# Graphical Markov Models

3 Lectures

Giovanni M. Marchetti
2024

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

Markov Properties

Conditional Indep. tests
Examples

2 DAGs and Regression chain graphs
MLE for Gaussian distributions

Examples

3 MLE for categorical (binary) data
Structure learning

Examples

# LECTURE 1

Summary

- Conditional independence
- Properties
- Graphical Markov Models

  Directed acyclic graphs

  Undirected graphs

  Regression graphs

References


- Højsgaard, Edwards, Lauritzen (2012) Graphical Models in R, Springer V.
- Scutari, Denis (2022) Bayesian Networks CRC Press.
- Cox, Wermuth (1996) Multivariate dependencies, Chapman & Hall.
- Roverato (2017) Graphical models for categorical variables.
- Whittaker (1990) Graphical models in applied multivariate statistics Wiley.

## Useful R packages

bnlearn (Scutari et al)

gRbase (Højsgaard et al)
gRim

ggm (Marchetti et al)

ct2 (J. Laug)

- Look at the CRAN Task View

https://CRAN.R-project.org/view=GraphicalModels

## Other References.

J. Pearl et al (2016) Causal Inference in Statistics. Wiley.

Hernán, Robins (2020). What if. CRC Press.

- Look at the CRAN Task View

https://cran.r-project.org/web/views/CausalInference.html

# Data

We will consider a system of $d$ variables

$$\underset{\sim}{X} = (X_1, X_2, X_3, \ldots, X_d)$$

can be
- continuous (Gaussian)
- categorical or binary.
- mixed

**Example**
- Gaussian $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma)$
- multinomial $\underset{\sim}{X} \sim \text{Mult}(n, \underset{\sim}{\pi})$

**Exercise 1**

Simulate 1000 observations from 3 binary variables $X_1\ X_2\ X_3$ with a joint prob.

$$\underset{\sim}{\pi} = (0.38, 0.01, 0.07, 0.01, 0.42, 0.02, 0.08, 0.01)$$

**Exercise 2**

Simulate 100 observations from $N_3(\underline{\mu}, \underline{\Sigma})$ with

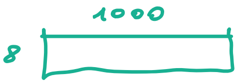
$$\Sigma^{-1} = \begin{bmatrix} 1.2 & 0.9 & 0 \\ & 2.7 & 1.4 \\ & & 1.6 \end{bmatrix}$$

# Solutions

## Exercise 1

$\underset{\sim}{\pi} > 0$

$$\begin{matrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{matrix}$$




```
n = 1000
p <- c(0.38, 0.01, 0.07, 0.01,
       0.42, 0.02, 0.08, 0.01)
X<- expand.grid(X1 = factor(0:1),
                X2 = factor(0:1),
                X3 = factor(0:1))
Z <- rmultinom(n, size = 1, prob = p)
cell <- apply(Z, 2, function(x) which(x==1))
data <- X[cell,]
rownames(data) <- 1:n

head(data)
tail(data)

table(data)
as.data.frame(table(data))
```

## Exercise 2

```
library(mnormt)
K <- matrix(c(1.2, 0.9,   0,
              0.9, 2.7, 1.4,
              0,   1.4, 1.6), 3, 3)
dimnames(K) <- list(c("X1","X2", "X3"),
                    c("X1","X2", "X3"))
Sigma <- solve(K)
dimnames(Sigma) <- list(c("X1","X2", "X3"),
                        c("X1","X2", "X3"))
round(Sigma, 2)
round(cov2cor(Sigma),2)
X <- rmnorm(n = 100, varcov = Sigma)
```

# Conditional independence

Random vector $(X_1, \dots, X_d)$ $\nearrow$ all contin. $\searrow$ all discrete

Denote the joint distribution by

$$p_V(x_1, \dots, x_d) \begin{array}{l} \nearrow \text{pmf} \\ \searrow \text{pdf} \end{array}$$

