Propensity Score Overlap Weighting

Ben Rohlfing

Southern Illinois University Edwardsville

brohlfi@siue.edu

December 12, 2019

Overview

Background

- Treatment group - Group in an observational study that receives treatment
- Everybody else placed in control group
- Propensity score of an individual is the conditional probability of treatment, given the individual's background covariates
- Concept introduced in 1984 by Paul Rosenbaum and Donald Rubin in order to estimate causal effects of smoking on one's mortality rate
- Can be applied to other observational studies with non-randomized treatment assignment

Potential Outcome Framework

Given:

- Sample size n
- $Z_i = z$, where $z = 0, 1$, indicates group membership
- $X_i = (X_{i1}, \dots, X_{iK})^T$ indicates vector of K covariates
- $Y_i(Z_i)$ indicates potential outcome for i^{th} individual

Observed Response

$$Y_i = Y_i(Z_i) = Z_i \cdot Y_i(1) + (1 - Z_i) \cdot Y_i(0) \tag{1}$$

Potential Outcome Framework (Cont.)

Average Treatment Effect (ATE)

The individual treatment effect on the i^{th} individual is

$$Y_i(1) - Y_i(0)$$

Cannot be directly measured, so instead consider ATE as follows:

$$ATE = E[Y(1) - Y(0)]. \tag{2}$$

A naive estimate of the ATE is given as follows:

$$\widehat{ATE}_{nv} = \frac{\sum_{i=1}^{n} Z_i Y_i}{\sum_{i=1}^{n} Z_i} - \frac{\sum_{i=1}^{n}(1 - Z_i) Y_i}{\sum_{i=1}^{n}(1 - Z_i)}. \tag{3}$$

Propensity Score Assumptions

Assumption 1 (Unconfoundedness [2])

For any unit $i = 1, \ldots, n$,

$$P(Z_i = 1 \mid Y_i(0), Y_i(1), X_i) = P(Z_i = 1 \mid X_i) \tag{4}$$

or, using conditional independence notation

$$Z_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid X_i \tag{5}$$

In other words, the treatment variable Z_i is independent of the potential outcomes, $Y_i(0)$ and $Y_i(1)$, after conditioning on X_i

Assumption 2 (Probabilistic Assignment or Positive Overlap [2])

For any unit i,

$$0 < P(Z_i = 1 \mid X_i) < 1$$

Balancing Scores

Definition 1 (Balancing Score [2])

For every unit i in the sample, a balancing score $b(X_i)$ is a function of the covariate X_i such that

$$Z_i \perp\!\!\!\perp X_i \mid b(X_i),$$

or, in terms of a probability statement,

$$P(Z_i = 1 | \; X_i, \; b(X_i)) = P(Z_i = 1 | \; b(X_i)).$$

Balancing Scores (Cont.)

Theorem (Unconfoundedness Given Any Balancing Score)

Suppose Assumption 1 is true. Then, treatment assignment is unconfounded given any balancing score,

$$P(Z_i = 1 \mid Y_i(0), Y_i(1), b(X_i)) = P(Z_i = 1 \mid b(X_i)), \tag{6}$$

or, using conditional independence notation

$$Z_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid b(X_i)). \tag{7}$$

Propensity Scores

Definition 2 (Propensity Score)

The propensity score of unit i, with covariate measurement X_i, is defined as the conditional probability of treatment assignment

$$e(X_i) = P(Z_i = 1|X_i) = E_Z[Z_i|X_i].$$

Theorem (Propensity Score is a balancing score [2])

For every unit i, the propensity score $e(X_i)$ is a balancing score, i.e.,

$$P(Z_i = 1 \mid X_i, e(X_i)) = P(Z_i = 1 \mid e(X_i)). \tag{8}$$

Note: Propensity scores are commonly estimated using logistic regression model

Intuition on Propensity Score Weighting

- Treatment units with low propensity scores are upweighted by the reciprocal of its propensity score
- Control units with high propensity scores are upweighted by the reciprocal of one minus its propensity score

