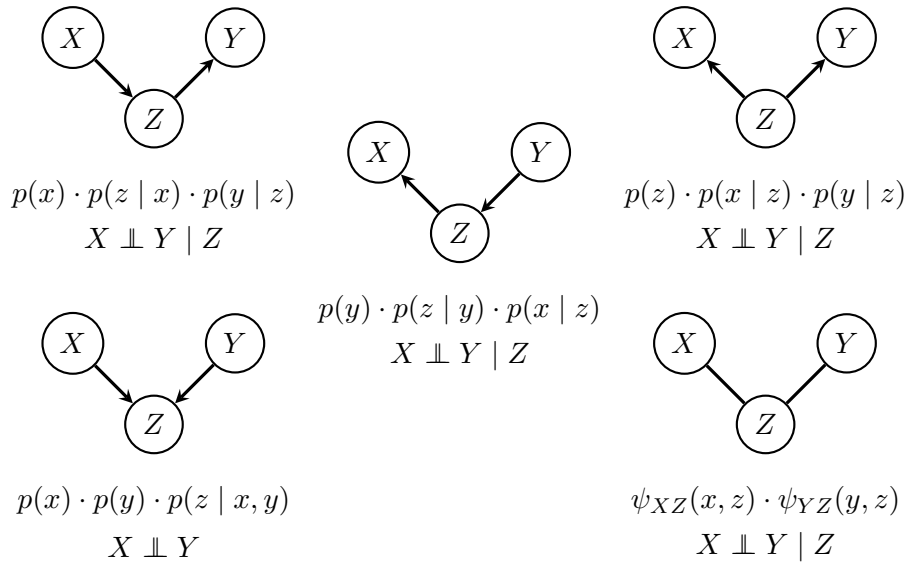
All undirected graphs induce distinct models.

$$v \not\sim w \iff X_v \perp\!\!\!\perp X_w | X_{V \setminus \{v,w\}} \text{ implied}$$

The same is not true for directed graphs:



Markov Equivalence



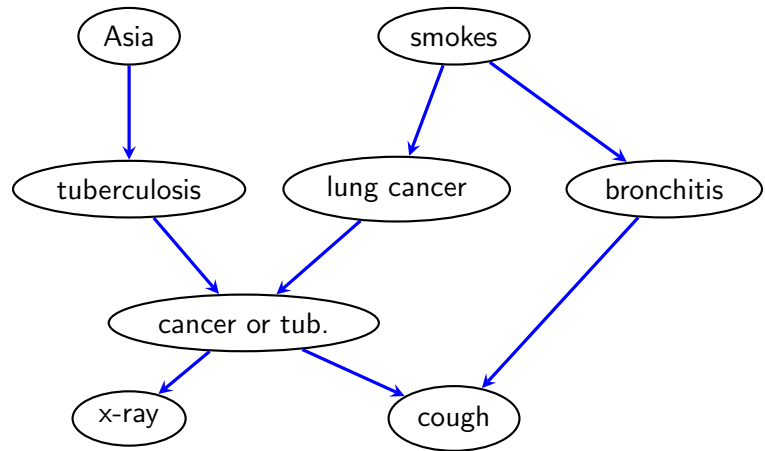
73

Expert Systems

74

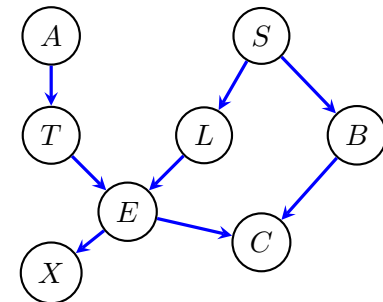
Expert Systems

Variables



The 'Chest Clinic' network, a fictitious diagnostic model.

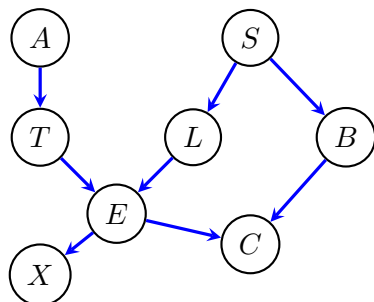
75



- A has the patient recently visited southern Asia?
- S does the patient smoke?
- T,C,B Tuberculosis, lung cancer, bronchitis.
- E logical: Tuberculosis OR lung cancer.
- X shadow on chest X-ray?
- C does the patient have a persistent cough?

76

Conditional Probability Tables



We have our factorization:

$$p(a, s, t, l, b, e, x, c) = p(a) \cdot p(s) \cdot p(t | a) \cdot p(l | s) \cdot p(b | s) \cdot p(e | t, l) \cdot p(x | e) \cdot p(c | e, b).$$

Assume that we are given each of these factors. How could we calculate $p(l | x, c, a, s)$?

