Causal Inference: prediction, explanation, and intervention

Lecture 5: Causality and Graphical Models

Samantha Kleinberg
samantha.kleinberg@stevens.edu

Next few weeks

- October 11 (Tuesday class!): Time series
- October 17: Lecture + midterm review/Q&A
- October 24: Midterm exam
- October 31: Project proposal due

- November 29 and December 6: final presentations

Final Project

- Project types (not exhaustive!)
  - Analyze data, adapt causal inference method to particular domain
  - Compare causal inference methods
  - Theoretical work on causality (methods or meaning)
- Proposal
  - What's the goal? How will you accomplish it? How will you know the project was successful? What resources are needed (and do you have them)? **1 page**

Today

- What makes a graphical model causal?
- How to use graphical models to answer causal questions
  - Predicting effects of actions
  - Counterfactual queries
  - Explanation

Application: infant mortality

Table 1: The output of BLCD

Cause	Effect
1. Heart malformations	Infant outcome
2. Hydrocephalus	Infant outcome
3. Weight gain during preg-nancy	Infant circulatory/respira-tory anomalies
4. Microcephalus	Ultrasound
5. Diaphragmatic hernia	Plurality of birth
6. Five minute Apgar score	Infant heart malformations

The probability distributions associated with relationships 1 and 2 in Table 1 are as follows:

$$P(\text{infant alive at one year} \mid \text{heart malformations}) = 0.797$$

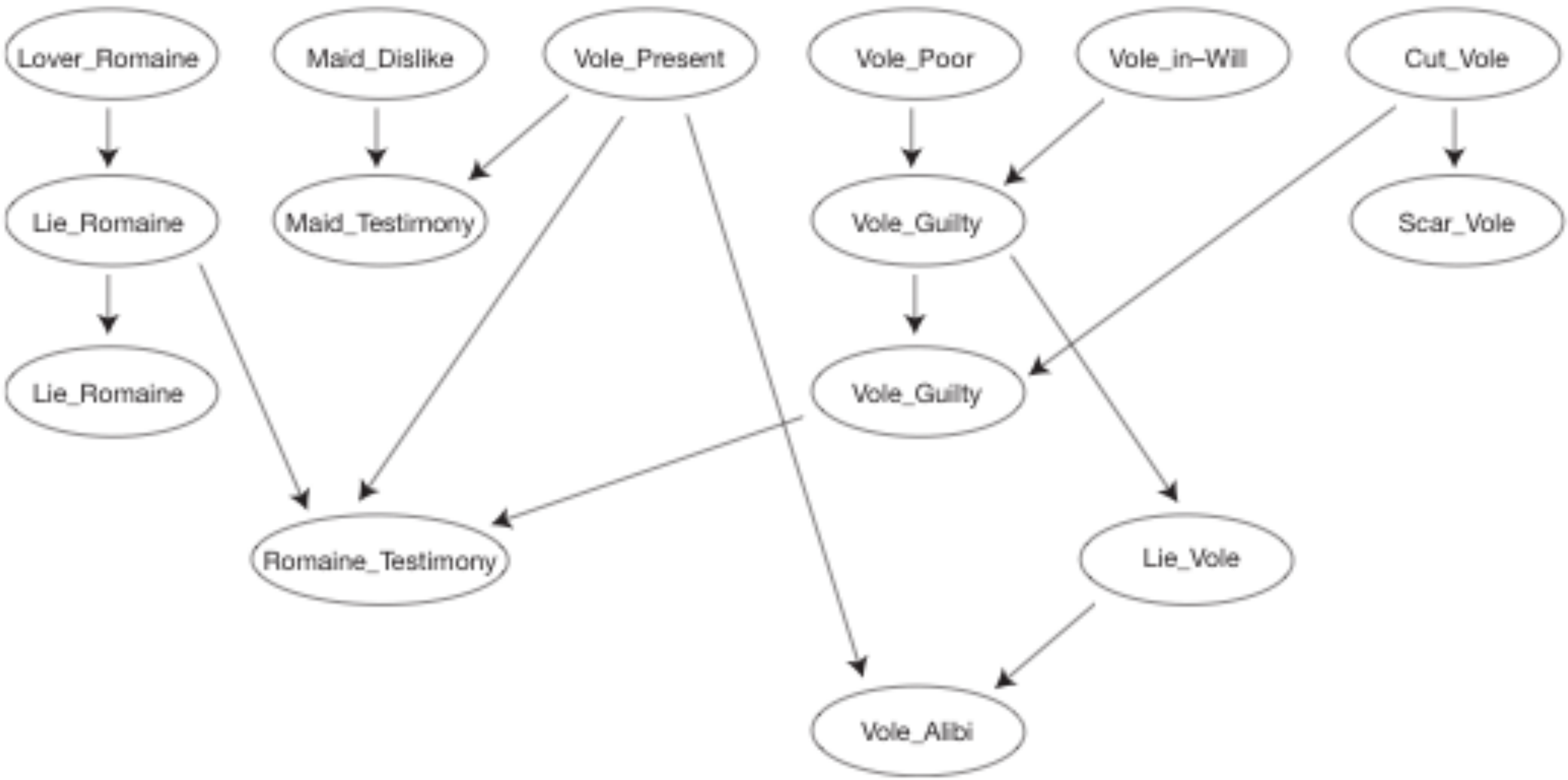
$$P(\text{infant alive at one year} \mid \text{no heart malformations}) = 0.992$$

$$P(\text{infant alive at one year} \mid \text{hydrocephalus}) = 0.5$$

$$P(\text{infant alive at one year} \mid \text{no hydrocephalus}) = 0.992$$

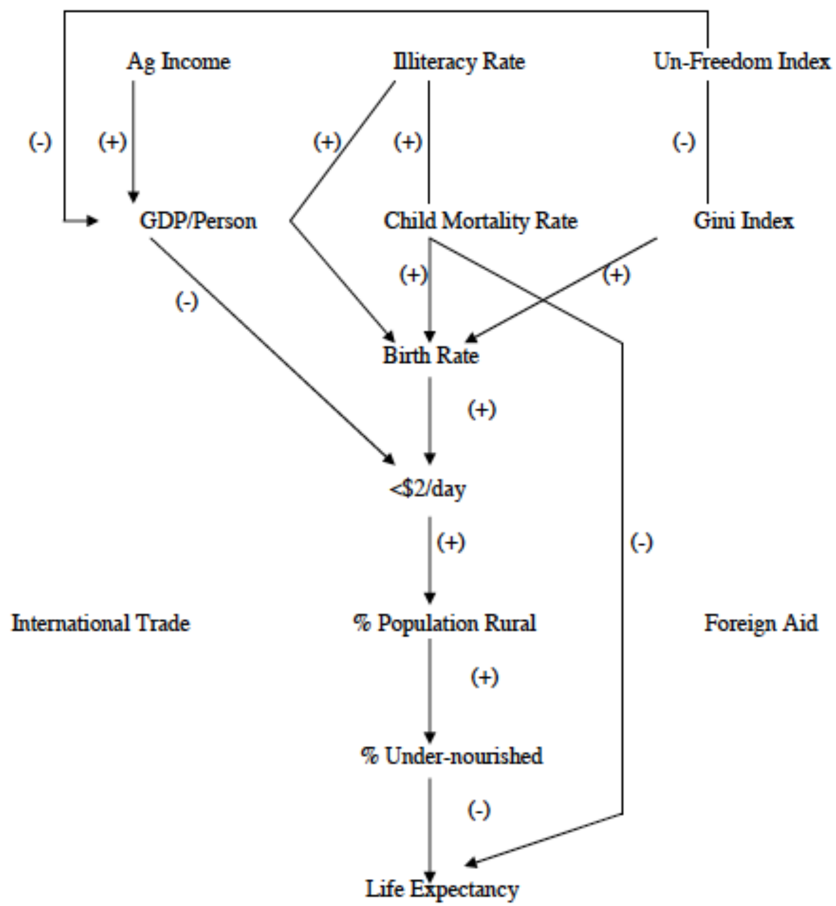
Mani, S., & Cooper, G. F. (2004). Causal Discovery Using a Bayesian Local Causal Discovery Algorithm. *Proceedings of MedInfo*, 731-735

Application: Organizing evidence



Lagnado, D. (2011). Thinking about Evidence. In *Proceedings of the British Academy* (Vol. 171, pp. 183-223)

Application: Causes and effects of poverty



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

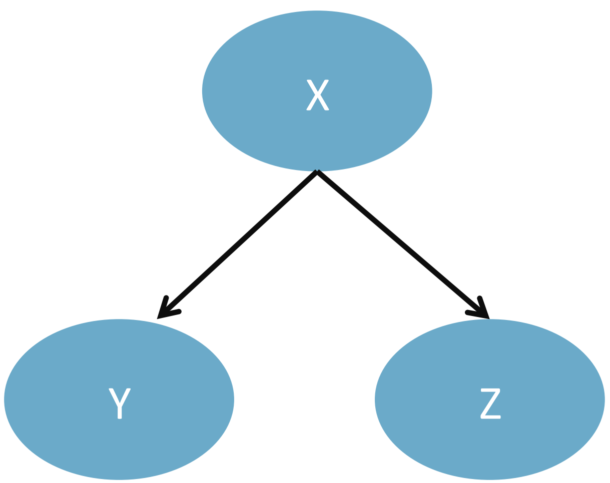
Applications of causal BNs

- Biomedical informatics – diagnosis, prognosis
- Psychology – representation of causal learning
- Economics/finance

Markov condition

(from last week)

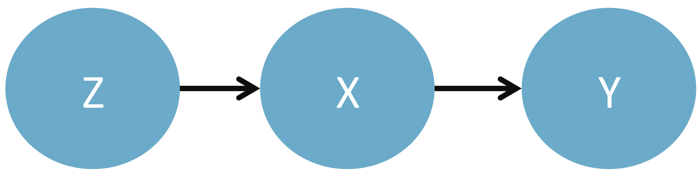
Node independent of non-descendants given its parents



$$Y \perp Z \mid X$$

d-separation

DAG G



$\longrightarrow$

d-separation

Set of independencies

$$Y \perp Z \mid X$$

d-separation

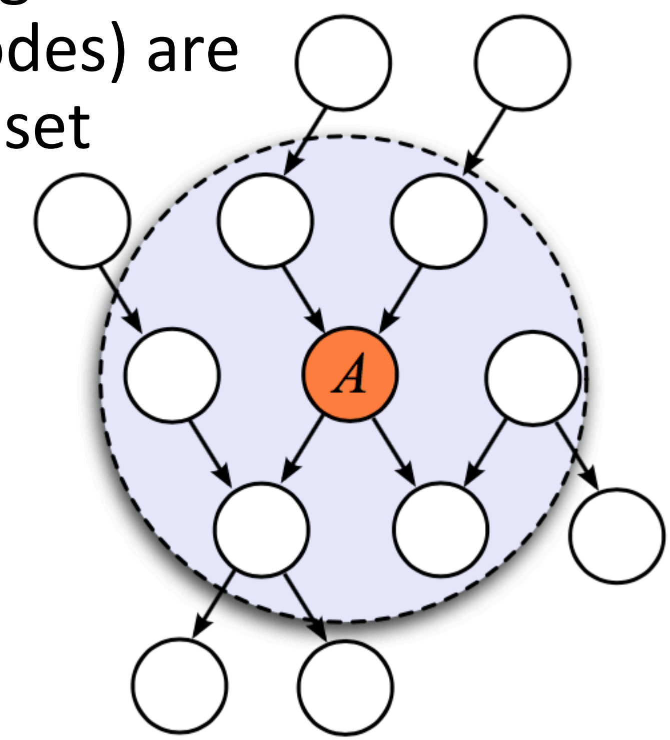
Equivalent statements, for sets of nodes X, Y, Z in graph G:

- X and Y are d-separated by Z (Z can be node or set of nodes) in G
- X and Y are conditionally independent given Z
- Z blocks all paths between X and Y

d-separation and Markov blanket

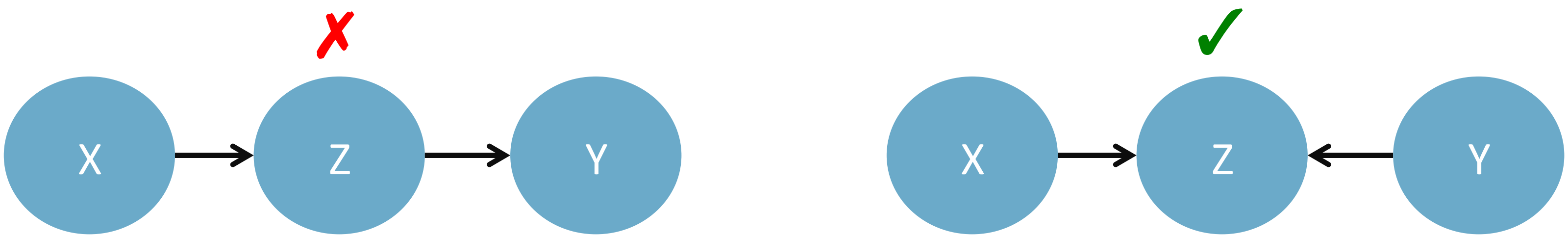
Markov blanket: set of nodes that separate a node from all others

d-separation: Method for determining whether a pair of nodes (or sets of nodes) are independent conditioned on another set



Definition: d-separation

- Node v is a **collider** if two arrowheads meet at v



- X and Y are **d-connected** by Z in graph G iff
  - Exists an undirected path between a vertex in X and vertex in Y s.t. for every collider C on the path, C or descendant of C is in Z and no non-collider on path is in Z
- X and Y are **d-separated** by Z in G iff they are not d-connected by Z in G

Example 1

- $X \rightarrow Y \rightarrow Z$

- $X \leftarrow Y \rightarrow Z$

In both cases, X,Z d-separated by Y: no colliders on path from X to Z, and X and Z not d-connected by Y

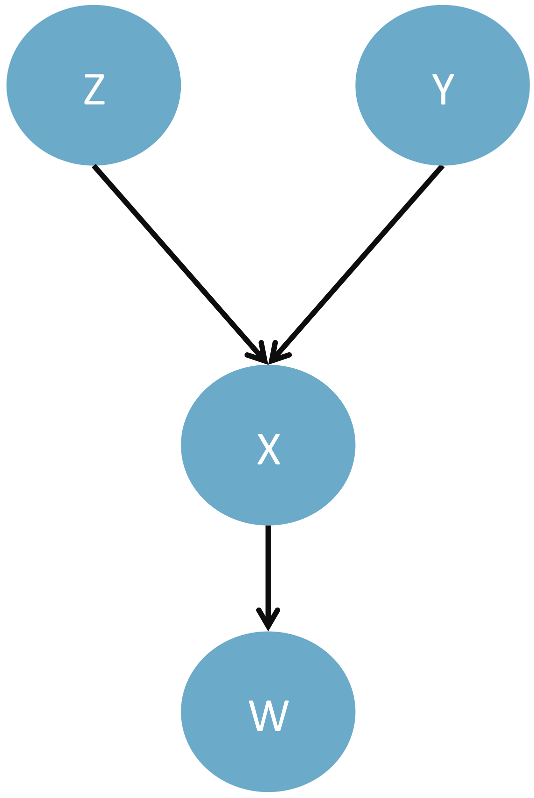
Example 2

- $X \rightarrow Y \leftarrow Z$

- X,Z d-separated by a set of nodes only if Y NOT in that set. X,Z d-connected by Y