where

$$V = \{1, 2, \dots, d\}$$

## Example

If $(X_1, X_2)$ $\quad p_{12}(x_1, x_2)$

## Conditional Distributions

$X_1 \mid X_2 \qquad p_{1|2}(x_1 | x_2)$

$X_1 \mid X_2 \; X_3 \qquad p_{1|23}(x_1 | x_2, x_3)$

## Conditional independence

Given 3 variables $X_1, X_2, X_3$

$X_1$ conditionally independent of $X_2$ given $X_3$

if

$$P_{123}(x_1, x_2 | x_3) = P_{1|3}(x_1 | x_3) P_{2|3}(x_2, x_3)$$

for all $(x_1, x_2, x_3)$ such that $p_3(x_3) > 0$

## Properties of CI

Notation $X_1 \perp\!\!\!\perp X_2 \mid X_3$

abbreviated $1 \perp\!\!\!\perp 2 \mid 3$

Equivalent (omit the suffix)

$$p(x_1 \mid x_2\, x_3) = p(x_1 \mid x_3) \quad p(x_2, x_3) > 0$$

Factorizations

$$p(x_1, x_2, x_3) = p(x_1 \mid x_3)\, p(x_2 \mid x_3)\, p(x_3)$$

$$= p(x_1 \mid x_3)\, p(x_3 \mid x_2)\, p(x_2)$$

Marginal independence

$X_1 \perp\!\!\!\perp X_2$ if $p_{12}(x_1, x_2) = p_1(x_1)\, p_2(x_2)$

for all $x_1, x_2$.

It is a quite different constraint.

## Exercise 3

Two binary variables have marginal probabilities

$p_1(x_1) = \{0.2, 0.8\}$ $\quad p_2(x_2) = \{0.6, 0.4\}$

and odds-ratio = 1 $\quad odr = \dfrac{\pi_{11}\ \pi_{22}}{\pi_{12}\ \pi_{21}}$

What is the joint distribution ?

## Exercise 4

Does $X_1 \perp\!\!\!\perp X_2 \Rightarrow X_1 \perp\!\!\!\perp X_2 \mid X_3$ ?

Does $X_1 \perp\!\!\!\perp X_2 \mid X_3 \Rightarrow X_1 \perp\!\!\!\perp X_2$ ?

# Solutions

## Exercise 3

$$\text{odr} = 1 \iff X_1 \perp\!\!\!\perp X_2$$

Thus, $P_{12}(x_1, x_2) = \begin{array}{cc|c} 0.12 & & 0.2 \\ & & 0.8 \\ \hline 0.6 & 0.4 & \end{array}$

## Exercise 4

Counterexample 1

$X_3 = 1$

| $X_1$ \ $X_2$ | 0 | 1 |
|---|---|---|
| 0 | 37 | 8 |
| 1 | 3 | 2 |

odr = 3.08

$X_3 = 2$

| $X_1$ \ $X_2$ | 0 | 1 |
|---|---|---|
| 0 | 9 | 38 |
| 1 | 1 | 2 |

odr = 0.47

Marginal $X_1, X_2$

| | 0 | 1 |
|---|---|---|
| 0 | 46 | 46 |
| 1 | 4 | 4 |

odr = 1

$X_1 \perp\!\!\!\perp X_2$

Counterexample 2

$X_3 = 0$

| $X_1$ \ $X_2$ | 0 | 1 |
|---|---|---|
| 0 | 384 | 96 |
| 1 | 16 | 4 |

odr = 1

$X_3 = 1$

| | 0 | 1 |
|---|---|---|
| 0 | 40 | 360 |
| 1 | 10 | 90 |

odr = 1

| | |
|---|---|
| 424 | 456 |
| 26 | 94 |

odr = 3.36

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

## A fundamental property of CI

Let $d=4$ $(X_1, X_2, X_3, X_4)$ and consider the joint (conditional) indep.

$$X_1 \perp\!\!\!\perp X_2 X_3 \mid X_4 \qquad 1 \perp\!\!\!\perp 23 \mid 4$$

**Proposition**

$$1 \perp\!\!\!\perp 23 \mid 4 \iff \begin{array}{c} 1 \perp\!\!\!\perp 2 \mid 4 \\ \text{and} \\ 1 \perp\!\!\!\perp 3 \mid 24 \end{array}$$

This property makes it possible to decompose a CI involving random vectors into a set of pairwise CIs

**Semi graphoids**

→ look at

Pearl, Judea. 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, CA: Morgan Kaufmann.

Further properties

- Intersection
- Composition

... later on

## Concentration matrix

Let $\underline{X} = (X_1, \dots, X_d)$ be a Gaussian r.v.

