SC6/SM9 Graphical Models

Michaelmas Term, 2018

Robin Evans

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

November 29, 2018

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.

Course Information

There will be four problem sheets and four associated classes.

Part C/OMMS students, your classes are weeks 2, 5, 8 and HT1. Sign up online for one of the two sessions (Wednesday 9am or Thursday 2pm).

Hand in your work by Monday, 5pm.

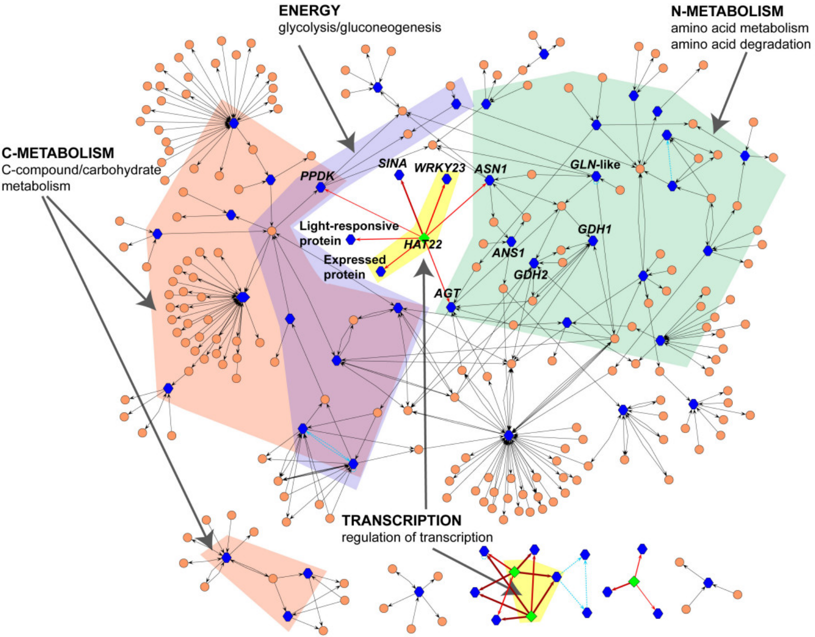
MSc students, classes are on Thursdays, weeks 2, 5, 8 and HT1 in here (LG.01) at 4pm (*except* for week 5, 11am)

Books

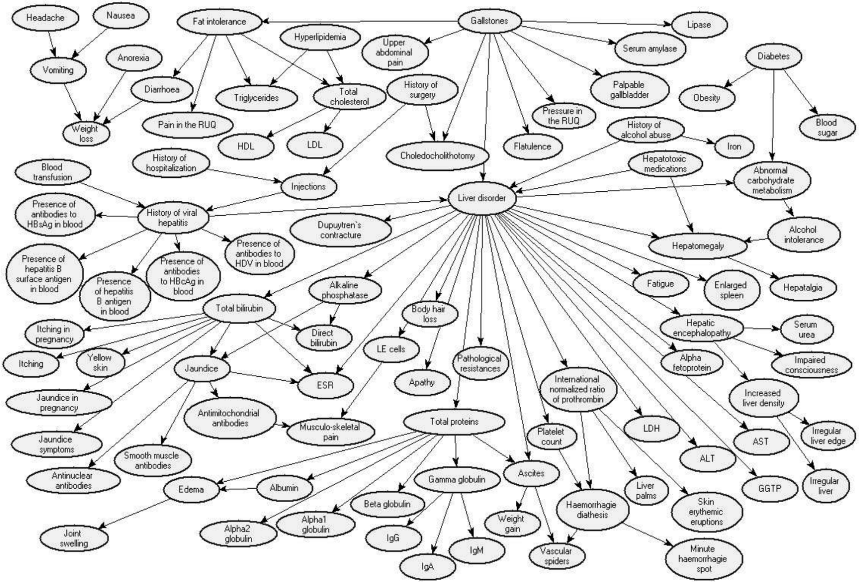
These books might be useful.

- 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.
- Agresti (2002). *Categorical Data Analysis*, (2nd edition), John Wiley & Sons.

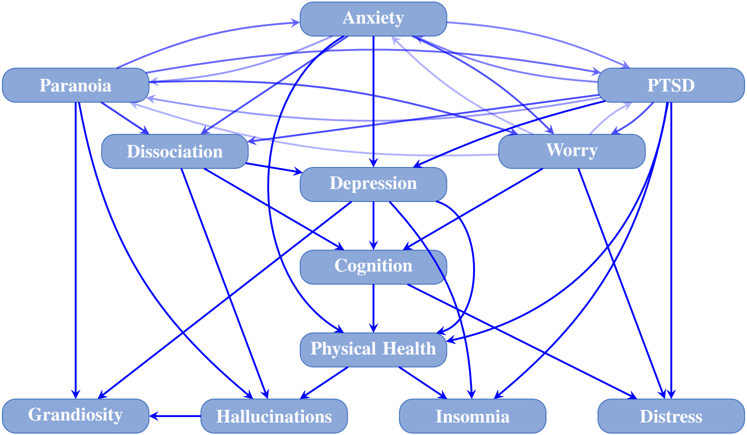
Gene Regulatory Networks



Medical Diagnosis



Mental Health



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:

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

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

Conditional Independence

Simpson's Paradox

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

Simpson's Paradox

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

Contingency Tables: Some Notation

We will consider multivariate systems of vectors $X_V \equiv (X_v : v \in V)$ for some set $V = \{1, \dots, p\}$.

Write $X_A \equiv (X_v : v \in A)$ for any $A \subseteq V$.

We assume that each $X_v \in \{1, \dots, d_v\}$ (usually $d_v = 2$).

