Lecs 16-17

10/18/17 11:05

Recap: Observational Studies Under Unconfoundedness

Setup units $i = 1, \cdots, n$

observed treatment $Z_i \in \{0,1\}$ $\quad I_z = \{i : Z_i = z\}$

Outcomes $Y_i(0), Y_i(1)$ observed $Y_i \equiv Y_i(Z_i)$

Covars X_i

Want to estimate $SATT = \frac{1}{n_1} \Sigma_{i \in I_1} (Y_i(1) - Y_i(0))$

$$SATE = \frac{1}{n} \Sigma_i (Y_i(1) - Y_i(0))$$

Soln: compare like-to-like

Need: like is available — overlap

like to be sufficient — ignorability

Discussed

Regression-based approaches

$$\mathbb{E}[Y(z)] = \mathbb{E}[\mathbb{E}[Y \mid X, Z = z]]$$

Matching-Based Approaches

Find a subset, multisubset, or any redistribution of the raw samples that is better matched on the covariates

For now, focus on estimating ATT:

Keep the treated sample as is

$$\left(\frac{1}{n_1} \Sigma_{i \in I_1} Y(1) = \frac{1}{n_1} \Sigma_{i \in I_1} Y_i \right)$$

& create a matched control sample.

One-to-one matching

For each treated unit $i \in I_1$ find one unique control unit pair $m(i) \in I_0 : Z_{m(i)} = 0$

Matching as an estimator:

$$\hat{\tau} = \frac{1}{n_1} \Sigma_{i \in I_1} (Y_i - Y_{m(i)})$$

$$= \frac{1}{n_1} \Sigma_{i \in I_1} Y_i - \frac{1}{n_1} \Sigma_{i \in \{m(j) : Z_j = 1\}} Y_i$$

$I_0^{\text{matched}} = \{m(j) : Z_j = 1\} \subseteq I_0$

Matching variants:

1. Greedy – pass through treated units in random order, find closest $\leftarrow \delta(x_i, x')$ control that hasn't been matched yet & match it
2. Caliper – same but pick any control w/ $\delta(X_i, X_{m(i)}) \leq \delta_0$ at random.
3. Optimal – solve bipartite matching

Matching as pre-processing

Estimate $\hat{f}_0(x)$ on $\{(X_i, Y_i) : i \in I_0^{\text{matched}}\}$

& set $\hat{\tau} = \frac{1}{n_1} \sum_{i \in I_1} (Y_i - \hat{f}_0(x_i))$

E.g. for linear same as

$$Y_i \sim \beta^T X_i + \delta^T Z_i (X_i - \bar{X}_1) + \tau Z_i + \varepsilon$$

only on $I_1 \cup I_0^{\text{matched}}$

One-to-many: may use control units more than once – e.g NNM

$\rightarrow$ result is multisubset

Exact matching & Coarsened Exact Matching

Reweight samples to have the exact same covariate distributions

E.g.

$X_i \backslash Z_i$	Treated	Control
Diabetic	10	2
Non-diabetic	5	15
	2-to-1	2-to-15

either: prune any 14 non diabetic controls

or: give weight 5/15 to diabetic controls

& 1/45 to non-diabetic controls

(instead of uniform $1/17$)

Only possible if $\{X_i : Z_i = 1\} \subseteq \{X_i : Z_i = 0\}$

1. impossible if continuous → Coarsen
2. Unlikely for many factors → Coarsen even more

Fine-balance, near fine-balance, & mean-matching

Fine balance: Match marginal distributions instead of joint

E.g.