Example 3

- Are Y,Z d-separated by W?
- No, d-connected by X and W is descendent

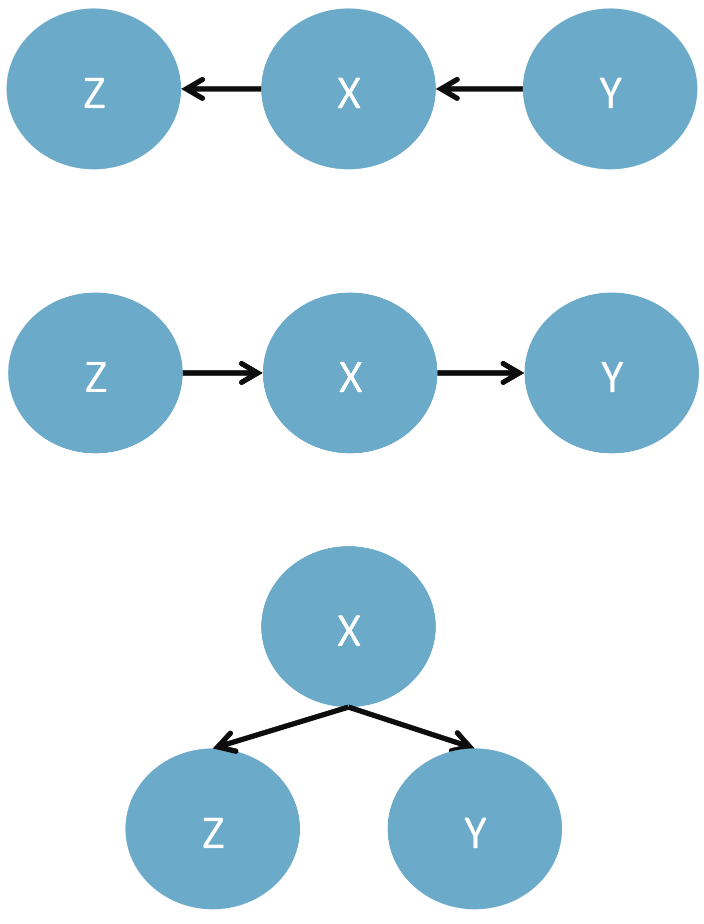


But…

Indep -> many networks

Network -> 1 set indep

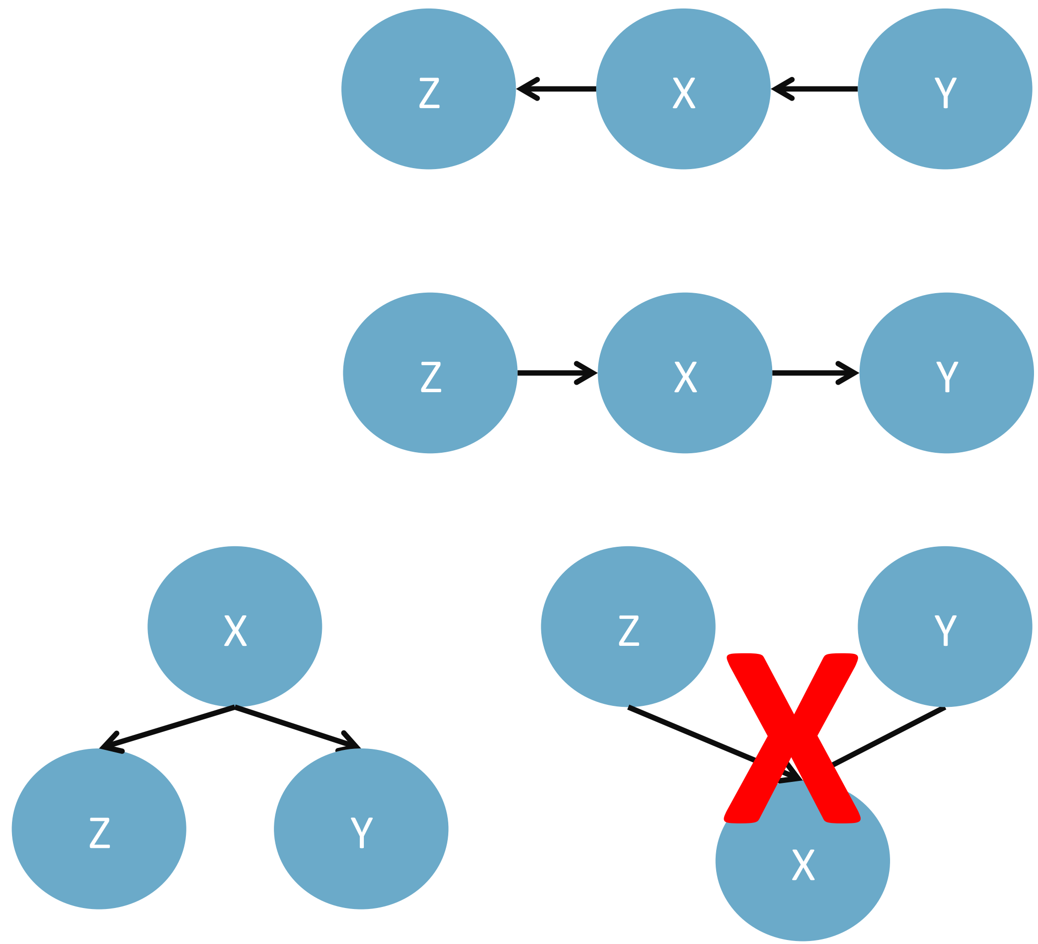
$$Y \perp Z \mid X$$



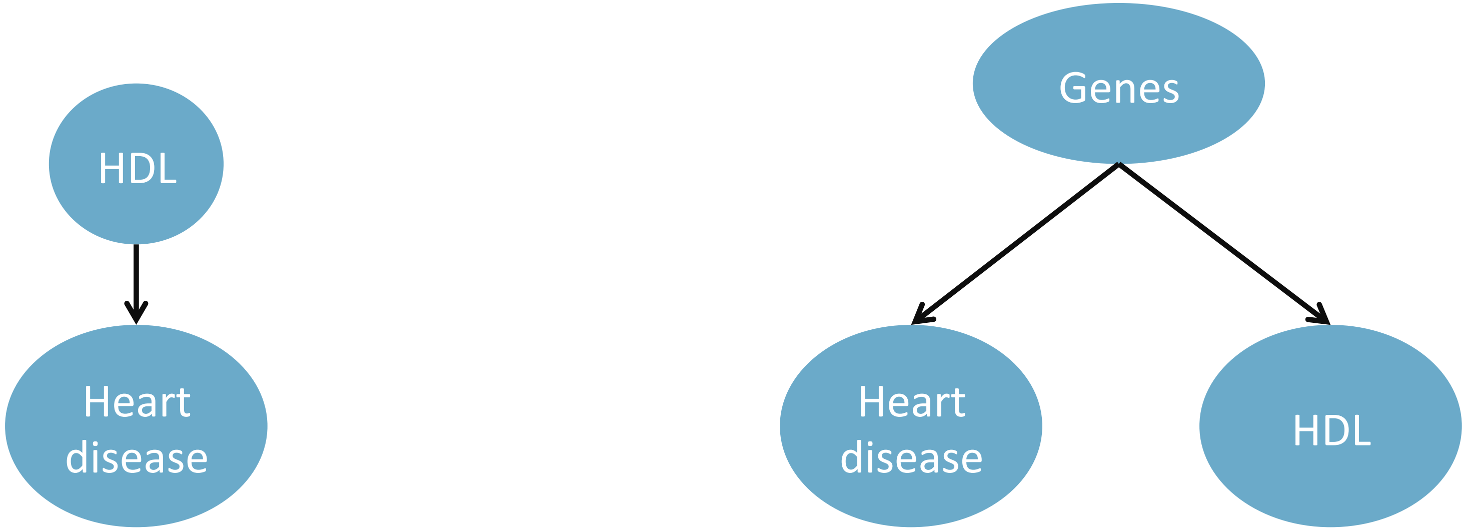
But…

Can rule out some

$$Y \perp Z \mid X$$



…And



…Also

Coin 1	Coin 2	# obs.
H	H	5
T	T	3
H	T	1
T	H	1

$$P(C_1 = H \wedge C_2 = H) > P(C_1 = H)P(C_2 = H)$$

$$5/10 > 6/10 * 6/10$$

$$C_1 \not\perp C_2$$

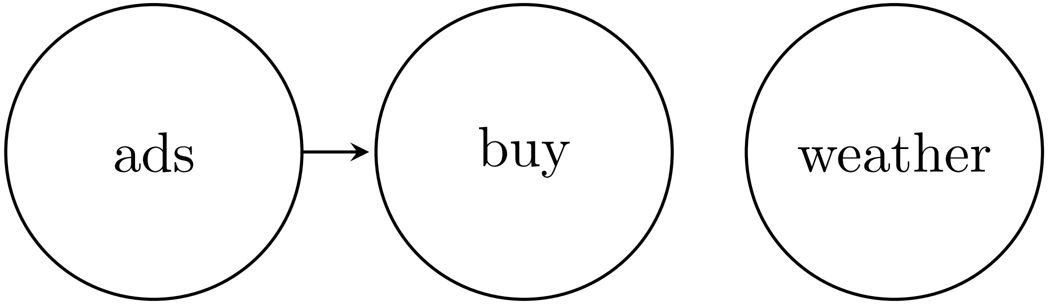
Causal interpretation

- Causal Markov condition
- Faithfulness
- Causal sufficiency

+ a few others, e.g. variables "correctly" specified

Causal graph

- Arrows denote **direct** causes
  - Edge from X to Y means X causes Y
- DAG



Causal Markov condition (CMC)

Node in the graph is independent of all of its non-descendants (direct and indirect effects) given its direct causes

CMC and screening off

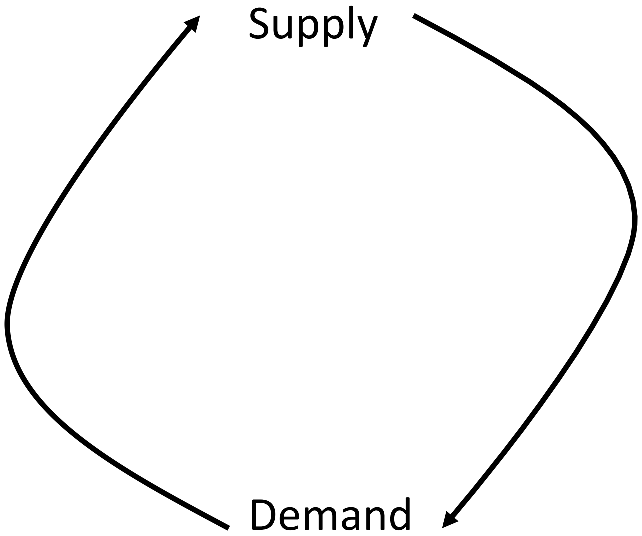
Recall Common Cause Principle (CCP)

- If $P(X \wedge Y) > P(X)P(Y)$ then either X causes Y (or vice versa) or they have a common cause

Now: if $P(X \wedge Y) > P(X)P(Y)$ and they have a common cause C, it means X ind Y | C

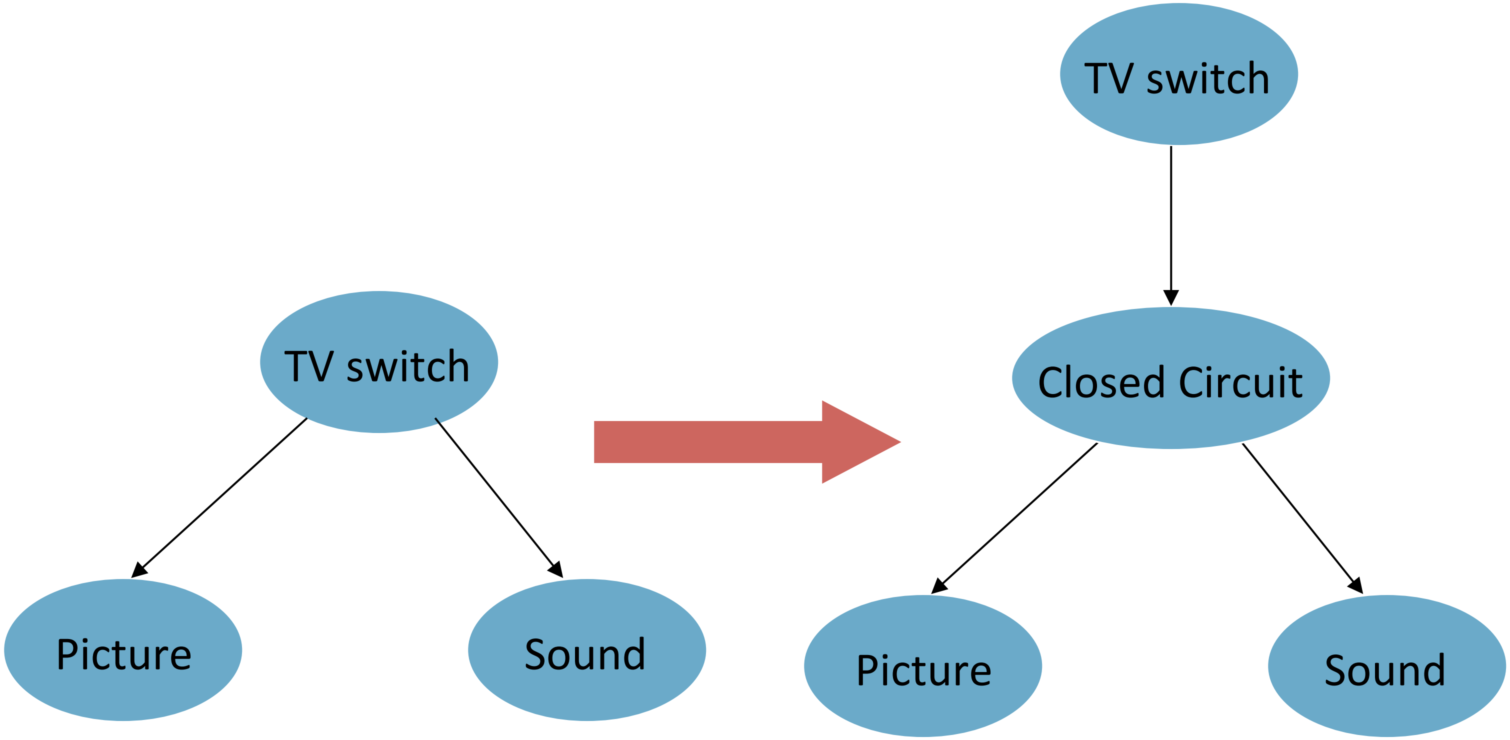
Note that CCP seeks single common cause. CMC allows for sets of nodes.

Problems: feedback



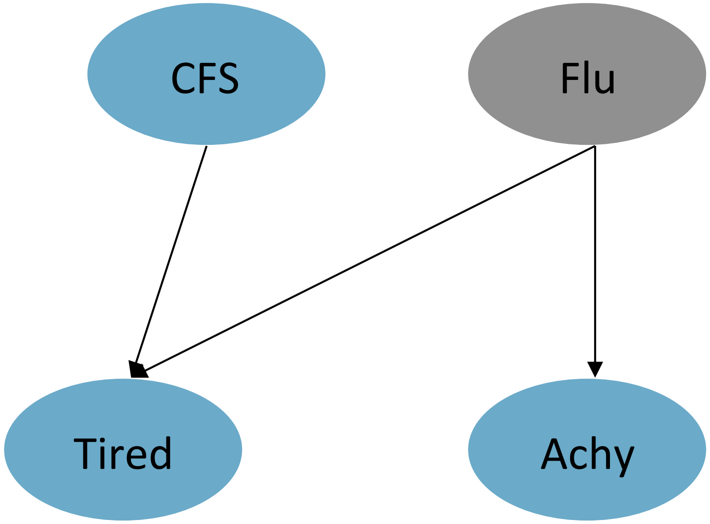
Problems: Indeterminism

P(picture|switch) < P(picture|switch,sound)



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

Problems: hidden common causes



Completeness of graph

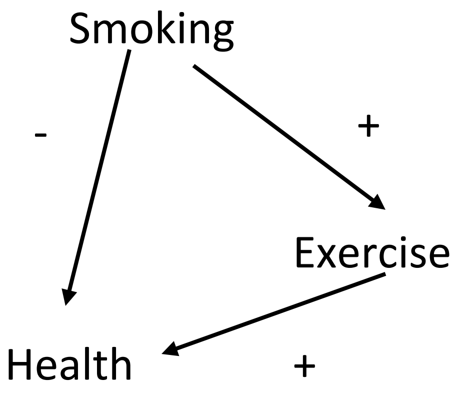
- Complete: all common causes included, all causal relations among variables included
- Incomplete: not all intermediate factors necessarily included

Faithfulness

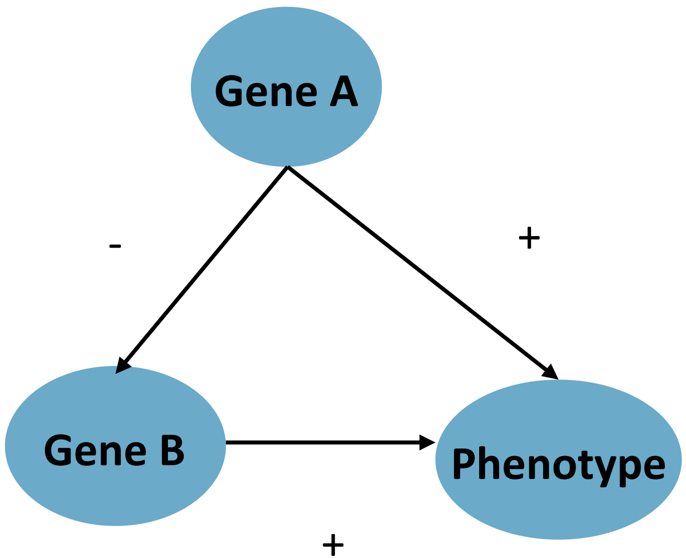
Exactly the dependencies in the underlying structure hold in the data

- i.e. Independence relations not from chance but from structure
- No canceling out

