CompSci 590.6

Understanding Data:

Theory and Applications

Lecture 15

Causality in AI

Instructor: Sudeepa Roy

Email: *sudeepa@cs.duke.edu*

Today's Reading

1. Causes and Explanations: A Structural-Model Approach – Part I: Causes
   - Halpern-Pearl, UAI 2001
2. Causes and Explanations: A Structural-Model Approach – Part Ii: Explanations
   - Halpern-Pearl, IJCAI 2001
3. Probabilities of Causation: Three Counterfactual Interpretations and their Identification
   - Pearl 1999, UCLA Technical Report R-260
   - A revised version in Tian-Pearl, 2000, UCLA Technical Report R-271-A

- Also see numerous surveys/talks by Pearl
  - An excellent survey of history of causality can be found in his lecture slides (with transcripts): 1996 Faculty Research Lecture: "The Art and Science of Cause and Effect"

Acknowledgement:

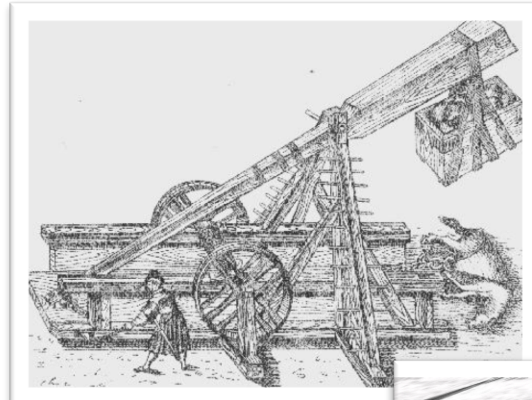
Some of the following slides are from the tutorial “Causality and Explanations in Databases”, VLDB 2014, Meliou-Roy-Suciu, and are due to Dr. Alexandra Meliou, University of Massachusetts-Amherst.

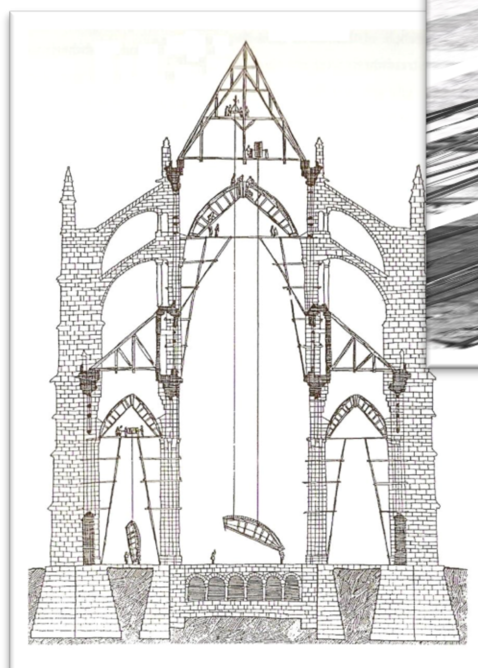
Why Care About Causality?

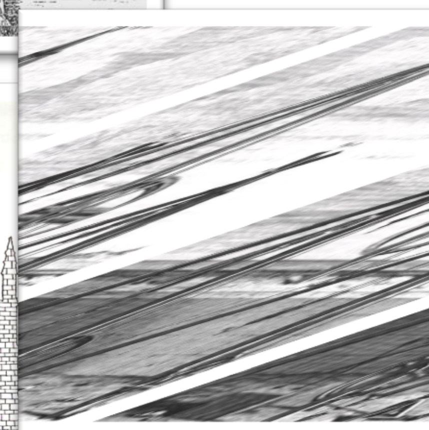
- Databases
  - Why is my program taking so long?
  - Why do/don't I see this result?
- Education
  - Does this special program in a high school encourage higher studies?
- Politics
  - Why did X win the presidential election in year Y?
- Healthcare/Medicine
  - Why did patient X survive cancer while patient Y did not?
  - Does this drug help cure melanoma?
- Agriculture
  - Does this fertilizer help increase plant growth?
- Social science
  - Does the new housing initiative encourage a population buy houses?

