# Asking Good Causal Questions: Positivity and Consistency

Cornell STSCI / INFO / ILRST 3900
Fall 2023
causal3900.github.io

19 Sep 2024

# Learning goals for today

At the end of class, you will be able to ask good causal questions.

Good causal questions

- involve precise treatments (consistency assumption)
- with clarity about interference (consistency assumption)
- involve treatments that exist (positivity assumption)

After class:

- Hernán and Robins 2020 Chapter 3.3-3.6
- Optional: Does water kill? A call for less casual causal inferences. (Hernán, M. 2016)

Good causal questions involve **precise treatments**:

$$\text{If } A_i = a, \text{ then } Y_i = Y^a$$

(consistency)

# Good causal questions involve **clarity about interference**

# Clarity about interference

You see an cool advertisement for some the new upside down whoppers and the next day you buy it for lunch.

$$Y_{\text{You}} = \text{Purchase}$$

Maybe, if you had not see the advertisement, you would not have bought the product.

$$Y_{\text{You}}^{\text{No ad}} = \text{No Purchase}$$

Your friend who *has not* seen the ad, but sees you enjoying your meal, so they get it for lunch the next day.

- What are your friend's potential outcomes?

# Potential Outcomes under Interference

- Your friend's potential outcomes depend on their own treatment assignment as well as yours!

| $a_{\text{you}}$ | $a_{\text{friend}}$ | $Y_{\text{friend}}^{a_{\text{you}}, a_{\text{friend}}}$ |
|---|---|---|
| ad | ad | Purchase |
| no ad | no ad | No Purchase |
| ad | no ad | Purchase |
| no ad | ad | No Purchase |

What is the problem if I write

$$Y_{\text{friend}}^{\text{ad}} \quad \text{or} \quad Y_{\text{friend}}^{\text{no ad}}?$$

- Your treatment "interfered" with their outcome
- In general, when an individual's outcome can depend on the treatment of others in the population, we call this *interference*

# Activity: Interference

In what other applications might interference be something to think about?

# Interference

- In most general settings, everyone can affect everyone else's outcomes!

$$Y_i^{a_1, a_2, \cdots, a_n}$$

- $2^n$ possible potential outcomes per person!
- Interference is present in many settings
- Not all settings have interference!
- Good causal questions account for interference when interference is present

Good causal questions involve

**treatments that exist**

# 1. Treatments that exist (positivity)

| Employer 1 | Employer 2 |
| --- | --- |
| 100 employees | 200 employees |
| Face-to-face interaction | Work in individual offices |
| 75% randomized to vaccine<br>25% randomized to no vaccine | 50% randomized to vaccine<br>50% randomized to no vaccine |

How do you estimate the conditional average treatment effect (CATE)?

# 1. Treatments that exist (positivity)

| Employer 1 | Employer 2 |
| --- | --- |
| 100 employees | 200 employees |
| Face-to-face interaction | Work in individual offices |
| 100% randomized to vaccine<br>0% randomized to no vaccine | 50% randomized to vaccine<br>50% randomized to no vaccine |

How do you estimate the conditional average treatment effect (CATE)?

# 1. Treatments that exist

If units are exchangeable given a confounder $L$,
then to estimate $E(Y^a)$ we need **positivity** to hold

$$\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0 \text{ for all } a, \vec{\ell}$$

where $\vec{L} := (L_1, L_2, \cdots, L_m) \in \mathbb{R}^m$ is a vector of covariates.

Recall Inverse Probability Weighting:

$$\widehat{ATE}_{\mathrm{IPW}} = \frac{1}{N} \sum_{i:A_i=1} \frac{Y_i}{\pi_i^1} \; - \; \frac{1}{N} \sum_{i:A_i=0} \frac{Y_i}{\pi_i^0}$$

Some treatments simply do not exist in some populations.


Source: Wikimedia A, B, C

Would the bulbs in Ithaca bloom if it did not freeze all winter?

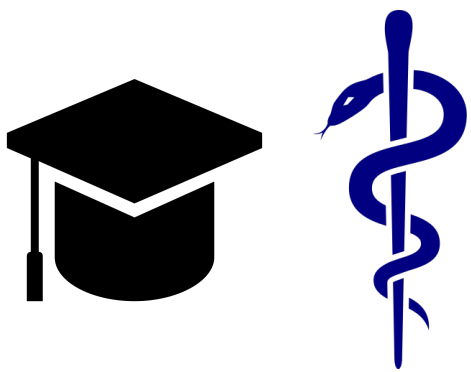
| | |
|---|---|
| Confounder $L$ | Ithaca |
| Treatment $a$ | Did not freeze |
| Outcome $Y^a$ | Blooms? |

Some treatments simply do not exist in some populations.


Source: Wikimedia A, B, C

Would the bulbs in Ithaca bloom if it did not freeze all winter?

| | |
|---|---|
| Confounder $L$ | Ithaca |
| Treatment $a$ | Did not freeze |
| Outcome $Y^a$ | Blooms? |



Source: Wikimedia A and B

Sarah has no MD training. Would Sarah earn more money if she were a surgeon?

| | |
|---|---|
| Confounder $L$ | No MD training |
| Treatment $a$ | Surgeon |
| Outcome $Y^a$ | Earnings |

# 1. Treatments that exist

We can choose causal questions so that positivity holds.

$$\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$$

- in each population subgroup $\vec{L} = \vec{\ell}$
- only study treatment values $a$ that can actually happen

# Good causal questions: In math

We should study treatments that exist (positivity)

$$\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$$

with potential outcomes that are well-defined (consistency)

$$Y = Y^a \text{ when } A = a$$

Well-defined potential outcomes involve precise treatments

$$\mathrm{BA}^{\text{Binghamton}} \quad \text{instead of} \quad \mathrm{BA}^{\text{SUNY}}$$

and incorporate interference when it exists

$$Y^{a_{\text{you}}, a_{\text{your friend}}} \quad \text{instead of} \quad Y^{a_{\text{you}}}$$

# Course Project: Overview

- Walk through an entire causal analysis: starting with defining a causal question all the way to communicating the results of your analysis
- Project split up into different tasks you'll complete throughout the semester, with feedback
- Some parts you will complete individually, some as a group
- We'll post details of all this on the website

# Course Project: Task One

- Released today
- **Due Thursday, October 3rd by 5pm**
- Pick a causal question that is interesting to you and a dataset that you can use to study this question
- 25 point assignment
  - 10 points: well-defined and clear causal question + clear potential outcomes notation
  - 5 points for explaining how the fundamental problem of causal inference applies
  - 10 points for having a dataset that you can use to answer your question
- We will post detailed descriptions (including a more detailed rubric) of all this
- Descriptions will include pointers to possible datasets

# Learning goals for today

At the end of class, you will be able to ask good causal questions.

Good causal questions

- involve precise treatments (consistency assumption)
- with clarity about interference (consistency assumption)
- involve treatments that exist (positivity assumption)

After class:

- Hernán and Robins 2020 Chapter 3.3-3.6
- Optional: Does water kill? A call for less casual causal inferences. (Hernán, M. 2016)