Potential Outcomes

Kosuke Imai

Harvard University

Spring 2021

Goals of this Lecture

1. Define various causal effects
2. Introduce the fundamental problem of causal inference
3. Formalize the basic assumptions underlying causal effects
4. Understand how principal stratification addresses the truncation-by-death problem

A Study of Political Canvassing

- The study has n voters
  - n_1 voters are canvassed
  - $n_0 = n - n_1$ voters are not canvassed

- The study may be experimental or observational

- For each voter $i = 1, 2, \ldots, n$, we observe the following:
  - Treatment variable:

$$T_i = \begin{cases} 1 & \text{if treated (canvassed)} \\ 0 & \text{if control (not canvassed)} \end{cases}$$

  - Outcome variable (turnout): Y_i
  - Pre-treatment covariates (voter characteristics): $\mathbf{X}_i$

- Does the canvassing causally affect turnout?

Defining Causal Effects

- Potential outcomes
  - $Y_i(1)$: outcome under the treatment condition
  - $Y_i(0)$: outcome under the control condition
- Relationship between the observed and potential outcomes

$$Y_i = Y_i(T_i)$$

- Causal effect for voter i:

$$\tau_i \ = \ Y_i(1) - Y_i(0)$$

Voters	Age	Gender	Canvassed	Turnout		Causal effect
i	X_{i1}	X_{i2}	T_i	$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$
1	40	M	1	1	?1	?0
2	55	F	0	?1	0	?1
3	20	F	0	?0	1	?−1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	62	M	1	0	?0	?0

Fundamental Problem of Causal Inference

- We only observe one potential outcome for each unit
  - Yet, any causal quantity is a function of two potential outcomes

$$Y_i(1) - Y_i(0),\ \log Y_i(1) - \log Y_i(0),\ \frac{Y_i(1)}{Y_i(0)},\ \frac{Y_i(1) - Y_i(0)}{Y_i(0)} \times 100,\ etc.$$

  - Only marginal, not joint, distribution is identifiable
- Non-binary treatment:
  - categorical: $Y_i(t)$ for $t = 0, 1, \ldots, K-1$
  - continuous: $Y_i(t)$ for any $t \in \mathbb{R}$
  - multivariate: $Y_i(t_1, \ldots, t_K)$ for $t_k \in \mathcal{T}_k$
- Causal inference as a severe missing data problem
- Potential outcomes can be thought as fixed for a given unit
  - potential outcomes as characteristics of units
  - potential outcomes do have a distribution across units
  - treatment variable determines which potential outcome is observed
  - observed outcomes are random because the treatment is random

The Key Assumptions

1. Causal ordering:

$$T_i \longrightarrow Y_i$$

   - no reverse causality: $T_i \not\longleftarrow Y_i$
   - no simultaneity: $T_i \not\longleftrightarrow Y_i$

2. Consistency:

$$Y_i = Y_i(t) \text{ whenever } T_i = t$$

   - no hidden multiple versions of treatment
   - no hidden different administration of treatment
   - we can redefine treatment to satisfy this assumption

3. No interference between units:

$$Y_i(T_1, T_2, \ldots, T_n) \ = \ Y_i(T_i)$$

   - treatments of no other units affect one's outcome

Interference between Units

- Interference within a two-voter household
  - both treated: $Y_{me}(T_{me} = 1, T_{partner} = 1)$
  - neither treated: $Y_{me}(T_{me} = 0, T_{partner} = 0)$
  - one treated: $Y_{me}(T_{me} = 1, T_{partner} = 0)$, $Y_{me}(T_{me} = 0, T_{partner} = 1)$
- Within-household spillover effects:

$$Y_{me}(T_{me} = 0, T_{partner} = 1) - Y_{me}(T_{me} = 0, T_{partner} = 0)$$
$$Y_{me}(T_{me} = 1, T_{partner} = 1) - Y_{me}(T_{me} = 1, T_{partner} = 0)$$

- Direct effects:

$$Y_{me}(T_{me} = 1, T_{partner} = 0) - Y_{me}(T_{me} = 0, T_{partner} = 0)$$
$$Y_{me}(T_{me} = 1, T_{partner} = 1) - Y_{me}(T_{me} = 0, T_{partner} = 1)$$

- General definitions

$$\text{spillover effects}: \quad Y_i(T_i = t, \mathbf{T}_{-i} = \mathbf{t}) - Y_i(T_i = t, \mathbf{T}_{-i} = \mathbf{t}')$$
$$\text{direct effects}: \quad Y_i(T_i = 1, \mathbf{T}_{-i} = \mathbf{t}) - Y_i(T_i = 0, \mathbf{T}_{-i} = \mathbf{t})$$