Causality in science

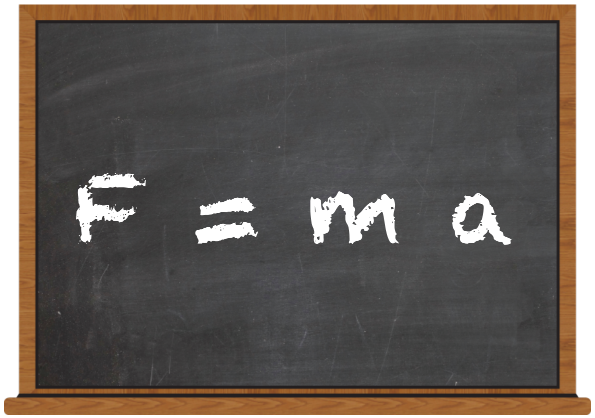
- Science seeks to understand and explain physical observations
  - *Why* doesn’t the wheel turn?
  - *What if* I make the beam half as thick, will it carry the load?
  - *How* do I shape the beam so it will carry the load?

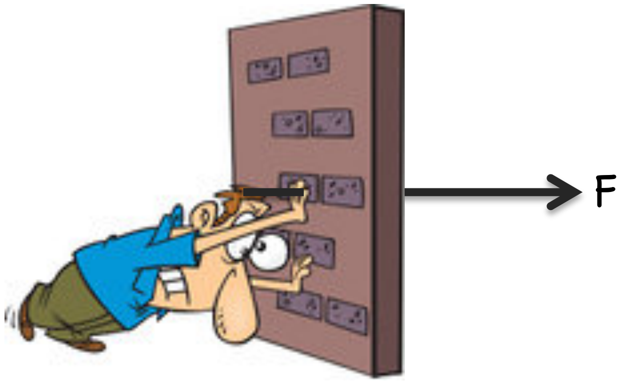






But laws in science do not tell us about causality





- Does acceleration cause the force?
- Does the force cause the acceleration?
- Does the force cause the mass?

Today: Causality in AI

- Pearl's causality model

- Next two lectures
  - Adoption of Pearl's model for causality in databases
  - Causality in Statistics (Rubin's potential outcome model)

Pearl'1996, 1999

Hume's Legacy

David Hume
(1711-1776)

Scottish Philosopher

The author of
"A Treatise of
Human Nature" (1738)

- Analytical vs. Empirical Claims
  - Analytical = product of thoughts
  - Empirical = matter of facts
- Causal claims are empirical
  - human experience is the source
  - "*Thus we remember to have seen that species of object we call Flame, and to have felt that species of sensation we call Heat. We likewise call to mind their constant conjunction in all past instances. Without any farther ceremony, we call the one Cause and the other Effect, and infer the existence of the one from that of the other.*"
- Leads to two riddles of causation

Pearl'1996, 1999

Two Riddles of Causation

1. Learning of causal connection

   What empirical evidence legitimizes a cause-effect connection?

   – How do people ever acquire knowledge of causation
   – e.g. does a rooster cause the sun to rise?
   – succession, correlations are not sufficient
   – e.g. roosters crow before dawn, both ice cream sales and crime rate increase at the same time (in summer months)

2. Usage of causal connection

   What inferences can be drawn from causal information and how?

   – e.g. what would change if the rooster were to cause the sun to rise, can we make the night shorter by waking him up early?

- Major focus of Pearl's work is in (2)
  – More of (1) in Rubin's model

Main Concepts in Pearl's model

Concept	Formalization
Causation	Encoding of behavior under intervention
Intervention	Surgeries on mechanisms
Mechanisms	Functional relationships by equations and graphs

- Devise a computational scheme for causality to facilitate prediction of the effects of "actions"
- Use "Intervention" for "Action"
  - as actions are external entities originating "outside" the theory

Main Concepts in Pearl's model

Concept	Formalization
Causation	Encoding of behavior under intervention
Intervention	Surgeries on mechanisms
Mechanisms	Functional relationships by equations and graphs

Mechanism:

- Autonomous physical laws or mechanisms of interest
  - we can change one without changing the others
  - e.g. logic gates of a circuit, mechanical linkages in a machine

Main Concepts in Pearl's model

Concept	Formalization
Causation	Encoding of behavior under intervention
Intervention	Surgeries on mechanisms
Mechanisms	Functional relationships by equations and graphs

Intervention

- Breakdown of a mechanism = surgery

Main Concepts in Pearl's model

Concept	Formalization
Causation	Encoding of behavior under intervention
Intervention	Surgeries on mechanisms
Mechanisms	Functional relationships by equations and graphs

Causality

- Which mechanism is to be surgically modified by a given action
- Q. Why are these non-trivial?
- A. A number of factors to take into account

Pearl’1996, 1999

Example-1

- If the grass is wet, then it rained
- If we break this bottle, the grass gets wet

Conclusion

- If we break this bottle, then it rained (!)

Lin'1995
Pearl'1999

Example-2

- A suitcase will open iff both locks are open
- The right lock is open
- What happens if we open the left lock?
- Not sure – the right lock might get closed (!)

