Causal Inference

Xin Zhang
Peking University

Most of the content is from Chapter 1 of "Causality" second edition by Judea Pearl and "Actual Causality: A Survey" by Joseph Halpern

Recap of Last Lecture

- Probabilistic Logic Programming
  - Logic programming + probabilities
  - Unifying logic and probabilities
    - Logic: Expressiveness
    - Probabilities: Handling uncertainty

Recap of Last Lecture

- Representative language: Problog
  - Problog = Datalog + Probabilities + Additional features

Problog: Example Program

```
0.5 :: stayUp.
0.7 :: drinkCoffee :- stayUp.
0.5 :: drinkCoffee :- \+ stayUp.
0.9 :: fallSleep :- \+ drinkCoffee, stayUp.
0.3 :: fallSleep :- drinkCoffee, stayUp.
0.1 :: fallSleep :- \+stayUp.

evidence(fallSleep).

query(stayUp).
```

Problog: Semantics

- First, ground the program into a Boolean program

- The Boolean program describes a distribution of Datalog program, which in turn defines a distribution of outputs

Semantics of Problog

- Ground

  Constants: 0, 1, 2, 3 4

  path(A,C) :- path(A,B), edge(B,C), r(A,B,C).

  Generates

path(0,0) :- path(0,0), edge(0,0), r(0,0,0).	A=0, B=0, C=0
path(0,1) :- path(0,0), edge(0,1), r(0,0,1).	A=0, B=0, C=1
path(0,1) :- path(0,0), edge(0,1), r(0,0,1).	A=0, B=0, C=1

  …

Semantics of Problog

- From a Problog program, we can sample a Datalog program by sampling the facts

```
0.5 :: stayUp.
0.7 :: drinkCoffee :- stayUp.
0.3 :: fallSleep :- drinkCoffee, stayUp.
```

=

```
0.5 :: stayUp.
0.7 :: r1.
0.3 :: r2.
drinkCoffee :- stayUp, r1.
fallSleep :- drinkCoffee, stayUp, r2.
```

sample →

```
stayUp.
r1.
r2.
drinkCoffee :- stayUp, r1.
fallSleep :- drinkCoffee, stayUp, r2.
```

Probability: 0.5*0.7*0.3

Solving

- Once we have a grounded program, we can leverage existing techniques

- Idea 1: convert the program into a Bayesian network

- Idea 2: convert the program into a Boolean formula with weights (MaxSAT)

Solving: Converting into a MaxSAT

- Finding the most likely solution becomes solving the MaxSAT

- Computing marginal probabilities becomes weighted model counting

This Class

- Causal inference
  - Structural equation model (Pearl)
  - Causal inference in probabilistic programming
  - Actual causality
- Not causal discovery
  - Assume we have a model
  - How to use the model to represent causality
  - How to reason with the model

Motivating Example

- If a person has long hair, they are likely to be a girl

- If we change a boy from short hair to long air, would he become a girl?

Intervention

Question

- Can we separate causality from correlation without intervention?

Motivating Example

- Xiaoming was late for the lecture. Would he still be late for the lecture if he had got up at 6am?