Weighted ATE Estimator

First step to find weighted ATE is to rewrite $E[ZY]$ as

$$\begin{aligned} E[Z \cdot Y] &= E_X(E[Z|X] \cdot E[Y(1)|X]) \quad \text{(Assumption 1)} \\ &= E_X(e(X) \cdot E[Y(1)|X]) \quad \text{(Definition 2)} \end{aligned}$$

By Assumption 2, we can rewrite $E[Y(1)]$ as

$$E\left\{\frac{Z \cdot Y}{e(X)}\right\} = E_X(E[Y(1)|X]) = E[Y(1)].$$

An unbiased estimator for $E[Y(1)]$ can then be

$$\widehat{E[Y(1)]} = \frac{1}{n}\sum_{i=1}^{n}\frac{Z_i Y_i}{e(X_i)}. \tag{9}$$

Weighted ATE Estimator(Cont.)

Using a similar argument, we can rewrite $E[Y(0)]$ as

$$E\left\{\frac{(1-Z)\cdot Y}{1-e(X)}\right\} = E[Y(0)]$$

Its natural estimator would be

$$\widehat{E[Y(0)]} = \frac{1}{n}\sum_{i=1}^{n}\frac{(1-Z_i)Y_i}{1-e(X_i)}. \tag{10}$$

Weighted ATE (Cont.)

The ATE can be written as

$$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$
$$= E\left\{\frac{Z \cdot Y}{e(X)}\right\} - E\left\{\frac{(1-Z) \cdot Y}{1-e(X)}\right\}. \qquad (11)$$

Consequently, an unbiased estimator of the ATE (11), based on (9) and (10), can be written as

$$\widehat{ATE}_w = E\widehat{[Y(1)}] - E\widehat{[Y(0)}]$$
$$= \frac{1}{n}\sum_{i=1}^{n}\frac{Z_i Y_i}{e(X_i)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-Z_i)Y_i}{1-e(X_i)} \qquad (12)$$

Weighted ATT

Using similar arguments as the ATE, we can write the ATT as

$$
\begin{aligned}
E[Y(1) - Y(0)|Z = 1] &= E[Y(1)|Z = 1] - E[Y(0)|Z = 1] \\
&= E[ZY] - E\left[\frac{(1 - Z)Ye(X)}{1 - e(X)}\right]. \qquad (13)
\end{aligned}
$$

Consequently, an unbiased estimator of the ATT is

$$
\begin{aligned}
\widehat{ATT}_w &= E[Y\widehat{(1)|Z} = 1] - E[Y\widehat{(0)|Z} = 1] \\
&= \frac{1}{n}\sum_{i=1}^{n} Z_i Y_i - \frac{1}{n}\sum_{i=1}^{n} \frac{(1 - Z_i)Y_i \cdot e(X_i)}{1 - e(X_i)}. \qquad (14)
\end{aligned}
$$

Average Controlled Difference (ACD)

- Let X be a vector of covariates with PDF $f(X_i)$
- Let $f(x)h(x)$ be the target population density, where $h(x)$ is the weight function of x
- The Average Controlled Difference (ACD) is defined as

$$\begin{aligned}\tau(x) &= E[Y(1) - Y(0)|X = x] \\ &= E[Y|Z = 1, X = x] - E[Y|Z = 0, X = x]\end{aligned}$$

Average Controlled Difference (ACD) (Cont.)

For the continuous case, the weighted ACD can be defined as

$$\tau_h = \frac{\int \tau(x) f(x) h(x) dx}{\int f(x) h(x) dx} \tag{15}$$

Let $f_z(x) = P(X = x|Z = z)$. Note that,

$$f_1(x) = P(X = x|Z = 1) = \frac{f(x) \cdot e(x)}{P(Z = 1)},$$

implying that

$$f_1(x) \propto f(x) \cdot e(x), \quad \text{and, similarly,}$$
$$f_0(x) \propto f(x) \cdot (1 - e(x)).$$

Balancing Weights

- For $Z = 1$, $f(x)h(x) \propto \frac{f_1(x)}{e(x)} \cdot h(x) = f_1(x)\omega_1(x)$
- For $Z = 0$, $f(x)h(x) \propto \frac{f_0(x)}{1-e(x)} \cdot h(x) = f_0(x)\omega_0(x)$
- The omegas are the balancing weights, where

$$\omega_1(x) = \frac{h(x)}{e(x)},\ \omega_0(x) = \frac{h(x)}{(1 - e(x))} \tag{16}$$

