Lecture 8: Causal Inference
CS486/686 Intro to Artificial Intelligence

2023-6-6

Pascal Poupart
David R. Cheriton School of Computer Science

UNIVERSITY OF WATERLOO

Outline

- Models
  - Causal Bayesian Networks
  - Structural Causal Models
- Causal inference
  - Interventions
  - Counterfactuals

Causality

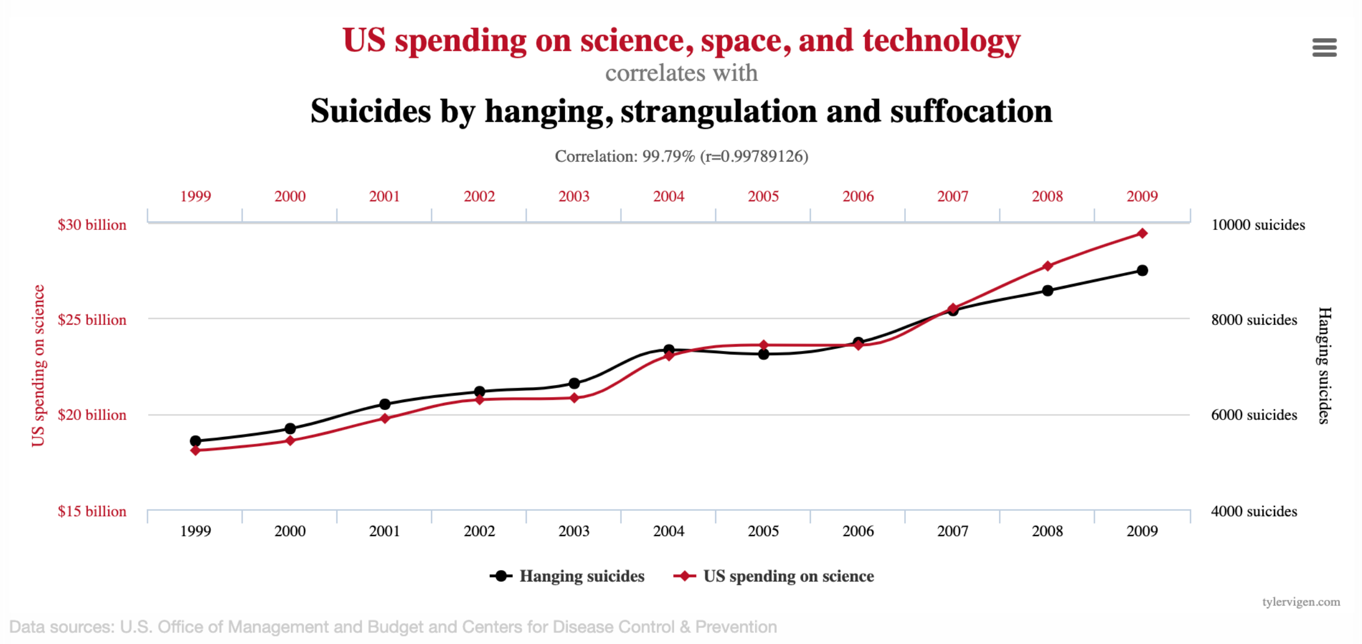
- **Causality** is the study of **how things influence one other**, **how causes lead to effects**.

- **Causal dependence**: X causes Y iff changes to X induce changes to Y
  - Example: Diseases cause symptoms, but symptoms do not cause diseases

