Balancing Covariates via Propensity Score Weighting

Kari Lock Morgan

Department of Statistics
Penn State University
klm47@psu.edu

Stochastic Modeling and Computational Statistics Seminar
October 17, 2014

Joint work with Fan Li (Duke) and Alan Zaslavsky (Harvard)

Outline

1. Causal inference in observational studies - a brief overview
2. Introduce a general class of balancing weights
3. Propose new overlap weights and show some optimality properties
4. Illustrate with examples

Causal Inference in Observational Studies

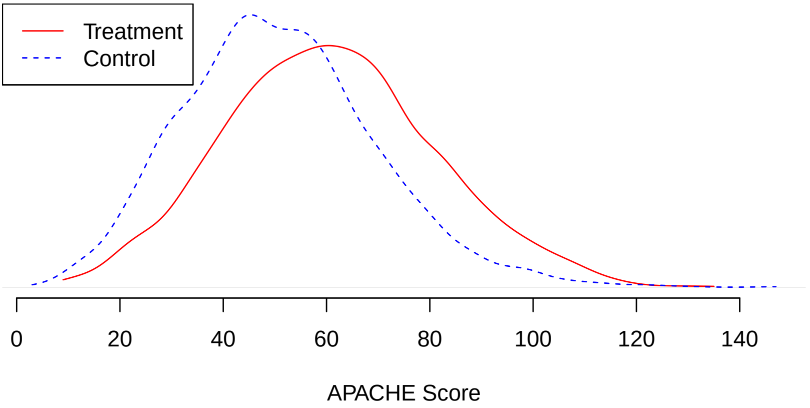
- Ideal goal: estimate the causal effect of a treatment using observational data
- Problem: Without randomization to treatment groups, severe covariate imbalance is likely
- Realistic goal: Balance observed covariates between treatment groups

Example: Right Heart Catheterization

- Right heart catheterization (RHC) is an invasive diagnostic procedure to assess cardiac function
- What is the causal effect of right heart catheterization on survival?
- 2184 treatment (RHC), 3551 control (no RHC)
- Observational data (Murphy and Cluff, 1990)
- Covariate imbalance

Example: Right Heart Catheterization

Figure: Imbalance in APACHE, Acute Physiology and Chronic Health Evaluation Score, measured before procedure.



Example: Right Heart Catheterization

- APACHE scores for a random subset of 20 patients:
  - Treatment: 34, 41, 42, 43, 49, 79, 80, 84
  - Control: 9, 19, 35, 40, 44, 45, 48, 50, 51, 53, 53, 60
- Get rid of units with no comparable units in other group?
  - Treatment: 34, 41, 42, 43, 49
  - Control: 35, 40, 44, 45, 48, 50, 51, 53, 53
- Options for covariate balance:
  - Matching - match each unit in treatment group to similar unit(s) in control group, discard the rest
    - (34, 35), (41, 40), (42, 44), (43, 45), (49, 48)
  - Subclassification - compare units within similar subclasses, then take a (weighted) average across subclasses
    - (34, 35), (40, 41, 42, 43, 44, 45), (48, 49, 50, 51, 53, 53)
  - Weighting - weight each unit so the weighted covariate distributions are similar

Propensity Scores

- This gets much more complicated with many covariates.
- Wouldn't it be nice if we could just balance on a single number summarizing all covariates...
- The propensity score allows us to do this!
- The propensity score, $e(x)$, is the probability a unit belongs to the treatment group, based on observed covariates:

$$e(x) = \Pr(Z_i = 1 | X_i = x),$$

  where Z_i indicates treatment group ($Z_i = 1$ for treatment and $Z_i = 0$ for control) and X_i denotes the covariates.
- Amazing fact: balancing just the propensity score yields balance on all covariates included in the propensity score model!

Covariate Distributions

- Population density of the covariates X is $f(x)$
- Density for group $Z = z$ is $f_z(x) = P(X = x|Z = z)$
- Then

$$f_z(x) = P(X = x \mid Z = z) = \frac{P(Z = z|X = x)P(X = x)}{P(Z = z)},$$

- So $f_1(x) \propto f(x)e(x)$ and $f_0(x) \propto f(x)(1 - e(x))$
- GOAL: make $f_1(x) \propto f_0(x)$
- One solution: Use weights, $w_z(x)$, such that $f_1(x)w_1(x) \propto f_0(x)w_0(x)$

Balancing weights

- We propose the following class of balancing weights:

$$\begin{cases} w_1(x) \propto \frac{h(x)}{e(x)}, \\ w_0(x) \propto \frac{h(x)}{1-e(x)}, \end{cases}$$

  where $h(\cdot)$ is a pre-specified function.

