When is Correlation Causation?

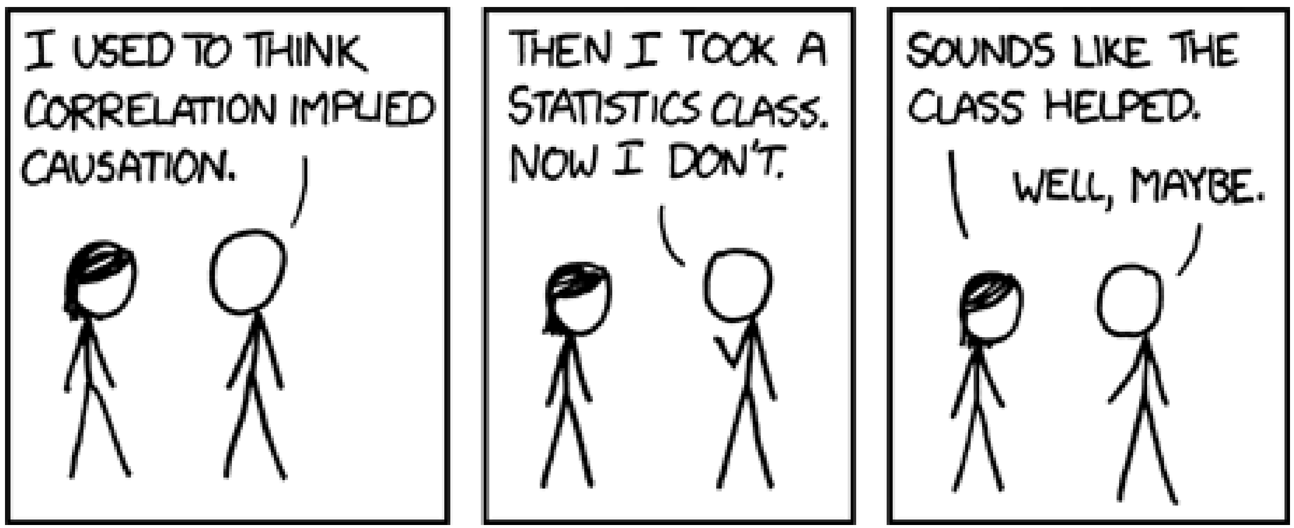
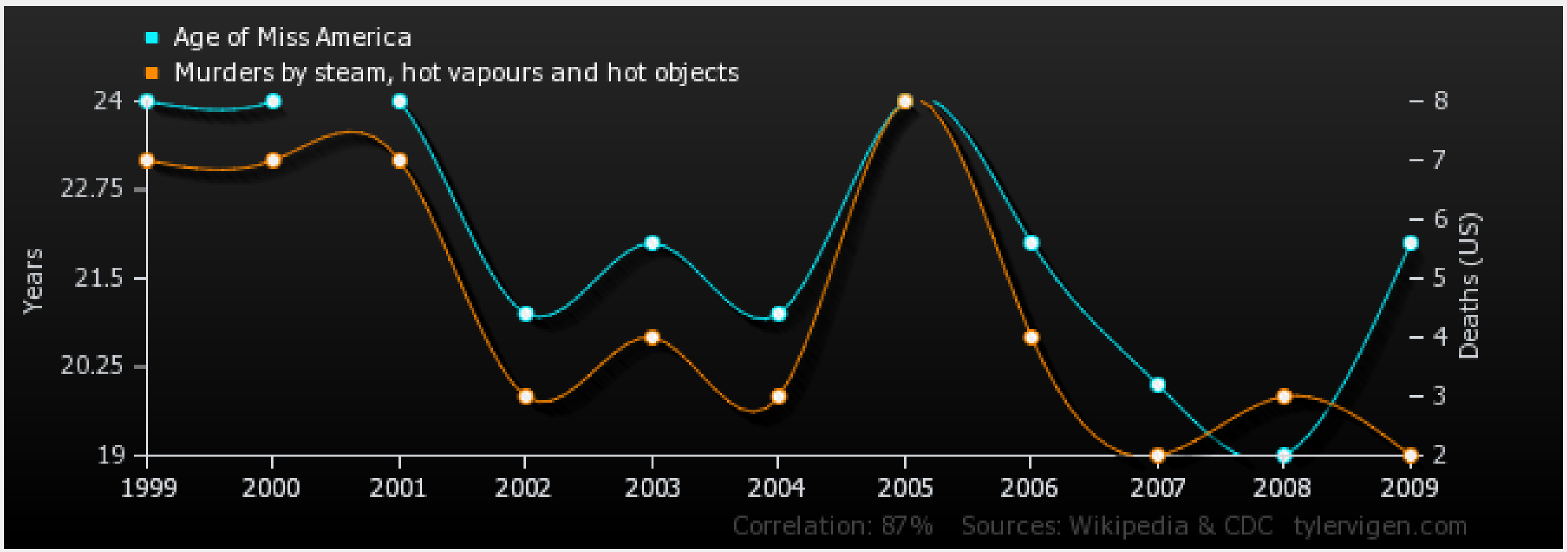


Image Source: http://xkcd.com/552/

When is Correlation Causation?

Age of Miss America
correlates with
Murders by steam, hot vapours and hot objects

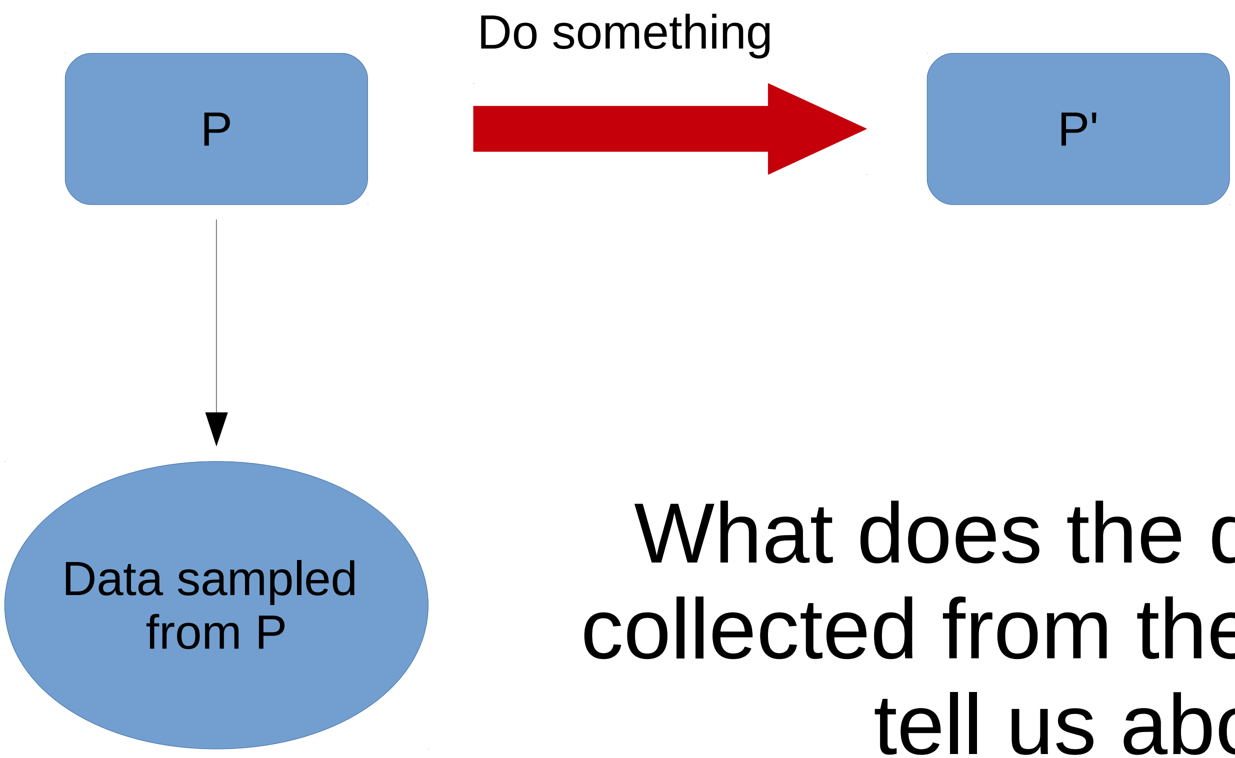


	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Age of Miss America Years (Wikipedia)	24	24	24	21	22	21	24	22	20	19	22
Murders by steam, hot vapours and hot objects Deaths (US) (CDC)	7	7	7	3	4	3	8	4	2	3	2

