

# Graphical Markov Models

3 Lectures

Giovanni M. Marchetti  
2024

- 1 Conditional independence  
Graphs (DAG, UG, BG, RCG)  
Markov Properties  
Conditional Indep. tests  
Examples
- 2 Undirected graphs.  
Markov properties  
Separation, decomposable graphs  
Fitting by ML, Examples
- 3 Bidirected graphs  
Regression chain graphs  
Markov properties  
Parametrizations  
Fitting graphs to data  
Examples

# 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)

cct (J. Lang)

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

$$\underline{x} = (x_1, x_2, x_3, \dots, x_d)$$

can be

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

### Example

→ Gaussian  $\underline{x} \sim N(\underline{\mu}, \Sigma)$

→ Multinomial  $\underline{x} \sim \text{Mult}(n, \underline{\pi})$

### Exercise 1

Simulate 1000 observations  
from 3 binary variables  
 $x_1, x_2, x_3$  with a joint prob.

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

### Exercise 2

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

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

# Solutions

## Exercise 1

```
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))
```

$\pi \sim \text{Dir}(0, 0, 0, 0)$

8

## 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  $(X_1, \dots, X_d)$  ↗ all contin.  
↗ all discrete

Denote the joint distribution by

$$p_V(x_1, \dots, x_d) \xrightarrow{\text{pmf}} \frac{pmf}{polf}$$

where

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

### Example

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

## Conditional Distributions

$$X_1 | X_2 \quad P_{1|2}(x_1 | x_2)$$

$$X_1 | X_2, X_3 \quad P_{1|23}(x_1 | x_2, x_3)$$

## Conditional independence

Given 3 variables  $X_1, X_2, X_3$

$X_1$  conditionally independent of  $X_2$

given  $X_3$

if

$$P_{12|3}(x_1, x_2 | x_3) = P_{1|3}(x_1 | x_3) P_{2|3}(x_2 | x_3)$$

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

$$\begin{aligned} p(x_1, x_2, x_3) &= p(x_1 \mid x_3) p(x_2 \mid x_3) p(x_3) - \\ &= p(x_1 \mid x_3) p(x_3 \mid x_2) p(x_2) - \end{aligned}$$

### Marginal independence

$$\begin{aligned} X_1 \perp\!\!\!\perp X_2 \text{ if } p_{12}(x_1, x_2) &= p_1(x_1) p_2(x_2) \\ \text{for all } x_1, x_2. \end{aligned}$$

It's 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 odr =  $\frac{\pi_{11}}{\pi_{12}} \frac{\pi_{22}}{\pi_{21}}$

What is the joint distribution?

### Exercise 4

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

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

# Solutions

## Exercise 3

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

Thus,  $P_{12}(x_1, x_2) =$

0.12	0.2
0.8	
0.6 0.4	

## Exercise 4

### Counterexample 1

		$x_3 = 1$	
		0	1
$x_1$	$x_2$	0	37 8
		1	3 2

$$\text{odr} = 3.08$$

		$x_3 = 2$	
		0	1
$x_1$	$x_2$	0	9 38
		1	1 2

$$\text{odr} = 0.47$$

		Marginal $X_1, X_2$	
		0	1
$x_1$	$x_2$	0	46 46
		1	4 4

$$\text{odr} = 1$$

### Counterexample 2

		$x_3 = 0$	
		0	1
$x_1$	$x_2$	0	384 96
		1	16 4

$$\text{odr} = 1$$

		$x_3 = 1$	
		0	1
$x_1$	$x_2$	0	40 360
		1	10 90

$$\text{odr} = 1$$

		424 456	
		26	94

$$\text{odr} = 3.36$$

$X_1 \perp\!\!\!\perp X_2 | X_3$

## A fundamental property of CI

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

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

1

Proposition

$$1 \perp\!\!\!\perp 23 \mid 4 \iff$$

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

and

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

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

Semigraphoids

→ look at

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

Further properties

- Intersection ... later on
- Composition

## Concentration matrix

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

Its covariance matrix is

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

Its concentration matrix is

$$K = \Sigma^{-1} = \begin{bmatrix} K_{11} & K_{12} & \dots & K_{1d} \\ K_{21} & \dots & K_{2d} \\ \vdots & \ddots & \ddots & K_{dd} \end{bmatrix} > 0$$

## Block covariance matrix

Sometimes we consider a partition of the variables. Example

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

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

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

$$K_{aa} \neq \Sigma_{aa}^{-1}$$

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

Conditional independence  
in Gaussian distributions.

