Causal Inference: prediction, explanation, and intervention

Lecture 5: Causality and Graphical Models

Samantha Kleinberg
samantha.kleinberg@stevens.edu

Next few weeks

- October 7: Time series
- October 15 (Tuesday class!): Lecture + midterm review/Q&A
- October 21: Midterm
- October 28: Project proposal due

- November 25 and December 2: 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**

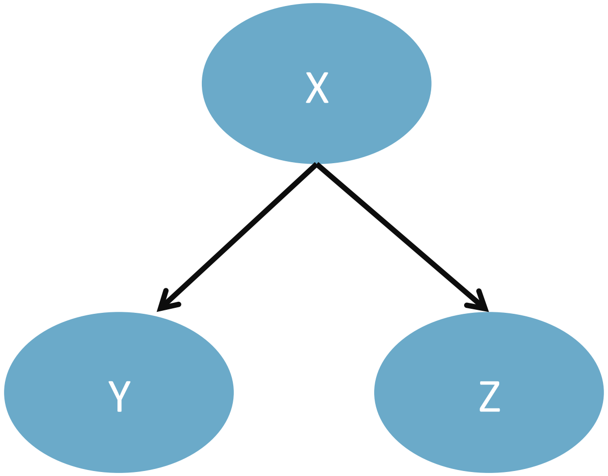
Today

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

Markov condition

(from last week)

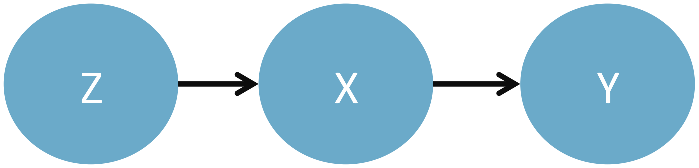
Node independent of non-descendants given its parents



$$Y \perp Z \mid X$$

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

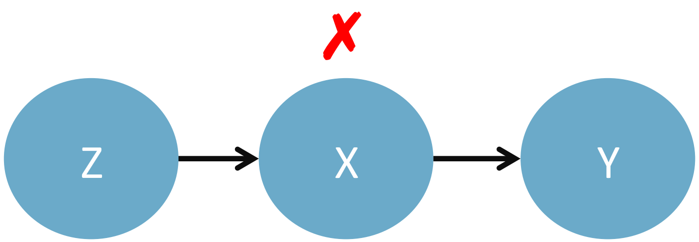
d-separation and Markov blanket

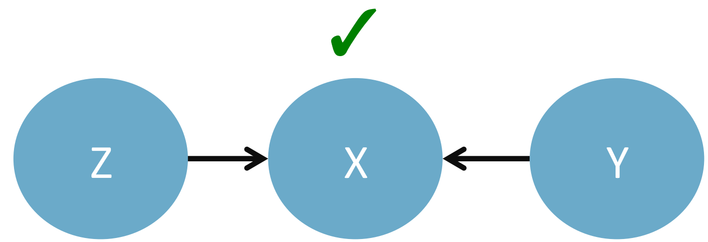
Markov blanket: set of nodes that separate a node from all others

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

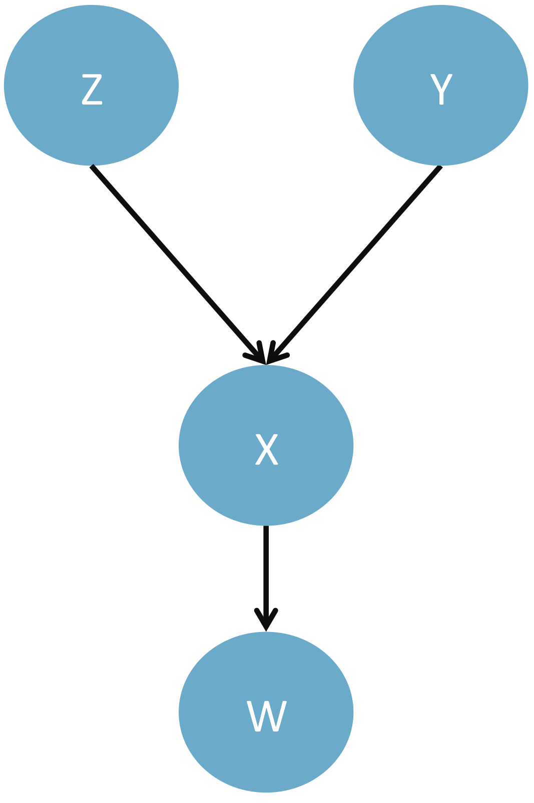
Example 2

- X → Y ← Z

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

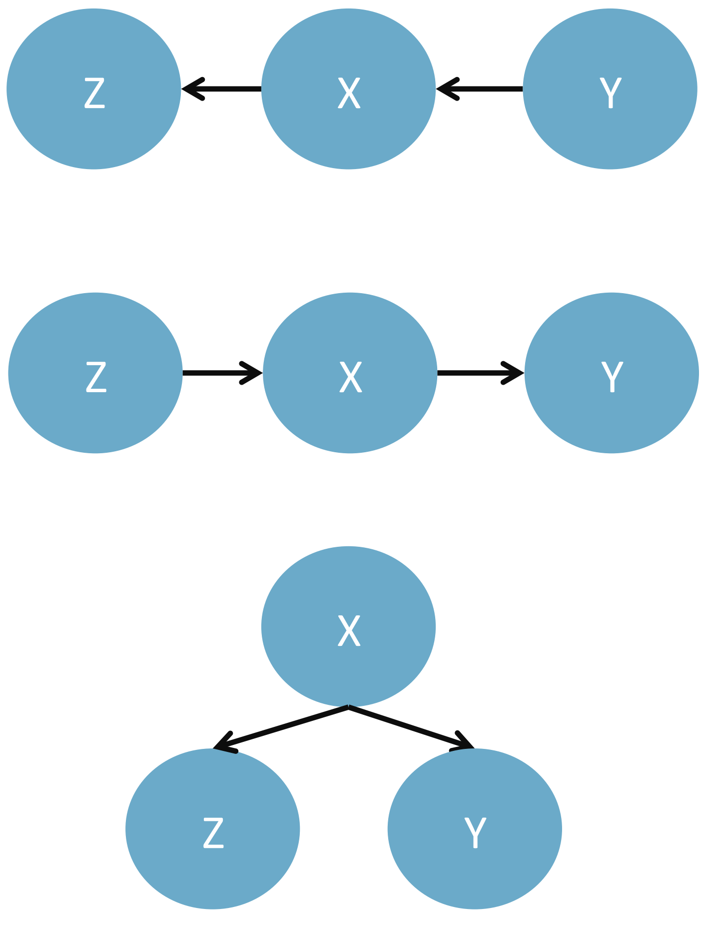
Example 3

- Are Y,Z d-separated by W?



