Causal Discovery

Richard Scheines

Peter Spirtes, Clark Glymour,
and many others

Dept. of Philosophy & Machine Learning
Carnegie Mellon

Outline

1. Motivation
2. Representation
3. Connecting Causation to Probability (Independence)
4. Searching for Causal Models

1. Motivation

Non-experimental Evidence

	Day Care	Aggressivenes
John	A lot	A lot
Mary	None	A little
⋮	⋮	⋮

Typical Predictive Questions

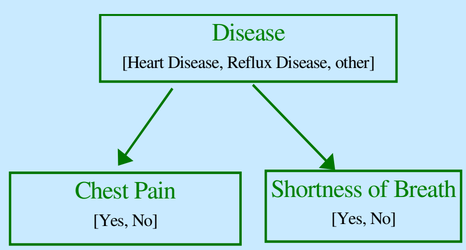
- Can we predict aggressiveness from the amount of violent TV watched
- Can we predict crime rates from abortion rates 20 years ago

Causal Questions:

- Does watching violent TV cause Aggression?
- I.e., if we change TV watching, will the level of Aggression change?

Bayes Networks

Qualitative Part:
Directed Graph



Quantitative Part:
Conditional Probability Tables

P(Disease = Heart Disease) = .2
P(Disease = Reflux Disease) = .5
P(Disease = other) = .3

P(Chest Pain = yes | D = Heart D.) = .7
P(Shortness of B = yes | D= Hear D.) = .8

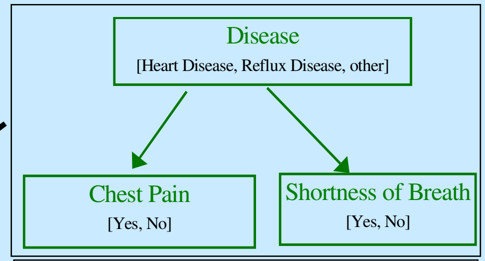
P(Chest Pain = yes | D = Reflux) = .9
P(Shortness of B = yes | D= Reflux) = .2

P(Chest Pain = yes | D = other) = .1
P(Shortness of B = yes | D= other) = .2

Bayes Networks: Updating

Given: **Data on Symptoms**

Chest Pain = yes



P(D = Heart Disease) = .2
P(D = Reflux Disease) = .5
P(D = other) = .3

P(Chest Pain = yes | D = Heart D.) = .7
~~**P(Shortness of B = yes | D= Hear D.) = .8**~~

P(Chest Pain = yes | D = Reflux) = .9
~~**P(Shortness of B = yes | D= Reflux) = .2**~~

P(Chest Pain = yes | D = other) = .1
~~**P(Shortness of B = yes | D= other) = .2**~~

Updating

Wanted:

P(Disease | Chest Pain = yes)

Causal Inference

Given: **Data on Symptoms**

Chest Pain = yes

Updating

P(Disease | Chest Pain = yes)

Causal Inference

P(Disease | Chest Pain set= yes)

Causal Inference

When and how can we use non-experimental data to tell us about the effect of an intervention?

Manipulated Probability P(Y | X set= x, **Z=z**)

from

Unmanipulated Probability P(Y | X = x, **Z=z**)

Conditioning ≠ Intervening

P(Y | X = x_1) vs. P(Y | X set= x_1)

Teeth Slides

2. Representation

1. Association & causal structure - qualitatively
2. Interventions
3. Statistical Causal Models
   1. Bayes Networks
   2. Structural Equation Models

Causation & Association

X and Y are associated ($X \not\!\perp\!\!\!\perp Y$) iff

$$\exists x_1 \neq x_2 \; P(Y \mid X = x_1) \neq P(Y \mid X = x_2)$$

Association is symmetric: $X \not\!\perp\!\!\!\perp Y \Leftrightarrow Y \not\!\perp\!\!\!\perp X$

X is a cause of Y iff

$$\exists x_1 \neq x_2 \; P(Y \mid X \text{ set}= x_1) \neq P(Y \mid X \text{ set}= x_2)$$

