Introducing the Overlap Weights in Causal Inference

Fan Li

Department of Statistical Science
Duke University

November 12, 2018
Introduction

- Population-based observational data increasingly used for causal inference
- Essential for causal comparisons: Balancing covariate distributions across groups to remove confounding
- One common approach is weighting
  - Main idea: weigh the treatment and control groups to create a pseudo-population—the target population—where the two groups are balanced, in expectation
- The dominant weighting approach: inverse probability weighting (IPW), originated from the Horvitz-Thompson estimator in survey
Example: Framingham Heart Study
(Thomas, Lorenzi, et al. 2018)

- **Goal:** evaluate the effect of statins on health outcomes

- **Patients:** cross-sectional population from the offspring cohort with a visit 6 (1995-1998)

- **Treatment:** statin use at visit 6 vs. no statin use

- **Outcomes:** CV death, myocardial infarction (MI), stroke

- **Confounders:** sex, age, body mass index, diabetes, history of MI, history of PAD, history of stroke...

- Significant imbalance between treatment and control groups in covariates
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.
- For each unit $i$, two potential outcomes $(Y_i(0), Y_i(1))$, but only $Y_i(Z_i)$ is observed.

- **Estimand:** Average Treatment Effect (ATE)

  $$\tau_{ATE} = \mathbb{E}[Y(1) - Y(0)]$$

- Assuming strong ignorability:
  (i) $\Pr(Z = 1|Y(1), Y(0), X) = \Pr(Z = 1|X)$
  (ii) $0 < \Pr(Z = 1|X) < 1$ for all units

- Then ATE is identified from the observed data:
  $$\mathbb{E}(Y(z)|X) = \mathbb{E}(Y|X, Z = z) \text{ for } z = 0, 1$$
Inverse Probability Weights (IPW)

- The propensity score: \( e(X) = \Pr(Z = 1 | X) \)
- Base of IPW
  \[
  \mathbb{E} \left[ \frac{ZY}{e(X)} - \frac{(1 - Z)Y}{1 - e(X)} \right] = \tau^{ATE}.
  \]
- Inverse probability weights:
  \[
  \begin{cases}
  w_1(X_i) = \frac{1}{e(X_i)}, & \text{for } Z_i = 1 \\
  w_0(X_i) = \frac{1}{1 - e(X_i)}, & \text{for } Z_i = 0.
  \end{cases}
  \]
- IPW balances, in expectation, the weighted distribution of covariates in the two groups
- An unbiased nonparametric estimator of ATE is the difference in the mean of the weighted outcomes between groups
IPW: Conceptual Challenges
(Thomas, Li, Pencina, 2018)

- Target population of IPW: the “whole” population – the combined treatment and control groups
- Key but often forgotten question: what population does the study sample is representative of?
- In observational studies, the study sample is often a convenience sample – does not represent any natural population of scientific interest
- Applying IPW to such a sample does NOT lead to an ATE on any meaningful population
- IPW (equivalently ATE) may correspond to the effect of an infeasible intervention
IPW: Operational Challenges

- Prone to adverse finite-sample consequences due to extreme probabilities (0 or 1) – Basu’s elephant: severe bias and variance
- Normalization of weights helps, but not a lot
- Common remedy is trimming (remove extreme propensities): ad hoc, sensitive to cut off points, ambiguous target population
- Core problem: lack of overlap in the tail of the propensity distribution – causal comparisons of these units are highly uncertain
Two main points of this talk

1. Provide a unified framework—the balancing weights—to allow different user-specified target populations.

2. Propose a new weighting scheme—the overlap weighting—to capture the “overlapped population” and possess statistical optimality.
A General Framework: Defining Estimands
(Li, Morgan, Zaslavsky, 2018)

- Define conditional average treatment effect (CATE)

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

- Assume density of the covariates of the sample, \( f(x) \), exists wrt a base measure \( \mu \)

- Target population density: \( g(x) = f(x)h(x) \), with pre-specified \( h(\cdot) \)

