Causal Inference
Chapter 16. Trimming to Improve Balance in Covariate Distributions

Yuha Park

Aug 9, 2022

Seoul National University

Contents

Introduction

Trimming to Improve Balance in Covariate Distributions

- **Trimming the sample** by discarding units with propensity score values close to zero or one, with the exact threshold determined by the joint distribution of covariates and treatment status in order to optimize asymptotic precision

- A covariate-and-treatment-indicator-dependent criterion for determining a threshold, denoted by α, such that all units with estimated propensity score values in the intervals $[0, \alpha]$ and $[1 - \alpha, 1]$ are discarded, and causal effects are estimated only for units with values for the estimated propensity score in the interval $[\alpha, 1 - \alpha]$

An example with a single binary covariate

An example with a single binary covariate

Notations:

- X_i: a single pre-treatment variable taking on two values
  (e.g., $X_i = x$, $x \in \{f, m\}$)
- W_i: the observed treatment indicator for unit i $\quad (W_i \in \{0, 1\})$
- N: the number of a random sample from an infinite super-population
- $N(x)$: the sample size for the subsample with $X_i = x$
  $$(e.g., N = N(f) + N(m),\ N(x) = \sum_{i=1}^{N} \mathbb{I}(X_i = x))$$
- q: the super-population share of $X_i = m$ units
  $(q = \mathbb{E}_{sp}[N(m)/N])$

An example with a single binary covariate

- The population average treatment effect conditional on $X_i = x$:
  $$\tau_{sp}(x) = \mathbb{E}_{sp}[Y_i(1) - Y_i(0)|X_i = x]$$
- The super-population average treatment effect:
  $$\tau_{sp} = \mathbb{E}_{sp}[Y_i(1) - Y_i(0)] = (1-q)\cdot\tau_{sp}(f) + q\cdot\tau_{sp}(m)$$
- The number of control and treated units with covariate value $X_i = x$:
  $$N_c(x) = \sum_{i:X_i=x}(1-W_i) \quad \text{and} \quad N_t(x) = \sum_{i:X_i=x} W_i$$
- The propensity score at x: $e(x) = N_t(x)/N(x)$
- The average outcome within each of the four subpopulations defined by treatment status and covariate value:
  $$\bar{Y}_c^{obs}(x) = \frac{1}{N_c(x)} \sum_{i:X_i=x} Y_i^{obs}\cdot(1-W_i) \quad \text{and}$$
  $$\bar{Y}_t^{obs}(x) = \frac{1}{N_t(x)} \sum_{i:X_i=x} Y_i^{obs}\cdot W_i, \qquad \text{for} \quad x = f, m$$

An example with a single binary covariate

Assume that the super-population variance of $Y_i(w)$ given $X_i = x$ is σ^2 for all x and w.

Natural estimators for the average treatment effects for each of the two subpopulations, $X_i = f, m$:

$$\hat{\tau}^{dif}(f) = \bar{Y}_t^{obs}(f) - \bar{Y}_c^{obs}(f) \quad \text{and} \quad \hat{\tau}^{dif}(m) = \bar{Y}_t^{obs}(m) - \bar{Y}_c^{obs}(m).$$

Then the asymptotic sampling variance:

$$N \cdot \mathbb{V}\left(\hat{\tau}^{dif}(f)\right) = N \cdot \sigma^2 \cdot \left(\frac{1}{N_c(f)} + \frac{1}{N_t(f)}\right) \longrightarrow \frac{\sigma^2}{(1-q)} \cdot \frac{1}{e(f)\cdot(1-e(f))} = \mathbb{AV}\left(\hat{\tau}^{dif}(f)\right),$$

$$N \cdot \mathbb{V}\left(\hat{\tau}^{dif}(m)\right) = N \cdot \sigma^2 \cdot \left(\frac{1}{N_c(m)} + \frac{1}{N_t(m)}\right) \longrightarrow \frac{\sigma^2}{q} \cdot \frac{1}{e(m)\cdot(1-e(m))} = \mathbb{AV}\left(\hat{\tau}^{dif}(m)\right).$$

An example with a single binary covariate

Natural estimator for the population average treatment effect, $\tau_{sp} = \mathbb{E}_{sp}[Y_i(1) - Y_i(0)]$:

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

Since $\hat{\tau}^{dif}(f)$ and $\hat{\tau}^{dif}(m)$ are independent, the sampling variance of the population average treatment effect:

$$\mathbb{V}\left(\hat{\tau}^{strat}\right) = \left(\frac{N(f)}{N(f)+N(m)}\right)^2 \cdot \mathbb{V}\left(\hat{\tau}^{dif}(f)\right) + \left(\frac{N(m)}{N(f)+N(m)}\right)^2 \cdot \mathbb{V}\left(\hat{\tau}^{dif}(m)\right).$$

Thus, the normalized sampling variance for $\hat{\tau}$ converges to

$$N \cdot \mathbb{V}\left(\hat{\tau}^{strat}\right) \longrightarrow \sigma^2 \cdot \left(\frac{q}{e(m)\cdot(1-e(m))} + \frac{1-q}{e(f)\cdot(1-e(f))}\right) = \mathbb{AV}\left(\hat{\tau}^{strat}\right).$$

An example with a single binary covariate

Suppose that $\frac{e(m)\cdot(1-e(m))}{e(f)\cdot(1-e(f))} \leq \frac{1-q}{1-2\cdot q}$. Then,

$$\mathbb{AV}\big(\hat{\tau}^{dif}(f)\big) \leq \mathbb{AV}\big(\hat{\tau}^{strat}\big) \leq \mathbb{AV}\big(\hat{\tau}^{dif}(m)\big).$$

On the other hand, if, $\frac{1+q}{q} \leq \frac{e(m)\cdot(1-e(m))}{e(f)\cdot(1-e(f))}$, then

$$\mathbb{AV}\big(\hat{\tau}^{dif}(m)\big) \leq \mathbb{AV}\big(\hat{\tau}^{strat}\big) \leq \mathbb{AV}\big(\hat{\tau}^{dif}(f)\big).$$

Otherwise, i.e., $\frac{1-q}{1-2\cdot q} < \frac{e(m)\cdot(1-e(m))}{e(f)\cdot(1-e(f))} < \frac{1+q}{q}$, then

$$\mathbb{AV}\big(\hat{\tau}^{strat}\big) \leq \min\Big(\mathbb{AV}\big(\hat{\tau}^{dif}(m)\big), \mathbb{AV}\big(\hat{\tau}^{dif}(f)\big)\Big).$$

An example with a single binary covariate

The general idea behind the trimming approach is based on the estimation of average effects for a subpopulation of units with $X_i \in \mathbb{C}$:

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

for a subset of the covariate space, $\mathbb{C} \subset \mathbb{X}$.

We look for an optimal subset $\mathbb{C}^*$ of the covariate space $\mathbb{X}$ where the average treatment effect is most precisely estimable.

- covariate space $\mathbb{X} = \{f, m\}$
- the set of possible subsets of $\mathbb{X}$ is $\{\{f, m\}, \{f\}, \{m\}, \emptyset\}$

An example with a single binary covariate

Choose the subset $\mathbb{C}^*$ of the covariate space as

$$\mathbb{C}^* = \begin{cases} \{f\}, & \text{if } \frac{e(m)\cdot(1-e(m))}{e(f)\cdot(1-e(f))} \leq \frac{1-q}{1-2\cdot q}, \\ \{m\}, & \text{if } \frac{1+q}{q} \leq \frac{e(m)\cdot(1-e(m))}{e(f)\cdot(1-e(f))}, \\ \{f, m\}, & \text{otherwise.} \end{cases}$$

We then discard all units with $X_i \notin \mathbb{C}^*$, and thus focus on estimating

$$\tau_{\mathbb{C}^*} = \mathbb{E}_{sp}[Y_i(1) - Y_i(0) | X_i \in \mathbb{C}^*],$$

based solely on the subsample of units with $X_i \in \mathbb{C}^*$.

Selecting a subsample based on the propensity score

Selecting a subsample based on the propensity score

- Now let us look at the general case, which allows for multi-component and continuous covariates
- We cannot simply list all subsets of the covariate space and compare within-subset sampling variances because there are infinitely many such subsets
- For a given subset, we cannot even calculate the exact sampling variance the way we did for the binary covariate case
- We focus on the asymptotic sampling variance for the efficient estimator for the average treatment effect for each subset

Selecting a subsample based on the propensity score

The asymptotic sampling variance for the efficient estimator normalized by the sample size:

$$\mathbb{AV}_{fs}^{eff} = \mathbb{E}_{sp}\left[\frac{\sigma_t^2(X_i)}{e(X_i)} + \frac{\sigma_c^2(X_i)}{1-e(X_i)}\right],$$

for the finite-sample average treatment effect τ_{fs}.

