Section 10: An Overview of Causal Inference[1]

April 4, 2012

[1] Credit to Brandon Stewart, Patrick Lam, and Maya Sen

Outline

1. Causal Inference is Hard
2. Classical Randomized Experiments
3. Observational Data
   - Matching
   - Instrumental Variables
   - Regression Discontinuity
   - Difference-in-Differences

Outline

1. Causal Inference is Hard
2. Classical Randomized Experiments
3. Observational Data
   - Matching
   - Instrumental Variables
   - Regression Discontinuity
   - Difference-in-Differences

Summary

Causal inference is hard due to the fundamental problem of causal inference. You have to make assumptions to gain leverage.

Fundamental Problem of Causal Inference.

The approach we will discuss is known as the *Rubin Causal Model* or the *Potential Outcomes framework*. Here's the terminology:

- Let $Y_i(0)$ be the outcome of interest under control – i.e., no treatment.
- Let $Y_i(1)$ be the outcome of interest under treatment.
- The **causal effect** we are after is $Y_i(1) - Y_i(0)$.
- And $Y_i(0)$ and $Y_i(1)$ are the **potential outcomes**.

Let's look at an example.

Fundamental Problem of Causal Inference (ctd).

In an ideal world, we would see this:

$Unit_i$	X_i^1	X_i^2	X_i^3	T_i	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
1	2	1	50	0	69	75	6
2	3	1	98	0	111	108	-3
3	2	2	80	1	92	102	10
4	3	1	98	1	112	111	-1

Fundamental Problem of Causal Inference (ctd).

But in the real world, we see this:

$Unit_i$	X_i^1	X_i^2	X_i^3	T_i	$Y_i(0)$	$Y_i(1)$	$Y_i(1) - Y_i(0)$
1	2	1	50	0	69	?	?
2	3	1	98	0	111	?	?
3	2	2	80	1	?	102	?
4	3	1	98	1	?	111	?

Fundamental Problem of Causal Inference (ctd).

- The **fundamental problem of causal inference** is that at most only one of the two potential outcomes $Y_i(0)$ or $Y_i(1)$ can be observed for each unit i.
- For control units, $Y_i(1)$ is the **counterfactual** (i.e., unobserved) potential outcome.
- For treatment units, $Y_i(0)$ is the **counterfactual**.
- For this reason, some people (including Don Rubin) call causal inference a *missing data* problem.

Differing Views of a Cause

Within the potential outcomes framework there is some debate over how to think about casual inference:

- Effects of Causes vs. Causes of Effects
- "No Causation without Manipulation"
- Envision the ideal experiment

Quantities of Interest

Before we talk about solutions, note that we might be interested in the following quantities of interest

- The individual treatment effect: $Y_i(1) - Y_i(0)$
- The average treatment effect (ATE): $E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$
- The treatment effect on the treated (ATT): $E[Y(1|T = 1) - Y(0|T = 1)] = E[Y_t(1) - Y_t(0)]$

At this point, all of them fail because of the fundamental problem.

Proposed Solutions

So here are some possible solutions

1. Randomized or Experimental Studies
2. Observational studies
    - Matching
    - Instrumental Variables
    - Regression Discontinuity Design
    - Difference-in-Differences

Estimating ATE

We can estimate the ATE in the following way:

$$\begin{aligned}\hat{\text{ATE}} &= E[Y_t(1) - Y_c(0)] \\ &= E[Y_t(1)] - E[Y_c(0)]\end{aligned}$$

Both quantities are observed.

We basically find the average Y for observations that received treatment and average Y for observations that received control.