Causation is asymmetric: $X \rightarrow Y \not\Leftrightarrow X \leftarrow Y$

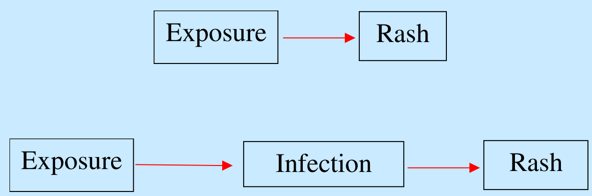
Causal Graphs

Causal Graph G = {**V**,**E**}

Each edge X → Y represents a direct causal claim:

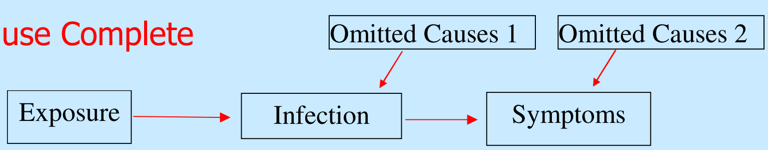
X is a direct cause of Y relative to **V**

Chicken Pox

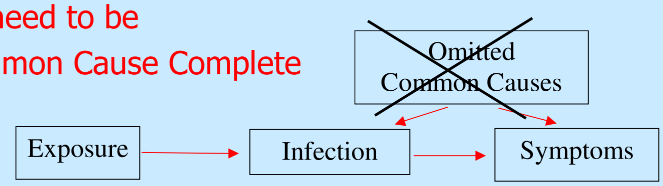


Causal Graphs

Do *Not* need to be
Cause Complete



Do need to be
Common Cause Complete



Modeling Ideal Interventions

Ideal Interventions (on a variable X):

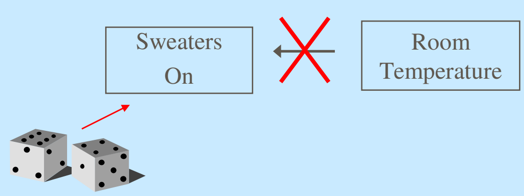
- **Completely *determine* the value or distribution of a variable X**

- **Directly Target only X**
  (no “fat hand”)
  E.g., Variables: Confidence, Athletic Performance
  Intervention 1: hypnosis for confidence
  Intervention 2: anti-anxiety drug (also muscle relaxer)

Modeling Ideal Interventions

Interventions on the Effect

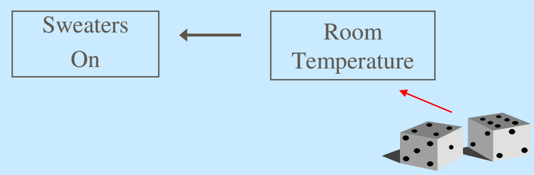
Post experimental System



Modeling Ideal Interventions

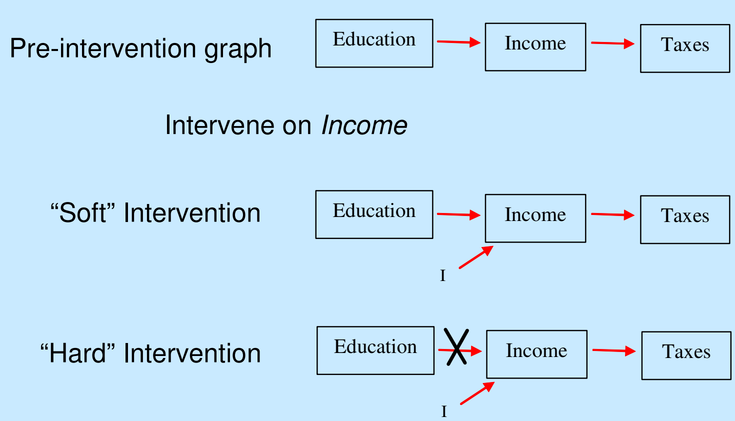
Interventions on the Cause

Post-experimental System

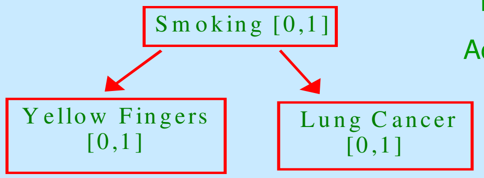


Interventions & Causal Graphs

Model an ideal intervention by adding an “intervention” variable outside the original system as a direct cause of its target.



