Matching to Improve Balance in Covariate Distributions

DongYoon Yang

September 19, 2022

Seoul National University

Table of Contents

Table of Contents

Notation

- N_t : the number of treated units.
- N_c : the number of controlled units.
- $\mathrm{I}'_c = \{N_t + 1, \cdots, N_t + N_c\}$
- Choose a subset $\mathrm{I}_c \subset \mathrm{I}'_c$
- $\mathcal{M}^c_1 := \left\{ j \in \mathrm{I}'_c : d(X_1, X_j) = \min_{j' \in \mathrm{I}'_c} d(X_1, X'_j) \right\}$
- $\mathcal{M}^c_i :=$
  $$\left\{ j \in \mathrm{I}'_c - \cup_{i'=1}^{i-1} \mathcal{M}^c_{i'} : d(X_1, X_j) = \min_{j \in \mathrm{I}'_c - \cup_{i'=1}^{i-1} \mathcal{M}^c_{i'}} d(X_1, X'_j) \right\}$$

The choice of distance metric d

- Mahalanobis Metric

$$d_M(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x} - \boldsymbol{x}')^\top \widehat{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{x}')$$

  where $\widehat{\Sigma} = \frac{N_c \cdot \widehat{\Sigma}_c + N_t \cdot \widehat{\Sigma}_t}{N_c + N_t}$

- $e(\boldsymbol{x})$: propensity score

- Linearized propensity score

$$d_\ell(\boldsymbol{x}, \boldsymbol{x}') = (\ell(\boldsymbol{x}) - \ell(\boldsymbol{x}'))^2$$

  where $\ell(\boldsymbol{x}) = \log \dfrac{e(\boldsymbol{x})}{1 - e(\boldsymbol{x})}$

Rejecting and Caliper Matching

- Rejecting Matching

- In some cases, even the closest match may not be close enough.

- Drop treated units if

$$|\ell(\boldsymbol{x}_i) - \ell(\boldsymbol{x}_{m_i})| > d_{\max}$$

- Caliper Matching

- Assign to treated unit $i = 1$ all control units $j \in I'_c$ such that

$$d(X_1, X_j) \leq d_{cal}$$

Table of Contents

Equal Percentage Bias Reduction (EPBR)

- $X_t \sim N(\mu_t, \Sigma)$
- $X_c \sim N(\mu_c, \Sigma)$
- $\mathbb{E}[Y_i(0)|X_i = \boldsymbol{x}_i] = \boldsymbol{x}_i^\top \boldsymbol{\beta} \; \forall i$
- μ_c^M : mean of control groups after matching.
- $\tau_{sp,t} = \mathbb{E}[Y_i(1) - Y_i(0)|W_i = 1]$

The expected bias of $\hat{\tau}^{\text{diff}}$ in observed data is

$$\mathbb{E}[\underbrace{\bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}}_{\hat{\tau}^{\text{diff}}} -\tau_{sp,t}] = (\mu_t - \mu_c)^\top \boldsymbol{\beta} \tag{1}$$

Definition

Matching method is called equal percentage bias reducing (EPBR) if $\exists c_M$ with $0 < c_M < 1$ such that
$(\mu_t - \mu_c^M)\boldsymbol{\beta} = c_M(\mu_t - \mu_c)\boldsymbol{\beta} \; \forall \boldsymbol{\beta}$.

- The only EPBR matching methods improves the bias for every linear predictor.

- Both propensity score and Mahalanobis matching methods are epbr.

- Propensity score Matching

- scalar matching, small samples, non-robust to model misspecification(parametric).

- Mahalanobis matching

- vector matching, large samples, robust to model misspecification(nonparametric).