Its covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \cdot & \sigma_{22} & \cdots & \sigma_{2d} \\ & & \ddots & \\ & & & \sigma_{dd} \end{bmatrix} > 0$$

Its concentration matrix is

$$K = \Sigma^{-1} = \begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1d} \\ & K_{22} & \cdots & K_{2d} \\ & & \ddots & \\ & & & K_{dd} \end{bmatrix} > 0$$

## Block covariance matrix

Sometimes we consider a partition of the variables. Example

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

$\underline{X} = (X_a, X_b)$

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \cdot & \Sigma_{bb} \end{pmatrix} \qquad K = \Sigma^{-1} = \begin{pmatrix} K_{aa} & K_{ab} \\ \cdot & K_{bb} \end{pmatrix}$$

$$K_{aa} \neq \Sigma_{aa}^{-1} \qquad K_{aa} = \left(\Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba}\right)^{-1}$$

# Conditional independence in Gaussian distributions.

Let $\underset{\sim}{X} = (X_1, \dots, X_d) \sim N_d(\overset{\to 0}{\underline{\mu}}, \Sigma)$

Well-known $X_1 \perp\!\!\!\perp X_2 \iff \sigma_{12} = 0$

$\boxed{\text{When } X_1 \perp\!\!\!\perp X_2 \mid X_3 \ ?}$

① Conditional covariance $\operatorname{cov}(X_1, X_2 \mid X_3) = 0$

$$\sigma_{12} - \sigma_{13}\sigma_{32}/\sigma_{33} = 0$$

② Conditional regression coefficient $\beta_{12} = 0$

$$X_1 \mid X_2, X_3 \sim N\big(\underbrace{E(X_1 \mid X_2, X_3)}_{\text{linear}},\ \underbrace{\operatorname{var}(X_1 \mid X_2, X_3)}_{\text{constant}}\big)$$

linear: $\underset{\substack{\downarrow \\ -\frac{K_{12}}{K_{11}}}}{\beta_{12}} X_2 + \underset{\substack{\downarrow \\ -\frac{K_{13}}{K_{11}}}}{\beta_{13}} X_3$

constant: $\frac{1}{K_{11}}$

$$X_1 \perp\!\!\!\perp X_2 \mid X_3 \iff E(X_1 \mid X_2, X_3) = \beta_{13} X_3$$

$$\iff \beta_{12} = 0$$

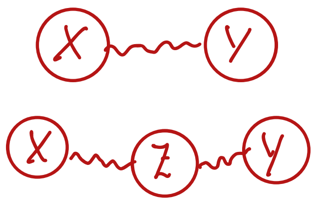
$$\iff K_{12} = 0$$

# Graphical models

It is useful to regard conditional indep. as expressing the notion of irrelevance

| | |
|---|---|
| $X \perp\!\!\!\perp Y \mid Z$ | if we know $Z$, information about $Y$ is irrelevant for knowledge of $X$ |
| $X \perp\!\!\!\perp Y$ | if we ignore all other variables $Y$ looks irrelevant for $X$ |

Irrelevance  separation

Relevance  connection

Idea: use graphs to represent a set of conditional or marginal ind. expressed by separation

# Graphs

$G = (V, E)$

$V \rightarrow$ nodes $= \{1, 2, \dots, d\}$

$E \rightarrow$ edges connecting two nodes $(i, j)$

Edges: arrows $i \rightarrow j$

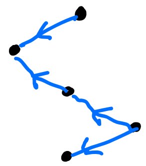
full lines $i - j$

arcs $i \leftrightarrow j$

## Types of graph

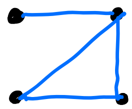
Directed acyclic DAG

only arrows – no cyles



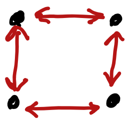
Undirected UG

only full lines



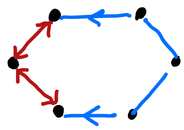
Bidirected BG

only bi-directed edges



Mixed MG

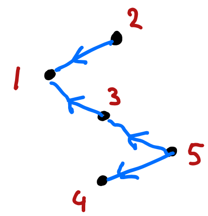
All three types of edge



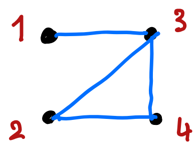
require(ggm)

