When is Correlation Causation?

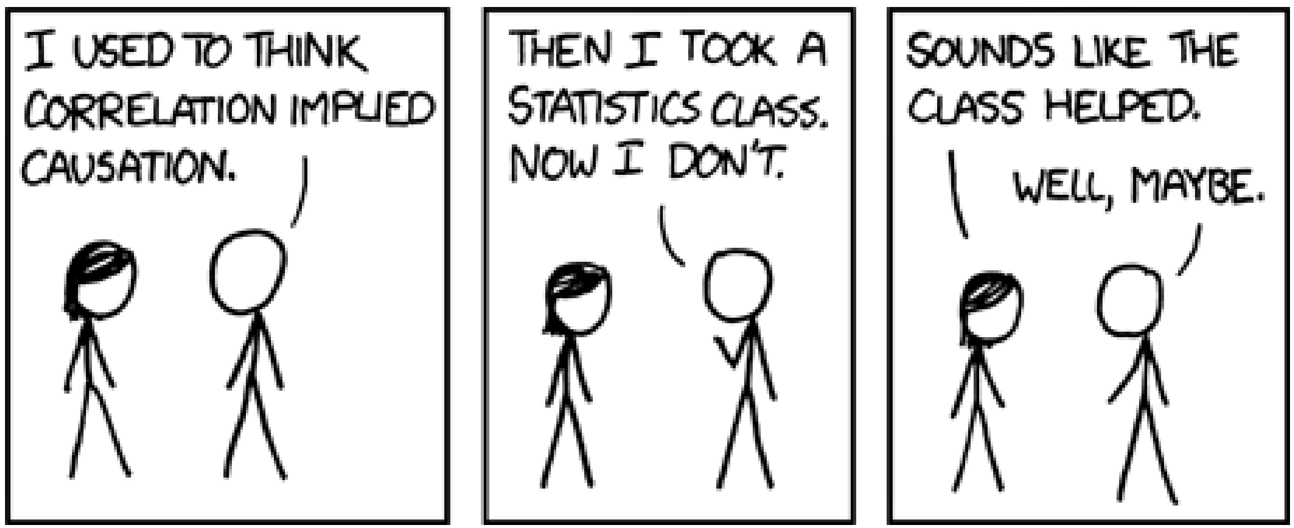
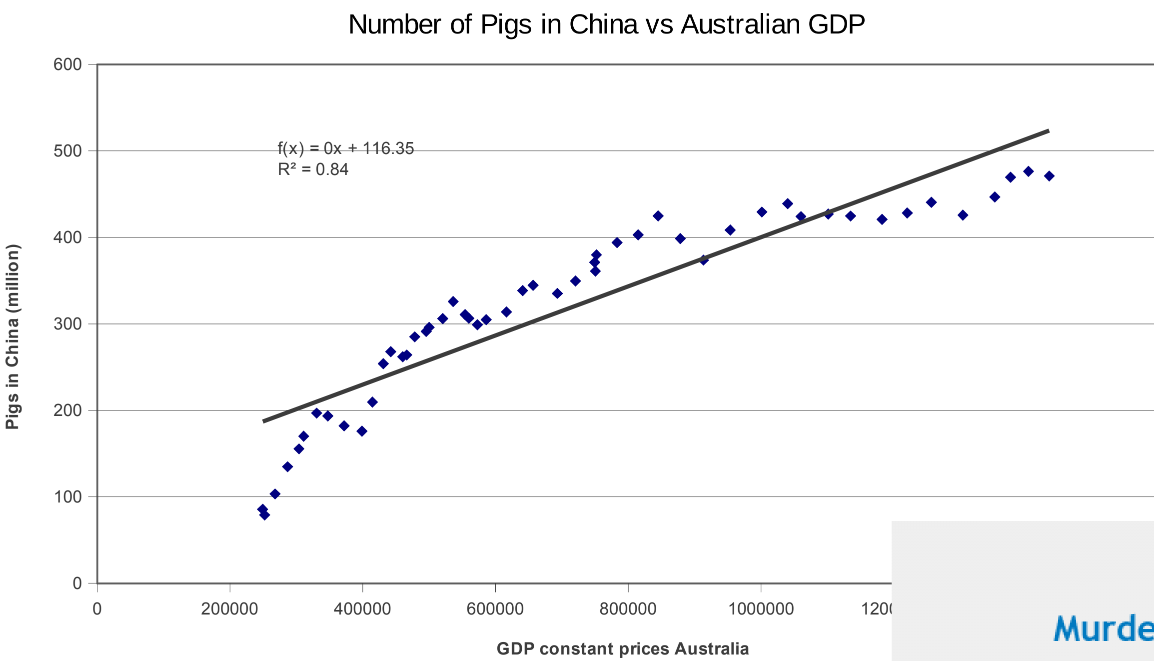
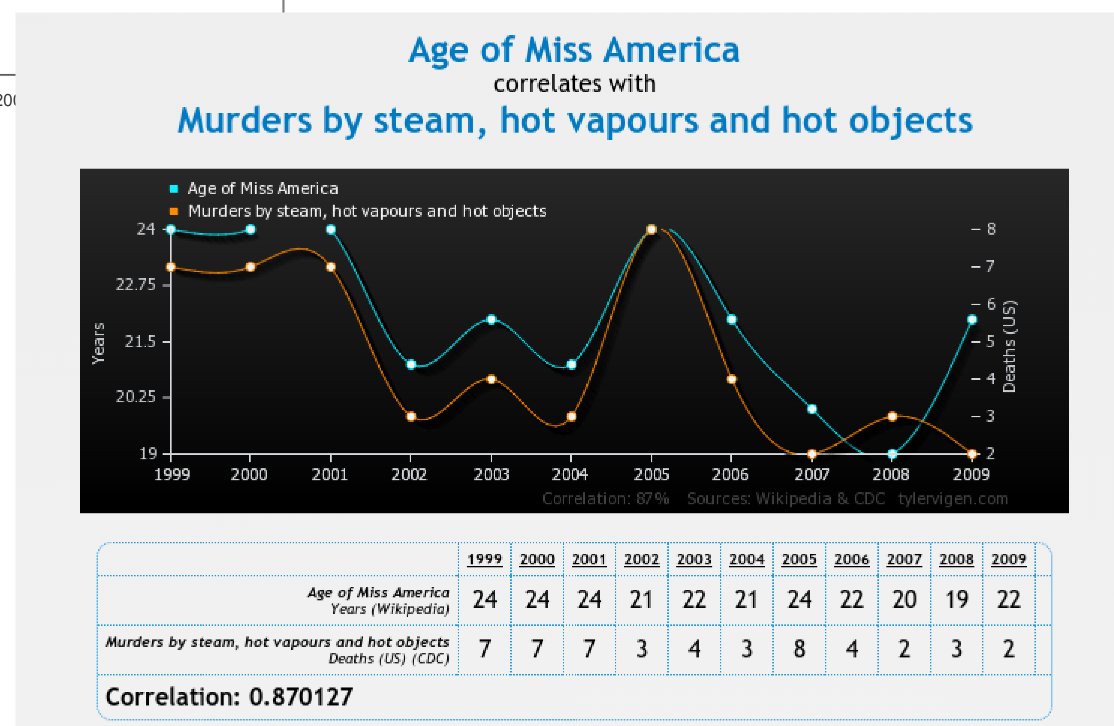


Image Source: http://xkcd.com/552/

When is Correlation Causation?





	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Age of Miss America *Years (Wikipedia)*	24	24	24	21	22	21	24	22	20	19	22
Murders by steam, hot vapours and hot objects *Deaths (US) (CDC)*	7	7	7	3	4	3	8	4	2	3	2

Correlation: 0.870127

Image source: www.tylervigen.com/

We care about causality

Chocolate 'may help keep people slim'

COMMENTS (251)

By Michelle Roberts
Health reporter, BBC News

People who eat chocolate regularly tend to be thinner, new research suggests.

The findings come from a study of nearly 1,000 US people that looked at diet, calorie intake and body mass index (BMI) - a measure of obesity.

It found those who ate chocolate a few times a week were, on average, slimmer than those who ate it occasionally.