Counterfactual

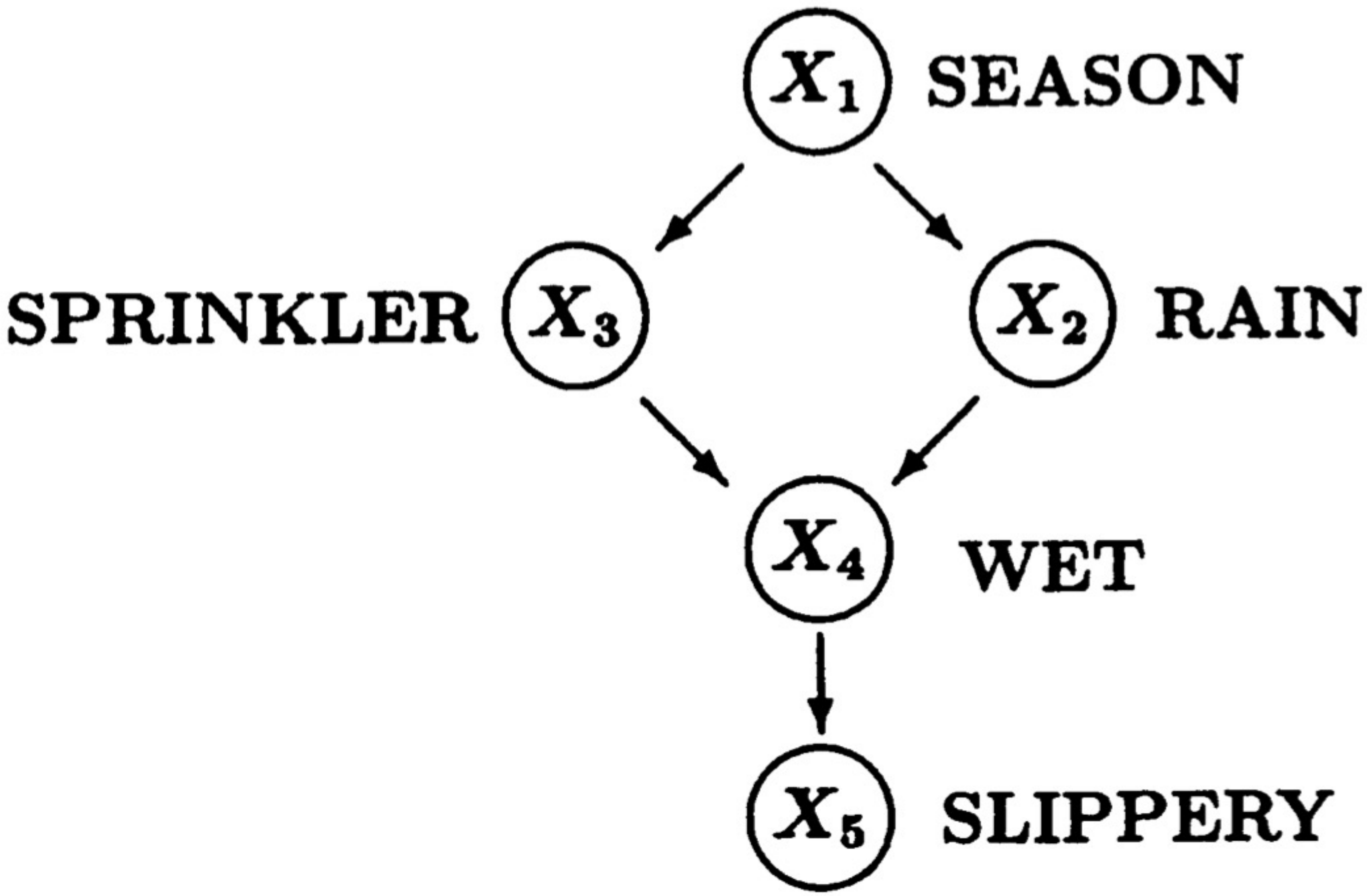
Pearl's Causal Hierarchy

- L1: Predictions: What if I observe … ?
- L2: Interventions: What if I change … ?
- L3: Counterfactuals: What if we did … given … ?

What models can be used to answer these questions?

Causal Bayesian Network: Handling Interventions

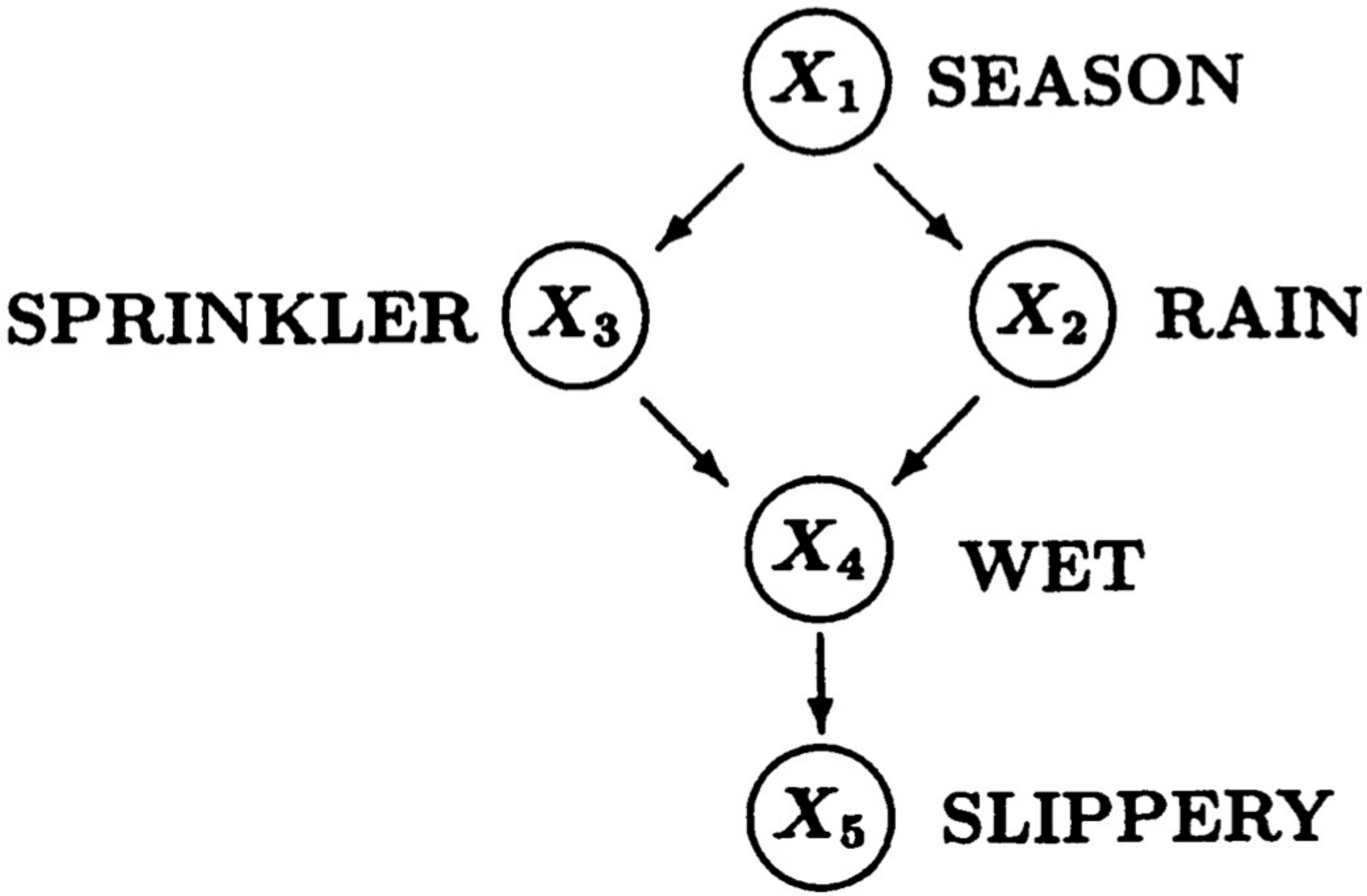
What is the joined probability distribution if we observe the sprinkler is on?



