Inferring Causality from Observational Data: a review

NICTA

Finnian Lattimore, Australian National University

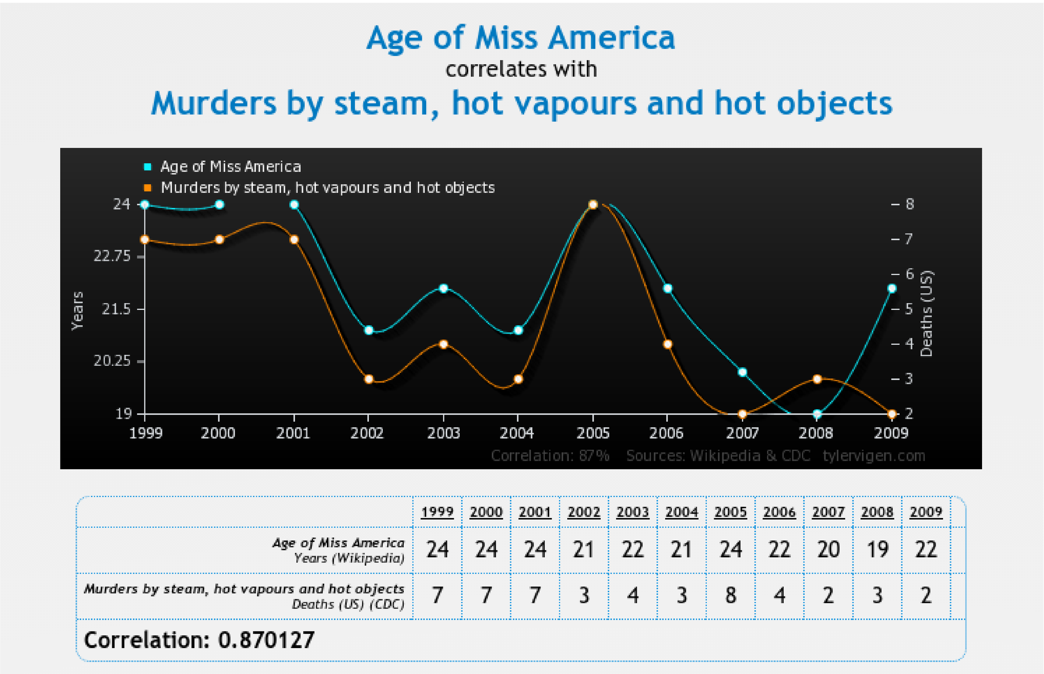
Introduction

Predicting the consequences of actions or interventions is central to many areas of science and policy; what would be the impact on the number of assaults if pubs were required to close at midnight, does living near a park reduce heart disease.

This kind of causal question cannot be addressed by standard machine learning techniques based on observational data, since the process of intervening changes the system. The ideal solution is to run randomised experiments. However, there are many situations in which this is prohibitively expensive, unethical or impossible. Determining causality from observational data requires assumptions. We review how a simple Occam's razor type assumption allows us to infer some aspects of causal structure without requiring interventional data.

Correlation is not causation

- A correlation may be strong and predictive but not a causal relationship (due to latent confounding variables)
- A correlation may be strong but not predictive, due to over-fitting/insufficient data.



	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

Image source: www.tylervigen.com/

- We consider only the former case and assume the data sets are sufficiently large that we can correctly infer dependencies between variables.

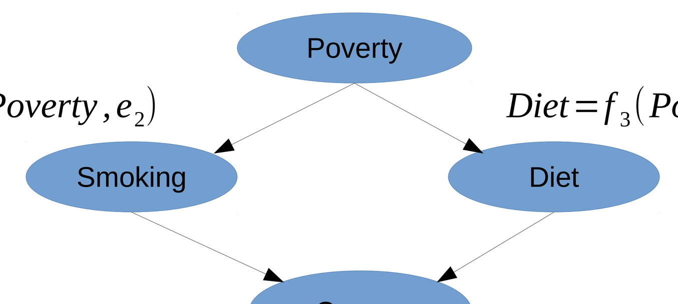
A framework for causal inference

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

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

$Poverty = f_1(e_1)$

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



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

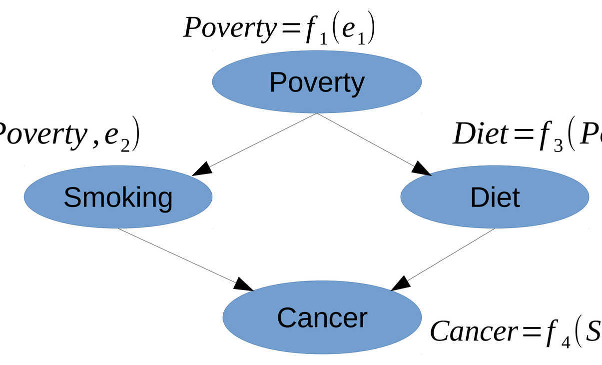
- Variables are independent of their non-effects given their direct causes

The Do calculus

- Provides a mechanism for estimating causal effects given a known causal structure.