Correlation: 0.870127

Image source: www.tylervigen.com/

Causal inference



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

A framework for causal inference

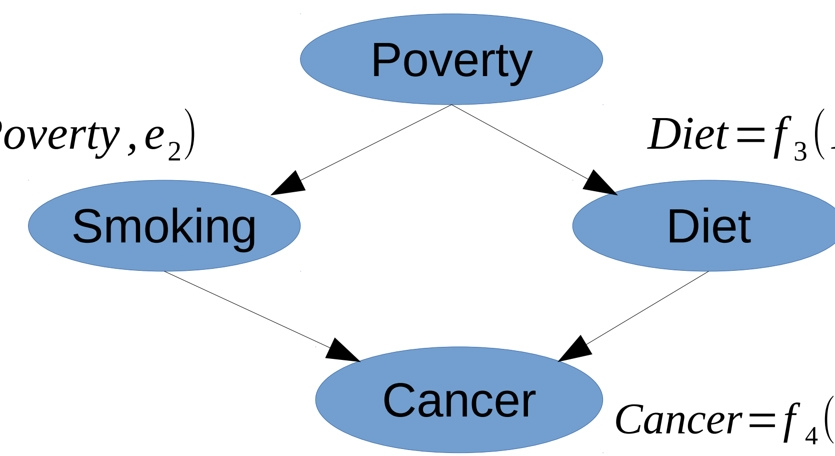
- Represent each variable as deterministic function of its direct causes and noise
- Noise terms must be mutually independent

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

$Poverty = f_1(e_1)$

$Smoking = f_2(Poverty, e_2)$

$Diet = f_3(Poverty, e_3)$



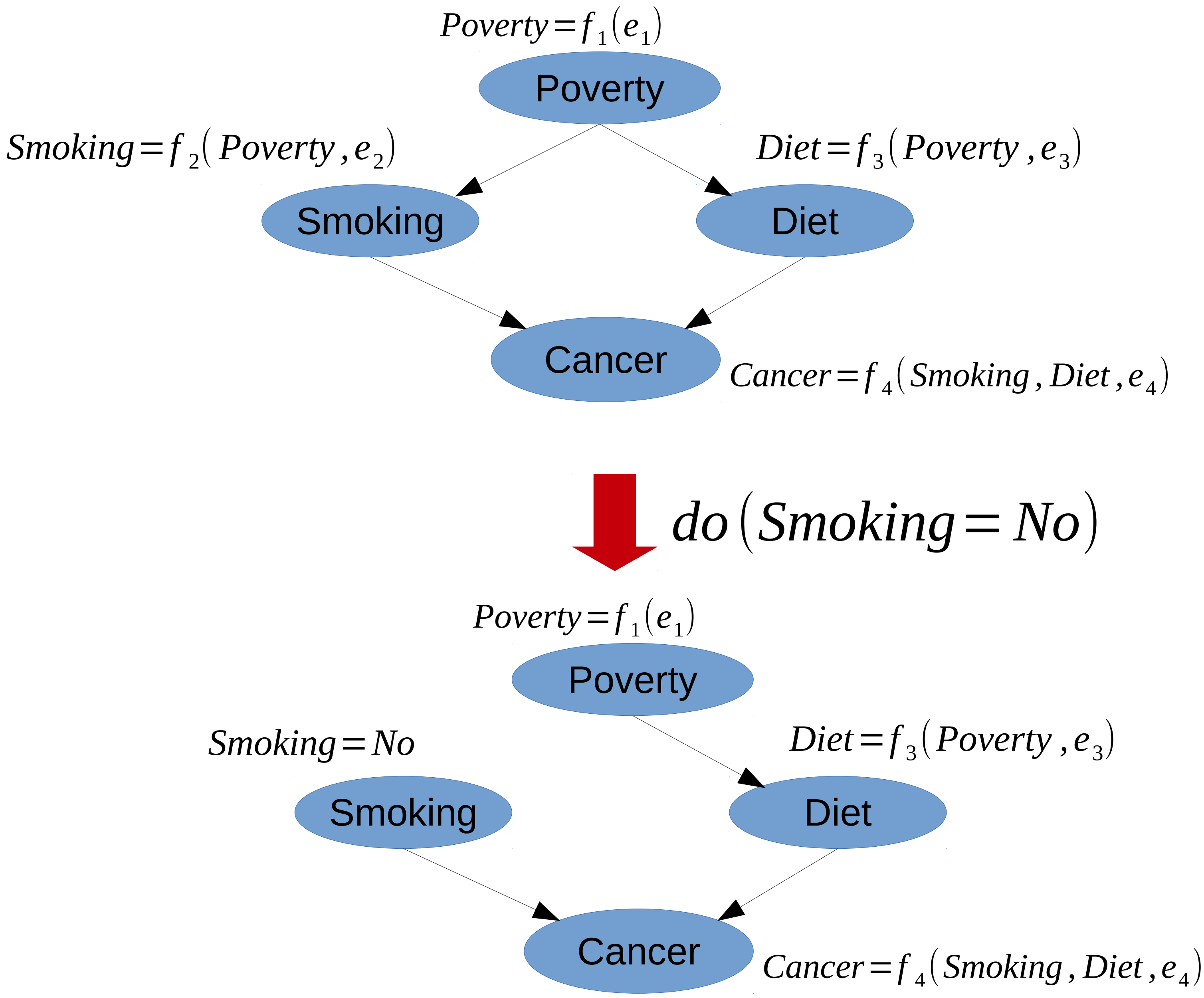
$Cancer = f_4(Smoking, Diet, e_4)$

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

=> The factorisation of the full distribution can be simplified

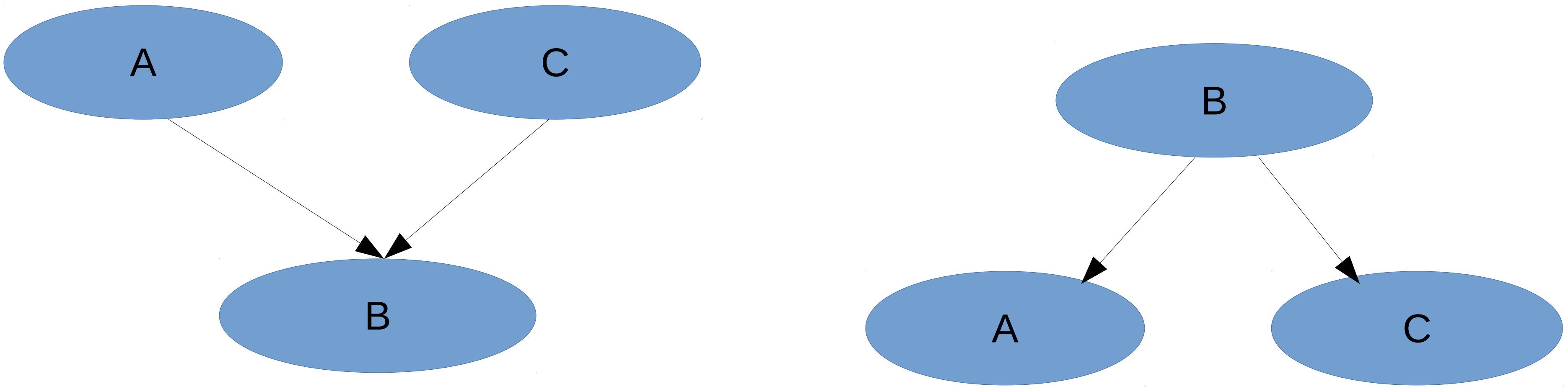
$$P(Po, S, D, C) = P(Po)P(S|Po)P(D|Po, S)P(C|Po, S, D) = P(Po)P(S|Po)P(D|Po)P(C|S, D)$$