Causal and Non-Causal Correlations

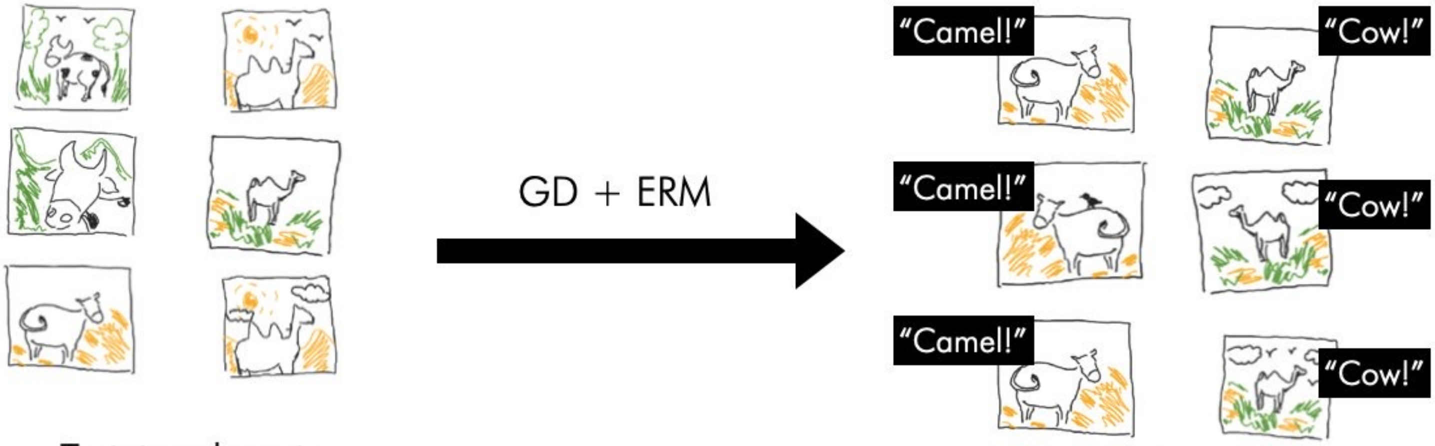
- A **joint distribution** $\boldsymbol{P(X,Y)}$ **captures correlations** between X and Y, but does not indicate whether a causal relation exists between X and Y nor the direction of the causal relation when it exists.
- A **conditional distribution** $\boldsymbol{P(Y|X)}$ **does not necessarily indicate** $\boldsymbol{X}$ **causes** $\boldsymbol{Y}$
  - Recall Bayes' rule: $P(Y|X) = \frac{\Pr(X|Y)P(Y)}{P(X)}$
  - Since we can transform $P(X|Y)$ into $P(Y|X)$, conditional distributions do not always indicate causal dependences, otherwise Y would cause X and X would cause Y

Spurious Correlations



From https://www.tylervigen.com/spurious-correlations

Spurious Correlations



Training domain
Cows: 90% green background
Camels: 90% yellow background

Test domain
Cows: 0% green background
Camels: 0% yellow background

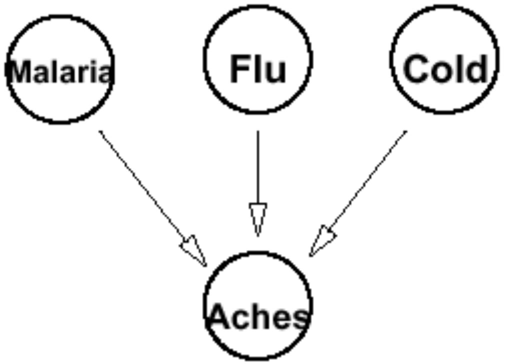
Standard example (Beery et al., '18 Arjovsky et al., '19)

UNIVERSITY OF WATERLOO

Causal Bayesian Network

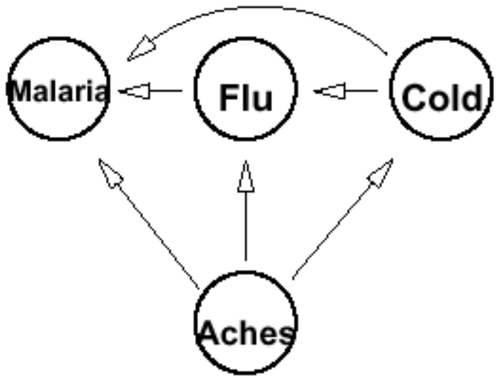
Definition: Bayesian network where all edges indicate direct causal effects.