Overview: Causal Model

- Action sentences
  - B (would be true) if we do A
- Counterfactuals
  - ¬B would change to B (B would be different) if it were A
- Explanation
  - B occurred because of A

Overview: Causal Model Extension

- Action sentences
  – B if we do A With probability p
- Counterfactuals
  – ¬B would change to B if it were A With probability p
- Explanation
  – B occurred because of A With probability p

Pearl’s Model: At a glance

- Modeling causality
  - Causal networks and structural equations
- Reasoning about causality
  - Counterfactual causes
  - Actual causes (Halpern & Pearl)
- Measuring causality
  - Responsibility
  - Probability of necessity, Probability of sufficiency

Modeling Causality

Causal Model M = (U, V, F)

- U = Exogenous variables
  - Values are determined by factors outside the model
- V = Endogenous variables
  - values are described by structural equations
- F is a set of structural equations $\{F_X \mid X \text{ in } V\}$ (endogenous)
  - F_X is a mapping, tells us the value of X given the values of all the other variables in U and V
  - represents a mechanism or law in the world

Example

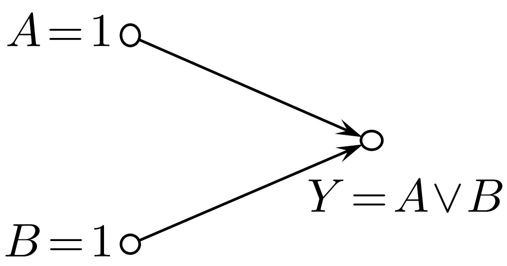
A forest fire could be caused by either lightning or a lit match by an arsonist

- Endogenous variables (Boolean)
  - F for fire
  - L for lightning
  - ML for match lit
- Exogenous variables U
  - whether the wood is dry
  - whether there is enough oxygen in the air
- F_F (U, L, ML) is such that F = 1 if L= 1 or ML = 1

Pearl' 2000

Structural Equations as Causal networks

- Causal structural models:
  - Variables: A, B, Y
  - Structural equations: Y = A v B

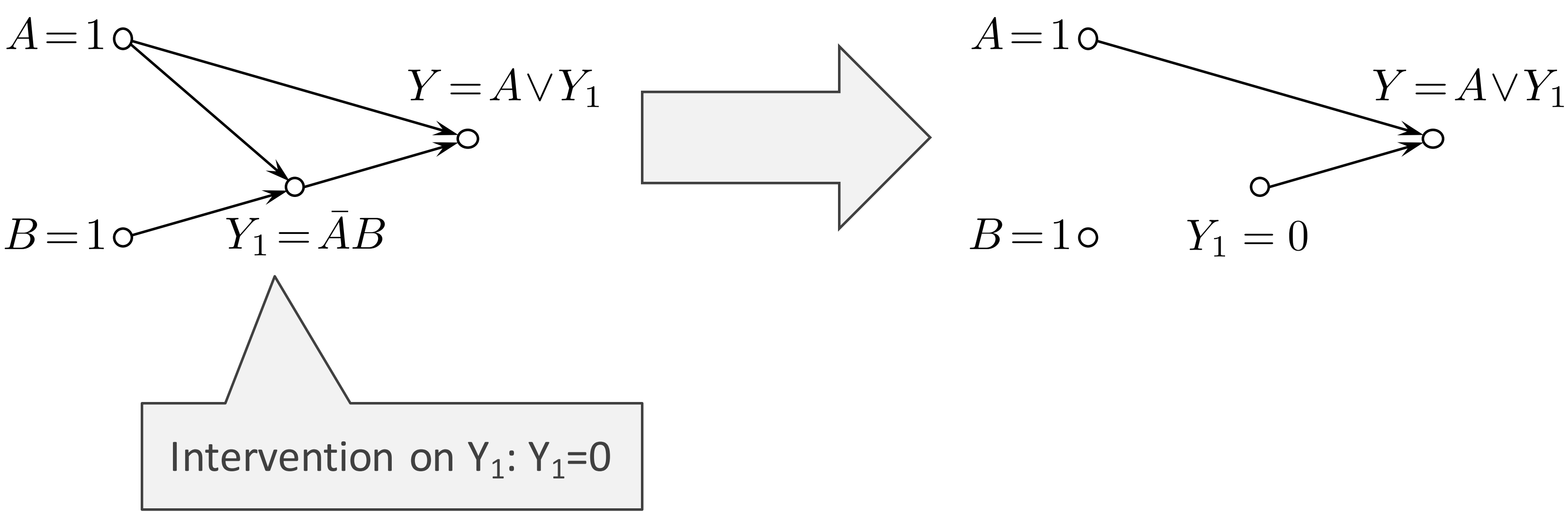


- Modeling problems:
  - *E.g., A bottle breaks if either Alice or Bob throw a rock at it.*
  - Endogenous variables:
    - Alice throws a rock (A)
    - Bob throws a rock (B)
    - The bottle breaks (Y)
  - Exogenous variables:
    - Alice's aim, speed of the wind, bottle material etc.

