Weighting Methods

Kosuke Imai

Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

Fall 2018

Motivation

- Matching methods for improving covariate balance
- Potential limitations of matching methods:
  1. inefficient $\rightsquigarrow$ it may throw away data
  2. ineffective $\rightsquigarrow$ it may not be able to balance covariates
- Recall that matching is a special case of weighting:

$$\begin{aligned}\hat{\tau}_{\text{match}} &= \frac{1}{n_1}\sum_{i=1}^{n} T_i \left(Y_i - \frac{1}{|\mathcal{M}_i|}\sum_{i' \in \mathcal{M}_i} Y_{i'}\right) \\ &= \frac{1}{n_1}\sum_{i:T_i=1} Y_i - \frac{1}{n_0}\sum_{i:T_i=0} \underbrace{\left(\frac{n_0}{n_1}\sum_{i':T_{i'}=1} \frac{\mathbf{1}\{i \in \mathcal{M}_{i'}\}}{|\mathcal{M}_{i'}|}\right)}_{w_i} Y_i\end{aligned}$$

- Idea: weight each observation in the control group such that it looks like the treatment group (i.e., good covariate balance)

Inverse Propensity Score Weighting

- Weighting for surveys: down-weight over-sampled respondents
- Sampling weights inversely proportional to samplig probability
- Horvitz-Thompson estimator (1952. *J. Am. Stat. Assoc.*):

$$\widehat{\mathbb{E}(Y_i)} = \frac{1}{N}\sum_{i=1}^{N} \frac{S_i Y_i}{\Pr(S_i = 1)}$$

- Inverse probability-of-treatment weighting estimators (IPW):

$$\begin{aligned}
\widehat{\text{ATE}} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i}{\hat{\pi}(\mathbf{X}_i)} - \frac{(1-T_i)Y_i}{1-\hat{\pi}(\mathbf{X}_i)}\right\} \\
\widehat{\text{ATT}} &= \frac{1}{n_1}\sum_{i=1}^{n}\left\{T_i Y_i - \frac{\hat{\pi}(\mathbf{X}_i)(1-T_i)Y_i}{1-\hat{\pi}(\mathbf{X}_i)}\right\} \\
\widehat{\text{ATC}} &= \frac{1}{n_0}\sum_{i=1}^{n}\left\{\frac{(1-\hat{\pi}(\mathbf{X}_i))T_i Y_i}{\pi(\mathbf{X}_i)} - (1-T_i)Y_i\right\}
\end{aligned}$$

- Identical propensity score $\rightsquigarrow$ difference-in-means

Normalized Weights

- Survey sampling when the population size is unknown
- Hajek Estimator:

$$\widehat{\mathbb{E}(Y_i)} = \frac{\sum_i S_i Y_i / \Pr(S_i = 1)}{\sum_i S_i / \Pr(S_i = 1)}$$

- Weights are normalized but no longer unbiased
- Normalization of weights may be important when propensity score is estimated

$$\widehat{\text{ATE}} = \frac{\sum_{i=1}^n T_i Y_i / \hat{\pi}(\mathbf{X}_i)}{\sum_{i=1}^n T_i / \hat{\pi}(\mathbf{X}_i)} - \frac{\sum_{i=1}^n (1 - T_i) Y_i / \{1 - \hat{\pi}(\mathbf{X}_i)\}}{\sum_{i=1}^n (1 - T_i) / \{1 - \hat{\pi}(\mathbf{X}_i)\}}$$

- Weighted least squares:

$$(\hat{\alpha}_{\text{wls}}, \hat{\beta}_{\text{wls}}) = \underset{\alpha,\beta}{\text{argmin}} \sum_{i=1}^n \frac{T_i(1 - \hat{\pi}(\mathbf{X}_i)) + (1 - T_i)\hat{\pi}(\mathbf{X}_i)}{\hat{\pi}(\mathbf{X}_i)\{1 - \hat{\pi}(\mathbf{X}_i)\}} (Y_i - \alpha - \beta T_i)^2$$

Variance

- IPW estimator as the method of moments estimator:
  1. moment condition from the propensity score model (e.g., score)
  2. moment conditions from the weighting estimator

$$\textit{Horvitz/Thompson}: \quad \frac{1}{n}\sum_{i=1}^{n}\frac{T_i Y_i}{\hat{\pi}(\mathbf{X}_i)} - \mu_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{(1-T_i)Y_i}{1-\hat{\pi}(\mathbf{X}_i)} - \mu_0 = 0$$

$$\textit{Hajek}: \quad \frac{1}{n}\sum_{i=1}^{n}\frac{T_i(Y_i-\mu_1)}{\hat{\pi}(\mathbf{X}_i)} = \frac{1}{n}\sum_{i=1}^{n}\frac{(1-T_i)(Y_i-\mu_0)}{1-\hat{\pi}(\mathbf{X}_i)} = 0$$