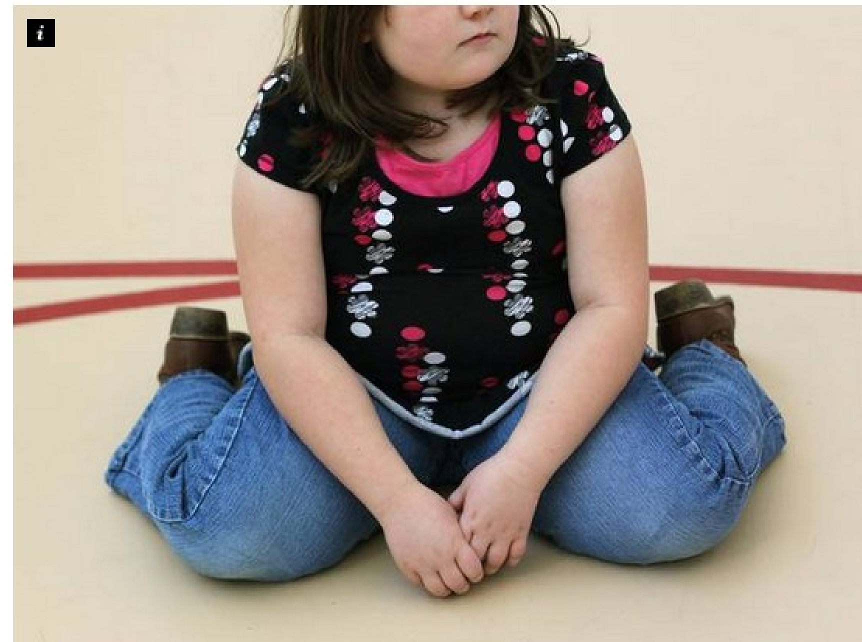
Even though chocolate is loaded with calories, it contains ingredients that may favour weight loss rather than fat synthesis, scientists believe.

Chocolate contains antioxidants but is also high in fat and sugar

Related Stories

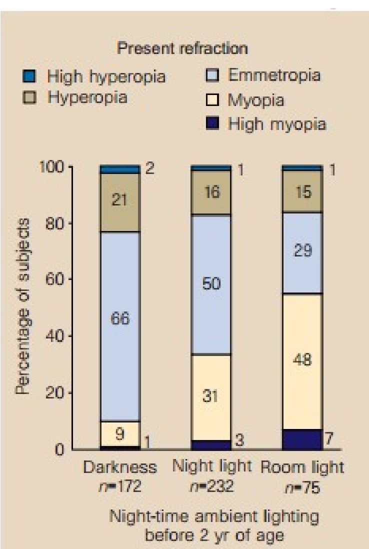
http://www.bbc.com/news/health-17511011

Childhood obesity partly caused by strict parenting, say scientists



Parents who struck a balance between being strict and kind were less likely to bring up obese children

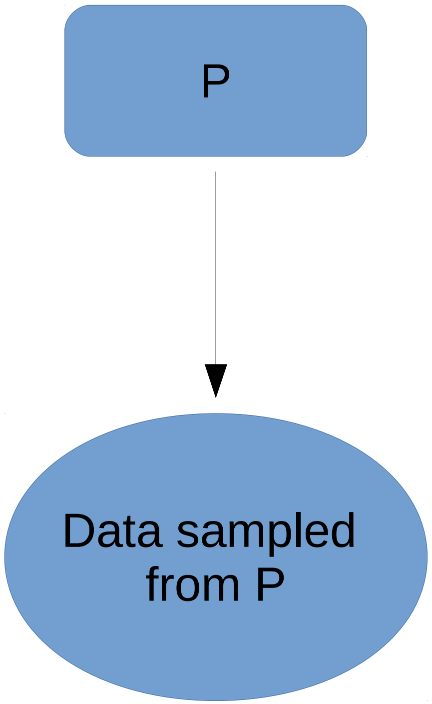
http://www.independent.co.uk/life-style/health-and-families/health-news/childhood-obesity-partly-caused-by-strict-parenting-say-scientists-9206147.html



Sleeping with the light on is associated with short-sightedness in kids.

Quinn, Graham E., et al. "Myopia and ambient lighting at night." Nature 399.6732 (1999): 113-114

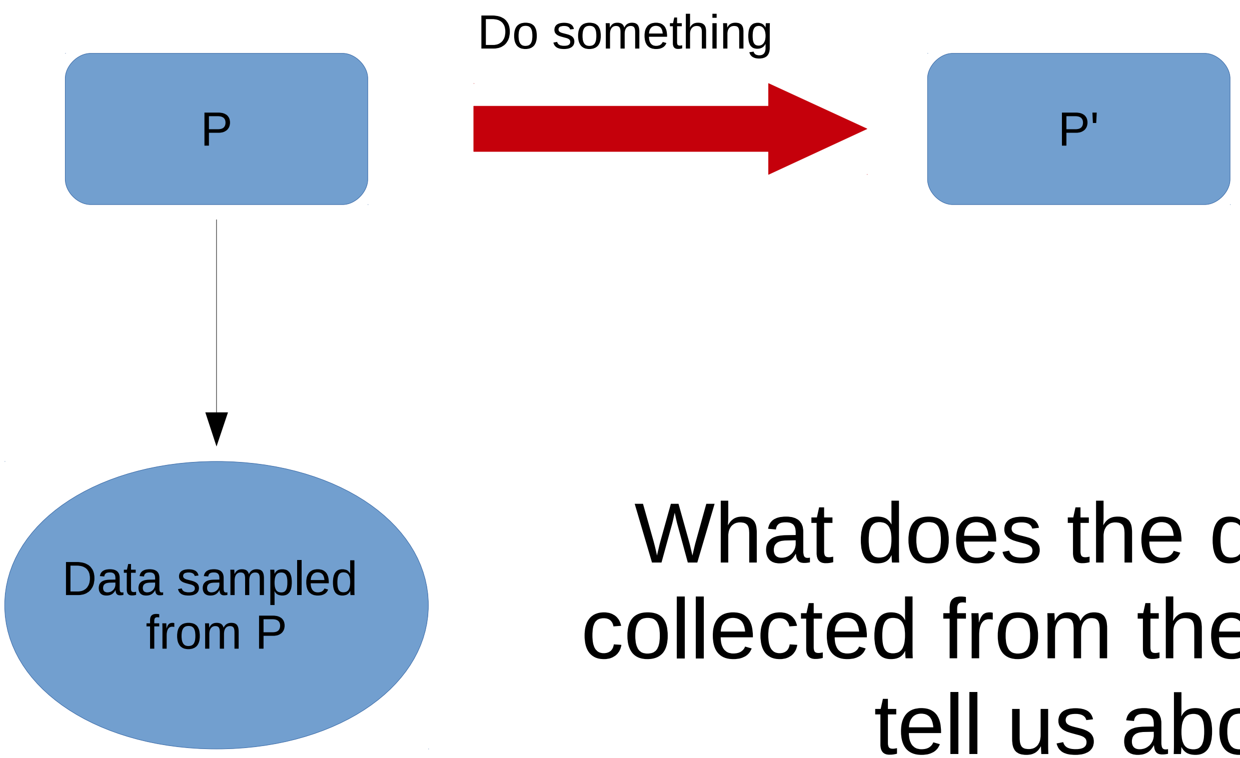
Machine Learning



Infer some properties of P from the sample.

- Supervised ML, given a set of training examples y and features $\underline{\boldsymbol{x}}$ from an underlying distribution P infer $E(y|\underline{\boldsymbol{x}})$

Causal inference



What does the data we have collected from the distribution P tell us about P'?

