Weighting beyond Horvitz-Thompson in Causal Inference

Fan Li

Department of Statistical Science
Duke University

May 5, 2014

Joint work with Alan Zaslavsky and Kari Lock Morgan

Introduction

- Population-based observational data increasingly used for comparative effectiveness studies.
- Balancing covariate distributions across groups for descriptive and causal analyses.
- One common approach is weighting, usually based on inverse probability (Horvitz-Thompson) weights.
- Several other weights are available: for the treated population (Hirano and Imbens, 2001), or for a subpopulation with sufficient overlap (Crump et al., 2009)
- Our goals:
  1. Introduce a general framework for a class of balancing weights that unifies the existing weights;
  2. Propose the new overlap weights and show some optimality properties

Standard Setup

- Data: a random sample of n units from a population.
- Treatment status $Z_i (= 0, 1)$ and covariates $X_i = (X_{i1}, ...X_{ip})$ are observed.
- Potential outcomes are $Y_i(0)$, $Y_i(1)$, but only $Y_i(Z_i)$ is observed.
- Estimand: Average Treatment Effect (ATE)

$$\tau^{\text{ATE}} = \mathbb{E}[Y(1) - Y(0)],$$

- Under the assumption of “unconfoundedness",

$$\Pr(Y(z)|X) = \Pr(Y|X, Z = z), \quad z = 0, 1.$$

Thus ATE can be identified from observed data.

Horvitz-Thompson Weighting

- Propensity score is $e(X) = \Pr(Z = 1 \mid X)$, then

$$\mathbb{E}\left[\frac{ZY}{e(X)} - \frac{(1-Z)Y}{1-e(X)}\right] = \tau^{\text{ATE}}.$$

- Define inverse-probability or Horvitz-Thompson (HT) weight:

$$\begin{cases} w_1(X_i) = \frac{1}{e(X_i)}, \text{for } Z_i = 1 \\ w_0(X_i) = \frac{1}{1-e(X_i)}, \text{for } Z_i = 0. \end{cases}$$

- HT weighting balances (in expectation) the weighted distribution of covariates in the two groups.
- An unbiased nonparametric estimator of ATE is:

$$\hat{\tau}^{\text{ATE}} = \frac{\sum_i Y_i Z_i w_1(X_i)}{\sum_i Z_i w_1(X_i)} - \frac{\sum_i Y_i(1-Z_i)w_0(X_i)}{\sum_i (1-Z_i)w_0(X_i)}. \qquad (1)$$

Extension Beyond Horvitz-Thompson

- The target population of ATE is the combined treated and control groups.
- A well known problem of HT: explosion of the weights corresponding to extreme propensity scores (0 or 1).
- A common remedy is trimming, but what target population is the resulting subsample corresponding to?
- Other target populations: the treated group, or subgroup defined by a covariate, e.g. age or sex.

Extension Beyond Horvitz-Thompson

- Sometimes the available data do not represent to a natural population
- Often “marginal subjects" who may be exposed to either treatment are of interest (Rosenbaum, 2012).
- Examples:
  1. Racial disparity studies
  2. Under clinical equipoise, the target population are units with the most “overlap” observed characteristics.
- Motivating questions:
  1. Is there a unified framework for weighting for different target populations?
  2. Is there an intrinsic approach to define such “marginal population” while avoiding the pitfall of HT?

A General Framework: Defining Estimands

- Conditional average treatment effect

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

- Population density of the covariates X is $f(x)$
- Estimand is average $\tau(x)$ over a target population with density $\propto f(x)h(x)$, where $h(\cdot)$ is pre-specified function:

$$\tau_h \equiv \frac{\int \tau(dx)f(x)h(x)\mu(dx)}{\int f(x)h(x)\mu(dx)}. \tag{2}$$

- τ_h represents a general class of weighted ATE (WATE) estimands.