If we have n i.i.d. observations write

$$X_V^{(i)} \equiv (X_1^{(i)}, \dots, X_p^{(i)})^T, \qquad i = 1, \dots, n.$$

Contingency Tables: Some Notation

We typically summarize categorical data by counts:

aspirin	heart attack
Y	N
Y	Y
N	N
N	N
Y	N
$\vdots$	$\vdots$

	heart attack	
	Y	N
no aspirin	28	656
aspirin	18	658

Write

$$n(x_V) = \sum_{i=1}^{n} \mathbb{1}\{X_1^{(i)} = x_1, \dots, X_p^{(i)} = x_p\}$$

A **marginal table** only counts some of the variables.

$$n(x_A) = \sum_{x_{V \setminus A}} n(x_A, x_{V \setminus A}).$$

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

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
```

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.)

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
```

Undirected Graphical Models

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
```

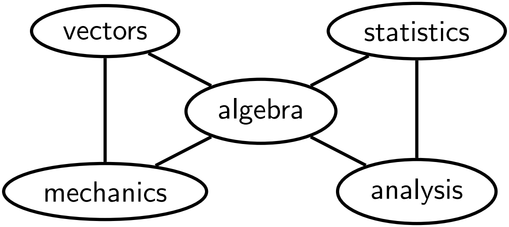
Multivariate Data

```
> 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
```

Multivariate Data

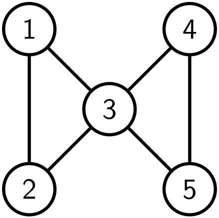
```
> conc <- solve(cov(marks))  # concentration matrix
> round(1000*conc, 2)

            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
```

Undirected Graphs



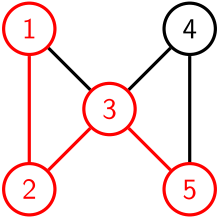
	mech	vecs	alg	anlys	stats
mechanics	5.24	-2.43	-2.72	0.01	-0.15
vectors	-2.43	10.42	-4.72	-0.79	-0.16
algebra	-2.72	-4.72	26.94	-7.05	-4.70
analysis	0.01	-0.79	-7.05	9.88	-2.02
statistics	-0.15	-0.16	-4.70	-2.02	6.45

Undirected Graphs



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

Paths

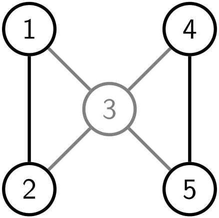


Paths:

$$\pi_1 : 1 - 2 - 3 - 5$$
$$\pi_2 : 3$$

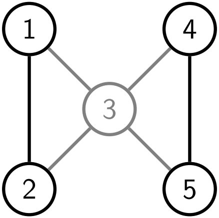
Note that paths may consist of one vertex and no edges.

Induced Subgraph



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

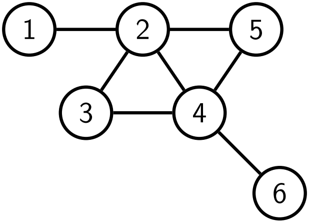
Separation



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

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

Cliques and Running Intersection



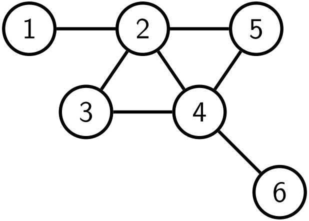
Cliques:

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

Separator sets:

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

Cliques and Running Intersection



A different ordering of the cliques:

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

Separator sets:

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

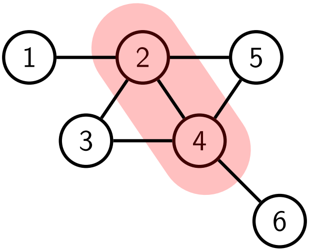
Any ordering works in this case as long $\{1,2\}$ and $\{4,6\}$ aren't the first two entries.

Estimation

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

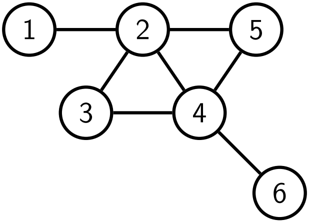
$$X_1, X_3 \perp\!\!\!\perp X_5, X_6 \mid X_2, X_4$$

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



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

Estimation



Repeating this process on each subgraph we obtain:

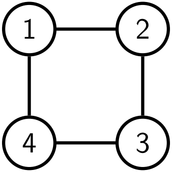
$$p(x_{123456}) \cdot p(x_{24}) \cdot p(x_2) \cdot p(x_4) = p(x_{12}) \cdot p(x_{234}) \cdot p(x_{245}) \cdot p(x_{46}).$$

i.e.

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

Non-Decomposable Graphs

But can't we do this for any factorization?



No! Although

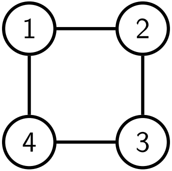
$$p(x_{1234}) = \psi_{12}(x_{12}) \cdot \psi_{23}(x_{23}) \cdot \psi_{34}(x_{34}) \cdot \psi_{14}(x_{14}),$$

the ψs are constrained by the requirement that

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

There is no nice representation of the ψ_Cs in terms of p.

Non-Decomposable Graphs



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

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

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

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 $\max_i \max_{x_{C_i}} |p(x_{C_i}) - q(x_{C_i})| >$ tol **do**
        for i in $1, \ldots, k$ **do**
            update $p(x_V)$ to $p(x_{V \setminus C_i} \mid x_{C_i}) \cdot q(x_{C_i})$;
        end for
    end while
    return distribution p with margins $p(x_{C_i}) \approx q(x_{C_i})$.
end function

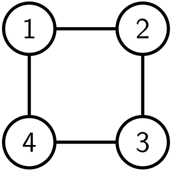
If any distribution satisfying $p(x_{C_i}) = q(x_{C_i})$ for each $i = 1, \ldots, k$ exists, then the algorithm converges to the **unique distribution** with those margins and which is Markov with respect to the graph with cliques $C_1, \ldots, C_k$.

Some Data

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	9	9	0	8
	1	6	4	4	3
$X_4 = 1$	0	22	0	2	6
	1	5	3	10	5

Margins

Suppose we want to fit the 4-cycle model:



The relevant margins are:

$n(x_{12})$	$X_2 = 0$	1
$X_1 = 0$	42	16
1	16	22

$n(x_{23})$	$X_3 = 0$	1
$X_2 = 0$	40	18
1	16	22

$n(x_{34})$	$X_4 = 0$	1
$X_3 = 0$	26	30
1	17	23

$n(x_{14})$	$X_4 = 0$	1
$X_1 = 0$	19	39
1	24	14

Start with a Uniform Table

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	6	6	6	6
	1	6	6	6	6
$X_4 = 1$	0	6	6	6	6
	1	6	6	6	6

Set Margin X_1, X_2 to Correct Value

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	10.5	4	4	5.5
	1	10.5	4	4	5.5
$X_4 = 1$	0	10.5	4	4	5.5
	1	10.5	4	4	5.5

Replace

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

Set Margin X_2, X_3 to Correct Value

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	14.48	5.52	3.37	4.63
	1	6.52	2.48	4.63	6.37
$X_4 = 1$	0	14.48	5.52	3.37	4.63
	1	6.52	2.48	4.63	6.37

Replace

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

Set Margin X_3, X_4 to Correct Value

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	13.45	5.12	3.13	4.3
	1	5.54	2.11	3.94	5.41
$X_4 = 1$	0	15.52	5.91	3.61	4.96
	1	7.49	2.86	5.33	7.32

Replace

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

Set Margin X_1, X_4 to Correct Value

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	9.81	7.26	2.28	6.09
	1	4.04	2.99	2.87	7.67
$X_4 = 1$	0	18.94	3.93	4.41	3.3
	1	9.15	1.9	6.5	4.87

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 41.94.

Set Margin X_1, X_2 to Correct Value

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	9.82	7.27	2.28	6.1
	1	4.02	2.97	2.86	7.63
$X_4 = 1$	0	18.87	3.92	4.39	3.29
	1	9.18	1.91	6.52	4.89

Eventually:

Waiting for this process to converge leads to the MLE:

		$X_2 = 0$		$X_2 = 1$	
		$X_1 = 0$	1	0	1
$X_4 = 0$	$X_3 = 0$	10.07	7.41	2.29	6.23
	1	3.87	2.85	2.77	7.51
$X_4 = 1$	0	18.7	3.83	4.26	3.22
	1	9.36	1.91	6.68	5.04

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.

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

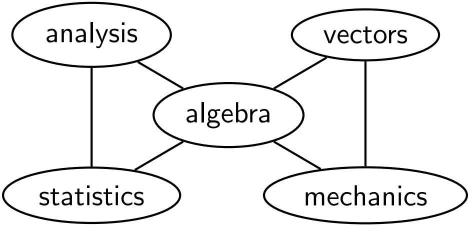
The log-likelihood for Σ is

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

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

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}

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.:

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

Equating terms gives:

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

Maximum Likelihood Estimation

Iterating this process with a decomposable graph shows that:

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

For maximum likelihood estimation, using Theorem 4.24 we have

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

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

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
```