$Poverty = f_1(e_1)$

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

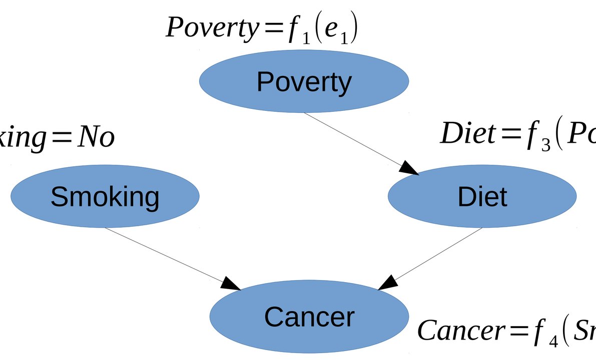


$Cancer = f_4(Smoking, Diet, e_4)$

$do(Smoking = No)$

$Poverty = f_1(e_1)$

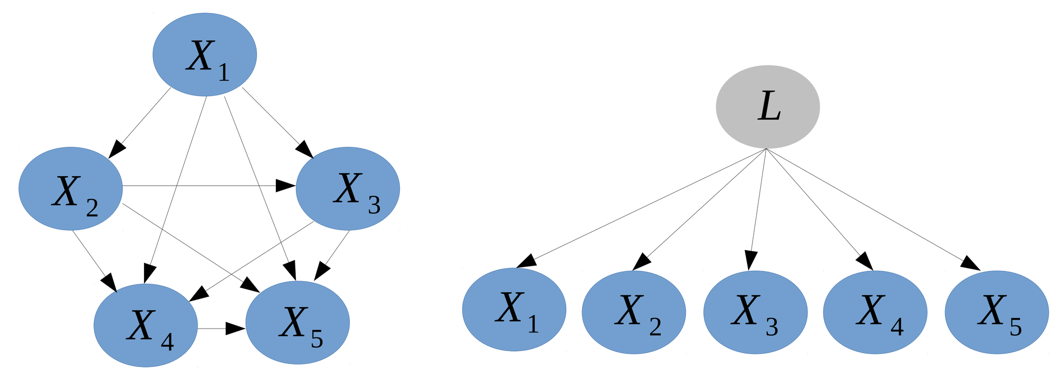
$Smoking = No$ $Diet = f_3(Poverty, e_3)$



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

Inferring causal structure

- Inferring causality is impossible without assumptions.
- The networks below can represent capture any set of conditional independence assertions over five observed variables.

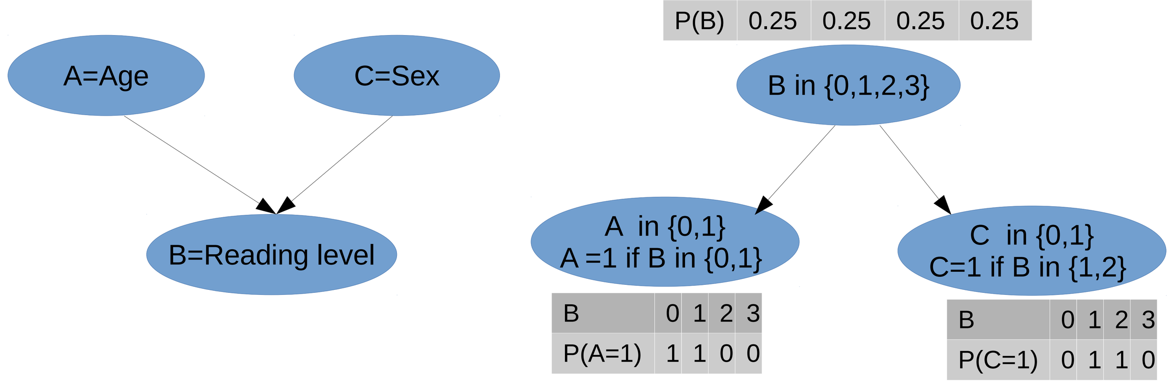


The faithfulness assumption

- Some causal structure can be inferred from observational data, given the assumption that all the conditional independences in the data ***P*** are represented by the true causal graph that generated the data, ***G***: $I(P) = I(G)$

Example: The dependencies below are faithfully represented by the network on the left. They are also represented by the network on the right, but not faithfully.