- $h(x)$ can be set to anything. For example, in the previous slide
  - Set $h(x) = 1$, ACD $=$ ATE
  - Set $h(x) = e(x)$, ACD $=$ ATT
  - Set $h(x) = 1 - e(x)$, ACD $=$ ATC

Weighted Average Treatment Effect (WATE)

The Weighted Average Treatment Effect (WATE) is defined as

$$\hat{\tau}_h = \frac{\sum_i \omega_1(x_i) Z_i Y_i}{\sum_i \omega_1(x_i) Z_i} - \frac{\sum_i \omega_0(x_i)(1 - Z_i) Y_i}{\sum_i \omega_0(x_i)(1 - Z_i)}, \tag{17}$$

and will be our estimator of the ACD τ_h,

$$\tau_h = \frac{\int \tau(x) f(x) h(x) dx}{\int f(x) h(x) dx}$$

Consistency of an Estimator

Definition 3 (Consistency)

Let $Z_1, Z_2, \ldots$ be iid random variables and Z be a random variable. The random variable Z_n converges in probability to Z, or $Z_n \xrightarrow{p} Z$ if

$$\lim_{n\to\infty} P\big(|Z_n - Z| \leq \epsilon\big) = 1$$

Definition 4 (Consistency of an Estimator)

Let $X_1, X_2, \ldots$ be a sequence of iid random variables drawn from a distribution with parameter θ and $\hat{\theta}$ as its estimator. This estimator $\hat{\theta}$ is a consistent estimator of θ if

$$\hat{\theta} \xrightarrow{p} \theta \quad \text{or} \quad \lim_{n\to\infty} P\big(|\hat{\theta}(X_1, \ldots, X_n) - \theta| \leq \epsilon\big) = 1.$$

Limiting Distribution Lemmas

Lemma 1 (Strong Law of Large Numbers)

The sample average converges almost surely to the expected value,

$$\bar{X}_n \xrightarrow{a.s.} \mu, \quad \text{when} \quad n \to \infty,$$

or, in other words,

$$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1.$$

Lemma 2 (Slutsky's Theorem)

Let $\hat{\theta} \xrightarrow{p} \theta$ and $\hat{\eta} \xrightarrow{p} \eta$. Then, for any continuous multivariate valued function g,

$$g(\hat{\theta}, \hat{\eta}) \xrightarrow{p} g(\theta, \eta).$$

Consistency of $\hat{\tau}_h$

Theorem

$\hat{\tau}_h$ is a consistent estimator of τ_h.

Proof: We start by considering the continuous version of the ACD:

$$\tau_h = \frac{\int \tau(x) f(x) h(x) dx}{\int f(x) h(x) dx}$$

Consistency of $\hat{\tau}_h$ (Cont.)

First, look at the top of the ACD
Rewrite $\tau(x)$ as

$$\begin{aligned}\tau(x) &= E_{Y,Z|X}[Y(1) - Y(0)|X = x] \\ &= E_{Y,Z|X}\left[\frac{Z \cdot Y}{e(x)}\middle|X = x\right] - E_{Y,Z|X}\left[\frac{(1-Z) \cdot Y}{1-e(x)}\middle|X = x\right]\end{aligned}$$

Plug this back into the original integral to get

$$\int \tau(x)f(x)h(x)dx = \int \left(E_{Y,Z|X}\left[\frac{h(x)}{e(x)} \cdot ZY\middle|x\right] - E_{Y,Z|X}\left[\frac{h(x)}{1-e(x)} \cdot (1-Z)Y\middle|x\right]\right) f(x)dx$$

Consistency of $\hat{\tau}_h$ (Cont.)

For the bottom part of the ACD, rewrite $h(x)$ into a piecewise function in terms of conditional expectations since $Z|x$ is a Bernoulli random variable

$$h(x) = \begin{cases} E_{Z|X}\left[\frac{h(x)}{e(x)} \cdot Z \middle| x\right], & \text{or} \\ E_{Z|X}\left[\frac{h(x)}{1-e(x)} \cdot (1-Z) \middle| x\right]. \end{cases}$$

Consistency of $\hat{\tau}_h$ (Cont.)