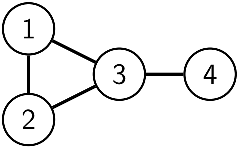
Example

```
> round(W, 3)          # sample covariance

       [,1]   [,2]   [,3]   [,4]
[1,]  1.158 -0.374 -0.242  0.036
[2,] -0.374  1.099  0.227 -0.065
[3,] -0.242  0.227  1.169 -0.378
[4,]  0.036 -0.065 -0.378  1.085

> round(solve(W), 3)   # sample concentration

      [,1]   [,2]   [,3]  [,4]
[1,] 0.995  0.308  0.160 0.040
[2,] 0.308  1.044 -0.138 0.004
[3,] 0.160 -0.138  1.026 0.344
[4,] 0.040  0.004  0.344 1.040
```

Note that these are fairly close to the true values.

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,] 0.993  0.308  0.146 0.000
[2,] 0.308  1.044 -0.139 0.000
[3,] 0.146 -0.139  1.021 0.336
[4,] 0.000  0.000  0.336 1.038
```

Note this is close to the true concentration matrix.

Directed Graphical Models

Directed Graphs

The graphs considered so far are all **undirected**. Directed graphs give each edge an orientation.

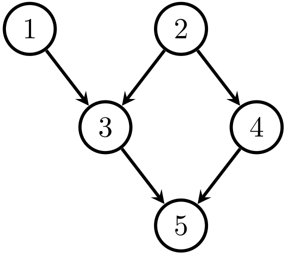
A **directed graph** $\mathcal{G}$ is a pair (V, D), where

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

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

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

If $i \to j$ or $i \leftarrow j$ we say i and j are **adjacent** and write $i \sim j$.



Acyclicity

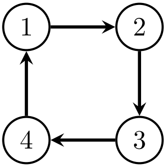
Paths are sequences of adjacent vertices, without repetition:

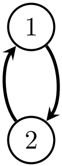
$$1 \to 3 \leftarrow 2 \to 4 \to 5 \qquad\qquad 1 \to 3 \to 5.$$

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 $j \to i$.





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

All the directed graphs we consider are acyclic.

Happy Families

$$i \to j \quad \begin{cases} i \in \mathrm{pa}_{\mathcal{G}}(j) & i \text{ is a parent of } j \\ j \in \mathrm{ch}_{\mathcal{G}}(i) & j \text{ is a child of } i \end{cases}$$

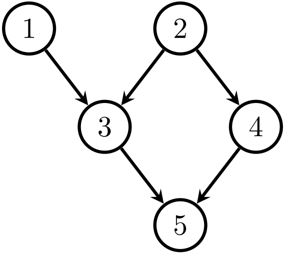
$$\begin{array}{c} a \to \cdots \to b \\ \text{or } a = b \end{array} \quad \begin{cases} a \in \mathrm{an}_{\mathcal{G}}(b) & a \text{ is an ancestor of } b \\ b \in \mathrm{de}_{\mathcal{G}}(a) & b \text{ is a descendant of } a \end{cases}$$

If $w \notin \mathrm{de}_{\mathcal{G}}(v)$ then w is a **non-descendant** of v:

$$\mathrm{nd}_{\mathcal{G}}(v) = V \setminus \mathrm{de}_{\mathcal{G}}(v).$$

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

Examples



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

$$\mathrm{ch}_{\mathcal{G}}(5) = \emptyset \qquad \mathrm{de}_{\mathcal{G}}(1) = \{1, 3, 5\}$$

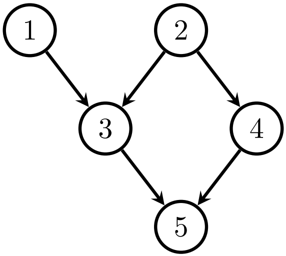
$$\mathrm{nd}_{\mathcal{G}}(1) = \{2, 4\}.$$

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:

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



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;

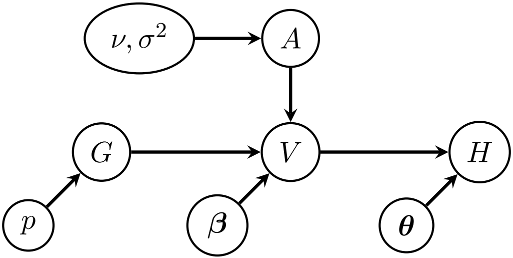
$$
\begin{array}{ccccc}
 & & A & & \\
 & & \downarrow & & \\
G & \rightarrow & V & \rightarrow & H
\end{array}
$$

Parameter Estimation

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

$$
\begin{aligned}
\text{group} &: G_i \sim \text{Bernoulli}(p); \\
\text{age} &: A_i \sim N(\nu, \sigma^2); \\
\text{vaccine} &: V_i \mid A_i, G_i \sim \text{Bernoulli}(\mu_i) \text{ where} \\