A framework for causal inference



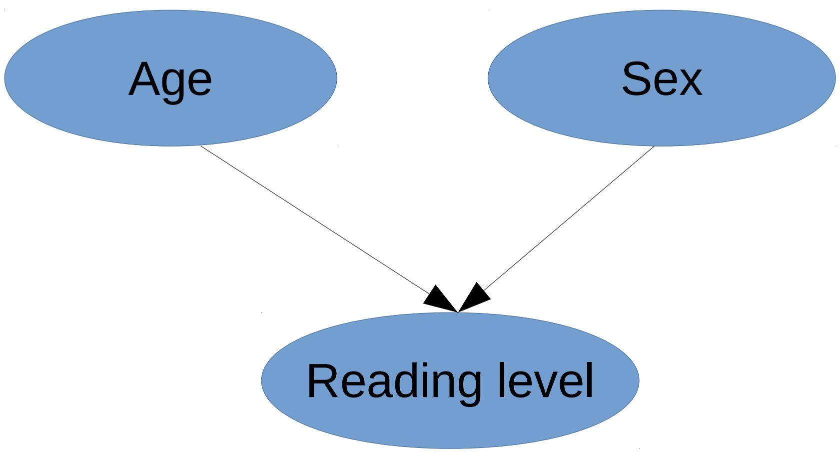
Some intuition

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

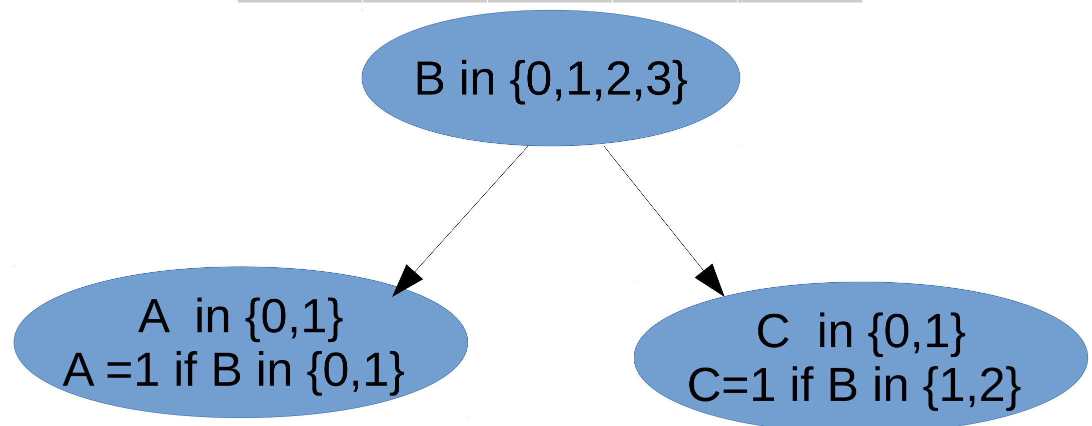


Some intuition

- 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

$$C \perp A$$

$$C \perp A | B$$

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

Bayesian Network Definition: A directed acyclic graph, whose nodes represent variables, where each variable is independent of its non-desendents given its parents (Markov independences)

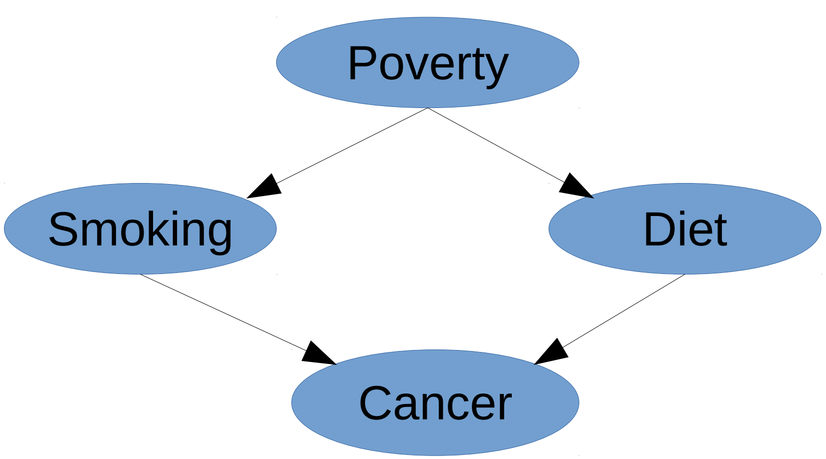
A graph G represents a distribution P if P can be factorised over G. In other words, if all the Markov independences implied by the graph structure are true in P, then G represents P.

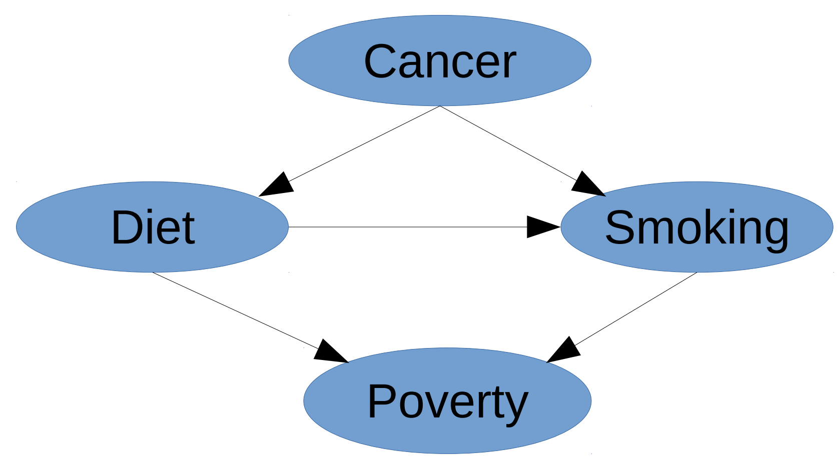
Note: The distribution can have additional independences not required by G, in other words if G represents a given distribution P, we can add any edges to G and it will still represent P.