Causal Bayes net



Probabilistic Inference
Causal Inference

Non-causal Bayes net



Probabilistic Inference

Causal Inference

Intervention: What is the effect of an action?

E.g., What is the effect of a treatment?

Causal networks can easily support intervention queries, but not non-causal networks do not.

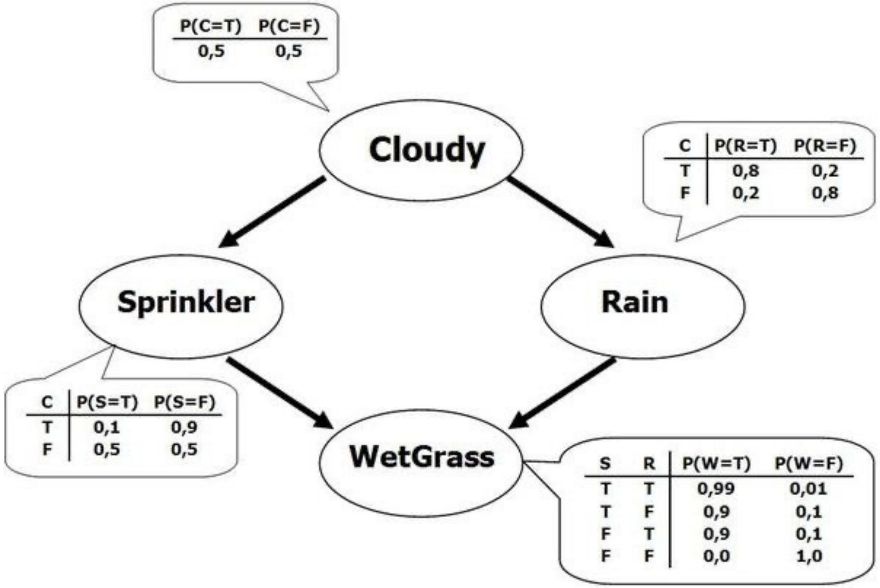
Observation versus Intervention

Observation: What is the likelihood that the grass is wet when the sprinkler is observed to be on?

$$P(WG|S = true)?$$

Intervention: How does turning on the sprinkler affect the grass?

$$P(WG|\boldsymbol{do}(S = true))?$$



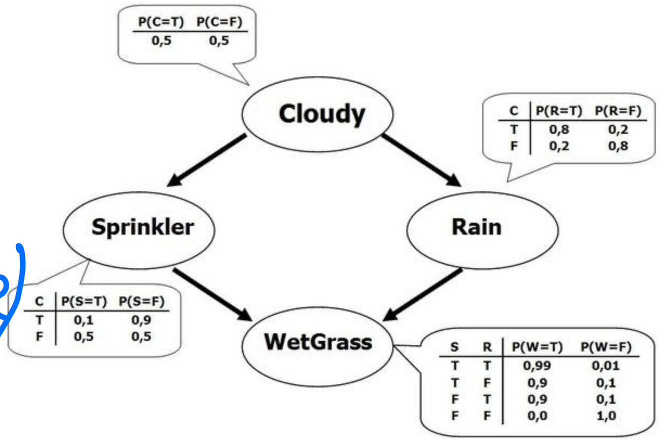
UNIVERSITY OF WATERLOO

Do Operator

Observational query:

$P(WG|S = true)$?