&\qquad \operatorname{logit} \mu_i = \beta_0 + \beta_1 A_i + \beta_2 G_i. \\
\text{hospital} &: H_i \mid V_i \sim \text{Bernoulli}(\operatorname{expit}(\theta_0 + \theta_1 V_i)).
\end{aligned}
$$

Assuming independent priors:

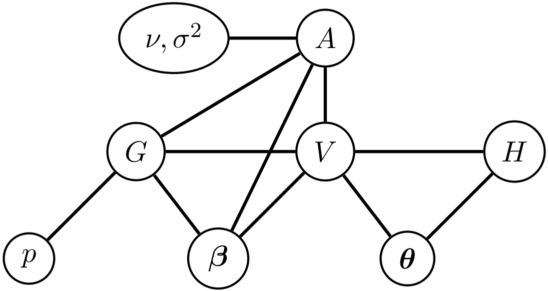


Bayesian Inference

From our argument, we have

$$\begin{aligned}\pi(\boldsymbol{\beta} \mid G, A, V, H) &= \pi(\boldsymbol{\beta} \mid G, A, V) \\ &\propto p(V \mid A, G, \boldsymbol{\beta}) \cdot \pi(\boldsymbol{\beta}).\end{aligned}$$

Looking at the moral graph we see

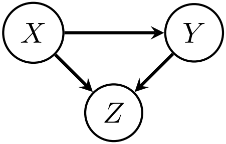


Markov Equivalence

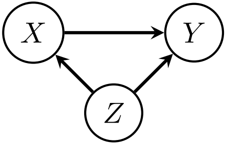
All undirected graphs induce distinct models.

$$v \not\sim w \qquad \Longleftrightarrow \qquad X_v \perp\!\!\!\perp X_w \mid X_{V \setminus \{v,w\}} \text{ implied}$$

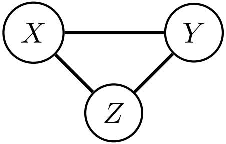
The same is not true for directed graphs:



$p(x) \cdot p(y \mid x) \cdot p(z \mid x, y)$

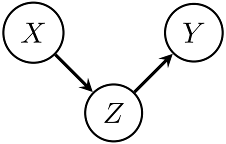


$p(z) \cdot p(x \mid z) \cdot p(y \mid x, z)$

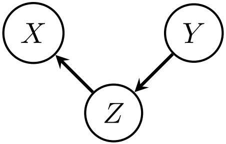


$\psi_{XYZ}(x, y, z)$

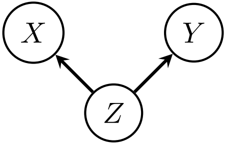
Markov Equivalence



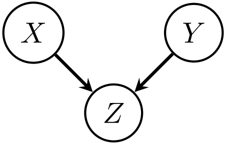
$$p(x) \cdot p(z \mid x) \cdot p(y \mid z)$$
$$X \perp\!\!\!\perp Y \mid Z$$



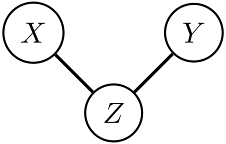
$$p(y) \cdot p(z \mid y) \cdot p(x \mid z)$$
$$X \perp\!\!\!\perp Y \mid Z$$



$$p(z) \cdot p(x \mid z) \cdot p(y \mid z)$$
$$X \perp\!\!\!\perp Y \mid Z$$



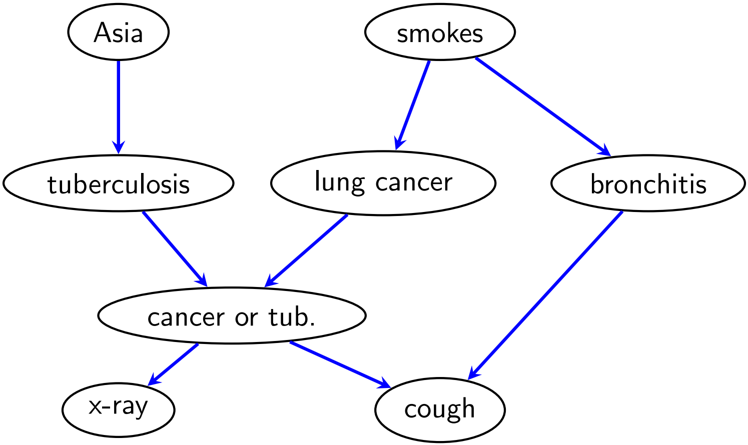
$$p(x) \cdot p(y) \cdot p(z \mid x, y)$$
$$X \perp\!\!\!\perp Y$$



$$\psi_{XZ}(x, z) \cdot \psi_{YZ}(y, z)$$
$$X \perp\!\!\!\perp Y \mid Z$$

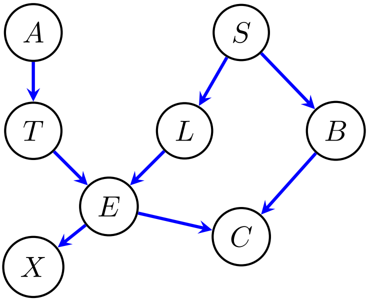
Expert Systems

Expert Systems



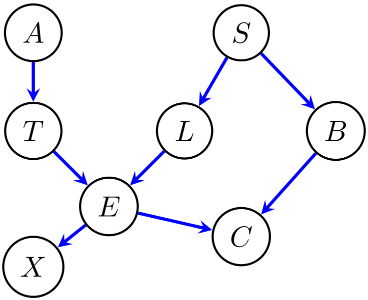
The ‘Chest Clinic’ network, a fictitious diagnostic model.

Variables



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?

Conditional Probability Tables



We have our factorization:

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

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

Probabilities

$p(a) =$

yes	no
0.01	0.99

$p(s) =$

yes	no
0.5	0.5

$p(t \mid a) =$

A	yes	no
yes	0.05	0.95
no	0.01	0.99

$p(l \mid s) =$

S	yes	no
yes	0.1	0.9
no	0.01	0.99

$p(b \mid s) =$

S	yes	no
yes	0.6	0.4
no	0.3	0.7

$p(x \mid e) =$

E	yes	no
yes	0.98	0.02
no	0.05	0.95

$p(c \mid b, e) =$

B	E	yes	no
yes	yes	0.9	0.1
	no	0.8	0.2
no	yes	0.7	0.3
	no	0.1	0.9

Factorizations

$$p(l \mid x, c, a, s) = \frac{p(l, x, c \mid a, s)}{\sum_{l'} p(l', x, c \mid a, s)}$$

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

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

But this is:

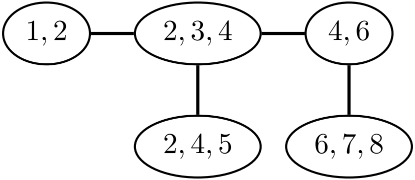
$$p(l \mid s) \sum_{e} p(x \mid e) \left(\sum_{b} p(b \mid s) \cdot p(c \mid e, b) \right) \left(\sum_{t} p(t \mid a) \cdot p(e \mid t, l) \right).$$

Junction Trees

A **junction tree**:

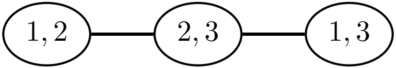
- is a (connected) undirected graph without cycles (a tree);
- has vertices C_i that consist of **subsets** of a set V;
- satsifies the property that if $C_i \cap C_j = S$ then every vertex on the (unique) path from C_i to C_j contains S.