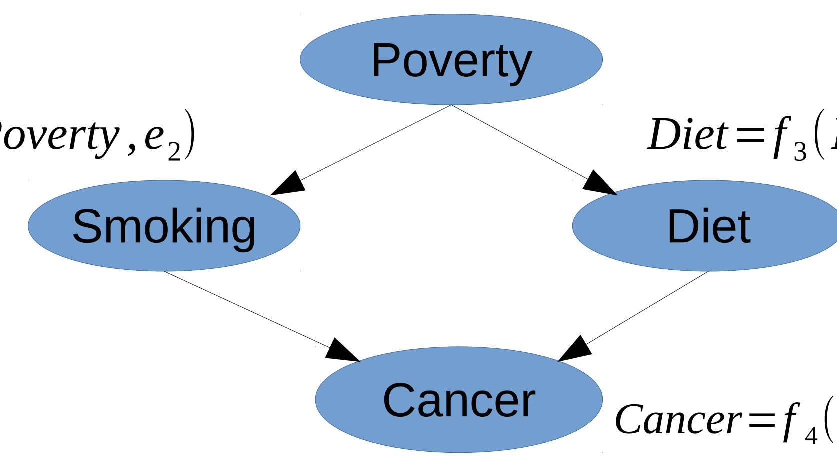
A framework for causal inference

- Represent each variable as deterministic function of its direct causes and noise
- Noise terms must be mutually independent

$$x_i = f_i(Parents_i, e_i) \quad \text{where} \quad \begin{array}{l} \{f_1 \ldots f_n\}\ deterministic\ functions \\ \{e_1 \ldots e_n\}\ mutually\ independent \end{array}$$

$$Poverty = f_1(e_1)$$

$$Smoking = f_2(Poverty, e_2) \qquad Diet = f_3(Poverty, e_3)$$



$$Cancer = f_4(Smoking, Diet, e_4)$$

=> Variables are independent of their non-effects given their direct causes

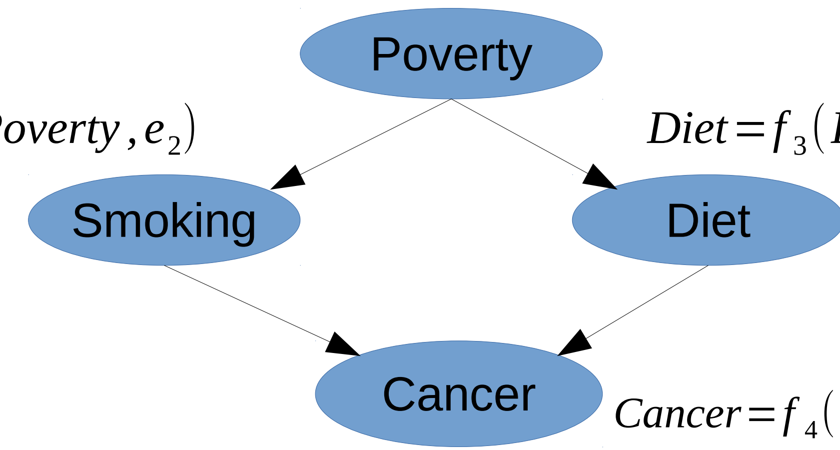
=> The factorisation of the full distribution can be simplified

$$P(Po, S, D, C) = P(Po)P(S|Po)P(D|Po, S)P(C|Po, S, D) = P(Po)P(S|Po)P(D|Po)P(C|S, D)$$

A framework for causal inference

$$Poverty = f_1(e_1)$$

$$Smoking = f_2(Poverty, e_2) \qquad Diet = f_3(Poverty, e_3)$$

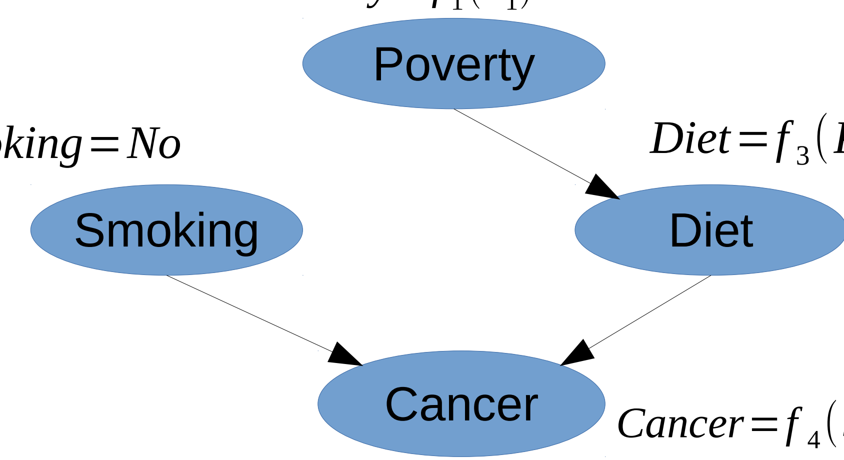


$$Cancer = f_4(Smoking, Diet, e_4)$$

$$do(Smoking = No)$$

$$Poverty = f_1(e_1)$$

$$Smoking = No \qquad Diet = f_3(Poverty, e_3)$$

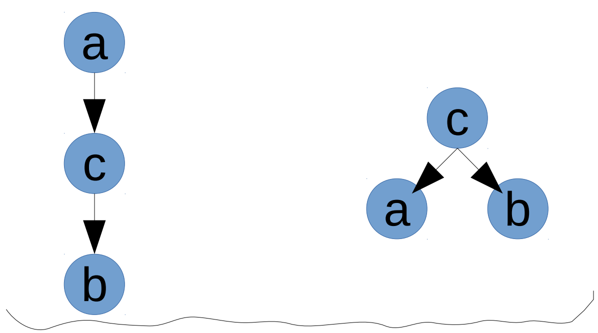


$$Cancer = f_4(Smoking, Diet, e_4)$$

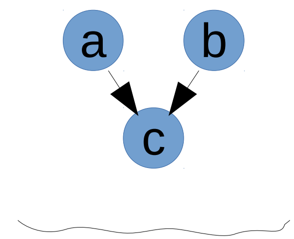
Conditional Independence

a conditionally independent of **b** given **c,** $a \perp b | \{c\}$,means:

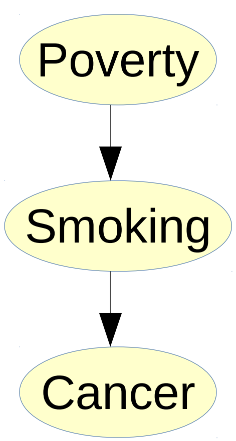
- Assuming you know **c**, learning the value of **b** tells you nothing about the value of **a**.
- Graphically, influence can't flow from **a** to **b** through **c**.

