Addressing Disparities in the Propensity Distributions for Treatment Comparisons from Observation Studies

Tingting Zhou, Michael Elliott, Roderick Little
University of Michigan

Contact Information:
Department of Biostastistics
University of Michigan
1415 Washington Heights, Ann Arbor, MI 48109
Email: tkzhou@umich.edu

Introduction

- Causal inference is valid when subjects are randomized
- In the absence of randomization, we need to do one of the following:
  - Adjust for confounders
  - Adjust for propensity score - the probability of treatment assignment given covariates
    1. Matching
    2. Regression adjustment
    3. Weighting
- Lack of overlap in the propensity score distributions for the compared groups is a common problem in nonrandomized studies
- But sufficient overlap is required for the methods to work reliably
  - Avoid extrapolation outside of the overlap region
  - Reduce model dependence
- GOAL: discuss alternative estimands to address limited overlap

Figure 1: (top) adequate overlap, and (bottom) severe lack of overlap

Notation

- N subjects $j = 1, \cdots, N$
- Baseline covariates: X
- Binary treatment: $Z \in \{0, 1\}$
- Potential outcome under treatment Z: Y^Z

Assumptions

- The stable unit treatment value assumption (SUTVA) (Angrist, Imbens and Rubin 1996)
  1. The observed $Y(z)$ equal to the potential outcome Y^z
  2. No interference between subjects
- Each subject has a positive probability of being assigned to each treatment
- Ignorability: $(Y^1, Y^0) \perp\!\!\!\perp Z|X$

Propensity Score

- The probability of receiving treatment z: $p_z(x) = \Pr(Z = z|x)$
- Balancing property: $Z \perp\!\!\!\perp X|p_z(x)$
- Ignorability: $(Y^1, Y^0) \perp\!\!\!\perp Z|p_z(x)$
- Dimension reduction

Weighting and Matching

- Four weighting schemes: inverse-probability-treatment-weighted estimator (IPTW); augmented IPTW estimator (AIPTW) (Scharfstein, Rotnitzky, and Robins, 1999); Matching weight and doubly robust matching weight estimators (Li and Greene, 2013); Overlap weight estimator (Li et al, 2017)
- Pair Matching: greedy one-to-one matching with a caliper and without replacement, can be used to preprocess the data to reduce model dependence.

Methods	Estimand	Weights(w_1, w_0)
IPTW (AIPTW)	ATE	$(\frac{1}{p_1(x)}, \frac{1}{p_0(x)})$
Truncated+IPTW (AIPTW)	Truncated	$(\frac{I(\alpha<p_1(x)<1-\alpha)}{p_1(x)}, \frac{I(\alpha<p_0(x)<1-\alpha)}{p_0(x)})$
Matching+IPTW (AIPTW)	ATM	$(\frac{1}{p*_1(x)}, \frac{1}{p*_0(x)})$
Matching Weight	ATM	$(\frac{min\{p_1(x), p_0(x)\}}{p_1(x)}, \frac{min\{p_1(x), p_0(x)\}}{p_0(x)})$
ATO	ATO	$(1 - p_0(x), p_1(x))$

Table 1: Estimand associated with each weighting method. ATE-the entire population; ATM-subpopulation that can be matched; Truncated-truncated subpopulation; ATO:combination of both treated and control populations, not easily recognizable. $p*$ reestimated propensity score after matching.

Penalized Spline of Propensity Methods for Treatment Comparison (PENCOMP)

- Robust multiple imputation based approach
- Key idea: impute the missing potential outcomes

Subjects	X	Z	Y^0	Y^1
1		0		?
$\cdots$		0		?
n_0		0		?
$n_0 + 1$		1	?	
$\cdots$		1	?	
$n = n_0 + n_1$		1	?	

Figure 2: Observed and missing outcomes in a single treatment.