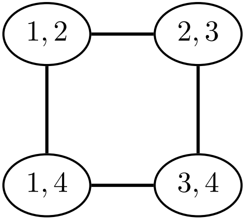
Example.



Junction Trees

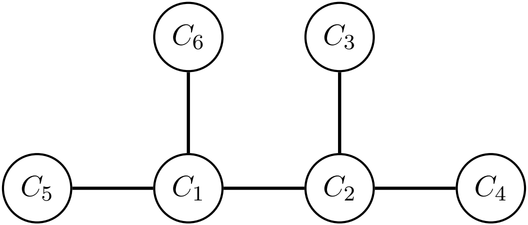
The following graphs are **not** junction trees:





Junction Trees

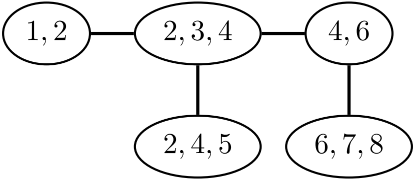
Junction trees can be constructed directly from sets of cliques satisfying running intersection.



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

Example: Junction Trees and RIP

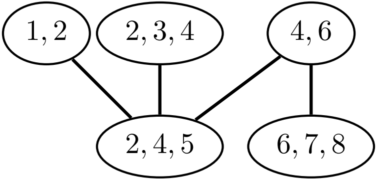
Given sets $\{1,2\}$, $\{2,3,4\}$, $\{2,4,5\}$, $\{4,6\}$, $\{6,7,8\}$, we can build this tree:



Example: Junction Trees and RIP

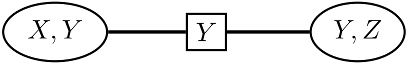
Equally, we could use a different ordering:

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



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
	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
	1	0.2	0.4

Initialize with

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

Updating / Message Passing

Suppose we have two vertices and one separator set.

$$
\begin{tikzcd}
(X,Y) \arrow[r] & \boxed{Y} \arrow[r] & (Y,Z)
\end{tikzcd}
$$

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

		$y=0$	1
x	0	0.3	0.9
	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
	1	0.2	0.4

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

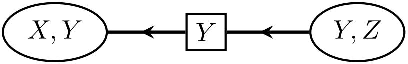
$$\psi'_Y(y) = \sum_x \psi_{XY}(x,y) = (1,1);$$

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

So in this case nothing changes.

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
	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
	1	0.2	0.4

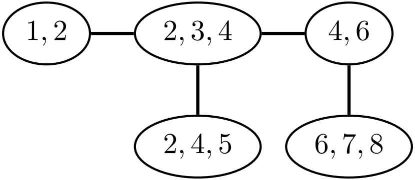
Pass message from Y, Z to X, Y. We set

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

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

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

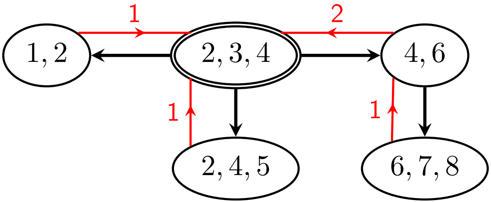
Rooting



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

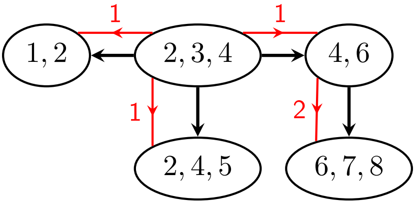
Collection and Distribution

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

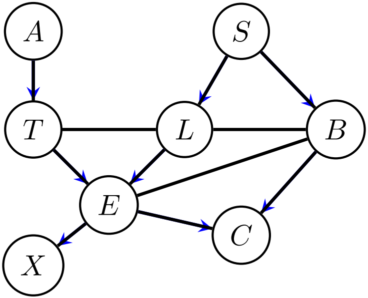


Collection and Distribution

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



Forming A Junction Tree



Steps to Forming a Junction Tree:

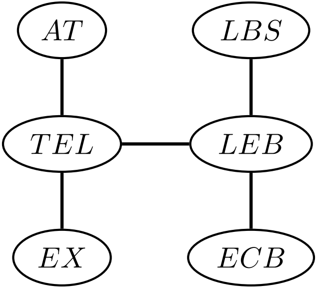
Moralize

Drop directions

Triangulate (add edges to get a decomposable graph)

Forming A Junction Tree

Finally, form the tree of cliques.



Initialization

$p(a) =$

yes	no
0.01	0.99

$p(s) =$

yes	no
0.5	0.5

$p(t \mid a) =$

A	yes	no
yes	0.05	0.95
no	0.01	0.99

$p(l \mid s) =$

S	yes	no
yes	0.1	0.9
no	0.01	0.99

$p(b \mid s) =$

S	yes	no
yes	0.6	0.4
no	0.3	0.7

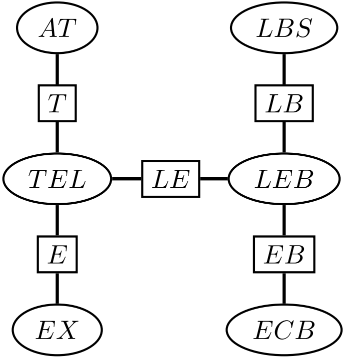
$p(x \mid e) =$

E	yes	no
yes	0.98	0.02
no	0.05	0.95

$p(c \mid b, e) =$

B	E	yes	no
yes	yes	0.9	0.1
	no	0.8	0.2
no	yes	0.7	0.3
	no	0.1	0.9

Initialization

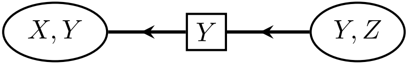


Can set, for example:

$$\psi_{AT}(a,t) = p(a)\cdot p(t \mid a) \qquad \psi_{LBS}(l,b,s) = p(s)\cdot p(l \mid s)\cdot p(b \mid s)$$
$$\psi_{TEL}(t,e,l) = p(e \mid t,l) \qquad \psi_{ELB}(e,l,b) = 1$$