But we have to make some assumptions

- SUTVA
- unconfoundedness/ignorability

Stable Unit Treatment Value Assumption

The **stable unit treatment value assumption (SUTVA)** assumes that

- the treatment status of any unit does not affect the potential outcomes of the other units (non-interference)
- the treatments for all units are comparable (no variation in treatment)

Violations:

- Job training for too many people may flood the market with qualified job applicants (interference)
- Some patients get extra-strength aspirin (variation in treatment)

Ignorability/Unconfoundedness

Unconfoundedness (strong ignorability):

$$(Y(1), Y(0)) \perp T$$

Treatment assignment is independent of the outcomes (Y).

Ignorability and Unconfoundedness are often used interchangeably. Technically, unconfoundedness is a stronger assumption. Most people just say ignorability.

Violations:

- Omitted Variable Bias

Outline

1. Causal Inference is Hard
2. Classical Randomized Experiments
3. Observational Data
   - Matching
   - Instrumental Variables
   - Regression Discontinuity
   - Difference-in-Differences

Classical Randomized Experiment

The gold standard of scientific research.

- The simplest way (conceptually) to compare treated and control units.
- Units are randomly assigned to receive treatment and control.
- We still can't estimate individual-level causal effects, but we can estimate the population **average treatment effect**, $E[Y_i(1) - Y_i(0)]$.
- Fisher's Fundamental Principles of Experimentation: Replication, Randomization, Blocking

Classical Randomized Experiments

Note again our assumptions:

1. **Ignorability.** There are no factors out there that affect both the probability of treatment and the outcome.
    - If the treatment was assigned in a truly random fashion, you are generally ok.
    - But if it wasn't, then you are in trouble!
2. **SUTVA.** We have assumed that assigning treatment to one unit doesn't affect the outcome for another unit.
    - Not always reasonable if there are potential peer effects!

.

Classical Randomized Experiments

We also need to think about compliance issues.

- Did the units "take" the treatment as they were supposed to?
  - Never-taker: Unit never takes treatment
  - Always-taker: Unit always takes treatment
  - Complier: Unit takes the treatment like they are supposed to
  - Defier: Unit takes treatment when not assigned and control when assigned
- Naive solution: Focus on "intent to treat" rather than on actual treatment.
- Another solution: Use an IV approach with the intent to treat as an instrument.

Other issues as well like missing data (measurement error and structural).

Outline

1. Causal Inference is Hard
2. Classical Randomized Experiments
3. Observational Data
   - Matching
   - Instrumental Variables
   - Regression Discontinuity
   - Difference-in-Differences

Observational Data

We have a dataset where we only observe after the experiment occurred and we have no control over treatment assignment.

This is the case with most of the sciences.

1. Gather dataset.
2. Estimate ATE or ATT with a model.

SUTVA: assumed (a problematic assumption most of the time)

Ignorability: include covariates to get conditional ignorability

$$(Y(1), Y(0)) \perp T | X$$

Treatment assignment is independent of the outcomes (Y) given covariates X.

Problems:

- SUTVA assumption
- Omitted variable bias
  - Don't include all the variables that makes treatment assignment independent of Y.
- Model Dependence
  - We try to alleviate the curse of dimensionality and problem of continuous covariates by specifying a model.
  - Estimates of ATE or ATT may differ depending on the model you specify.

Observational Studies

Here's how people have thought about this problem:

- What if we condition on every variable that might affect both T_i and Y_i?
- That way, it will appear for all intents and purposes that T_i has been randomly assigned.
- In plain English: We want to try to mimic a randomized study. We do so by conditioning on all possible confounding variables.
- Note: We still have to satisfy SUTVA, the stable unit treatment value assumption.

In all circumstances we will make (often unstated) assumptions to help generate the inference. These assumptions are often untestable *by definition.*

A Quick Aside for Emphasis

What do we do 99% of the time in the literature

We run a regression and interpret every coefficient

- What is the treatment? We can only interpret one coefficient causally.
- Post-Treatment Bias
- SUTVA

We are usually estimating the conditional mean, not the causal effect. **Like I said, causal inference is hard!**

Matching: A way to Ameliorate Model Dependence

- If we had pairs of observations that had the exact same covariate values (perfect **balance**) and differed only on treatment assignment, then we would have perfect conditional ignorability – i.e., conditional on the covariates, the treatment is independent of the outcome.
- Then we will get the same results regardless of the model.
- Matching is a method of trying to achieve better balance on covariates and reduce model dependence. Goal: Balance on covariates
- But remember: Don't match on post-treatment variables!