77

Probabilities

$$p(a) = \begin{array}{c|cc} & \text{yes} & \text{no} \\ \hline & 0.01 & 0.99 \end{array}$$

$$p(s) = \begin{array}{c|cc} & \text{yes} & \text{no} \\ \hline & 0.5 & 0.5 \end{array}$$

$$p(t | a) = \begin{array}{c|cc} A & \text{yes} & \text{no} \\ \hline \text{yes} & 0.05 & 0.95 \\ \text{no} & 0.01 & 0.99 \end{array}$$

$$p(l | s) = \begin{array}{c|cc} S & \text{yes} & \text{no} \\ \hline \text{yes} & 0.1 & 0.9 \\ \text{no} & 0.01 & 0.99 \end{array}$$

$$p(b | s) = \begin{array}{c|cc} S & \text{yes} & \text{no} \\ \hline \text{yes} & 0.6 & 0.4 \\ \text{no} & 0.3 & 0.7 \end{array}$$

$$p(x | e) = \begin{array}{c|cc} E & \text{yes} & \text{no} \\ \hline \text{yes} & 0.98 & 0.02 \\ \text{no} & 0.05 & 0.95 \end{array}$$

$$p(c | b, e) = \begin{array}{cc|cc} B & E & \text{yes} & \text{no} \\ \hline \text{yes} & \text{yes} & 0.9 & 0.1 \\ & \text{no} & 0.8 & 0.2 \\ \text{no} & \text{yes} & 0.7 & 0.3 \\ & \text{no} & 0.1 & 0.9 \end{array}$$

78

Factorizations

$$p(l | x, c, a, s) = \frac{p(l, x, c | a, s)}{\sum_l p(l, x, c | a, s)}$$

From the graph $p(l, x, c | a, s)$ is

$$\sum_{t, e, b} p(t | a) \cdot p(l | s) \cdot p(b | s) \cdot p(e | t, l) \cdot p(x | e) \cdot p(c | e, b).$$

But this is:

$$p(l | s) \sum_e p(x | e) \left(\sum_b p(b | s) \cdot p(c | e, b) \right) \left(\sum_t p(t | a) \cdot p(e | t, l) \right).$$

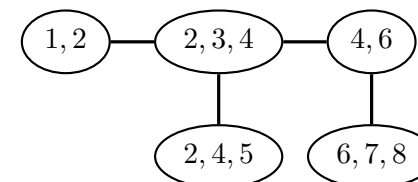
79

Junction Trees

A **junction tree**:

- is a connected undirected graph without cycles (a tree);
- has vertices C_i that consist of **subsets** of a set V ;
- satisfies the property that if $C_i \cap C_j = S$ then every vertex on the (unique) path from C_i to C_j contains S .

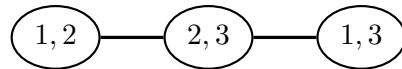
Example.



80

Junction Trees

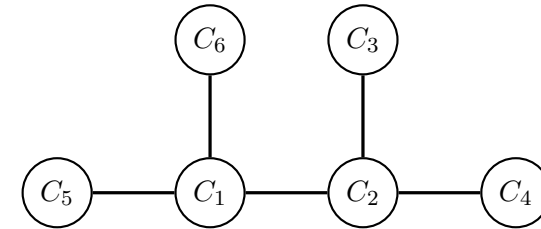
The following graph is **not** a junction tree:



81

Junction Trees

Junction trees can be constructed directly from sets of cliques satisfying running intersection.

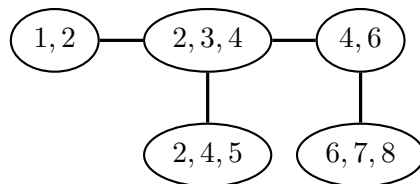


$$C_i \cap \bigcup_{j < i} C_j = C_i \cap C_{\sigma(i)}.$$

82

Example: Junction Trees and RIP

Given sets $\{1, 2\}$, $\{2, 3, 4\}$, $\{2, 4, 5\}$, $\{4, 6\}$, $\{6, 7, 8\}$, we can build this tree:

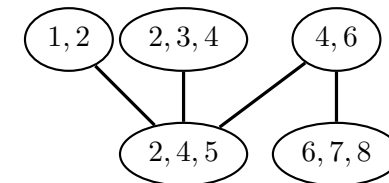


83

Example: Junction Trees and RIP

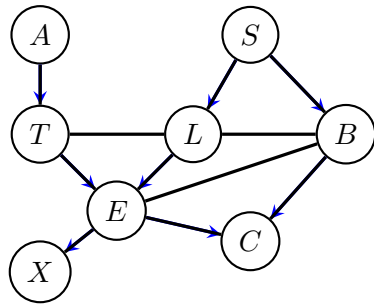
Equally, we could use a different ordering:

$$\{6, 7, 8\}, \{4, 6\}, \{2, 4, 5\}, \{1, 2\}, \{2, 3, 4\}.$$



84

Forming A Junction Tree



Steps to Forming a Junction Tree:

Moralize

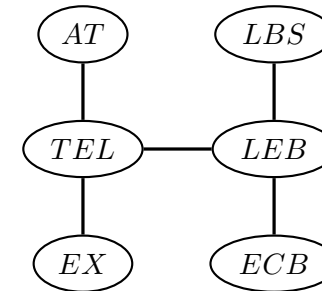
Drop directions

Triangulate (add edges to get a decomposable graph)

85

Forming A Junction Tree

Finally, form the tree of cliques.