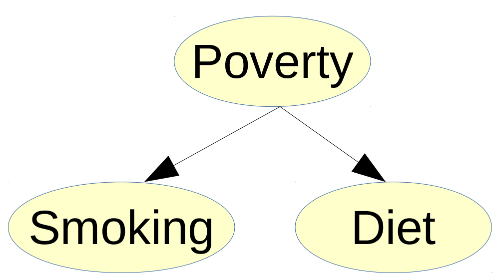


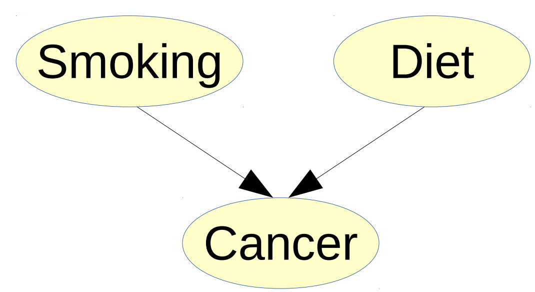
$$a \perp b | \{c\}$$



$$a \perp b | \{\} \quad a \not\perp b | \{c\}$$

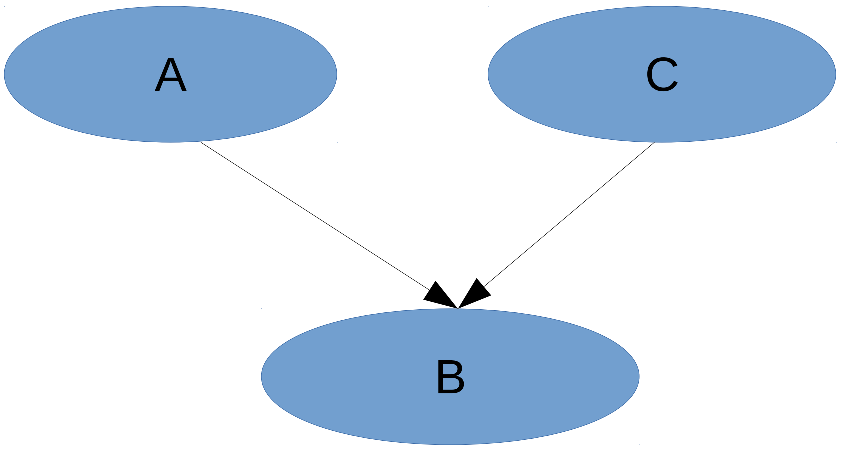


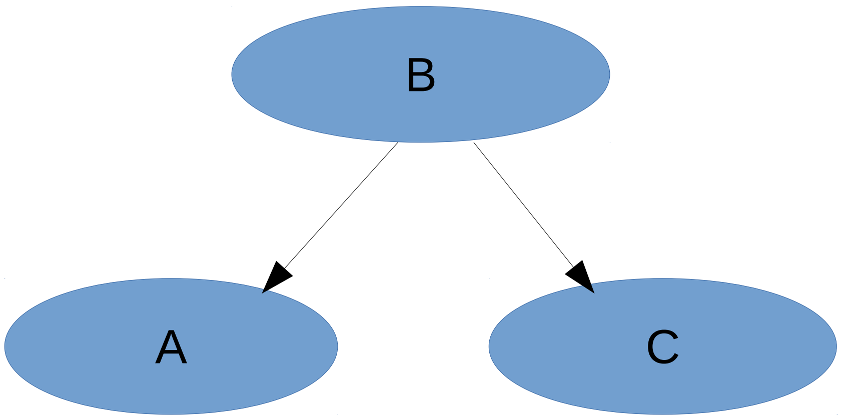




Some intuition

- A correlated with B
- B correlated with C
- C not correlated with A

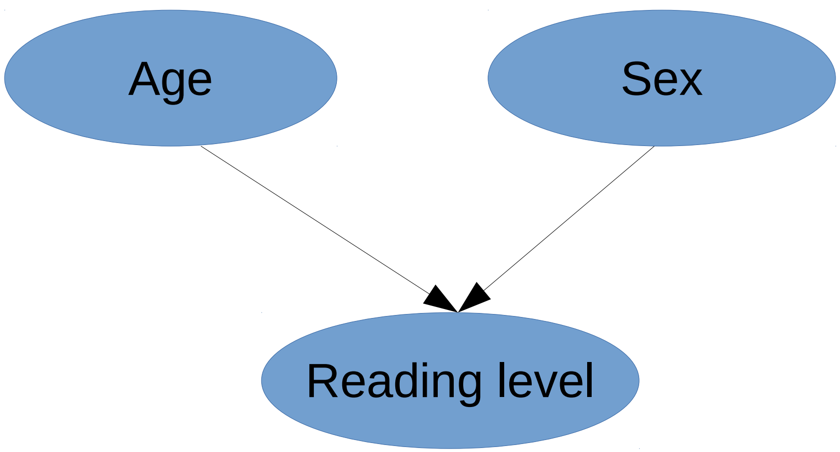


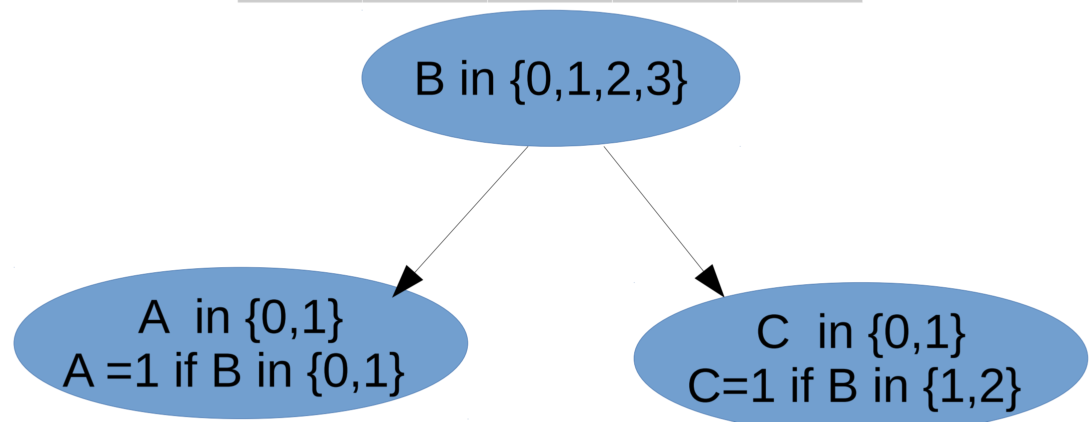


Some intuition

- A correlated with B
- B correlated with C
- C not correlated with A

B	0	1	2	3
P(B)	0.25	0.25	0.25	0.25





B	0	1	2	3
P(A=1)	1	1	0	0

B	0	1	2	3
P(C=1)	0	1	1	0

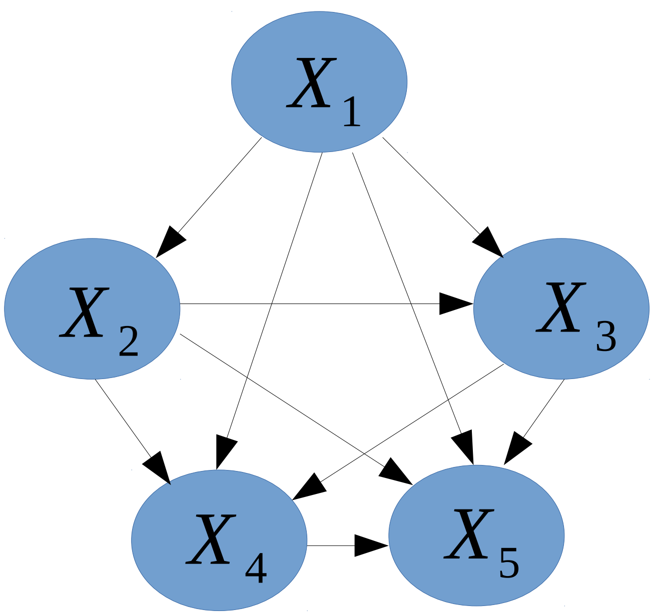
$$C \perp A$$

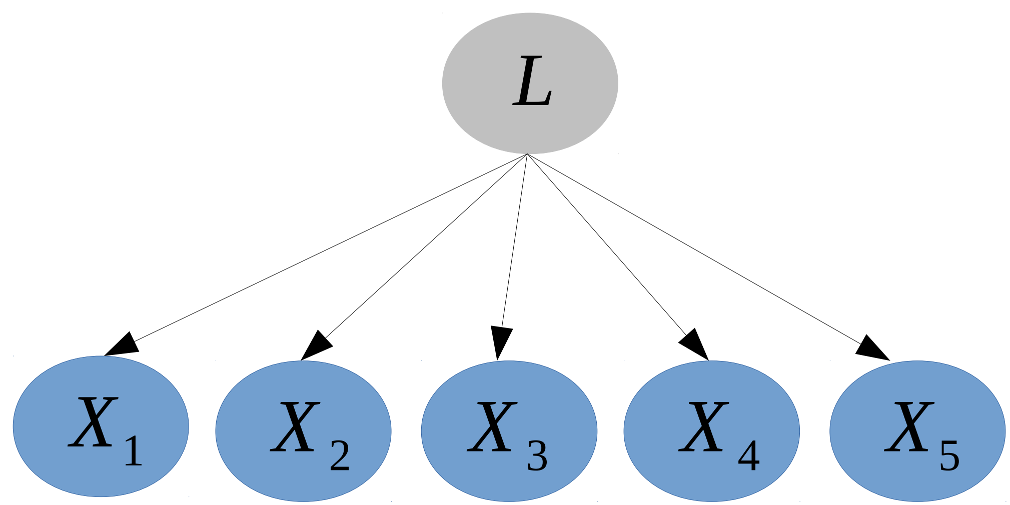
$$C \perp A \mid B$$

$$P(C=1 \mid A=1)=0.5=P(C) \Rightarrow C \perp A$$

When is correlation causation?

- Can we infer causal structure from observed conditional independences?