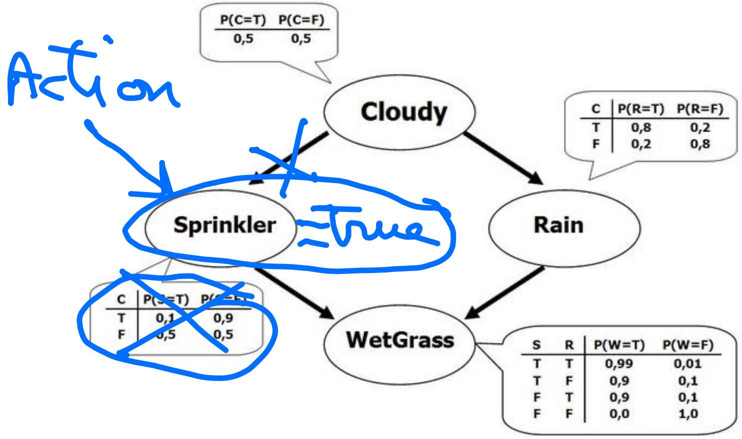
- Factors: $P(C)\,P(R|C)\,P(S|C)\,P(WG|S,R)$
- Evidence: $S = true$
- Eliminate: C, R



Intervention query:

$P(WG|\boldsymbol{do}(S = true))$?

- Factors: $P(C)\,P(R|C)\,P(WG|S,R)$
- Evidence: $S = true$
- Eliminate: C, R



Inference with Do Operator

$$P(X|do(Y = y), Z = z)$$

In a causal graph:

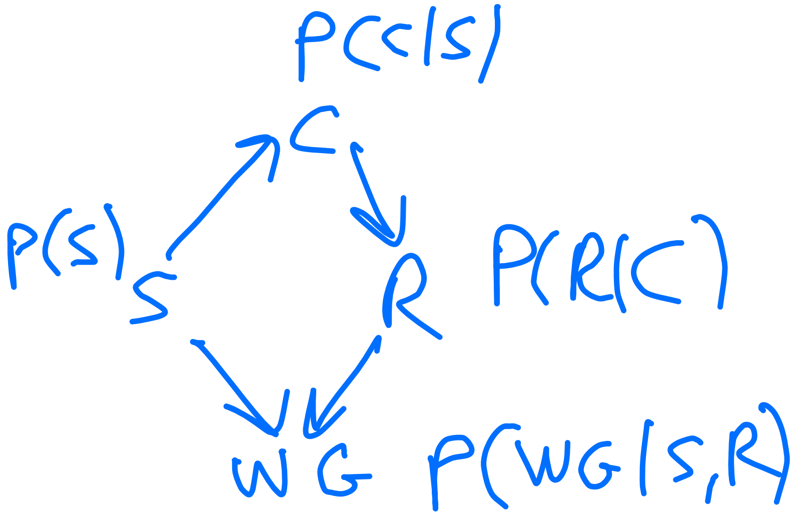
1) Remove edges pointing to Y and $P(Y|parents(Y))$
2) Perform variable elimination on remaining graph:
   a) Restrict factors to evidence: $Y = y$ and $Z = z$
   b) Eliminate variables
   c) Multiply remaining factors and normalize

Non-Causal Graph

Observational query:

$P(WG|S = true)$?

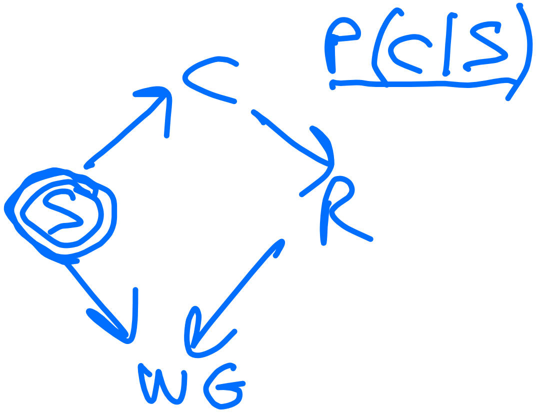
- Factors: $P(S)\,P(C|S)\,P(R|C)\,P(WG|S,R)$
- Evidence: $S = true$
- Eliminate: C, R



Intervention query:

$P(WG|\boldsymbol{do}(S = true))$?

- Factors:
- Evidence:
- Eliminate:



Counterfactual Analysis

Intervention: What is the effect of an action?

E.g., What is the effect of a treatment?