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

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

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



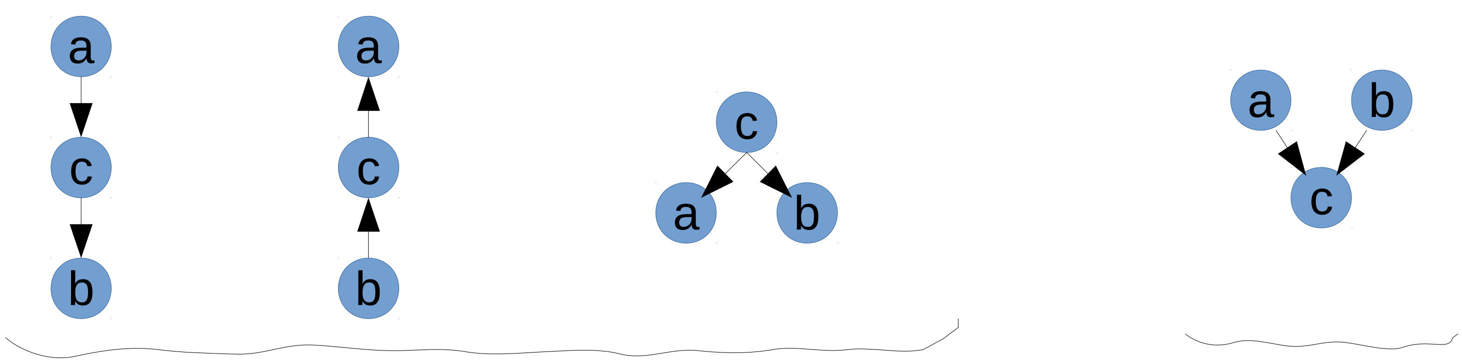


D-separation

- A given set of Markov independence relations, as represented by a graph G, often imply additional independence relations that are true for **all** distributions that factories over G

- D-separation gives us a method to read these dependencies directly off the graph.

Consider how ***a*** could be connected to ***b*** through ***c***.
- Is a independent of b given c is known?
- Is a independent of b if c is not known.

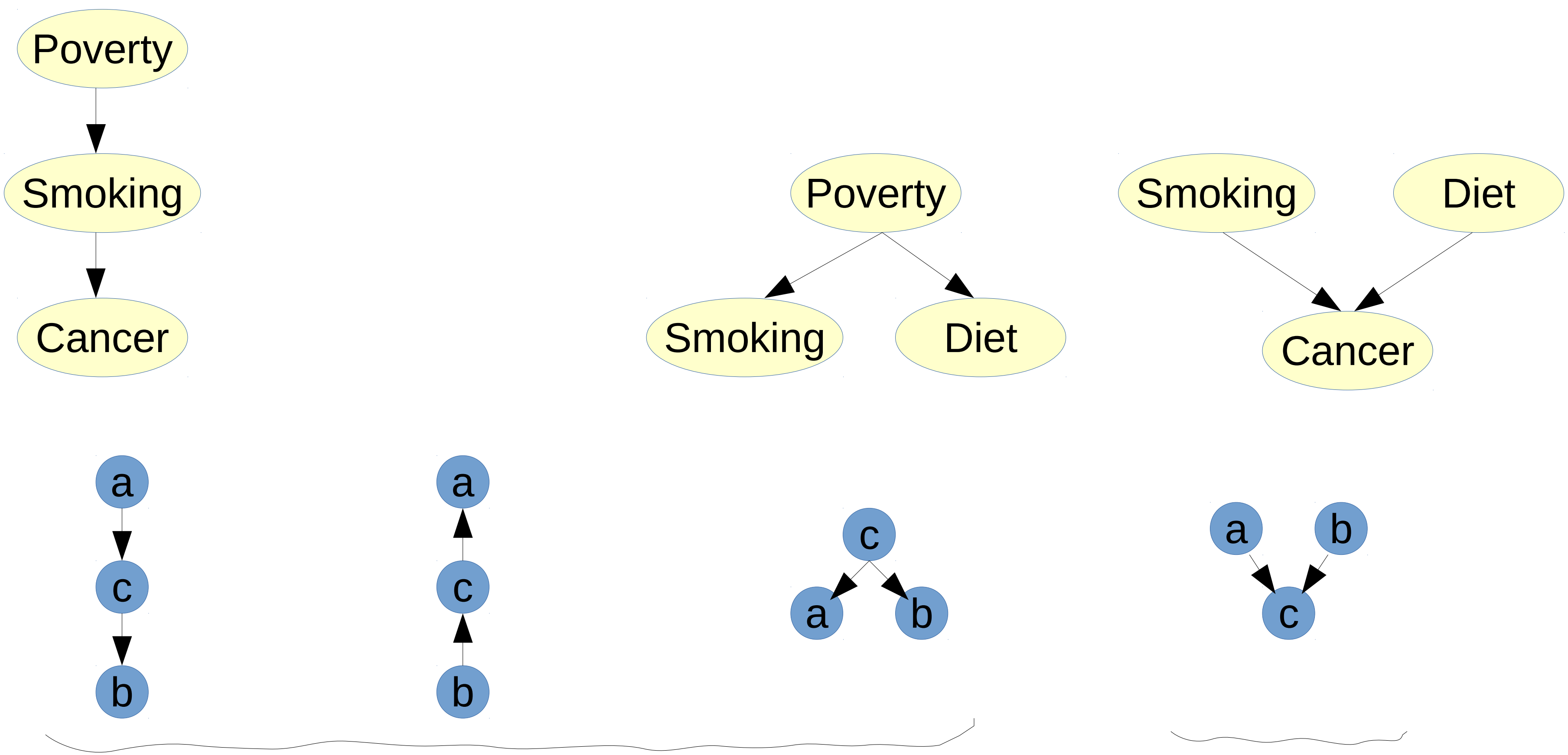


$$a \perp b \,|\, \{c\}$$

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

D-separation

When can influence flow from ***a*** to ***b*** through ***c*** ?



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

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

Influence can flow <=> trail is active if c is not observed

Influence can flow if c is observed

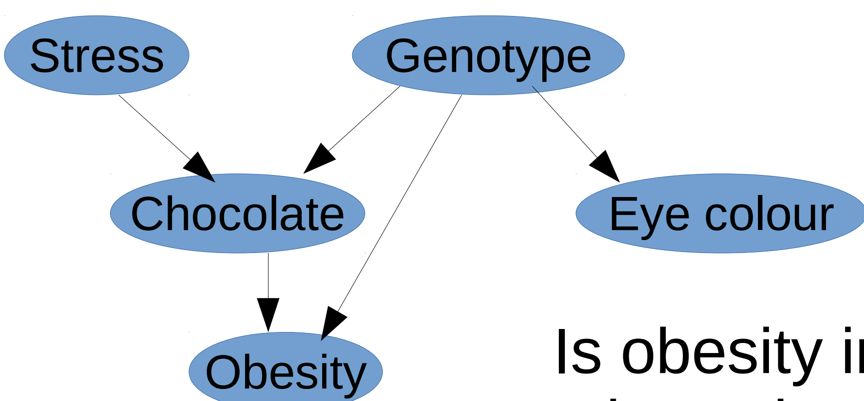
D-separation

A trail through variables $X_1 \ldots X_n$ is active if:

- Whenever we have a v-structure ($X_{i-1} \searrow \; \swarrow X_{i+1}$), Then X_i or a descendent is observed.
- No other variables in $X_1 \ldots X_n$ are observed

If there are no active trails from **a** to **b**, then they are D-separated with respect to the observed variables

Theorem: let *I(G)* represent the set of independences implied by d-separation. If *P* factorises over *G* then $I(G) \subseteq I(P)$ that is independences implied by d-separation apply to all distributions that factorise over that graph.



Is obesity independent of stress, given chocolate consumption?

