Balancing vs modeling approaches to weighting in practice

Ambarish Chattopadhyay[1], *Christopher H. Hase*[1] & *José R. Zubizarreta*[1,2,3]

Harvard University

September 26, 2023

Contents

The need to weight

- In observational study, treated and control samples may suffer from imbalances due to baseline covariates

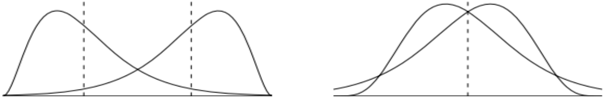


Figure 10.1 *Imbalance in distributions across treatment and control groups. (a) In the left panel, the groups differ in their averages (dotted vertical lines) but cover the same range of x. (b) The right panel shows a more subtle form of imbalance, in which the groups have the same average but differ in their distributions.*

- This imbalance introduces bias in estimating the treatment effects
- Remedy is to recreate the observational study as if it was a randomized experiment, using information only on observed covariates

- Two existing weighting approaches:
  1. propensity score modeling: maximize $P(Z = 1|\mathbf{X})$
  2. Balancing: optimizes certain features of the weights
- Modeling approach is popular because it provides a one-dimensional summary of the (possibly high dimensional) vector of observed covariates
- Any guesses what should be the problem with this modeling approach?

Motivation for Balancing approach

- The problem is incorrect PS model which introduces inconsistent estimates and imbalance in covariates
- Mild violations of overlap condition gives rise to unstable weights
- It is important to consider the dispersion of weights because it is related to effective sample size
- We want a method which balances covariates, as well as, reduce dispersion of weights

What should be a good weighting strategy?

- Construct weights so that it
  - balances covariates
  - decreases dispersion i.e., increases effective sample size
  - weights should be non-negative to satisfy sample boundedness condition
- Although, modeling and balancing serves different purpose but they are connected

Framework

- Let $Z_i \in \{0, 1\}$ be the treatment indicator
- n_t and n_c be the sample sizes in treatment and control groups, respectively. where $n_t + n_c = n$
- Observed data: $(\mathbf{X}_i, Z_i, Y_i^{\text{obs}}); i = 1, 2, ..., n$
- Estimands:

$$\text{ATE} = E[Y_i(1) - Y_i(0)]$$

$$\text{ATT} = E[Y_i(1) - Y_i(0) | Z_i = 1]$$

Assumptions:

- SUTVA: $Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$
- Unconfoundedness/ignorability: $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp Z_i | \mathbf{X_i}$
- Positivity: $0 < P(Z_i = 1 | \mathbf{X}_i = \mathbf{x}) < 1$
- Unconfoundedness assumption can be rewritten as:

$$\{Y_i(1), Y_i(0)\} | Z_i = 1, \mathbf{X_i} \sim \{Y_i(1), Y_i(0)\} | Z_i = 0, \mathbf{X_i}$$

Estimators

$$\hat{T}_{\text{HT}}^{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\frac{Z_i Y_i^{\text{obs}}}{\hat{e}(X_i)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-Z_i)Y_i^{\text{obs}}}{1-\hat{e}(X_i)} \tag{1}$$

$$\hat{T}_{\text{HT}}^{\text{ATT}} = \frac{1}{n}\sum_{i=1}^{n} Z_i Y_i^{\text{obs}} - \frac{1}{n}\sum_{i=1}^{n}(1-Z_i)\frac{\hat{e}(X_i)Y_i^{\text{obs}}}{1-\hat{e}(X_i)} \tag{2}$$

Weighting wish list

- Balance the distribution of observed covariates
- Assess the covariate balance in terms of the absolute standardized mean difference (ASMD) for each covariate:

$$\text{ASMD}(x) = \frac{\left|\bar{\mathbf{x}}_{w,t} - \bar{\mathbf{x}}_{w,c}\right|}{\sqrt{\frac{s_t^2+s_c^2}{2}}} \tag{3}$$