- The weighted covariate distributions in the two groups have the same target density $\propto f(x)h(x)$:

$$f_1(x)w_1(x) \propto f(x)e(x)\frac{h(x)}{e(x)} = f(x)h(x),$$

$$f_0(x)w_0(x) \propto f(x)\left(1 - e(x)\right)\frac{h(x)}{1 - e(x)} = f(x)h(x).$$

Examples of target population and balancing weights

target population	$h(x)$	estimand	weight (w_1, w_0)
combined	1	ATE	$\left(\frac{1}{e(x)}, \frac{1}{1-e(x)}\right)$ [HT]
treated	$e(x)$	ATT	$\left(1, \frac{e(x)}{1-e(x)}\right)$
control	$1-e(x)$	ATC	$\left(\frac{1-e(x)}{e(x)}, 1\right)$
truncated combined	$\mathbf{1}(\alpha < e(x) < 1-\alpha)$	ATTrunc	$\left(\frac{\mathbf{1}(\alpha<e(x)<1-\alpha)}{e(x)}, \frac{\mathbf{1}(\alpha<e(x)<1-\alpha)}{1-e(x)}\right)$
overlap	$e(x)(1-e(x))$	ATO	$(1-e(x), e(x))$

Estimands and Estimators

- Potential outcome framework: $Y_i(1)$, $Y_i(0)$
- Conditional average treatment effect (ATE)

$$\tau(x) \equiv \mathbb{E}(Y(1)|X = x) - \mathbb{E}(Y(0)|X = x).$$

- Estimand is average (ATE) over a target population with density $\propto f(x)h(x)$:

$$\tau_h \equiv \frac{\int \tau(dx) f(x) h(x) \mu(dx)}{\int f(x) h(x) \mu(dx)}.$$

- τ_h can be estimated by weighted averages:

$$\hat{\tau}_h^w = \frac{\sum_{i:Z_i=1} Y_i w_1(x_i)}{\sum_{i:Z_i=1} w_1(x_i)} - \frac{\sum_{i:Z_i=0} Y_i w_0(x_i)}{\sum_{i:Z_i=0} w_0(x_i)}.$$

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

Theorem

Given the normalizing constraint $\int f(x)h(x)\mu(dx) = 1$, the large-sample variance of the estimator $\hat{\tau}_h$ is:

$$\mathbb{V}[\hat{\tau}_h] = \int f(x)h(x)^2 \left[\frac{v_1(x)}{e(x)} + \frac{v_0(x)}{1-e(x)}\right] \mu(dx)/N,$$

where $v_z(x)$ is the variance of Y in a neighborhood dx of x in the $Z = z$ group.

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

Theorem

Assuming $v_0(x) \equiv v_1(x) \equiv v$, the function $h(x) = e(x)(1 - e(x))$ gives the smallest asymptotic variance for the weighted estimator $\hat{\tau}_h$, and

$$min\{\mathbb{V}[\hat{\tau}_h]\} = \frac{v}{N} \int f(x)e(x)(1 - e(x))\mu(dx).$$

Overlap weights

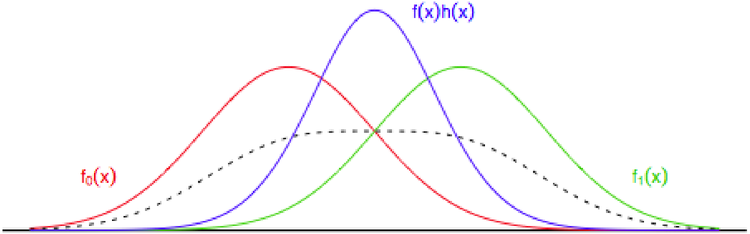
- We propose a new weight by letting $h(x) = e(x)(1 - e(x))$, leading to the overlap weights:

$$w_1(x) \propto 1 - e(x),$$
$$w_0(x) \propto e(x).$$

- Target population $f(x)e(x)(1 - e(x))$
  - "Marginal" units who may or may not receive the treatment (Rosenbaum, 2012).
  - Defined by overlap of covariates...

Overlap Weights

Figure: Densities for the treatment group, $f_1(x)$, control group, $f_0(x)$, and overlap population, $f(x)h(x)$.



Exact Balance

Theorem

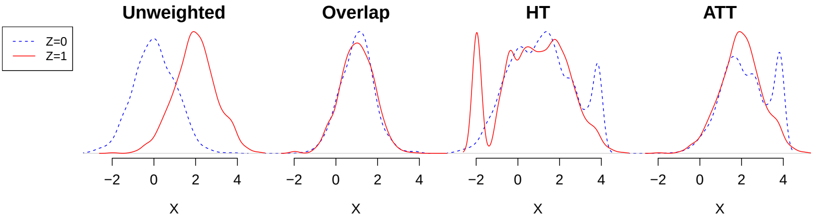
When the propensity scores are estimated from a logistic regression model with main effects, logit$\{e(x_i)\} = \beta_0 + \beta' x_i$, the overlap weights lead to exact balance in any included covariate between treatment and control groups. That is,

$$\frac{\sum_i x_{i,k} Z_i (1 - \hat{e}_i)}{\sum_i Z_i (1 - \hat{e}_i)} = \frac{\sum_i x_{i,k} (1 - Z_i) \hat{e}_i}{\sum_i (1 - Z_i) \hat{e}_i}.$$

Simulated Example

$n_0 = n_1 = 1000$ units, with $X_i \sim N(0,1) + 2Z_i$.