# Define graphs inth ggm

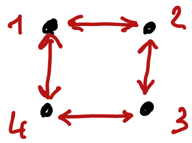
```
DAG(X1 ~ X2 + X3, X3 ~ X5, X4 ~ X5)

   X1 X2 X3 X5 X4
X1  0  0  0  0  0
X2  1  0  0  0  0
X3  1  0  0  0  0
X5  0  0  1  0  1
X4  0  0  0  0  0
```



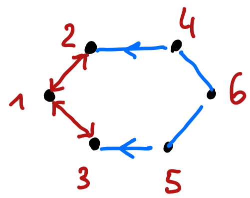

```
UG(~ X1*X3 + X2*X3*X4 )

   X1 X3 X2 X4
X1  0  1  0  0
X3  1  0  1  1
X2  0  1  0  1
X4  0  1  1  0
```




```
100 * UG(~ X1*X2 + X2*X3 + X3*X4 + X1*X4)

    X1  X2  X3  X4
X1   0 100   0 100
X2 100   0 100   0
X3   0 100   0 100
X4 100   0 100   0
```




```
makeMG(bg = UG(~ X1*X2 + X1*X3),
       dg = DAG(X2 ~ X4, X3 ~ X5),
       ug = UG(~ X4*X6 + X5*X6))

    X2 X4  X3 X5 X6  X1
X2   0  0   0  0  0 100
X4   1  0   0  0 10   0
X3   0  0   0  0  0 100
X5   0  0   1  0 10   0
X6   0 10   0 10  0   0
X1 100  0 100  0  0   0
```



# Directed acyclic graph models.

## Example from German labor market

| $A$, successful job placem. | $B$, field of qualification | | | |
|---|---|---|---|---|
| | home economics | | mechan. engineering | |
| | $C$, gender | | $C$, gender | |
| | female | male | female | male |
| yes | 15 (3.61%) | 2 (3.64%) | 4 (20.0%) | 95 (21.1%) |
| no | 400 | 53 | 16 | 355 |
| sum | 415 | 55 | 20 | 450 |
| odds-ratio | 0.99 | | 0.93 | |

A : successful job placement — response

B : Field of qualification — intermediate

C : Gender — Context variable

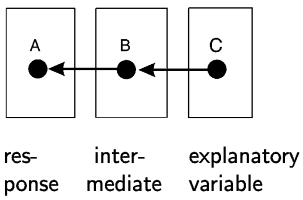
A depends on B and C

B depends on C (both explanatory and response)

C is an intrinsic v.

Data show a CI : $A \perp\!\!\!\perp C \mid B$

Graph :




The missing edge $A \leftarrow C$ implies a CI

## Exercise

What happens if we ignore B = Field of qualification ?

Find the marginal distribution of A and C and the odds-ratio

$$A \xleftarrow{?} C$$

What's the interpretation ?

## Solution

Discrimination against women?

| $A$, successful job placement | $C$, gender: female | male |
|---|---|---|
| yes | 19 (4.4%) | 97 (19.2%) |
| no | 416 | 408 |
| sum | 435 | 505 |
| odds-ratio | 0.19 | |

# Factorization according to a DAG

$$A \longleftarrow B \longleftarrow C$$

The **parents** of a node $i$ are the nodes that are directly connected to $i$

$$pa(A) = \{B\}$$
$$pa(B) = \{C\}$$
$$pa(C) = \emptyset$$

The joint distribution of $(A, B, C)$ can be factorized recursively

$$p(a, b, c) = p(a|bc)\, p(b|c)\, p(c)$$

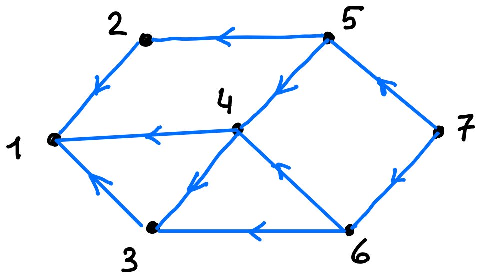
and it simplifies if there are independencies

$$p(a,b,c) = p(a|\underset{pa(A)}{\underset{\downarrow}{b}})\, p(b|\underset{pa(B)}{\underset{\downarrow}{c}})\, \underset{pa(C)}{\underset{\downarrow}{p(c)}}$$

## General formula

$$\begin{aligned} p(x_1, x_2 \cdots x_d) &= p(x_1 \mid pa(x_1)) \times \\ &\quad\; p(x_2 \mid pa(x_2)) \times \\ &\quad\; \vdots \\ &\quad\; p(x_d) \\ &= \prod_{i=1}^{d} p(x_i \mid pa(x_i)). \end{aligned}$$

## Exercise




well ordered DAG

Find the recursive factorization.

