Identification and Inference for Randomized Experiments

Yuki Shiraito

POLSCI 699 Statistical Methods in Political Research II

University of Michigan

Treatment Assignment and Observed Outcomes

- *Observed* outcomes $\{Y_i\}_{i=1}^n$ depends on treatment assignment
- Example with $n = 200$
  - Potential outcomes:
    1. $(Y_i(0), Y_i(1)) = (1, 1)$ for $i = 1, \ldots, 100$
    2. $(Y_i(0), Y_i(1)) = (0, 0)$ for $i = 101, \ldots, 200$
  - If
    1. $T_i = 1$ for $i = 1, \ldots, 100$
    2. $T_i = 0$ for $i = 101, \ldots, 200$

    then

    1. $Y_i = 1$ for the treated
    2. $Y_i = 0$ for the control
  - If
    1. $T_i = 0$ for $i = 1, \ldots, 100$
    2. $T_i = 1$ for $i = 101, \ldots, 200$

    then

    1. $Y_i = 0$ for the treated
    2. $Y_i = 1$ for the control
- "Correlation does not imply causation"

Completely Randomized Experiments

- Setup:
  1. Random sample of size n from a superpopulation
  2. Binary treatment $T_i \in \{0, 1\}$
  3. Pretreatment covariate vector X_i may be observed
- Randomized experiments:
  1. $p_i \equiv \mathbb{P}(T_i = 1) \in (0, 1)$
  2. Researcher sets p_i
- Assignment mechanism: joint distribution of the treatment, i.e.,
$$p(\mathrm{T} \mid \mathbf{X}, \mathrm{Y}(0), \mathrm{Y}(1)) \text{ where } \mathrm{T} \equiv (T_1, T_2, \ldots, T_n)^\top$$
- Complete randomization: with fixed n_1,
$$p(\mathrm{T} \mid \mathbf{X}, \mathrm{Y}(0), \mathrm{Y}(1)) = \begin{cases} \binom{n}{n_1}^{-1} & \text{if } \sum_{i=1}^{n} T_i = n_1 \\ 0 & \text{otherwise} \end{cases}$$
  - Unconfounded: $p(\mathrm{T} \mid \mathbf{X}, \mathrm{Y}(0), \mathrm{Y}(1)) = p(\mathrm{T} \mid \mathbf{X})$ for any $\mathrm{Y}(0), \mathrm{Y}(1)$
- Difference-in-means estimator:
$$\hat{\tau} \equiv \frac{1}{n_1}\sum_{i=1}^{n} T_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - T_i) Y_i \text{ where } n_0 = n - n_1$$

Unbiased Estimation of SATE

- Key idea (Neyman 1923): Randomness comes from treatment assignment (plus sampling for PATE) alone
- Design-based (randomization-based) rather than model-based
- Statistical properties of $\hat{\tau}$ based on design features

- Define $\mathcal{O} \equiv \{Y_i(0), Y_i(1)\}_{i=1}^n$
- Within sample, randomness of T conditional on $\mathcal{O}$
- Unbiasedness (over repeated treatment assignments):

$$\begin{aligned}
\mathbb{E}(\hat{\tau} \mid \mathcal{O}) =& \frac{1}{n_1}\sum_{i=1}^n \mathbb{E}(T_i \mid \mathcal{O})Y_i(1) - \frac{1}{n_0}\sum_{i=1}^n \{1 - \mathbb{E}(T_i \mid \mathcal{O})\}Y_i(0) \\
=& \frac{1}{n}\sum_{i=1}^n (Y_i(1) - Y_i(0)) \\
=& \text{SATE}
\end{aligned}$$

Randomization Inference for SATE

- Variance of $\hat{\tau}$:
$$\mathbb{V}(\hat{\tau} \mid \mathcal{O}) = \frac{1}{n}\left(\frac{n_0}{n_1}S_1^2 + \frac{n_1}{n_0}S_0^2 + 2S_{01}\right),$$
where for $t = 0, 1$,
$$S_t^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(t) - \overline{Y(t)})^2 \quad \text{sample variance of } Y_i(t)$$
$$S_{01} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(0) - \overline{Y(0)})(Y_i(1) - \overline{Y(1)}) \quad \text{sample covariance}$$
- Derivation: Adam's section

- S_{01} is *not identifiable*: cannot be estimated even with infinite amount of data
- Therefore $\mathbb{V}(\hat{\tau} \mid \mathcal{O})$ is not identifiable