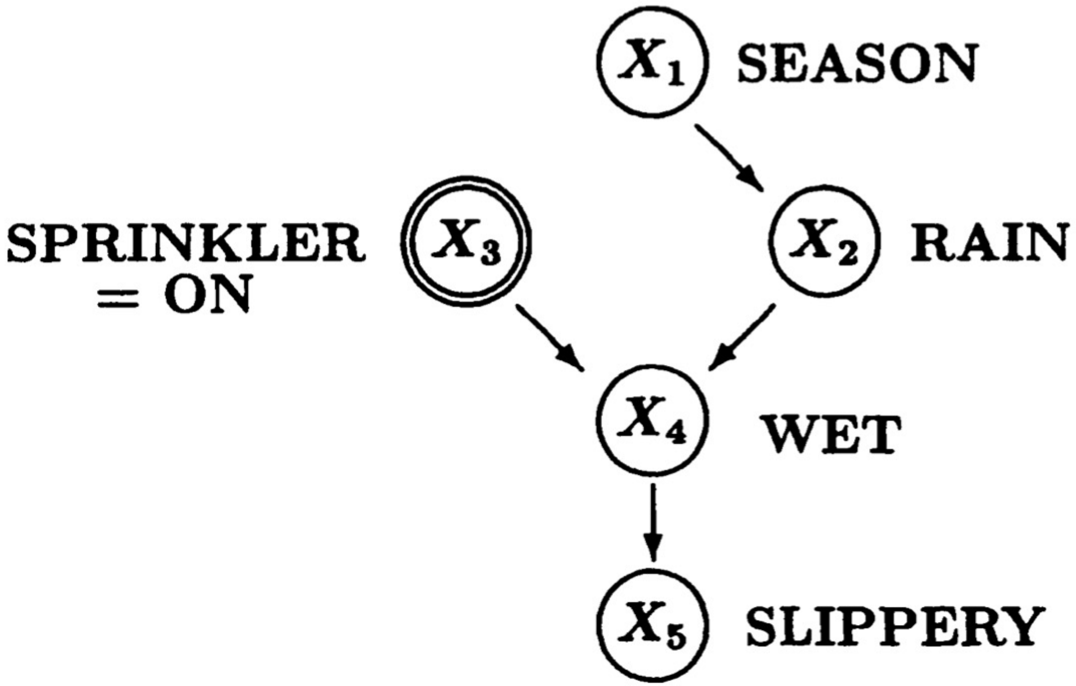
Causal Bayesian Network: Handling Interventions

What is the joint probability if we intervene on the sprinkler by turning it on?

do(X_3=On)



Causal Bayesian Network: Handling Interventions



$$P_{X_3=\mathrm{On}}(x_1, x_2, x_4, x_5) = P(x_1)\, P(x_2 \mid x_1)\, P(x_4 \mid x_2, X_3 = \mathrm{On}) P(x_5 \mid x_4),$$