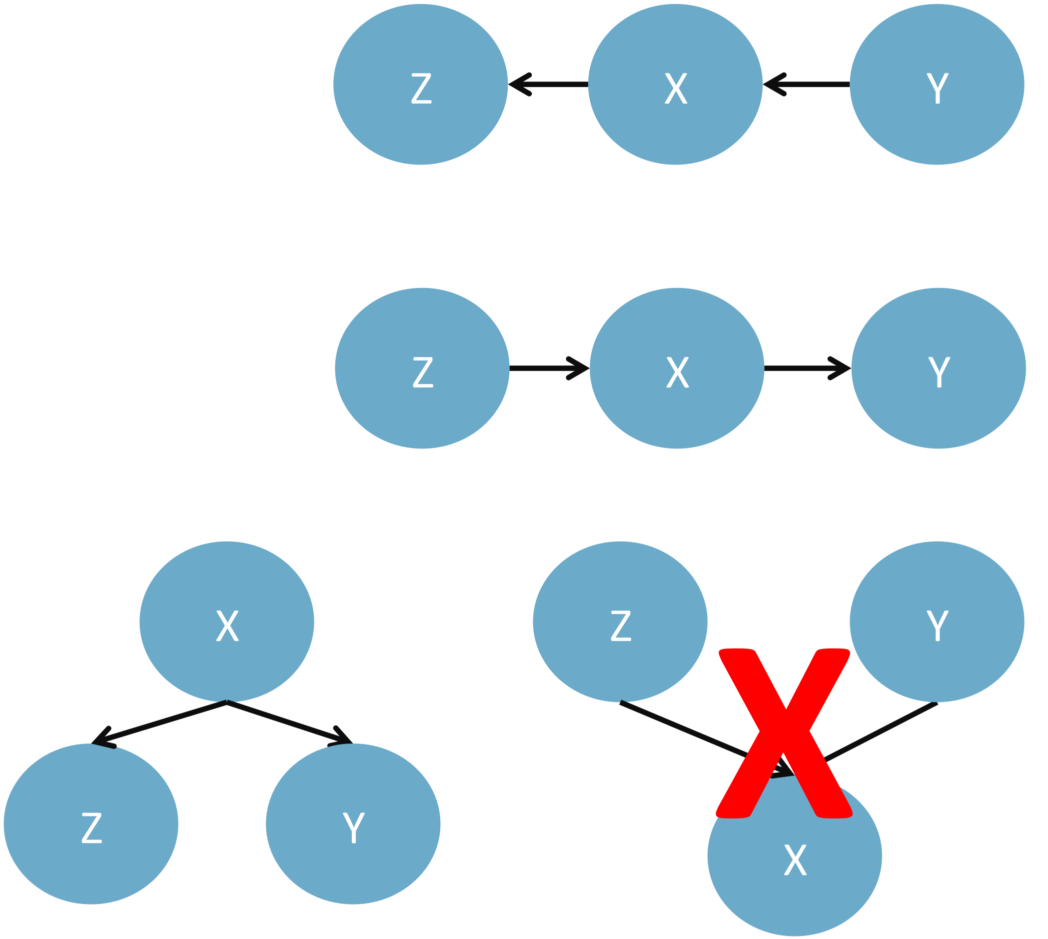
But…

$$Y \perp Z \mid X$$

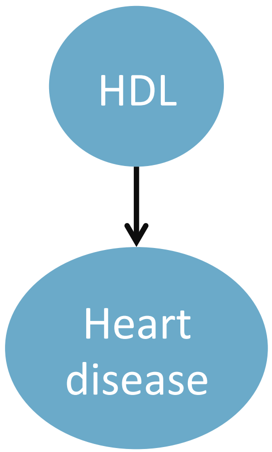


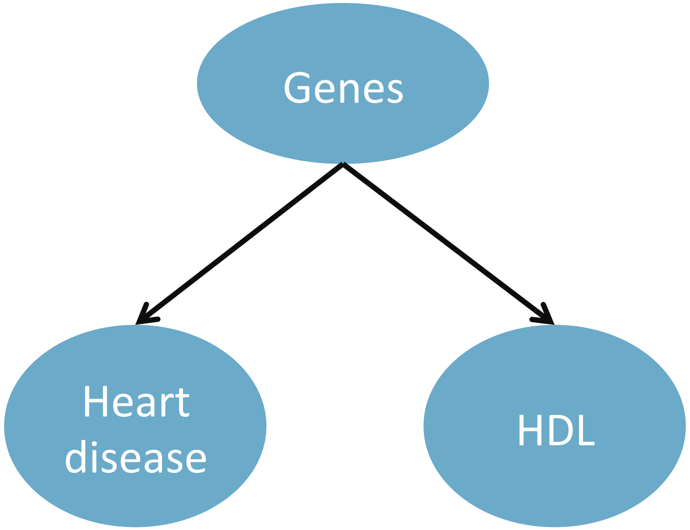
But...

$$Y \perp Z \mid X$$



…And





…Also

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

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

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

$$C_1 \not\perp C_2$$

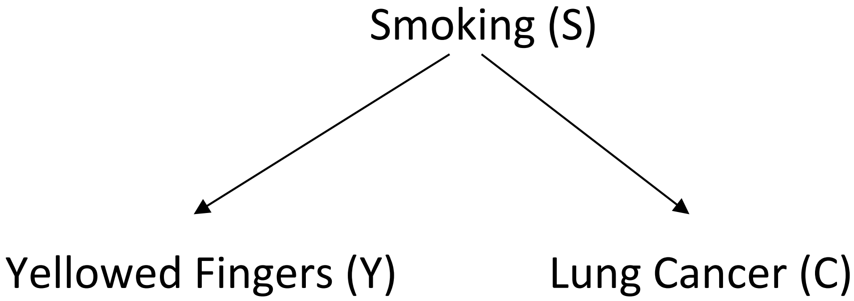
Causal interpretation

- 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

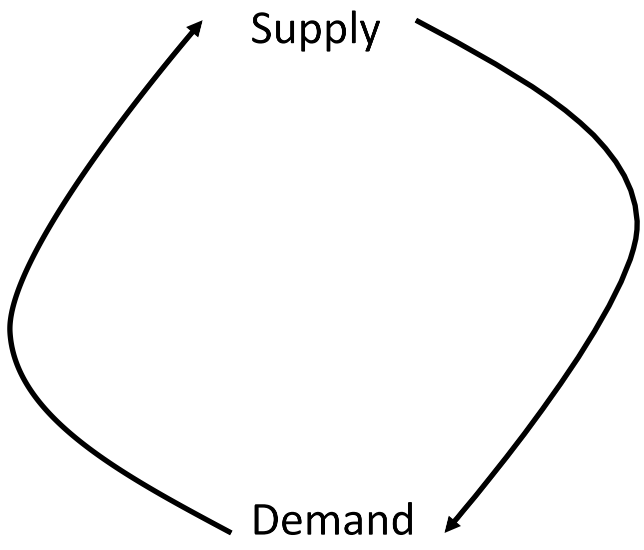
Recall Common Cause Principle (CCP)

– If P(X^Y)>P(X)P(Y) then either X causes Y (or vice versa) or they have a common cause

Now: if P(X^Y)>P(X)P(Y) and they have a common cause C   X, Y | C

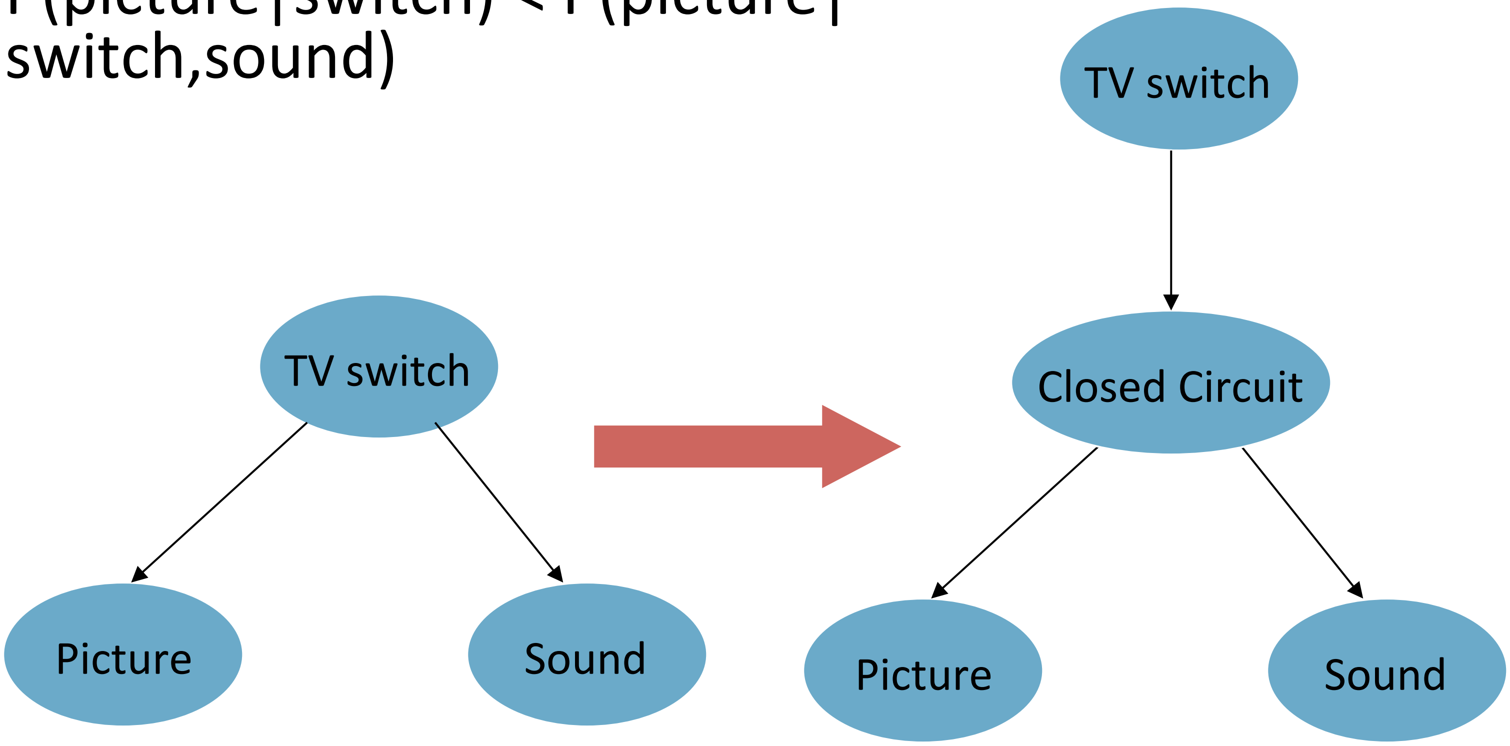
Note that CCP seeks single common cause. CMC allows for sets of nodes.