## Solution

$$
\begin{aligned}
p(x_1, \dots, x_7) = \ & p(x_1 \mid x_2, x_3, x_4) \times \\
& p(x_2 \mid x_5) \times \\
& p(x_3 \mid x_4, x_6) \times \\
& p(x_4 \mid x_5, x_6) \times \\
& p(x_5 \mid x_7) \times \\
& p(x_6 \mid x_7)\ p(x_7).
\end{aligned}
$$

## Exercise

Which indepencies are present?

## Solution

$$
\begin{aligned}
& X_1 \perp\!\!\!\perp X_5 X_6 X_7 \mid X_2 X_3 X_4 \\
& X_2 \perp\!\!\!\perp X_3 X_4 X_6 X_7 \mid X_5 \\
& X_3 \perp\!\!\!\perp X_5 X_7 \mid X_4 X_6 \\
& X_4 \perp\!\!\!\perp X_7 \mid X_5 X_6 \\
& X_5 \perp\!\!\!\perp X_6 \mid X_7
\end{aligned}
$$

# Markov property.

$G \longleftrightarrow \underline{X}_V$

graph — variables

V, E
properties
(adjacencies,
missing edges)

conditional
independencies

Markov
property
rules

Specifies the rules that translate
the properties of the graph into
conditional independencies

## Ordered Markov property for DAGs

Every DAG can be well-ordered
so that given two nodes

$i$ = head $\bullet \longleftarrow \bullet$ tail $= j > i$

lower number — higher number

The predecessors of a node : pre($i$)
are the set of nodes $j$ that are $> i$

Then the rule is:

$$X_i \perp\!\!\!\perp X_{pre(i) \setminus pa(i)} \mid X_{pa(i)}$$

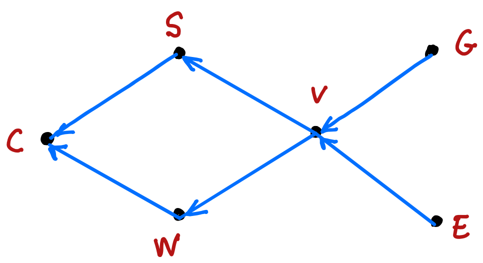
Simulated
(Scutari et al)

# Example : cropdata

In the analysis of a specific plant
a simple model is

| | | | |
|---|---|---|---|
| C: crop (grain mass) | S: number of seeds<br>W: seeds weight | V: Vegetative organs | G: genetic potential<br>E: environm. potential |

A DAG model shows a generating process.



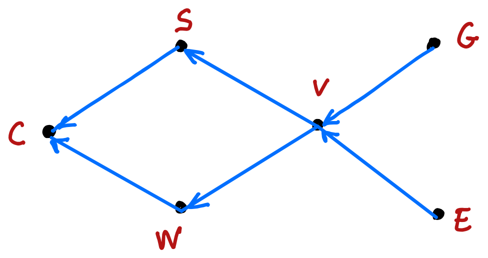

Assume that the joint $p_V$ is Gaussian

Factorization:

$$p_V = p_{C|S W} \times p_{S|V} \times p_{W|V} \times p_{V|G E} \times p_G \times p_E$$

All the factors are univariate
conditional normal distributions

## Recursive univariate regression models




Order:
C
S
W
V
G
E

Simulation $n = 200$

Exercise!

$C \mid SW \sim N(0.3S + 0.3W;\ 6.25^2)$

$S \mid V \sim N(45 + 0.1V;\ 9.94^2)$

$W \mid V \sim N(15 + 0.7V;\ 7.14^2)$

$V \mid GE \sim N(-10.3 + 0.5G + 0.77E;\ 5^2)$

$G \mid E \sim N(50;\ 10^2)$

$E \sim N(50;\ 10^2)$

## Parameters, variation independence

Each density depends on 2 parameters

- $\underset{\sim}{\beta}$ regression coefficients
- $\sigma^2$ conditional variance

The parameter space of the model is the Cartesian product of the separate ranges of its components

# Model fitting.

- Maximum likelihood estimation
- Done by separate fit of each regression model.

```
m_full<- lm(C ~ S+W+V+G+E, data = crop)
```