Causal Effects of Immutable Characteristics

- "No causation without manipulation" (Holland, 1986. *J. Amer. Stat. Assoc*)
- Immutable characteristics or attributes: gender, race, age, etc.
- Can immutable characteristics have meaningful causal effects?
- Strategies:
  1. Causal effects of perceived characteristics:
     - Causal effect of a job applicant's gender/race on call-back rates (Bertrand and Mullainathan, 2004. *Am. Econ. Rev*)
  2. Reinterpretation:
     - Causal effect of having a female politician on policy outcomes (Chattopadhyay and Duflo, 2004. *Q. J. Econ*)
  3. Redefinition:
     - Race as a "bundle of sticks": skin color, neighborhood, socio-economic status, etc. (Sen and Wasow, 2016. *Annu. Rev. Polit. Sci*)
  4. Group-level intervention:
     - Would racial disparity go away if we equalize socio-economic status of blacks and whites? (VanderWeele and Robinson, 2014. *Epidemiology*)

Average Treatment Effects

- Unit causal effects are difficult to estimate
- We can average them over a sample of units
  1. sample average treatment effect:

  $$\text{SATE} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}$$

     $\rightsquigarrow$ difference in the average outcome between two scenarios: everyone is treated vs. nobody is treated
  2. sample average treatment effect for the treated

  $$\text{SATT} = \frac{1}{n_1}\sum_{i=1}^{n} T_i\{Y_i(1) - Y_i(0)\} = \frac{1}{n_1}\sum_{i=1}^{n} T_i\{Y_i - Y_i(0)\}$$

- Population average treatment effects:

$$\begin{aligned} \text{PATE} &= \mathbb{E}(Y_i(1) - Y_i(0)) \\ \text{PATT} &= \mathbb{E}(Y_i(1) - Y_i(0) \mid T_i = 1) = \mathbb{E}(Y_i - Y_i(0) \mid T_i = 1) \end{aligned}$$

Other Common Causal Quantities of Interest

- Heterogenous effects:
  - Conditional average treatment effect (CATE)

$$\tau(\mathbf{x}) = \mathbb{E}(Y_i(1) - Y_i(0) \mid \mathbf{X}_i = \mathbf{x})$$

  - Applications to precision medicine and micro-targeting
  - Do not confuse it with the individual treatment effect $Y_i(1) - Y_i(0)$

- Non-additive effects:
  - Quantile treatment effects
    - example: $\text{median}(Y_i(1)) - \text{median}(Y_i(0))$
    - it does not necessarily equal $\text{median}(Y_i(1) - Y_i(0))$
  - Odds ratio

$$\frac{\Pr(Y_i(1) = 1) / \Pr(Y_i(1) = 0)}{\Pr(Y_i(0) = 1) / \Pr(Y_i(0) = 0)}$$

Truncation by Death (Zhang and Rubin, 2003. *J. Educ. Behav. Stat.*)

- Setup
  1. Units: patients
  2. Treatment: new medicine
  3. Outcome: cholesterol level
  4. Truncation: patient death
- Truncation by death problem:
  - cholesterol level undefined for the dead
  - survivors in the treatment group are not comparable to those in the control group
  - Post-treatment bias: treatment may affect survival
  - If the treatment saves the lives of people with high cholesterol, it may appear that the treatment increases cholesterol
- In general, one should not adjust for post-treatment variables
- Other examples:
  - drop-out in program evaluation
  - registration, turnout, and vote choice in get-out-the-vote studies

Principal Stratification (Frangakis and Rubin, 2002. *Biometrics*)

- Observed data
  - Binary treatment: $T_i \in \{0, 1\}$
  - Survival variable: $W_i \in \{0, 1\}$
  - Observed outcome: Y_i only when $W_i = 1$
- Potential variables
  - Potential truncation variable: $W_i(1), W_i(0)$
  - Potential outcomes: $Y_i(t, w) \rightsquigarrow Y_i(0, 0)$ and $Y_i(1, 0)$ do not exist
- Four principal strata defined by $(W_i(0), W_i(1))$
  1. (1, 1): always survivor
  2. (0, 0): always non-survivor
  3. (0, 1): survivor only when treated
  4. (1, 0): survivor only when untreated
- We do not know which principal stratum each patient belongs to
- Causal effect is only defined for always-survivors
  $\mathbb{E}(Y(1) - Y(0) \mid W(1) = W(0) = 1)$

Summary

- Causal effects $=$ function of potential outcomes
- Fundamental problem of causal inference
  - only one potential outcome is observed
  - causal inference as a missing data problem
- Basic assumptions
  1. causal ordering
  2. consistency
  3. no interference between units
- Causal effects
  - individual effects, average effects
  - spillover effects, direct effects
- Principal stratification
  - “types” of units based on potential outcomes
  - wide applicability $\rightsquigarrow$ truncation by death