$\rightsquigarrow$ large sample variances are readily available

- If the propensity score model is correctly specified, these variances are smaller than those with the true propensity score

Doubly Robust Estimator (Robins et al. 1994. *J. Am. Stat. Assoc.*)

- Augmented IPW (AIPW) estimator:

$$\begin{aligned}
\hat{\tau}_{\mathrm{DR}} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{T_i Y_i}{\hat{\pi}(\mathbf{X}_i)} - \frac{T_i - \hat{\pi}(\mathbf{X}_i)}{\hat{\pi}(\mathbf{X}_i)}\hat{\mu}(1, \mathbf{X}_i)\right. \\
&\qquad \left. - \frac{(1 - T_i)Y_i}{1 - \hat{\pi}(\mathbf{X}_i)} + \frac{T_i - \hat{\pi}(\mathbf{X}_i)}{1 - \hat{\pi}(\mathbf{X}_i)}\hat{\mu}(0, \mathbf{X}_i)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\mu}(1, \mathbf{X}_i) + \frac{T_i(Y_i - \hat{\mu}(1, \mathbf{X}_i))}{\hat{\pi}(\mathbf{X}_i)}\right. \\
&\qquad \left. - \hat{\mu}(0, \mathbf{X}_i) - \frac{(1 - T_i)(Y_i - \hat{\mu}(0, \mathbf{X}_i))}{1 - \hat{\pi}(\mathbf{X}_i)}\right\}
\end{aligned}$$

- Consistent if either the propensity score model or the outcome model is correct $\rightsquigarrow$ you get two chances to be correct
- Efficient: smallest asymptotic variance among estimators that are consistent when the propensity score model is correct
- Estimator may not behave well when both models are incorrect

A Simulation Study (Kang and Schafer. 2007. *Statistical Science*)

- The deteriorating performance of propensity score weighting methods when the model is misspecified
- Led to improvements of doubly robust estimators $\rightsquigarrow$ Cao et al. (2009), Tan (2010), Rotnitzky et al. (2012), Han and Wang (2013) *Biometrika*. etc.
- Setup:
  - 4 covariates X_i^*: all are *i.i.d.* standard normal
  - Outcome model: linear model
  - Propensity score model: logistic model with linear predictors
  - Misspecification induced by measurement error:
    - $X_{i1} = \exp(X_{i1}^*/2)$
    - $X_{i2} = X_{i2}^*/(1 + \exp(X_{1i}^*) + 10)$
    - $X_{i3} = (X_{i1}^* X_{i3}^*/25 + 0.6)^3$
    - $X_{i4} = (X_{i1}^* + X_{i4}^* + 20)^2$
- Weighting estimators to be evaluated:
  1. Horvitz-Thompson
  2. Inverse-probability weighting with normalized weights
  3. Weighted least squares regression with covariates
  4. Doubly-robust least squares regression with covariates

Weighting Estimators Do Fine If the Model is Correct

Sample size	Estimator	Bias logit	Bias True	RMSE logit	RMSE True
(1) Both models correct					
$n = 200$	HT	0.33	1.19	12.61	23.93
	IPW	−0.13	−0.13	3.98	5.03
	WLS	−0.04	−0.04	2.58	2.58
	DR	−0.04	−0.04	2.58	2.58
$n = 1000$	HT	0.01	−0.18	4.92	10.47
	IPW	0.01	−0.05	1.75	2.22
	WLS	0.01	0.01	1.14	1.14
	DR	0.01	0.01	1.14	1.14
(2) Propensity score model correct					
$n = 200$	HT	−0.05	−0.14	14.39	24.28
	IPW	−0.13	−0.18	4.08	4.97
	WLS	0.04	0.04	2.51	2.51
	DR	0.04	0.04	2.51	2.51
$n = 1000$	HT	−0.02	0.29	4.85	10.62
	IPW	0.02	−0.03	1.75	2.27
	WLS	0.04	0.04	1.14	1.14
	DR	0.04	0.04	1.14	1.14

Weighting Estimators are Sensitive to Misspecification

		Bias		RMSE	
Sample size	Estimator	logit	True	logit	True
(3) Outcome model correct					
$n = 200$	HT	24.25	−0.18	194.58	23.24
	IPW	1.70	−0.26	9.75	4.93
	WLS	−2.29	0.41	4.03	3.31
	DR	−0.08	−0.10	2.67	2.58
$n = 1000$	HT	41.14	−0.23	238.14	10.42
	IPW	4.93	−0.02	11.44	2.21
	WLS	−2.94	0.20	3.29	1.47
	DR	0.02	0.01	1.89	1.13
(4) Both models incorrect					
$n = 200$	HT	30.32	−0.38	266.30	23.86
	IPW	1.93	−0.09	10.50	5.08
	WLS	−2.13	0.55	3.87	3.29
	DR	−7.46	0.37	50.30	3.74
$n = 1000$	HT	101.47	0.01	2371.18	10.53
	IPW	5.16	0.02	12.71	2.25
	WLS	−2.95	0.37	3.30	1.47
	DR	−48.66	0.08	1370.91	1.81