Treated

	Diabetic	Non-D	
Age ≥ 50	5	2	7
Age < 50	2	3	5
	7	5	

Control

10	0	10
0	10	10
10	10	

prune 3 & 5

or re-weight

7	0	7
0	5	5
7	5	

Near-fine balance: approximate fine balance by minimizing some norm (e.g. l_1 or l_2) in the marginal discrepancies

$$\sum_j \sum_{v \in V_j} \left| \frac{1}{n_1} \sum_{i \in I_1} \mathbb{I}[X_{ij} = z] - \frac{1}{n_0} \sum_{i \in I_0} \mathbb{I}[X_{ij} = z] \right|$$

Equivalent: encode each factor as a one hot vector & cocatenate to make X_i

after transforming X

$$= \sum_j \left| \frac{1}{n_1} \sum_{i \in I_1} X_{ij} - \frac{1}{n_0} \sum_{i \in I_0} X_{ij} \right|$$

$$= \left\| \frac{1}{n_1} \sum_{i \in I_1} X_i - \frac{1}{n_0} \sum_{i \in I_0} X_i \right\|_1$$

mean matching

Can have more general (e.g. cont) X

Propensity Score (Rosenbaum & Rubin '83)

Def: The propensity function is $e(x) = \mathbb{P}(Z_i = 1 \mid X_i = x)$

The propensity score of unit is $e(X_i)$

Propensity score is a sufficient statistic for matching

Def: $b(x)$ is a balancing score if $X_i \perp\!\!\!\perp Z_i \mid b(X_i)$

$p(X_i \mid b(X_i)) = p(X_i \mid Z_i, b(X_i))$

E.g. $b(x) = x \to X_i \perp\!\!\!\perp Z_i \mid X_i$

$\boxed{p(Z_i \mid b(X_i)) = p(Z_i \mid X_i)}$

Thm Suppose $b(x)$ is a balancing score.

Weak ignorability wrt X $\Rightarrow$ Weak ignorability wrt $b(X)$

$Y(z) \perp\!\!\!\perp Z \mid X \ \forall z$ $\qquad$ $Y(z) \perp\!\!\!\perp Z \mid b(X) \ \forall z$

Pf: Fix z.

$$\mathbb{P}(Z_i = 1 \mid Y_i(z), b(X_i)) \quad \left(\overset{?}{=} \mathbb{P}(Z_i = 1 \mid b(X_i))\right)$$

$$= \mathbb{E}\left[\mathbb{P}(Z_i = 1 \mid Y_i(z), \cancel{b(X_i)}, X_i) \mid Y_i(z), b(X_i)\right]$$

$$= \mathbb{E}\left[\mathbb{P}(Z_i = 1 \mid X_i) \mid Y_i(z), b(X_i)\right]$$

$$= \mathbb{E}\left[\mathbb{P}(Z_i = 1 \mid X_i, b(X_i)) \mid Y_i(z), b(X_i)\right]$$

$\varphi(b) = \mathbb{P}(Z_i = 1 \mid b(X_i) = b)$

$\mathbb{E}[\varphi(b(X_i)) \mid b(X_i), Y_i(z)]$

$$= \mathbb{E}\left[\boxed{\mathbb{P}(Z_i = 1 \mid b(X_i))} \mid Y_i(z), \boxed{b(X_i)}\right]$$

$g(X, Z)$

$\mathbb{E}[\cancel{f(X)} \mid X, Z]$

$$= \mathbb{P}(Z_i = 1 \mid b(X_i))$$

Thm The propensity score is a balancing score.

Pf: recall $e(x) = \mathbb{P}(Z_i = 1 \mid X_i = x)$ PS is $e(X_i)$

$$\mathbb{P}(Z_i = 1 \mid e(X_i)) \quad \left(\overset{?}{=} \mathbb{P}(Z_i = 1 \mid X_i, e(X_i))\right)$$

$$= \mathbb{E}\left[\mathbb{P}(Z_i = 1 \mid \cancel{e(X_i)}, X_i) \mid e(X_i)\right]$$

$$
\begin{aligned}
&= \mathbb{E}[e(X_i) \mid e(X_i)] \\
&= e(X_i) \\
&= \mathbb{P}(Z_i = 1 \mid X_i) \\
&= \mathbb{P}(Z_i = 1 \mid X_i, e(X_i))
\end{aligned}
$$

Thm: The propensity score is the coarsest balancing score:
If $b(x)$ is a balancing score
then $\exists g : e(X_i) = g(b(X_i))$