If the propensity score is close to zero or one, the sampling variance bound will be relatively large. Therefore, **dropping units for which the propensity score is close to zero or one** may improve to estimate average treatment effects.

Selecting a subsample based on the propensity score

The average treatment effect given that the covariate value X in some subset $\mathbb{C}$ of the covariate space:

$$\tau_{\mathbb{C}} = \mathbb{E}_{sp}[\tau(X_i)|X_i \in \mathbb{C}]$$

The asymptotic sampling variance of the efficient estimator for this average treatment effect normalized by the sample size N:

$$\mathbb{AV}^{eff}_{fs}(\mathbb{C}) = \frac{1}{q(\mathbb{C})} \cdot \mathbb{E}_{sp}\left[\frac{\sigma^2_t(X_i)}{e(X_i)} + \frac{\sigma^2_c(X_i)}{1-e(X_i)}\middle| X \in \mathbb{C}\right], \qquad (\star)$$

where $q(\mathbb{C}) = \Pr_{sp}(X_i \in \mathbb{C})$ is the probability of the covariate being in the subset $\mathbb{C}$ in the super-population (i.e., the effective sample size).

$\Longrightarrow$ The question is how to minimize equation ($\star$).

Selecting a subsample based on the propensity score

If we assume homoskedasticity, $\mathbb{V}(Y_i(w)|X_i = x) = \sigma^2$, for all w and x, the optimal sampling variance:

$$\mathbb{AV}_{fs}^{eff}(\mathbb{C}) = \frac{\sigma^2}{q(\mathbb{C})} \cdot \mathbb{E}_{sp}\left[\frac{1}{e(X_i)} + \frac{1}{1-e(X_i)}\middle| X \in \mathbb{C}\right].$$

Now we look for the optimal subset $\mathbb{C}^*$ which is the set $\mathbb{C}$ such that minimizes the asymptotic sampling variance among all subsets $\mathbb{C}$ of $\mathbb{X}$.

Selecting a subsample based on the propensity score

– If $\sup_{x \in \mathbb{X}} \frac{1}{e(x) \cdot (1-e(x))} \leq 2 \cdot \mathbb{E}_{sp}\left[\frac{1}{e(X_i) \cdot (1-e(X_i))}\right]$, then the optimal $\mathbb{C}$ is equal to the entire covariate space, $\mathbb{C}^* = \mathbb{X}$.

– Otherwise, the optimal set $\mathbb{C}^*$ has the form $\mathbb{C}^* = \{x \in \mathbb{X} | \, \alpha \leq e(x) \leq 1 - \alpha\}$, where the threshold $\alpha = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{\gamma}}$, where γ is a solution to

$$\gamma = 2 \cdot \mathbb{E}_{sp}\left[\frac{1}{e(X_i) \cdot (1-e(X_i))} \middle| \frac{1}{e(X_i) \cdot (1-e(X_i))} \leq \gamma\right].$$

Selecting a subsample based on the propensity score

To implement this procedure we conduct the following calculations.

(1) Estimate the propensity score using the methods in ch.13

(2) Given the estimated propensity score $\hat{e}(x)$, we check whether

$$\max_{i:1,\cdots,N} \frac{1}{\hat{e}(X_i)\cdot(1-\hat{e}(X_i))} \leq 2 \cdot \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\hat{e}(X_i)\cdot(1-\hat{e}(X_i))}$$

(3) If the inequality holds, then $\hat{\mathbb{C}} = \mathbb{X}$

(3′-1) If the inequality does not hold, then solve for a value of γ, denoted by $\hat{\gamma}$, satisfying

$$\frac{\gamma}{N}\sum_{i=1}^{N} 1_{(\hat{e}(X_i)\cdot(1-\hat{e}(X_i)))^{-1}\leq\gamma} = \frac{2}{N}\sum_{i=1}^{N} \frac{1}{\hat{e}(X_i)\cdot(1-\hat{e}(X_i))} 1_{(\hat{e}(X_i)\cdot(1-\hat{e}(X_i)))^{-1}\leq\gamma}$$

(3′-2) Calculate $\hat{\alpha} = 1/2 - \sqrt{1/4 - 1/\hat{\gamma}}$ and $\hat{\mathbb{C}} = \{x \in \mathbb{X} |\, \hat{\alpha} \leq \hat{e}(x) \leq 1 - \hat{\alpha}\}$

(3′-3) Exclude units i with $\hat{e}(X_i)$ outside $\hat{\mathbb{C}}$