- Estimand: average \( \tau(x) \) over the target population \( g(x) \)

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

- \( \tau_h \) represents a general class of weighted ATE (WATE) estimands.
Balancing weights

- Let \( f_z(x) = \Pr(X = x|Z = z) \), we have

\[
\begin{align*}
    f_1(x) &\propto f(x)e(x), \\
    f_0(x) &\propto f(x)(1 - e(x))
\end{align*}
\]

- For a given \( h(x) \), to estimate \( \tau_h \), we can weight \( f_z(x) \) to the target population using weights

\[
\begin{align*}
    w_1(x) &\propto \frac{f(x)h(x)}{f_1(x)} = \frac{f(x)h(x)}{f(x)e(x)} = \frac{h(x)}{e(x)}, \\
    w_0(x) &\propto \frac{f(x)h(x)}{f_0(x)} = \frac{f(x)h(x)}{f(x)(1 - e(x))} = \frac{h(x)}{1 - e(x)}.
\end{align*}
\]

- We call the class of weights \((w_0, w_1)\) balancing weights: they balance the distributions of the weighted covariates between comparison groups.
Examples: target population \((h)\) and balancing weights

- Choice of \(h(x)\) determines the target population, estimand, weights.
- Statistical, scientific and policy considerations all come into play in specifying \(h(x)\).

<table>
<thead>
<tr>
<th>target population</th>
<th>(h(x))</th>
<th>estimand</th>
<th>weight ((w_1, w_0))</th>
</tr>
</thead>
<tbody>
<tr>
<td>combined</td>
<td>1</td>
<td>ATE</td>
<td>(\left(\frac{1}{e(x)}, \frac{1}{1-e(x)}\right)) [IPW]</td>
</tr>
<tr>
<td>treated</td>
<td>(e(x))</td>
<td>ATT</td>
<td>(\left{\frac{e(x)}{1-e(x)}, 1\right})</td>
</tr>
<tr>
<td>control</td>
<td>(1-e(x))</td>
<td>ATC</td>
<td>((1-e(x), e(x)))</td>
</tr>
<tr>
<td>overlap</td>
<td>(e(x)(1-e(x)))</td>
<td>ATO</td>
<td></td>
</tr>
<tr>
<td>trunc combined</td>
<td>(1(\alpha &lt; e(x) &lt; 1-\alpha))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>matching</td>
<td>(\min{e(x), 1-e(x)})</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sample estimator of WATE

\[ \hat{\tau}_h = \frac{\sum_i w_1(x_i) Z_i Y_i}{\sum_i w_1(x_i) Z_i} - \frac{\sum_i w_0(x_i)(1 - Z_i) Y_i}{\sum_i w_0(x_i)(1 - Z_i)} \]  

**Theorem 1.** \( \hat{\tau}_h \) is a consistent estimator of \( \tau_h \).
Theorem 2. As \( n \to \infty \), the expectation (over possible samples of covariate values) of the conditional variance of the estimator \( \hat{r}_h \) given the sample \( X = \{x_1, \ldots, x_n\} \) converges:

\[
n \cdot \mathbb{E}_X \mathbb{V} \left[ \hat{r}_h \mid X \right] \to \int f(x)h(x)^2 \left[ \frac{v_1(x)}{e(x)} + \frac{v_0(x)}{1 - e(x)} \right] \mu(dx)/C_h^2,
\]

where \( v_z(x) = \mathbb{V}[Y(z) \mid X] \) and \( C_h = \int h(x)f(x)\mu(x) \) is a normalizing constant.

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

\[
n \cdot \min_h \{\mathbb{V}[\hat{r}_h]\} \to v/\int f(x)e(x)(1 - e(x))\mu(dx).
\]
Overlap Weights

- Based on Corollary 1, we propose a new type of weights, the overlap weights, by letting $h(x) = e(x)(1 - e(x))$,

$$
\begin{align*}
    w_1(x) &\propto 1 - e(x), \quad \text{for } Z = 1, \\
    w_0(x) &\propto e(x), \quad \text{for } Z = 0.
\end{align*}
$$

- 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.