Example

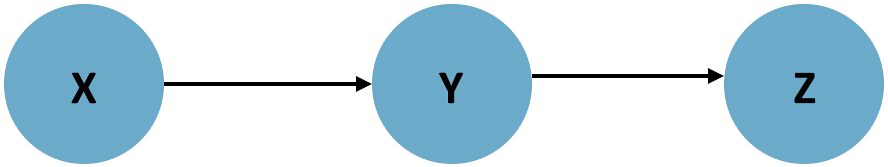


Another example

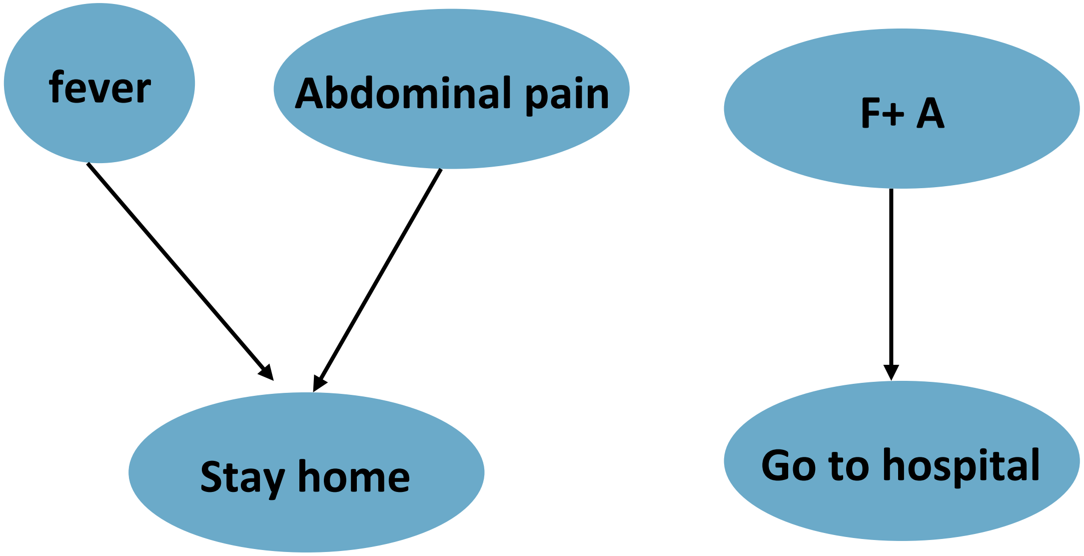


A final example (deterministic chain)

$$X \perp Z \mid Y$$



Selection bias



Cooper, G. F. (1999). An overview of the representation and discovery of causal relationships using bayesian networks. In C. Glymour & G. F. Cooper (Eds.), *Computation, causation, and discovery.* AAAI Press and MIT Press

How big of a problem is this?

Effect of Study Criteria on Recruitment and Generalizability of the Results

Khan, Ahsan Y. MD*; Preskorn, Sheldon H. MD†; Baker, Bryan MS, CCRC‡

Abstract

Objective: Clinical trials are indispensable to drug approval process. This research examined the effect of a specific study criteria on recruitment and generalizability of the results.

Methods: The following were reviewed: (a) the usual inclusion and exclusion criteria for the antipsychotic trials performed at the Institute; (b) epidemiologic data, to determine the effect of study criteria on the target population; and (c) the recruitment procedures/strategies used to identify potential candidates. A survey was conducted to determine the percentage of schizophrenic patients in a conventional outpatient psychiatric clinic conforming to the usual enrollment criteria for antipsychotic trials.

Results: Intensive recruitment efforts in a general population of 3.6 million would have been expected to yield only 632 eligible subjects out of 36,000 suffering from schizophrenia. Out of 632, only 50 contacted the research site after an intensive recruitment effort. From those 50, 30 were excluded during a telephone interview. Of the 20 remaining, 6 were excluded for a variety of reasons during a face-to-face interview. Thus, only 14 subjects out of a population of 3.6 million met the study criteria.

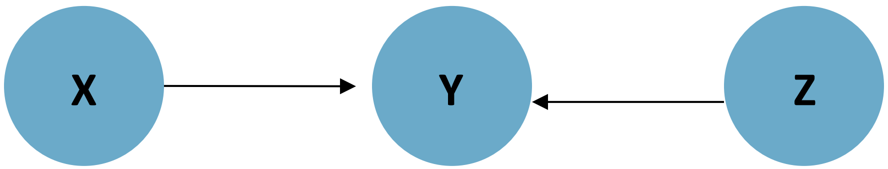
Conclusions: These results emphasize the rarified nature of patients-volunteers who enter a clinical trial. Inclusion and exclusion study criteria can severely restrict the number of eligible subjects, dictate recruitment strategies, and in turn affect generalizability of the results.

Recap of problems for faithfulness

- Only true in large sample limit
- Simpson's paradox
- Selection bias
- Statistical tests

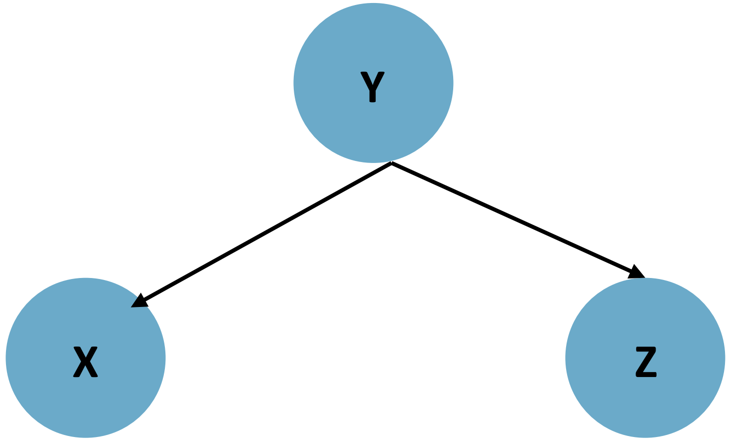
Quick recap

- CMC: population produced by structure has these independencies
- Faithfulness: population has *only* these independencies Why do we need both?



Causal sufficiency

- All common causes of pairs of variables measured
- Not sufficient if Y not measured

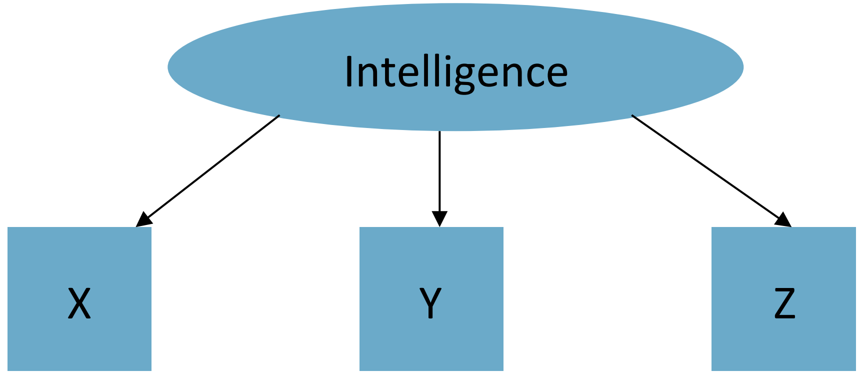


Completeness vs. sufficiency

Completeness: common causes are included in causal graph

Sufficiency: all common causes have been measured

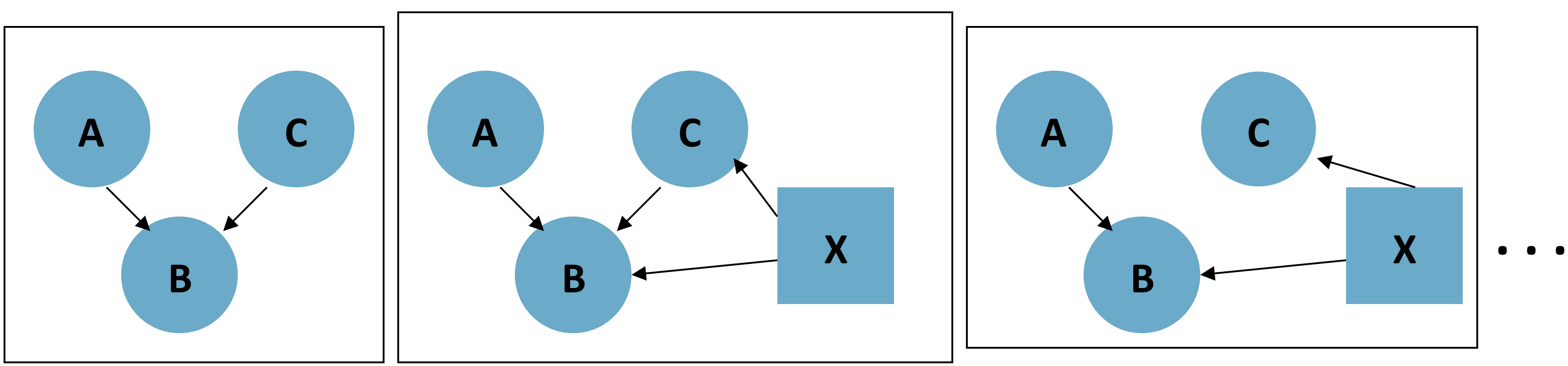
Example



Scheines, R. (1997). An introduction to causal inference. *Causality in Crisis*, 185-99

What if no causal sufficiency?

- Need to include all graphs with unmeasured common causes.
- Ex: measured A, B, C. Found $A \perp\!\!\!\perp C$. With CMC, faithfulness but no causal sufficiency, the following graphs are all possible (X is unmeasured):



In absence of sufficiency…

- Can still learn something
  - Some relationships may appear in all graphs
  - Can find set of all graphs representing independence relations, with nodes for possible hidden variables
- Timing information helps

Things to beware of with inference

- Sample size
- Missing data (not just variables)
- Multiple testing (and FDR)
- What structures DAG can/cannot represent (e.g. time series and feedback)
- Variable representation

The good news

- Can add time
- Can experiment
- Methods for testing assumptions

Recap of causal inference with BN

What makes a Bayesian network causal?