- This paper proposed using the target absolute standardized mean difference (TASMD)

$$\text{TASMD}(x_g) = \frac{\left|\bar{\mathbf{x}}_{w,g} - \bar{\mathbf{x}}^{*}\right|}{s_g} \tag{4}$$

- In principle, it is desirable to assess balance beyond marginal means

$$\text{TASMD}\{B_k(\mathbf{X})_g\} = \frac{|B_k\bar{(}\mathbf{X})_{w,g} - B_k\bar{(}\mathbf{X})^*|}{s_{g\{Bk(\mathbf{X})\}}} \tag{5}$$

Weighting for Stability

- Balance covariates to remove bias, as well as, to produce an estimator that has low variance
- Kang and Schafer (2007) exemplified how weights obtained through a misspecified PS model can be highly variable and produce weighted estimators that are highly unstable
- Let us consider the problem of estimating the ATE using the Hajek estimator:

$$\text{MSE}[T_{\text{lin}}] = (\text{Bias}[T_{\text{lin}}])^2 + \text{Var}[T_{\text{lin}}] \tag{6}$$

$$
\begin{aligned}
\text{Bias}[T_{\text{lin}}] &= E\left[E[T_{\text{lin}}|X,Z]\right] - E\left[\frac{1}{n}\sum_{i=1}^{n} E[Y_i(1) - Y_i(0)|X,Z]\right] \\
&= E\left[\sum_{i=1}^{n} Z_i w_i m_1(X_i) - \frac{1}{n}\sum_{i=1}^{n} m_1(X_i)\right] \\
&- E\left[\sum_{i=1}^{n}(1 - Z_i) w_i m_0(X_i) - \frac{1}{n}\sum_{i=1}^{n} m_0(X_i)\right]
\end{aligned} \tag{7}
$$

Under the balancing conditions bias tends to zero:

$$
\sum_{i=1}^{n} Z_i w_i m_1(X_i) = \frac{1}{n}\sum_{i=1}^{n} m_1(X_i)
$$

$$
\sum_{i=1}^{n}(1 - Z_i) w_i m_0(X_i) = \frac{1}{n}\sum_{i=1}^{n} m_0(X_i)
$$

- According to the variance term:

$$\mathrm{Var}[T_{\mathrm{lin}}] = E[\mathrm{Var}[T_{\mathrm{lin}}|X, Z]] + \mathrm{Var}[E[T_{\mathrm{lin}}|X, Z]] \tag{8}$$

$$\begin{aligned}\mathrm{Var}[T_{\mathrm{lin}}|X, Z] &= \sum_{i=1}^{n} Z_i w_i^2 \mathrm{Var}[Y_i(1)|X_i] - \sum_{i=1}^{n}(1 - Z_i) w_i^2 \mathrm{Var}[Y_i(0)|X_i] \\ &= \sigma_1^2 \sum_{i=1}^{n} Z_i w_i^2 + \sigma_0^2 \sum_{i=1}^{n}(1 - Z_i) w_i^2\end{aligned} \tag{9}$$

- Controlling the sum of squares of the weights in the two groups is equivalent to controlling the variance of the weights in the two groups

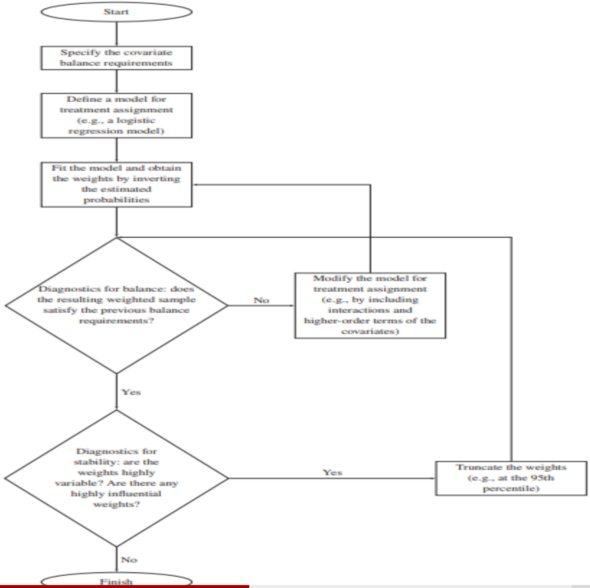
Violations of overlap

- High variability in the weights often stems from a few observations
- With modeling approaches to weighting, the presence of a few very large weights often relates to violations of the positivity assumption
- Diagnostics for stability:
  1. Coefficient of variation of the weights
  2. Effective sample size:

