Causality and Counterfactuals

Matti Sarvimäki

Mini-Course on Causal Inference
Lecture 1

Causal questions

- We often want to evaluate the **impact** of X on Y, e.g.
  - education on earnings
  - marketing campaing on sales
  - carbon tax on emissions
  - R&D subsidy on innovation
  - fiscal stimulus on unemployment

Causal questions

- We often want to evaluate the **impact** of X on Y, e.g.
  - education on earnings
  - marketing campaing on sales
  - carbon tax on emissions
  - R&D subsidy on innovation
  - fiscal stimulus on unemployment
- These are **causal** questions
  - requires evaluating *counterfactual* states of the world
  - "how would Y change if we changed X?"

Causal questions

- We often want to evaluate the **impact** of X on Y, e.g.
  - education on earnings
  - marketing campaing on sales
  - carbon tax on emissions
  - R&D subsidy on innovation
  - fiscal stimulus on unemployment
- These are **causal** questions
  - requires evaluating *counterfactual* states of the world
  - "how would Y change if we changed X?"
- Compare to **descriptive** questions
  - requires measuring the actual state of the world
  - "what is joint distribution of X and Y?"

Randomized experiments

- The next four lectures will focus on answering causal questions using research designs based on **randomization**
  - the simplest context for learning relevant statistical concepts
- The prime example is randomized controlled trials (RCT)
  - RCTs have become an important part of economits' toolkit
  - you might end up running them for living
  - you will definitely end up interpretting results from other people's RCTs

Randomized experiments

- The next four lectures will focus on answering causal questions using research designs based on **randomization**
  - the simplest context for learning relevant statistical concepts
- The prime example is randomized controlled trials (RCT)
  - RCTs have become an important part of economits' toolkit
  - you might end up running them for living
  - you will definitely end up interpretting results from other people's RCTs
- Even when we can't run an experiment, it is often helpful to ask: what would be the **ideal experiment** for answering this question?
  - helpful benchmark for "naturally occurring" or "quasi" experiments
    - we'll discuss an example of a "natural experiment" involving actual randomization already in the next class
    - you'll see other types of quasi-experimental approaches Ciprian's and Kristiina's parts

Today's learning objectives

- Good understanding of **why randomization eliminates selection bias** and the content and importance of the following concepts:
  1. causality
  2. counterfactual
  3. potential outcomes
  4. treatment effect
  5. selection bias

Example: Impact of a new integration program

- Imagine that you have been asked to assist the government to evaluate a new type of integration program for immigrants
- How would you approach this task?

Example: Impact of a new integration program

- Imagine that you have been asked to assist the government to evaluate a new type of integration program for immigrants
- How would you approach this task?
- My take: helpful to break this into two parts
  - what is the question one needs to answer?
  - how to answer it?

Elements of causal questions

1. **Treatment**
   - impact *of* [...]
2. **Counterfactual**
   - impact *in comparison to* [...]
3. **Outcome** and **population**
   - impact *on* [...]

Elements of causal questions

1. **Treatment**
   - impact *of* [...]
2. **Counterfactual**
   - impact *in comparison to* [...]
3. **Outcome** and **population**
   - impact *on* [...]

- What is a well-defined question for our case study?

Elements of causal questions

1. **Treatment**
   - impact *of* [...]
2. **Counterfactual**
   - impact *in comparison to* [...]
3. **Outcome** and **population**
   - impact *on* [...]

