



Introduction to Causality

Graduate Quantitative Economics and Datascience

Jesse Perla

jesse.perla@ubc.ca

University of British Columbia

Table of contents

- Overview
- Introduction
- Potential Outcomes Framework

Overview

Summary

- Introduction and motivation for causal inference and randomization
- We will introduce the concepts of treatment effects, potential outcomes, and the fundamental problem of causal inference
- Material includes much adapted from **Causal Inference for the Brave and True: Introduction to Causality**
- Using the following packages and definitions

```
1 import pandas as pd
2 import numpy as np
3 from scipy.special import expit
4 import seaborn as sns
5 from matplotlib import pyplot as plt
6 from matplotlib import style
```



Introduction

Prediction and Inference

- Machine learning is often criticized as being only about “prediction” and sometimes “inference”
 - This isn’t quite true, but it provides a good starting point to ask what prediction really means
- “Inference” is used in different ways within ML and datascience
 - Sometimes the “point estimate” of some $\hat{f}(X)$ approximation even if we think $y = f(X) + \epsilon$ is the true model
 - Other times means the entire distribution of y given X (e.g., Bayesian inference) or some approximation around the mean with normal covariance (confidence intervals)

Thinking in Probabilities

- Prediction/Estimation/etc. are sometimes better interpreted with probability. If there was some true $f(\cdot)$ function,
- Take some X_1 and X_2 and want to find the distribution $y \sim \mathbb{P}(f(X_1, X_2) \mid X_2)$

Forecasts and Prediction

- The key becomes the distribution itself and what you can and can't condition on.
e.g. permissible X_2 values
- From this perspective, prediction is just an unconditional evaluation of the probability distribution, maybe the mean, a sample from it, or with confidence intervals - and not really special
 - The question is whether you have the right joint distribution!
- Forecasts typically just condition on the past observations, but could condition on future events
 - i.e., how might GDP grow if a tax cut is passed in 3 years

Counterfactuals: “What If?”

- Most interesting problems in economics are about counterfactuals in one way or another
 - What would have happened to the economy if the government had not intervened?
 - What would have been her income if she had not gone to college, or if she wasn't subjected to gender bias?
- By definition these are not observable. If we had the data we wouldn't need to ponder “What if?”. How? One way or another....

YOU HAVE TO MAKE \$HIT UP

The Role of Theory

- There is no data interpretation without some theory - even if it is sometimes implicit
- The role of both data and theory is then to help constrain the set of possible counterfactual
- So any criticisms of ML as “merely prediction” are basically a statement on whether the theory makes sense
 - i.e., if you fit $y = f(X) + \epsilon$ on data to find a $\hat{f}(X)$ function, then theory tells you if you made the right assumptions (e.g., that the X data is representative and wouldn't change for your counterfactual of interest, etc)
- Some models (e.g., random assignment) have easier to swallow assumptions than others.

Approaches

- Always remember: you need assumptions in one form or another because the counterfactuals are inherently not factual
- Broadly there are three approaches to conducting counterfactuals. They are not mutually exclusive
 1. Structural models: i.e. emphasize theory + data to put structure on the joint distribution of $\mathbb{P}(\mathbf{X}_1, \mathbf{X}_2)$
 2. Causal inference using matching, instrumental variables, etc. which use theoretical assumptions on independence to adjust for bias and missing latents
 3. Randomized Experiments/Treatment Effects where you can get good data which truly randomizes some sort of “treatment”.

Why do People Love Randomized Experiments?

- Because the assumptions are often easy to believe if you trust your random assignment
 - It often requires fewer assumptions beyond random assignment - for better or worse
- However:
 - They are not always possible, and even when they are, they are not always ethical
 - And even when possible and ethical, the inherent difficulty in randomization means it has limited scope and generalizability. i.e., you can learn an effect in one circumstance, but how common are those exact circumstances?

Potential Outcomes Framework

Treatments

- A coherent approach, which fits well with randomized trials, is to emphasize “treatment”. This means conditioning on binaries. Language/tools best thought of in terms of pharmaceutical trials
 - Call the value $T_i \in \{0, 1\}$ as the treatment
 - Let $Y_i(T_i)$ be the observed outcome
 - Let $Y_i(0)$ be the outcome if $T_i = 0$
 - Let $Y_i(1)$ be the outcome if $T_i = 1$
- The key: **you never get to see both**. One is always counterfactual

Potential Outcomes

- Many economic questions posed as: what would have happened if T_i was different for person i ? (or country i , etc)
- A “structural” model might be able to help answer that question, but might require a lot of assumptions on the underlying structure of i
- Alternatively, maybe we can make fewer (or different) assumptions and ask:
 - **Average Treatment Effect:** $\mathbb{E}[Y_i(1) - Y_i(0)]$
 - **Average Treatment Effect on the Treated:** $\mathbb{E}[Y_i(1) - Y_i(0) \mid T_i = 1]$
- Note here that we are taking expectations over the distribution of i . Hides lots of probability.

Potential Outcomes Framework

- The potential outcomes framework is a way to formalize causal inference
- It involves defining potential outcomes Y_{0i} and Y_{1i} for each unit under different treatment conditions
- The treatment variable T_i is a binary variable that indicates whether unit i receives the treatment ($T_i = 1$) or not ($T_i = 0$)
- The treatment effect on a unit of type i is the difference between the potential outcomes under different treatment conditions: $\tau_i = Y_{1i} - Y_{0i}$

Treatment Effects

- We are generally interested in treatment effects of the form $\tau_i = Y_{1i} - Y_{0i}$. However, we cannot observe *both* potential outcomes for a given unit. Instead, we can estimate
 - The **average treatment effect (ATE)**, which is the average of the treatment effects across all units: $\tau = E[Y_{1i} - Y_{0i}]$
 - The **average treatment effect on the treated (ATT)**, which is the average of the treatment effects for units that receive the treatment: $\tau_T = E[Y_{1i} - Y_{0i} | T_i = 1]$
- In randomized experiments, we can estimate the ATE and ATT using the difference in means between the treatment and control groups
 - Why is randomization important? To find out, we will move on to the second part of the course