86

Initialization

$$p(a) = \begin{array}{c|cc} & \text{yes} & \text{no} \\ \hline & 0.01 & 0.99 \end{array}$$

$$p(s) = \begin{array}{c|cc} & \text{yes} & \text{no} \\ \hline & 0.5 & 0.5 \end{array}$$

$$p(t | a) = \begin{array}{c|cc} A & \text{yes} & \text{no} \\ \hline \text{yes} & 0.05 & 0.95 \\ \text{no} & 0.01 & 0.99 \end{array}$$

$$p(l | s) = \begin{array}{c|cc} S & \text{yes} & \text{no} \\ \hline \text{yes} & 0.1 & 0.9 \\ \text{no} & 0.01 & 0.99 \end{array}$$

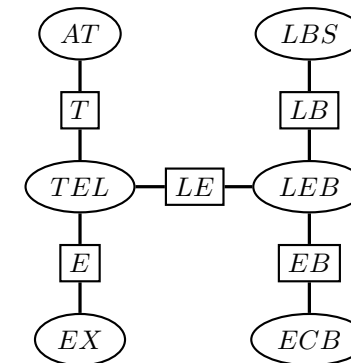
$$p(b | s) = \begin{array}{c|cc} S & \text{yes} & \text{no} \\ \hline \text{yes} & 0.6 & 0.4 \\ \text{no} & 0.3 & 0.7 \end{array}$$

$$p(x | e) = \begin{array}{c|cc} E & \text{yes} & \text{no} \\ \hline \text{yes} & 0.98 & 0.02 \\ \text{no} & 0.05 & 0.95 \end{array}$$

$$p(c | b, e) = \begin{array}{cc|cc} B & E & \text{yes} & \text{no} \\ \hline \text{yes} & \text{yes} & 0.9 & 0.1 \\ & \text{no} & 0.8 & 0.2 \\ \text{no} & \text{yes} & 0.7 & 0.3 \\ & \text{no} & 0.1 & 0.9 \end{array}$$

87

Initialization



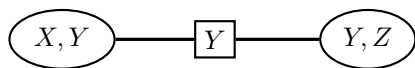
Can set, for example:

$$\begin{aligned} \psi_{AT}(a, t) &= p(a) \cdot p(t | a) & \psi_{LBS}(l, b, s) &= p(s) \cdot p(l | s) \cdot p(b | s) \\ \psi_{TEL}(t, e, l) &= p(e | t, l) & \psi_{ELB}(e, l, b) &= 1 \\ \psi_{EX}(e, x) &= p(x | e) & \psi_{ECB}(e, c, b) &= p(c | e, b). \end{aligned}$$

88

Updating / Message Passing

Suppose we have two vertices and one separator set.



$$\psi_{XY}(x, y)$$

	$y = 0$	1
$x = 0$	0.3	0.9
$x = 1$	0.7	0.1

$$\psi_Y(y)$$

$y = 0$	1
1	1

$$\psi_{YZ}(y, z)$$

	$z = 0$	1
$y = 0$	0.3	0.1
$y = 1$	0.2	0.4

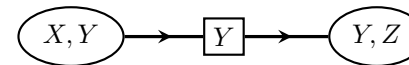
Initialize with

$$\psi_{XY}(x, y) = p(x | y) \quad \psi_{YZ}(y, z) = p(z | y) \cdot p(y) \quad \psi_Y(y) = 1.$$

89

Updating / Message Passing

Suppose we have two vertices and one separator set.



$$\psi_{XY}(x, y)$$

	$y = 0$	1
$x = 0$	0.3	0.9
$x = 1$	0.7	0.1

$$\psi_Y(y)$$

$y = 0$	1
1	1

$$\psi_{YZ}(y, z)$$

	$z = 0$	1
$y = 0$	0.3	0.1
$y = 1$	0.2	0.4

Pass message from X, Y to Y, Z . We set

$$\psi'_Y(y) = \sum_x \psi_{XY}(x, y) = (1, 1);$$

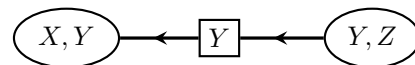
$$\psi'_{YZ}(y, z) = \frac{\psi'_Y(y)}{\psi_Y(y)} \psi_{YZ}(y, z) = \psi_{YZ}(y, z).$$

So in this case nothing changes.

90

Updating / Message Passing

Suppose we have two vertices and one separator set.



$$\psi_{XY}(x, y)$$

	$y = 0$	1
$x = 0$	0.3	0.9
$x = 1$	0.7	0.1

$$\psi'_Y(y)$$

$y = 0$	1
1	1

$$\psi'_{YZ}(y, z)$$

	$z = 0$	1
$y = 0$	0.3	0.1
$y = 1$	0.2	0.4

Pass message from Y, Z to X, Y . We set

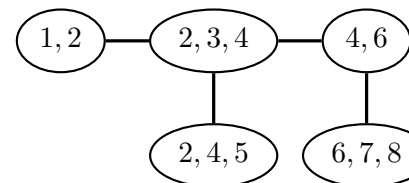
$$\psi''_Y(y) = \sum_z \psi'_{YZ}(y, z) = (0.4, 0.6);$$

$$\psi'_{XY}(x, y) = \frac{\psi''_Y(y)}{\psi'_Y(y)} \psi_{XY}(x, y) = \begin{pmatrix} 0.12 & 0.54 \\ 0.28 & 0.06 \end{pmatrix}.$$

And now we note that $\psi'_{XY}(x, y) = p(x, y)$ as intended.

91

Rooting



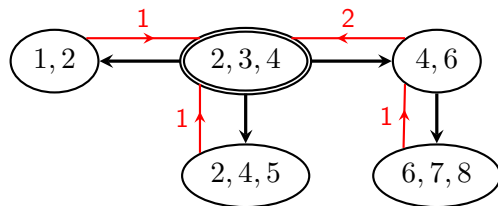
Given a tree, we can pick any vertex as a 'root', and direct all edges away from it.