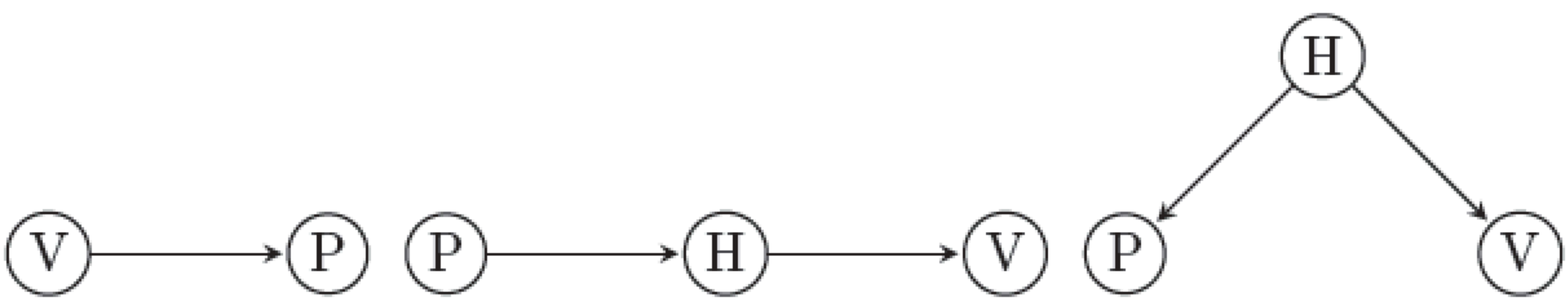
The assumptions: CMC, sufficiency, faithfulness

Assumptions+Data→Independencies → Causal BN(s) → effects of interventions

Learning BN

Same methods we discussed last week

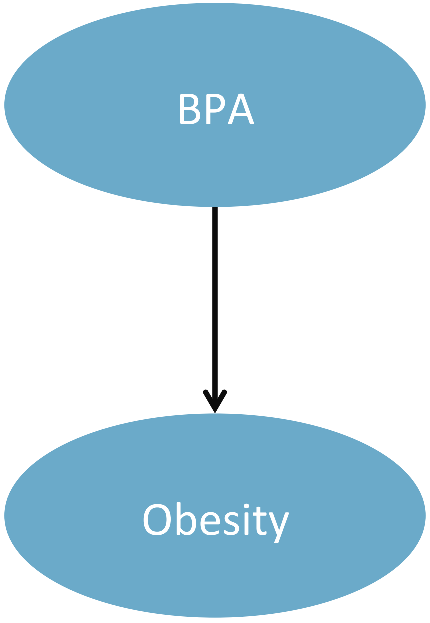
- Search and score
- Constraint based (e.g. PC algorithm)



Uses for BNs

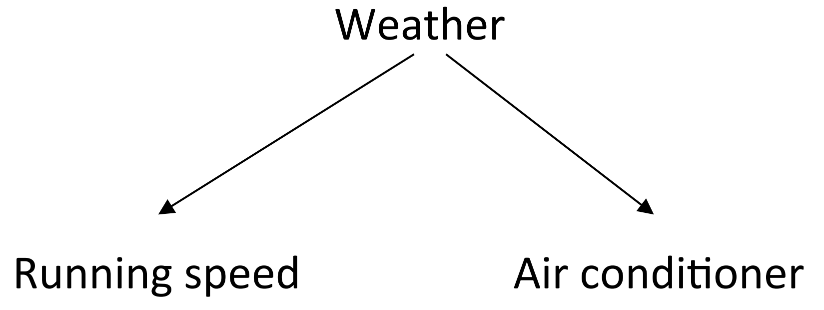
- Actions
  - What happens if we do X?
- Counterfactuals
  - What if things happened differently?
- Explanations
  - Why did X happen?

Manipulability

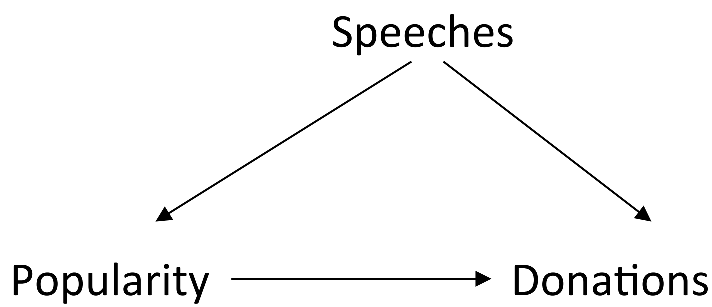


Ideal manipulations

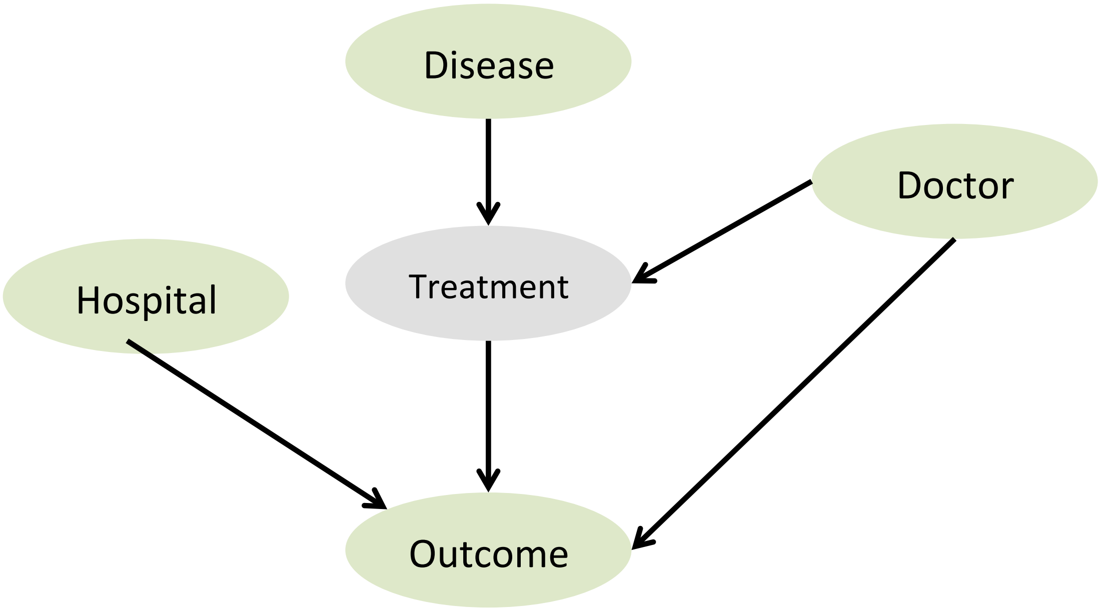
- Definition: change in value of a variable that does not introduce any other changes (except those produced by the change in variable)

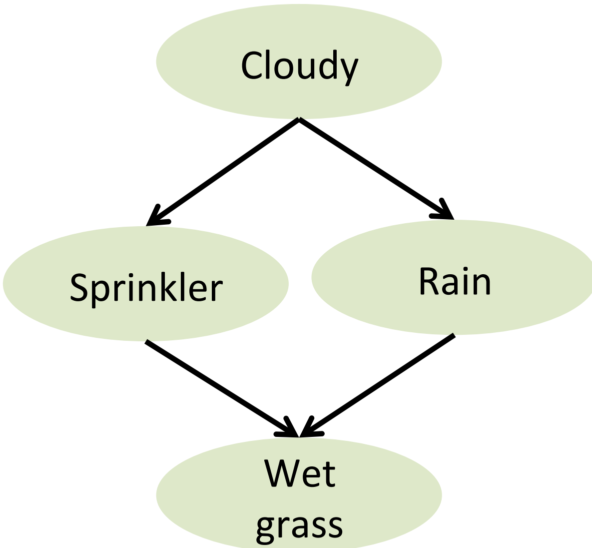


Testing popularity, how do we manipulate it's value?



Seeing versus doing





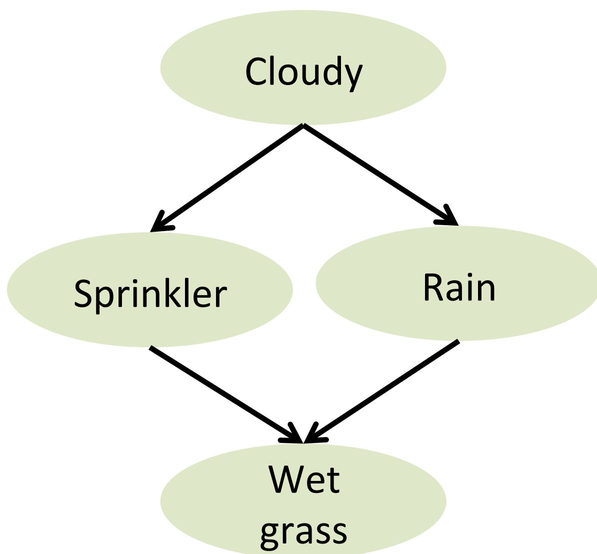
What’s P(C) if I turn the sprinkler on?

Is this the same as P(C|S=T)?

Intervention and joint probability

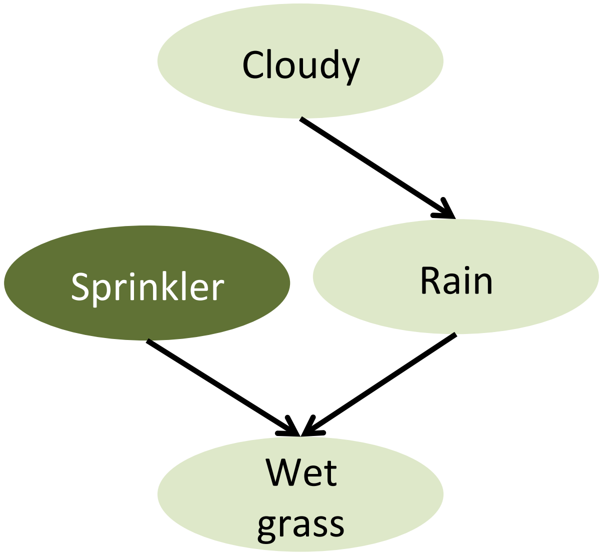
!= just incorporating evidence

- Last week: set value of observed values
- This week: set value by forcing variable to take value <u>independent of its parents' values</u>



If turn on sprinkler, the fact that it's on no longer gives info about C

Intervention and joint probability



$$P(C,S,W,R) = \sum_{C,S,W,R} P(c)P(s \mid c)P(r \mid c)P(w \mid s,r)$$

$$P(C,W,R \mid do(s)) = \sum_{C,W,R} P(c)\underline{P(s)}P(r \mid c)P(w \mid s,r)$$

do() operator

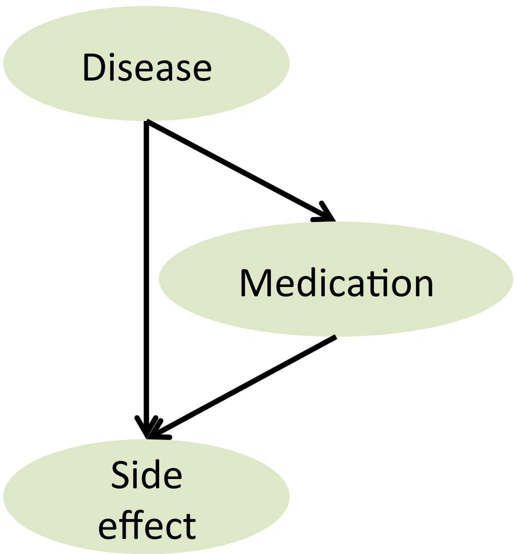
Model can help us determine the effect of interventions

P(X=x|Y=y) != P(X=x|set Y=y)

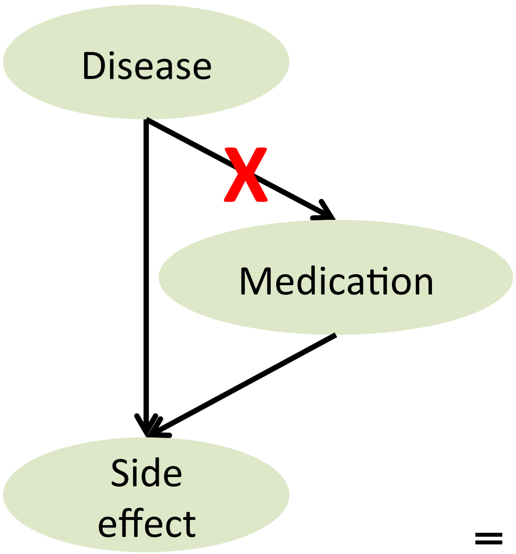
Big assumption: can set variable T/F!