Details of Variance Derivation

1. Let $Z_i = Y_i(1) + n_1 Y_i(0)/n_0$ and $D_i = nT_i/n_1 - 1$, and write
$$\mathbb{V}(\hat{\tau} \mid \mathcal{O}) = \frac{1}{n^2}\,\mathbb{E}\left\{\left(\sum_{i=1}^{n} D_i Z_i\right)^2 \,\middle|\, \mathcal{O}\right\}$$

2. Show
$$\mathbb{E}(D_i \mid \mathcal{O}) = 0, \quad \mathbb{E}(D_i^2 \mid \mathcal{O}) = \frac{n_0}{n_1}, \quad \mathbb{E}(D_i D_j \mid \mathcal{O}) = -\frac{n_0}{n_1(n-1)}$$

3. Use the above to show,
$$\mathbb{V}(\hat{\tau} \mid \mathcal{O}) = \frac{n_0}{n(n-1)n_1}\sum_{i=1}^{n}(Z_i - \overline{Z})^2$$

4. Substitute the potential outcome expressions for Z_i

Sharp Bounds on the Variance

- Cauchy-Shwartz inequality:

$$S_{01}^2 \leq S_1^2 S_0^2 \implies -S_1 S_0 \leq S_{01} \leq S_1 S_0 \text{ where } S_t = \sqrt{S_t^2}$$

- Sharp bounds on $\mathbb{V}(\hat{\tau} \mid \mathcal{O})$:

$$\frac{n_0 n_1}{n}\left(\frac{S_1}{n_1} - \frac{S_0}{n_0}\right)^2 \leq \mathbb{V}(\hat{\tau} \mid \mathcal{O}) \leq \frac{n_0 n_1}{n}\left(\frac{S_1}{n_1} + \frac{S_0}{n_0}\right)^2$$

- The upper bound when $\frac{S_{01}}{S_1 S_0} = 1$
- The lower bound when $\frac{S_{01}}{S_1 S_0} = -1$

- Under the constant additive unit causal effect assumption, i.e., $Y_i(1) - Y_i(0) = c$ for all i,

$$S_1^2 = S_0^2 = S_{01}$$

and letting $S^2 \equiv S_1^2 = S_0^2 = S_{01}$,

$$\mathbb{V}(\hat{\tau} \mid \mathcal{O}) = \frac{S^2}{n_1} + \frac{S^2}{n_0}$$

Estimation of the Sample Variance

- S_t is a function of $Y_i(t)$ of all i, hence unknown
- The usual variance estimator is conservative on average:
$$\mathbb{V}(\hat{\tau} \mid \mathcal{O}) \leq \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} = \mathbb{E}\left[\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0} \mid \mathcal{O}\right]$$
where
$$\hat{\sigma}_t^2 \equiv \frac{1}{n_t - 1} \sum_{i=1}^{n} 1\{T_i = t\}(Y_i - \bar{Y}_t)^2$$
for $t = 0, 1$
- Unbiased variance estimator under the constant additive effect assumption:
$$\widehat{\mathbb{V}(\hat{\tau} \mid \mathcal{O})} = \frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0} \text{ where } \mathbb{E}\left[\widehat{\mathbb{V}(\hat{\tau} \mid \mathcal{O})} \mid \mathcal{O}\right] = \mathbb{V}(\hat{\tau} \mid \mathcal{O})$$

Randomization Inference for PATE

- Randomness from sampling $\rightsquigarrow \mathcal{O}$ is r.v.
- Complete randomization implies strong ignorability: for all i,

$$\big(Y_i(0), Y_i(1)\big) \perp\!\!\!\perp T_i$$

- Unbiasedness (over repeated sampling and treatment assignment):

$$\begin{aligned}\mathbb{E}\big[\mathbb{E}[\hat{\tau} \mid \mathcal{O}]\big] &= \mathbb{E}[\text{SATE}] \\ &= \mathbb{E}\big[Y_i(1) - Y_i(0)\big] = \text{PATE}\end{aligned}$$

- Variance:

$$\begin{aligned}\mathbb{V}(\hat{\tau}) &= \mathbb{V}(\mathbb{E}(\hat{\tau} \mid \mathcal{O})) + \mathbb{E}(\mathbb{V}(\hat{\tau} \mid \mathcal{O})) \\ &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}\end{aligned}$$

  where $\sigma_t^2 \equiv \mathbb{V}(Y_i(t))$ for $t = 0, 1$
- Unbiased variance estimator:

$$\widehat{\mathbb{V}(\hat{\tau})} = \frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0} \quad \text{where} \quad \mathbb{E}\left[\widehat{\mathbb{V}(\hat{\tau})}\right] = \mathbb{V}(\hat{\tau})$$