92

Collection and Distribution

```

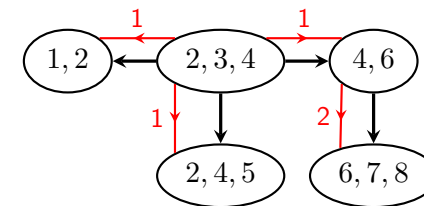
function COLLECT(rooted tree  $\mathcal{T}$ , potentials  $\psi_t$ )
  let  $1 < \dots < k$  be a topological ordering of  $\mathcal{T}$ 
  for  $t$  in  $k, \dots, 2$  do
    send message from  $\psi_t$  to  $\psi_{\sigma(t)}$ ;
  end for
  return updated potentials  $\psi_t$ 
end function
    
```



Collection and Distribution

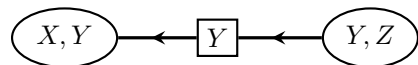
```

function DISTRIBUTE(rooted tree  $\mathcal{T}$ , potentials  $\psi_t$ )
  let  $1 < \dots < k$  be a topological ordering of  $\mathcal{T}$ 
  for  $t$  in  $2, \dots, k$  do
    send message from  $\psi_{\sigma(t)}$  to  $\psi_t$ ;
  end for
  return updated potentials  $\psi_t$ 
end function
    
```



Evidence

Now, suppose we want to calculate $p(x | z = 0)$.



		$\psi_{XY}(x, y)$	
		$y = 0$	1
x	0	0.12	0.54
	1	0.28	0.06

		$\psi_Y(y)$	
		$y = 0$	1
y	0	0.4	0.6
	1	0.4	0

		$\psi_{YZ}(y, z)$	
		$z = 0$	1
y	0	0.6	0
	1	0.4	0

Replace $\psi_{YZ}(y, z)$ with $p(y | z = 0)$.

Pass message from Y, Z to X, Y . We set

$$\psi_Y(y) = \sum_x \psi_{YZ}(y, z) = (0.60.4);$$

$$\psi'_{XY}(x, y) = \frac{\psi''_Y(y)}{\psi'_Y(y)} \psi_{XY}(x, y) = \begin{matrix} 0.18 & 0.36 \\ 0.42 & 0.04 \end{matrix}$$

And now calculate $\sum_y \psi_{XY}(x, y) = (0.54, 0.46)$.

From the Chest Clinic Network

Marginal Probability Tables:

ψ_{EX}		E	
		yes	no
yes	yes	0.06	0
	no	0.05	0.89

ψ_{LBS}		L	
		yes	no
yes	yes	0.03	0
	no	0.02	0
no	yes	0.27	0.15
	no	0.18	0.35

ψ_{TEL}		T	
		yes	no
yes	yes	5.72×10^{-4}	0
	no	0.01	0
no	yes	0.05	0
	no	0	0.94

ψ_{AT}		A	
		yes	no
yes	yes	5×10^{-4}	0.01
	no	0.01	0.98

ψ_{LEB}		L	
		yes	no
yes	yes	0.03	0.02
	no	0	0
no	yes	0	0.01
	no	0.41	0.52

ψ_{ECB}		B	
		yes	no
yes	yes	0.03	0
	no	0.02	0.01
no	yes	0.33	0.08
	no	0.05	0.47

From the Chest Clinic Network

Suppose now that we have a shadow on the chest X-ray:

	E	yes	no	
ψ_{EX}	yes	0.58	-	
	no	0.42	-	

	A	yes	no
ψ_{AT}	yes	0	0.01
	no	0.09	0.9

	L	B	yes	no
ψ_{LBS}	yes	yes	0.27	0.01
		no	0.18	0.03
	no	yes	0.15	0.08
		no	0.1	0.19

	L	E	yes	no
ψ_{LEB}	yes	yes	0.28	0.21
		no	0	0
	no	yes	0.04	0.05
		no	0.19	0.24

	T	E	yes	no
ψ_{TEL}	yes	yes	0.01	0
		no	0.09	0
	no	yes	0.48	0
		no	0	0.42

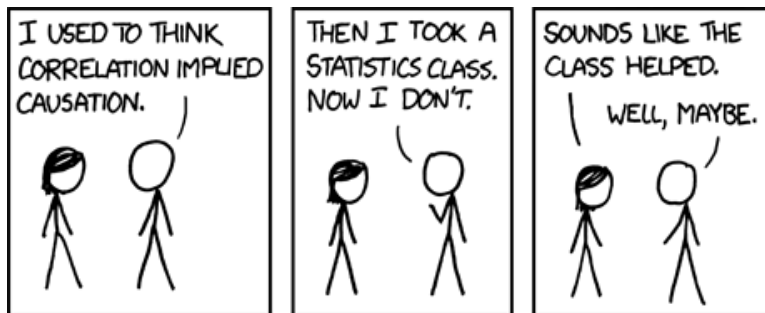
	B	E	yes	no
ψ_{ECB}	yes	yes	0.29	0.03
		no	0.18	0.08
	no	yes	0.15	0.04
		no	0.02	0.21