Faithfulness

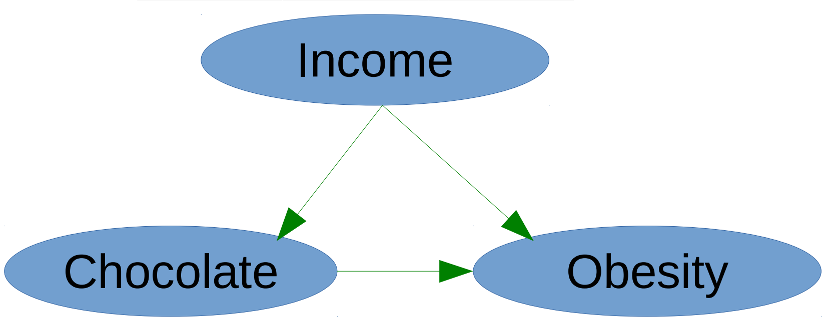
- We have some distribution P that was generated by some true (unknown) causal structure.
- We **assume** that all the conditional independences in P are represented in the true causal structure
- Finding the causal structure equates to finding the networks that encode exactly the dependencies we have observed in our data.

Sparsity and causal ordering

- Can you have a network that represents a distribution, and is more sparse than the true causal model that generated that distribution?

True causal structure generating P

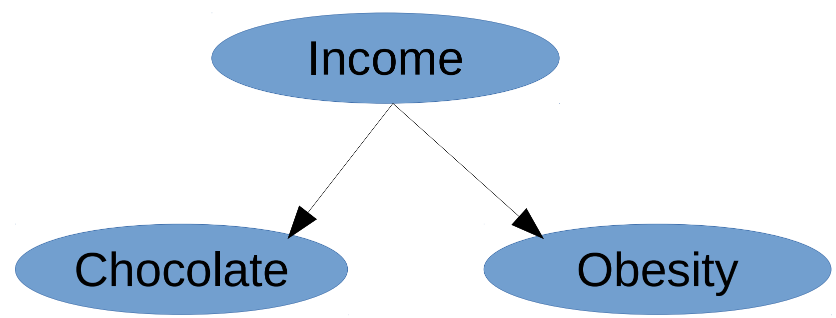
I	High	Low
P(I)	0.5	0.5



I	H	L
P(C=Y)	.863	.2

I	High		Low	
C	Y	N	Y	N
P(O=Y)	.4	.2	.7	.5

Bayesian network representing P



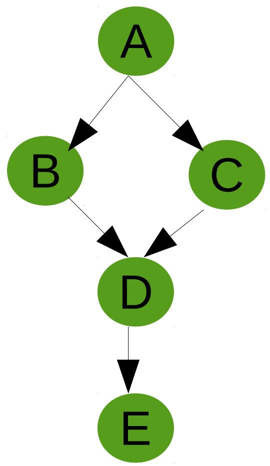
The IC algorithm

1. $for\ all\ pairs\ of\ variables\ a, b\ search\ for\ a\ set\ S_{ab}\ such\ that\ a \perp b | S_{ab}.$
   $If\ there\ is\ no\ such\ set,\ then\ draw\ an\ undirected\ link\ between\ them.$

2. $for\ all\ pairs\ of\ non-linked\ nodes\ with\ a\ common\ neighbour, c,:$
   $If\ c \notin S_{ab}\ direct\ links\ towards\ c$

3. *Orient any undirected edges so as to avoid creating cycles or additional v-structures*

True Causal Model



$$B \perp C | A$$
$$D \perp A | B, C$$
$$E \perp A, B, C | D$$

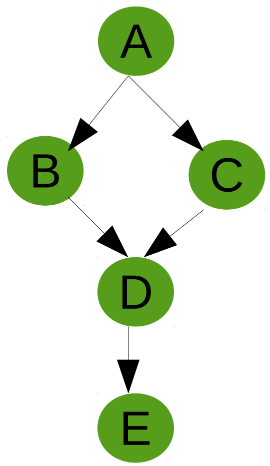
The IC algorithm

1. $for\ all\ pairs\ of\ variables\ a, b\ search\ for\ a\ set\ S_{ab}\ such\ that\ a \perp b | S_{ab}.$
   $If\ there\ is\ no\ such\ set, then\ draw\ an\ undirected\ link\ between\ them.$

2. $for\ all\ pairs\ of\ non-linked\ nodes\ with\ a\ common\ neighbour, c, :$
   $If\ c \notin S_{ab}\ direct\ links\ towards\ c$

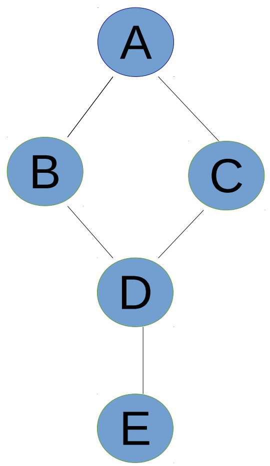
3. *Orient any undirected edges so as to avoid creating cycles or additional v-structures*

True Causal Model



$$B \perp C | A$$
$$D \perp A | B, C$$
$$E \perp A, B, C | D$$

Inferred at step 1



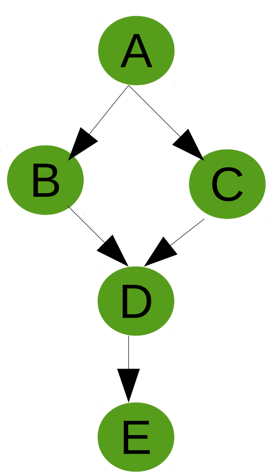
The IC algorithm

1. $for\ all\ pairs\ of\ variables\ a, b\ search\ for\ a\ set\ S_{ab}\ such\ that\ a \perp b | S_{ab}.$
   $If\ there\ is\ no\ such\ set, then\ draw\ an\ undirected\ link\ between\ them.$

2. $for\ all\ pairs\ of\ non-linked\ nodes\ with\ a\ common\ neighbour, c,:$
   $If\ c \notin S_{ab}\ direct\ links\ towards\ c$

3. *Orient any undirected edges so as to avoid creating cycles or additional v-structures*

True Causal Model

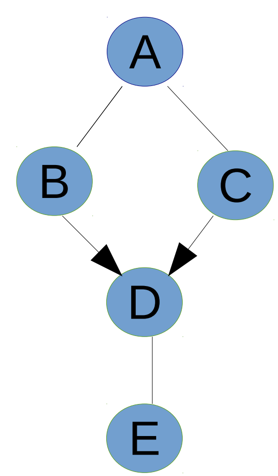


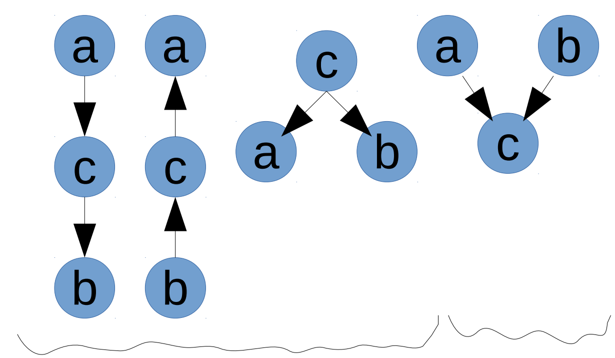
$B \perp C | A$

$D \perp A | B, C$

$E \perp A, B, C | D$

Inferred at step 2





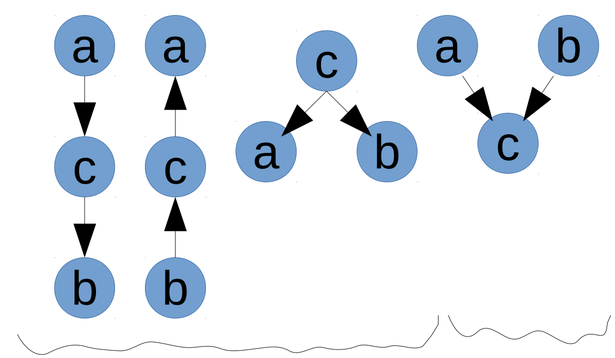
$a \perp b | \{c\}$ $\quad$ $a \perp b | \{\}$

The IC algorithm

1. *for all pairs of variables* a, b *search for a set* S_{ab} *such that* $a \perp b | S_{ab}$.
   If there is no such set, then draw an undirected link between them.