Putting the top and bottom together yields

$$\tau_h = \frac{\int E_{Y,Z|X}\left[\frac{h(x)}{e(x)} \cdot ZY \middle| x\right] f(x)dx}{\int f(x)h(x)dx} - \frac{\int E_{Y,Z|X}\left[\frac{h(x)}{1-e(x)} \cdot (1-Z)Y \middle| x\right] f(x)dx}{\int f(x)h(x)dx}$$
$$= \frac{\int E_{Y,Z|X}\left[\omega_1(x) \cdot ZY \middle| x\right] f(x)dx}{\int E_{Z|X}\left[\omega_1(x) \cdot Z \middle| x\right] f(x)dx} - \frac{\int E_{Y,Z|X}\left[\omega_0(x) \cdot (1-Z)Y \middle| x\right] f(x)dx}{\int E_{Z|X}\left[\omega_0(x) \cdot (1-Z) \middle| x\right] f(x)dx} \tag{18}$$

$$\hat{\tau}_h = \frac{\sum_i \omega_1(x_i)Z_iY_i}{\sum_i \omega_1(x_i)Z_i} - \frac{\sum_i \omega_0(x_i)(1-Z_i)Y_i}{\sum_i \omega_0(x_i)(1-Z_i)}, \tag{19}$$

- Note that each component in the WATE estimator (19) converges to each component in (18) by the Law of Large Numbers
- Also, by Slutsky's Theorem, the WATE estimator $\hat{\tau}_h$ will converge in probability to τ_h

Conditional Variance

Focus now shifts to variance of estimator $\hat{\tau}_h$. Start by expanding the variance using iterated expectations, given $\mathbf{X} = \{x_1, ..., x_n\}$:

$$Var(\hat{\tau}_h) = E(\hat{\tau}_h^2) - [E(\hat{\tau}_h)]^2 = E_X[Var(\hat{\tau}_h|\mathbf{X})] + Var_X(E[\hat{\tau}_h|\mathbf{X}])$$

- First term is expected variation directly due to variation in **X**, and is typically much larger than the second term
- Second term is unexplained variation coming from somewhere other than **X**, where the estimation of this term involves the outcome model
- Benefit of selecting weights that incorporate the outcome model don't justify risk of biasing model specification to attain variance results, so we only focus on first term

Conditional Variance (Cont.)

Theorem

As $n \to \infty$, the expectation of the conditional variance of the estimator $\hat{\tau}_h$, given the sample $\mathbf{X} = (x_1, ..., x_n)$ converges:

$$n \cdot E_X(Var[\hat{\tau}_h|\mathbf{X}]) \to \frac{\int f(x)h(x)^2[v_1(x)/e(x) + v_0(x)/1 - e(x))]dx}{\int [h(x)f(x)]^2 dx}$$

where $v_z(x) = Var[Y(z)|\mathbf{X}]$.

Conditional Variance Convergence (Cont.)

Proof: Start by conditioning further on $\mathbf{Z} = (z_1, \ldots, z_n)$, which yields

$$Var(\hat{\tau}_h | \mathbf{X}, \mathbf{Z}) = \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i}{e(x_i)}[(h(x_i))^2/e(x_i)] \cdot v_1(x_i)}{n[\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i}{e(x_i)} h(x_i)]^2} + \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{1-Z_i}{1-e(x_i)}[(h(x_i))^2/(1-e(x_i))] \cdot v_0(x_i)}{n[\frac{1}{n}\sum_{i=1}^{n} \frac{1-Z_i}{1-e(x_i)} h(x_i)]^2}$$

Note that

$$E_Z\left[\frac{Z_i}{e(x_i)}|x_i\right] = \frac{1}{e(x_i)} E_Z[Z_i|x_i] = \frac{1}{e(x_i)} \cdot e(x_i) = 1 \tag{20}$$

$$E_Z\left[\frac{1-Z_i}{1-e(x_i)}|x_i\right] = \frac{1}{1-e(x_i)} E_Z[1-Z_i|x_i] = \frac{1}{1-e(x_i)} \cdot (1-e(x_i)) = 1 \tag{21}$$

By Strong Law of Large Numbers, the sample version of LHS of (19) and (20) will both approach 1.