Pf: Suppose $\nexists g$. Then $\exists x, x'$ s.t.
$b = b(x) = b(x')$ but $e(x) \neq e(x')$.
Then

$$
\begin{array}{ccc}
\mathbb{P}(Z_i = 1 \mid X_i = x) & \neq & \mathbb{P}(Z_i = 1 \mid X_i = x') \\
\| & & \| \\
\mathbb{P}\left(Z_i = 1 \,\middle|\, \substack{X_i = x \\ b(X_i) = b(x)}\right) & & \vdots \\
\| & & \| \\
\mathbb{P}(Z_i = 1 \mid b(X_i) = b(x)) & & \mathbb{P}(Z_i = 1 \mid b(X_i) = b(x')) \\
\| & & \| \\
\mathbb{P}(Z_i \mid b(X_i) = b) & = & \mathbb{P}(Z_i = 1 \mid b(X_i) = b). \; \square
\end{array}
$$

Matching on the estimated propensity score

- How to estimate propensity score? / Dimensionality reduction?
- Continuity of outcome of effect given propensity
- Correct propensity model vs correct regression model
  - Double robustness
- Interesting fact:
  If logistic model is right then matching on the estimated propensity score is strictly better than matching on the real propensity score.
  (Abadie & Imbens)
  Why: induces actual covariate balance

just randomization vs covariate balance

In practice of PSM, iterate over propensity models until a metric of covariate balance small

Why propensity score? (King & ....)

Propensity Score Weighting

Original idea of comparing like to like: Want to make each control unit (in the redistribution) look like it actually came from the treated distribution.

→ reweight each control unit by

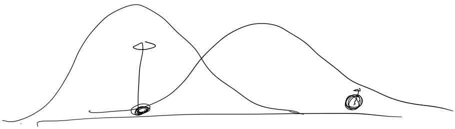
$$\frac{1}{n_1} \quad \frac{e(x_i)}{1-e(x_i)}$$

given $\hat{e}$:

$$\hat{\tau}^{IPW} = \frac{1}{n_1} \sum_{i \in I_1} Y_i - \frac{1}{n_1} \sum_{i \in I_0} \frac{\hat{e}(x_i)}{1-\hat{e}(x_i)} Y_i$$

for ATT

$$= \frac{1}{n_1} \sum_{i=1}^{n} \left(Z_i - (1-Z_i) \frac{\hat{e}(x_i)}{1-\hat{e}(x_i)} \right) Y_i$$



for ATE:

$$\hat{\tau}^{IPW} = \frac{1}{n} \sum_i \left(\frac{Z_i}{\hat{e}(x_i)} - \frac{1-Z_i}{1-\hat{e}(x_i)} \right) Y_i$$

really need overlap

$$\mathbb{E}\left[\left(\frac{Z_i}{e(x_i)} - \frac{1-Z_i}{1-e(x_i)} \right) Y_i \right]$$

$$\parallel$$

$$\mathbb{E}\left[\frac{Z_i \, Y_i(1)}{e(x_i)} - \frac{(1-Z_i) Y_i(0)}{1-e(x_i)} \right]$$

$\parallel$

$$\mathbb{E}\left[\mathbb{E}\left[\ -\!\!//\!\!-\ \middle|\, X_i\right]\right]$$

$\parallel$ $\qquad Z_i = \mathbb{I}[Z_i = 1] \qquad 1 - Z_i = \mathbb{I}[Z_i = 0]$

$$\mathbb{E}\left[\frac{\mathbb{E}[Y_i(1) \mid X_i]\,\cancel{\mathbb{E}[Z_i \mid X_i]}}{\cancel{e(X_i)}} - \frac{\mathbb{E}[Y_i(0)]\,\cancel{\mathbb{E}[1 - Z_i \mid X_i]}}{\cancel{1 - e(X_i)}}\right]$$

$$= \mathbb{E}\left[\mathbb{E}[Y_i(1) \mid X_i] - \mathbb{E}[Y_i(0) \mid X_i]\right]$$

$$= \mathbb{E}\left[Y_i(1) - Y_i(0)\right]$$

for ATT

$$\mathbb{E}\left[\left(Z_i - (1 - Z_i) \cdot \frac{e(X_i)}{1 - e(X_i)}\right) Y_i\right] \overset{\text{Want}}{=} \mathbb{E}\left[Z_i\left(Y_i(1) - Y_i(0)\right)\right]$$