Figure: Original covariate distributions within each treatment group, and weighted covariate distributions with overlap, HT, ATT weights.



	Unweighted	Overlap	HT	ATT
$\bar{x}_1$	1.98	1.01	0.74	1.98
$\bar{x}_0$	0.03	1.01	1.19	2.22

Simulated Example

- A single covariate: $X_i \sim N(0,1) + 2Z_i$.
- Outcome model with additive treatment effect: $Y_i \sim X_i + \tau Z_i + N(0,1)$, with $\tau = 1$.
- Use the nonparametric estimator $\hat{\tau}_h^w$ with different weights:

	Unweighted	Overlap	HT	ATT
$\hat{\tau}$	2.945	1.000	0.581	0.640
$SE(\hat{\tau})$	0.054	0.038	0.386	0.402

Racial Disparity in Medical Expenditure

- Goal: estimate racial disparity in medical expenditures after balancing covariates (Le Cook et al., 2010)
- Race is not manipulable so comparisons are descriptive, not causal
- Data: 2009 Medical Expenditure Panel Survey: 9830 non-Hispanic Whites, 4020 Blacks, 1446 Asians, 5280 Hispanics
- Three independent comparisons; comparing non-Hispanic Whites to each minority group
- Logistic regression to estimate propensity scores, 31 covariates (5 continuous, 26 binary)
- Ignore survey weights here, but weighting allows easy incorporation of survey weights

Racial Disparity in Medical Expenditure

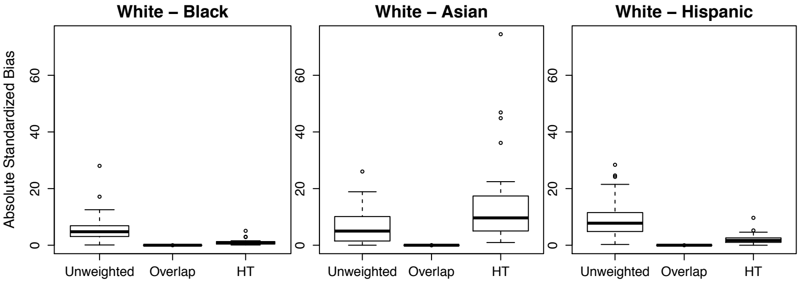


Figure: Covariate balance (absolute standardized bias) with no weights, overlap weights, and HT weights.

Racial Disparity in Medical Expenditure

- One Asian woman has over 30% of the weight! (out of 1446 Asians)
- 78 year old Asian lady with a BMI of 55.4: e(x) = 0.9998
- Common practice:
  - Eliminate cases with propensity scores close to 0 or 1
  - Truncate propensity scores or weights
  - Can lead to ad hoc changes to target population
  - Results can be very sensitive to truncation choice
- The overlap weights avoid these extreme weights and avoid an abrupt threshold for elimination or truncation

Racial Disparity in Medical Expenditure

Table: Unweighted, overlap weights, and HT weighting estimates (SE) for difference in yearly medical expenditure.

	Unweighted	Overlap	HT
White - Black	$786 (222)	$824 (185)	$856 (200)
White - Asian	$2764 (209)	$1227 (205)	$2167 (640)
White - Hispanic	$2599 (174)	$1212 (171)	$596 (323)

Right heart catheterization (RHC)

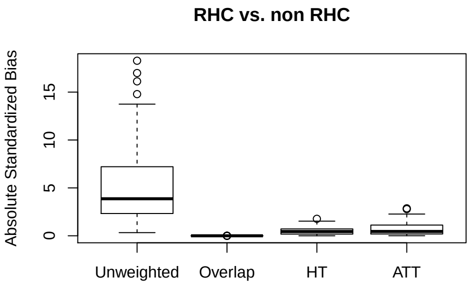


Table: Estimated treatment effect (in %) with different weights

	unweighted	overlap	HT	ATT
$\hat{\tau}_h$	7.36	6.54	5.93	5.81
$SE(\hat{\tau}_h)$	1.27	1.32	2.46	2.67

Advantages of the overlap weight

Statistical advantages

- Minimizes asymptotic variance of the weighted average estimator among all balancing weights.
- Perfect (exact small-sample) balance for means of included covariates in logistic propensity score model.
- Weights are bounded (unlike HT, etc.).
- Avoids artificially truncating weights or eliminating cases.

Scientific advantages

- Clinical equipoise.
- The “marginal units” are likely the group who are responsive to policy intervention.

Summary

- Unified framework for use of weighting to balance covariates for any target population.
- The general class of balancing weights balance covariates and include many of the existing weights.
- A new weighting method, the overlap weights, have desirable properties

Arxiv: `http://arxiv.org/abs/1404.1785`