$$\psi_{EX}(e,x) = p(x \mid e) \qquad \psi_{ECB}(e,c,b) = p(c \mid e,b).$$

Evidence

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



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

		$y = 0$	1
x	0	0.12	0.54
	1	0.28	0.06

$\psi_Y(y)$

$y = 0$	1
0.4	0.6

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

		$z = 0$	1
y	0	0.6	0
	1	0.4	0

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

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

$$\psi_Y(y) = \sum_x \psi_{YZ}(y, z) = (0.6, 0.4);$$

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

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

From the Chest Clinic Network

Conditional Probability Tables:

$p(x \mid e)$:

E	yes	no
yes	0.98	0.02
no	0.05	0.95

$p(c \mid e, b)$:

E	B	yes	no
yes	yes	0.9	0.1
	no	0.7	0.3
no	yes	0.8	0.2
	no	0.1	0.9

From the Chest Clinic Network

Marginal Probability Tables:

ψ_{EX} :

E	yes	no
yes	0.064	0.0013
no	0.047	0.89

ψ_{AT} :

A	yes	no
yes	0.0005	0.0095
no	0.0099	0.98

ψ_{LBS} :

L	B	yes	no
yes	yes	0.03	0.0015
	no	0.02	0.0035
no	yes	0.27	0.15
	no	0.18	0.35

ψ_{LEB} :

L	E	yes	no
yes	yes	0.032	0.024
	no	0	0
no	yes	0.0044	0.0055
	no	0.41	0.52

ψ_{TEL} :

T	E	yes	no
yes	yes	0.00057	0
	no	0.0098	0
no	yes	0.054	0
	no	0	0.94

ψ_{ECB} :

B	E	yes	no
yes	yes	0.032	0.0036
	no	0.02	0.0087
no	yes	0.33	0.083
	no	0.052	0.47

From the Chest Clinic Network

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

ψ_{EX} :

E	yes	no
yes	0.58	-
no	0.42	-

ψ_{AT} :

A	yes	no
yes	0.0044	0.0087
no	0.088	0.9

ψ_{LBS} :

L	B	yes	no
yes	yes	0.27	0.013
	no	0.18	0.031
no	yes	0.15	0.08
	no	0.097	0.19

ψ_{LEB} :

L	E	yes	no
yes	yes	0.28	0.21
	no	0	0
no	yes	0.039	0.049
	no	0.19	0.24

ψ_{TEL} :

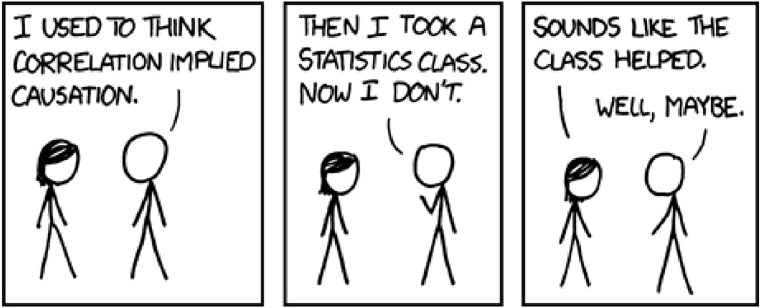
T	E	yes	no
yes	yes	0.0051	0
	no	0.087	0
no	yes	0.48	0
	no	0	0.42

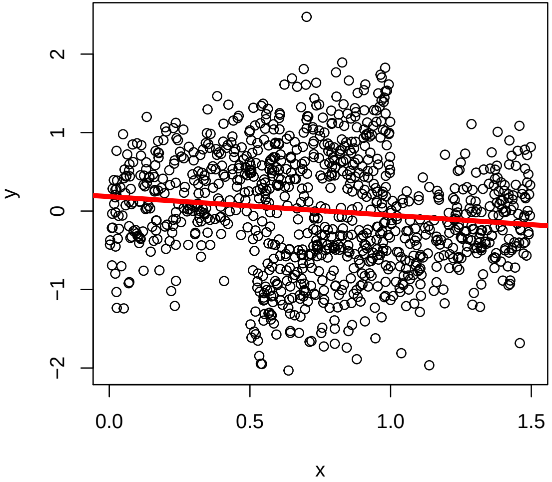
ψ_{ECB} :

B	E	yes	no
yes	yes	0.29	0.032
	no	0.18	0.077
no	yes	0.15	0.038
	no	0.024	0.21

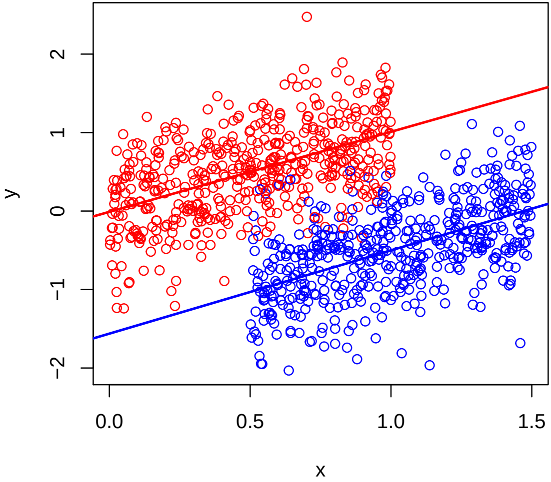
Causal Inference

Correlation



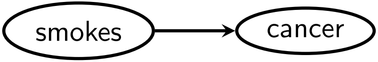

y
x
−2
−1
0
1
2
0.0
0.5
1.0
1.5

Controlling for Covariates



Causation

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

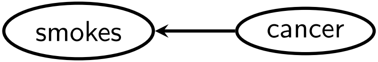


BUT: how do we know that this is a causal relationship? And what do we mean by that?

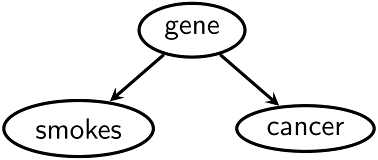
The central question is: "if we stop people from smoking, will they be less likely to get lung cancer?"

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

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.

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

$$P(D = 1 \mid S = 1) = \frac{5}{8} \qquad P(D = 1 \mid S = 0) = \frac{5}{32}.$$

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 estimate 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.

‘do’ notation

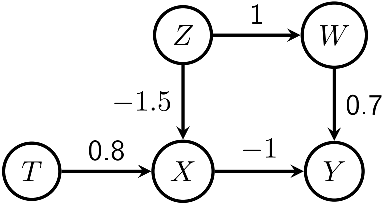
In this example there is a difference between predicting damage when we ‘observe’ that someone smokes ...

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

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

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

Linear Gaussian Causal Models