Causal Bayes Networks



The Joint Distribution Factors

According to the Causal Graph,

i.e., for all X in **V**

P(**V**) = ΠP(X|Immediate Causes of(X))

P(S = 0) = .7

P(S = 1) = .3

P(S,YF, L) = P(S) P(YF | S) P(LC | S)

P(YF = 0 | S = 0) = .99

P(YF = 1 | S = 0) = .01

P(YF = 0 | S = 1) = .20

P(YF = 1 | S = 1) = .80

P(LC = 0 | S = 0) = .95

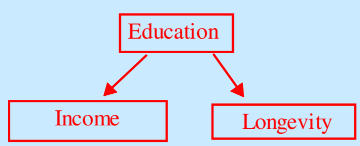
P(LC = 1 | S = 0) = .05

P(LC = 0 | S = 1) = .80

P(LC = 1 | S = 1) = .20

Structural Equation Models

Causal Graph



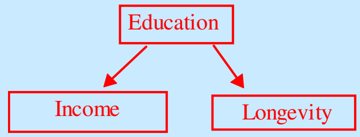
Statistical Model

1. Structural Equations

2. Statistical Constraints

Structural Equation Models

Causal Graph



- **Structural Equations:**

  One Equation for each variable V in the graph:

  $V = f(\text{parents}(V), \text{error}_V)$

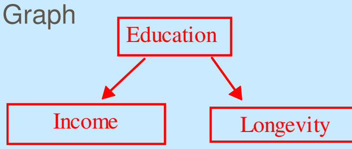
  for SEM (linear regression) f is a linear function

- **Statistical Constraints:**

  Joint Distribution over the Error terms

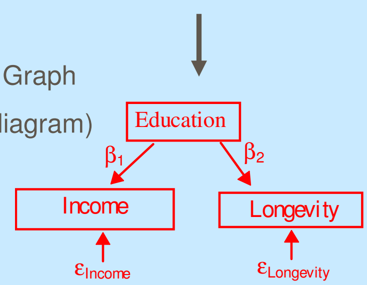
Structural Equation Models

Causal Graph



SEM Graph

(path diagram)



Equations:

$\text{Education} = \varepsilon_{ed}$

$\text{Income} = \beta_1 \text{ Education} + \varepsilon_{income}$

$\text{Longevity} = \beta_2 \text{ Education} + \varepsilon_{Longevity}$

Statistical Constraints:

$(\varepsilon_{ed}, \varepsilon_{Income}, \varepsilon_{Income}) \sim N(0, \Sigma^2)$

– Σ^2 diagonal

- no variance is zero

3. *Connecting*

Causation to Probability

Semantics of Causation

- Choice 1: Define X → Y, or X Y in terms of intervention, i.e., (hypothetical) treatment)
- Choice 2: Causal systems over V ⇒ Probabilistic Independence Relations in P(V)

Choice 1: Define Causation from Manipulation

X is a cause of Y iff

$\exists x_1 \neq x_2 \; P(Y \mid X \text{ set= } x_1) \neq P(Y \mid X \text{ set= } x_2)$

Causation is asymmetric: $X \rightarrow Y \;\; \not\Leftrightarrow \;\; X \leftarrow Y$

X and Y are associated ($X \not\perp\!\!\!\perp Y$) iff

$\exists x_1 \neq x_2 \; P(Y \mid X = x_1) \neq P(Y \mid X = x_2)$

Association is symmetric: $X \not\perp\!\!\!\perp Y \;\; \Leftrightarrow \;\; Y \not\perp\!\!\!\perp X$

Choice 1: Define *Direct* Causation from Intervention

X is a direct cause of Y relative to **S**, iff

$\exists \mathbf{z}, x_1 \neq x_2 \; P(Y \mid X \text{ set= } x_1, \mathbf{Z} \text{ set= } \mathbf{z})$

$\neq P(Y \mid X \text{ set= } x_2, \mathbf{Z} \text{ set= } \mathbf{z})$

where **Z** = **S** - {X,Y}

$X \longrightarrow Y$

Choice 2: Causal Markov Axiom

If G is a causal graph, and P a probability distribution over the variables in G, then in P:

every variable V is independent of its non-effects, conditional on its immediate causes.