Causal Bayesian Network: Handling Interventions

Definition 1.3.1 (Causal Bayesian Network) Let $P(v)$ be a probability distribution on a set V of variables, and let $P_x(v)$ denote the distribution resulting from the intervention $do(X = x)$ that sets a subset X of variables to constants x. Denote by $\boldsymbol{P}_*$ the set of all interventional distributions $P_x(v)$, $X \subseteq V$, including $P(v)$, which represents no intervention (i.e., $X = \emptyset$). A DAG G is said to be a **causal Bayesian network** compatible with $\boldsymbol{P}_*$ if and only if the following three conditions hold for every $P_x \in \boldsymbol{P}_*$:

Causal Bayesian Network: Handling Interventions

(i) $P_x(v)$ is Markov relative to G;

Conditional Independence

(ii) $P_x(v_i) = 1$ for all $V_i \in X$ whenever v_i is consistent with $X = x$;

(iii) $P_x(v_i | pa_i) = P(v_i | pa_i)$ for all $V_i \notin X$ whenever pa_i is consistent with $X = x$.

Defining Effects of Interventions

The distribution $P_x(v)$ resulting from the intervention $do(X = x)$ is given as a **truncated-factorization**

$$P_x(v) = \prod_{\{i|V_i \notin X\}} P(v_i|pa_i) \text{ for all } v \text{ consistent with } x, \quad \textbf{(1.37)}$$

Defining Effects of Interventions

- On the graph:
  - Cut the connections from the parents to the intervened nodes
  - Set the intervened nodes to the corresponding values

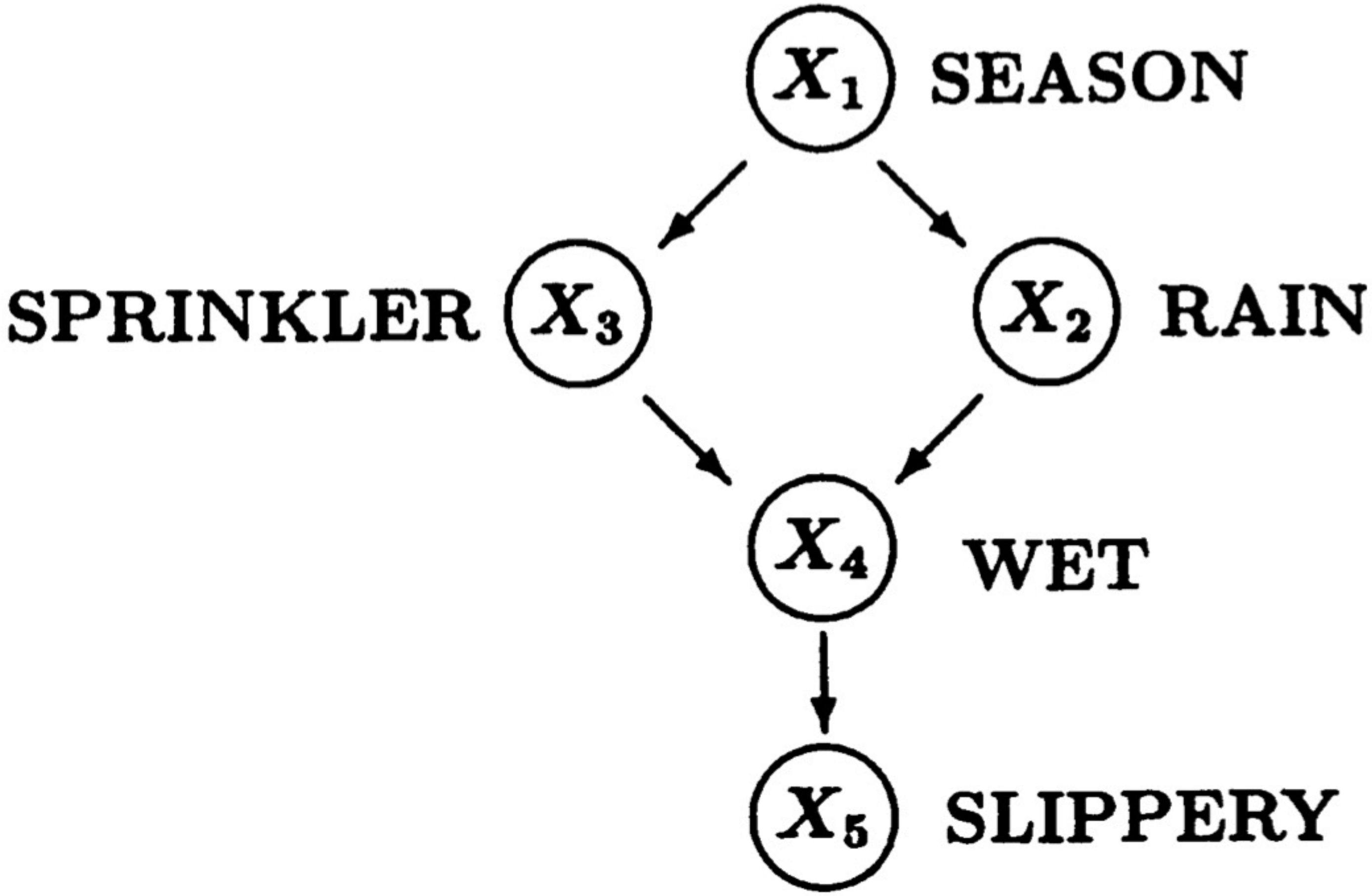
Advantages of Using a Graphical Model

- Modular

- Can use tools like d-separation to reason about the impact of interventions

What About Counterfactuals?