Woodward'2003, Hagmeyer'2007

Intervention / contingency

- External interventions modify the structural equations or values of the variables.



Counterfactual vs. Actual Cause

- Is Counterfactual the right notion of cause?

- Example 1:
  - Fire A approaches house. The house is burnt. In absence of the fire A, the house would not be burnt. Fire A is the cause

- Example 2:
  - Fire A and B approach house. Fire A reaches first. The house is burnt. In absence of the fire A, the house would still be burnt.
  - But still Fire A will be a cause (e.g. for legal purposes, replace Fire by a Burglar)
  - Not counterfactual but "Cause in Fact"
  - Fire B = 0 is the contingency
  - Resemblance with necessary and sufficient conditions

Counterfactual Cause

- If *not A* then *not φ*
  - In the absence of a cause, the effect doesn’t occur

$$C = A \wedge B, \quad A = 1 \wedge B = 1$$ ⟵ Both counterfactual

- Problem: Disjunctive causes
  - If Alice doesn’t throw a rock, the bottle still breaks (because of Bob)
  - Neither Alice nor Bob are counterfactual causes

$$C = A \vee B, \quad A = 1 \wedge B = 1$$ ⟵ No counterfactual causes

Actual Cause

[simplification]

A variable X is an actual cause of an effect Y if there exists a contingency that makes X counterfactual for Y.

$$C = A \vee B, \quad A = 1 \wedge B = 1 \longleftarrow$$

A is a cause under the contingency B=0

A Formal Definition of Actual Cause

- Actual causes are of the form
  - $X_1 = x_1 \wedge X_2 = x_2 \wedge ..... \wedge X_k = X_k$
  - In short, $X = x$
- For $X = x$ to be an actual cause of event Z The following three conditions should hold
  - Both $X = x$ and Z are true in the actual world
  - Changing X to x’ and some other variables W from w to w’ changes Z from true to false
  - Setting W to w’ does not have an effect on Z
- X is minimal- no subset of X satisfies the above two conditions

Example 1

$$Y = X_1 \wedge X_2$$

$X_1 = 1$ is counterfactual for $Y = 1$

$X_1 = 1, X_2 = 1 \Rightarrow Y = 1$

$X_1 = 0, X_2 = 1 \Rightarrow Y = 0$

Example 2

$$Y = X_1 \vee X_2$$

$X_1 = 1$ is not counterfactual for $Y=1$

$X_1 = 1$ is an actual cause for $Y = 1$, with contingency $X_2 = 0$

$X_1 = 1, X_2 = 1 \Rightarrow Y = 1$

$X_1 = 1, X_2 = 0 \Rightarrow Y = 1$

$X_1 = 0, X_2 = 0 \Rightarrow Y = 0$

Example 3

$$Y = (\neg X_1 \wedge X_2) \vee X_3$$

$X_1 = 1$ is not counterfactual for $Y = 1$

$X_1 = 1$ is not an actual cause for $Y = 1$

$X_1 = 1, X_2 = 1, X_3 = 1 \Rightarrow Y = 1$

$X_1 = 0, X_2 = 1, X_3 = 1 \Rightarrow Y = 1$

$X_1 = 1, X_2 = 0, X_3 = 1 \Rightarrow Y = 1$

$X_1 = 0, X_2 = 0, X_3 = 1 \Rightarrow Y = 1$

....

Y never changes by flipping X_1 for all combinations of X_2, X_3

Is the causality definition circular?

- Are we assuming causes and then inferring causes?
  - No!

