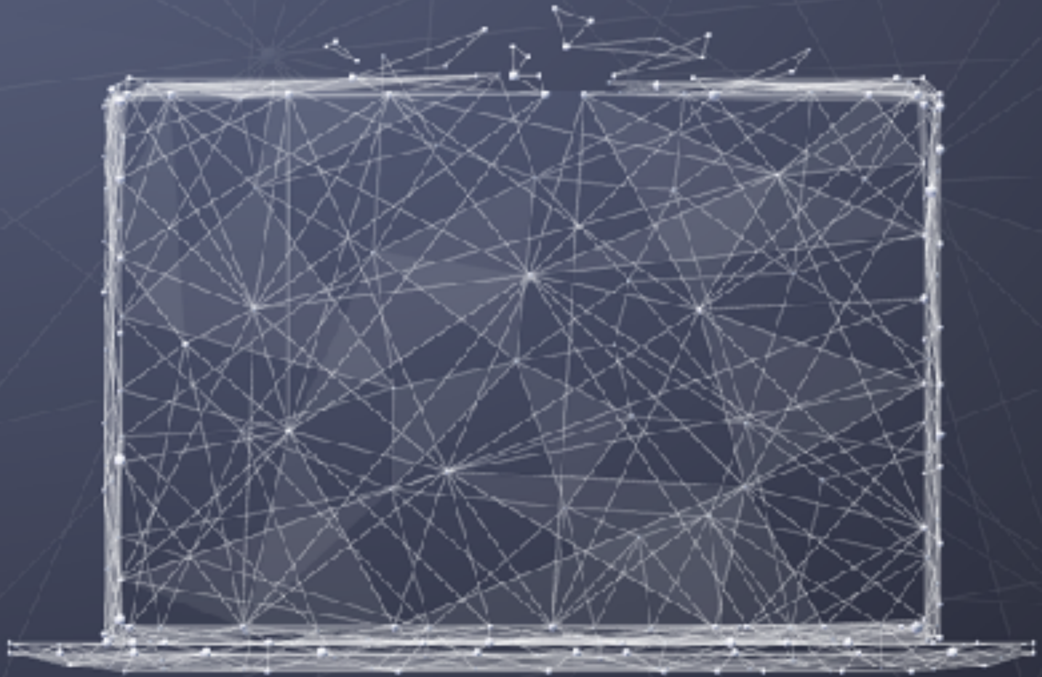
Data Science
Foundations of Decision Making

Causal reasoning



PURDUE UNIVERSITY | College of Science

Simpson’s paradox

- Simpson's paradox occurs when groups of data show one particular trend, but this trend is reversed when the groups are combined together
  - Specifically it refers to the scenario when marginal and conditional associations are opposing
- Understanding and identifying this paradox is important for correctly interpreting data
- This result is is particularly problematic when frequency data is unduly given causal interpretations

Example

- In the early 1970s, the University of California, Berkeley was sued for gender discrimination over admission to graduate school
- Of the 8442 male applicants for the fall of 1973, 44% were admitted, but only 35% of the 4351 female applicants were accepted
- Assuming the applicants' qualifications were similar, this pattern indeed appeared consistent with gender discrimination
- However, when researchers looked more closely within specific departments, this bias against women went away, and even reversed in several cases

Example

Department	#male applicants	#female applicants	%male admit	%female admit
A	825	108	62	82
B	560	25	63	68
C	325	593	37	34
D	417	375	33	54

- Normally one would expect that higher admittance rates across all groups would lead to a higher admittance rate overall. However, this is only guaranteed to be the case if the group sizes are equal. When group sizes differ, the totals for each side might be dominated by particular groups

- The explanation is that women applied in larger numbers to departments that had lower admittance rates

Explanation

- Mathematically, Simpson’s pardox is not paradoxical
  - Statistically, it says that the apparent relationship between two variables can change in the light or absence of a third
- Variables that are correlated with both the explanatory and response variables can distort the estimated effect
  - Application rate was correlated with gender and admit rate
- How do we avoid this type of paradox? Have a well−designed study, which identifies and properly accounts for all possible hidden variables

Causal inference

- Prediction and causation are very different. Typical questions are:
  Prediction: Predict Y after observing X = x
  Causation: Predict Y after setting X = x