Conditional Variance Convergence (Cont.)

Next, let us average the $Var(\hat{\tau}_h|\mathbf{X},\mathbf{Z})$ over the distribution of $\mathbf{Z}$ and Strong Law of Large Numbers together with Slutsky's Theorem results to

$$\begin{aligned} n \cdot E_X Var(\hat{\tau}_h|\mathbf{X}) &= n \cdot E_X E_Z[Var(\hat{\tau}_h|\mathbf{X},\mathbf{Z})] \\ &\longrightarrow \frac{\int \left[\frac{v_1(x)}{e(x)} + \frac{v_0(x)}{1-e(x)}\right] \cdot h(x)^2 f(x)dx}{\left(\int h(x)f(x)dx\right)^2} \end{aligned}$$

If $v_1(x) = v_0(x) = v$, then the asymptotic variance of $\hat{\tau}_h$ can simplify to

$$n \cdot E_X Var[\hat{\tau}_h|\mathbf{X}] \to v \cdot \frac{\int f(x)h(x)^2/[e(x)(1-e(x))]dx}{\left(\int h(x)f(x)dx\right)^2}$$

Smallest Asymptotic Variance of $\hat{\tau}_h$

Lemma 3 (Cauchy-Schwarz Inequality)

If X and Y are random variables, then $[E(XY)]^2 \leq E(X^2)E(Y^2)$

Corollary 1

The function $h(x) \propto e(x)(1 - e(x))$ gives the smallest asymptotic variance for the weighted estimator $\hat{\tau}_h$ among all h's under homoscedasticity, and as $n \to \infty$,

$$n \cdot \min_h (E_X Var[\hat{\tau}_h | \mathbf{X}]) \to \frac{v}{C_h^2} \cdot \int f(x)e(x)(1 - e(x))dx,$$

where $C_h = \int h(x)f(x)dx$.

Smallest Asymptotic Variance of $\hat{\tau}_h$ (Cont.)

Using the results of Lemma 3, we have

$$\left(E[h(x)]\right)^2 = \left[E\left(\frac{h(x)}{\sqrt{e(x)(1-e(x))}}\sqrt{e(x)(1-e(x))}\right)\right]^2$$
$$\leq E\left[\frac{h(x)^2}{e(x)(1-e(x))}\right]E[e(x)(1-e(x))]$$

or

$$E\left[\frac{h(x)^2}{e(x)(1-e(x))}\right] \geq \frac{\left(E[h(x)]\right)^2}{E[e(x)(1-e(x))]}$$

Smallest Asymptotic Variance of $\hat{\tau}_h$ (Cont.)

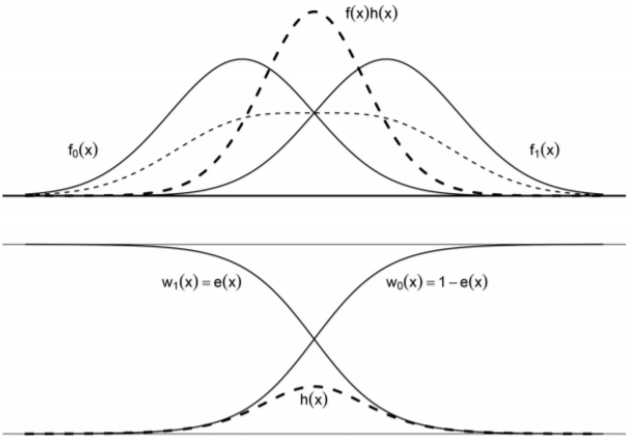
Applying this to the right hand side of the last theorem yields

$$
\begin{aligned}
n \cdot E_X Var[\hat{\tau}_h | \mathbf{X}] &\to \frac{v}{C_h^2} \cdot E\left[\frac{h(x)^2}{e(x)(1-e(x))}\right] \\
&\geq \frac{v}{C_h^2} \cdot \frac{\big(E[h(x)]\big)^2}{E[e(x)(1-e(x))]} \\
&= \frac{v}{C_h^2} \cdot E[e(x)(1-e(x))],
\end{aligned}
$$

The last equality is true if $h(x) = e(x)(1-e(x))$.