Problems: feedback



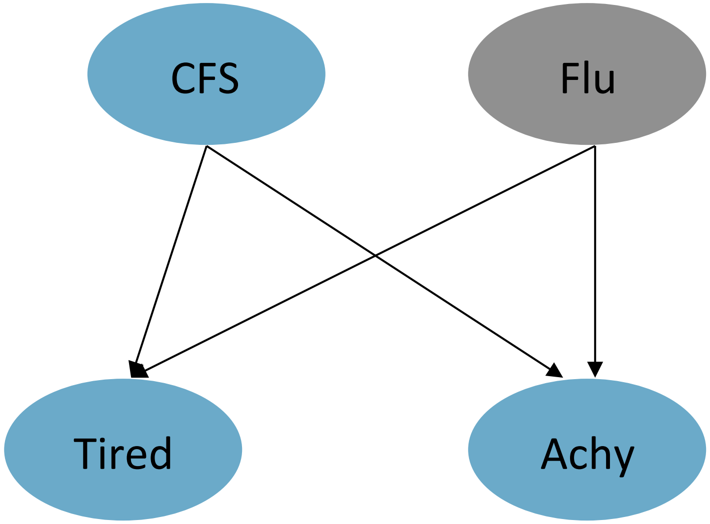
Problems: Indeterminism

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

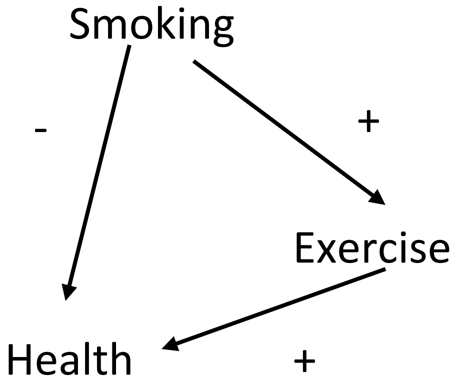
- Complete: all common causes included, all causal relations among variables included
- Incomplete: not all intermediate factors necessarily included

Faithfulness

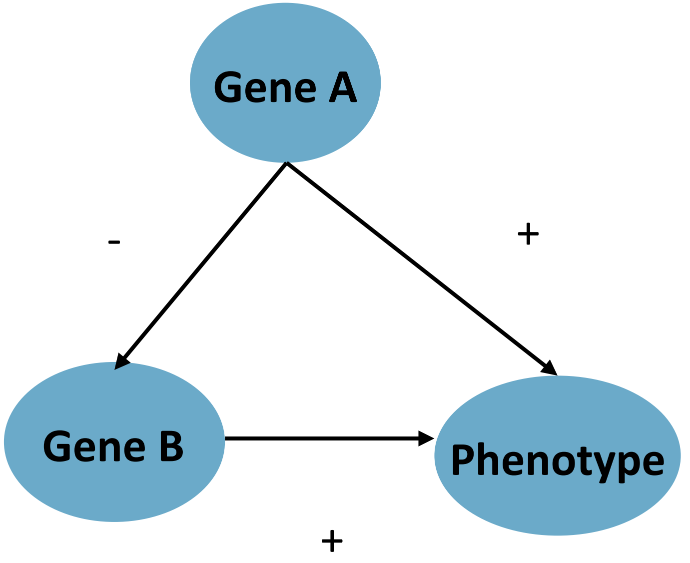
Exactly the dependencies in the underlying structure hold in the data

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

Example

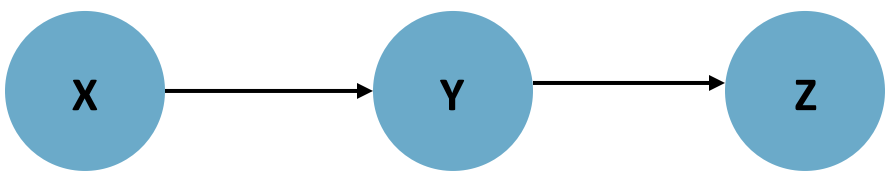


Another example

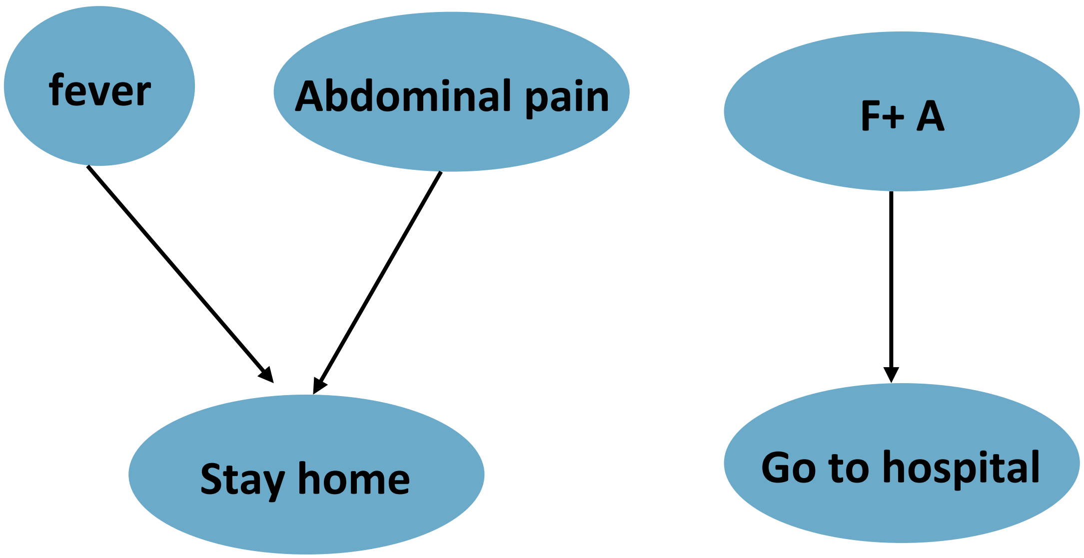


A final example (deterministic chain)

$$X \perp Z \mid Y$$



Selection bias



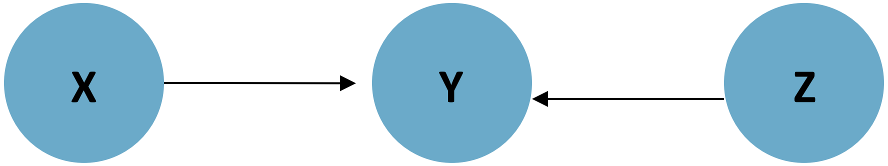
Cooper, G. F. (1999). An overview of the representation and discovery of causal relationships using bayesian networks. In C. Glymour & G. F. Cooper (Eds.), *Computation, causation, and discovery.* AAAI Press and MIT Press

Recap of problems for faithfulness

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

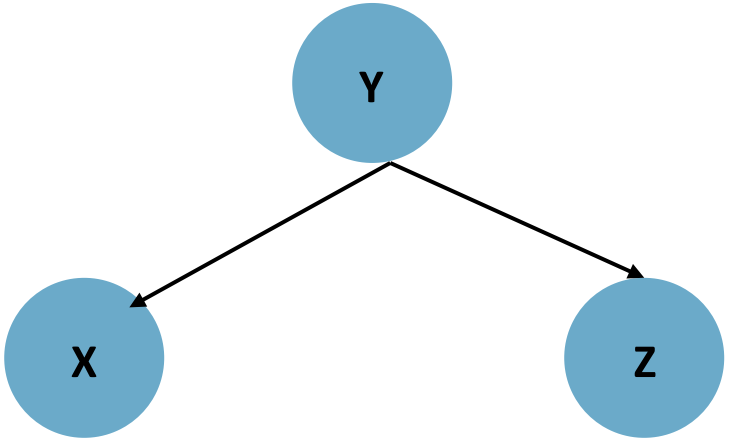
Quick recap

- 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

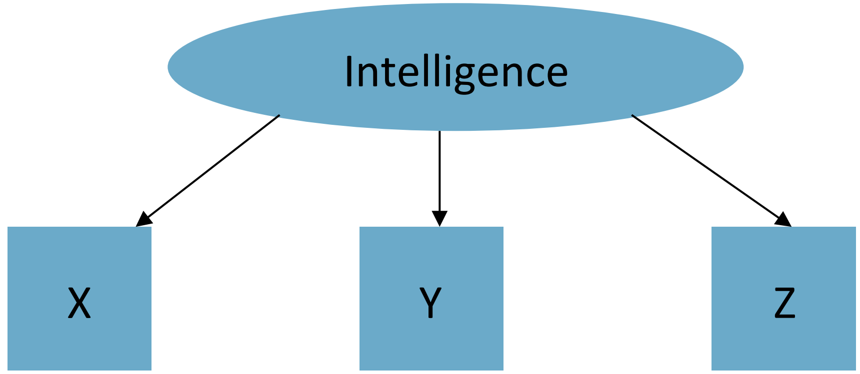


Completeness vs. sufficiency

Completeness: common causes are included in causal graph

Sufficiency: all common causes have been measured

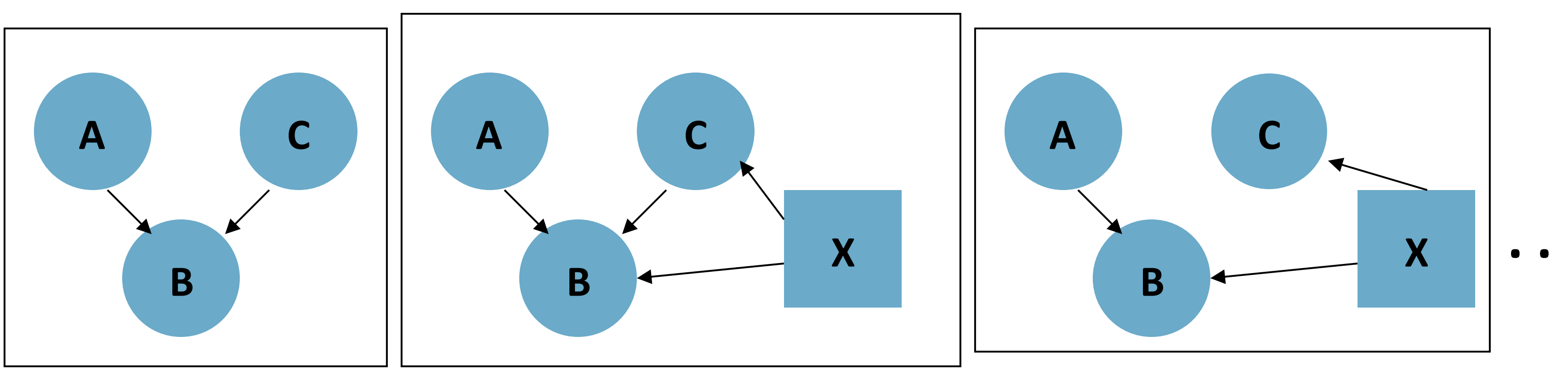
Example



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:



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)
- Complexity
- Multiple testing
- What structures DAG can/cannot represent
- Variable representation

The good news

- Can add time
- Can experiment
- Methods for testing assumptions
- Inference with latent variables

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

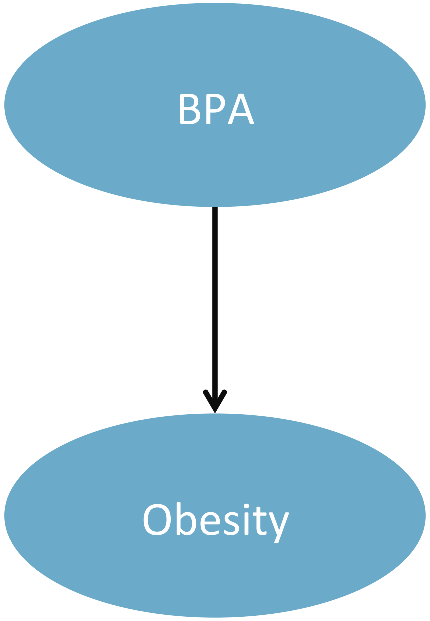
Same methods we discussed last week

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

Pearl on causality

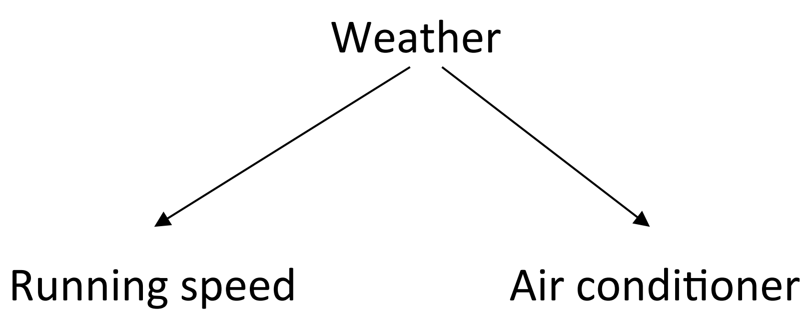
- Actions
  - What happens if we do X?
- Counterfactuals
  - What if things happened differently?
- Explanations
  - Why did X happen?

Manipulability

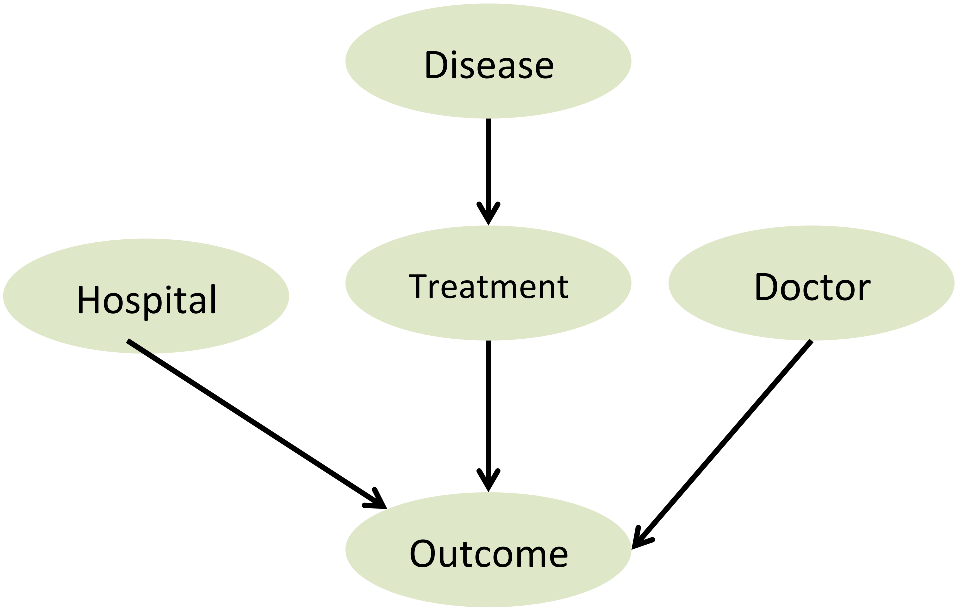


Ideal manipulations

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



Seeing versus doing



Predicting the effects of actions

Will an Aspirin a Day Keep Cancer Away?

Data suggesting that regular aspirin use lowers cancer risk has accumulated to the point where some argue that it's time to recommend that many more people take the drug

IN THE LATE 1970S, A SURGEON IN Melbourne, Australia, wanted to figure out why his country had a relatively high rate of colorectal cancer. After he and colleagues interviewed more than 700 cancer patients and a comparable number of healthy people, they reported in 1987 and 1988 that Australians' penchant for beer, fatty foods, and red meat all seemed to predispose them to the disease. The researchers also found a surprising protective factor: People who regularly used aspirin were 40% less likely to develop colorectal cancer than those who didn't take the drug.

That first hint that the age-old headache remedy also blocks intestinal tumors helped spur a wave of research in animals and clinical trials that established that aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) protect against colon cancer. And now, 2 decades later, aspirin is generating new excitement among cancer researchers. A series of studies from the United Kingdom in the last 2 years has offered the first evidence from placebo-controlled clinical trials that regularly taking low doses of aspirin wards off other types of cancer as well.

The studies, which tallied cancer cases among people who had been taking aspirin for years to prevent vascular events such as heart attack and stroke, found that death rates from several tumor types were as much as 37% lower. And even in the people who developed a cancer, taking aspirin seemed to slow the spread of tumors to other parts of the body. "It's just about the first proof of principle that a simple compound of any kind can reduce the risk of several cancers," says lead researcher Peter Rothwell of the University of Oxford in the United Kingdom.

These reports have raised the tantalizing possibility that aspirin could serve as the first anticancer drug for the general population. "It reshapes the debate about the risks and benefits of aspirin for cancer prevention," says colorectal cancer researcher Andrew Chan of Massachusetts General Hospital in Boston.

CREDIT: GETTY IMAGES

Because aspirin can cause stomach upset and dangerous internal bleeding, U.S. guidelines now recommend that only people at elevated risk for heart disease or stroke take low doses of the medicine, typically 81 milligrams a day. But Chan and others suggest that medical societies and policymakers should also consider aspirin's general cancer-fighting effects. Rothwell, who is 48, is so convinced by his team's data, for example, that he's begun taking aspirin daily even though he has no risk factors for vascular disease.

Versatile. Aspirin may block cancer as well as vascular disease.

The new British aspirin studies are also fueling basic research on NSAIDs to ward off cancer, a field that lost momentum in the past decade when one NSAID, Vioxx, was pulled off the market because of safety concerns. Partly because of the U.K. results, U.S. National Cancer Institute (NCI) Director Harold Varmus last year added to his list of 24 "Provocative Questions" one that asks what is the mechanism by which aspirin and other NSAIDs protect against cancer (see sidebar, p. 1472). NCI hopes the attempt to resolve the question—there are several competing theories—will lead to next-generation anticancer drugs and biomarkers showing who will respond to them.

But some cancer researchers say health policymakers needn't wait for a better aspirin and should advocate wider use of the current one. "This is a really extraordinary opportunity if everything bears out, because you have a drug that costs a penny a pill at a low dose, that can prevent the two major causes of death in the Western world," says epidemiologist Michael Thun of the American Cancer Society in Atlanta.

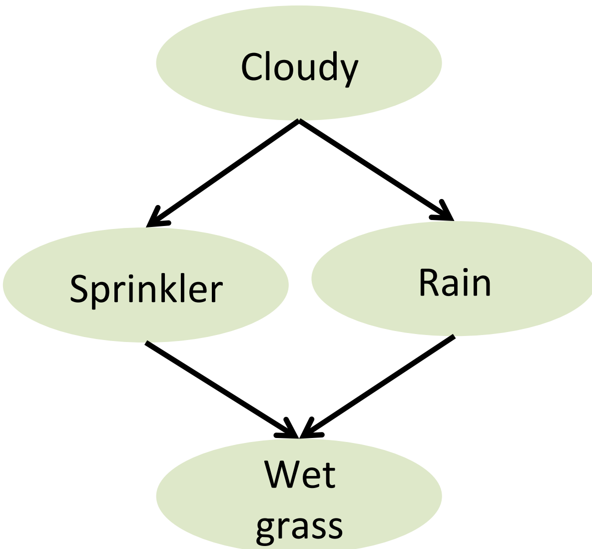
Comeback

Thun's optimism is the latest high in the up-and-down story of NSAIDs as a cancer preventive. Aspirin and some other NSAIDs first bore out their promise in trials published starting in 2000 among people who had had precancerous colon polyps removed and others genetically prone to colorectal cancer. The drugs protected them from polyps and premalignant tumors that precede full-blown cancer. Such people are now sometimes advised by their doctors to take NSAIDs as an adjunct to surgery to prevent polyps from recurring.