- What is a well-defined question for our case study?
  - my take: "what is the impact of the **new program** in comparison to **business-as-usual programs** on **participants'** **cumulative unemployment benefits during their first three years in Finland**?
- Next: formal definitions using the potential outcomes framework

Potential Outcomes

- We focus on binary (0/1) **treatments** and denote treament status of individual i as

$$D_i = \begin{cases} 1 & \text{if she receives the treament} \\ 0 & \text{if she doesn't} \end{cases}$$

Potential Outcomes

- We focus on binary (0/1) **treatments** and denote treament status of individual i as

$$D_i = \begin{cases} 1 & \text{if she receives the treament} \\ 0 & \text{if she doesn't} \end{cases}$$

- We denote **outcomes** by y and define

$$\text{potential outcome} = \begin{cases} y_{1i} & \text{if } D_i = 1 \\ y_{0i} & \text{if } D_i = 0 \end{cases}$$

Potential Outcomes

- We focus on binary (0/1) **treatments** and denote treament status of individual i as

$$D_i = \begin{cases} 1 & \text{if she receives the treament} \\ 0 & \text{if she doesn't} \end{cases}$$

- We denote **outcomes** by y and define

$$\text{potential outcome} = \begin{cases} y_{1i} & \text{if } D_i = 1 \\ y_{0i} & \text{if } D_i = 0 \end{cases}$$

in words: y_{1i} is the outcome of individual i in the state of the world where she is treated and y_{0i} is her outcome in the state of the world where she was *not* treated (note: only one state of the world occurs)

Treatment effect

- The treatment effect for individual i is:

$$y_{1i} - y_{0i}$$

in words: difference in the potential outcomes with and without the treatment

Treatment effect

- The treatment effect for individual i is:

$$y_{1i} - y_{0i}$$

in words: difference in the potential outcomes with and without the treatment

- The fundamental challenge of causal inference is that we cannot observe both y_{1i} and y_{0i} for the same individual. Instead, we observe

$$y_i = \begin{cases} y_{1i} & \text{if } D_i = 1 \\ y_{0i} & \text{if } D_i = 0 \end{cases}$$

Treatment effect

- We can never identify the treatment effect for an individual person, but sometimes we can estimate average treatment effects:

$$\text{Average treatment effect (ATE)} \quad = \quad \mathbb{E}\left[y_{i1} - y_{0i}\right]$$

Treatment effect

- We can never identify the treatment effect for an individual person, but sometimes we can estimate average treatment effects:

$$\begin{aligned} \text{Average treatment effect (ATE)} &= \mathbb{E}\left[y_{i1} - y_{0i}\right] \\ \text{ATE for the treated (ATT)} &= \mathbb{E}\left[y_{i1} - y_{0i} | D_i = 1\right] \end{aligned}$$

where $\mathbb{E}[a|b]$ is the expectation of a conditional on b

Treatment effect

- We can never identify the treatment effect for an individual person, but sometimes we can estimate average treatment effects:

$$\begin{aligned} \text{Average treatment effect (ATE)} &= \mathbb{E}\left[y_{i1} - y_{0i}\right] \\ \text{ATE for the treated (ATT)} &= \mathbb{E}\left[y_{i1} - y_{0i} | D_i = 1\right] \end{aligned}$$

where $\mathbb{E}\left[a|b\right]$ is the expectation of a conditional on b

- Why ATE *and* ATT?
  - treatment effect may be different for those getting the treatment than it would be for those not getting it (e.g. specific integration policy)
  - internal validity: do we learn the true effect for the treated population?
  - external validity: can we extrapolate to other populations?

Research designs and control groups

- We use a *comparison* or *control* group to approximate what would have happened to the treated in the absence of the treatment
  - that is, we *estimate* the counterfactual $\mathbb{E}[y_{0i}|D_i = 1]$

Research designs and control groups

- We use a *comparison* or *control* group to approximate what would have happened to the treated in the absence of the treatment
  - that is, we *estimate* the counterfactual $\mathbb{E}[y_{0i}|D_i = 1]$
- In economics parlance, this approach is know as "**design-based**" or "**reduced form**" or "**experimental**" approach
  - the alternative is the "structural" approach, where we use quantitative economic models to simulate counterfactual states of the world