- 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

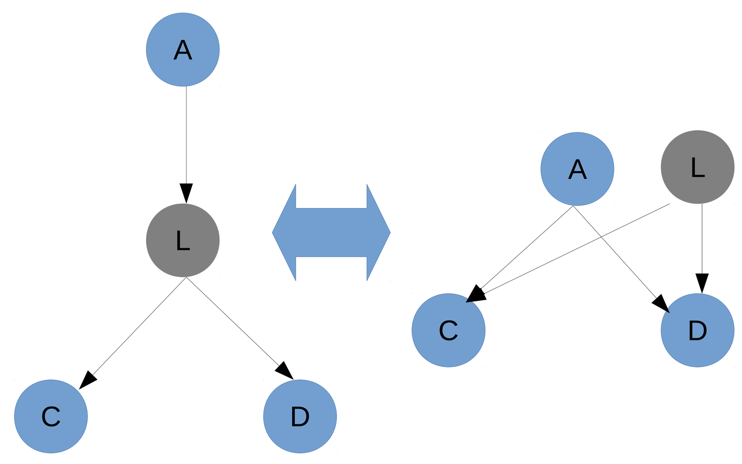
B	0	1	2	3
P(C=1)	0	1	1	0

$C \perp A$
Faithful

$C \perp A$ and $C \perp A | B$
Not Faithful

Latent variables

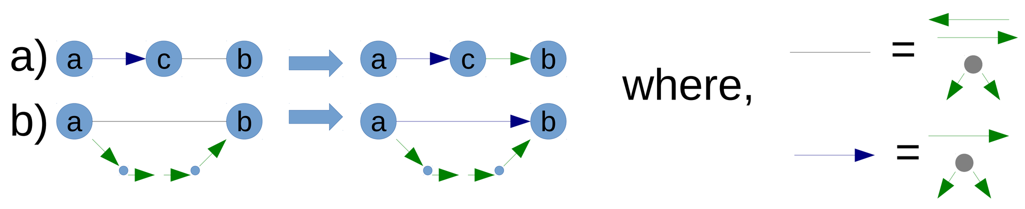
- **Theorem (Verma 1993):** for any latent structure there is an equivalent structure such that every latent variable is a root node with exactly 2 children.