- Causation involves predicting the effect of an intervention, e.g:
  Prediction: Predict health given that a person takes vitamin C
  Causation: Predict health if I give a person vitamin C

- The difference between passively observing X and actively intervening and setting X is significant and requires different techniques and, typically, much stronger assumptions

Association vs. causation example

- Consider the conjecture that a new diet is linked to lower risk of inflammatory arthritis
- You observe that in a given sample:
  - A smaller fraction of individuals on the diet have inflammatory arthritis
  - A larger fraction of individuals not on the diet have inflammatory arthritis
- So you recommend that everyone pursue this new diet, but rates of inflammatory arthritis are unaffected. Why?

Association vs. causation example

- Consider the conjecture that a new e-mail promotion is useful to your business
- You observe that:
  - Those who received the e-mail promotion did not convert at substantially higher rates than those who did not receive the e-mail.
- So you give up...and later, another product manager runs an experiment with a similar idea, and conclusively demonstrates that the email promotion raises conversion rates. Why?

Association vs. causation

- In each case, we were unable to see what would have happened to each individual if the alternative action had been applied
  - In the arthritis scenario, suppose only individuals predisposed to being healthy do the diet in the first place. Then you cannot see either what happens to an unhealthy person who follows the diet, or a healthy person who does not follow the diet
  - In the e-mail scenario, suppose only individuals who are unlikely to convert received the e-mail. Then you cannot see either what happens to an individual who is likely to convert who receives the promotion, or an individual who is not likely to convert who does not receive the promotion
- The lack of information is what prevents inference about causation from association data

Counterfactuals

- In these examples, the unseen information about each individual is the counterfactual
- Without reasoning about counterfactuals, we can't draw causal inferences—or worse, we draw the wrong causal inferences
- The potential outcomes model is a way to formally think about counterfactuals and causal inference.

Potential outcomes

- Suppose there are two possible actions that can be applied to an individual:
  - 1 (“treatment”)
  - 0 (“control”)
- For each individual in the population, there are two associated potential outcomes:
  - Y(1): outcome if treatment applied
  - Y(0): outcome if control applied

Causal effects

- The causal effect of the action for an individual is the difference between the outcome if they are assigned treatment or control:
  causal effect = $Y(1) - Y(0)$
- The fundamental problem of causal inference is this:
  - In any example, for each individual, we only get to observe one of the two potential outcomes
  - Causal inference is a problem of missing data

Assignment

- We can’t observe both potential outcomes for each individual, so we have to get around it in some way. Eg:
  - Observe the same individual at different points in time
  - Observe two individuals who are nearly identical to each other, and give one treatment and the other control (e.g., AB testing)
- The assignment mechanism is what decides which outcome we get to observe. We let $W = 1$ (resp., 0) if an individual is assigned to treatment (resp., control).
  - In the arthritis example, individuals self-assigned themselves
  - In the e-mail example, there was a bias in the assignment

Average treatment effect

- When we can't observe both outcomes for each individual, we can estimate the average treatment effect (ATE) in the population:
  $ATE = E[Y(1)] - E[Y(0)]$
- In doing so we lose individual information, but now we have a reasonable chance of getting an estimate of both terms in the expectation
- However, to estimate ATE accurately, assignment to treatment should be uncorrelated with the outcome
  - This requirement is automatically satisfied if W is assigned randomly, since then W and the outcomes are independent—this is the case in a randomized experiment

Randomized assignment

- Goal: Construct two groups who are nearly identical in every way, then measure differences between the two groups.
- Randomization provides a way of creating two groups that are as similar as possible prior to treatment:
  - If subjects are randomly assigned to groups, there shouldn't be any systematic difference (e.g., selection bias)
  - Since only difference between the groups is that one gets treated and the other doesn't, we ascribe differences in outcomes to the treatment

Selection bias

- If assignment is correlated with the outcome, then estimated average effects may not be accurate
- Selection bias is rampant in conflating association and causation
- Mechanisms for avoiding selection biases include:
  - Using randomization to select subgroups from populations
  - Ensuring that the selected subgroups are representative of the population in terms of their key characteristics (this method is less of a protection than the first, since the key characteristics may be unknown)