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

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

When  $X_1 \perp\!\!\!\perp X_2 | X_3$  ?

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

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

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

$$X_1 | X_2, X_3 \sim N(E(X_1 | X_2, X_3), \text{var}(X_1 | X_2, X_3))$$

$$\begin{aligned} & \text{linear} \leftarrow & \text{constant} \leftarrow \\ & \beta_{12} X_2 + \beta_{13} X_3 & \downarrow \\ & \frac{-K_{12}}{K_{11}} & \frac{-K_{13}}{K_{11}} & \frac{1}{K_{11}} \end{aligned}$$

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

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

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

## 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  $\textcircled{X} \quad \textcircled{Y}$  separation

Relevance



connection



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

# Graphs

$$G = (V, E)$$

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

edges connecting  
two nodes  $(i, j)$

Edges: arrows  $i \rightarrow j$

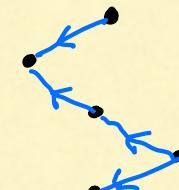
full lines  $i - j$

arcs  $i \leftrightarrow j$

## Types of graph

Directed acyclic DAG

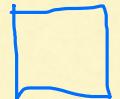
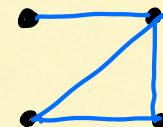
only arrows - no cycles



Undirected

UG

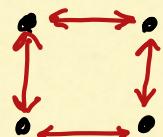
only full lines



Bidirected

BG

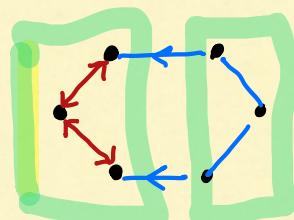
only bi-directed edges



Mixed

MG

All three types of edge

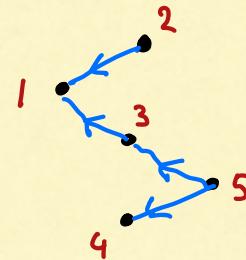


require(ggm)

## Define graphs with ggm

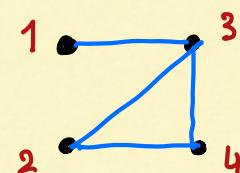
DAG( $X_1 \sim X_2 + X_3$ ,  $X_3 \sim X_5$ ,  $X_4 \sim X_5$ )

	$X_1$	$X_2$	$X_3$	$X_5$	$X_4$
$X_1$	0	0	0	0	0
$X_2$	1	0	0	0	0
$X_3$	1	0	0	0	0
$X_5$	0	0	1	0	1
$X_4$	0	0	0	0	0



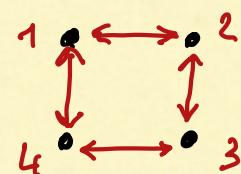
UG( $\sim X_1 \cdot X_3 + X_2 \cdot X_3 \cdot X_4$ )

	$X_1$	$X_3$	$X_2$	$X_4$
$X_1$	0	1	0	0
$X_3$	1	0	1	1
$X_2$	0	1	0	1
$X_4$	0	1	1	0



$100 * \text{UG}(\sim X_1 \cdot X_2 + X_2 \cdot X_3 + X_3 \cdot X_4 + X_1 \cdot X_4)$