Causal Markov Condition

Two Intuitions:

1) Immediate causes make effects independent of remote causes (Markov).

2) Common causes make their effects independent (Reichenbach).

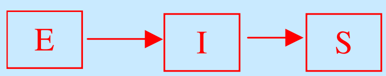
Causal Markov Condition

1) Immediate causes make effects independent of remote causes (Markov).

E = Exposure to Chicken Pox

I = Infected

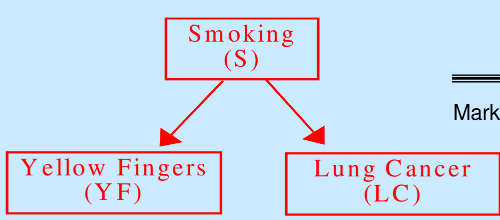
S = Symptoms



Markov Cond. ⟹ $E \perp\!\!\!\perp S \mid I$

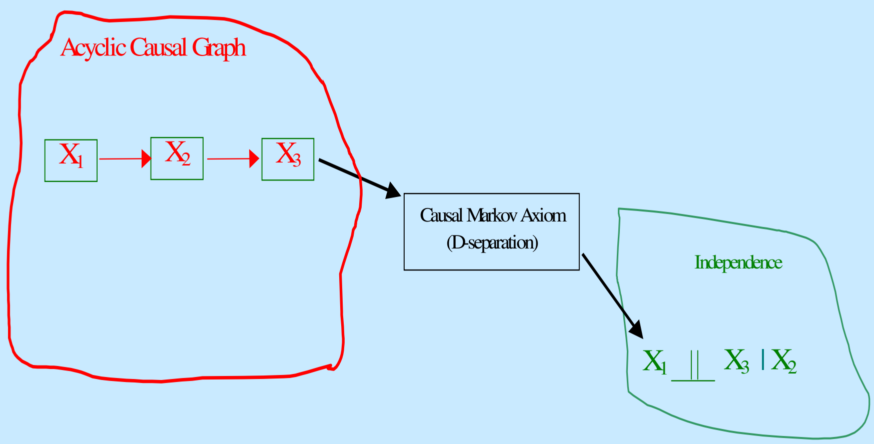
Causal Markov Condition

2) Effects are independent conditional on their common causes.



Markov Cond. ⟹ $YF \perp\!\!\!\perp LC \mid S$

Causal Structure ⇒ Statistical Data



Causal Markov Axiom

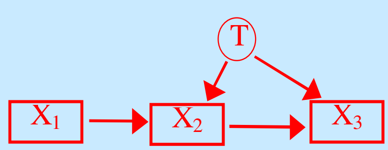
In SEMs, d-separation follows from assuming independence among error terms that have no connection in the path diagram -

i.e., assuming that the model is common cause complete.

Causal Markov and D-Separation

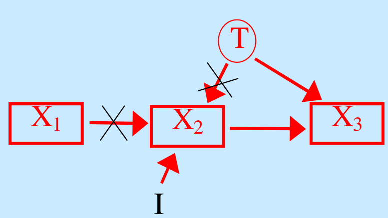
- In acyclic graphs: equivalent
- Cyclic Linear SEMs with uncorrelated errors:
  - D-separation correct
  - Markov condition incorrect
- Cyclic Discrete Variable Bayes Nets:
  - If equilibrium --> d-separation correct
  - Markov incorrect

