Potential Outcomes and Causal Estimands

Yuki Shiraito

POLSCI 699 Statistical Methods in Political Research II
University of Michigan

Counterfactuals

- "Correlation does not imply causation"—but what is causation?
- One (not the only one) conceptualization of causation: counterfactuals
    - "What would have happened if…"
    - "What would have been the US presidential election outcome if the Democratic Party had nominated Bernie Sanders instead of Hillary Clinton?"
    - "What would have been Saudi Arabia's political regime if oil had not existed?"
    - "What would have happend to the Russo-Ukranian conflict in 2022 if Ukraine had joined NATO?"
- Specific to:
    1. unit
    2. counterfactual scenario
- Fundamental problem of causal inference
    - Impossible to observe the outcome that did not happen

Formalization: Potential Outcomes

- Voter turnout with/without a get-out-the-vote (GOTV) message
- Units (= voters): $i = 1, \ldots, n$
- "Treatment": $T_i = 1$ if treated, $T_i = 0$ otherwise
- *Observed* outcome: Y_i
- Pre-treatment covariates: X_i
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$ where $Y_i = Y_i(T_i)$

Voters	Contact	Turnout		Age	Party ID
i	T_i	$Y_i(1)$	$Y_i(0)$	X_i	X_i
1	1	1	?	20	D
2	0	?	0	55	R
3	0	?	1	40	R
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	1	0	?	62	D

- (Individual) Causal effect: $Y_i(1) - Y_i(0)$

Key Assumptions

- The notation $Y_i(t)$ implies three assumptions:
  1. No simultaneity (different from endogeneity): $T_i = T_i(Y_i)$ for any Y_i
  2. No interference between units: $Y_i(T_1, T_2, \ldots, T_n) = Y_i(T_i)$
  3. Same version of the treatment: $Y_i(T_i) = Y_i(t)$ whenever $T_i = t$
- Stable Unit Treatment Value Assumption (SUTVA)
- Examples of SUTVA violations:
  1. feedback effects
  2. spill-over effects, carry-over effects
  3. different treatment administration
- Multi-valued treatment: more potential outcomes for each unit

- Randomness in statistical causal inference
  - Potential outcome for each unit $(Y_i(0), Y_i(1))$ is "fixed"; data cannot distinguish fixed and random potential outcomes
  - Observed outcome for each unit $Y_i = Y_i(T_i)$ is random because the treatment is random
  - Potential outcomes across units have a joint distribution of $(Y_i(0), Y_i(1))$: randomness from sampling

Manipulation of the Treatment

- "No causation without manipulation" (Holland, 1986)
- Medical trials: technically feasible to manipulate dosage, vaccination, etc.
- Social science: infeasible to manipulate immutable characteristics such as gender, race, age, etc.
- What does the causal effect of gender mean?

- Causal effect of having a female politician on policy outcomes (Chattopadhyay and Duflo, 2004 *QJE*)
- Causal effect of having a discussion leader with certain preferences on deliberation outcomes (Humphreys *et al*. 2006 *WP*)
- Causal effect of a job applicant's gender/race on call-back rates (Bertrand and Mullainathan, 2004 *AER*)

- Problem: confounding

Common Causal Estimands

- Individual effects are never observed $\rightsquigarrow$ average effects of the common treatment across units
- Sample Average Treatment Effect:

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

  - Sample average of individual causal effects
  - Still unobservable due to missing data of potential outcomes
- Population Average Treatment Effect:

$$\text{PATE} \equiv \mathbb{E}[Y_i(1) - Y_i(0)]$$

  - Population average of individual causal effects
  - Unobservable due to both missingness and sampling
- Population Average Treatment Effect for the Treated:

$$\text{PATT} \equiv \mathbb{E}[Y_i(1) - Y_i(0) \mid T_i = 1]$$

  - Conditional population average given treated
  - Often used policy evaluation

Problems

- Causal identification—how to unbiasedly or consistently estimate:
$$\frac{1}{n}\sum_{i=1}^{n} Y_i(t) \text{ for } i \text{ s.t. } T_i \neq t$$
$$\mathbb{E}[Y_i(t)] \text{ for all } t$$
$$\mathbb{E}[Y_i(0) \mid T_i = 1]$$
- Equivalent to predicting missing outcome data
- Finding identifiable estimand (e.g., principal effects)
- Treatment effect heterogeneity:
$$Y_i(1) - Y_i(0) = 0 \,\forall i \implies \text{ATE} = 0$$
$$\text{ATE} = 0 \;\not\Longrightarrow\; Y_i(1) - Y_i(0) = 0 \,\forall i$$
- Conditional average treatment effect:
$$\mathbb{E}\big[Y_i(1) - Y_i(0) \mid X\big] = \mathbb{E}\big[Y_i(1) \mid X\big] - \mathbb{E}\big[Y_i(0) \mid X\big]$$