Research designs and control groups

- We use a *comparison* or *control* group to approximate what would have happened to the treated in the absence of the treatment
  - that is, we *estimate* the counterfactual $\mathbb{E}[y_{0i}|D_i = 1]$
- In economics parlance, this approach is know as "**design-based**" or "**reduced form**" or "**experimental**" approach
  - the alternative is the "structural" approach, where we use quantitative economic models to simulate counterfactual states of the world
- Invalid control group leads to **selection bias**
  - whether the control group provides a good counterfactual or not is the key question of all design-based causal inference

Selection Bias

- As the amount of data increases, the sample averages approach the population average (expectations)

$$\underbrace{Avg[y_i|D=1]}_{\text{treatment group}} - \underbrace{Avg[y_i|D=0]}_{\text{control group}} \rightarrow \mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0]$$

Selection Bias

- As the amount of data increases, the sample averages approach the population average (expectations)

$$\underbrace{Avg[y_i|D=1]}_{\text{treatment group}} - \underbrace{Avg[y_i|D=0]}_{\text{control group}} \rightarrow \mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0]$$

$$= \mathbb{E}[y_{1i}|D=1] - \mathbb{E}[y_{0i}|D=0]$$

Selection Bias

- As the amount of data increases, the sample averages approach the population average (expectations)

$$\underbrace{Avg[y_i|D=1]}_{\text{treatment group}} - \underbrace{Avg[y_i|D=0]}_{\text{control group}} \rightarrow \mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0]$$

$$= \mathbb{E}[y_{1i}|D=1] - \mathbb{E}[y_{0i}|D=0]$$

- Where the second row emphasizes that we observe y_{0i} only for the control group, while our objective is to estimate ATT, i.e.

$$\mathbb{E}\left[y_{i1} - y_{0i}|D_i = 1\right] = \mathbb{E}[y_{1i}|D=1] - \underbrace{\mathbb{E}[y_{0i}|D=1]}_{\text{never observed}}$$

Selection Bias

- As the amount of data increases, the sample averages approach the population average (expectations)

$$\underbrace{Avg[y_i|D=1]}_{\text{treatment group}} - \underbrace{Avg[y_i|D=0]}_{\text{control group}} \rightarrow \mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0]$$
$$= \mathbb{E}[y_{1i}|D=1] - \mathbb{E}[y_{0i}|D=0]$$

- Where the second row emphasizes that we observe y_{0i} only for the control group, while our objective is to estimate ATT, i.e.

$$\mathbb{E}\left[y_{i1} - y_{0i}|D_i = 1\right] = \mathbb{E}[y_{1i}|D=1] - \underbrace{\mathbb{E}[y_{0i}|D=1]}_{\text{never observed}}$$

- Selection bias arises when a control group leads to an incorrect estimate of the counterfactual, i.e. $\mathbb{E}[y_{0i}|D=0] \neq \mathbb{E}[y_{0i}|D=1]$

Selection Bias

- A particularly informative way to illustrate selection bias is:

$$\mathbb{E}[y_i|D=1] - \mathbb{E}[y_i|D=0] \quad = \quad \mathbb{E}[y_{1i}|D=1] - \mathbb{E}[y_{0i}|D=0]$$

Selection Bias

- A particularly informative way to illustrate selection bias is:

$$\begin{aligned}\mathbb{E}[y_i|D=1]-\mathbb{E}[y_i|D=0] &= \mathbb{E}[y_{1i}|D=1]-\mathbb{E}[y_{0i}|D=0] \\ &= \underbrace{\mathbb{E}[y_{1i}|D=1]-\mathbb{E}[y_{0i}|D=1]}_{\text{ATT}}+ \\ &\quad \underbrace{\mathbb{E}[y_{0i}|D=1]-\mathbb{E}[y_{0i}|D=0]}_{\text{Selection bias}}\end{aligned}$$