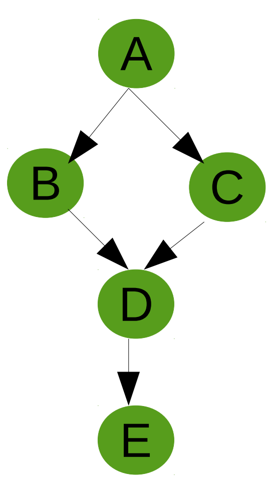
2. *for all pairs of non−linked nodes with a common neighbour,* c*,:*
   If $c \notin S_{ab}$ *direct links towards* c

3. *Orient any undirected edges so as to avoid creating cycles or additional v-structures*



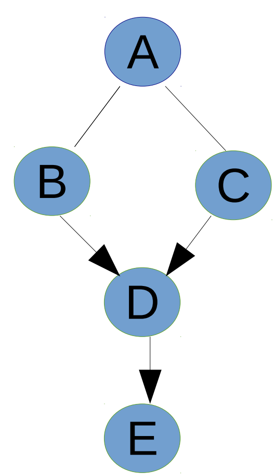
$$a \perp b | \{c\} \qquad a \perp b | \{\}$$

True Causal Model



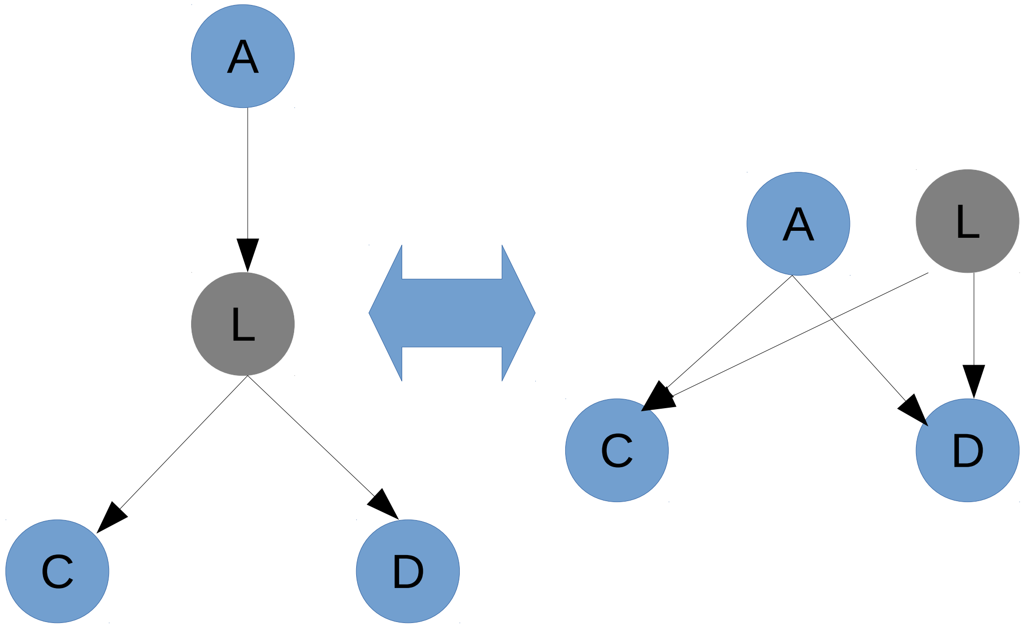
$$B \perp C | A$$
$$D \perp A | B, C$$
$$E \perp A, B, C | D$$

Inferred output



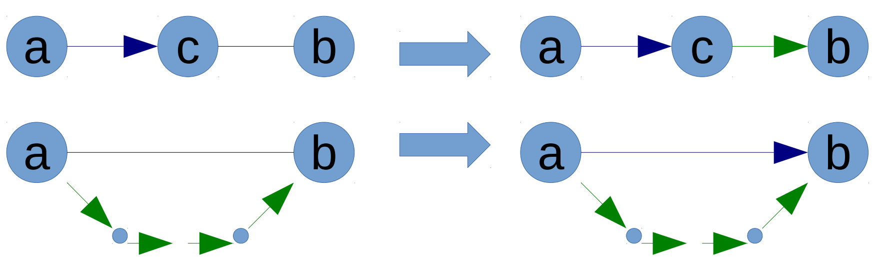
Latent variables

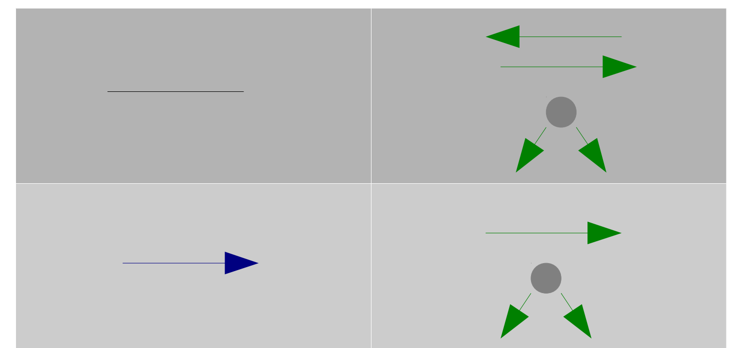
Theorem (Verma 1993):
for any latent structure there is an equivalent structure such that every latent variable is a root node with exactly 2 children.



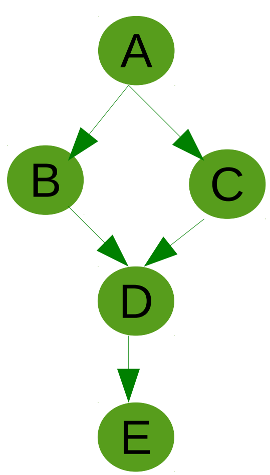
The IC* algorithm

1. *for all pairs of variables* a, b *search for a set* S_{ab} *such that* $a \perp b \mid S_{ab}$.
   If there is no such set, then draw an undirected link between them.
2. *for all pairs of non−linked nodes with a common neighbour,* c*,:*
   If $c \notin S_{ab}$ *direct links towards* c
3. *Recursively add arrowheads/mark edges according to:*



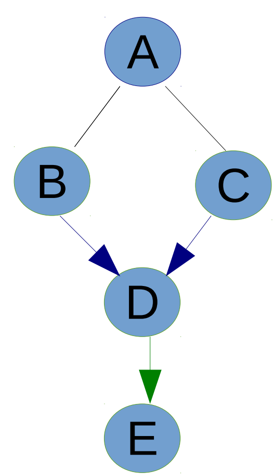


True Causal Model



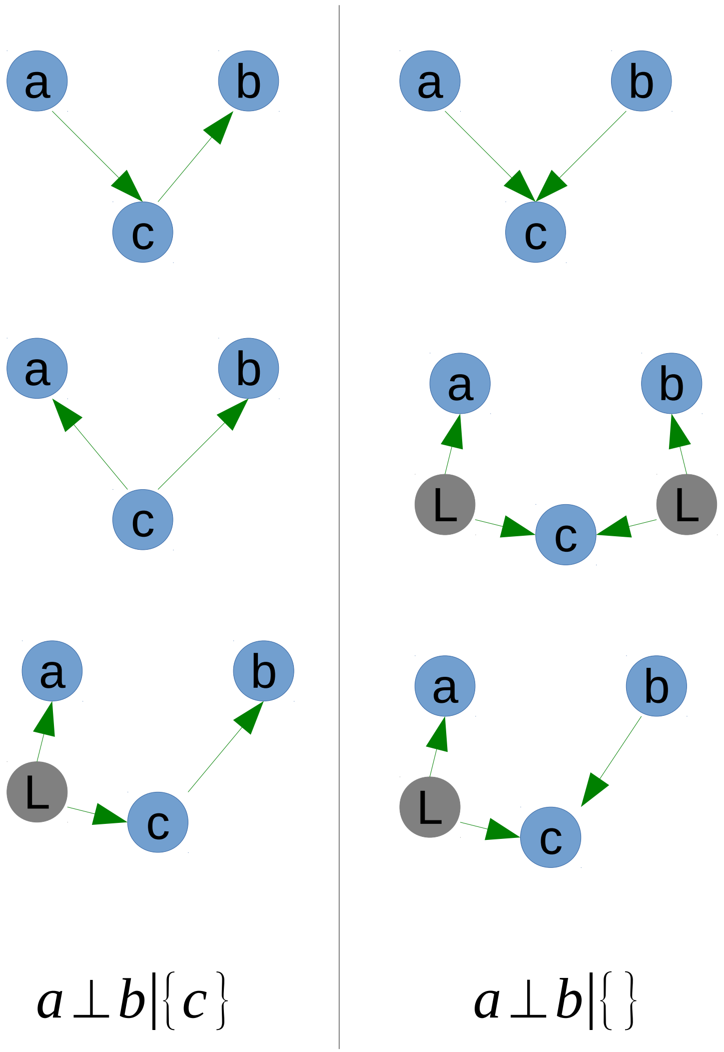
$$B \perp C \mid A$$
$$D \perp A \mid B, C$$
$$E \perp A, B, C \mid D$$