$\parallel$

$$\mathbb{E}\left[Z_i Y_i(1) - (1 - Z_i) Y_i(0) \cdot \frac{e(X_i)}{1 - e(X_i)}\right]$$

$\parallel$

$$\mathbb{E}\left[\mathbb{E}\left[\ -\!\!//\!\!-\ \middle|\, X_i\right]\right]$$

$\parallel$

$$\mathbb{E}\left[e(X_i)\mathbb{E}[Y_i(1) \mid X_i] - \cancel{(1 - e(X_i))}\,\mathbb{E}[Y_i(0) \mid X_i]\frac{e(X_i)}{\cancel{1 - e(X_i)}}\right]$$

$\parallel$

$$\mathbb{E}\left[\mathbb{P}(Z_i = 1 \mid X_i)\,\mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i\right]\right]$$

$\parallel$

$$\mathbb{E}\left[\mathbb{E}\,Z_i\left(Y_i(1) - Y_i(0)\right) \mid X_i\right]$$

$\parallel$

$$\mathbb{E}\left[Z_i\left(Y_i(1) - Y_i(0)\right)\right]$$

Problems: instability of weights
plugging a nuisance param in the denom
model specification
no overlap is hard

Common tricks to improve IPW:

Trimming/clipping: $\min\left(M, \dfrac{\hat{e}(X_i)}{1 - \hat{e}(X_i)}\right)$ (ATT)

$$Z_i \min\left(M, 1/\hat{e}(X_i)\right) - (1-Z_i)\min\left(M, 1/1-\hat{e}(X_i)\right)$$

$$\updownarrow \; 1/\max(1/M, \hat{e}(X_i))$$

forcing estimated propensities to have strong overlap

some sort of bias-var control

Normalization: for ATT weight controls by

$$\frac{1}{n_1} \cdot \frac{\hat{e}(X_i)/1-\hat{e}(X_i)}{\sum_{j \in I_0} \hat{e}(X_j)/1-\hat{e}(X_j)}$$

May introduce bias but is still consistent

$$\begin{aligned} & \mathbb{E}\left[\sum_{j \in I_0} e(X_j)/1-e(X_j)\right] \\ & \quad \| \\ & \sum_{i=1}^n \mathbb{E}\left[\cancel{(1-Z_i)} \cdot e(X_i)/\cancel{1-e(X_i)}\right] \\ & \quad \| \\ & \sum_{i=1}^n \mathbb{E}[e(X_i)] \\ & \quad \| \\ & \sum_{i=1}^n \mathbb{E}[Z_i] = \sum_{i=1}^n \mathbb{P}(Z_i=1) \longrightarrow n \cdot \frac{n_1}{n} = n_1 \end{aligned}$$

By slutzky we get consistency.

For ATE: normalize on each side

make the treated weights sum to 1

& —"— control —"—

More advanced bias-var control: shrinkage

toward 0 or toward eq weights

$$W_i \leftarrow (1-\lambda)W_i$$

$W_i \leftarrow (1-\lambda)W_i + \lambda \frac{1}{n_0}$

more direct way to control variance as we will see.