Although epidemiological evidence has suggested that aspirin could have broader anticancer effects, those results aren't conclusive. They come from studies in which people answered questions about their past use of medications—a design prone to bias in part because memories aren't reliable. And hopes for aspirin fell in 2005 when a huge prospective study—a randomized controlled trial called the Women's Health Study (WHS)—failed to show a reduction in the incidence of cancer in nearly 40,000 women who took low-dose aspirin every other day for 10 years.

The field of NSAIDs for cancer prevention also suffered a black eye when two arthritis drugs developed to avoid the side effects of aspirin, the COX-2 inhibitors Vioxx and Celebrex, were tested in trials to prevent colon polyps. In one large trial, Vioxx cut the risk of polyps but raised the chances of heart attack so much that risk arguably outweighed the cancer benefit. Vioxx was pulled off the market in 2004 and Celebrex, although still used for preventing colorectal cancer in people at high risk of such disease, now carries a

Downloaded from www.sciencemag.org on October 1, 2012



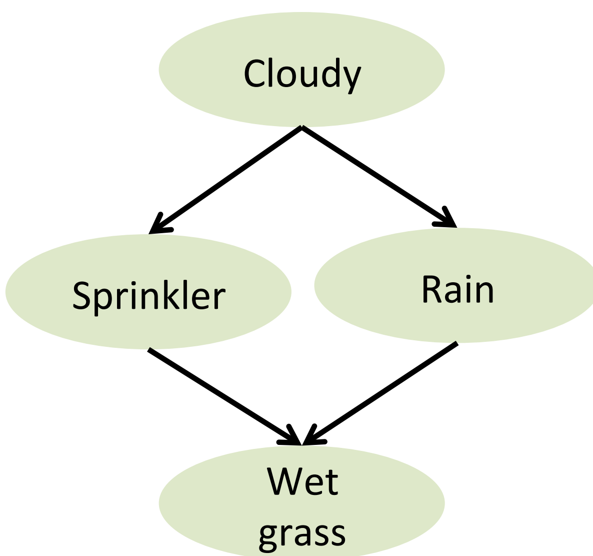
What's P(C) if I turn the sprinkler on?

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

Intervention and joint probability

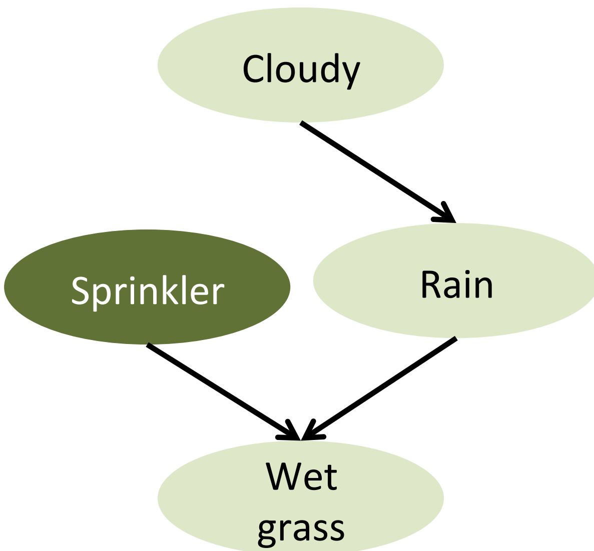
!= just incorporating evidence

- 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



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

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

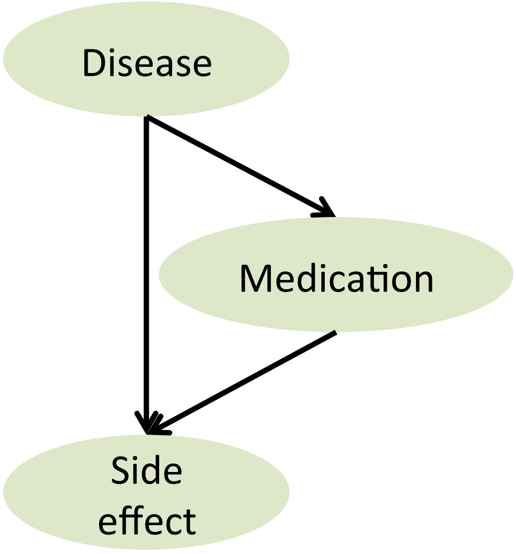
do() operator

Model can help us determine the effect of interventions

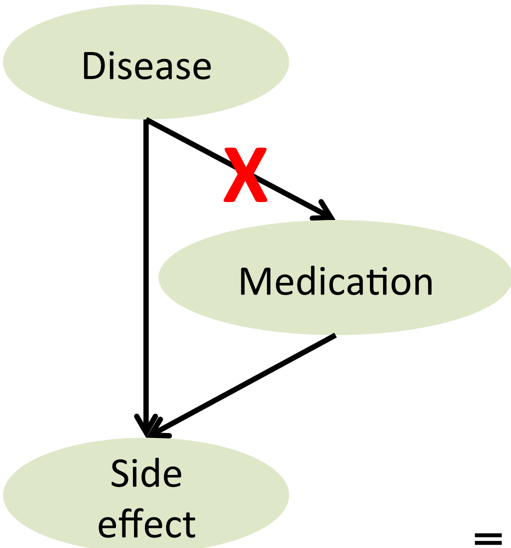
P(X=x|Y=y) != P(X=x|set Y=y)

Example

$P(S|do(M))$



Example



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

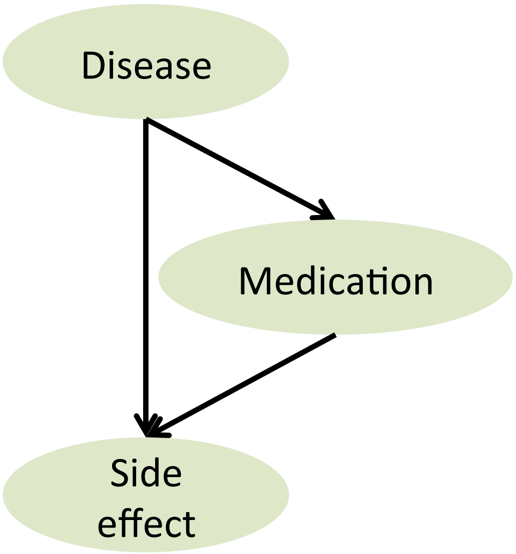
$$= \sum_d P(s, d, \hat{m}) / P(\hat{m}) = \sum_d P(s \mid d, \hat{m}) P(d \mid \hat{m}) P(\hat{m}) / P(\hat{m})$$

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

do-calculus rules

1. Insertion/deletion of observations

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

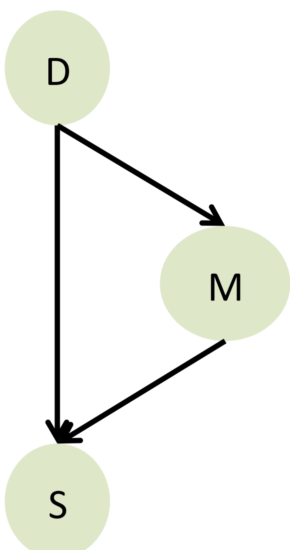


$G_{\overline{X}}$ Means edges into X deleted

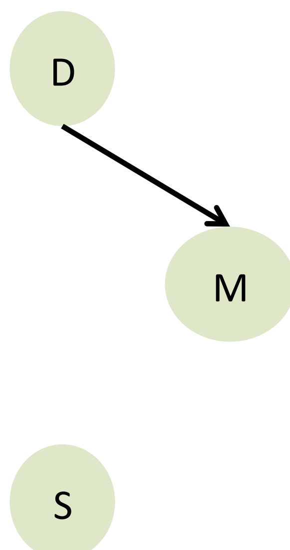
do-calculus rules

1. Insertion/deletion of observations

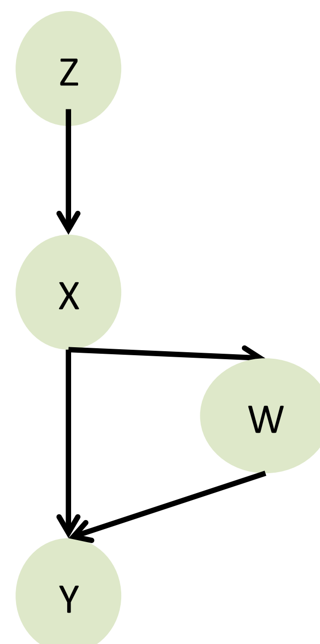
$$P(y \mid do(x), z, w) = P(y \mid do(x), w) \text{ if } (\mathrm{Y} \perp \mathrm{Z} \mid \mathrm{X}, \mathrm{W})_{G_{\overline{X}}}$$