Counterfactual analysis (or counterfactual thinking): explores outcomes that did not actually occur, but which could have occurred under different conditions. It's a kind of what if? analysis and is a useful way for testing cause-and-effect relationships.

E.g., Would the patient have died if he was not treated?

E.g., Would a goal be scored had the player not tripped?

Counterfactual Analysis

How can we answer counterfactual questions with a causal Bayes net?

Treatment → Dead

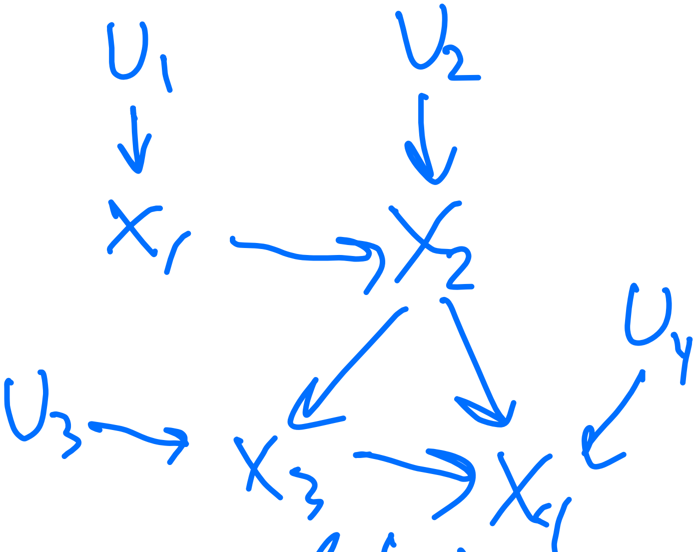
Fact: patient was treated and then died

Counterfactual question: Had the patient not been treated, would the patient have survived?

Can't answer this question since we can't revive the patient to try no treatment…

Structural Causal Models

Idea: separate causal relations from noise



Structural Causal Model contains:

- **X:** endogenous variables (domain variables)
- **U:** exogenous variables (noise)
- Only **deterministic** relations given by equations
  - $X_i = f(parents(X_i), U_i)$

$X_1 = f_1(U_1)$

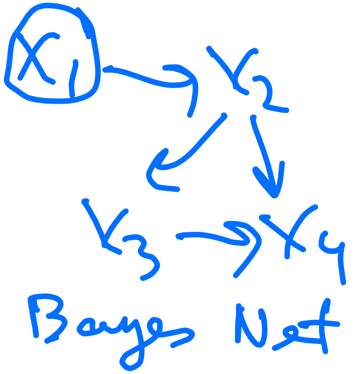
$X_2 = f_2(X_1, U_2)$

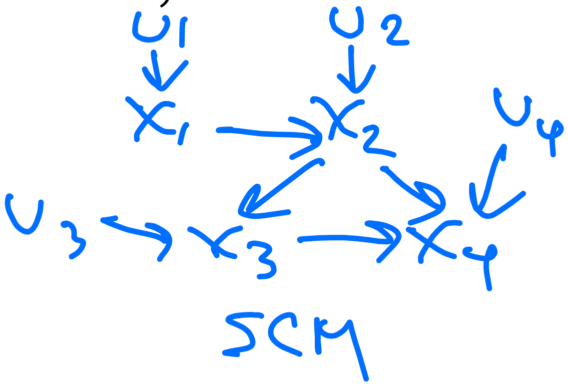
$X_3 = f_3(X_2, U_3)$

$X_4 = f_4(X_3, X_2, U_4)$

Conversion

- Structural Causal Models (SCMs) can be converted into equivalent Causal Bayesian Network, but not the other way around