97

Causal Inference

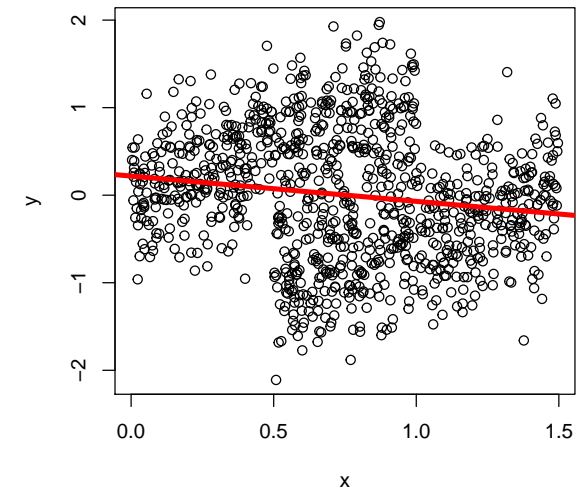
98

Correlation



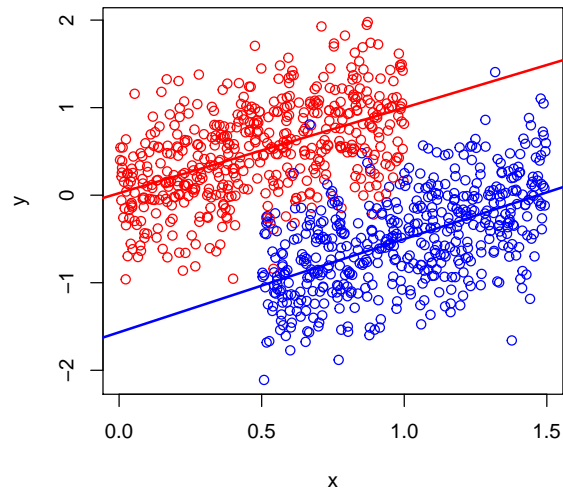
99

Controlling for Covariates



100

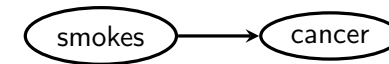
Controlling for Covariates



101

Causation

Example. Smoking is strongly predictive of lung cancer. So maybe smoking causes lung cancer to develop.



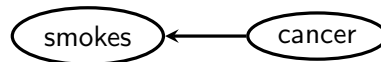
BUT: how do we know that this is a causal relationship? And what do we mean by that?

The central question is: “if we stop people from smoking, will they be less likely to get lung cancer?”

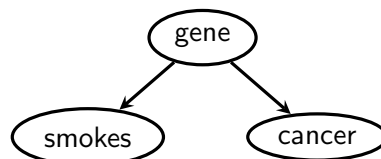
That is: does this ‘intervention’ on one variable change the distribution of another variable?

102

Alternative Explanations



Reverse Causation. Lung cancer causes smoking: people with (undiagnosed) lung cancer smoke to soothe irritation in the lungs.



Confounding / Common Cause. There is a gene that makes people likely to smoke, and also more likely to get lung cancer.

103

Example

Suppose we take 32 men and 32 women, ask them whether they smoke and check for lung damage.

	women		men	
	not smoke	smoke	not smoke	smoke
no damage	21	6	6	6
damage	3	2	2	18

Marginally, there is clearly a strong relationship between smoking and damage

	not smoke	smoke
no damage	27	12
damage	5	20

$$P(D = 1 | S = 1) = \frac{5}{8} \quad P(D = 1 | S = 0) = \frac{5}{32}$$

104

Example

This might suggest that if we had prevented them all from smoking, only $\frac{5}{32} \times 64 = 10$ would have had damage, whereas if we had made them all smoke, $\frac{5}{8} \times 64 = 40$ would have damage.

But: both smoking and damage are also correlated with gender, so this effect may be inaccurate. If we repeat this separately for men and women:

no-one smoking:

$$\frac{3}{21+3} \times 32 + \frac{2}{6+2} \times 32 = 12$$

everyone smoking

$$\frac{2}{6+2} \times 32 + \frac{18}{18+6} \times 32 = 32.$$

Compare these to 10 and 40.

105

'do' notation

In this example there is a difference between predicting damage when we 'observe' that someone smokes ...

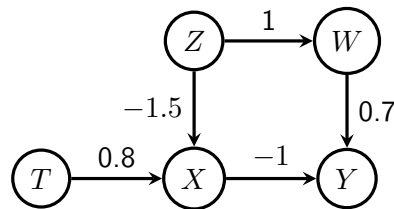
$$P(D = 1 \mid S = 1) = \frac{5}{8},$$

... and predicting damage when we intervene to make someone smoke:

$$P(D = 1 \mid do(S = 1)) = \frac{32}{64} = \frac{1}{2}.$$

106

Gaussian



```
> set.seed(513)
> n <- 1e3
> Z <- rnorm(n)
> T <- rnorm(n)
> W <- Z + rnorm(n)
> X <- 0.8*T - 1.5*Z + rnorm(n)
> Y <- 0.7*W - X + rnorm(n)
```

107

Back-Door Paths

```
> summary(lm(Y ~ X))$coefficients[,1:2]
              Estimate Std. Error
(Intercept)  0.035         0.04
X            -1.285         0.02

> summary(lm(Y ~ X + Z))$coefficients[,1:2]
              Estimate Std. Error
(Intercept)  0.043         0.038
X            -1.024         0.032
Z             0.645         0.062

> summary(lm(Y ~ X + W))$coefficients[,1:2]
              Estimate Std. Error
(Intercept)  0.029         0.031
X            -1.011         0.019
W             0.668         0.027
```

108

Instruments

Adding in unnecessary variables to the regression generally increases the variance.

```
> summary(lm(Y ~ X + W + T))$coefficients[,1:2]
```