1. Imputation model:

$$E(Y^z|X, Z = z, \theta_z, \beta_z) = s(\hat{P*}_z|\theta_z) + g_z(X; \beta_z),$$

- where $\hat{P*}_z = \text{logit}\,(\hat{P}_z(X))$.
- $s(\hat{P*}_z) = \theta_0 + \theta_1 \hat{P*}_z + \sum_{k=1}^{K} \theta_{1k}(\hat{P*}_z - k_k)_+$, for knots $k_1, \cdots, k_K$
- Truncated linear bases $(\hat{P*}_{z_1} - k_1)_+, \cdots, (\hat{P*}_{z_1} - k_K)_+$ treated as random effects
- $g_{z_1}()$ increases efficiency by including covariates predictive of outcome

2. Combine the imputed and observed outcomes of Y using Rubin's Combining Rules over D complete datasets
   - Causal effect: $\Delta = \bar{\Delta}_D = \frac{1}{D}\sum_{d=1}^{D} \hat{\Delta}_d$
   - Variance: $T_D = \bar{W}_D + (1 + 1/D)B_D$, where
   (a) $\bar{W}_D = \sum_{d=1}^{D} W^d/D$
   (b) $B_D = \sum_{d=1}^{D} \left(\hat{\Delta}^{(d)} - \bar{\Delta}_D\right)^2/(D-1)$
3. Can restrict inference to corresponding target population to obtain ATE, truncated, and ATM (by preprocessing the data via pair matching)

Property of PENCOMP

- If (1) or (2) is true, the marginal mean from our imputation model is consistent (Zhang and Little 2009, Zhou, Elliott and Little 2018):
  1. $g()$ correctly specified
  2. $P_z(X)$ correctly specified, and the relationship between the outcome and the propensity score are correctly specified.
- Extend to multiple time points (Zhou, Elliott and Little 2018)

Simulations

- We consider a setting where the predictors of the outcome and the propensity score are different (misaligned) and the effects of the covariate on the treated and the control are the same (homogeneity)

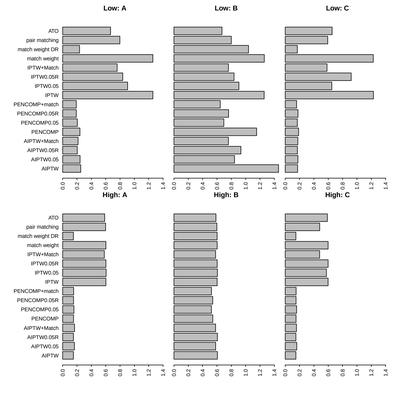

Figure 3: Empirical RMSE, sample size of 200.

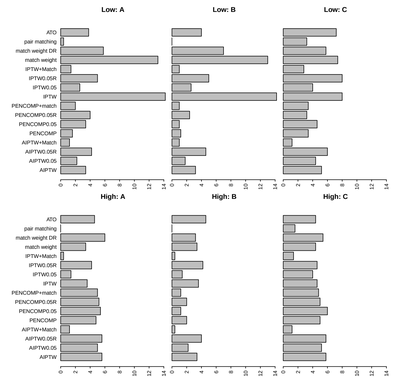

Figure 4: 100 * 95% non coverage rate, sample size of 200. (A)-Both propensity and prediction models are correct; (B) Prediction models are incorrect; (C) Propensity models are incorrect. Top Panel-Low overlap in the propensity distributions; Bottom Panel-high overlap in the propensity distributions.

Application

- We applied our method to the Multicenter AIDS Cohort study (MACS) to analyze the the short-term (6-month) effects of antiretroviral treatments on CD4 counts for HIV+ subjects
- Blood test results and treatment histories were collected every 6-month
- The treatment assigment was modeled as logistic regression with the blood count and treatment histories from the previous two visits

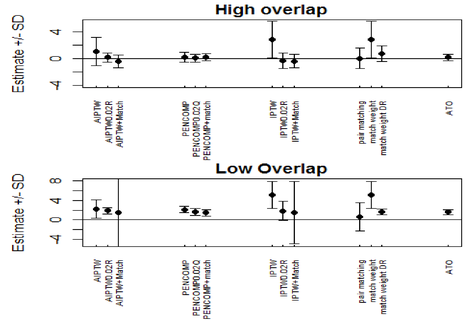

Figure 5: Estimates of treatment effects and 95% confidence intervals based on 1000 bootstraps (or complete datasets for PENCOMP) for two visits: low and high overlap in the propensity score distributions

Discussion

- PENCOMP has the flexibility of estimating different estimands
- PENCOMP tends to be more stable compared to weighting approaches-less sensitive to extreme weights