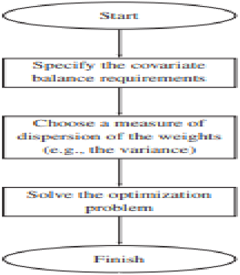
$$n_{\text{eff}} = \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{n} w_i^2} \tag{10}$$

  3. Measures of extremity: maximum of the weights, 95th and 99th percentile of weights
  4. Graphical diagnostics: box plot of the weights

Flowchart of PS Modeling



Flowchart of Balancing



Balancing approach

- Suppose we are interested in estimating the ATT. In that case the strategy should be to weight the individuals in the control group such that the resulting weighted control units have similar covariate distributions as the treated units
- Minimal weighting method chooses an optimal w by solving the following convex optimization problem:

$$\underset{w}{\text{Minimize}} \quad \sum_{i:Z_i=0} \psi(w_i) \tag{11}$$

$$\text{subject to} \quad \left| \sum_{i:Z_i=0} w_i B_k(X_i) - \frac{1}{n_t} \sum_{i:Z_i=1} B_k(X_i) \right| \leq \delta_k, \quad k = 1, 2, \ldots, K \tag{12}$$

Connection between the two approaches

- Wang and Zubizarreta (2020) showed that the optimization problem (11) implicitly fits a model for the propensity score
- In particular, investigating the dual formulation of (11) enables to identify this connection between the two approaches
- Let $\mathbf{w}^* = (w_1^*, w_2^*, \ldots, w_n^*)$ be the optimal set of weights obtained by solving the primal problem in (11)

$$w_i^* = \rho' \left(\lambda_1^* B_1(X_i) + \ldots + \lambda_K^* B_K(X_i) \right) - \frac{1}{n_t}, \quad i = 1, 2, \ldots, n \quad (13)$$

- where $\boldsymbol{\lambda}^* = (\lambda_1^*, \ldots, \lambda_K^*)$ is the solution to the corresponding dual problem of (11) and ρ is a real-valued differentiable function

- The dual problem can equivalently be written in the following form:

$$\underset{\boldsymbol{\lambda}}{\text{Minimize}} \quad \sum_{i=1}^{n} \left[-(1 - Z_i)\rho\{B(X_i)^\top \boldsymbol{\lambda}\} + \frac{B(X_i)^\top \boldsymbol{\lambda}}{n_t} \right] + |\boldsymbol{\lambda}|^T \boldsymbol{\delta} \quad (14)$$

Conclusion and Future direction

- Should use balancing approach over modeling approach
- Future direction:
  1. In an observational study, the investigator may not have enough substantive knowledge on which functions of the covariates to balance
  2. So we can consider some non-parametric form of covariate balance because true forms of conditional mean may not be known
  3. There is a wide scope for research on developing weights in the balancing approach, which targets balance for a sufficiently general class of functions
  4. Choice of the tuning parameter for approximate covariate balance

References

- Chattopadhyay, Ambarish, Christopher H. Hase, and José R. Zubizarreta. "Balancing vs modeling approaches to weighting in practice." Statistics in Medicine 39, no. 24 (2020): 3227-3254.
- Kang JDY, Schafer JL. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data (with discussion). Stat Sci. 2007;22(4):523-539
- Wang Y, Zubizarreta JR. Minimal dispersion approximately balancing weights: asymptotic properties and practical considerations. Biometrika. 2020;107(1):93-105

Thank You !

Any Questions?