Given the grass is slippery, will it still be slippery if we had turned off the sprinkler?



Structural Equation (Functional) Model

- Functional causal model
  - Can answer all three questions

- Expressed using deterministic functional equations
  - Probabilities are introduced by assuming certain variables are unobserved
  - Follows Laplace's conception of natural phenomena

- Advantages over stochastic representations
  - More general
  - More in tune with human intuition
  - Counterfactuals

Structural Equations

- A functional causal model consists a set of equations:

$$x_i = f_i(\underline{pa_i}, \underline{u_i}), \quad i = 1, \ldots, n,$$

parents

Errors due to omitted factors. Random.

Structural Equations: Example I

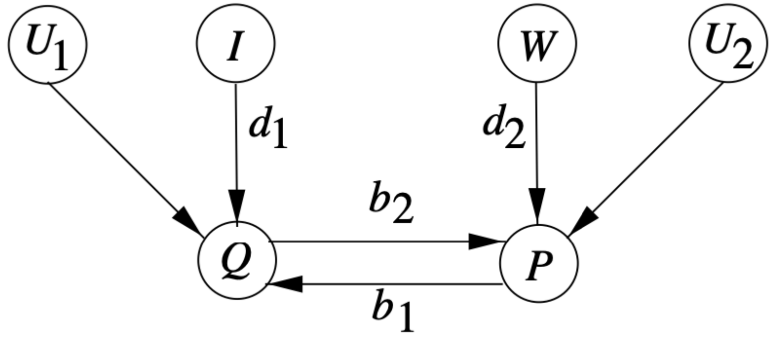


Figure 1.5: Causal diagram illustrating the relationship between price (P), demand (Q), income (Z), and wages (W).

$$q = b_1 p + d_1 i + u_1, \quad \textbf{(1.42)}$$
$$p = b_2 q + d_2 w + u_2, \quad \textbf{(1.43)}$$

Structural Equations: Example II

Explicitly separate deterministic parts from the stochastic parts

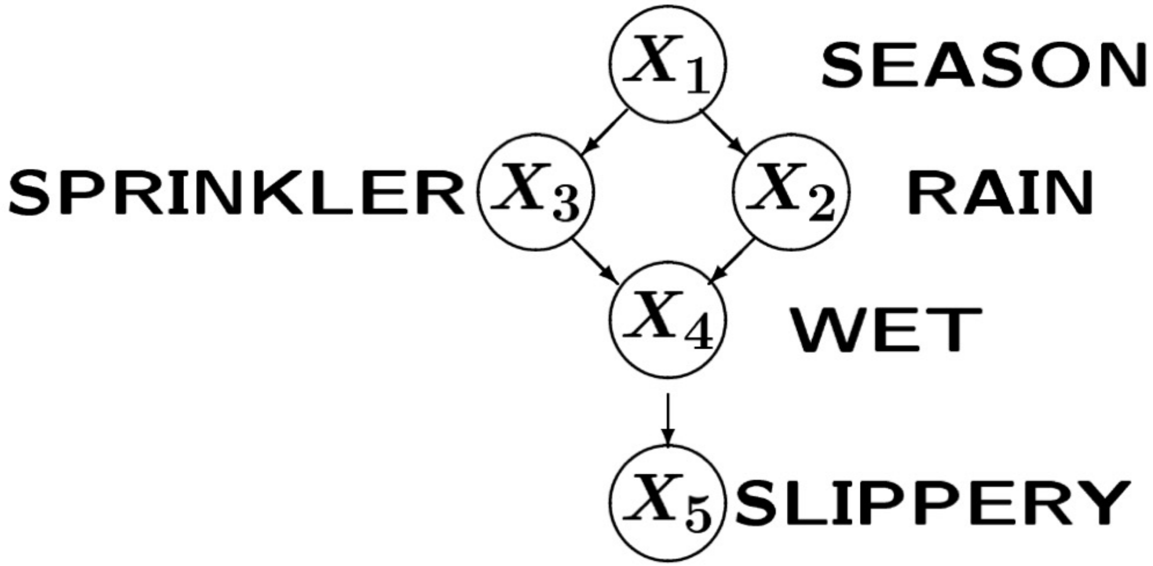


Figure 1.2

$$
\begin{aligned}
x_1 &= u_1, \\
x_2 &= f_2(x_1, u_2), \\
x_3 &= f_3(x_1, u_3), \\
x_4 &= f_4(x_3, x_2, u_4), \\
x_5 &= f_5(x_4, u_5).
\end{aligned}
$$

$$
\begin{aligned}
x_2 &= [(X_1 = \text{winter}) \vee (X_1 = \text{fall}) \vee u_2] \wedge \neg u'_2, \\
x_3 &= [(X_1 = \text{summer}) \vee (X_1 = \text{spring}) \vee u_3] \wedge \neg u'_3, \\
x_4 &= (x_2 \vee x_3 \vee u_4) \wedge \neg u'_4, \\
x_5 &= (x_4 \vee u_5) \wedge \neg u'_5, \qquad \textbf{(1.45)}
\end{aligned}
$$