- Target population: the units whose characteristics could appear with substantial probability in either treatment group (most overlap).
Densities for two groups and overlap population

\[ f(x)h(x) \]

\[ f_0(x) \]

\[ f_1(x) \]

Propensity score overlap weights and \( h(x) = e(x)(1 - e(x)) \)

\[ w_1(x) = e(x) \]

\[ w_0(x) = 1 - e(x) \]

\[ h(x) \]
Overlap Weights: Exact Balance

**Theorem 3.** When the propensity scores are estimated by maximum likelihood under a logistic regression model,

\[
\text{logit}\{e(x_i)\} = \beta_0 + x_i'\beta,
\]

the overlap weights lead to exact balance in the means of any included covariate between treatment and control groups:

\[
\frac{\sum_i x_{ij}Z_i(1 - \hat{e}_i)}{\sum_i Z_i(1 - \hat{e}_i)} = \frac{\sum_i x_{ij}(1 - Z_i)\hat{e}_i}{\sum_i(1 - Z_i)\hat{e}_i}, \quad \text{for } j = 1, \ldots, p, \tag{4}
\]

where \(\hat{e}_i = \{1 + \exp[-(\hat{\beta}_0 + x_i'\hat{\beta})]\}^{-1}\) and \(\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_j)\) is the MLE for the regression coefficients.

**Remark:** the exact balance property applies to any included covariate and derived covariate, including high order terms and interaction terms of the covariates.
Corollary 2. If the postulated propensity score model includes any interaction term of a binary covariate, then the overlap weights lead to exact balance in the means in the subgroups defined by that binary covariate.

Remarks:
- If the true PS model has interaction terms, then overlap weights using PS estimated from any model that nests the true model gives exact balance in the subgroups defined by the interaction terms.
Overlap Weights: Statistical Advantages

- **Minimum variance** of the nonparametric estimator among all balancing weights
- **Exact balance** for means of included covariates in logistic propensity score model
- Weights are **bounded** (unlike IPW)
- Avoids *ad hoc* eliminating cases: continuously down-weigh units in the tail
Overlap Weights: Statistical Advantages

- Overlap weights are adaptive, ATO approximate
  - ATE: if treatment and control groups are nearly balanced in size and distribution (for $e(x) \approx 1/2$,
    $$(1 - e(x), e(x)) \approx \left(\frac{.25}{e(x)}, \frac{.25}{1 - e(x)}\right)$$
  - ATT: if propensity to treatment is always small (for $e(x) \approx 0$,
    $$(1 - e(x), e(x)) \approx \left(1, \frac{e(x)}{1 - e(x)}\right)$$
  - ATC: if propensity to control is small
Overlap Weights: Scientific Relevance

- Overlap weights focus on the (sub)population closest to the population in a randomized clinical trial.
- Overlap weights put emphasis on internal validity.
- The overlap population is of intrinsic substantive interest, for example:
  - In medicine, patients in clinical equipoise.
  - In policy, units whose treatment assignment would be most responsive to a policy shift as new information is obtained.
- Better **transportability** than ATE: always focus on the most overlapped population regardless of what the study sample is representative of.
Variance Estimator
(Li, Thomas, Li, 2018)

- A sandwich variance estimator for $\tau_h$ using the OW when the PS is estimated from a logistic regression

\[
\nabla(\tau_{ow}) = \frac{\sum_{i=1}^{n} \psi_i^2}{\left(\sum_{i=1}^{n} \hat{e}_i(1 - \hat{e}_i)\right)^2},
\]

where

\[
\psi_i = Z_i(Y_i - \hat{\tau}_1)(1 - \hat{e}_i) - (1 - Z_i)(Y_i - \hat{\tau}_0)\hat{e}_i - (Z_i - \hat{e}_i)\hat{H}'_{\beta} \hat{E}_{\beta\beta}^{-1} x_i,
\]

and

\[
\hat{H}_{\beta} = n^{-1}\sum_{i=1}^{n} \{Z_i(Y_i - \hat{\tau}_1) - (1 - Z_i)(Y_i - \hat{\tau}_0)\} \hat{e}_i(1 - \hat{e}_i)x_i,
\]

and $\hat{E}_{\beta\beta}^{-1}$ is the information matrix.
Based on M-estimation, account for the uncertainty in estimating the propensity scores