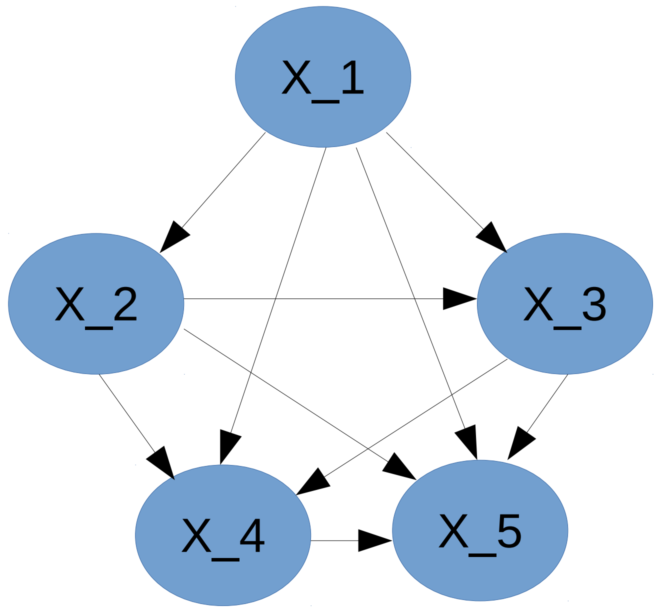
Faithfulness

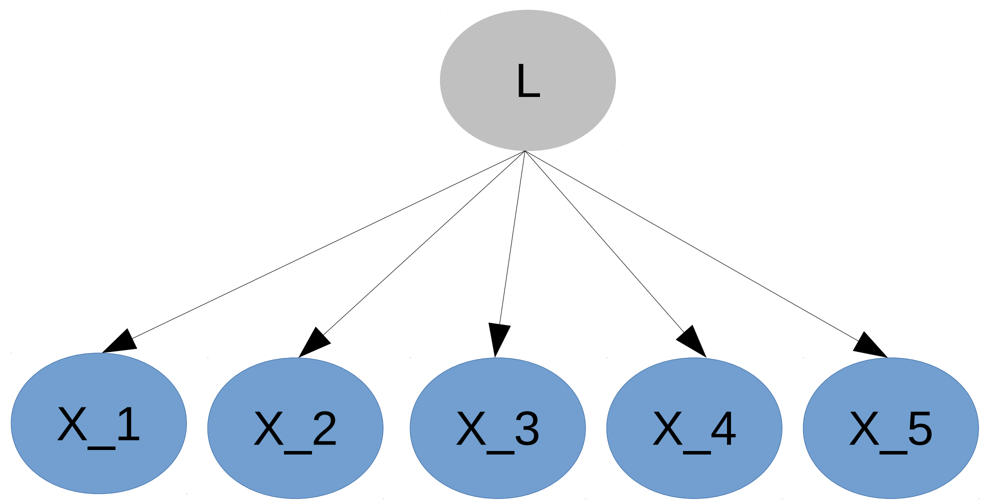
A distribution P is faithful to G if $I(P) = I(G)$

Theorem: For almost all distributions P that factorise over G, i.e. for all distributions except a set of measure zero in the space of conditional probability distribution parameterizations, $I(P) = I(G)$

When is correlation causation?

- Can we infer causal structure from observed conditional independences?



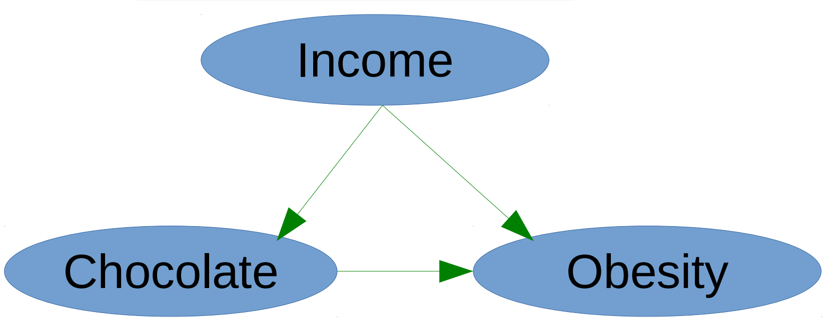


Sparsity and causal ordering

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

True causal structure generating P

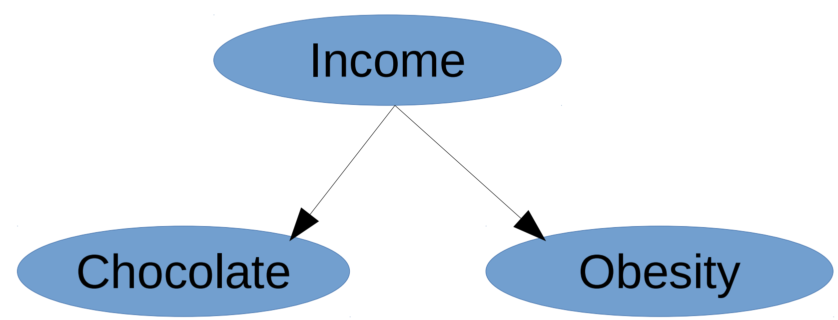
I	High	Low
P(I)	0.5	0.5



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



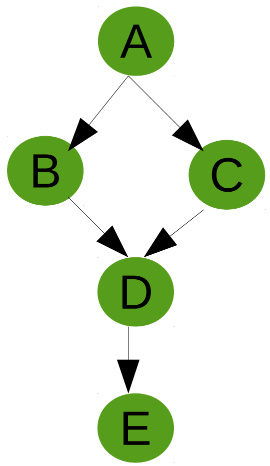
A perfect I-Map; the IC algorithm

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



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

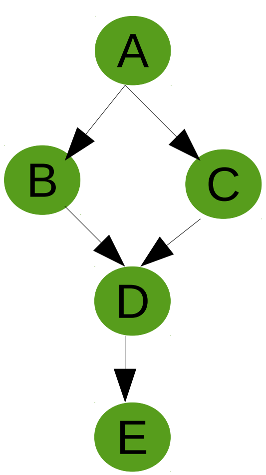
The IC algorithm

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

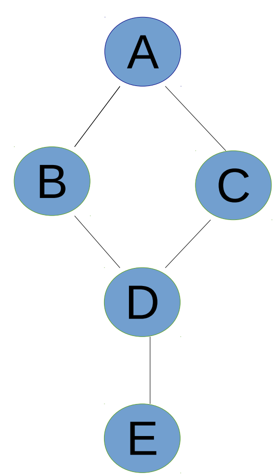
3. *Orient any undirected edges so as to avoid creating cycles or additional v-structures*

True Causal Model



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

Inferred at step 1



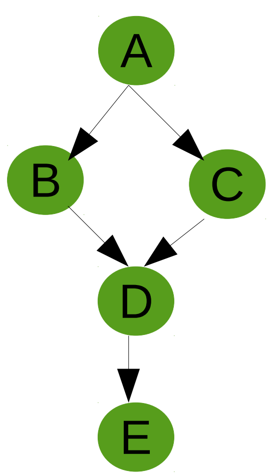
The IC algorithm

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

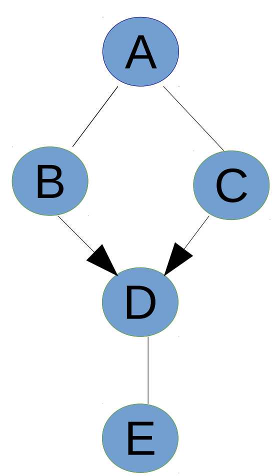
3. *Orient any undirected edges so as to avoid creating cycles or additional v-structures*

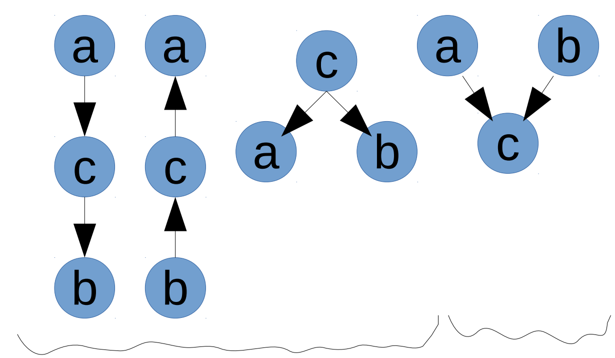
True Causal Model



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

Inferred at step 2





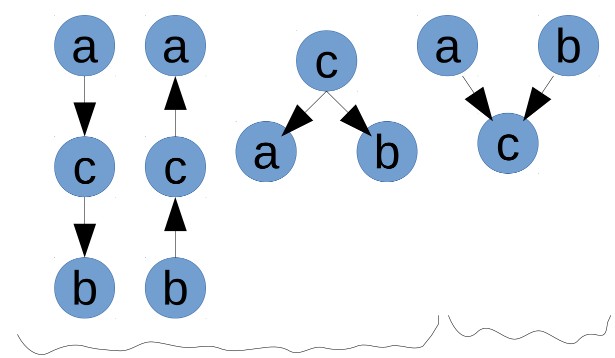
$a \perp b | \{c\}$ $\quad$ $a \perp b | \{\}$

A perfect I-Map

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.