Example

P(S|do(M))



Example



$$P(s \mid \mathrm{do}(m)) = P(s \mid \hat{m}))$$

$$= \sum_d P(s,d,\hat{m}) / P(\hat{m}) = \sum_d P(s \mid d,\hat{m}) P(d \mid \hat{m}) P(\hat{m}) / P(\hat{m})$$

BUT!

$$P(d \mid \hat{m}) = P(d)$$

$$P(m) / P(m) = 1$$

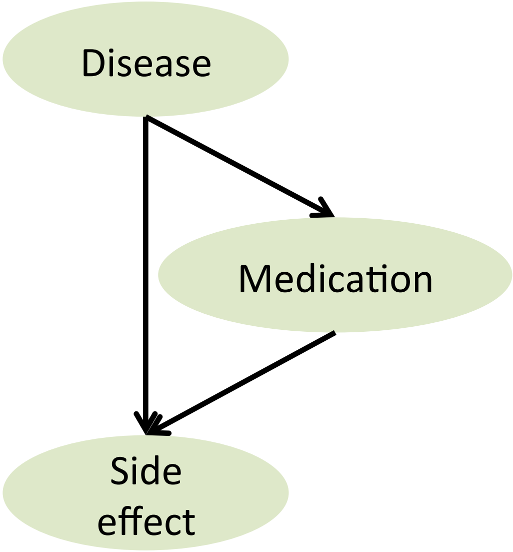
SO

$$= \sum_d P(s \mid d,\hat{m}) P(d)$$

do-calculus rules

1. Insertion/deletion of observations

$$P(y \mid do(x), z, w) = P(y \mid do(x), w) \text{ if } (\mathrm{Y} \perp \mathrm{Z} \mid \mathrm{X}, \mathrm{W})_{G_{\overline{X}}}$$

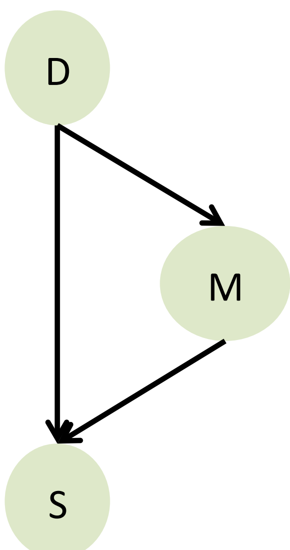


$G_{\overline{X}}$ Means edges into X deleted

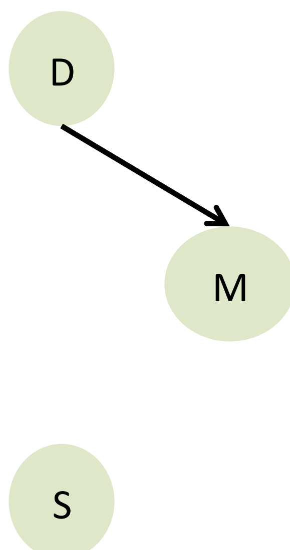
do-calculus rule 1 examples

1. Insertion/deletion of observations

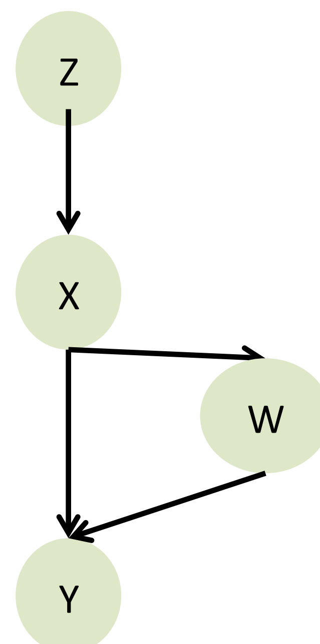
$$P(y \mid do(x), z, w) = P(y \mid do(x), w) \text{ if } (Y \perp Z \mid X, W)_{G_{\overline{X}}}$$



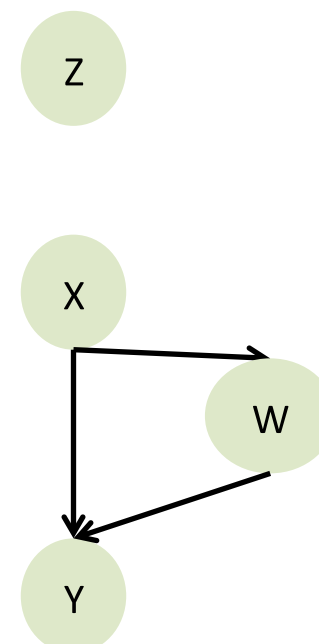
G



$G_{\overline{S}}$



G



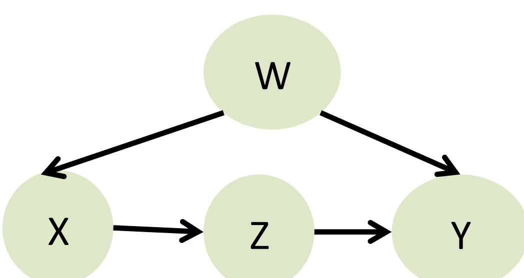
$G_{\overline{X}}$

do-calculus rules

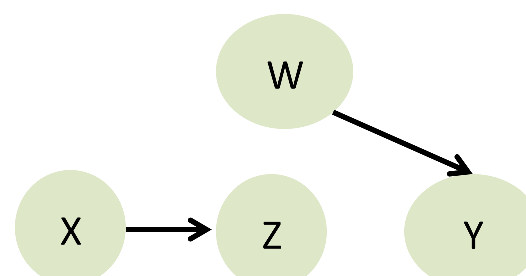
2. Action/Observation exchange

$$P(y \mid do(x), do(z), w) = P(y \mid do(x), z, w) \text{ if } (\mathrm{Y} \perp \mathrm{Z} \mid \mathrm{X},\mathrm{W})_{G\overline{X}\underline{Z}}$$

If remove edges **from** Z, and independent, can replace doing with observing



G



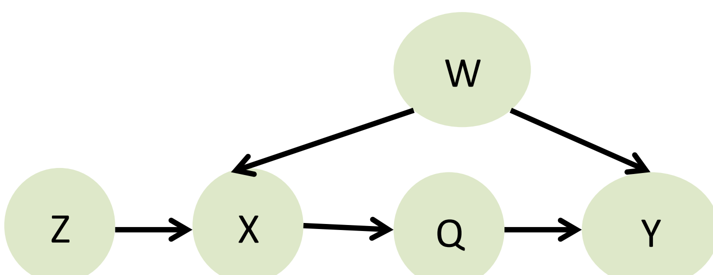
G-X$\underline{\text{Z}}$

do-calculus rules

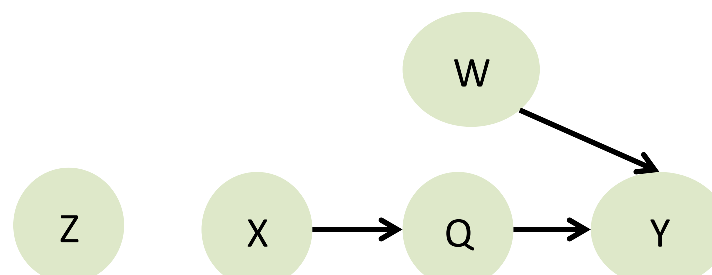
3. Insertion/deletion of actions

$$P(y \mid do(x), do(z), w) = P(y \mid do(x), w) \text{ if } (Y \perp Z \mid X, W)_{G_{\overline{X}, \overline{Z(W)}}}$$

Z(W) is set of Z nodes that are not ancestors of W-nodes in $G_{\overline{X}}$



G



G –X,Z(W)

Summary of do-calculus

1. Insertion/deletion of observations
2. Action/Observation exchange
3. Insertion/deletion of actions

In general, may have unobserved/hidden variables

Some caveats

- Time
- Modularity
- Possibility of intervening
- Efficacy

Counterfactuals reminder

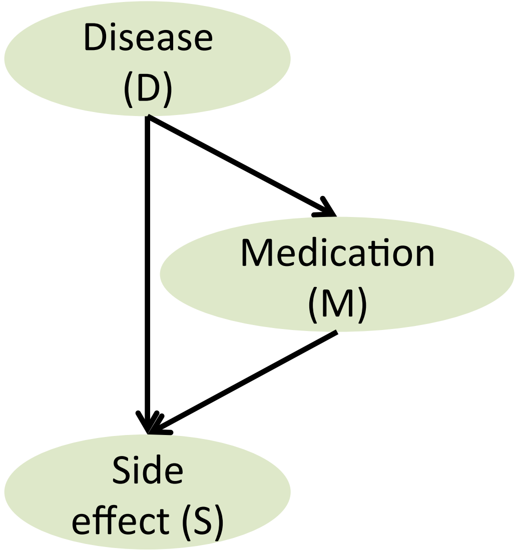
- If I had not gone running, I would not have gotten a sunburn
- If the patient had taken the drug, she would have recovered
- Had I bought shares of Apple stock in 2004, I would have made a large profit

Pearl on Counterfactuals

Like do(), except backward looking and changing value of variable

Three steps

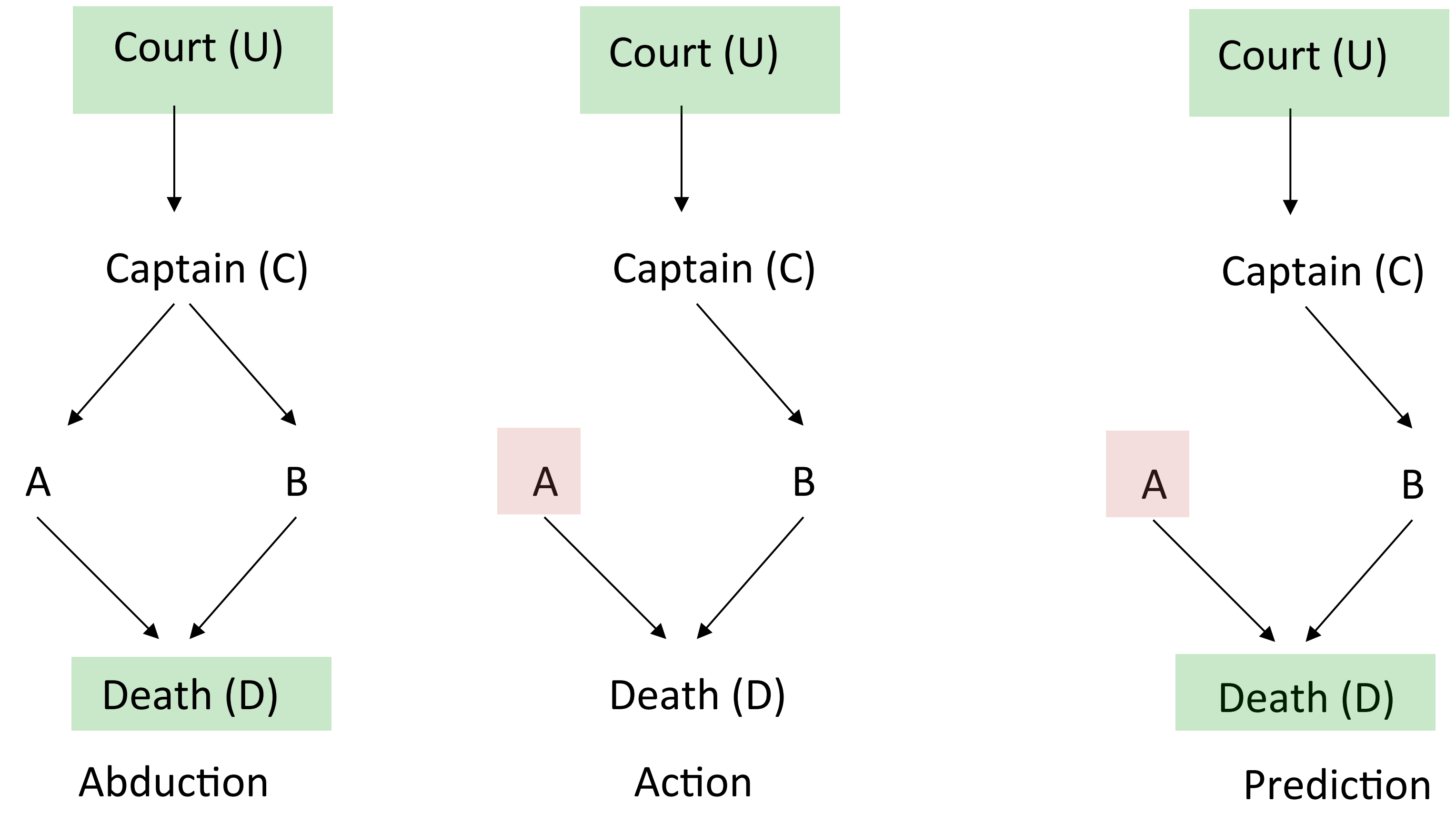
1. Abduction: use evidence to interpret past
2. Action: change to hypothetical values
3. Prediction: see consequences of actions