FIRST EQUATION

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.001      4.349   0.000
S              0.276      0.047   5.838    0.000
W              0.706      0.067  10.606    0.000
V             -0.098      0.098  -0.997    0.320
G              0.078      0.062   1.275    0.204
E              0.043      0.079   0.552    0.581


m_red <- lm(C ~ S + W, data = crop)
anova(m_red, m_full, test = "F")

Model 1: C ~ S + W
Model 2: C ~ S + W + V + G + E
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    197 7851.9
2    194 7770.3  3    81.565 0.6788  0.566


        w        df         p
2.0884556 3.0000000 0.5542518
```

LRT

$C \perp\!\!\!\perp VGE \mid SW$

```
m_full <- lm(S ~ W + V + G + E, data = crop)
```

SECOND EQUATION

```
           Estimate Std. Error t value Pr(>|t|)
(Intercept)  54.374      5.313  10.235    0.000
W             0.020      0.101   0.199    0.842
V             0.014      0.148   0.097    0.923
G            -0.081      0.093  -0.874    0.383
E            -0.048      0.119  -0.402    0.688

m_red <- lm(S ~ 1, data = crop)
        w        df         p
0.8098683 3.0000000 0.8471052
        w        df         p
1.2909126 4.0000000 0.8629154
```

LRT

$S \perp\!\!\!\perp WGE \mid V$

$S \perp\!\!\!\perp WVGE$

(continued)

```
m_full <- lm(W ~ V + G + E, data = crop)

             Estimate Std. Error t value Pr(>|t|)
(Intercept)   18.702      3.518   5.317    0.000
V              0.593      0.096   6.168    0.000
G              0.062      0.066   0.947    0.345
E             -0.035      0.084  -0.420    0.675

        w        df          p
2.2415432 2.0000000 0.3260281
```

THIRD EQUATION

LRT

$W \perp\!\!\!\perp G\,E \mid V$

```
m_full <- lm(V ~ G + E, data = crop)

             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -10.455      2.500  -4.182        0
G              0.455      0.036  12.501        0
E              0.743      0.033  22.357        0
```

4th EQUATION

no independence

```
m_full <- lm(G ~ E, data = crop)

             Estimate Std. Error t value Pr(>|t|)
(Intercept)   49.692      3.365  14.766    0.000
E              0.009      0.065   0.137    0.891

m_red <- lm(G ~ 1, data = crop)

         w         df           p
0.01905948 1.00000000 0.89019611
```

5th EQUATION

LRT

$G \perp\!\!\!\perp E$

# Overall fit using ggm

Adjacency matrix

```
G <- DAG(C ~ S + W, S ~ V, W ~ V, V ~ G + E)
G
  C S W V G E
C 0 0 0 0 0 0
S 1 0 0 0 0 0
W 1 0 0 0 0 0
V 0 1 1 0 0 0
G 0 0 0 1 0 0
E 0 0 0 1 0 0
```

covariance matrix in the same order of data

Fit the DAG

```
ord <- colnames(G)
S <- cov(crop[,ord])
fitDag(G, S, n = 200)
```

fitted covariance matrix

```
$Shat
       C      S      W       V      G       E
C 84.539 22.814 56.638  42.640 17.156  33.639
S 22.814 90.605 -2.860  -4.848 -1.950  -3.824
W 56.638 -2.860 83.725  64.107 25.792  50.574
V 42.640 -4.848 64.107 108.656 43.716  85.719
G 17.156 -1.950 25.792  43.716 96.019   0.000
E 33.639 -3.824 50.574  85.719  0.000 115.416
```

fitted reg. coeff with sign changed

```
$Ahat
  C      S      W      V      G      E
C 1 -0.273 -0.686  0.000  0.000  0.000
S 0  1.000  0.000  0.045  0.000  0.000
W 0  0.000  1.000 -0.590  0.000  0.000
V 0  0.000  0.000  1.000 -0.455 -0.743
G 0  0.000  0.000  0.000  1.000  0.000
E 0  0.000  0.000  0.000  0.000  1.000
```

partial variances

```
$Dhat
      C      S      W      V      G       E
 39.457 90.388 45.902 25.090 96.019 115.416
```

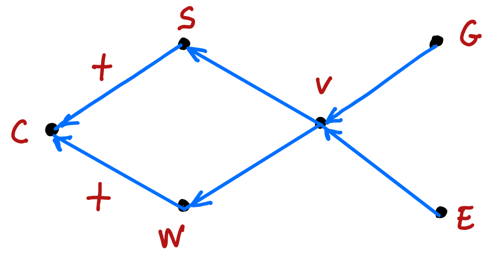
deviance

```
$dev
[1] 5.159
```

d.f.