The last subtraction in $\psi_i$ is an orthogonal projection term that accounts for the uncertainty in estimating the propensity scores, i.e., $\psi_i = \tilde{\psi}_i - \Pi(\tilde{\psi}_i|\Lambda)$.

Connection to Matching

- Matching: link “similar” cases in two samples, discard unmatched cases (bottom-up approach)

- Weighting: apply weights to entire samples, designed to create global balance (top-down approach)

- **Intrinsic connection**: Overlap weighting approaches many-to-many matching as the propensity score model becomes increasingly complex.

- The limit is a saturated model with a fixed effect for each design point.

- The nonparametric weighted estimate $\hat{\tau}_h$ using the overlap weights is the same as that from a LS model for the outcome with a fixed effect for each design point.
A hybrid approach to combine the benefits of matching and overlap weighting

1. Obtain a matched sample using any preferred approach (e.g., Mahalanobis distance)

2. Estimate the propensity scores a logistic regression with all main effects \textit{within} the matched sample

3. Apply the overlap weights to the matched sample to estimate the treatment effect

Retain the nearness of matched cases in multivariate space, and adjust for residual imbalance in matching via overlap weighting
A Simulated Example

- Simulate $n_0 = n_1 = 1000$ units.
- A single covariate: $X_i \sim N(0, 1) + 2Z_i$.
- Outcome model: $Y_i(z) \sim N(X_i, 1) + \tau z$, and $\tau = 1$.
- Use the nonparametric estimator $\hat{\tau}_h$ with different weights

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

<table>
<thead>
<tr>
<th></th>
<th>Unweighted</th>
<th>Overlap</th>
<th>IPW</th>
<th>ATT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\tau}$</td>
<td>2.945</td>
<td>1.000</td>
<td>0.581</td>
<td>0.640</td>
</tr>
<tr>
<td>$SE(\hat{\tau})$</td>
<td>0.054</td>
<td>0.038</td>
<td>0.386</td>
<td>0.402</td>
</tr>
</tbody>
</table>
Framingham revisited
Weighted Distribution

All Patients

- Among Statins
- Among Control
- Overlap
- IPW

Propensity Score Across patients

Density
Weighted Distribution

All Patients

- Among Statins
- Among Control
- Overlap
- IPW

Density

Framingham Risk Score

Across patients
Results: CV death

**Figure:** IPW 1: No trimming; IPW 2: trimming ps between (.10, 0.90); IPW 3: asymmetric trimming 5th% ps of trt, 95th% of ps for control
Results: composite of non-death endpoints

Figure: IPW 1: No trimming; IPW 2: trimming ps between (.10, 0.90); IPW 3: asymmetric trimming 5th% ps of trt, 95th% of ps for control
Conclusion and extension

▸ We proposed a unified framework of balancing weights to balance covariates for any target population

▸ We proposed the overlap weights: efficiency and exact balance

▸ Takeaways: (1) Important to consider scientific appropriate target population in practice; (2) should not automatically focus on IPW (ATE)

▸ Extensions
  ▸ multiple treatments/groups (Li and Li, 2018)
  ▸ time-varying treatments
  ▸ variance reduction in randomized experiments
  ▸ clustering structure
Acknowledgements

- Alan Zaslavsky (Harvard)
- Laine Thomas (Duke)
- Frank Li (Duke)
- Elizabeth Lorenzi (Duke)
- Michael Pencina (Duke)
- Kari Lock Morgan (Penn State)