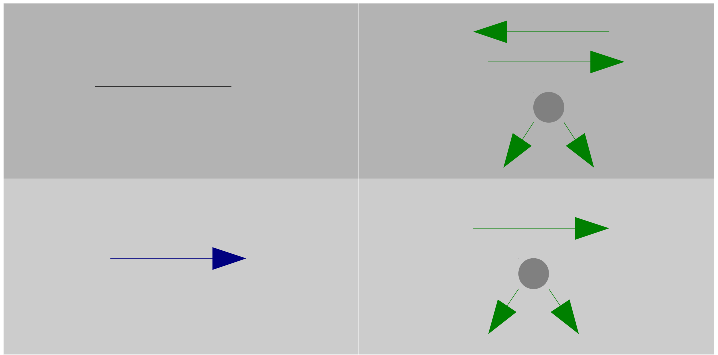
Inferred output



The IC* algorithm

2. $for\ all\ pairs\ of\ non-linked\ nodes\ with\ a\ common\ neighbour\ , c\ ,:$
$If\ c \notin S_{ab}\ direct\ links\ towards\ c$



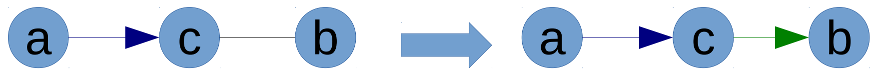


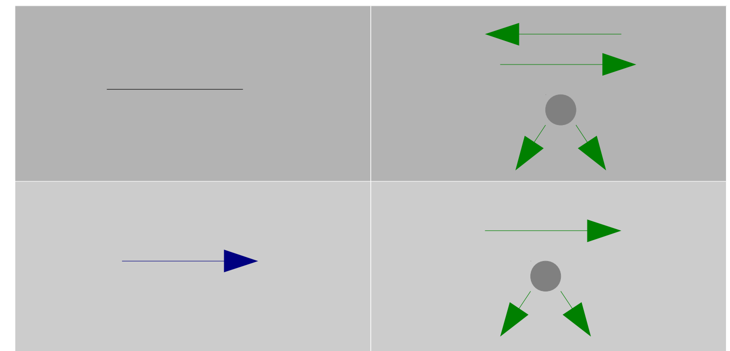
$a \perp b | \{c\}$

$a \perp b | \{\}$

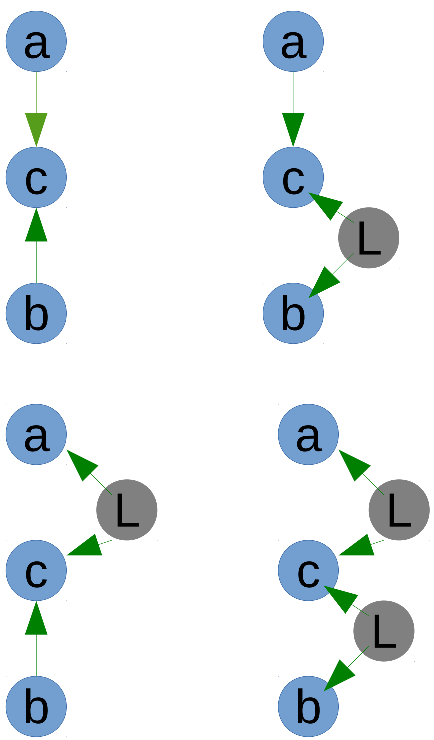
The IC* algorithm

3. *Add arrowheads/mark edges according to:*

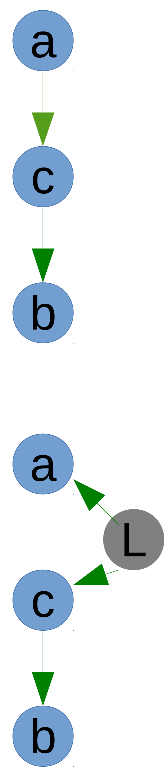




We know $a \perp b | c$ and $a \not\perp b | S$

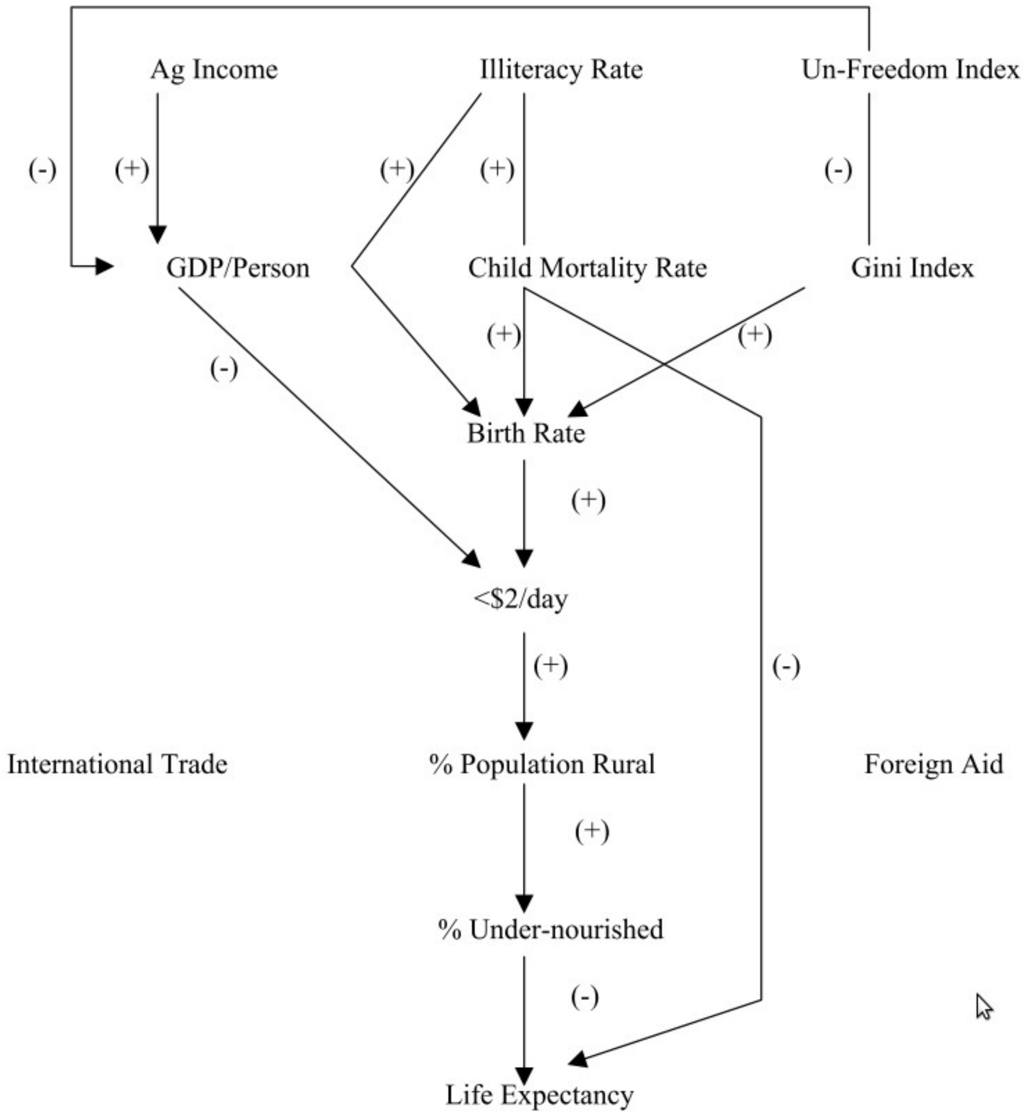


$a \not\perp b | c$



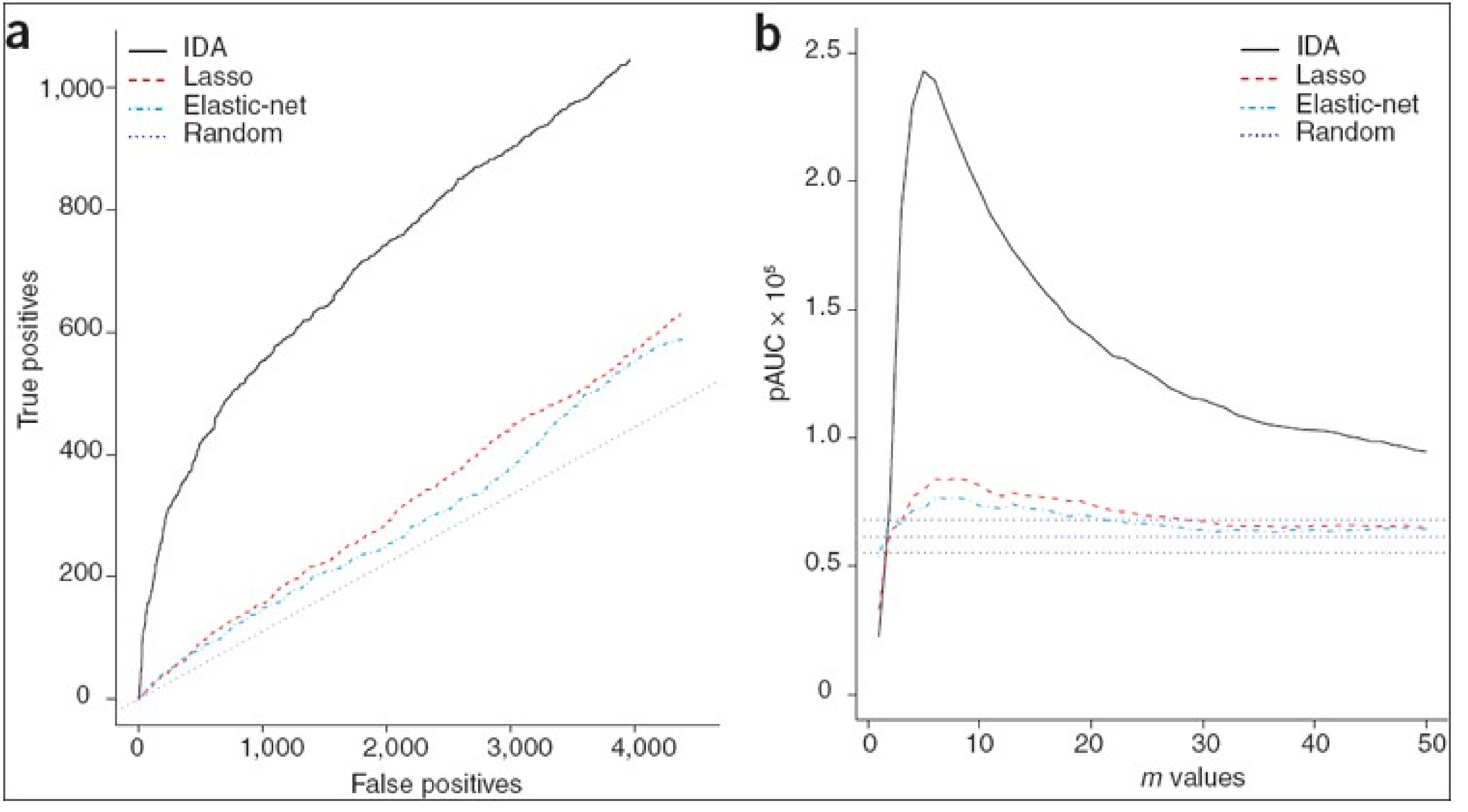
$a \perp b | c$

Causes of poverty



Bessler, David A. "On world poverty: Its causes and effects." Food and Agricultural Organization (FAO) of the United Nations, Research Bulletin, Rome(2003).

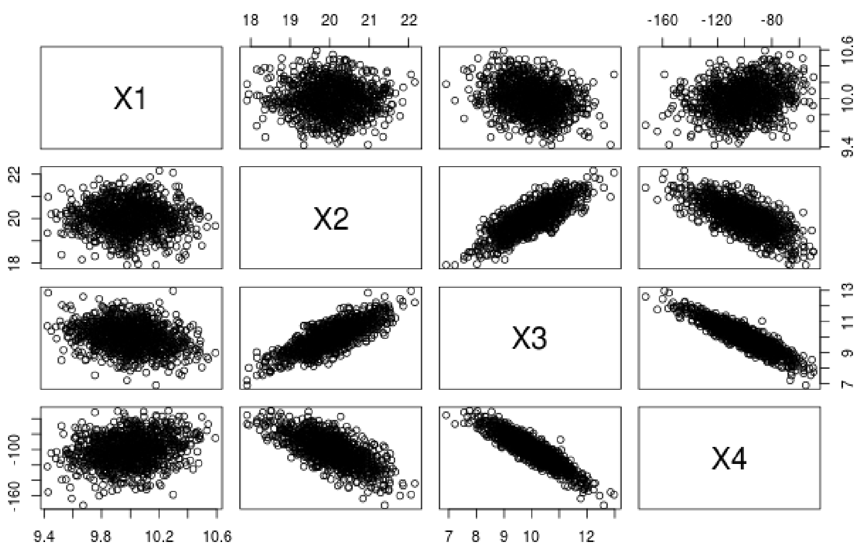
Uncovering causal links between genes and gene expression

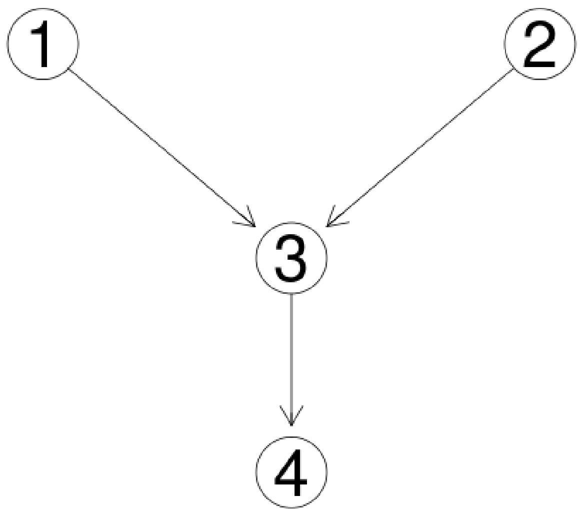


Maathuis, Marloes H., et al. "Predicting causal effects in large-scale systems from observational data." Nature Methods 7.4 (2010): 247-248.