```
$df
[1] 9
```

# Separation in DAGs and its connection with Conditional independence




We derived some CI from this graph

however

What can you say about $S \perp\!\!\!\perp W \mid C$ ?

It depends on $\sigma_{SW|C} \stackrel{?}{=} 0$ or $\rho_{SW|C} \stackrel{?}{=} 0$

We can estimate it from data:

$$\hat{\rho}_{SW|C} = -0.28$$

Or test it from data

```
ci.test("S", "W", "C", data = crop)

cor = -0.2876, df = 197, p-value = 3.808e-05
```

← used bnlearn

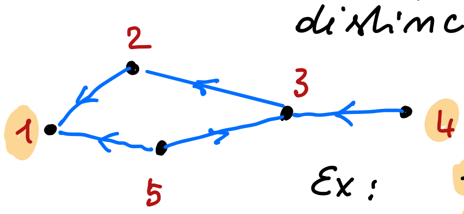
Reject $H_0: S \perp\!\!\!\perp W \mid C$

Can we say something without data ?

# d-separation in DAGs

> It is possible to use d-separation to answer to any CI statement
>
> $A \perp\!\!\!\perp B \mid C$ for any triple of subsets of $V$

**Paths in DAG:** a sequence of consecutive distinct edges




Ex: $1 \leftarrow 2 \rightarrow 3 \leftarrow 4$

$1 \leftarrow 5 \rightarrow 3 \leftarrow 4$

**Blocked paths:** by a set of nodes $C$ iff

- the path contains a chain $a \rightarrow m \rightarrow c$ or a fork $a \leftarrow m \rightarrow c$ such that $m$ is in $C$
- or the path contains a collider $a \rightarrow m \leftarrow c$ such that $m$ is not in $C$ and no descendants of $m$ are in $C$

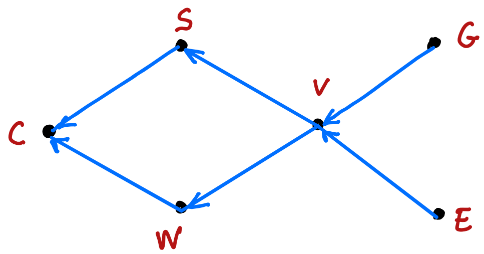
Two subsets of nodes $A$ an $B$ are said d-separated given $C$ if every path between $A$ and $B$ is blocked by $C$

# Global Markov Property in DAGs

**Pearl 1988**

Given a DAG, if $A, B, C \subset V$ are disjoint, and $A$ and $B$ are d-separated given $C$ then $A \perp\!\!\!\perp B \mid C$

## Example




$S \perp\!\!\!\perp W \mid C$ ?   NO because the path $S \to C \leftarrow W$ contains a collider node $C$ that is inside the conditioning set.

$S \perp\!\!\!\perp W \mid VG$ ?   YES   Paths

$S \to C \leftarrow W$   $C$ not in $VG$
$S \leftarrow V \to W$   $V$ is in $VG$