```
> 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)
```

Back-Door Paths

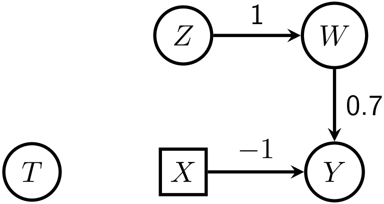
```
> summary(lm(Y ~ X))$coefficients[,1:2]

             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
```

Instruments

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

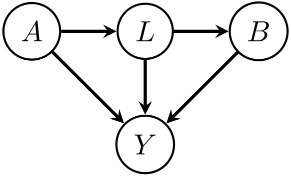
```
> summary(lm(Y ~ X + W + T))$coefficients[,1:2]

            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
```

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
```

Example: HIV Treatment



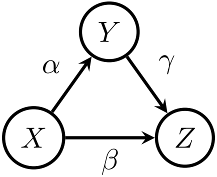
A treatment with AZT (an HIV drug);
L opportunisitic infection;
B treatment with antibiotics;
Y survival at 5 years.

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

Structural Equation Models

Covariance Matrices

Let $\mathcal{G}$ be a DAG with variables V.



$$X = \varepsilon_x \qquad\qquad Y = \alpha X + \varepsilon_y \qquad\qquad Z = \beta X + \gamma Y + \varepsilon_z.$$

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ \alpha & 0 & 0 \\ \beta & \gamma & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{pmatrix}.$$

Covariance Matrices

Rearranging:

$$\begin{pmatrix} 1 & 0 & 0 \\ -\alpha & 1 & 0 \\ -\beta & -\gamma & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{pmatrix}.$$

Now, you can check that:

$$(I-B)^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -\alpha & 1 & 0 \\ -\beta & -\gamma & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ \beta + \alpha\gamma & \gamma & 1 \end{pmatrix},$$

so

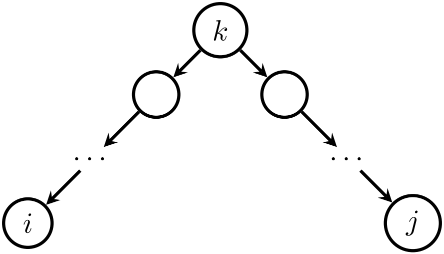
$$\begin{aligned} \Sigma &= (I-B)^{-1}(I-B)^{-T} \\ &= \begin{pmatrix} 1 & \alpha & \beta + \alpha\gamma \\ \alpha & 1 + \alpha^2 & \alpha\beta + \gamma + \alpha^2\gamma \\ \beta + \alpha\gamma & \alpha\beta + \gamma + \alpha^2\gamma & 1 + \gamma^2 + \beta^2 + 2\alpha\beta\gamma + \alpha^2\gamma^2 \end{pmatrix}. \end{aligned}$$

Treks

Let $\mathcal{G}$ be a DAG with variables V.

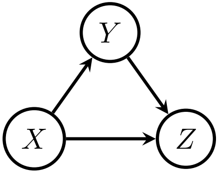
A **trek** from i to j with source k is a pair (π_l, π_r) of directed paths.

- π_l (the **left side**) is directed from k to i;
- π_r (the **right side**) is directed from k to j.



Trek Examples

Consider this DAG:



The treks from Z to Z are:

$$Z \qquad Z \leftarrow Y \rightarrow Z$$
$$Z \leftarrow X \rightarrow Z \qquad Z \leftarrow Y \leftarrow X \rightarrow Z$$
$$Z \leftarrow X \rightarrow Y \rightarrow Z \qquad Z \leftarrow Y \leftarrow X \rightarrow Y \rightarrow Z.$$

Note that:

- A vertex may be in both the left and right sides.
- We may have $i = k$ or $j = k$ or both.

Treks

Let Σ be Markov with respect to a DAG $\mathcal{G}$, so that

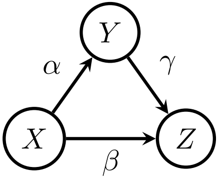
$$\Sigma = (I - B)^{-1} D (I - B)^{-T}.$$

Let $\tau = (\pi_l, \pi_r)$ be a trek with source k. The **trek covariance** associated with τ is:

$$c(\tau) = d_{kk} \left(\prod_{(i \to j) \in \pi_l} b_{ji} \right) \left(\prod_{(i \to j) \in \pi_r} b_{ji} \right).$$

Trek Covariance Examples

Consider this DAG:



Trek covariances include:

$$c(Z) = 1 \qquad c(Z \leftarrow X) = \beta$$
$$c(Z \leftarrow X \to Y \to Z) = \beta \cdot \alpha \cdot \gamma \qquad c(Y \to Z) = \gamma.$$

Note that an empty product is 1 by convention.