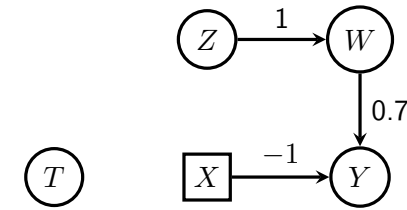
	Estimate	Std. Error
(Intercept)	0.029	0.031
X	-1.006	0.022
W	0.671	0.027
T	-0.018	0.036

```
> summary(lm(Y ~ X + W + Z))$coefficients[,1:2]
```

	Estimate	Std. Error
(Intercept)	0.028	0.031
X	-1.026	0.026
W	0.682	0.031
Z	-0.053	0.061

109

Simulating Intervention



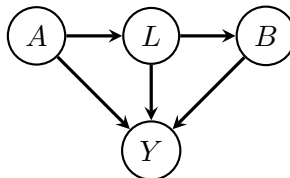
```

> Z <- rnorm(n)
> T <- rnorm(n)
> W <- Z + rnorm(n)
> X <- rnorm(n) # set X independently
> Y <- 0.7*W - X + rnorm(n)
> summary(lm(Y ~ X))$coefficients[,1:2]
  
```

	Estimate	Std. Error
(Intercept)	-0.04	0.045
X	-1.05	0.044

110

Example: HIV Treatment



A treatment with AZT (an HIV drug);
L opportunistic infection;
B treatment with antibiotics;
Y survival at 5 years.

$$\begin{aligned}
 p(a, l, b, y) &= p(a) \cdot p(l | a) \cdot p(b | l) \cdot p(y | a, l, b) \\
 p(l, y | do(a, b)) &= p(l | a) \cdot p(y | a, l, b) \\
 p(y | do(a, b)) &= \sum_l p(l | a) \cdot p(y | a, l, b).
 \end{aligned}$$

111

Model Selection and Causal Discovery

112

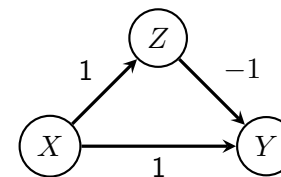
Finding Graphs

```
> head(dat)
      X      Y      Z
1  1.15  0.20  0.38
2  0.90 -1.57  1.26
3  0.79  1.52 -0.03
4  1.15 -0.74  2.47
5  0.74  0.78  0.40
6 -0.15  1.14 -1.04

> round(cov(dat), 2)
      X      Y      Z
X  0.99  0.01  1.02
Y  0.01  1.98 -0.98
Z  1.02 -0.98  2.06
```

113

Extra Independences



```
> n <- 1000
> X <- rnorm(n)
> Z <- X + rnorm(n)
> Y <- X - Z + rnorm(n)
> summary(lm(Y ~ X))$coefficients # no significant effect of X on Y

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.036      0.046    0.78   0.43
X              -0.086      0.045   -1.88   0.06
```

114

The PC Algorithm

Start with a complete undirected graph, \mathcal{H} .

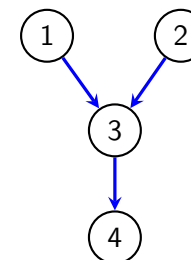
- For every pair of vertices i, j , test the marginal independence $X_i \perp\!\!\!\perp X_j$.
If it holds, remove $i - j$ from \mathcal{H} .
- For every remaining edge i, j and $k \in V \setminus \{i, j\}$, test $X_i \perp\!\!\!\perp X_j \mid X_k$.
If it holds, remove $i - j$ from \mathcal{H} .
- For every remaining edge i, j and $\{k, l\} \subseteq V \setminus \{i, j\}$, test $X_i \perp\!\!\!\perp X_j \mid X_k, X_l$.
If it holds, remove $i - j$ from \mathcal{H} .
- ...

In fact, it suffices to consider subsets of the neighbours of i , or the neighbours of j .

115

Example

Consider the following true graph and its implied pairwise independences.

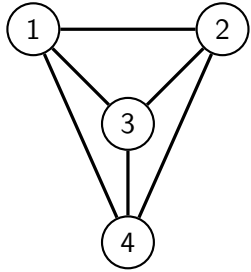


$X_1 \perp\!\!\!\perp X_2$
 $X_1 \perp\!\!\!\perp X_4 \mid X_3$
 $X_1 \perp\!\!\!\perp X_4 \mid X_2, X_3$
 $X_2 \perp\!\!\!\perp X_4 \mid X_3$
 $X_2 \perp\!\!\!\perp X_4 \mid X_1, X_3$

116

Example

Start with a complete graph:

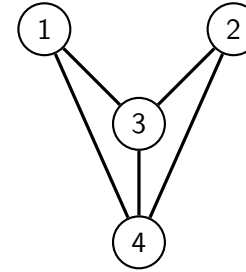


We have $X_1 \perp\!\!\!\perp X_2$, but no other marginal independences.

117

Example

Start with a complete graph:

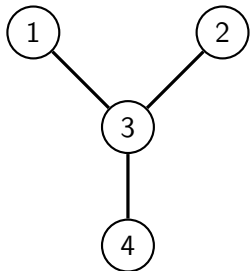


Now test $X_i \perp\!\!\!\perp X_j \mid X_k$ for each remaining edge $i - j$ and other variable k .
This gives $X_1 \perp\!\!\!\perp X_4 \mid X_3$ and $X_2 \perp\!\!\!\perp X_4 \mid X_3$.

118

Example

We have recovered the correct skeleton.