```
> dSep(G, "S", "W", "C")
[1] FALSE

> dSep(G, "S", "W", c("V", "G"))
[1] TRUE
```

check using R package ggm

$G \perp\!\!\!\perp E \mid S$ ?   Exercise

# Fit DAG models with categorical data

## Canadian Women Labour-Force participation

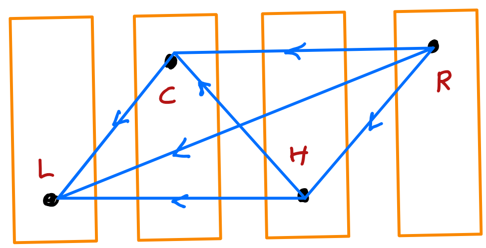
Data on 263 married women ages 21–30 (1977)

data from R carData

| H | C | L | Atlantic | BC | Ontario | Prairie | Quebec |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | 4 | 6 | 0 | 5 |
| | | 1 | 1 | 4 | 11 | 3 | 10 |
| | 1 | 0 | 11 | 6 | 25 | 15 | 18 |
| | | 1 | 2 | 0 | 4 | 4 | 4 |
| 1 | 0 | 0 | 1 | 4 | 6 | 1 | 5 |
| | | 1 | 1 | 2 | 10 | 1 | 3 |
| | 1 | 0 | 11 | 8 | 44 | 7 | 19 |
| | | 1 | 2 | 1 | 2 | 0 | 1 |

L Full time work
C Presence of children
H Husband's income (1 = > median \$14000)
R Region

Ordering of the variables
(L, C, H, R)




Multinomial model (cross-classified)

2 × 2 × 2 × 5 Table

$$\pi_{lchr} = \pi_{l|chr}\,\pi_{c|hr}\,\pi_{h|r}\,\pi_r$$

# Logistic regression

$L \mid CHR$ $\qquad \pi_{l|chr} = P(L=1 \mid c, h, r)$

$$\text{logit}(\pi) = \log \frac{\pi}{1-\pi} \qquad (0,1) \mapsto (-\infty, +\infty)$$

generalized linear model

$L \sim C + H + R \quad (\text{link} = \text{logit})$

$L \sim C * H * R \quad (\text{link} = \text{logit})$

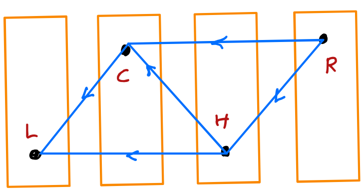
choose selected interactions!

```
glm(L ~ C + H + R, family = binomial, data =
wlfdata)

            Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.041      0.607   1.715    0.086
C1            -2.609      0.361  -7.234    0.000
H1            -0.768      0.348  -2.210    0.027
RBC           -0.944      0.745  -1.266    0.206
ROntario      -0.254      0.590  -0.430    0.667
RPrairie       0.168      0.695   0.241    0.809
RQuebec       -0.342      0.627  -0.545    0.586

glm(L ~ C + H, family = binomial, data = wlfdata)
        w        df         p
2.8117211 4.0000000 0.5898111
```

LRT test of

$L \perp\!\!\!\perp R \mid CH$

+ non-independence constraints.




the constraints are not visible in the graph

# Recursive logistic model

C | H R generalized linear model

C ~ H + R (link = logit)

(interaction not significant)

```
glm(C~ H+R, family = binomial, data = wlfdata)
```

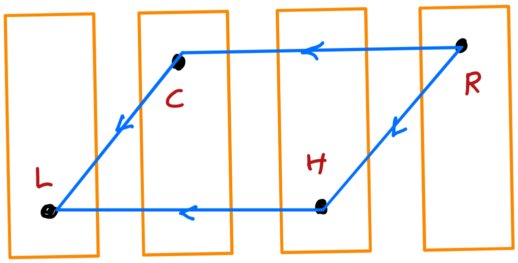
| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 1.671 | 0.551 | 3.030 | 0.002 |
| H1 | 0.437 | 0.282 | 1.546 | 0.122 |
| RBC | -1.827 | 0.656 | -2.785 | 0.005 |
| ROntario | -1.092 | 0.579 | -1.886 | 0.059 |
| RPrairie | -0.136 | 0.730 | -0.186 | 0.852 |
| RQuebec | -1.250 | 0.598 | -2.089 | 0.037 |

```
glm(C~ R, family = binomial, data = wlfdata)
```

| w | df | p |
|---|---|---|
| 2.4134751 | 1.0000000 | 0.1202951 |

LRT of CI

$C \perp\!\!\!\perp H \mid R$




```
glm(H ~ R, family = binomial, data = wlfdata)
```

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 0.000 | 0.365 | 0.000 | 1.000 |
| RBC | 0.069 | 0.521 | 0.132 | 0.895 |
| ROntario | 0.298 | 0.414 | 0.721 | 0.471 |
| RPrairie | -0.894 | 0.538 | -1.660 | 0.097 |
| RQuebec | -0.279 | 0.443 | -0.629 | 0.529 |

```
glm(H ~ 1, family = binomial, data = wlfdata)
```

| w | df | p |
|---|---|---|
| 9.19254035 | 4.00000000 | 0.05646299 |

$H \not\!\perp\!\!\!\perp R$