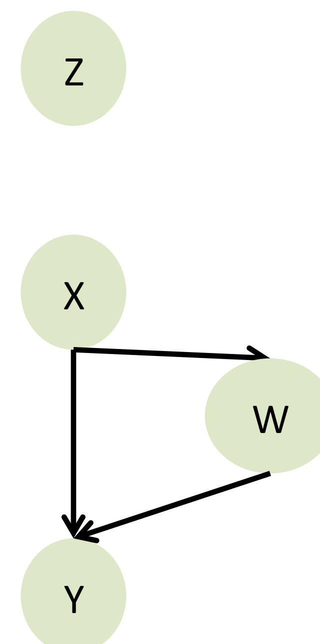
G



$G_{\overline{S}}$



G



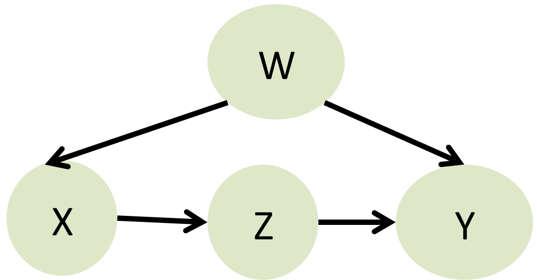
$G_{\overline{X}}$

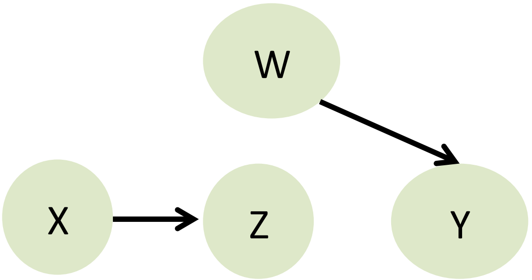
do-calculus rules

2. Action/Observation exchange

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

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



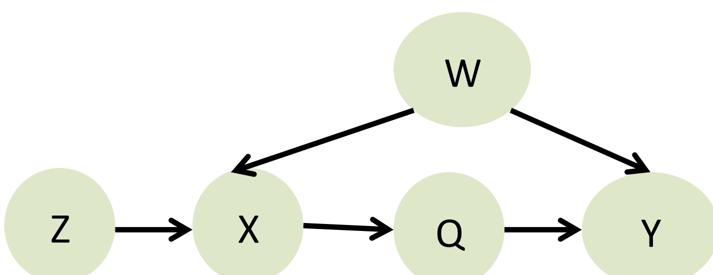


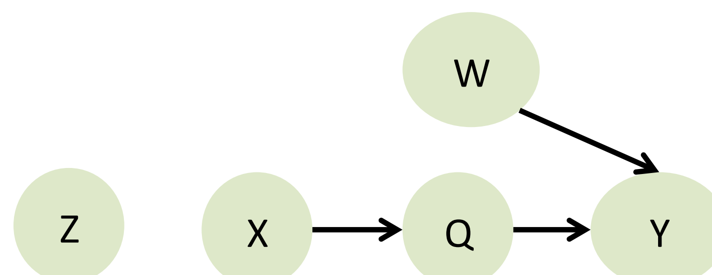
do-calculus rules

3. Insertion/deletion of actions

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

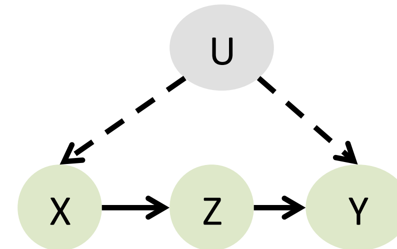
Z(W) is set of Z nodes that are not ancestors of W-nodes in $G_{\overline{X}}$





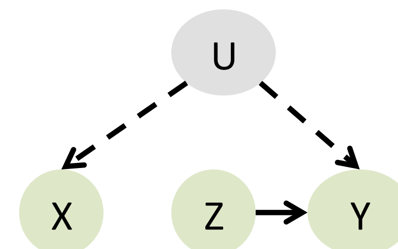
Example

$$P(y \mid do(z)) = P(y \mid \hat{z})$$



G

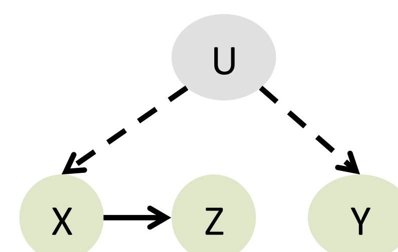
$$P(y \mid \hat{z}) = \sum_x P(y \mid x, \hat{z}) P(x \mid \hat{z})$$



$G_{\overline{Z}}$

With rule 3 (insert/delete action):

$$P(x \mid \hat{z}) = P(x) \text{ since } (\mathrm{Z} \perp \mathrm{X})_{\mathrm{G}\overline{\mathrm{Z}}}$$



$G_{\underline{Z}}$

With rule 2 (action/observation exchange):

$$P(y \mid x, \hat{z}) = P(y \mid x, z) \text{ since } (Z \perp Y \mid X)_{G\underline{Z}}$$

$$P(y \mid \hat{z}) = \sum_x P(y \mid x, z) P(x)$$

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

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

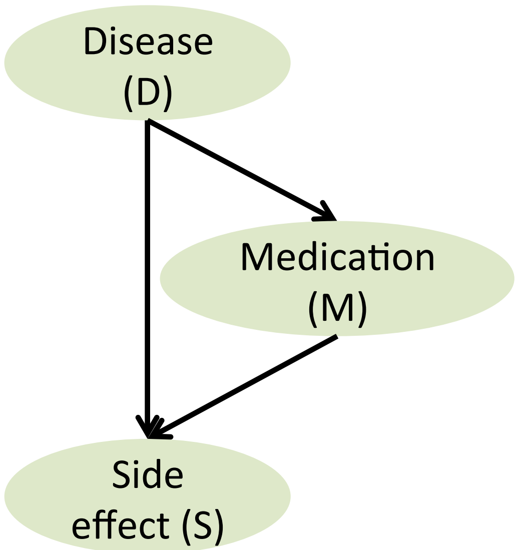
- Time
- Modularity
- Possibility of intervening
- Efficacy

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

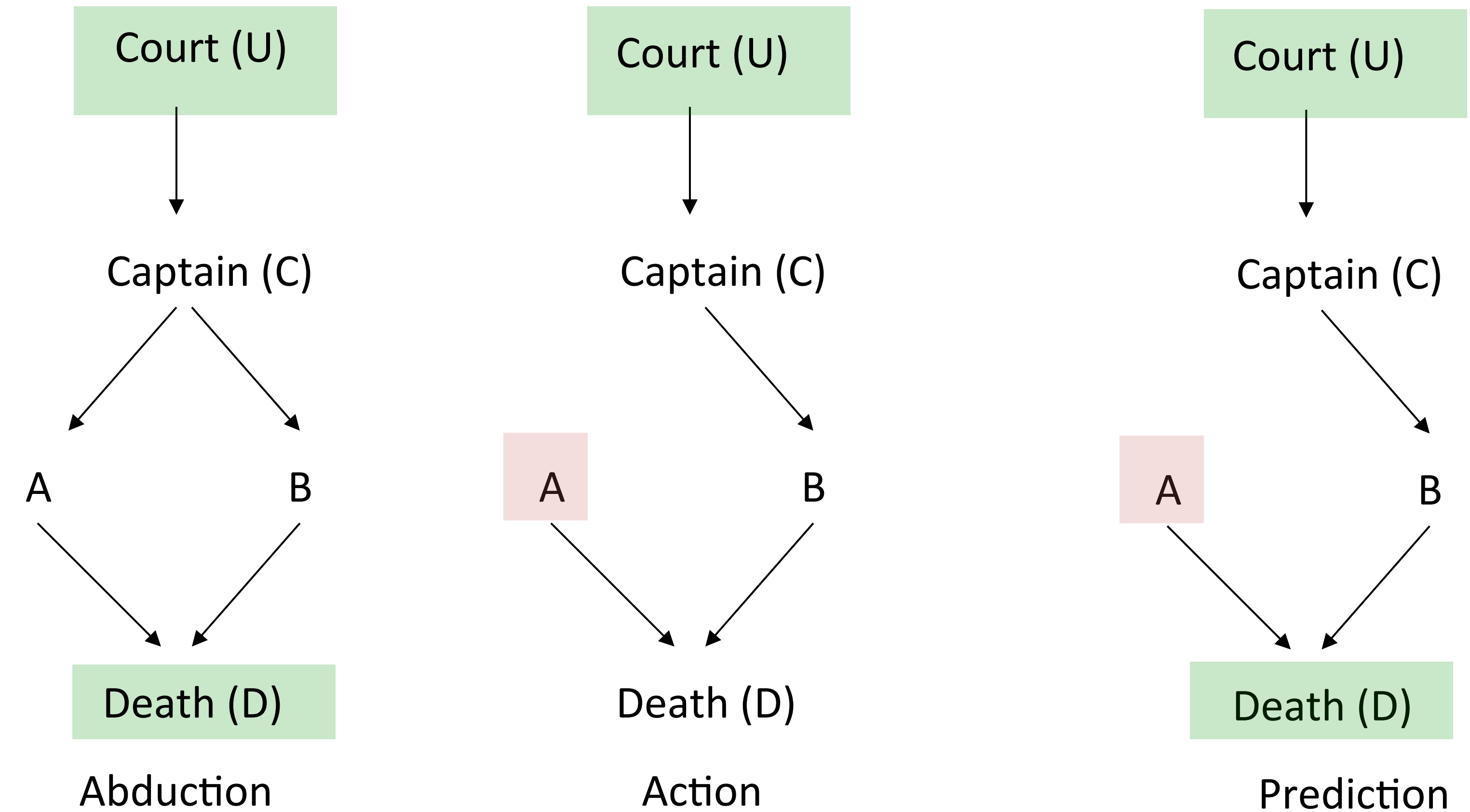
- Like do(), except backward looking and changing value of variable
- Three steps
  1. Abduction
  2. Action
  3. Prediction