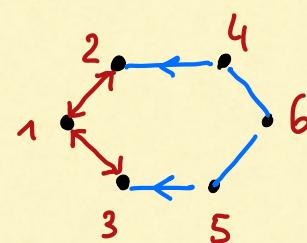
	$X_1$	$X_2$	$X_3$	$X_4$
$X_1$	0	100	0	100
$X_2$	100	0	100	0
$X_3$	0	100	0	100
$X_4$	100	0	100	0



```

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

	$X_2$	$X_4$	$X_3$	$X_5$	$X_6$	$X_1$
$X_2$	0	0	0	0	0	100
$X_4$	1	0	0	0	10	0
$X_3$	0	0	0	0	0	100
$X_5$	0	0	1	0	10	0
$X_6$	0	10	0	10	0	0
$X_1$	100	0	100	0	0	0



# Directed acyclic graph models.

Example from German labor market

A, suc- cessful job placem.	B, field of qualification			
	home economics		mechan. engineering	
	C, gender		C, gender	
female	male	female	male	
yes	15 (3.61%)	2	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

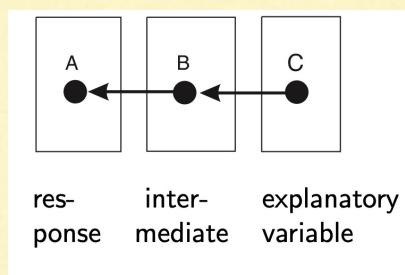
A depends on B and C

B depends on C (both explanatory and response)

C is an intrinsic v.

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

Graph :



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

## Exercise

What happens if we ignore  
 $B = \text{Field of qualification}$  ?

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



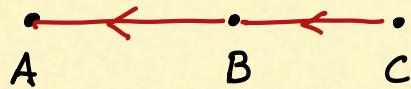
What's the interpretation ?

## 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



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

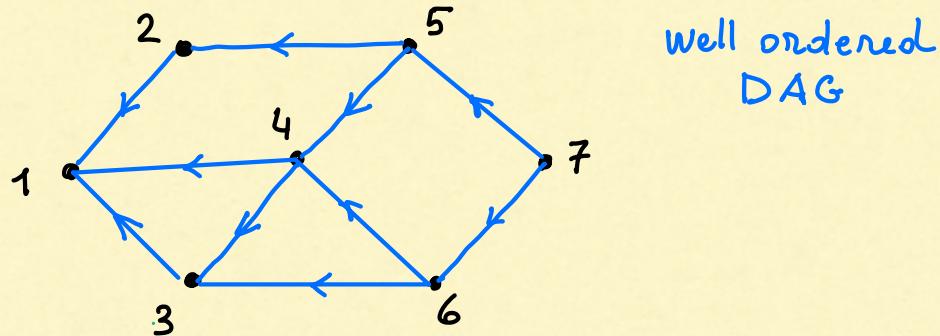
$$P(a, b, c) = P(a|b) P(b|c) P(c)$$

$\downarrow$                      $\downarrow$                      $\downarrow$   
 $pa(A)$        $pa(B)$        $pa(C)$ 
 $P \in \{\neq, \perp\}$

### General formula

$$\begin{aligned}
 P(x_1, x_2, \dots, x_d) &= P(x_1 | pa(x_1)) \times \\
 &\quad P(x_2 | pa(x_2)) \times \\
 &\quad \vdots \\
 &\quad P(x_d) \\
 &= \prod_{i=1}^d P(x_i | pa(x_i)).
 \end{aligned}$$

## Exercise



well ordered DAG

Find the recursive factorization.

## Solution

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

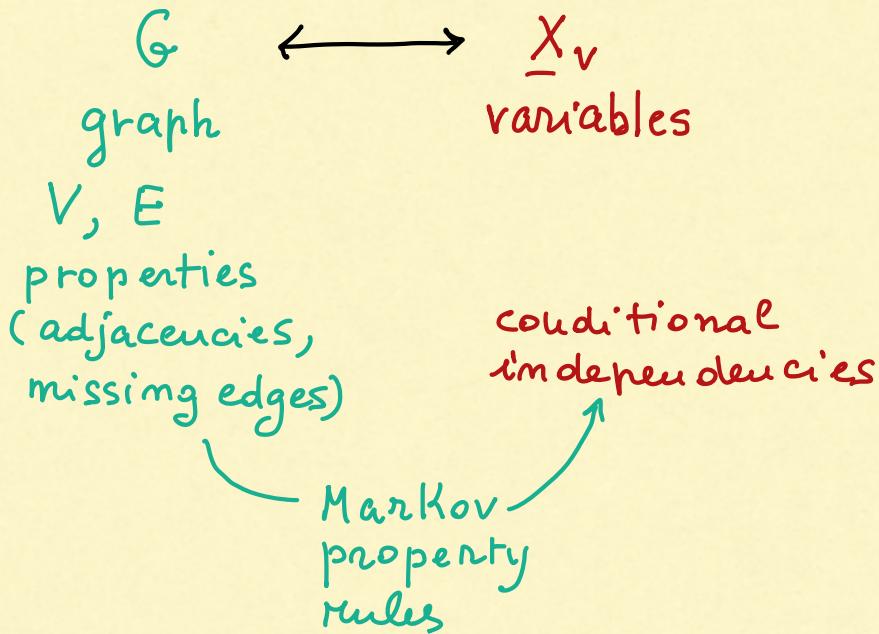
## Exercise

Which independencies are present ?

## Solution

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

## Markov property.



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



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

Then the rule is :

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

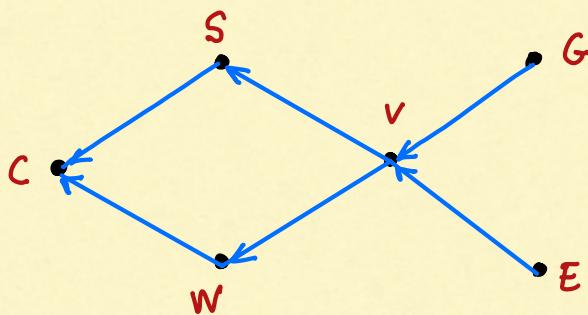
Simulated  
(Scutari et al.)

## Example: cropdata

In the analysis of a specific plant  
a simple model is

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

A DAG model shows a generating process.



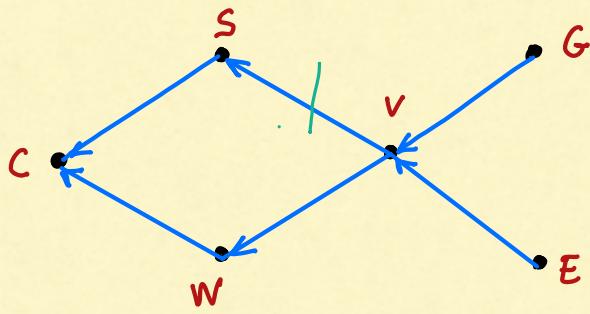
Assume that the joint p is Gaussian

Factorization:

$$P = P_{C|SW} \times P_{S|V} \times P_{W|V} \times P_{V|GE} \times P_G \times P_E$$

All the factors are univariate  
conditional normal distributions

## Recursive univariate regression models



Order:

- C
- S
- W
- V
- G
- E

Simulation

$n = 200$

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

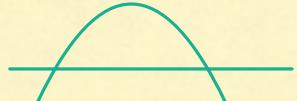
$$S | V \sim N(45 + 0.1V; 9.94^2)$$

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

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

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

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



Exercise

Parameters, variation independence

Each density depends on 2 parameters

- $\beta$  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

LRT

w	df	p
2.0884556	3.0000000	0.5542518

C || VGE | 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

LRT

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

w	df	p
0.8098683	3.0000000	0.8471052
1.2909126	4.0000000	0.8629154

S || WGE | v

S || 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

WILGE / 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

5th  
EQUATION

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

w	df	p
0.01905948	1.0000000	0.89019611

LRT

GLE

# Overall fit using ggm