- Causal model represents physical laws or potential causes
- The goal is to find the cause of a single event
  - in terms of values assigned to the variables
  - e.g. whether arson caused the fire on 6/10/2000, given what is known or assumed about this particular fire
- The causes are (variable, value) pairs, assuming a causal network on the variables

- Difference with Rubin's causal model and causal inference in statistics by randomized experiments or observational data

Chockler-Halpern' 2004

Responsibility

A measure of the degree of causality

$$\rho = \frac{1}{1 + \min_{\Gamma} |\Gamma|}$$

size of the contingency set

Example

$Y = A \wedge (B \vee C)$ $\quad A = B = C = 1 \Rightarrow Y = 1$

A=1 is counterfactual for Y=1 (ρ=1)

B=1 is an actual cause for Y=1, with contingency C=0 (ρ=0.5)

Eiter- Lukasiewicz' 2002

Complexity

- Actual Cause
  - NP-complete for binary variables
  - Σ_2^P-complete for non-binary variables

Proof sketch: Reduction from SAT.
Given F, F is satisfiable iff X is an actual cause for $X \wedge F$

Tian-Pearl' 2000

Submodel

- Model M = (U, V, F)
  - exogenous variables U
  - endogenous variables V
  - structural equations F
- Submodel M_X
  - Set endogenous variables X to x as constants in F_X
  - Remaining variables/equations remain the same
- do(X = x)
  - "do algebra" of Pearl

Tian-Pearl’ 2000

Potential Response in a submodel

- Model (U, V, F)
- Y = an endogenous variable in V
- The solution of Y from the structural equations F_X gives the potential response of the action X = x

- Denoted by Y_X

  = The value that Y would have obtained, had X been x

Probabilistic Causal Model

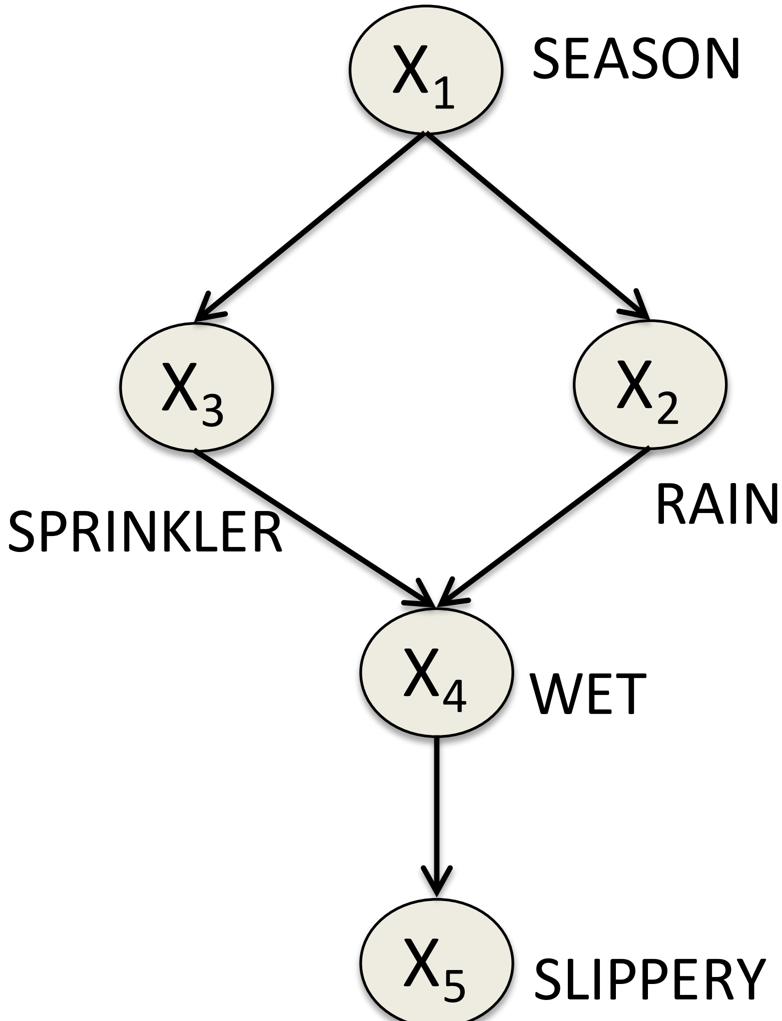
- A pair (M, P(u))
- P(u) is a probability function defined over the exogenous variables U
- Each endogenous variable in V is a function of exogenous variables U
  - also gives a distribution on V
- In turn gives the probability of counter factual statement $Pr(Y_{X=x} = y)$ or simply $Pr(Y_X = y)$

Tian-Pearl' 2000

Probabilistic model

- probability of necessity
  $= \Pr(Y_{X=x'} = y' \mid X = x, Y = y)$
  $= \Pr(y'_{x'} \mid x, y)$ ---- in short
  - the probability that event y would not have occurred in the absence of event x, $(= y'_{x'})$, given that x and y did in fact occur
- probability of sufficiency
  $= \Pr(Y_{X=x} = y \mid X = x', Y = y')$
  $= \Pr(y_x \mid x', y')$ ---- in short
  - the probability that setting x would produce y in a situation where x and y are in fact absent
  - captures the capacity of x to "produce" y