D-separation: Conditioning vs. Intervening



$P(X_3 \mid X_2) \neq P(X_3 \mid X_2, X_1)$

$X_3 \not\perp\!\!\!\perp X_1 \mid X_2$



$P(X_3 \mid X_2 \text{ set= }) = P(X_3 \mid X_2 \text{ set=}, X_1)$

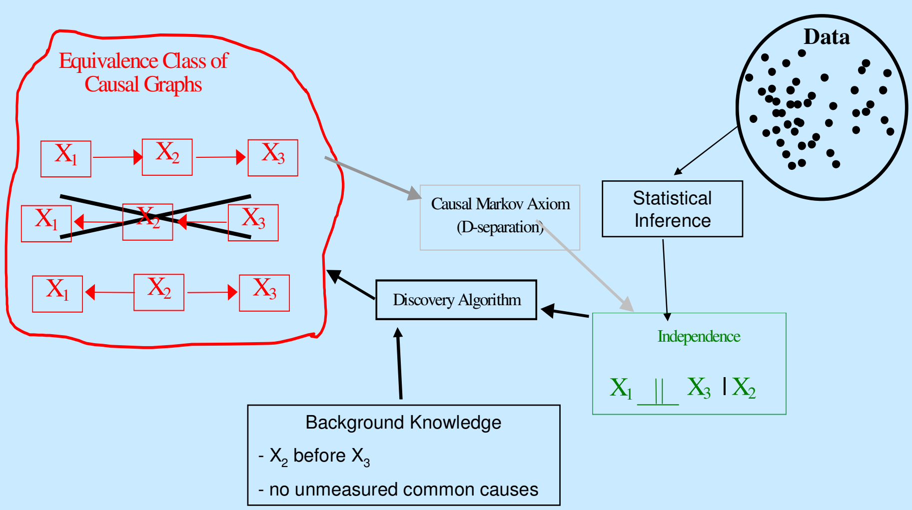
$X_3 \perp\!\!\!\perp X_1 \mid X_2 \text{ set=}$

4. *Search*

From Statistical Data
to Probability
to Causation

Causal Discovery

Statistical Data ⇒ Causal Structure



Representations of D-separation Equivalence Classes

We want the representations to:

- Characterize the Independence Relations Entailed by the Equivalence Class

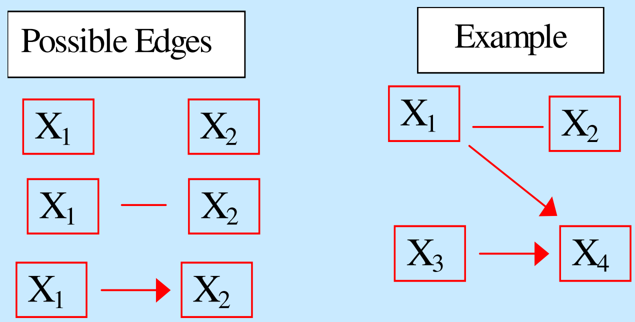
- Represent causal features that are shared by every member of the equivalence class

Patterns & PAGs

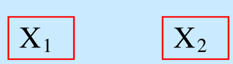
- Patterns (Verma and Pearl, 1990): graphical representation of an acyclic d-separation equivalence - no latent variables.

- PAGs: (Richardson 1994) graphical representation of an equivalence class including *latent variable models* and *sample selection bias* that are d-separation equivalent over a set of measured variables **X**

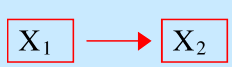
Patterns



Patterns: What the Edges Mean



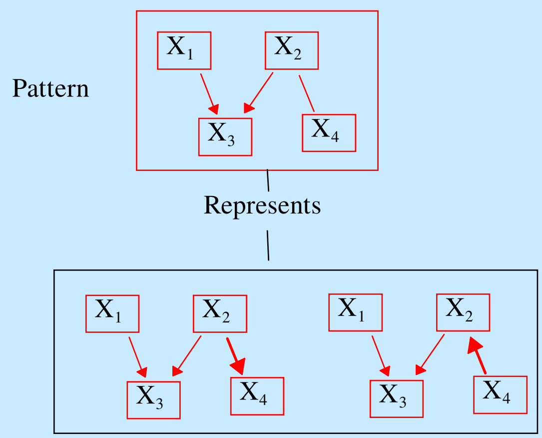
X_1 and X_2 are not adjacent in any member of the equivalence class



$X_1 \rightarrow X_2$ (X_1 is a cause of X_2) in every member of the equivalence class.