Covariate Balancing Propensity Score (CBPS)

(Imai and Ratkovic. 2015. *J. Royal Stat. Soc. B.*)

- How can we improve the estimation of propensity score?
- Estimate the propensity score such that covariates are balanced
- Covariate balance conditions:

$$\mathbb{E}\left\{\frac{T_i\, f(\mathbf{X}_i)}{\pi_{\boldsymbol{\beta}}(\mathbf{X}_i)} - \frac{(1-T_i)\, f(\mathbf{X}_i)}{1-\pi_{\boldsymbol{\beta}}(\mathbf{X}_i)}\right\} \quad = \quad 0$$

- Usual score condition: $f(\mathbf{X}_i) = \pi'_{\boldsymbol{\beta}}(\mathbf{X}_i)$
- Optimal choice (Fan et al. 2016. *Working Paper*):

$$f(\mathbf{X}_i) \;=\; \pi(\mathbf{X}_i)\mu(0,\mathbf{X}_i) + (1-\pi(\mathbf{X}_i))\mu(1,\mathbf{X}_i)$$

  1. double robustness
  2. smallest asymptotic variance when the propensity score is correct
- Estimation via the (generalized) method of moments

More Robust Weighting Methods

		Bias				RMSE			
Sample size	Estimator	GLM	CBPS1	CBPS2	True	GLM	CBPS1	CBPS2	True
(3) Outcome model correct									
$n = 200$	HT	24.25	1.09	−5.42	−0.18	194.58	5.04	10.71	23.24
	IPW	1.70	−1.37	−2.84	−0.26	9.75	3.42	4.74	4.93
	WLS	−2.29	−2.37	−2.19	0.41	4.03	4.06	3.96	3.31
	DR	−0.08	−0.10	−0.10	−0.10	2.67	2.58	2.58	2.58
$n = 1000$	HT	41.14	−2.02	2.08	−0.23	238.14	2.97	6.65	10.42
	IPW	4.93	−1.39	−0.82	−0.02	11.44	2.01	2.26	2.21
	WLS	−2.94	−2.99	−2.95	0.20	3.29	3.37	3.33	1.47
	DR	0.02	0.01	0.01	0.01	1.89	1.13	1.13	1.13
(4) Both models incorrect									
$n = 200$	HT	30.32	1.27	−5.31	−0.38	266.30	5.20	10.62	23.86
	IPW	1.93	−1.26	−2.77	−0.09	10.50	3.37	4.67	5.08
	WLS	−2.13	−2.20	−2.04	0.55	3.87	3.91	3.81	3.29
	DR	−7.46	−2.59	−2.13	0.37	50.30	4.27	3.99	3.74
$n = 1000$	HT	101.47	−2.05	1.90	0.01	2371.18	3.02	6.75	10.53
	IPW	5.16	−1.44	−0.92	0.02	12.71	2.06	2.39	2.25
	WLS	−2.95	−3.01	−2.98	0.19	3.30	3.40	3.36	1.47
	DR	−48.66	−3.59	−3.79	0.08	1370.91	4.02	4.25	1.81

Calibration Methods

- Forget about the propensity score $\rightsquigarrow$ just balance covariates
  - avoid modeling assumptions and balance certain moments
  - in theory, propensity score balances the entire distributions
  - validation and interpretation are more difficult

- Entropy balancing (Hainmueller. 2012. *Political Anal.*)

$$\{w_1^*, w_2^*, \ldots, w_{n_0}\} \;=\; \underset{w}{\operatorname{argmin}} \sum_{i:T_i=0} w_i \log(w_i/q_i)$$

  where $w_i \geq 0$, $\sum_{i:T_i=0} w_i = 1$, $\sum_{i:T_i=0} w_i f(\mathbf{X}_i) = \frac{1}{n_1}\sum_{i:T_i=1} f(\mathbf{X}_i)$
  - exact balance in moments
  - extreme weights

- Stable weights (Zubizarreta. 2015. *J. Am. Stat. Assoc.*)

Summary

- Weighting methods as a generalization of matching methods

- Propensity score weighting
- Doubly robust estimation
- Robust estimation of propensity score for balancing covariates
- Calibration methods

- Recommended readings:
  - Imbens and Rubin. Chapter 17 (Section 8)
  - Lunceford and Davidian. 2004. “Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study.” *Statistics in Medicine*