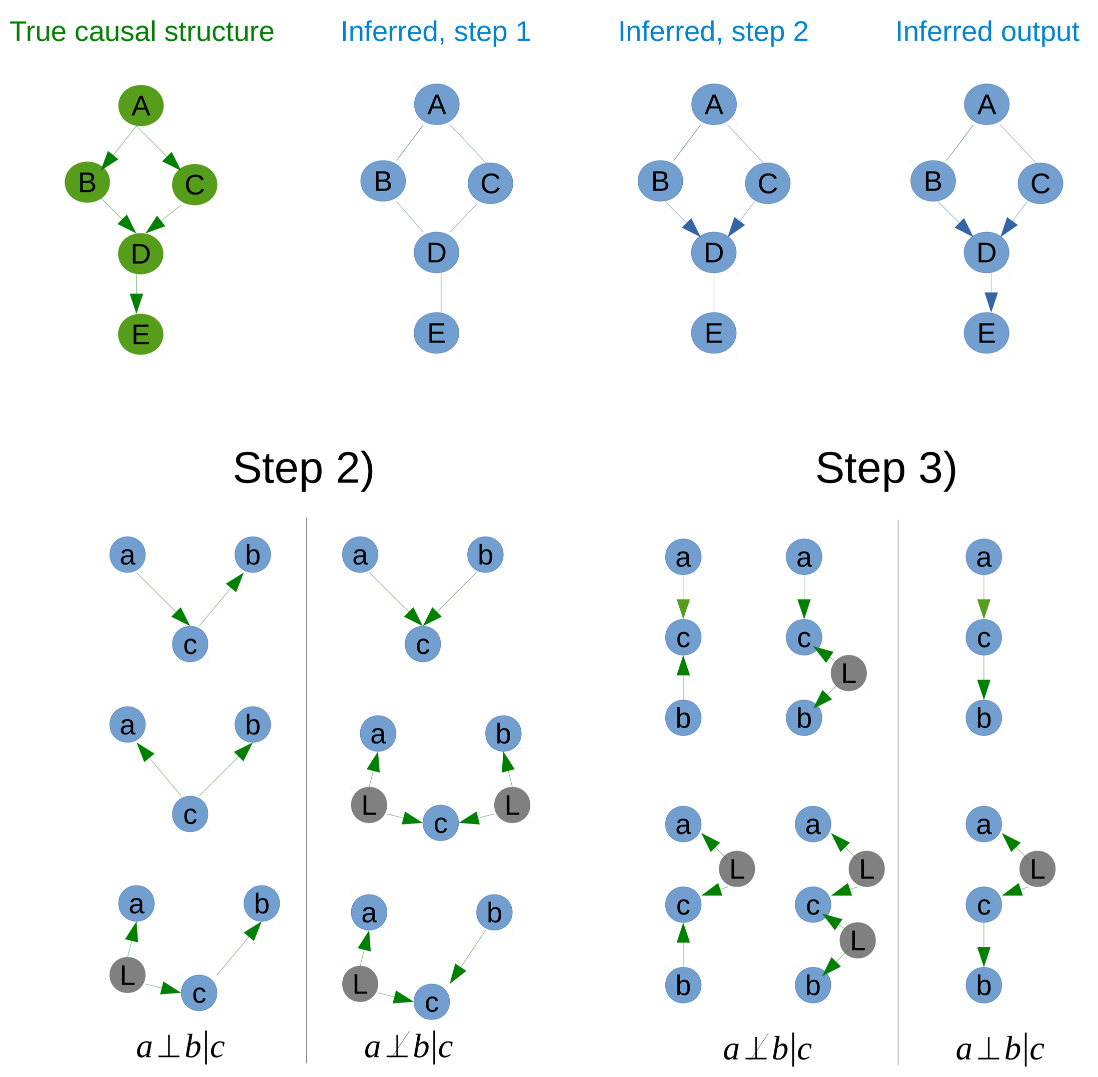


The IC* algorithm

Returns the set of causal networks that could have generated the observed set of conditional independences under the faithfulness assumption.

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. *Recursively add arrowheads/mark edges according to:*