Causal structure learning in R (pcalg)

```
library('pcalg')
n = 1000
X1 = rnorm(n,mean=10,sd=.2)
X2 = rnorm(n,mean=20,sd=.7)
X3 = X2-X1+rnorm(n,mean=0,sd=.5)
X4 = -X3^2+rnorm(n,mean=0,sd=8)
df = data.frame(X1,X2,X3,X4)
plot(df)
suffStat <- list(C = cor(df),n=nrow(df))
pc.3var = pc(suffStat,indepTest=gaussCItest,p=ncol(df),alpha=0.01)
plot(pc.3var, main = "")
```





References

Pearl, J. (2000). *Causality: models, reasoning and inference* (particularly chapter 2)

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. (chapters 3 & 21)

Verma 1993 *Graphical aspects of causal models Technical* Report. UCLA

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*.

Maathuis, Marloes H., et al. (2010) *Predicting causal effects in large-scale systems from observational data.* Nature Methods 7.4 : 247-248.

Kalisch, Markus, et al. (2012) Causal inference using graphical models with the R package pcalg. Journal of Statistical Software 47.11 : 1-26.

Bessler, David A. (2003) *On world poverty: Its causes and effects*. Food and Agricultural Organization (FAO) of the United Nations, Research Bulletin, Rome .