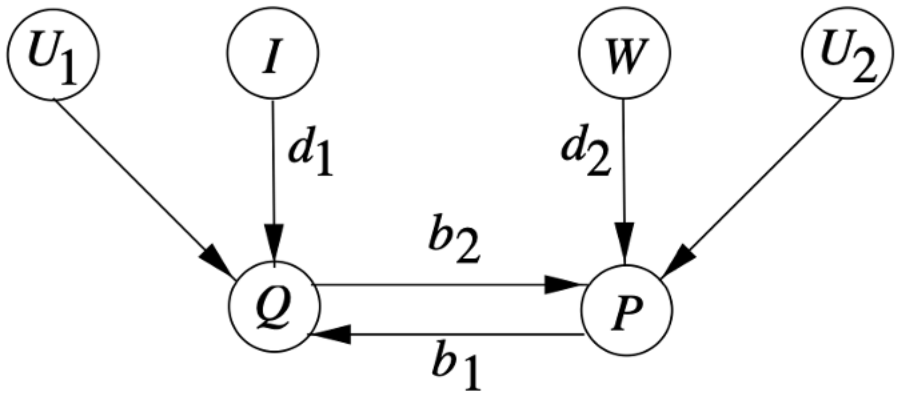
Goal: Handle the Whole Pearl's Causal Hierarchy

- L1: Predictions: What if I observe … ?
- L2: Interventions: What if I change … ?
- L3: Counterfactuals: What if we did … given … ?

What models can be used to answer these questions?

Probabilistic Predictions in Causal Models

- Causal diagram:



- Semi-Markovian model: the diagram is acyclic
- Markovian model: the diagram is acyclic and the errors are independent

The Causal Markov Condition

Theorem 1.4.1 (Causal Markov Condition)
Every Markovian causal model M induces a distribution $P(x_1, \ldots, x_n)$ that satisfies the parental Markov condition relative the causal diagram G associated with M; that is, each variable X_i is independent on all its non-descendants, given its parents PA_i in G (Pearl and Verma 1991)

The Causal Markov Condition Follows two Causal Assumptions

- Include every variable that is the cause of two or more variables in the model (not in the background)

- Reichenbach's common-cause assumption
  - No correlation without causation
  - If any two variables are dependent, then one is the cause of the other or there is a third variable causing both (confounder)

Interventions and Causal Effects in Functional Models

- Simply modify the corresponding equations

$$x_3 = f(x_1, u_3) \rightarrow x_3 = On$$

  - More formally: fix the intervened variables to their specified values, and removing equations defining them

- Intervening on a causal Markovian model is the same as intervening on a causal Bayesian network

Advantages Over Causal Bayesian Networks

- Extensions to feedback systems and non-Markovian models
- Modifications of parameters are meaningful
  - Functions generate the joint distribution, conditional probabilities are then inferred
- Simplifying the analysis of causal effects
- Permit the analysis of context-specific actions and policies

Last Point Explained

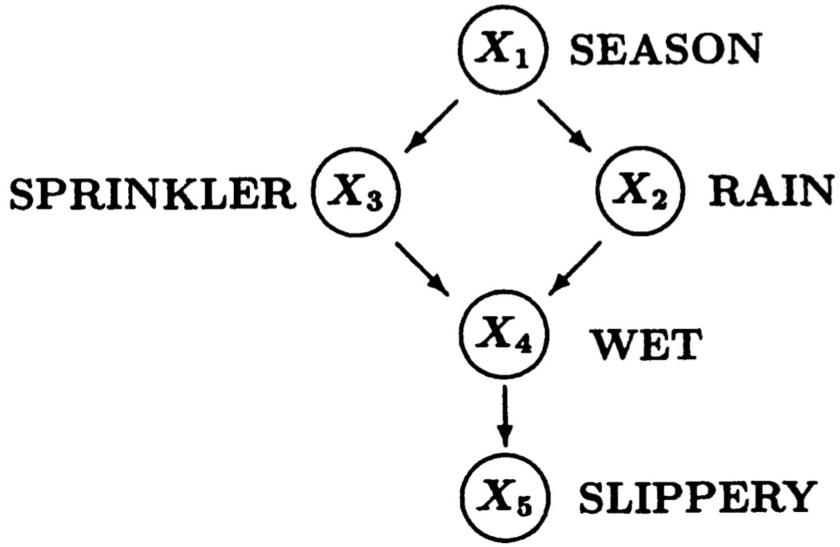
- Interventions affect contexts
  - Example: the patient has been examined by the doctor and he has some symptoms, but now the new intervention will affect these symptoms