$$P(x_1) = \sum_{U_1} P(U_1)\, \underbrace{f(U_1)}_{x_1}$$

$$P(x_2|x_1) = \sum_{U_2} P(U_2)\, \underbrace{f(x_1, U_2)}_{x_2}$$

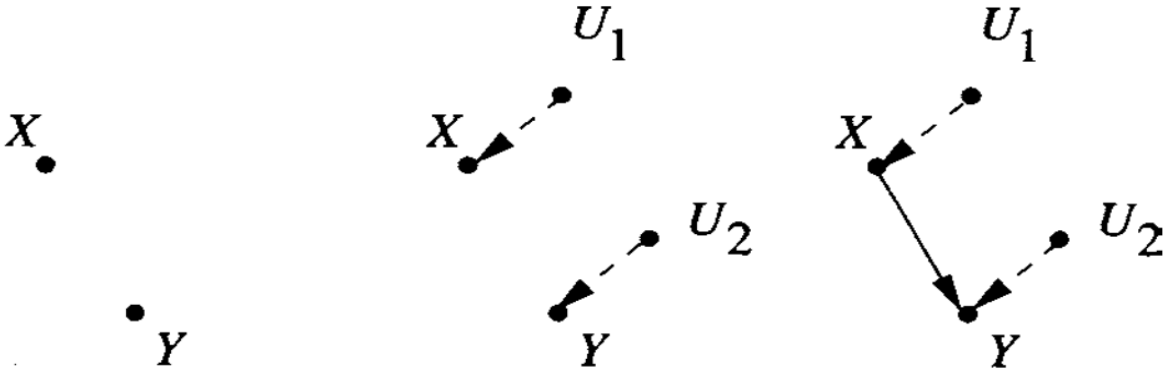
- SCMs separate causal relations from the noise and therefore provide more information

Example

Let $P(X,Y)$ be uniformly distributed i.e., $P(X=x, Y=y) = 0.25\ \ \forall x, y$

X: treatment

Y: dead



(a) (b) (c)

$X = U_1$
$Y = U_2$

$X = U_1$
$Y = XU_2 + (1-X)(1-U_2)$

Example

Model B	$u_2 = 0$		$u_2 = 1$		Marginal	
	$x = 1$	$x = 0$	$x = 1$	$x = 0$	$x = 1$	$x = 0$
$y = 1$ (death)	0	0	0.25	0.25	0.25	0.25
$y = 0$ (recovery)	0.25	0.25	0	0	0.25	0.25

Model C	$u_2 = 0$		$u_2 = 1$		Marginal	
	$x = 1$	$x = 0$	$x = 1$	$x = 0$	$x = 1$	$x = 0$
$y = 1$ (death)	0	0.25	0.25	0	0.25	0.25
$y = 0$ (recovery)	0.25	0	0	0.25	0.25	0.25

Counterfactual Analysis

These three steps can be generalized to any causal model M as follows. Given evidence e, to compute the probability of $Y = y$ under the hypothetical condition $X = x$ (where X is a subset of variables), apply the following three steps to M.

***Step 1* (*abduction*):** Update the probability $P(u)$ to obtain $P(u \mid e)$.

***Step 2* (*action*):** Replace the equations corresponding to variables in set X by the equations $X = x$.

***Step 3* (*prediction*):** Use the modified model to compute the probability of $Y = y$.

Example

Model b:

$$\begin{tikzcd} & U_1 \arrow[dl] \\ X & \\ & U_2 \arrow[dl] \\ Y & \end{tikzcd}$$

Evidence : Treat = true, Dead = true

$P(U_2 = 1 \mid \text{evidence}) = 1$

$Y = U_2 = 1$

Model c:

$$\begin{tikzcd} & U_1 \arrow[dl] & \\ X \arrow[dr] & & U_2 \arrow[dl] \\ & Y & \end{tikzcd}$$

Evidence : Treat = true, Dead = true

$P(U_2 = 1 \mid \text{evidence}) = 1$

$$Y = \underset{0}{X}\,\underset{1}{U_2} + (1-\underset{0}{X})(1-\underset{1}{U_2}) = 0$$

DoWhy Library (Microsoft)

- https://github.com/py-why/dowhy