Asymptotic Inference for PATE

- Hold $k = n_1/n$ constant
- Rewrite the difference-in-means estimator as

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\left(\frac{T_i Y_i(1)}{k} - \frac{(1-T_i)Y_i(0)}{1-k}\right)}_{\text{i.i.d. with mean PATE \& variance } n\mathbb{V}(\hat{\tau})}$$

- Consistency:

$$\hat{\tau} \xrightarrow{p} \text{PATE}$$

- Asymptotic normality via the Central Limit Theorem (CLT):

$$\frac{\hat{\tau} - \text{PATE}}{\sqrt{\sigma_1^2/n_1 + \sigma_0^2/n_0}} \xrightarrow{d} \mathcal{N}(0,1)$$

Two-Sample Test

- $H_0 : \text{PATE} = \tau_0$ and $H_1 : \text{PATE} \neq \tau_0$
- Often $\tau_0 = 0$
- Difference-in-means estimator: $\hat{\tau}$
- Asymptotic reference distribution:

$$Z\text{-statistic} = \frac{\hat{\tau} - \tau_0}{\text{s.e.}} = \frac{\hat{\tau} - \tau_0}{\sqrt{\hat{\sigma}_1^2/n_1 + \hat{\sigma}_0^2/n_0}} \xrightarrow{d} \mathcal{N}(0, 1)$$

- $(1 - \alpha) \times 100\%$ Confidence intervals:

$$[\hat{\tau} - \text{s.e.} \times z_{\alpha/2},\ \hat{\tau} + \text{s.e.} \times z_{\alpha/2}]$$

- Is Z_{obs} unusual under the null?
    - Reject the null when $|Z_{obs}| > z_{1-\alpha/2}$
    - Retain the null when $|Z_{obs}| \leq z_{1-\alpha/2}$

Error and Power of Hypothesis Test

- Two types of errors:

	Reject H_0	Retain H_0
H_0 is true	Type I error	Correct
H_0 is false	Correct	Type II error

- Size (level) of test: probability of Type I error
- Hypothesis tests control the level
- They do not control the probability of Type II error

- Tradeoff between the two types of error
- Power of test: probability that a test rejects the null
- Typically, we want a most powerful test with the proper size

Power Analysis

- Null hypotheses are often uninteresting
- But, hypothesis testing may indicate the strength of evidence for or against your theory
- Power analysis: What sample size do I need in order to detect a certain departure from the null?
- Power $= 1 - \Pr(\text{Type II error})$
- Four steps:
  1. Specify the null hypothesis and the significance level a
  2. Choose a true value for the parameter of interest and derive the sampling distribution of test statistic
  3. Calculate the probability of rejecting the null hypothesis under this sampling distribution
  4. Find the smallest sample size such that this rejection probability equals a prespecified level

One-Sided Test Example

- $H_0 : \tau = \pi$ and $H_0 : \tau > \pi$
- Reject H_0 if $\hat{\tau} > \pi + z_{\alpha/2} \times \sqrt{\sigma_1^2/n_1 + \sigma_0^2/n_0}$

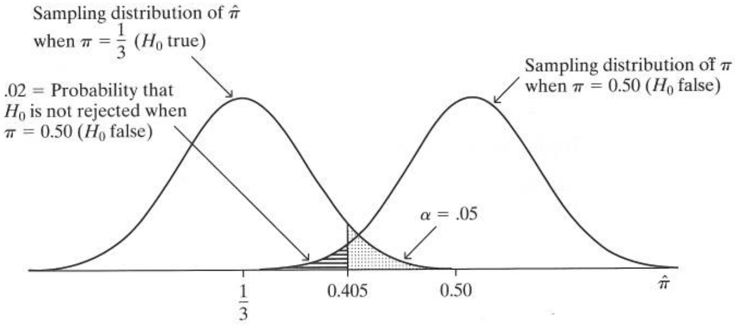

FIGURE 6.11: Calculation of P(Type II Error) for Testing H_0: $\pi = 1/3$ against H_a: $\pi > 1/3$ at $\alpha = 0.05$ Level, when True Proportion is $\pi = 0.50$. A Type II error occurs if $\hat{\pi} < 0.405$, since then P-value >0.05 even though H_0 is false.

Power Function ($\sigma_0^2 = \sigma_1^2 = 1$ and $n_1 = n_0$)