Causal Inference (Imbens, Rubin), Ch.19

A General Method for Estimating Sampling Variances for Standard Estimators for Average Causal Effects

presented by Jihu Lee

IDEA Lab
Department of Statistics
Seoul National University

August 16, 2022

Table of Contents

1. Introduction
2. Estimands
3. The Common Structure of Standard Estimators
4. A General Formula for the Conditional Sampling Variance
5. A Simple Estimator for the Unit-level Conditional Sampling Variance
6. An Estimator for the Sampling Variance for the Population ATE
7. Alternative Estimators for the Sampling Variance

Introduction

- Estimating sampling variances are important so that we can construct large-sample CIs

- Two issues
  1. Choice of estimand: average effect in the sample / super-population
  2. Choice of estimator: for specific method / more general way

Table of Contents

1. Introduction
2. Estimands
3. The Common Structure of Standard Estimators
4. A General Formula for the Conditional Sampling Variance
5. A Simple Estimator for the Unit-level Conditional Sampling Variance
6. An Estimator for the Sampling Variance for the Population ATE
7. Alternative Estimators for the Sampling Variance

Estimands

- Recall: Notations

$$\tau_{\mathsf{fs}} = \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)),\ \tau_{\mathsf{sp}} = \mathbb{E}_{\mathsf{sp}}[Y_i(1) - Y_i(0)] = \mathbb{E}_{\mathsf{sp}}[\tau_{\mathsf{fs}}]$$

$$\mathbb{V}_W(\hat{\tau}) = \mathbb{E}_W[(\hat{\tau} - \tau_{\mathsf{fs}})^2],\ \mathbb{V}(\hat{\tau}) = \mathbb{E}[(\hat{\tau} - \tau_{\mathsf{sp}})^2]$$

- Expectations and variances without "W" or "sp": taken over both the assignment and random sampling from the super-population
- The approximate difference will given as

$$\mathbb{V}(\hat{\tau}) - \mathbb{V}_W(\hat{\tau}) \approx \frac{\mathbb{V}_{\mathsf{sp}}(\tau(X_i))}{N} \tag{1}$$

Brief explanations for (1)

- **Simple example**: single pre-treatment variable - sex: $X_i \in \{f, m\}$
- $N(f), N(m)$: number of females and males

$$N_c(x) = \sum_{i:X_i=x} (1 - W_i),\ N_t(x) = \sum_{i:X_i=x} W_i,\ N(x) = N_c(x) + N_t(x)$$

$$\overline{Y}_c^{\text{obs}}(x) = \frac{1}{N_c(x)} \sum_{i:X_i=x} (1 - W_i) \cdot Y_i^{\text{obs}},\ \overline{Y}_t^{\text{obs}}(x) = \frac{1}{N_c(x)} \sum_{i:X_i=x} W_i \cdot Y_i^{\text{obs}}$$

$$\tau_{\text{fs}}(x) = \frac{1}{N(x)} \sum_{i:X_i=x} (Y_i(1) - Y_i(0))$$

$$\tau_{\text{sp}}(x) = \mathbb{E}_{\text{sp}}[Y_i(1) - Y_i(0) | X_i = x]$$

Brief explanations for (1)

- Suppose that treatment assignment is super-population unconfounded

$$W_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) | X_i$$

$$N_c(f), N_c(m), N_t(f), N_t(m) > 0$$

- Natural estimators:

$$\hat{\tau}^{\text{dif}}(x) = \bar{Y}_t^{\text{obs}}(x) - \bar{Y}_c^{\text{obs}}(x), \text{ for } x = f, m$$

$$\hat{\tau}^{\text{strat}} = \frac{N(f)}{N(f) + N(m)} \cdot \hat{\tau}^{\text{dif}}(f) + \frac{N(m)}{N(f) + N(m)} \cdot \hat{\tau}^{\text{dif}}(m)$$