Example from Pearl

If D, then D would still be true if A were false

$D \rightarrow D_{\neg A}$



Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge University Press

Actual causality

Pearl: token cause = actual cause

- What caused a person’s lung cancer?
- Who is responsible for an accident

Graphical models and explanation

- Graphical model represents relationships between variables
- Observations give truth value of variables in particular scenario
- Evaluate counterfactual queries using model + observations

Pearl's approach to actual cause

- Based on but attempts to solve problems with counterfactuals
  - Bob and Susie and the broken bottle
- Key idea: sustenance (mix of necessity and sufficiency)
  - If Bob's rock missed, would Susie's sustain the glass breaking?

Definitions

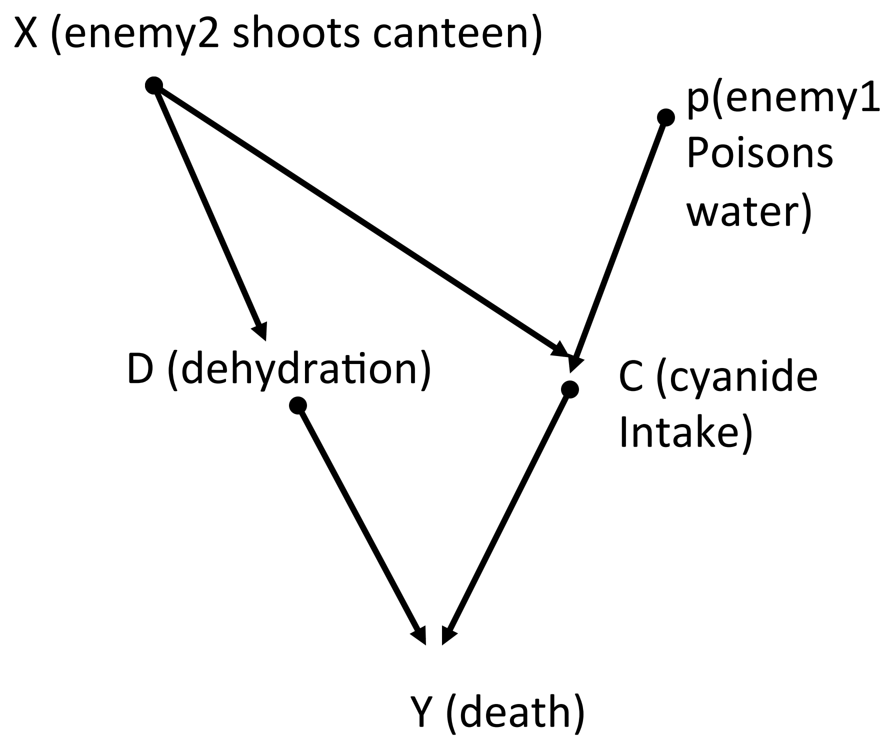
- Depends (necessity)
  - y depends on x, if x is necessary to maintain value of y
- Produces (sufficiency)
  - x can produce y if it can bring about effect when neither are present
- Sustains
  - x sustains y if there is at least one condition where Y will differ from y in absence of x AND Y=y is maintained in presence of x under any set of conditions

Causal Beams

- Causal beam: New model, where we remove all parents except those that minimally sustain their children. Set other parents to some w'.
- x is **actual cause** of y if x is necessary for y in that causal beam for some w'

