Causal inference

Wenda Zhou

June 22, 2017

Causal inference

Attempt to make causal claims about non-experiment data.

Example

Wish to understand if private school students perform better.

Causal inference

Attempt to make causal claims about non-experiment data.

Example

Wish to understand if private school students perform better.

Require care to avoid confounding (need to compare apples to apples).

Causality

Causality is a delicate philosophical topic.

- Abstraction on the progression of the world
- Usually though of as counterfactuals (what if?)
- In statistics: try to mimic a randomized experiment

Potential outcomes

Need to think about causal outcomes mathematically.

Potential outcome

Consider both counterfactuals of the outcome.

$$Y_i(0) = \text{outcome for individual } i \text{ under control,}$$
$$Y_i(1) = \text{outcome for individual } i \text{ under treatment.}$$

Potential outcomes

Need to think about causal outcomes mathematically.

Potential outcome

Consider both counterfactuals of the outcome.

$$
\begin{aligned}
Y_i(0) &= \text{outcome for individual } i \text{ under control,} \\
Y_i(1) &= \text{outcome for individual } i \text{ under treatment.}
\end{aligned}
$$

Only ever observe one counterfactual.

The potential outcome setup

For each observation, we may consider the following random variables:

$$\begin{aligned} Y_i(0), Y_i(1) &\text{ the counterfactual outcomes} \\ T_i &\text{ the treatment assignment} \\ X_i &\text{ other covariates} \end{aligned}$$

but we only ever observe $Y_i(T_i)$ but not the other one.

Estimating an average treatment effect

We want to understand the average effect of the treatment:

$$\text{ATE} = \mathbb{E}[Y(1) - Y(0)] \tag{1}$$

Estimating an average treatment effect

We want to understand the average effect of the treatment:

$$\mathrm{ATE} = \mathbb{E}[Y(1) - Y(0)] \tag{1}$$

If T_i is independent of everything, can write

$$\mathrm{ATE} = \mathbb{E}[Y(1) \mid T = 1] - \mathbb{E}[Y(0) \mid T = 0] \tag{2}$$

Uncounfoundness

Causal inference relies on an (unverifiable) assumption: uncounfoundness.
We must have that:

$$Y(0), Y(1) \perp T \mid X \tag{3}$$

that is: the treatment assignment does not depend on the outcome conditional on the covariates.

Uncounfoundness

Causal inference relies on an (unverifiable) assumption: uncounfoundness.
We must have that:

$$Y(0), Y(1) \perp T \mid X \tag{3}$$

that is: the treatment assignment does not depend on the outcome conditional on the covariates.
Example: given the socio-economic status, age and covariates of a person, their propensity to smoke does not depend on their risk of heart disease.

Comparing apples to apples

In order for our comparison to be valid, must compare apples to apples.

Comparing apples to apples

In order for our comparison to be valid, must compare apples to apples.

Matching

Idea: compare similar individuals, and average their differences.

Propensity score matching

In fact, enough to compare individuals with the same likelihood of being treated.

Propensity score

Define the propensity score to be:

$$\mathrm{P}(T = 1 \mid X = x) = p(x) \tag{4}$$

Propensity score matching

In fact, enough to compare individuals with the same likelihood of being treated.

Propensity score

Define the propensity score to be:

$$P(T = 1 \mid X = x) = p(x) \tag{4}$$

Propensity score matching

Compute propensity score, then compare observations with similar propensity scores.

Mediation

Suppose that we randomize the treatment T, and observe some outcome Y.

Mediation

Suppose that we also observe some variable M, and we wish to understand the effect of M on Y.

Mediation

Idea: model how M depends on T, and model how Y depends on M and T. Combine to obtain:

- **ACME** Average causal mediation effect: the causal effect of the mediator M on the outcome Y.
- **ADE** Average direct effect: the effect of the treatment on the outcome Y not going through M.

Mediation

Idea: model how M depends on T, and model how Y depends on M and T. Combine to obtain:

- **ACME** Average causal mediation effect: the causal effect of the mediator M on the outcome Y.
- **ADE** Average direct effect: the effect of the treatment on the outcome Y not going through M.

Decompose the total effect of T on Y into the ACME and ADE.

Instrumental variables

Attempt to deal with unmeasured confounding.
Idea: introduce additional source of variation that affects the treatment only through the variable of interest.

Instrumental variables

- One of the most popular methods in econometrics
- What constitutes a valid instrument can be subtle