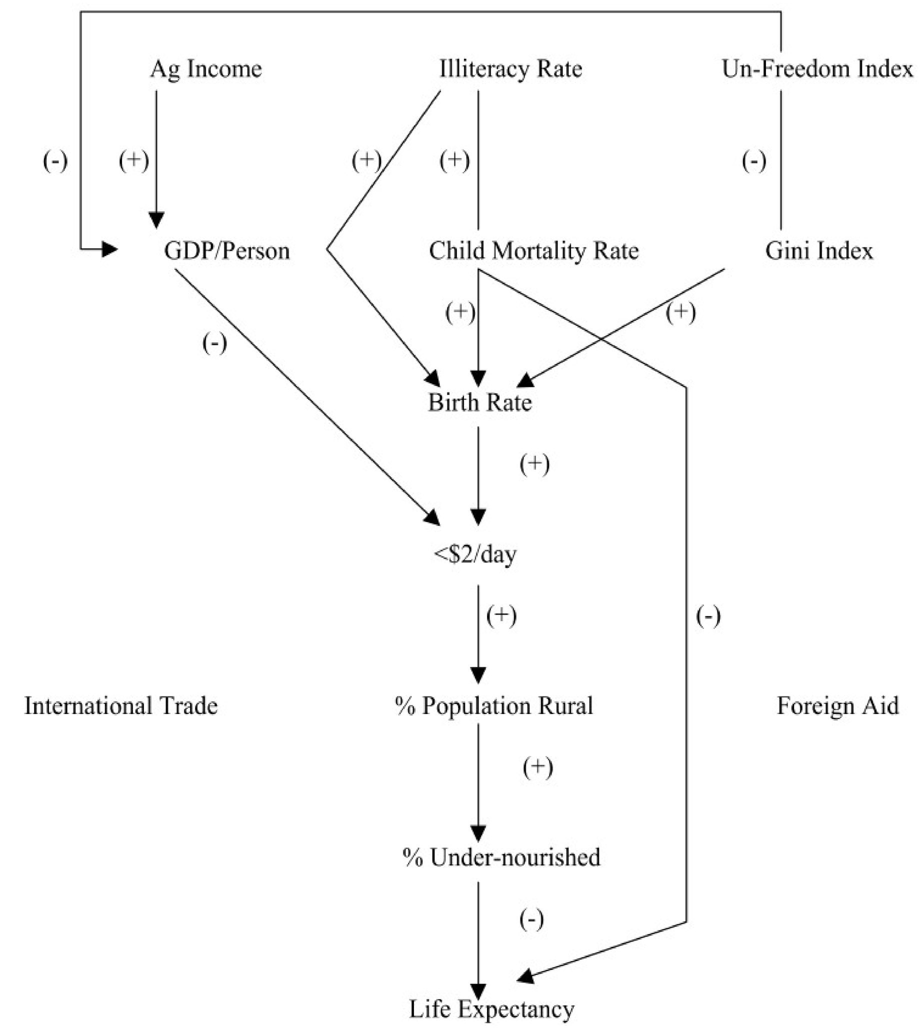




Applications

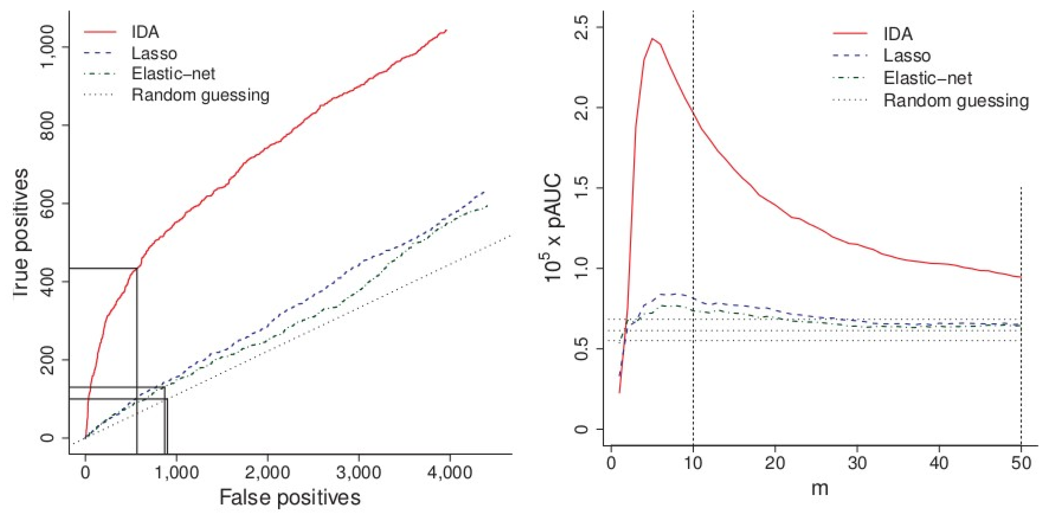
World poverty and its causes (Bessler 2003)

- Data on 13 variables on poverty, economic indicators, health, education and political freedom for 80 low income countries.



Effects of gene deletion on expression (Maathuis 2010)

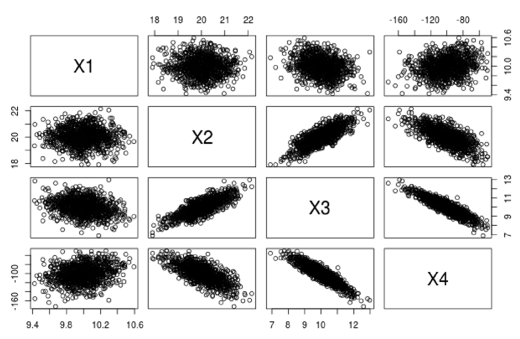
- Experimental data on the expression of 5361 genes in 234 strains with a single deleted gene is used to generate ground truth causal effects.
- Test how well these effects are recovered by the causal inference algorithm IDA, which combines the IC algorithm with the Do calculus. Substantial improvement is shown over the random and regularised feature selection benchmarks.

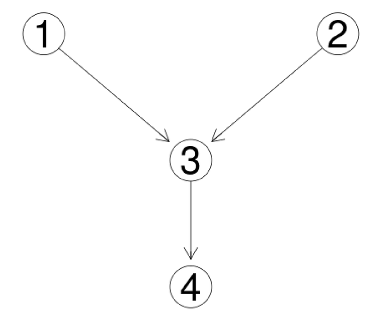


Pcalg, causal inference in R (Kalisch 2012)

- Correctly recovers the causal structure of a simple simulated data set in only 2 lines of code.

```
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 .