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 that represents relationships between variables
- Observations give truth value of variables in particular scenario
- Use model to evaluate counterfactual queries

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

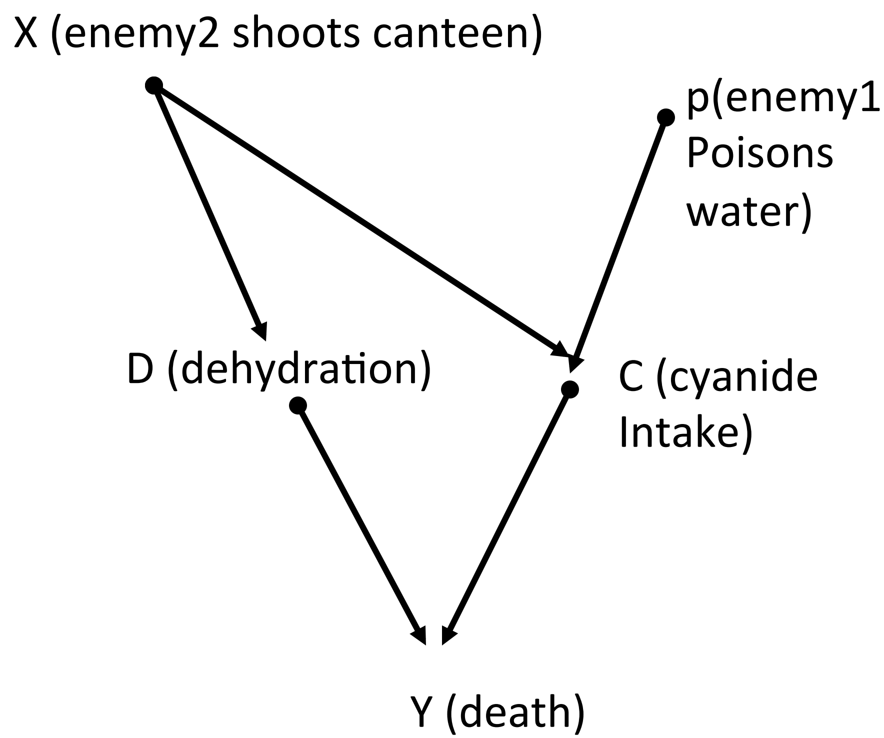
- Depends (necessity)
  - y depends on x, if x is necessary to maintain value of y
- Produces (sufficiency)
  - x can produce y if it can bring about effect when neither are present
- Sustains
  - x sustains y if there is at least one condition where Y will differ from y in absence of x AND Y=y is maintained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