Example of causal beam

- Did the traveler die of thirst or poisoning?
- Death = C v D

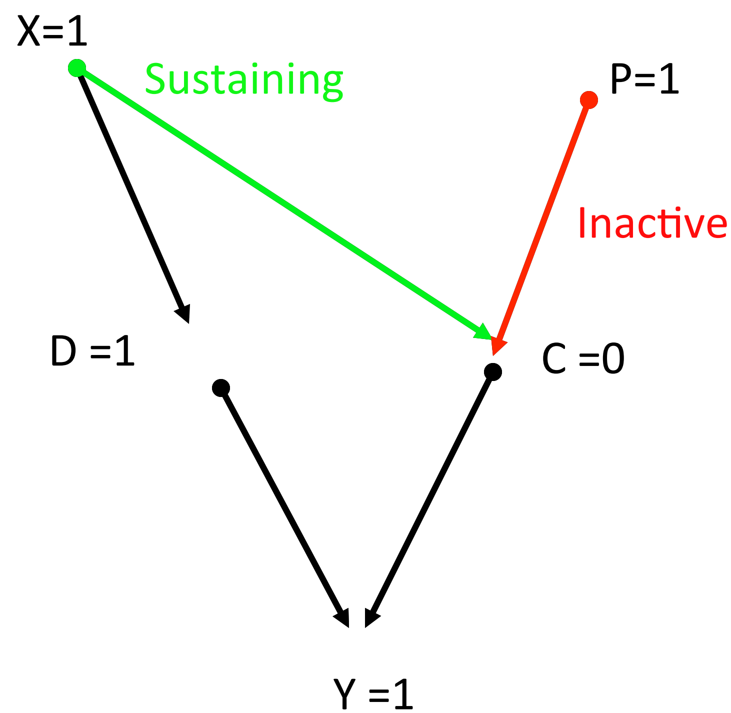


Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge University Press

Constructing the causal beam

- True: X,D,Y,P
- False: C

- Note: C=¬X^P

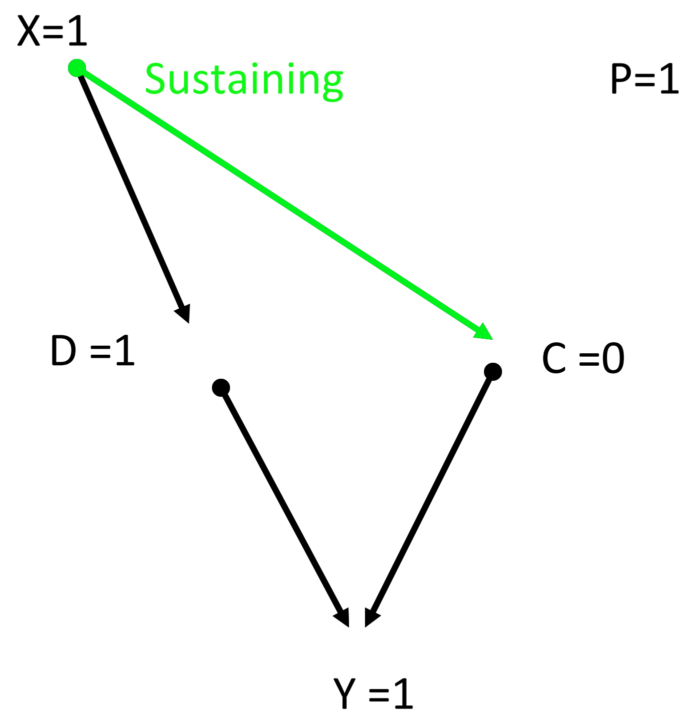


X= shoots canteen, D=dehydration, Y=Death, C=cyanide intake, P=poisons water

Constructing the causal beam

- True: X,D,Y,P
- False: C

- Now: C=¬X

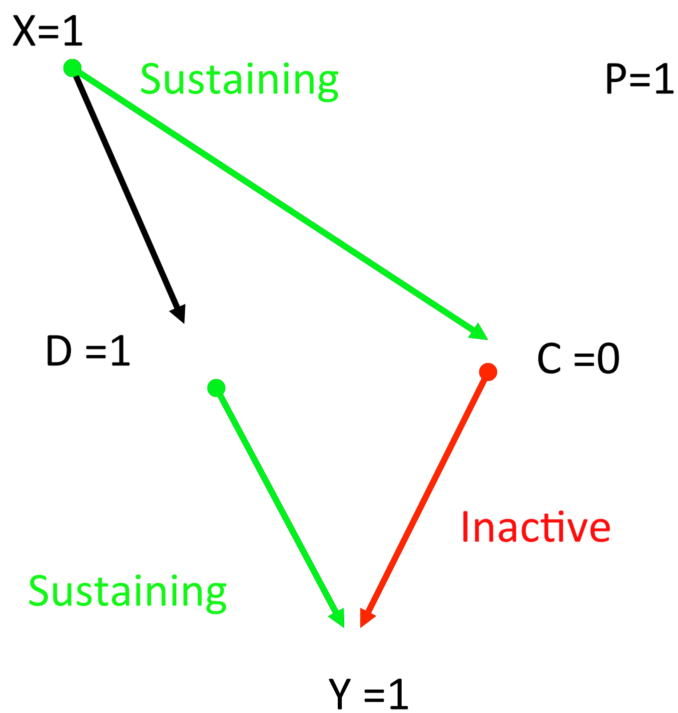


X= shoots canteen, D=dehydration, Y=Death, C=cyanide intake, P=poisons water

Constructing the causal beam

- True: X,D,Y,P
- False: C

- Next: Y=DvC

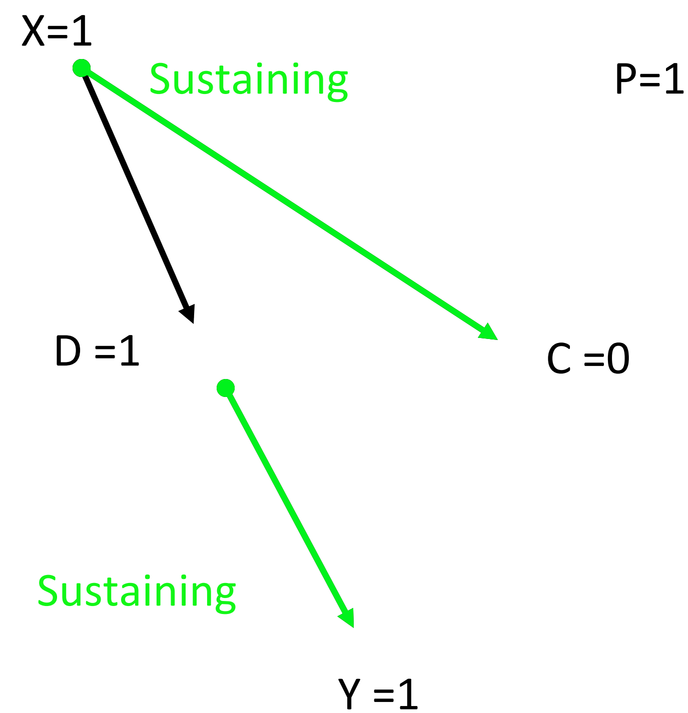


X= shoots canteen, D=dehydration, Y=Death, C=cyanide intake, P=poisons water

Constructing the causal beam

- True: X,D,Y,P
- False: C

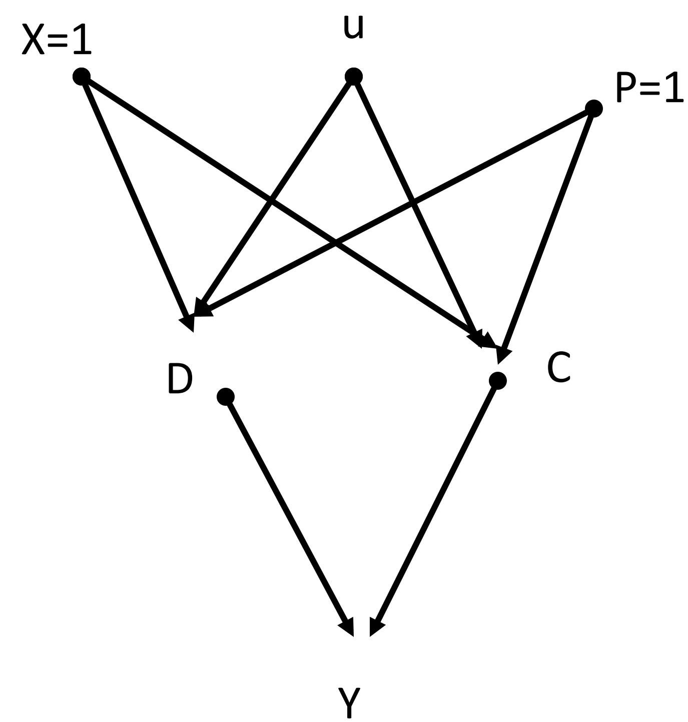
- Finally: Y=D, Y=X



X= shoots canteen, D=dehydration, Y=Death, C=cyanide intake, P=poisons water

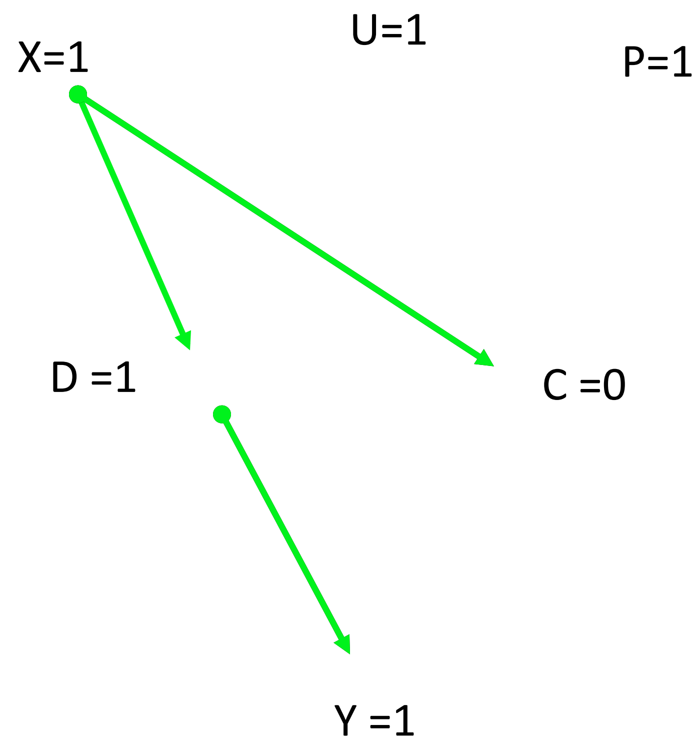
What if we are uncertain?

- U represents time till first drink
- U=1 if canteen emptied before drink
- U=0 otherwise



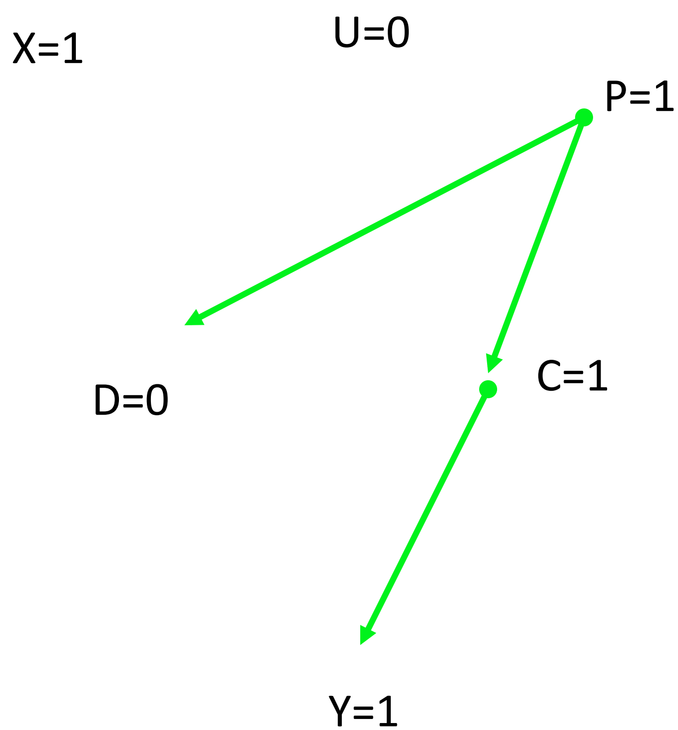
Case 1: u=1

- Canteen emptied before traveler drinks
- Same beam as before



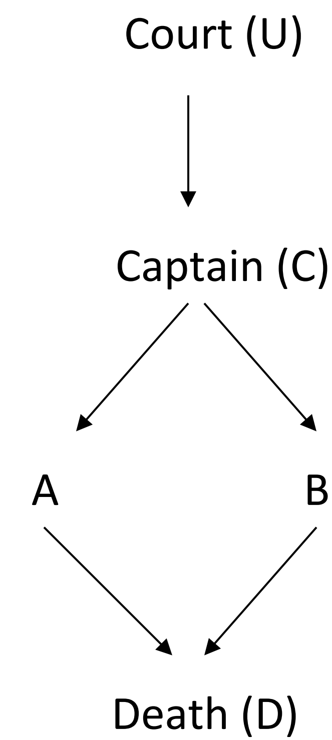
Case 2: u=0

- Drinks before canteen is emptied
- What if U is uncertain?
- Use P(u) to calculate
- P(x caused y)=
  - Sum of P(u) over u where x caused y in u



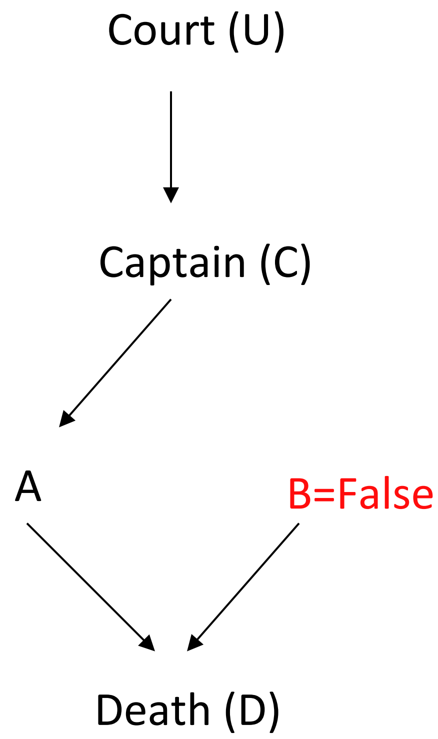
Problem: Over-determination

- Which rifleman caused the death?
- Counterfactual:
  - If not A, then B would have caused D



Problem: Over-determination

- Which rifleman caused the death?
- Counterfactual:
  - If not A, then B would have caused D
- Structural:
  - A sustains D against B



Partial solution: defaults

Example: Neither Jack nor Jill water a plant, and the plant dies. Jill usually waters it, Jack never does

Who's at fault?

Halpern, Joseph Y. "Defaults and Normality in Causal Structures." *KR*. 2008.

Joseph Y. Halpern and Christopher Hitchcock. Graded Causation and Defaults. The British Journal for the Philosophy of Science , 66(2):413–457, 2015.

- Default: what we think is true most of the time
- Typicality: what usually happens (frequency)
- Norms: judgment of what should usually happen

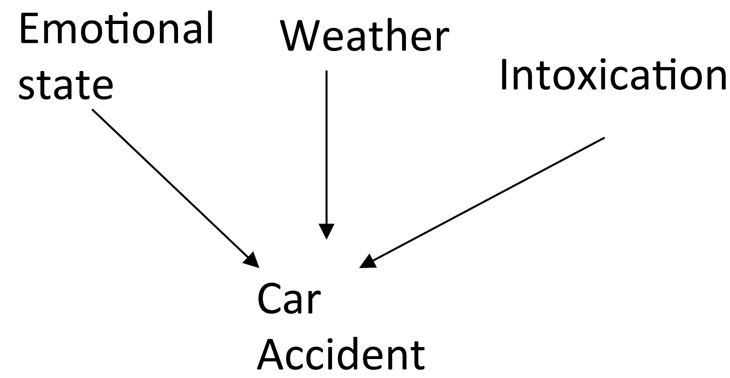
- Main idea: for counterfactuals we compare possible worlds, now we rank them
- Recall pen problem!

Challenges for the actual cause

- Type !=Token
- Subjectivity
- Timing
- Limited by completeness of model

Challenge: subjectivity

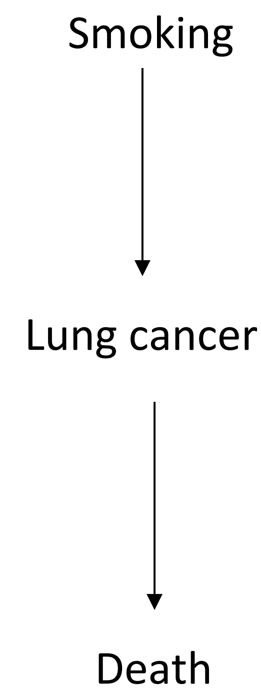
- Variables chosen and what values they are set to affect outcome
  - Smoking vs. duration of smoking
- Note: both queries and their answers may then differ



Challenge: time

- Bob starts smoking (S) Wednesday. He's diagnosed with lung cancer (LC) on Friday. Did his S cause his LC?

- Bob later dies. Was LC the cause?



Inference recap

	BN	DBN	Granger	Temporal logic
Results	Graph			
Time	No			
Data	C/D/M			
Cycles	No			
Latent vars.	Yes			
Prediction	Yes			
Token cause	Counterfactual-based			

Further reading

- Graphical models and causality
  - Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search.* MIT Press
  - Pearl, J. (2000/2009). *Causality: Models, reasoning, and inference.* Cambridge University Press.
- Actual cause
  - Pearl's book
  - Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science*, *56*(4), 843-887.

For next week

- How can we find how long it takes for smoking to cause lung cancer?
- When to buy/sell a stock after you hear some news?

- Read CPT 2.4.2