Overlap Weighting

- Overlap weight can be defined as $h(x) = e(x)(1 - e(x))$
- It follows that $\omega_1(x) = 1 - e(x)$ and $\omega_0(x) = e(x)$



Advantages of Overlap Weighting

Average treatment effect of overlap population (ATO) can adapt to any distribution of covariates and propensities

- For small propensity to treatment, $ATO \approx ATT$
- For small propensity to control, $ATO \approx ATC$
- For balanced treatment to control, $ATO \approx ATE$

Balance of Overlap Weights

Theorem

When the propensity scores are estimated by maximum likelihood under a logistic regression model, the overlap weights lead to exact balance in the means of any included covariate between treatment and control groups. In other words,

$$\frac{\sum_i x_{ik} Z_i (1 - \hat{e}(x_i))}{\sum_i Z_i (1 - \hat{e}(x_i))} = \frac{\sum_i x_{ik} (1 - Z_i) \hat{e}(x_i)}{\sum_i (1 - Z_i) \hat{e}(x_i)}, \quad k = 1, ..., K.$$

Example: Right Heart Catheterization (RHC)

- RHC is "a diagnostic procedure for directly measuring cardiac function in critically ill patients"
- Publicly available dataset contains data on 5735 adult patients
- 2184 patients underwent RHC procedure (Z=1)
- Remaining 3551 patients didn't undergo procedure (Z=0)
- Outcome is binary variable `dth30` which measured whether or not a patient survived 30 days after admission (Y=1 if they did, Y=0 if not)
- 72 Covariates - 21 continuous, 25 binary, 26 dummy formed by breaking up 6 categorical variables

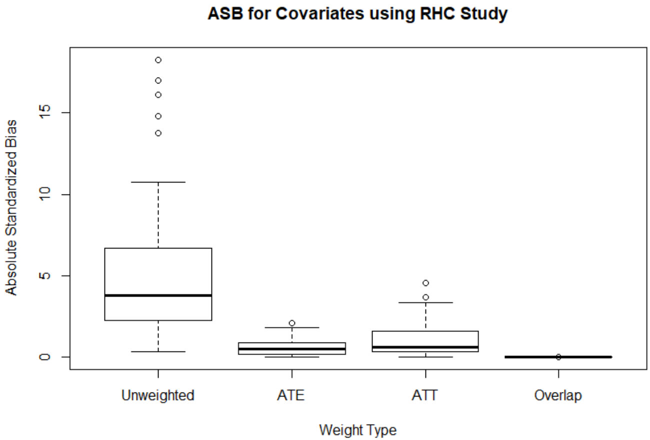
Analyzing the RHC Data - ASB

- Propensity score estimated under logistic model against treatment varaible `swang1`
- Covariate mean balance measured using Absolute Standardized Bias (ASB):

$$ASB = \left| \frac{\sum_{i=1}^{N} x_i Z_i \omega_i}{\sum_{i=1}^{N} Z_i \omega_i} - \frac{\sum_{i=1}^{N} x_i (1 - Z_i) \omega_i}{\sum_{i=1}^{N} (1 - Z_i) \omega_i} \right| \Big/ \sqrt{s_1^2 / N_1 + s_0^2 / N_0}$$

- Note that s_z^2 is the variance of the unweighted covariate of interest for treatment group z and N_z is the sample size of each treatment group z

Analyzing the RHC Data - ASB (Cont.)



Analyzing the RHC Data - WATE

- The next step is to estimate the WATE at each weight:

$$\widehat{WATE} = \frac{\sum_i \omega_1(x_i) Z_i Y_i}{\sum_i \omega_1(x_i) Z_i} - \frac{\sum_i \omega_0(x_i)(1 - Z_i) Y_i}{\sum_i \omega_0(x_i)(1 - Z_i)}.$$

- Along with every other weight, a truncated ATT WATE was also calculated using data points with propensity scores from 0.1 to 0.9
- Standard errors calculated using basic bootstrapping techniques

	Unweighted	ATE	ATT	Overlap	Trunc. ATT
Estimate $\cdot 10^2$	7.36	5.50	5.40	6.41	5.77
SE $\cdot 10^2$	1.39	1.82	2.36	1.47	1.67

Simulation Setup

- Six variables $V_1 - V_6$ are generated from a multivariate normal distribution with mean 0 and a covariance matrix with 1 for the diagonals and 0.5 for everything else
- $V_1 - V_3$ were then kept as continuous variables $X_1 - X_3$
- $V_4 - V_6$ were dichotomized into $X_4 - X_6$ by setting negative values to 1 and positive values to 0

Simulation Setup (Cont.)