Example: Conditional Probability vs. Action



Structural equations

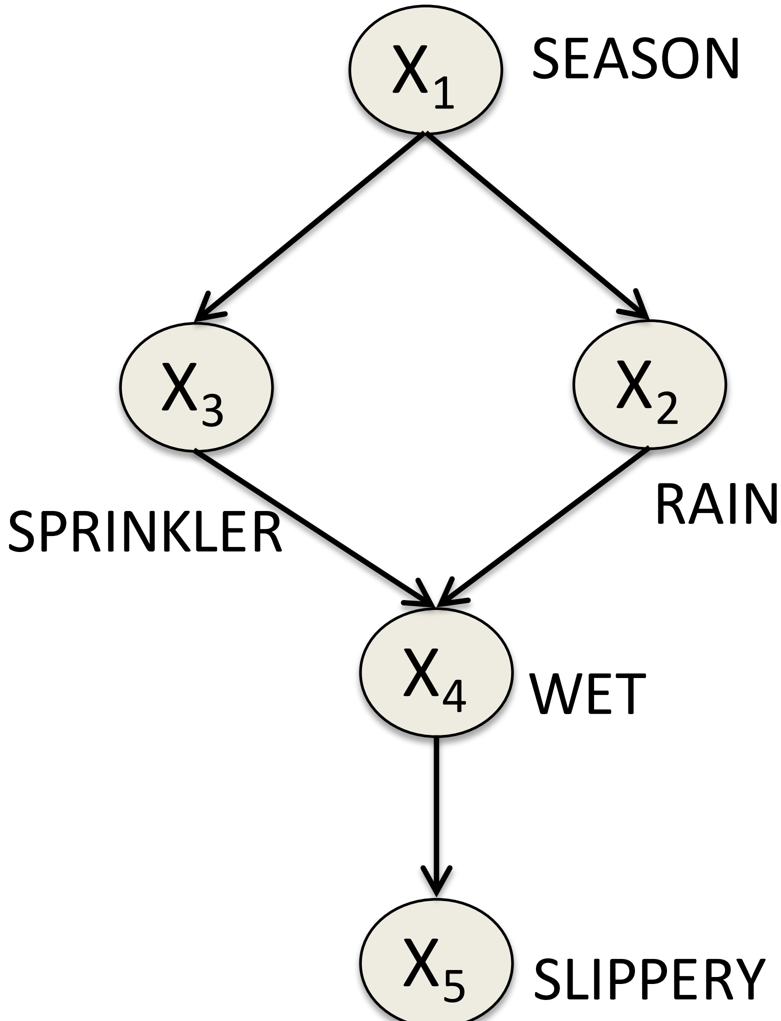
- $x_1 = u_1$
- $x_2 = f(x_1, u_2)$
- $x_3 = f(x_1, u_3)$
- $x_4 = f(x_2, x_3, u_4)$
- $x_5 = f(x_4, u_5)$

Exogenous vars

- $U = \{u_1, u_2, u_3, u_4, u_5\}$
- e.g. u_4 = a pipe is broken
- assumed to be independent