- We will see that counterfactuals are similar

Counterfactuals in Functional Models

- Causal Bayesian networks have trouble dealing with counterfactuals
  - The simplest example:
    - Consider two independent boolean variables x and y, we have p(x|y) = 0.5, given y = 1, what is P(y = 1 | do(x)= 0, y =1)?
  - A more complex example:

do(x_3=ON) , X_5=True



Understand Counterfactuals Better

- Counterfactuals can be seen as the combination of conditioning and interventions:
    - Use observations to infer the posterior distributions of the hidden variables
    - Based on the posterior distributions, predict under interventions

Three Steps for Computing

For computing $P(Y= y \mid do(X = x), e)$:

1. (abduction): Update the probability $P(u)$ to obtain $P(u|e)$
2. (action): Perform intervention $do(X) = x$
3. (prediction) Use the modified model to compute $P(Y=y)$

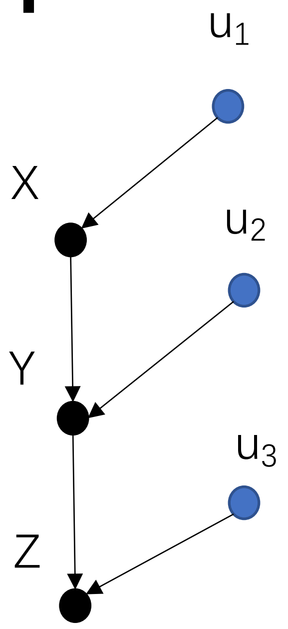
More on Computing Counterfactuals

- A major difficulty of the previous approach is the need to compute and store $p(u|e)$

- Can we overcome this problem by leveraging algorithms in graphical models?

The Twin Network Approach

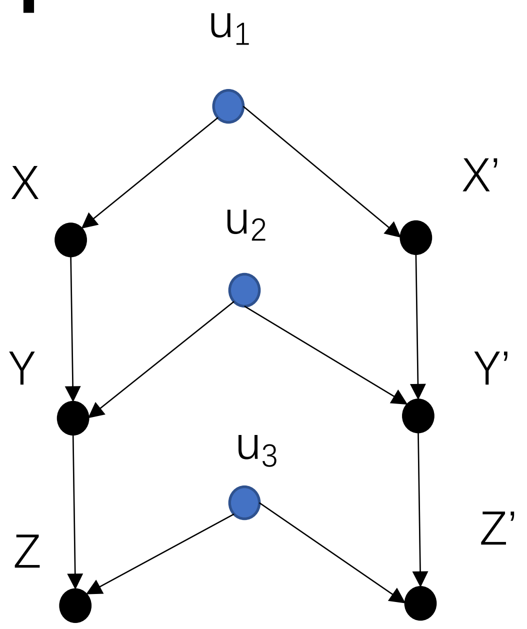
- Consider the following example
  - $X = u_1, Y = X + u_2, Z = Y + u_3$



- How to compute P(Z|do(X) = x, Z=z)?

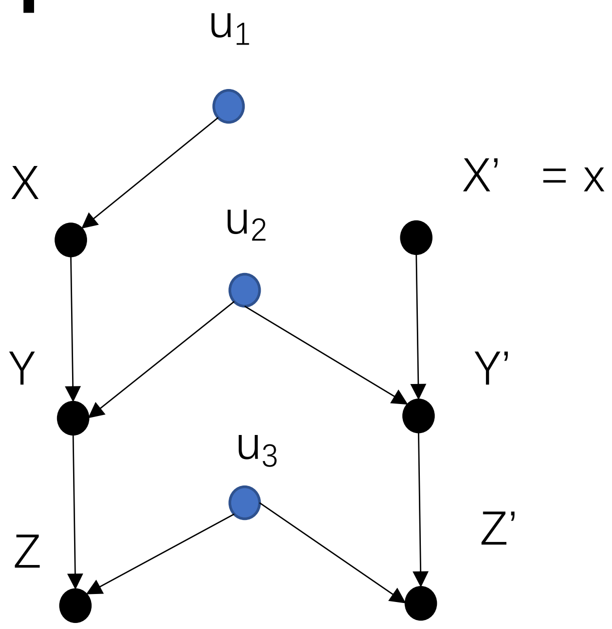
The Twin Network Approach

- P(Z | do(X) = x, Z=z) becomes P(Z' | do(X') = x, Z=z)



The Twin Network Approach

- P(Z|do(X) = x, Z=z) becomes P(Z'|X' = x, Z=z)



The Twin Network Approach

- Duplicate all the equations and observed variables
- Perform intervention on the copied part
- Keep observations on the original part

Can you apply the twin network approach to
causal Bayesian networks?