- Propensity scores calculated using the following logistic model:

$$e(X_n) = (1+exp[-(\alpha_0+\alpha_1 X_1+\alpha_2 X_2+\alpha_3 X_3+\alpha 4 X_4+\alpha_5 X_5+\alpha_6 X_6)])^{-1}$$

- The parameters are

$$(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0.15\gamma, 0.3\gamma, 0.3\gamma, -0.2\gamma, -0.25\gamma, -0.25\gamma)$$

- The γ values range from 1 (high overlap between groups) to 4 (low overlap between groups
- α_0 represents overall treatment prevalence in each sample (0.1 or 0.4)
- Each observation assigned to group by simulating a Bernoulli model based on its propensity score
- Outcome variable Y calculated as follows, with $\Delta = 0.75$:

$$E[Y|Z,X] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \Delta Z$$

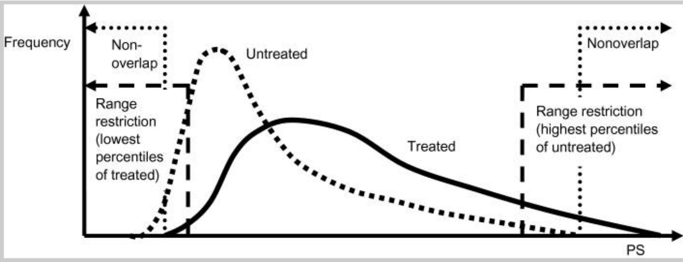
Weighting Methods Used in Simulation

- Crude estimate, where weights $\omega_i = 1$
- Overlap weighting, where $\omega_1 = 1 - \hat{e}(x_i)$ if $Z_i = 1$ and $\omega_0 = \hat{e}(x_i)$ if $Z_i = 0$
- Untrimmed IPW, where $\omega_1 = \frac{1}{\hat{e}(x_i)}$ if $Z_i = 1$ and $\omega_0 = \frac{1}{1-\hat{e}(x_i)}$ if $Z_i = 0$
- Symmetrically trimmed IPW, where individuals with propensity scores outside the range $[\alpha, 1-\alpha]$ are eliminated. Possible α values are $\alpha = 0.05, \alpha = 0.10$, and $\alpha = 0.15$

Weighting Methods Used in Simulation (Cont.)

Asymmetric trimming IPW method also analyzed

- Step 1: Remove individuals outside the overlap of propensity scores between treatment and control groups
- Step 2: Remove treatment units with PS below q quantile of treated units. Remove control units with PS above $1 - q$ quantile of all control units
- Possible q values are $q = 0, q = 0.01$, and $q = 0.05$



Bias of Estimators

- 1000 replications are performed to get 1000 WATE estimates
- Mean of estimates taken
- Mean then subtracted by treatment effect ($\Delta = 0.75$)

Bias of Estimators Results TP=0.4

- Bias increases with increasing levels of γ
- True for Crude estimator, Untrimmed IPW, and Asymmetric Trimmed IPW
- Overlap Weighting, and Symmetric Trimmed IPW show no bias

Estimator	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Crude	-2.00	-3.18	-3.77	-4.08
Overlap Weighting	0.00	-0.01	-0.02	-0.02
IPW				
No trimming	0.00	-0.04	-0.23	-0.54
Symmetric trimming				
α=0.05	0.00	-0.04	-0.05	-0.03
α=0.10	0.00	-0.02	-0.02	-0.04
α=0.15	-0.01	-0.01	-0.02	-0.02
Asymmetric trimming				
$q = 0$	0.18	0.44	0.74	0.90
$q = 0.01$	-0.25	-0.47	-0.54	-0.56
$q = 0.05$	-1.03	-1.55	-1.69	-1.60