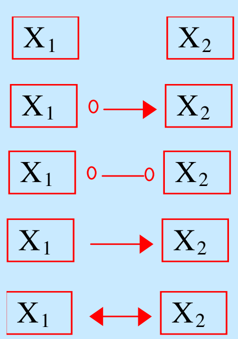
$X_1 \rightarrow X_2$ in some members of the equivalence class, and $X_2 \rightarrow X_1$ in others.

Patterns



PAGs: Partial Ancestral Graphs

What PAG edges mean.



X_1 and X_2 are not adjacent

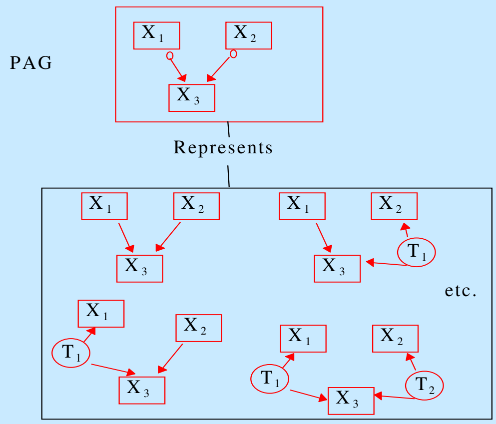
X_2 is not an ancestor of X_1

No set d-separates X_2 and X_1

X_1 is a cause of X_2

There is a latent common cause of X_1 and X_2

PAGs: Partial Ancestral Graph



Overview of Search Methods

- Constraint Based Searches
  - TETRAD
- Scoring Searches
  - Scores: BIC, AIC, etc.
  - Search: Hill Climb, Genetic Alg., Simulated Annealing
  - Very difficult to extend to latent variable models

Heckerman, Meek and Cooper (1999). "A Bayesian Approach to Causal Discovery" chp. 4 in Computation, Causation, and Discovery, ed. by Glymour and Cooper, MIT Press, pp. 141-166

Tetrad 4 Demo

www.phil.cmu.edu/projects/tetrad_download/

5. Regession and Causal Inference

Regression to estimate Causal Influence

- Let **V** = {**X**,Y,**T**}, where
  - Y : measured outcome
  - measured regressors: **X** = {X_1, X_2, …, X_n}
  - latent common causes of pairs in **X** U Y: **T** = {T_1, …, T_k}

- Let the true causal model over **V** be a Structural Equation Model in which each V $\in$ **V** is a linear combination of its direct causes and independent, Gaussian noise.

Regression and Causal Inference

- Consider the regression equation:

  $$Y = b_0 + b_1X_1 + b_2X_2 + \ldots b_nX_n$$

- Let the OLS regression estimate b_i be the *estimated* causal influence of X_i on Y.

- That is, holding **X**/Xi experimentally constant, b_i is an estimate of the change in E(Y) that results from an intervention that changes Xi by 1 unit.

- Let the *real* Causal Influence $X_i \rightarrow Y = \beta_i$

- When is the OLS estimate b_i an unbiased estimate of β_i ?

Linear Regression

Let the other regressors **O** = $\{X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$

$b_i = 0$ if and only if $\rho_{Xi,Y.\mathbf{O}} = 0$

In a multivariate normal distribuion,

$\rho_{Xi,Y.\mathbf{O}} = 0$ if and only if X_i _||_ Y | O

Linear Regression

So in regression:

$b_i = 0 \Leftrightarrow X_i$ _||_ Y | O

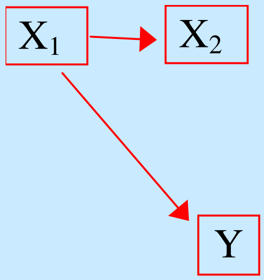
But provably :