```
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

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

Adjacency matrix

covariance matrix

in the same order of data

Fit the DAG

\$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

\$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

fitted covariance matrix

\$Dhat

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

\$dev

[1] 5.159

\$df

[1] 9

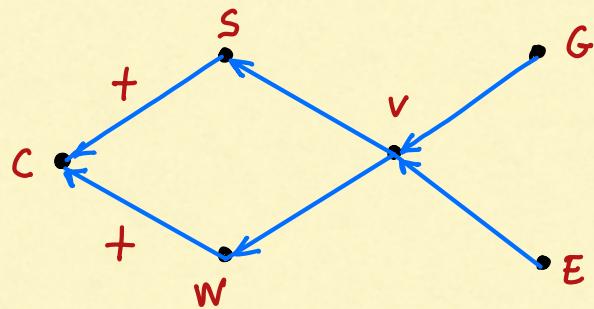
fitted reg. coeff with Sign change of

partial variances

deviance

dL.f.

# Separation in DAGs and its connection with Conditional independence



We derived some CI from this graph  
however

what can you say about  $S \perp\!\!\!\perp W | C$  ?

It depends on  $\sigma_{sw|c} = 0$  or  $p_{sw|c} = 0$

We can estimate it from data:

$$\hat{p}_{sw|c} = -0.28$$

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  
 $H_0: S \perp\!\!\!\perp W | C$

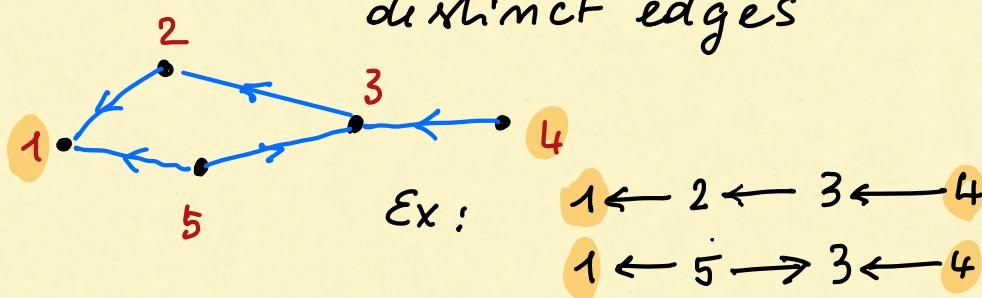
Can we say something without data?