Bias of Estimators Results TP=0.1

- Same general results as higher treatment prevalence

Estimator	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Crude	-2.01	-3.19	-3.78	-4.09
Overlap Weighting	0.00	-0.01	-0.02	-0.02
IPW				
No trimming	-0.01	-0.05	-0.23	-0.58
Symmetric trimming				
$\alpha = 0.05$	-0.01	-0.04	-0.05	-0.03
$\alpha = 0.10$	-0.01	-0.02	-0.03	-0.03
$\alpha = 0.15$	-0.01	-0.01	-0.03	-0.02
Asymmetric trimming				
$q = 0$	0.18	0.46	0.77	0.91
$q = 0.01$	-0.25	-0.41	-0.54	-0.56
$q = 0.05$	-1.03	-1.53	-1.68	-1.57

RMSE of Estimators

- Once again, 1000 WATE estimates based on 1000 replications are found
- RMSE of each group of replications found given by

$$RMSE(\hat{\theta}) = \sqrt{Var(\hat{\theta}) + \left[E[\hat{\theta}] - \theta\right]^2}$$

RMSE of Estimators Results TP=0.4

- RMSE results closely mirror bias results - RMSE increases with increasing γ for Crude, Untrimmed IPW,and Asymmetric Trimming
- No changes in RMSE for Overlap Weighting and Symmetric Trimming

Estimator	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Crude	2.02	3.19	3.78	4.08
Overlap Weighting	0.29	0.29	0.30	0.32
IPW				
No trimming	0.35	0.60	0.97	1.36
Symmetric trimming				
$\alpha = 0.05$	0.35	0.41	0.42	0.40
$\alpha = 0.10$	0.35	0.33	0.33	0.34
$\alpha = 0.15$	0.32	0.30	0.30	0.32
Asymmetric trimming				
$q = 0$	0.36	0.65	1.02	1.31
$q = 0.01$	0.41	0.62	0.73	0.76
$q = 0.05$	1.08	1.60	1.76	1.68

RMSE of Estimators Results TP=0.1

- Same general results as higher treatment prevalence results

Estimator	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
Crude	2.03	3.20	3.79	4.10
Overlap Weighting	0.29	0.30	0.30	0.30
IPW				
No trimming	0.35	0.64	1.03	1.41
Symmetric trimming				
$\alpha = 0.05$	0.35	0.43	0.41	0.38
$\alpha = 0.10$	0.35	0.34	0.32	0.34
$\alpha = 0.15$	0.33	0.30	0.30	0.33
Asymmetric trimming				
$q = 0$	0.37	0.69	1.08	1.33
$q = 0.01$	0.41	0.58	0.73	0.77
$q = 0.05$	1.09	1.59	1.75	1.65

95 Percent CI Coverage

- Generated 100 different datasets and bootstrapped each dataset 100 times
- Used to find mean, variance of any given estimator for each dataset
- Then figure out if CI of estimator contains 0.75
- Results line up with Bias and RMSE results

Estimator	γ=1	γ=2	$\gamma = 3$	$\gamma = 4$
Crude	0.00	0.00	0.00	0.00
Overlap Weighting	0.96	0.97	0.94	0.94
IPW				
No trimming	0.92	0.87	0.40	0.02
Symmetric trimming				
$\alpha = 0.15$	0.94	0.97	0.95	0.90

References

- Rosenbaum, Paul R., and Donald B. Rubin. "The central role of the propensity score in observational studies for causal effects." Biometrika 70.1 (1983): 41-55.
- Rosenbaum, Paul R., and Donald B. Rubin. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." Journal of the American Statistical Association 79, no. 387 (September 1984): 193-206.
- Rubin, Donald B. "Using propensity scores to help design observational studies: application to the tobacco litigation." Health Services and Outcomes Research Methodology 2.3-4 (2001): 169-188.
- Dietrich, Cecile C., and Eric J. Lichtenberger. "Using Propensity Score Matching to Test the Community College Penalty Assumption." The Review of Higher Education 38, no. 2 (Winter 2015): 193-219.
- Helmreich, James E., and Robert M. Pruzek. "PSAgraphics: AnRPackage to Support Propensity Score Analysis." Journal of

The End!