$\beta_i = 0 \Leftarrow \exists S \subseteq O, X_i$ _||_ Y | S

So $\exists S \subseteq O, X_i$ _||_ Y | $S \Rightarrow \beta_i = 0$

$\sim \exists S \subseteq O, X_i$ _||_ Y | $S \Rightarrow$ don't know (unless we're lucky)

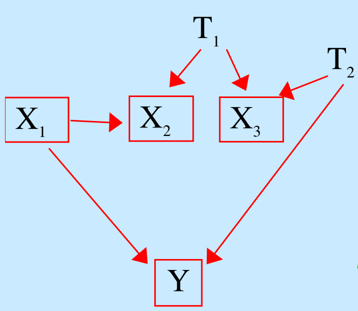
Regression Example



True Model

$X_1 \not\perp\!\!\!\perp Y \mid X_2$	$b_1 \neq 0$
$X_2 \perp\!\!\!\perp Y \mid X_1$	$b_2 = 0$
$\sim\exists S \subseteq \{X_2\}\quad X_1 \perp\!\!\!\perp Y \mid S$	Don't know
$\exists S \subseteq \{X_1\}\quad X_2 \perp\!\!\!\perp Y \mid \{X_1\}$	$\beta_2 = 0$

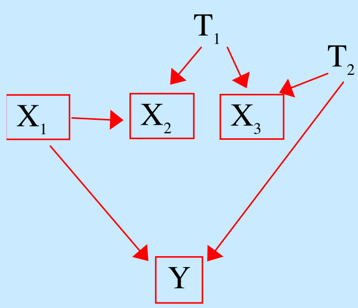
Regression Example



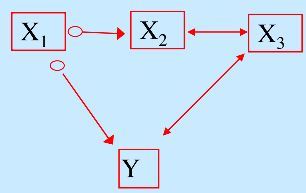
True Model

$X_1 \not\perp\!\!\!\perp Y \mid \{X_2, X_3\}$	$b_1 \neq 0$
$X_2 \not\perp\!\!\!\perp Y \mid \{X_1, X_3\}$	$b_2 \neq 0$
$X_3 \not\perp\!\!\!\perp Y \mid \{X_1, X_2\}$	$b_3 \neq 0$
$\sim\exists S \subseteq \{X_2, X_3\},\ X_1 \perp\!\!\!\perp Y \mid S$	DK
$\exists S \subseteq \{X_1, X_3\},\ X_2 \perp\!\!\!\perp Y \mid \{X_1\}$	$\beta_2 = 0$
$\sim\exists S \subseteq \{X_1, X_2\},\ X_3 \perp\!\!\!\perp Y \mid S$	DK

Regression Example



True Model



PAG

Regression Bias

If

- X_i is d-separated from Y conditional on **X**/X_i in the true graph after removing $X_i \rightarrow Y$, and
- **X** contains no descendant of Y, then:

b_i is an unbiased estimate of β_i

See Using Path Diagrams as a Structural Equation Modeling Tool, (1998). Spirtes, P., Richardson, T., Meek, C., Scheines, R., and Glymour, C., Sociological Methods & Research, Vol. 27, N. 2, 182-225

Ongoing Projects

- Finding Latent Variable Models (Ricardo Silva, Gatsby Neuroscience, former CALD PhD)
- Strong Faithfulness (Jiji Zhang, Philosophy)
- Educational Data Mining (Benjamin Shih, CALD)
- Sequential Experimentation (Active Discovery), (Frederick Eberhardt, CALD & Philosophy)
- Measurement Error ???
- Mixed Data (Discrete & Continuous Variables) ?????

References

- *Causation, Prediction, and Search*, 2nd Edition, (2000), by P. Spirtes, C. Glymour, and R. Scheines (MIT Press)
- *Causality: Models, Reasoning, and Inference*, (2000), Judea Pearl, Cambridge Univ. Press
- *Computation, Causation, & Discovery* (1999), edited by C. Glymour and G. Cooper, MIT Press
- *TETRAD IV:* www.phil.cmu.edu/projects/tetrad
- Causality Lab: www.phil.cmu.edu/projects/causality-lab
- Web Course: www.phil.cmu.edu/projects/csr/