Two Mainstream Causal Models

- Structural equation model (Pearl)
  - This class
- Potential outcomes (Neyman-Rubin)
- Two models are theoretically equivalent, but have their own advantages in practice

Causal inference in probabilistic programming

- A Language for Counterfactual Generative Models. Zenna Tavares, James Koppel, Xin Zhang, Ria Das, Armando Solar-Lezama. ICML 2021

- Implicitly implements the twin network approach
  - Lazy evaluation
  - Stores the program piece that computes a given variable

Actual Causality

- Interventions and counterfactuals basically tells how a things changes in response how another thing changes

- But it doesn't define what is the cause/reason of something.

- Causality answers this

Some Heads-Up

- Two notions of causality
  - Type (general) causality: smoking causes lung cancer
  - Actual causality: the fact that David smoked like a chimney for 30 years cased him to get cancer last year

- Actual causality is a long-debated problem in philosophy, math, and computer science

- We are not going to include philosophical discussions
  - No chicken-or-egg problems

- We assume there is a known model of the world and discuss how to define actual causalities according to it
  - Causes can be different if the modeling is different

The Big Picture on Actual Causalities

- The definition has changed many times
- No satisfying answers
- The new definitions are usually invented in response to counterexample

The Big Picture on Actual Causalities

- Attempts to define causality goes back to Aristotle
- Relatively recent trend (Lewis 1973) is to use counterfactuals
- More recent: capture counterfactuals using structural equations
- Pearl & Halpern definitions:
  - UAI 2001
  - BJPS 2005

But-For Causes

- Jimmy threw a ball to shatter the bottle
  - JimmyThrows = u_1
  - BottleShatters = JimmyThrows

- If Jimmy doesn't throw the ball, the bottle won't shatter
  - Therefore Jimmy throwing the ball is the cause for the bottle to shatter

But-For Causes

- Counter-example (preemption): Suzy and Jimmy both pick up rocks and throw them at a bottle. Suzy's rock gets there first, shattering the bottle. Since both throws are perfectly accurate, Billy's would have shattered the bottle if Suzy's throw had not preempted it

- JimmyThrows = u1, SuzyThrows = u2,

  SuzyShatters = SuzyThrows,

  JimmyShatters = JimmyThrows & !SuzyShatters,

  BottleShatters = SuzyShatters | JimmyShatters

Pearl and Halpern's: Problem Setting

- Represent the model using structural equations
- Remove all randomness by fixing the unobserved variables
  - In other words, the causes are defined for specific contexts
- The cause can be any conjunction of primitive events
- Arbitrary Boolean combinations of primitive events can be caused

Pearl and Halpern's Definition

- $\vec{X} = \vec{x}$ is an actual cause of ϕ in situation $(M, \vec{u})$ if

- AC1.$(M, \vec{u}) \vDash \left(\vec{X} = \vec{x}\right) \wedge \phi$
  - Both $\left(\vec{X} = \vec{x}\right)$ and ϕ are true in the actual world

- AC2. Complicated. Captures counterfactuals

- AC3. $\vec{X}$ is minimal; no subset of $\vec{X}$ satisfies AC1 and AC2.
  - No irrelevant conjuncts

Pearl and Halpern's Definition

- AC2. There is a set of $\overrightarrow{W}$ of variables in V and a setting $\vec{x}'$ of the variables in $\vec{X}$ such that if $(M, \vec{u}) \models (\overrightarrow{W} = \vec{w})$, then

$$(M, \vec{u}) \models \left(\vec{X} \leftarrow \overrightarrow{x'}, \overrightarrow{W} \rightarrow \vec{w}\right) \wedge \neg\phi$$

In words: keeping the variables in $\overrightarrow{W}$ fixed at their actual values, changing $\vec{X}$ can change the outcome ϕ

Example

- JimmyThrows = u1, SuzyThrows = u2,
  SuzyShatters = SuzyThrows,
  JimmyShatters = JimmyThrows & !SuzyShatters,
  BottleShatters = SuzyShatters | JimmyShatters

Let $\vec{X} = \{SuzyThrows\}, \overrightarrow{W} = \{\text{JimmyShatters}\}, \phi = BottleShatters$,
then $(M, \vec{u}) \vDash \left(\vec{X} \leftarrow \vec{x}, \overrightarrow{W} \rightarrow \overrightarrow{w}\right) \wedge \neg\phi$

Another Example

- Suppose in an election, Jim will be elected if two of the three voters vote for him.
- None of the voters voted for Jim. What is a cause of Jim not being elected?
- For more, watch https://www.youtube.com/watch?v=hXnCX2pJ0sg