Brief explanations for (1)

$$\hat{\mathbb{V}}_W(\hat{\tau}^{\text{dif}}(x)) = \frac{s_c^2(x)}{N_c(x)} + \frac{s_t^2(x)}{N_t(x)}$$

$$\hat{\mathbb{V}}_W(\hat{\tau^{\text{strat}}}) = \left(\frac{N(f)}{N(f)+N(m)}\right)^2 \cdot \left(\frac{s_c^2(f)}{N_c(f)} + \frac{s_t^2(f)}{N_t(f)}\right) + \left(\frac{N(m)}{N(f)+N(m)}\right)^2 \cdot \left(\frac{s_c^2(m)}{N_c(m)} + \frac{s_t^2(m)}{N_t(m)}\right)$$

$$\begin{aligned}\hat{\mathbb{V}}(\hat{\tau}^{\text{strat}}) &= \left(\frac{N(f)}{N(f)+N(m)}\right)^2 \cdot \left(\frac{s_c^2(f)}{N_c(f)} + \frac{s_t^2(f)}{N_t(f)}\right) + \left(\frac{N(f)}{N(f)+N(m)}\right)^2 \cdot \left(\frac{s_c^2(f)}{N_c(f)} + \frac{s_t^2(f)}{N_t(f)}\right) \\ &+ \frac{1}{N} \cdot \frac{N(f) \cdot N(m)}{(N(f)+N(m))^2} \cdot \left(\hat{\tau}^{\text{dif}}(f) - \hat{\tau}^{\text{dif}}(m)\right)^2 \\ &= \hat{\mathbb{V}}_W(\hat{\tau}^{\text{strat}}) + \frac{N(f) \cdot N(m)}{N^3} \cdot \left(\hat{\tau}^{\text{dif}}(f) - \hat{\tau}^{\text{dif}}(m)\right)^2\end{aligned}$$

Brief explanations for (1)

$$\mathbb{V}_{\mathsf{sp}}(\tau(X_i)) = \frac{N(f) \cdot N(m)}{N^2} \cdot (\tau(f) - \tau(m))^2$$

$$\mathbb{V}(\hat{\tau}^{\mathsf{strat}}) \approx \mathbb{V}_W(\hat{\tau}^{\mathsf{strat}}) + \frac{\mathbb{V}_{\mathsf{sp}}(\tau(X_i))}{N}$$

Table of Contents

1. Introduction
2. Estimands
3. The Common Structure of Standard Estimators
4. A General Formula for the Conditional Sampling Variance
5. A Simple Estimator for the Unit-level Conditional Sampling Variance
6. An Estimator for the Sampling Variance for the Population ATE
7. Alternative Estimators for the Sampling Variance

The Common Structure of Standard Estimators

- Most estimators for ATE (ex. those in Ch12, Ch17, Ch18) have a common structure
- Refer to this structure as **affine consistency**
  1. adding a constant c_t to all observed outcomes for treated units increases the estimated average causal effect by c_t
  2. adding a constant c_c to all observed outcomes for control units decreases the estimated average causal effect by c_c
  3. changing the scale of the outcome by multiplying all observed outcomes by a constant c_s changes the estimated average effect by a factor c_s
- Estimator that do not have this property often have particular unattractive features

Weights

$$\hat{\tau} = \hat{\tau}(\mathbf{Y}^{\mathsf{obs}}, \mathbf{W}, \mathbf{X}) = \frac{1}{N_t} \sum_{i: W_i=1} \lambda_i \cdot Y_i^{\mathsf{obs}} - \frac{1}{N_c} \sum_{i: W_i=0} \lambda_i \cdot Y_i^{\mathsf{obs}} \qquad (2)$$

- $\lambda_i = \lambda(W_i, X_i, \mathbf{W}_{(-i)}, \mathbf{X}_{(-i)})$: row exchangeable function in $(\mathbf{W}_{(-i)}, \mathbf{X}_{(-i)})$
- Summation restrictions

$$\frac{1}{N_c} \sum_{i: W_i=0} \lambda_i = 1, \; \frac{1}{N_t} \sum_{i: W_i=1} \lambda_i = 1 \qquad (3)$$

- (2), (3) capture affine consistency

Weights

$$\hat{\tau} = \frac{1}{N_t} \sum_{i: W_i = 1} \lambda_i \cdot Y_i^{\text{obs}} - \frac{1}{N_c} \sum_{i: W_i = 0} \lambda_i \cdot Y_i^{\text{obs}}$$