Covariance Matrices

$$
\begin{tikzcd}
 & Y \arrow[dr, "\gamma"] & \\
X \arrow[ur, "\alpha"] \arrow[rr, "\beta"'] & & Z
\end{tikzcd}
$$

$$
\begin{aligned}
&Z && Z \leftarrow Y \rightarrow Z \\
&Z \leftarrow X \rightarrow Z && Z \leftarrow Y \leftarrow X \rightarrow Z \\
&Z \leftarrow X \rightarrow Y \rightarrow Z && Z \leftarrow Y \leftarrow X \rightarrow Y \rightarrow Z.
\end{aligned}
$$

Recall that

$$\sigma_{zz} = 1 + \gamma^2 + \beta^2 + 2\alpha\beta\gamma + \alpha^2\gamma^2.$$

Theorem 8.15 (The Trek Rule)
Let $\mathcal{G}$ be a DAG and let X_V be Gaussian and Markov with respect to $\mathcal{G}$. Then

$$\sigma_{ij} = \sum_{\tau \in \mathcal{T}_{ij}} c(\tau),$$

where $\mathcal{T}_{ij}$ is the set of treks from i to j.

That is, the covariance between each X_i and X_j is the sum of the trek covariances over all treks between i and j.

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

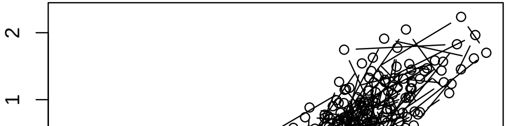
$$K = \Sigma^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}.$$

Then

$$X_1 \mid X_2 = x_2 \sim N\left(\rho x_2, (1-\rho)^2\right)$$
$$X_2 \mid X_1 = x_1 \sim N\left(\rho x_1, (1-\rho)^2\right)$$

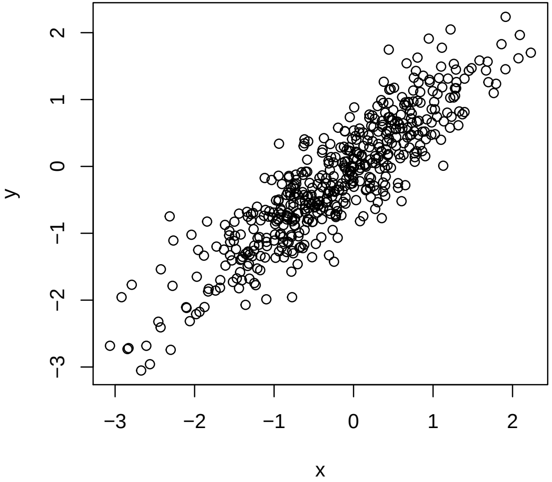
Gibbs Sampler

```
> ## 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="b")
```



Gibbs Sampler

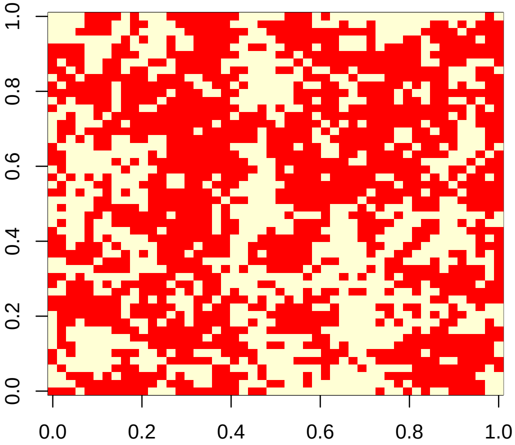
Gibbs Sampler



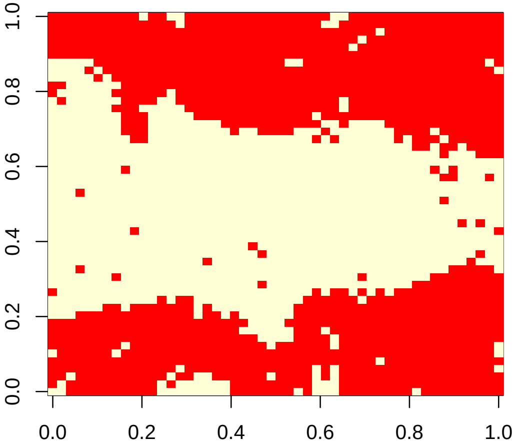
The Ising Model

$$
\begin{tikzcd}[arrows=dash]
 & \vdots \arrow[d] & \vdots \arrow[d] & \vdots \arrow[d] & \vdots \arrow[d] & \\
\cdots \arrow[r] & X_{i-1}^{j-1} \arrow[r] \arrow[d] & X_{i-1,j} \arrow[r] \arrow[d] & X_{i-1,j+1} \arrow[r] \arrow[d] & X_{i-1,j+2} \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & X_{i,j-1} \arrow[r] \arrow[d] & X_{i,j} \arrow[r] \arrow[d] & X_{i,j+1} \arrow[r] \arrow[d] & X_{i,j+2} \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & X_{i+1,j-1} \arrow[r] \arrow[d] & X_{i+1,j} \arrow[r] \arrow[d] & X_{i+1,j+1} \arrow[r] \arrow[d] & X_{i+1,j+2} \arrow[r] \arrow[d] & \cdots \\
 & \vdots & \vdots & \vdots & \vdots &
\end{tikzcd}
$$

The Ising Model



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



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

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
+ }
```

The Ising Model: Code

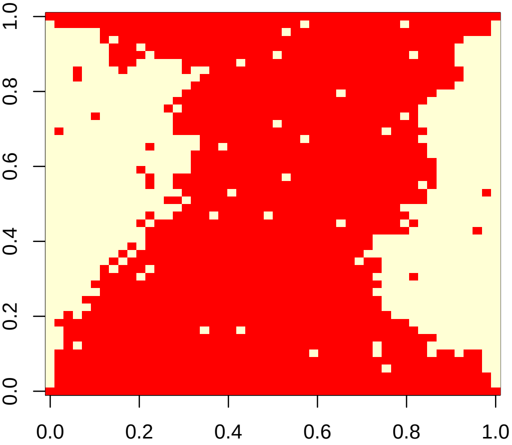
```
> ## generage data set
> set.seed(123)
> n <- 50; theta = 0.25
> x <- matrix(2*rbinom(n^2,1,.5)-1, n, n)
> x = iterate(x,100, theta=theta)
> image(x)
```

Introducing Evidence

Suppose that we know the border of the picture is all 1s.

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

+   for (it in 1:N) {
+     ## reduce scope of loop
+     for (i in 2:(n1-1)) for (j in 2:(n2-1)) {
+       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
+ }
```

Introducing Evidence



Chest Clinic