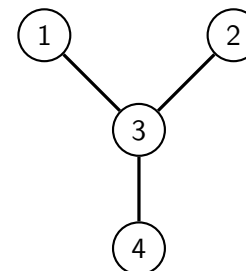


We would continue to test, but should find only dependence.
Can we orient the edges?
We found $X_1 \perp\!\!\!\perp X_2$ so we must have $1 \rightarrow 3 \leftarrow 2$.
Since we didn't find $X_1 \perp\!\!\!\perp X_4$, we must have $3 \rightarrow 4$ to avoid a v-structure.

119

Example

Start with a complete graph:



We would continue to test, but should find only dependence.

120

The PC Algorithm

```
function SKELETON(distribution  $p(x_V)$ )
  Start with complete undirected graph  $\mathcal{H}$ ;
  for  $k$  in  $0, 1, \dots, p - 2$  do
    for every  $i \sim j$  in  $\mathcal{H}$  do
      for  $C \subseteq \text{bd}_{\mathcal{H}}(i) \setminus \{j\}$  or  $C \subseteq \text{bd}_{\mathcal{H}}(j) \setminus \{i\}$  with  $|C| = k$  do
        if  $X_i \perp\!\!\!\perp X_j \mid X_C [p(x_V)]$  then
          remove  $i - j$  edge from  $\mathcal{H}$ ;
          record  $\text{SepSet}(i, j) = C$ ;
          exit loop over  $C$  and move to next edge in  $\mathcal{H}$ .
        end if
      end for
    end for
  end for
  return  $\mathcal{H}$ , collection of  $\text{Sepset}(i, j)$ s.
end function
```

121

Example

Consider the mathematics test marks again

```
> data(marks, package = "ggm")
> C <- cor(marks)
> n <- 88
> C
```

	mechanics	vectors	algebra	analysis	statistics
mechanics	1.00	0.55	0.55	0.41	0.39
vectors	0.55	1.00	0.61	0.49	0.44
algebra	0.55	0.61	1.00	0.71	0.66
analysis	0.41	0.49	0.71	1.00	0.61
statistics	0.39	0.44	0.66	0.61	1.00

123

The PC Algorithm

```
function ORIENT(Skeleton  $\mathcal{H}$ , collection of  $\text{Sepset}(i, j)$ s)
  for every triple  $i - k - j$  in  $\mathcal{H}$  with  $i \not\sim j$  do
    if  $k \notin \text{SepSet}(i, j)$  then
      orient  $i \rightarrow k \leftarrow j$ .
    end if
  end for
  return  $\mathcal{H}$ , collection of  $\text{Sepset}(i, j)$ s.
end function
```

122

Example

Consider the mathematics test marks again

```
> f <- function(x) 0.5*log((1+x)/(1-x))
> sqrt(n-3)*f(C)
```

	mechanics	vectors	algebra	analysis	statistics
mechanics	Inf	5.7	5.7	4.0	3.8
vectors	5.7	Inf	6.5	4.9	4.3
algebra	5.7	6.5	Inf	8.2	7.4
analysis	4.0	4.9	8.2	Inf	6.5
statistics	3.8	4.3	7.4	6.5	Inf

Comparing to a $N(0, 1)$, we would reject the possibility of any marginal independences.

124

Example

Now try conditional independences with sets of size one.

```
> S <- c(2,3,4) # pick subset of variables
> Cp <- solve(C[S,S]) # compute inverse matrix
> Cp <- Cp/sqrt(diag(Cp)*rep(diag(Cp),each=3)) # get partial cors
> sqrt(n-3-1)*f(Cp)
```

	vectors	algebra	analysis
vectors	Inf	-4.2	-0.85
algebra	-4.22	Inf	-6.34
analysis	-0.85	-6.3	Inf

Comparing to a $N(0,1)$, we would accept that Analysis is independent of Mechanics conditional on Vectors.

125

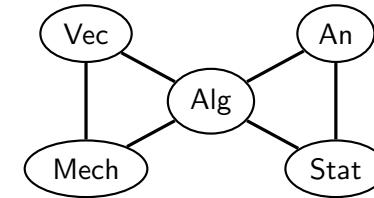
Gibbs Sampling

127

Example

Repeating this process we would obtain

Analysis \perp Mechanics | Vectors
Statistics \perp Mechanics | Vectors
Analysis \perp Vectors | Algebra
Statistics \perp Vectors | Algebra



There are two implied v-structures.

Two more edges can be oriented to avoid more v-structures and cycles.

126

Gibbs Sampling

Suppose

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2 \left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right)$$

so

$$K = \Sigma^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}.$$

Then

$$X_1 | X_2 = x_2 \sim N(\rho x_2, (1-\rho)^2)$$

$$X_2 | X_1 = x_1 \sim N(\rho x_1, (1-\rho)^2)$$

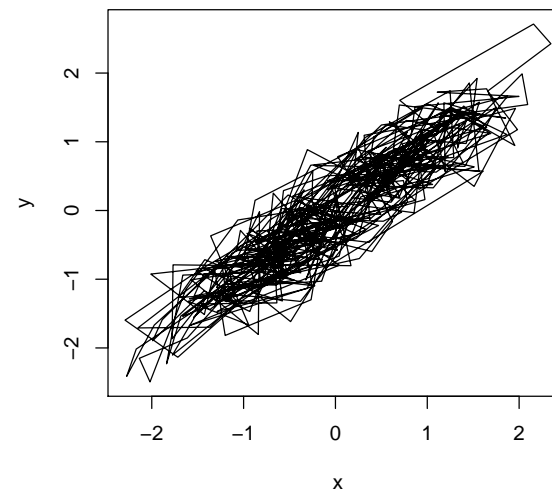
128

Gibbs Sampler

```
> ## Gaussian Gibbs sampler
> rho <- 0.9 ## correlation
> N <- 500 ## number of samples
> x <- y <- numeric(N)
> x[1] <- y[1] <- 0
>
> for (i in 2:N) {
+   x[i] <- rnorm(1, mean=rho*y[i-1], sd=sqrt(1-rho^2))
+   y[i] <- rnorm(1, mean=rho*x[i], sd=sqrt(1-rho^2))
+ }
>
> plot(x,y, type="l")
```