Synthetic Controls

Basic Idea: Assume that we can build matched-controls by weighting observations.

- Implemented in `R` and `MATLAB` in the `Synth` package
- Work by Abadie, Diamond and Hainmueller
- Key Paper: "The Economic Costs of Conflict: A Case Study of the Basque Country"

However: You essentially have an n of 1.

Synthetic Controls (Basque Country)

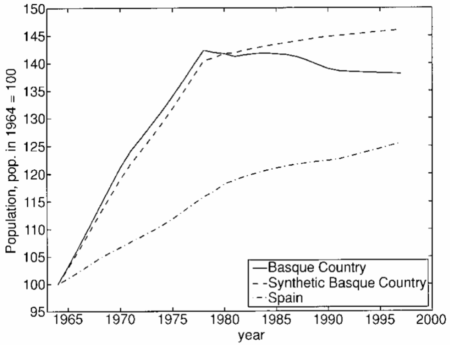


Figure: An Example from Abadie & Gareazabal (2003)

Synthetic Controls (California)

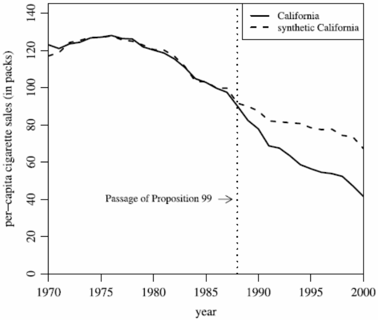


Figure: An Example from Abadie, Diamond and Hainmueller (2010)

Instrumental Variables

Goal: Estimate Causal Effects

Problem in Observational Data: Non-ignorability of treatment assignment (and SUTVA)

Solution so far: Include covariates and match

Another solution: **Instrumental Variables**

The idea: Find an instrument Z that is randomly assigned (or assignment is ignorable) and that affects Y only through T.

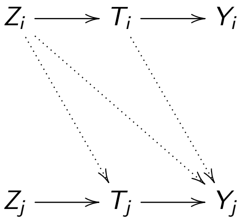
Example: $Y =$ post-Vietnam War civilian mortality; $T =$ serving in the military during Vietnam War; $Z =$ draft lottery

The Potential Outcomes Approach

Assumptions:

1. SUTVA: Z_i does not affect T_j and Y_j and T_i does not affect Y_j for all $i \neq j$ (non-interference) and there is no variation in the treatment or the instrument.

Figure: SUTVA Assumption implies absence of dotted arrows.



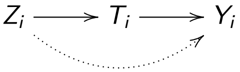
Example: The veteran status of any man at risk of being drafted in the lottery was not affected by the draft status of others at risk of being drafted, and, similarly, that the civilian mortality of any such man

2. Random (Ignorable) Assignment of the Instrument Z

Example: Assignment of draft status was random.

3. Exclusion Restriction: Any effect of Z on Y must be via an effect of Z on T.

Figure: Exclusion assumption implies absence of dotted arrow.



Example: Civilian mortality risk was not affected by draft status once veteran status is taken into account.

4. Nonzero Average Causal Effect of Z on T.

Example: Having a low lottery number increases the average probability of service.

5. Monotonicity: No Defiers

Example: There is no one who would have served if given a high lottery number, but not if given a low lottery number.