Pearl' 1999

Example: Conditional Probability vs. Action



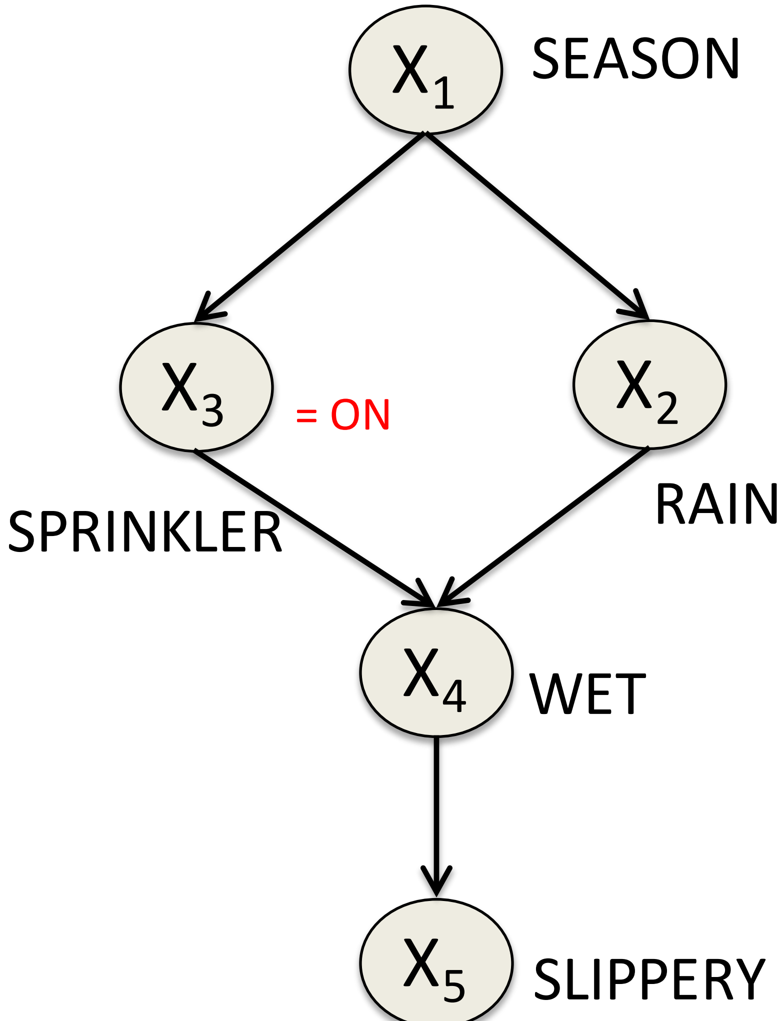
Joint probability distribution

$\Pr(x_1, x_2, x_3, x_4, x_5)$

$= \Pr(x_1) \Pr(x_2|x_1) \Pr(x_3|x_1)$
$\Pr(x_4|x_3, x_2) \Pr(x_5|x_4)$

Pearl' 1999

Example: Conditional Probability vs. Action



Joint probability distribution on "observing X_3 = ON"

$\Pr(x_1, x_2, x_3, x_4, x_5 \mid x_3 = 1)$

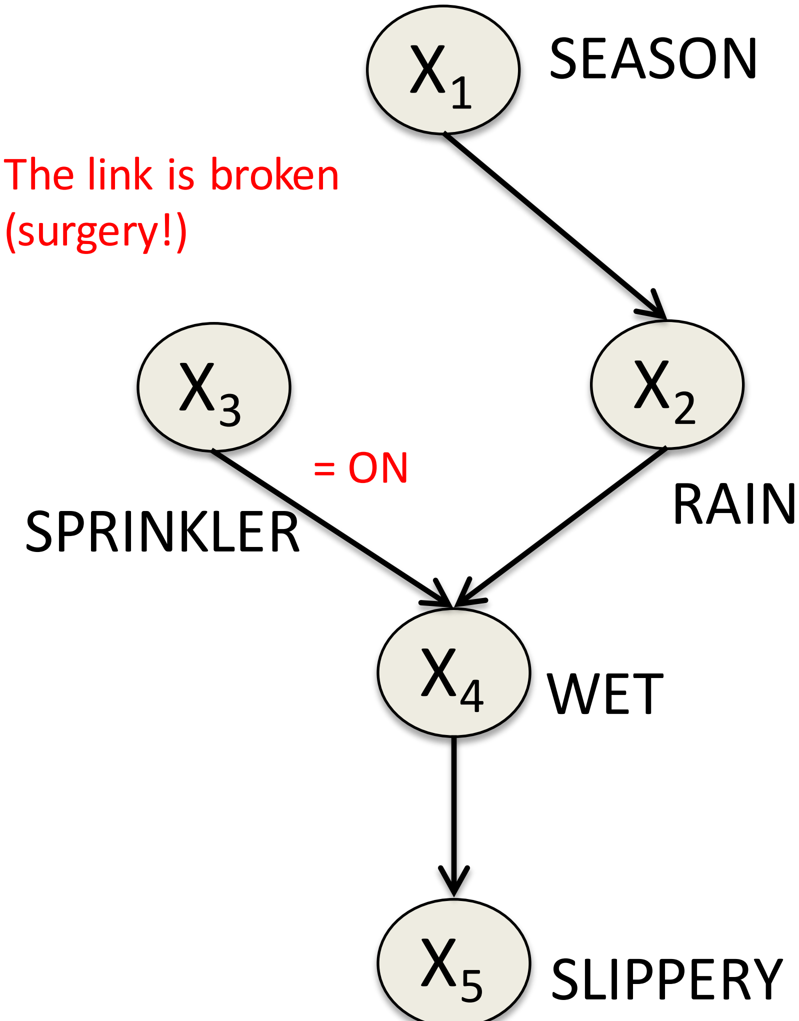
$= \Pr(x_1, x_2, x_3 = 1, x_4, x_5)$
$/ \Pr(x_3 = 1)$