Double Robustness

What is it? let's think about ATT

two types of estimators that we looked at were

$$\hat\tau^{reg} = \frac{1}{n_1}\sum_{i\in I_1}\left(Y_i - \hat f_0(X_i)\right) \qquad \hat\tau^{IPW} = \frac{1}{n_1}\sum_i\left(Z_i - (1-Z_i)\frac{\hat e(x_i)}{1-\hat e(x_i)}\right)Y_i$$

$$\hat\tau^{AIPW} = \frac{1}{n_1}\sum_{i=1}^n\left(Z_i-(1-Z_i)\frac{\hat e(x_i)}{1-\hat e(x_i)}\right)\left(Y_i - \hat f_0(x_i)\right)$$

$$= \underbrace{\frac{1}{n_1}\sum_{i=1}^n\left(Z_i-(1-Z_i)\frac{\hat e(x_i)}{1-\hat e(x_i)}\right)Y_i}_{\hat\tau^{IPW}} - \underbrace{\frac{1}{n_1}\left(Z_i-(1-Z_i)\frac{\hat e(x_i)}{1-\hat e(x_i)}\right)\hat f_0(x_i)}_{\text{Correction}}$$

$$= \underbrace{\frac{1}{n_1}\sum_{i\in I_1}\left(Y_i - \hat f_0(x_i)\right)}_{\hat\tau^{reg}} - \underbrace{\frac{1}{n_0}\sum_{i\in I_0}\frac{n_0\,\hat e(x_i)}{n_1(1-\hat e(x_i))}\underbrace{\left(Y_i-\hat f_0(x_i)\right)}_{\hat\varepsilon_i}}_{\text{Correction}}$$

$$\mathbb{E}[\hat\varepsilon_i \mid T_i=1]$$

for ATE:

$$\hat\tau^{AIPW} = \frac{1}{n}\sum_{i=1}^n\left(\hat f_1(x_i)-\hat f_0(x_i)\right) + \frac{1}{n}\sum_{i=1}^n\left(\frac{Z_i}{\hat e(x_i)} - \frac{1-Z_i}{1-\hat e(x_i)}\right)\left(Y_i - \hat f_{Z_i}(x_i)\right)$$

$$= \left(\frac{1}{n}\sum_{i=1}^n \hat f_1(x_i) + \frac{1}{n}\sum_{i\in I_1}\frac{1}{\hat e(x_i)}\left(Y_i - \hat f_1(x_i)\right)\right) \leftarrow \hat\mu_1^{AIPW}$$

$$- \left(\frac{1}{n}\sum_{i=1}^n \hat f_0(x_i) + \frac{1}{n}\sum_{i\in I_0}\frac{1}{1-\hat e(x_i)}\left(Y_i - \hat f_0(x_i)\right)\right) \leftarrow \hat\mu_0^{AIPW}$$

$$= \hat\mu_1^{AIPW} - \hat\mu_0^{AIPW}$$

ⓐ If $f_1 = \hat f_1$ then $\mathbb{E}[\text{thing inside sum}] = \mathbb{E}[Y(1)]$

ⓑ If $e = \hat e$ then =/=

$$\to \mathbb{E}\left[\hat f_1(X_i) + \frac{Z_i}{\hat e(X_i)}\left(Y_i - \hat f_1(X_i)\right)\right] \overset{?}{=} \mathbb{E}[Y(1)]$$

ⓐ $= \mathbb{E}\left[f_1(X) + \frac{Z}{\hat e(X)}\left(Y(1) - f_1(X)\right)\right]$

$\textcircled{=}\ \mathbb{E}\left[f_1(X) + \frac{\mathbb{E}[Z\mid X]}{\hat e(X)}\cancelto{0}{\left(f_1(X) - f_1(X)\right)}\right]$

$= \mathbb{E}[f_1(X)] = \mathbb{E}\left[\mathbb{E}[Y(1)\mid X]\right] = \mathbb{E}[Y(1)]$

ⓑ $= \mathbb{E}\left[\hat f_1(X) + \frac{Z}{e(X)}\left(Y(1) - \hat f_1(X)\right)\right]$

$= \mathbb{E}\left[\hat f_1(X) + \frac{e(X)}{e(X)}\cdot\left(f_1(X) - \hat f_1(X)\right)\right]$

$= \mathbb{E}\left[\hat f_1(X) - \hat f_1(X) + f_1(X)\right] = \mathbb{E}[Y(1)]$