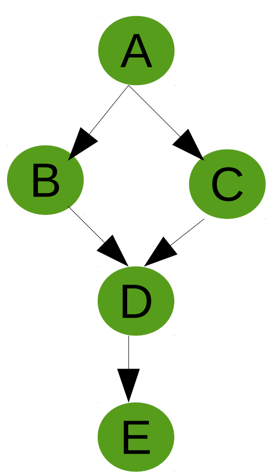
2. *for all pairs of non−linked nodes with a common neighbour,* c*,:*
   If $c \notin S_{ab}$ *direct links towards* c

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



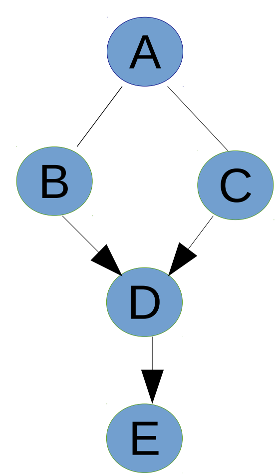
$a \perp b | \{c\}$ $\quad$ $a \perp b | \{\}$

True Causal Model



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

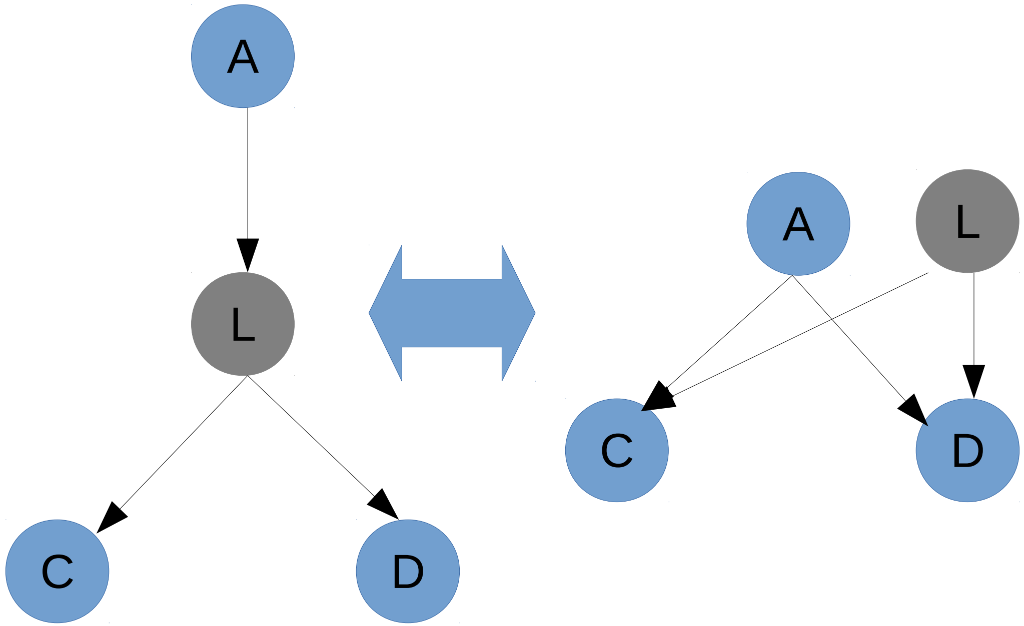
Inferred output



Latent variables

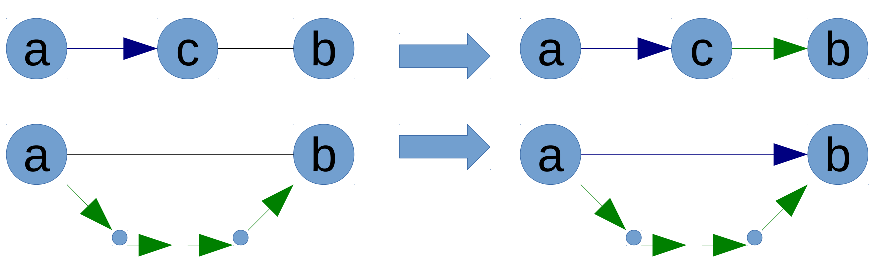
Theorem (Verma 1993):

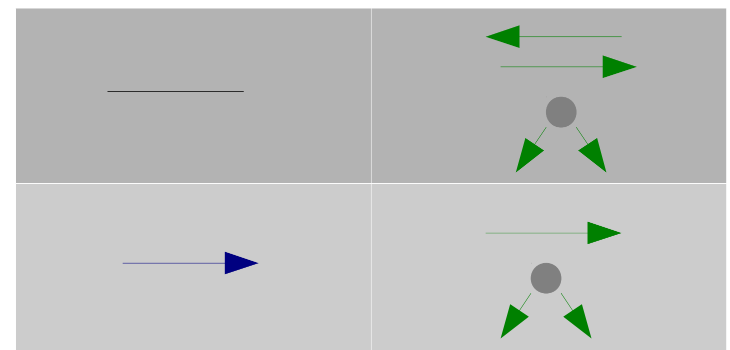
for any latent structure there is a dependency equivalent structure such that every latent variable is a root node with exactly 2 children.