- Difference Estimator

$$\hat{\tau}^{\text{dif}} = \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}$$

$$\lambda_i = 1, \text{ for all } i$$

Weights

- Regression Estimator

$$Y_i^{\text{obs}} = \alpha + \tau \cdot W_i + \beta \cdot X_i + \epsilon_i$$

$$\begin{aligned}\lambda_i^{\text{ols}} =& \bar{W}^{W_i} \cdot (1 - \bar{W})^{1-W_i} \\ & \cdot \frac{S_X^2(N-1)/N - (\bar{X}_t - \bar{X}_c) \cdot (X_i - \bar{X})}{S_X^2(N-1)/N - \bar{W} \cdot (1 - \bar{W}) \cdot (\bar{X}_t - \bar{X}_c)^2}\end{aligned}$$

Weights

- Weighting Estimator

$$\hat{\tau}^{\text{ht}} = \sum_{i:\,W_i=1} \frac{Y_i^{\text{obs}}}{e(X_i)} \Bigg/ \sum_{j:\,W_j=1} \frac{1}{e(X_j)} - \sum_{i:\,W_i=0} \frac{Y_i^{\text{obs}}}{1-e(X_i)} \Bigg/ \sum_{i':\,W_j=0} \frac{1}{1-e(X_j)}$$

$$\lambda_i^{\text{ht}} = \begin{cases} \frac{N_c}{1-e(X_i)} \Big/ \sum_{j:\,W_j=0} \frac{1}{1-e(X_j)} & \text{if } W_i = 0 \\ \frac{N_t}{e(X_i)} \Big/ \sum_{j:\,W_j=1} \frac{1}{e(X_j)} & \text{if } W_i = 1 \end{cases}$$

Weights

- Subclassification Estimator
  - $N(j)$: number of units in subclass j
  - $N_c(j), N_t(j)$: number of control / treated in subclass j
  - $B_i(j) \in \{0, 1\}$: binary indicator for unit i falling in subclass j

$$\lambda_i^{\text{strat}} = \begin{cases} \sum_{j=1}^{J} B_i(j) \cdot (N_c/N_c(j)) \cdot (N(j)/N) & \text{if } W_i = 0 \\ \sum_{j=1}^{J} B_i(j) \cdot (N_t/N_t(j)) \cdot (N(j)/N) & \text{if } W_i = 1 \end{cases}$$

Weights

- Matching Estimator
  - Simple matching estimator with M matches for each treated and control

$$\hat{\tau}^{\text{match}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{Y}_i(1) - \hat{Y}_i(0))$$

$$\hat{Y}_i(w) = \begin{cases} Yi^{\text{obs}} & \text{if } W_i = w \\ \sum_{j\in\mathcal{M}^c(i)} Y_j^{\text{obs}}/M & \text{if } W_i = 1, w = 0 \\ \sum_{j\in\mathcal{M}^t(i)} Y_j^{\text{obs}}/M & \text{if } W_i = 0, w = 1 \end{cases}$$

- $\hat{Y}_i(w)$ is a linear combination of Y_j^{obs}

Table of Contents

1. Introduction
2. Estimands
3. The Common Structure of Standard Estimators
4. A General Formula for the Conditional Sampling Variance
5. A Simple Estimator for the Unit-level Conditional Sampling Variance
6. An Estimator for the Sampling Variance for the Population ATE
7. Alternative Estimators for the Sampling Variance

Notations

- Super-population conditional statistics

$$\mu_c(x) = \mathbb{E}_{\text{sp}}[\,Y_i^{\text{obs}}|\,W_i = 0, X_i = x],\;\; \mu_t(x) = \mathbb{E}_{\text{sp}}[\,Y_i^{\text{obs}}|\,W_i = 1, X_i = x]$$

$$\sigma_c^2(x) = \mathbb{V}_{\text{sp}}(\,Y_i^{\text{obs}}|\,W_i = 0, X_i = x),\;\; \sigma_t^2(x) = \mathbb{V}_{\text{sp}}(\,Y_i^{\text{obs}}|\,W_i = 1, X_i = x)$$

- Unit-level conditional statistics

$$\mu_i = \mathbb{E}_{\text{sp}}[\,Y_i^{\text{obs}}|\,W_i, X_i] = \begin{cases} \mu_c(X_i) & \text{if } W_i = 0 \\ \mu_t(X_i) & \text{if } W_i = 1 \end{cases}$$

$$\sigma_i^2 = \mathbb{V}_{\text{sp}}[\,Y_i^{\text{obs}}|\,W_i, X_i] = \begin{cases} \sigma_c^2(X_i) & \text{if } W_i = 0 \\ \sigma_t^2(X_i) & \text{if } W_i = 1 \end{cases}$$