Local Average Treatment Effect Among Compliers

Never-Takers

$$T_i(1) - T_i(0) = 0$$

$$Y_i(1,0) - Y_i(0,0) = 0$$

By Exclusion Restriction, causal effect of Z on Y is zero

Defier

$$T_i(1) - T_i(0) = -1$$

$$Y_i(1,0) - Y_i(0,1) = Y_i(0) - Y_i(1)$$

By Monotonicity, no one in this group

Complier

$$T_i(1) - T_i(0) = 1$$

$$Y_i(1,0) - Y_i(0,1) = Y_i(1) - Y_i(0)$$

Average Treatment Effect among Compliers

Always-taker

$$T_i(1) - T_i(0) = 0$$

$$Y_i(1,1) - Y_i(0,1) = 0$$

By Exclusion Restriction, causal effect of Z on Y is zero

If all the assumptions hold, then the **Local Average Treatment Effect (LATE)** of T on Y is

$$\text{LATE} = \frac{\text{Effect of Z on Y}}{\text{Effect of Z on T}}$$

It is only a local average treatment effect because it's the effect of T on Y for the subpopulation of compliers, and not the whole population.

A Word of Caution: Be wary of people who use the same instrument for everything!

Regression Discontinuity Design

Sometimes called "natural experiments." Here's how it works:

- There is some pre-treatment variable x with a cutoff value – below the cut-off value the unit receives treatment and above the cutoff it receives control.
- The cutoff should be arbitrary and not related to the other covariates.
- There should be no real reason to think that the units below the cut-off are substantively that different from the units above the cut-off.
- We basically take care of ignorability. (Note: you still have to worry about SUTVA.)

Regression Discontinuity

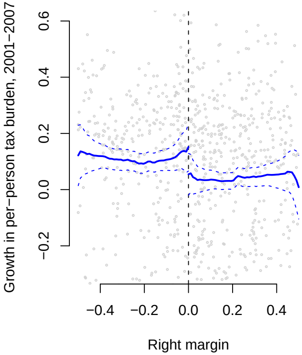
Here are some examples:

- Criminals who are 17 yrs old, 364 days are treated as juveniles, but those who are 18yrs old, 1 day are treated as adults. What is the effect of adult sentencing?
- European municipalities with populations of 3499 do not need to have dissenting party members sit on their councils, but those with populations of 3500 do. What is the effect of having dissenting party members on council decisions?
- HS standardized test takers who get a score of 90.5 get a small scholarship, but those that get 89.5 don't. What is the effect of the scholarship on college success?

It's sometimes easier to see this visually.

Regression Discontinuity (Eggers)

Example: In some French cities, the rightist party wins with 50.5 of the vote%; in other cities, they lose with 49.5% of the vote. What is the effect of rightist governments on tax policies?



Regression Discontinuity

Some thoughts about RDD designs:

- Always pay attention to the bandwidth size...
- If you are working on elections: "Are Close Elections Random?" Grimmer, Hersh, Feinstein, Carpenter (2011)
- Remember who you are generalizing to!

Difference-in-Differences

The Idea: We can use two units observed at two time points.

- Imagine two sets of units (treated and control, $T = 1, 0$) each observed at two time points ($D = 1, 2$)
- There is some intervention on the treated units between the two time points but not the controls
- Take First Differences and Compare

$$(E[Y(D=2)|T=1] - E[Y(D=2)|T=0]) - (E[Y(D=1)|T=1] - E[Y(D=1)|T=0])$$

Similarity is critical! E.g. Card and Krueger on the minimum wage (1994)

Conclusions

- There are a lot of causal inference techniques out there.
- Remember to ask yourself: where is the leverage coming from? Is it from a unique feature of the data? Or a heroic assumption?
- Causal Inference is hard: but it is worth doing!
- What we haven't covered: DAGs, SEMs, Selection Models etc.

References

Angrist, Joshua D., Guido W. Imbens and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434):444-455.

Angrist, Joshua D. and Jörn Steffen Pischke. *Mostly Harmless Econometrics.* Princeton University Press.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396): 945-960.

Pearl, Judea. *Causality: models, reasoning, and inference*.