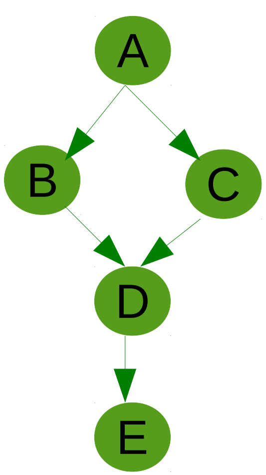
The IC* algorithm

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



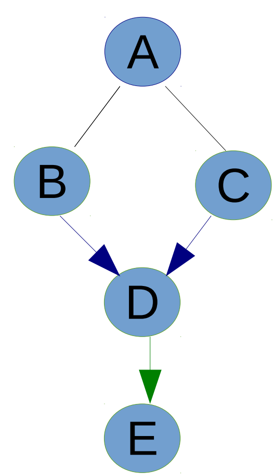


True Causal Model



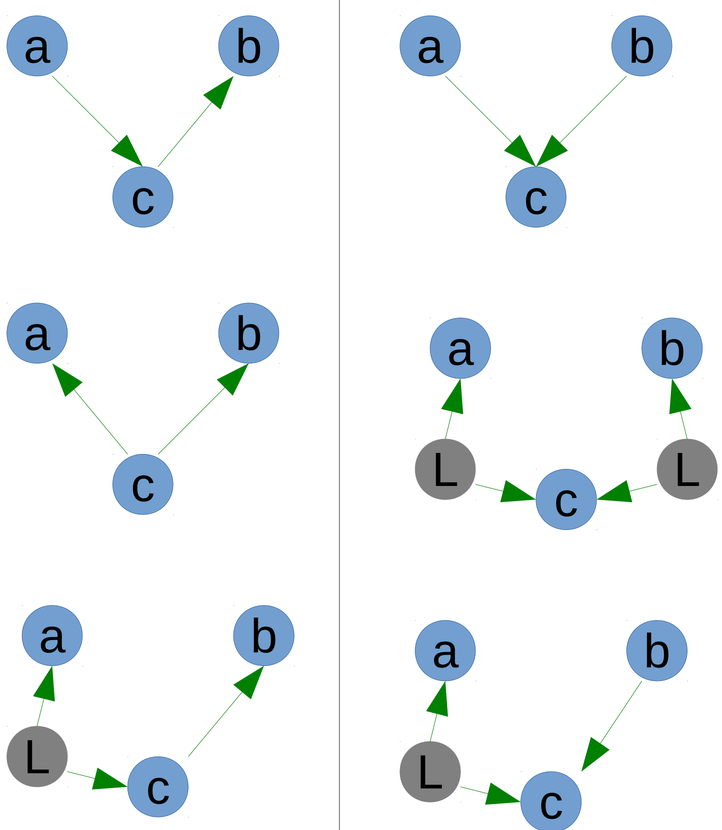
$$B \perp C | A$$
$$D \perp A | B, C$$
$$E \perp A, B, C | D$$

Inferred output

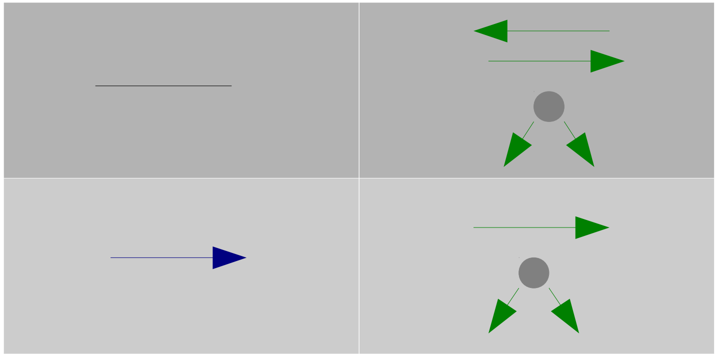


The IC* algorithm

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

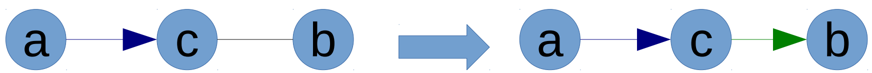


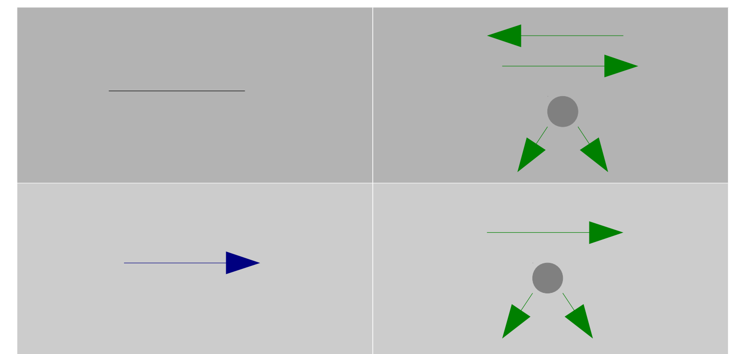
$a \perp b | \{c\}$ $\qquad$ $a \perp b | \{\,\}$