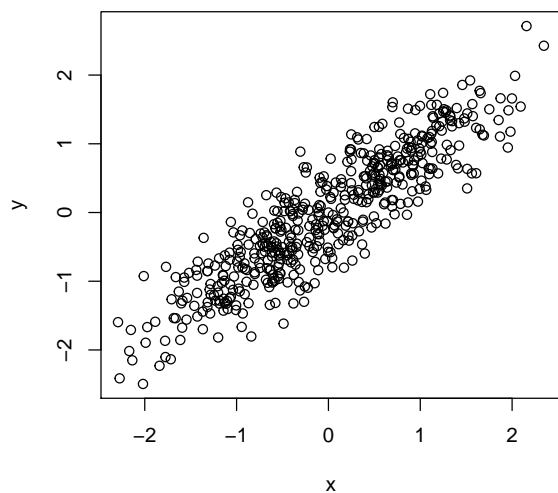
129

Gibbs Sampler



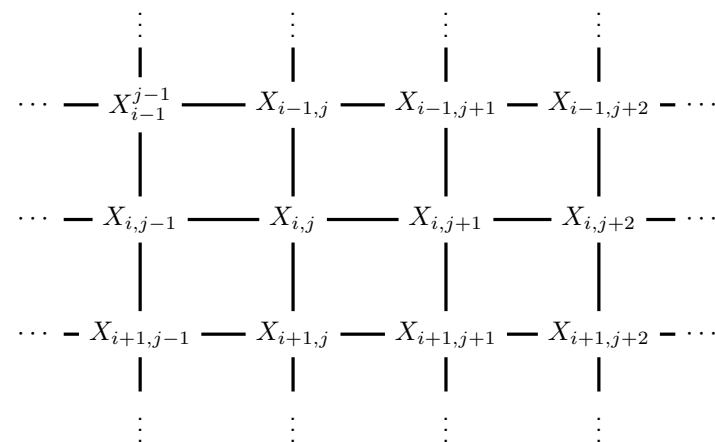
130

Gibbs Sampler



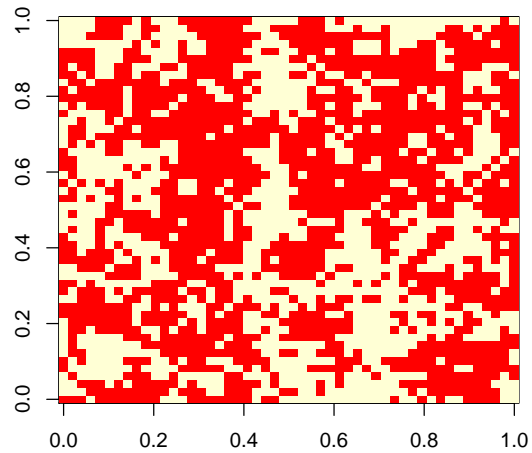
131

The Ising Model



132

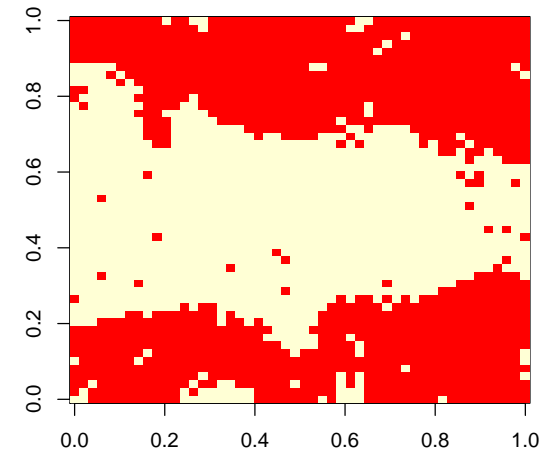
The Ising Model



50 × 50 grid, sample from $\theta = 0.15$.

133

The Ising Model



50 × 50 grid, sample from $\theta = 0.25$.

134

The Ising Model: Code

```
> ## function to perform Gibbs updates
> iterate = function(x, N, theta=0.5) {
+   n1 <- nrow(x); n2 <- ncol(x)

+   for (it in 1:N) {
+     for (i in 1:n1) for (j in 1:n2) {
+       rw <- (max(1,i-1):min(n1,i+1))
+       cl <- (max(1,j-1):min(n2,j+1))
+       cur <- sum(x[rw,cl]) - x[i,j]
+       prob = exp(cur*theta)/c(exp(cur*theta) + exp(-cur*theta))
+       x[i,j] <- 2*rbinom(1,1,prob)-1
+     }
+   }
+   x
+ }
```

135

The Ising Model: Code

```
> ## generate data set
> set.seed(123)
> n <- 50; theta = 0.25
> x <- matrix(2*rbinom(n^2,1,.5)-1, n, n)
> x = iterate(x,100, theta=theta)
> image(x)
```

136