Balancing weights

- Density for group $Z = z$ is $f_z(x) = \Pr(X = x | Z = z)$
- Then $f_1(x) \propto f(x)e(x)$ and $f_0(x) \propto f(x)(1 - e(x))$.
- Given $h(x)$, the weighted distributions with densities $\propto f_z(x)w_z(x)$ have the target density $\propto f(x)h(x)$, $z = 0, 1$, if

$$\begin{cases} w_1(x) \propto \frac{f(x)h(x)}{f_1(x)} = \frac{f(x)h(x)}{f(x)e(x)} = \frac{h(x)}{e(x)}, \\ w_0(x) \propto \frac{f(x)h(x)}{f_0(x)} = \frac{f(x)h(x)}{f(x)(1-e(x))} = \frac{h(x)}{1-e(x)}. \end{cases} \tag{3}$$

- The class of weights (w_0, w_1) are called balancing weights.
- Choice of $h(x)$ determines the target population, estimand, weights.
- Statistical, scientific and policy considerations all come into play in specifying $h(x)$.

Examples of target population and balancing weights

target population	$h(x)$	estimand	weights (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	$(1-e(x), e(x))$

Nonparametric Estimation

- A natural nonparametric estimator for $\mathbb{E}(Y|Z = z, X \in x)$ in a neighborhood dx of x is the sample mean $\bar{y}_z(dx)$.
- Let $\hat{\tau}(dx) = \bar{y}_1(dx) - \bar{y}_0(dx)$.
- A nonparametric estimator for the WATE τ_h is,

$$\hat{\tau}_h = \frac{\int \hat{\tau}(dx) f(x) h(x) \mu(dx)}{\int f(x) h(x) \mu(dx)}. \tag{4}$$

- In practice, $\hat{\tau}_h$ has the form

$$\hat{\tau}_h^w = \frac{\sum_i Y_i Z_i w_1(X_i)}{\sum_i Z_i w_1(X_i)} - \frac{\sum_i Y_i (1 - Z_i) w_0(X_i)}{\sum_i (1 - Z_i) w_0(X_i)}.$$

Two large-sample results on $\hat{\tau}_h$

Result 1. *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.

Result 2. *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*

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

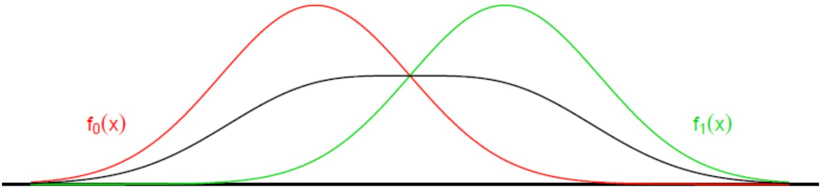
Overlap Weights

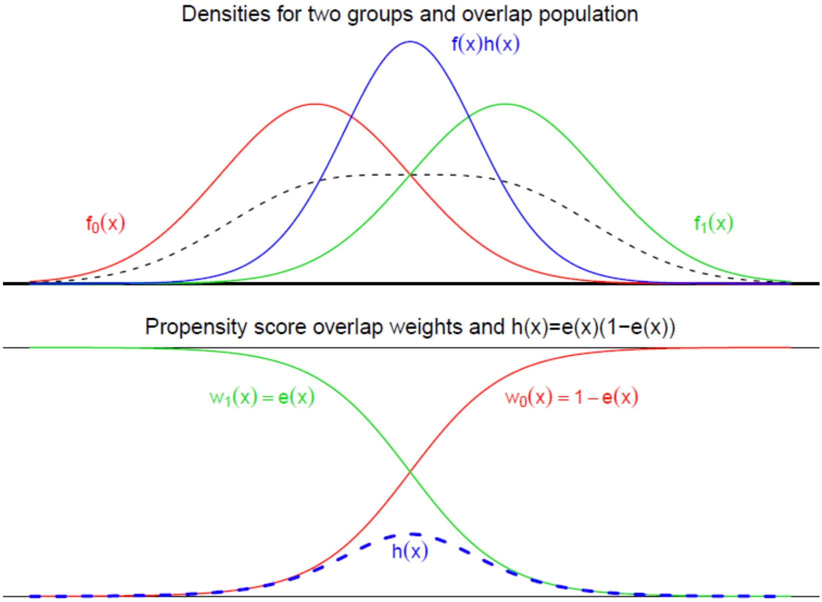
- Based on Result 2, we propose a new weight—the overlap weight—by letting $h(x) = e(x)(1 - e(x))$,

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

- Each unit is weighted by its probability of being assigned to the opposite group.
- Target population $f(x)e(x)(1 - e(x))$ is defined by overlap of covariates.
- “Marginal” units who may or may not receive the treatment.