## $d$ -separation in DAGs

It is possible to use  $d$ -separation to answer to any CI statement

$$A \perp\!\!\!\perp B \mid C \quad \text{for any triple of subsets of } V$$

Paths in DAG: a sequence of consecutive distinct edges



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

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

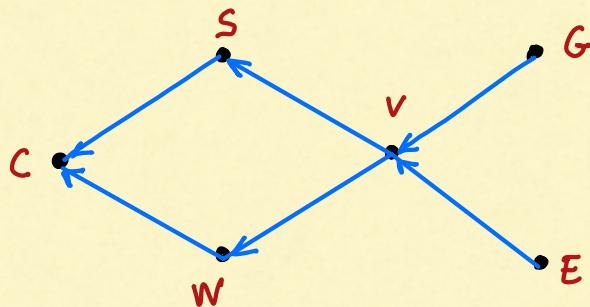
Two subsets of nodes  $A$  and  $B$  are said  $d$ -separated given  $C$  if every path between  $A$  and  $B$  is blocked by  $C$

# 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  $A \perp\!\!\!\perp B | C$

Example



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

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

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

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

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

} check  
using R package  
ggm

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

## Fit DAG models with categorical data

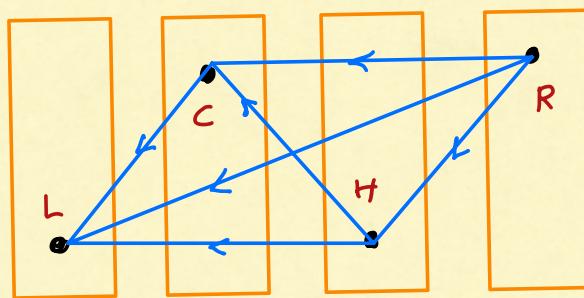
Canadian Women Labour Force participation

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

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

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

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



Multinomial model (cross-classified)

$2 \times 2 \times 2 \times 5$  Table

$$\pi_{\text{echr}} = \tilde{\pi}_{e|\text{chr}} \tilde{\pi}_{c|\text{hr}} \tilde{\pi}_{h|\text{r}} \tilde{\pi}_{\text{r}}$$

# Logistic regression

$$L \mid C H R \quad \pi_{\text{elchr}} = P(L=1 \mid c, h, r)$$

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

generalized linear model

$$\begin{array}{l} \xrightarrow{\quad} L \sim C + H + R \quad (\text{link} = \text{logit}) \\ \xrightarrow{\quad} L \sim C * H * R \quad (\text{link} = \text{logit}) \end{array}$$

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
RPrarie	0.168	0.695	0.241	0.809
RQuebec	-0.342	0.627	-0.545	0.586

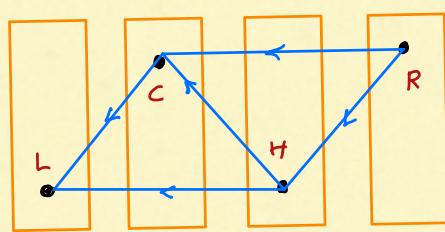
```
glm(L ~ C + H, family = binomial, data = wlfdata)
      w      df      p
2.8117211 4.0000000 0.5898111
```

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

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

(interaction not significant)

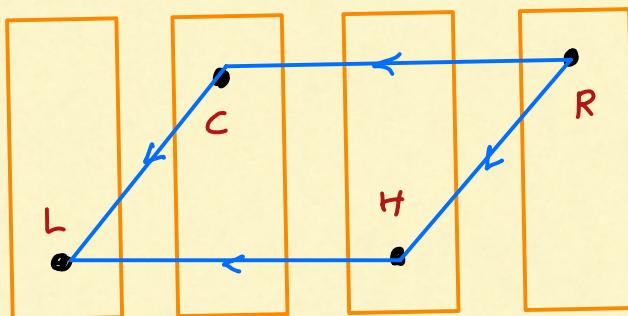
```
glm(C ~ H + R, family = binomial, data = wlfdata)
```

	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
RPrarie	-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 ⊥ L H | 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
RPrarie	-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	4.0000000	0.05646299

H ~~AK~~ R