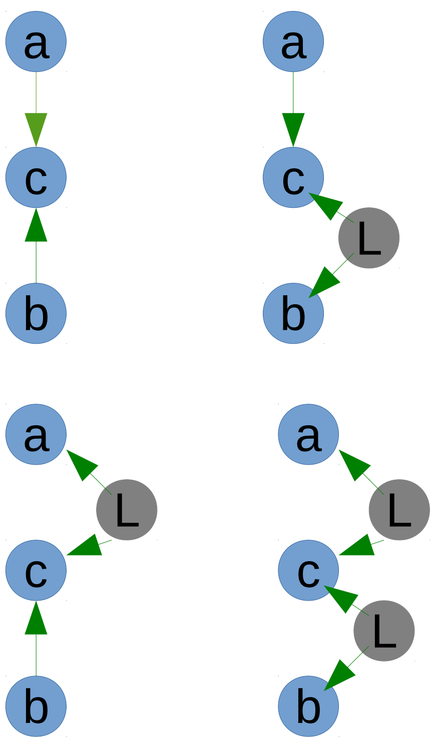
The IC* algorithm

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

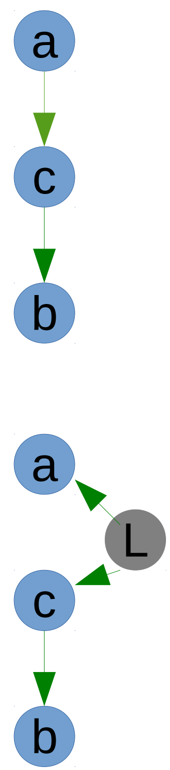




We know $a \perp b | c$ and $a \not\perp b | S$

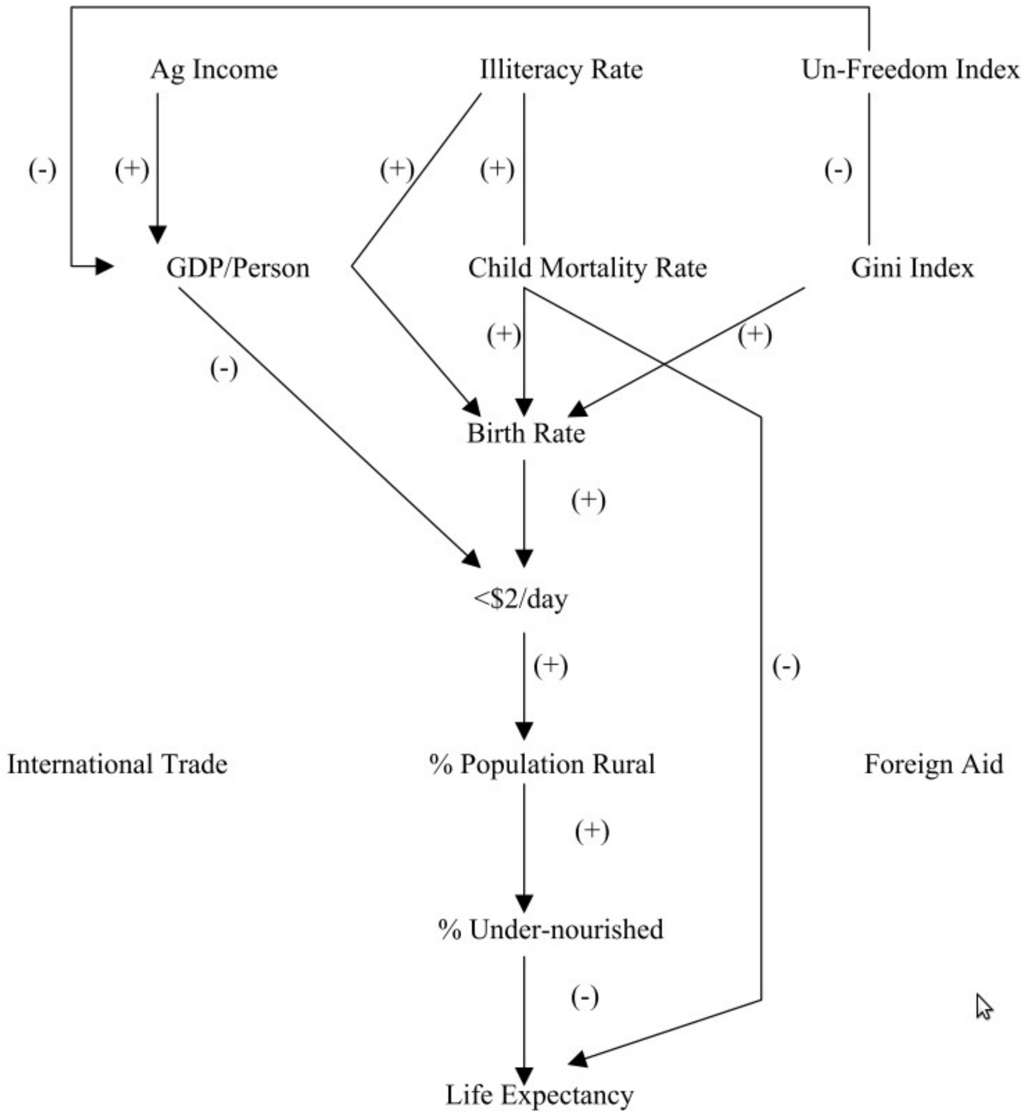


$a \not\perp b | c$



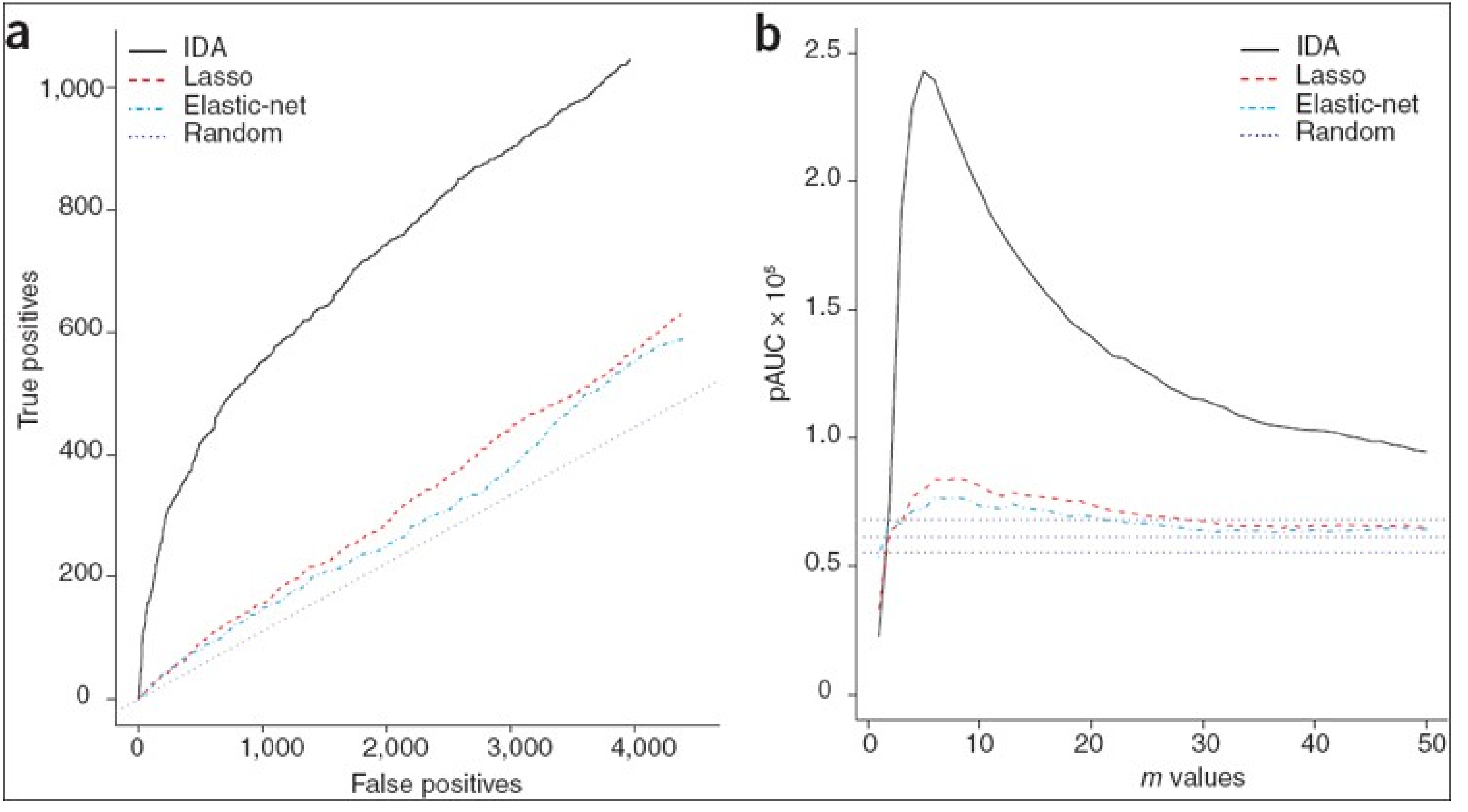
$a \perp b | c$

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).

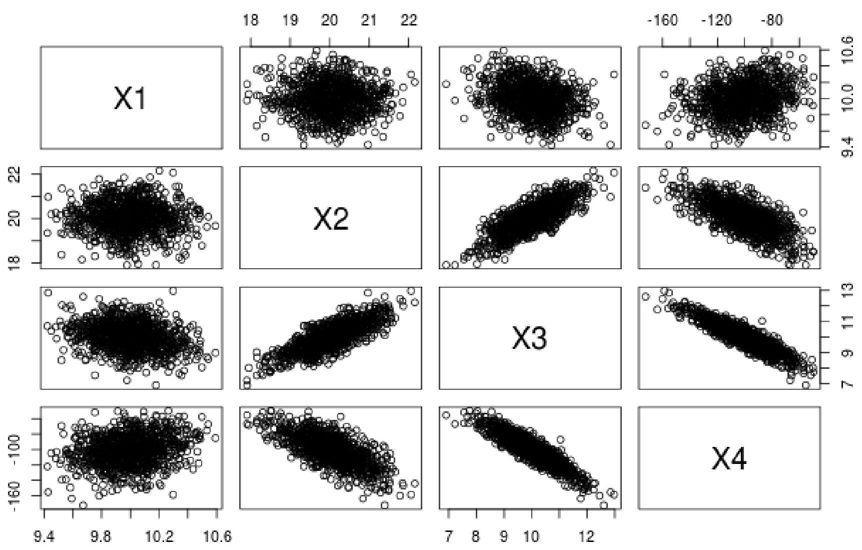
Uncovering causal links between genes and gene expression

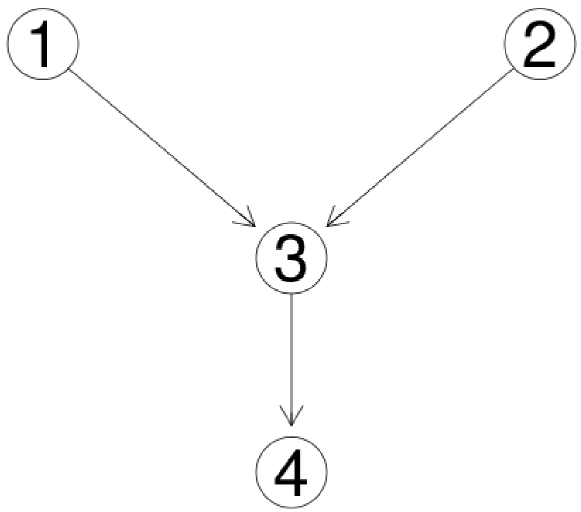


Maathuis, Marloes H., et al. "Predicting causal effects in large-scale systems from observational data." Nature Methods 7.4 (2010): 247-248.

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 .