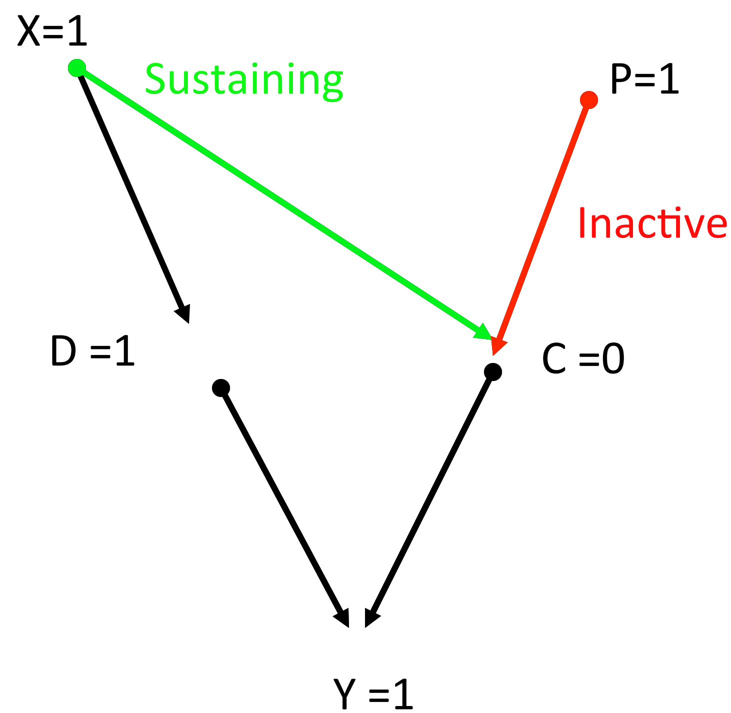


Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge University Press

Constructing the causal beam

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

- Note: $C = \neg X \wedge P$

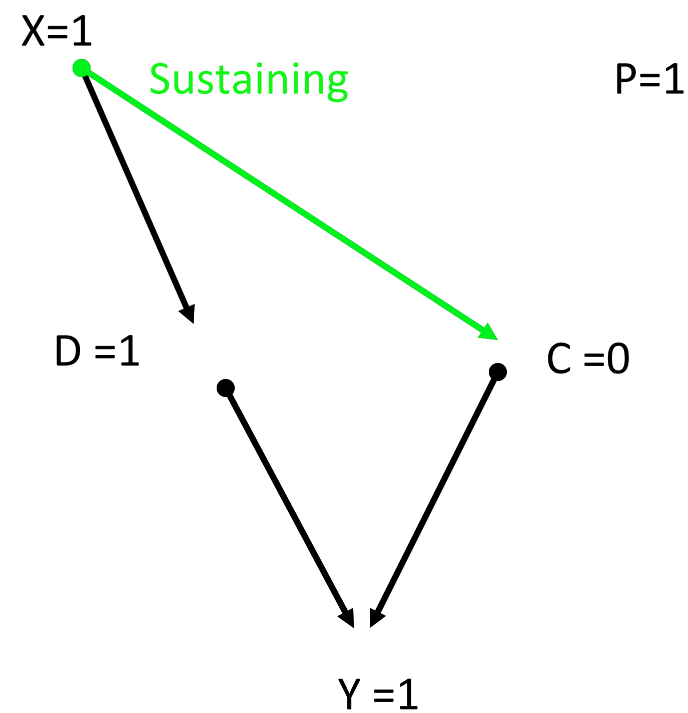


X= shoots canteen, D=dehydration, Y=Death, C=cyanide intake, P=poisons water

Constructing the causal beam

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

- Note: C=¬X

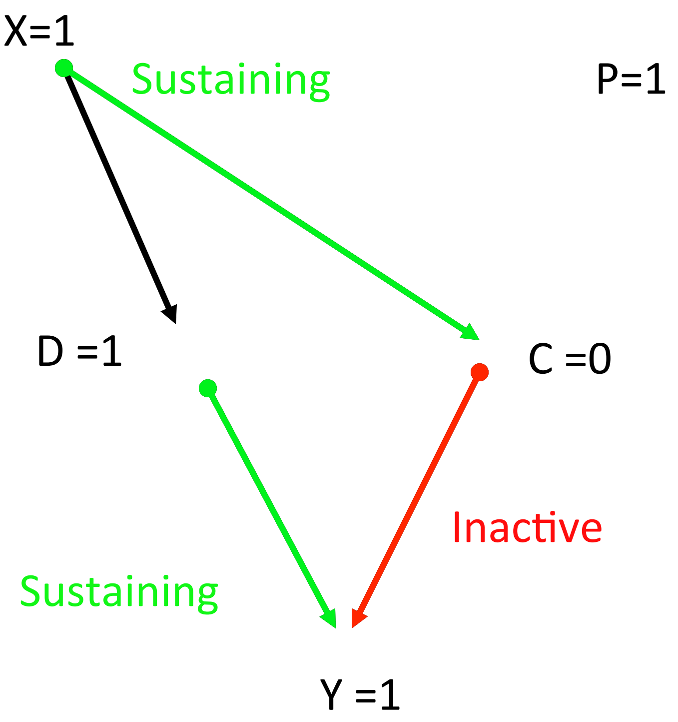


X= shoots canteen, D=dehydration, Y=Death, C=cyanide intake, P=poisons water

Constructing the causal beam

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

- Note: Y=DvC

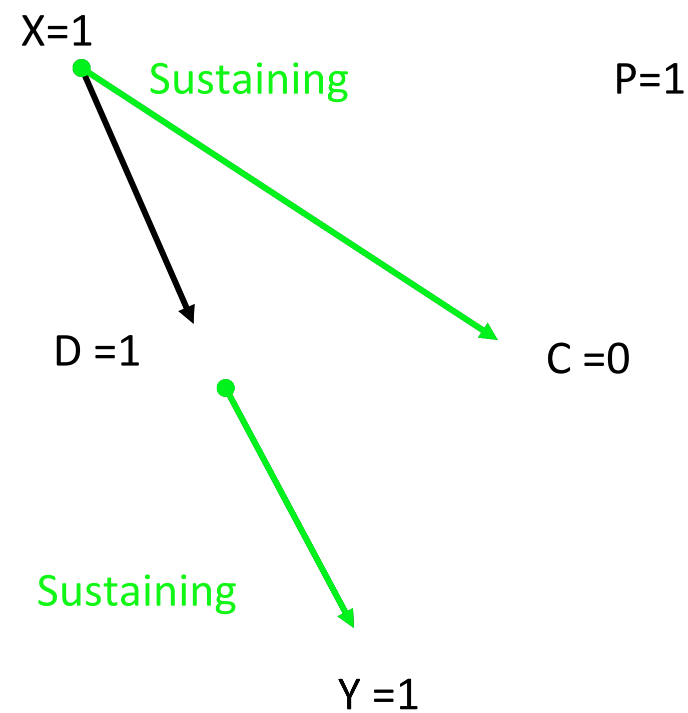


X= shoots canteen, D=dehydration, Y=Death, C=cyanide intake, P=poisons water

Constructing the causal beam

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

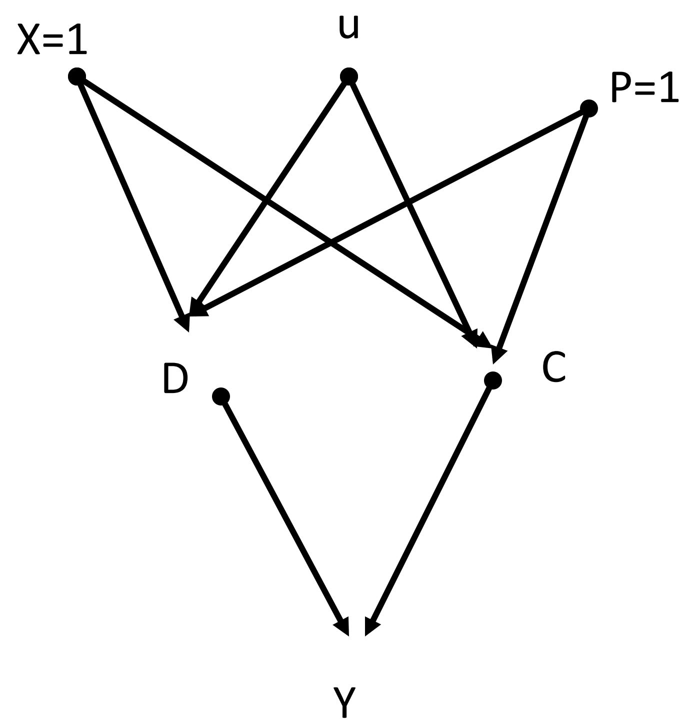
- Finally: Y=D, Y=X



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

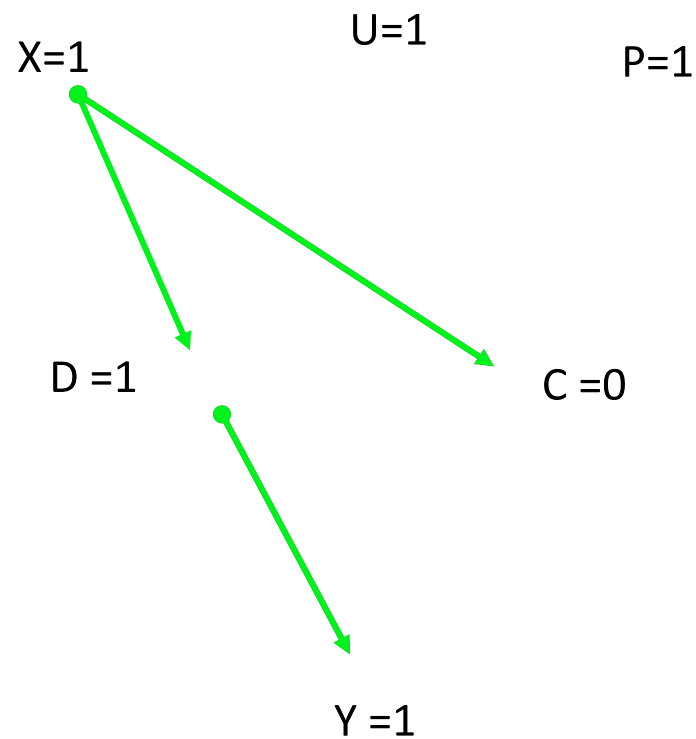
What if we are uncertain?

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



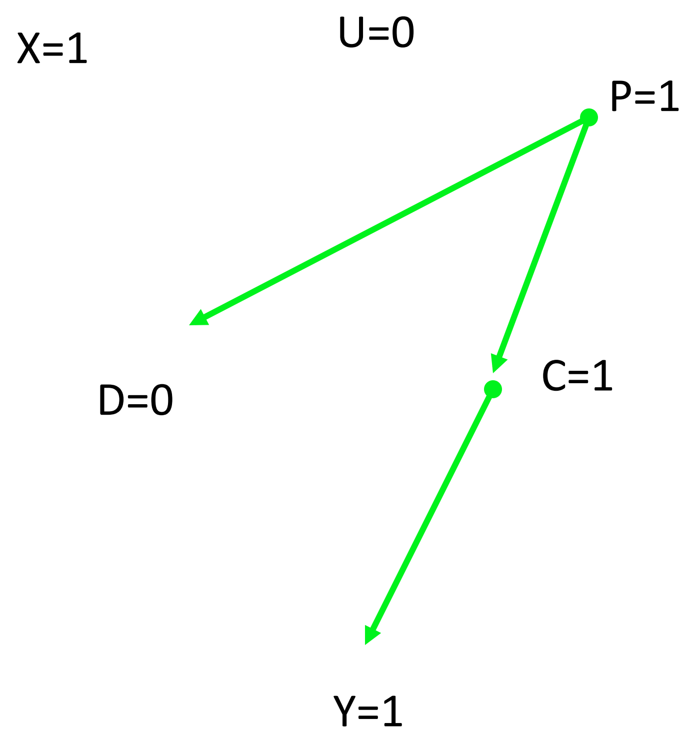
Case 1: u=1

- Canteen emptied before traveler drinks
- Same beam as before



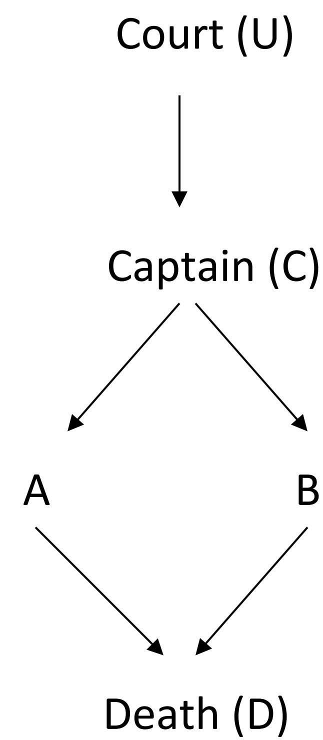
Case 2: u=0

- 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



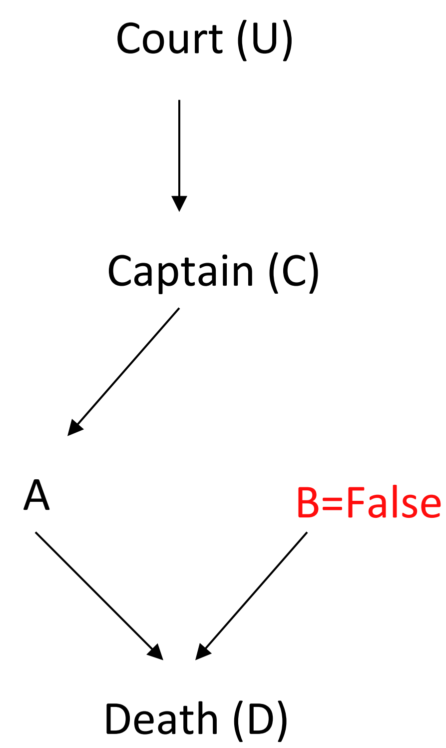
Problem: Over-determination

- 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

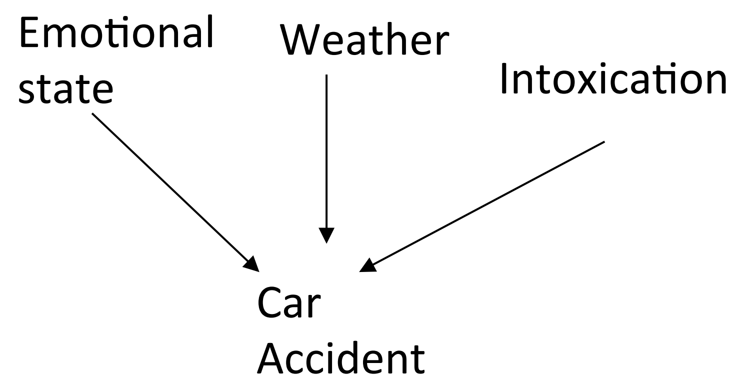


Challenges for the actual cause

- Type !=Token
- Subjectivity
- Timing

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?

$$\begin{tikzcd} \text{Smoking} \arrow[d] \\ \text{Lung cancer} \arrow[d] \\ \text{Death} \end{tikzcd}$$

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