Selection Bias

- A particularly informative way to illustrate selection bias is:

$$\begin{aligned}\mathbb{E}[y_i|D=1]-\mathbb{E}[y_i|D=0] &= \mathbb{E}[y_{1i}|D=1]-\mathbb{E}[y_{0i}|D=0] \\ &= \underbrace{\mathbb{E}[y_{1i}|D=1]-\mathbb{E}[y_{0i}|D=1]}_{\text{ATT}} + \\ &\quad \underbrace{\mathbb{E}[y_{0i}|D=1]-\mathbb{E}[y_{0i}|D=0]}_{\text{Selection bias}}\end{aligned}$$

where the first step is from the previous slide and the second step is taken by simply adding and substracting $\mathbb{E}[y_{0i}|D=1]$

  - i.e. $\mathbb{E}[y_{0i}|D=1]-\mathbb{E}[y_{0i}|D=1]=0$, so including it does not change the result, but allows us to rewrite the equation as ATT+SB

Selection Bias

- A particularly informative way to illustrate selection bias is:

$$\begin{aligned}\mathbb{E}[y_i|D=1]-\mathbb{E}[y_i|D=0] &= \mathbb{E}[y_{1i}|D=1]-\mathbb{E}[y_{0i}|D=0] \\ &= \underbrace{\mathbb{E}[y_{1i}|D=1]-\mathbb{E}[y_{0i}|D=1]}_{\text{ATT}} + \\ &\quad \underbrace{\mathbb{E}[y_{0i}|D=1]-\mathbb{E}[y_{0i}|D=0]}_{\text{Selection bias}}\end{aligned}$$

where the first step is from the previous slide and the second step is taken by simply adding and substracting $\mathbb{E}[y_{0i}|D=1]$
  - i.e. $\mathbb{E}[y_{0i}|D=1]-\mathbb{E}[y_{0i}|D=1]=0$, so including it does not change the result, but allows us to rewrite the equation as ATT+SB
  - in words: differences in the average outcomes between treatment and control groups include the treatment effect *and* the selection bias (the difference between the two groups if neither had been treated)

In-class discussion: Selection bias and integration policies

- Let's return to the case of new integration program and speculate about the likely selection bias in two alternative control groups:
  1. all immigrants not participating in the program
  2. all immigrants participating in the business-as-usual program
- What would be an ideal way to create a control group?

Randomization eliminates selection bias

- Random assignment into treatment/control ensures that the control groups is *comparable* to the treatment group

Randomization eliminates selection bias

- Random assignment into treatment/control ensures that the control groups is *comparable* to the treatment group
- Formally: their potential outcomes are in expectation the same, i.e.

$$\mathbb{E}[y_{1i}|D=1] = \mathbb{E}[y_{1i}|D=0]$$
$$\mathbb{E}[y_{0i}|D=1] = \mathbb{E}[y_{0i}|D=0]$$

Randomization eliminates selection bias

- Random assignment into treatment/control ensures that the control groups is *comparable* to the treatment group
- Formally: their potential outcomes are in expectation the same, i.e.

$$\mathbb{E}[y_{1i}|D=1] = \mathbb{E}[y_{1i}|D=0]$$
$$\mathbb{E}[y_{0i}|D=1] = \mathbb{E}[y_{0i}|D=0]$$

- Thus $\mathbb{E}[y_{0i}|D=1] - \mathbb{E}[y_{0i}|D=0] = 0$, i.e. no selection bias
  - in words: the control group tells us what would have happened to the treatment group in the absence of the treatment

Summary

- **Causality**: how one thing *affects* another thing
  - requires comparing counterfactual states of the world to each other ("how would Y change if we changed X?")
  - at most, one of them is observed
- **Control group** in an experimental research design
  - the outcomes of the control group are used to infer what would have happened to the treatment group in the absence of the treatment
- **Selection bias** occurs when the control group is not comparable to the treatment group, i.e. $\mathbb{E}[y_{0i}|D=0] \neq \mathbb{E}[y_{0i}|D=1]$

  $=$ potential outcomes differ between the treatment and control groups
- **Randomization** eliminates selection bias
  - on expectation, the only difference between the groups is that the treatment group gets the treatment and the control group does not

  $\rightarrow$ differences in average outcomes must be due to the treatment