Densities for two groups and combined population





Densities for two groups and overlap population
f(x)h(x)
f0(x)
f1(x)
Propensity score overlap weights and h(x)=e(x)(1−e(x))
w1(x) = e(x)
w0(x) = 1 − e(x)
h(x)

Exact Balance (Small-Sample) Property of the Overlap Weights

Result 3. *When the propensity scores are estimated from a logistic regression model with main effects,* $logit\{e(X_i)\} = \beta_0 + X_i\beta'$*, the overlap weights lead to exact balance in any included covariate between treatment and control groups.*

$$\frac{\sum_i X_{ij} Z_i (1 - \hat{e}_i)}{\sum_i Z_i (1 - \hat{e}_i)} = \frac{\sum_i X_{ij} (1 - Z_i) \hat{e}_i}{\sum_i (1 - Z_i) \hat{e}_i}, \quad \forall j.$$

Advantages of the Overlap Weights

Statistical advantages

- Minimum variances of the nonparametric 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, ATT, etc.)
- Avoids artificially truncating weights or eliminating cases

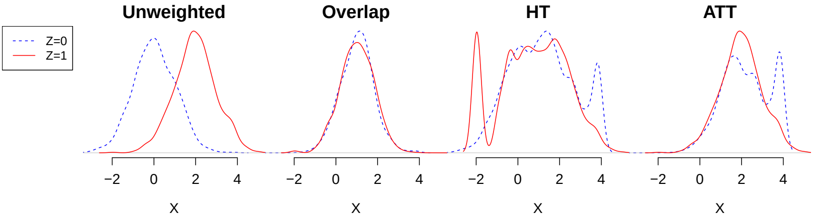
Scientific advantages

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

A Simulated Example

- Simulate $n_0 = n_1 = 1000$ units.
- A single covariate: $X_i \sim N(0, 1) + 2Z_i$.

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



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

A Simulated Example

- Two groups: $N_0 = N_1 = 1000$
- A single covariate: for control units $X_i \sim N(0,1)$; for treated units $X_i \sim N(2,1)$.
- Potential outcomes: $Y_i(z) \sim N(X_i, 1) + \tau z$, and $\tau = 1$.
- 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

Right Heart Catheterization: An Example of Causal Comparison

- Diagnostic to measure cardiac function.
- Observational data (Murphy and Cluff 1990).
  - $n = 5735$: 2184 RHC ($Z_i = 1$), 3551 control ($Z_i = 0$).
  - Outcome: survival at 30 days after admission.
  - Covariates: 53 binary/categorical variables.
- Extensively studied in literature.
  - Most focused on ATT: e.g. Connors et al. (1996), Hirano and Imbens (2001).
  - Crump et al. (2009): ATE for a truncated population with good overlap.
  - Rosenbaum (2012): ATE for a subpopulation with optimal overlap.

Figure : Boxplots for the absolute standardized differences for covariates under each weighting method.

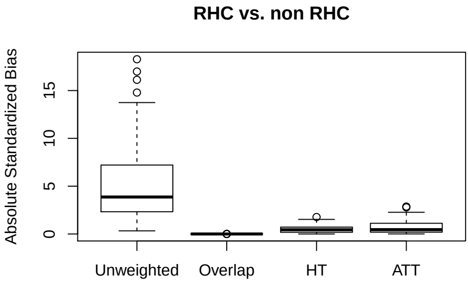


Table : Estimated treatment effect (in %) with different weights

	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

Limitations and caveats

- Relies on large-sample properties
- Fit of propensity score model is important to local balance
  - Flexible link function?
  - Splines (e.g. decile dummies) to improve fit
- Lose benefits of observation-level matching for variance reduction (paired tests)
  - Rely more on models in analysis stage
- Lose benefits of approximate matching of distributions in directions orthogonal to $\hat{e}(x)$
- May combine with matching to correct residual bias

Summary

- Propose a unified framework for weighting to balance covariates for any target population.
- The class of balancing weights include many of the existing weights and beyond.
- Propose a new type of weights—the overlap weights—with desirable statistical and scientific properties.
- The estimand targets the hypothetical population with the most overlap in covariates.
- The methods developed here are applicable to both causal and unconfounded descriptive comparisons.