General Formula

$$
\begin{aligned}
\hat{\tau} &= \frac{1}{N_t} \sum_{i:\, W_i=1} \lambda_i \cdot Y_i^{\text{obs}} - \frac{1}{N_c} \sum_{i:\, W_i=0} \lambda_i \cdot Y_i^{\text{obs}} \\
&= \left(\frac{1}{N_t} \sum_{i:\, W_i=1} \lambda_i \cdot \mu_i - \frac{1}{N_c} \sum_{i:\, W_i=0} \lambda_i \cdot \mu_i \right) \\
&\quad + \left(\frac{1}{N_t} \sum_{i:\, W_i=1} \lambda_i \cdot (Y_i^{\text{obs}} - \mu_i) - \frac{1}{N_c} \sum_{i:\, W_i=0} \lambda_i \cdot (Y_i^{\text{obs}} - \mu_i) \right)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathbb{V}_{\text{sp}}(\hat{\tau} | \mathbf{X}, \mathbf{W}) &= \frac{1}{N_t^2} \sum_{i:\, W_i=1} \lambda_i^2 \cdot \sigma_i^2 + \frac{1}{N_c^2} \sum_{i:\, W_i=0} \lambda_i^2 \cdot \sigma_i^2 \\
\hat{\mathbb{V}}_{\text{sp}}(\hat{\tau} | \mathbf{X}, \mathbf{W}) &= \frac{1}{N_t^2} \sum_{i:\, W_i=1} \lambda_i^2 \cdot \hat{\sigma}_i^2 + \frac{1}{N_c^2} \sum_{i:\, W_i=0} \lambda_i^2 \cdot \hat{\sigma}_i^2 \qquad (4)
\end{aligned}
$$

Table of Contents

1 Introduction

2 Estimands

3 The Common Structure of Standard Estimators

4 A General Formula for the Conditional Sampling Variance

5 A Simple Estimator for the Unit-level Conditional Sampling Variance

6 An Estimator for the Sampling Variance for the Population ATE

7 Alternative Estimators for the Sampling Variance

A Single Exact Match

- Particular i with $W_i = 1$, estimate σ_i^2
- Suppose $\exists i'$ s.t. $W_{i'} = W_i = 1$, $X_{i'} = X_i = x$
- Then,

$$\mathbb{E}_{\text{sp}}[Y_i^{\text{obs}} - Y_{i'}^{\text{obs}} | X_i = X_{i'} = x, W_i = W_{i'} = 1]$$
$$= \mathbb{E}_{\text{sp}}[(Y_i^{\text{obs}} - \mu_i) - (Y_{i'}^{\text{obs}} - \mu_{i'}) | X_i = X_{i'} = x, W_i = W_{i'} = 1] = 0$$

$$\mathbb{E}_{\text{sp}}[(Y_i^{\text{obs}} - Y_{i'}^{\text{obs}})^2 | X_i = X_{i'} = x, W_i = W_{i'} = 1]$$
$$= \mathbb{V}_{\text{sp}}(Y_i^{\text{obs}} | X_i = x, W_i = 1) + \mathbb{V}_{\text{sp}}(Y_{i'}^{\text{obs}} | X_{i'} = x, W_{i'} = 1) = 2\sigma_t^2(x)$$

$$\hat{\sigma}_i^2 = (Y_i^{\text{obs}} - Y_{i'}^{\text{obs}})^2 / 2 \tag{5}$$

A Single Approximate Match

- Look for the most similar unit (Ch18)

$$\begin{aligned}\hat{\sigma}_i^2 &= (Y_i^{\text{obs}} - Y_{l_i}^{\text{obs}})^2/2 \\ &= (\mu_i - \mu_{l_i} + (Y_i^{\text{obs}} - \mu_i) - (Y_{l_i}^{\text{obs}} - \mu_{l_i}))^2/2\end{aligned}$$

$$\mathbb{E}_{\text{sp}}[\hat{\sigma}_i^2|\mathbf{X},\mathbf{W}]/2 - \sigma_i^2 = (\mu_i - \mu_{l_i})^2/2 + (\sigma_{l_i}^2 - \sigma_i^2)/2$$