$= [\Pr(x_1) \Pr(x_2|x_1) \Pr(x_3=1|x_1)$
$\Pr(x_4|x_3=1, x_2) \Pr(x_5|x_4)]$
$/ \Pr(x_3 = 1)$

.... Equation (1)

Pearl' 1999

Example: Conditional Probability vs. Action

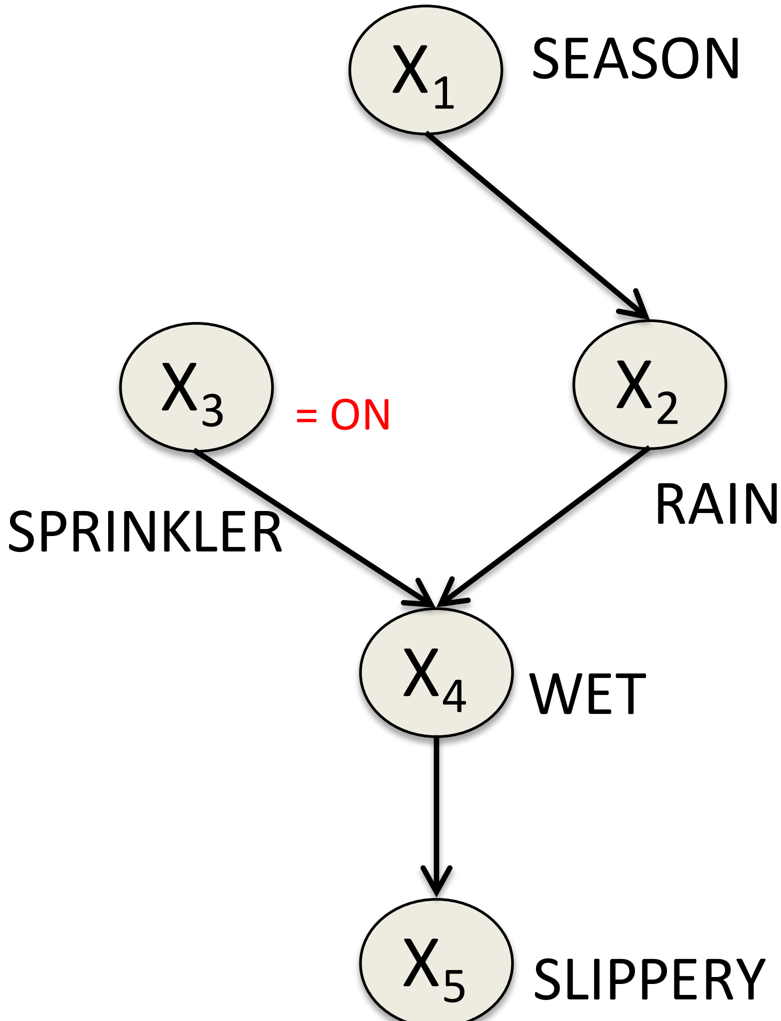


Structural equations for the "action do (X_3 = ON)"

- $x_1 = u_1$
- $x_2 = f(x_1, u_2)$
- $x_3 = 1$
- $x_4 = f(x_2, x_3, u_4)$
- $x_5 = f(x_4, u_5)$

Pearl' 1999

Example: Conditional Probability vs. Action



Joint probability distribution on "action do (X_3 = ON)"

$Pr(x_1, x_2, x_3, x_4, x_5 \mid do\ (X_3 = ON)$

$= Pr(x_1)\ Pr(x_2|x_1)\ Pr(x_4|x_3{=}1,\ x_2)\ Pr(x_5|x_4)]$

.... Equation (2)

- x_3 is treated as a constant

Pearl' 1999

Computing probability of a counterfactual sentence

- Equation 2 and 1 are very different
- But the probability in equation 2 can still be computed if the causal graph and probabilities are available
- However, sometimes the graph and probabilities are not sufficient
  - may need the functional forms f_i in structural equations
  - e.g. "the pavement would be slippery if the sprinkler were off, given that currently the pavement *is* slippery"

Pearl' 1999

Computing probability of a counterfactual sentence

In general, the conditional probability of a counterfactual sentence

"If it were A then B", given evidence e"

can be computed in three steps:

- Abduction
  - update P (u) by the evidence e, to obtain P (u | e).
- Action
  - Modify M by the action *do*(A), where A is the antecedent of the counterfactual, to obtain the submodel M_A.
- Deduction
  - Use the updated probability P (u | e) in conjunction with M_A to compute the probability of the counterfactual consequence B.