A Bias-Adjusted Variance Estimator

$$\mathbb{E}_{\mathsf{sp}}[\hat{\sigma}_i^2|\mathbf{X},\mathbf{W}]/2 - \sigma_i^2 = (\mu_i - \mu_{l_i})^2/2 + (\sigma_{l_i}^2 - \sigma_i^2)/2$$

- Number of covariates $\uparrow \Rightarrow$ bias $\uparrow$

$$\mathbb{E}_{\mathsf{sp}}[Y_i^{\mathsf{obs}}|X_i, W_i = 1] = X_i\beta_t,\ \mathbb{E}_{\mathsf{sp}}[Y_i^{\mathsf{obs}}|X_i, W_i = 0] = X_i\beta_c$$

$$\hat{\epsilon}_i = \begin{cases} Y_i^{\mathsf{obs}} - X_i\hat{\beta}_c & \text{if } W_i = 0 \\ Y_i^{\mathsf{obs}} - X_i\hat{\beta}_t & \text{if } W_i = 1 \end{cases}$$

$$\hat{\sigma}_i^{2,\mathsf{adj}} = (\hat{\epsilon}_i - \hat{\epsilon}_{l_i})^2/2$$

Multiple Matches

- Closest M units in terms of covariate values

$$\hat{\sigma}^2_{i,M} = \frac{1}{2M} \sum_{i' \in \mathcal{M}_i^t} (Y_{i'}^{\text{obs}} - Y_i^{\text{obs}})^2$$

Table of Contents

1. Introduction
2. Estimands
3. The Common Structure of Standard Estimators
4. A General Formula for the Conditional Sampling Variance
5. A Simple Estimator for the Unit-level Conditional Sampling Variance
6. An Estimator for the Sampling Variance for the Population ATE
7. Alternative Estimators for the Sampling Variance

An Estimator for the Sampling Variance for the Population ATE

$$\hat{\mathbb{V}}_W(\hat{\tau}) = \sum_{i=1}^{N} \hat{\sigma}_i^2 \cdot \lambda_i^2$$

$$\hat{\mathbb{V}}_{\mathsf{sp}}(\tau(X_i)) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\tau}_i^{\mathsf{match}} - \hat{\tau})^2 - \frac{2}{N}\sum_{i=1}^{N}\hat{\sigma}_i^2$$

$$\text{where } \hat{\tau}_i^{\mathsf{match}} = \hat{Y}_i(1) - \hat{Y}_i(0)$$

$$\hat{\mathbb{V}}_{\mathsf{sp}}(\hat{\tau}) = \hat{\mathbb{V}}_W(\hat{\tau}) + \frac{1}{N}\hat{\mathbb{V}}_{\mathsf{sp}}(\tau(X_i)) \tag{6}$$

$$= \sum_{i} \hat{\sigma}_i^2 \cdot (\lambda_i^2 - \frac{2}{N^2}) + \frac{1}{N^2}\sum_{i=1}^{N}(\hat{\tau}_i^{\mathsf{match}} - \hat{\tau})^2 \tag{7}$$

Table of Contents

1. Introduction
2. Estimands
3. The Common Structure of Standard Estimators
4. A General Formula for the Conditional Sampling Variance
5. A Simple Estimator for the Unit-level Conditional Sampling Variance
6. An Estimator for the Sampling Variance for the Population ATE
7. Alternative Estimators for the Sampling Variance

Least Squares Sampling Variance Estimators

- Consider the subclassification estimator
- J subclasses, $B_i(j)$: unit i belongs to subclass j, $\tau(j)$: average effect in j

$$\hat{\beta}(j) = (\sum_{i:B_i(j)=1} Z_i^T \cdot Z_i)^{-1} (\sum_{i:B_i(j)=1} Z_i^T \cdot Y_i^{\text{obs}})$$

$$\hat{\mathbb{V}}(\hat{\tau}^{\text{strat}}) = \sum_{j=1}^{J} (\frac{N_c)j + N_t(j)}{N})^2 \cdot \hat{\mathbb{V}}(\hat{\tau}^{\text{ols}}(j))$$

Bootstrap Sampling Variance Estimators

$$\hat{\mathbb{V}}^{\text{boot}} = \sum_b (\hat{\tau}_b - \bar{\tau}_{\text{boot}})/(B-1)$$

- No formal justification for the bootstrap for either the subclassification or the matching estimator

References

[1] Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015. DOI: 10.1017/CBO9781139025751.