

## Causal Inference Class Notes – 02/22/2010

- Payel Chowdhury

Clearing up a couple of points .....

### Propensity Score matching analysis:

Recall the weighting estimator (with known propensity score ( $e(x_i)$ ))

$$\hat{\tau}_{ATE,weighted} = \frac{\sum_{i=1}^N \frac{w_i Y_i^{obs}}{e(x_i)}}{\sum_{i=1}^N \frac{w_i}{e(x_i)}} - \frac{\sum_{i=1}^N \frac{(1-w_i) Y_i^{obs}}{1-e(x_i)}}{\sum_{i=1}^N \frac{(1-w_i)}{1-e(x_i)}}$$

We can also use weighting to estimate  $\tau_{ATT}$ . The key is to weight each case (both treatment and control) by the propensity score  $e(x_i)$ :

$$\hat{\tau}_{ATT,weighted} = \frac{\sum_{i=1}^N e(x_i) \frac{w_i Y_i^{obs}}{e(x_i)}}{\sum_{i=1}^N e(x_i) \frac{w_i}{e(x_i)}} - \frac{\sum_{i=1}^N e(x_i) \frac{(1-w_i) Y_i^{obs}}{1-e(x_i)}}{\sum_{i=1}^N e(x_i) \frac{(1-w_i)}{1-e(x_i)}} = \bar{Y}_t - \frac{\sum_{i=1}^N \frac{e(x_i)}{1-e(x_i)} (1-w_i) Y_i^{obs}}{\sum_{i=1}^N \frac{e(x_i)}{1-e(x_i)} (1-w_i)}$$

### Bias of unadjusted analysis – let's review once again the form of the bias

Reference: Chapter 17, page 7 (Sub-classification)

$$\hat{\tau}_{no-adjust} = \bar{Y}_t - \bar{Y}_c$$

Therefore,

$$E(\hat{\tau}_{no-adjust}) = E[Y_i(1) | W_i = 1] - E[Y_i(0) | W_i = 0]$$

$$Bias = E(\hat{\tau}_{no-adjust}) - \tau = E[Y_i(1) | W_i = 1] - E[Y_i(0) | W_i = 0] - (E[Y_i(1)] - E[Y_i(0)])$$

$$\text{Where, } E[Y_i(1)] = E[Y_i(1) | W_i = 1] \cdot \Pr[W_i = 1] + E[Y_i(1) | W_i = 0] \cdot \Pr[W_i = 0]$$

Thus,

$$\begin{aligned} Bias &= E(\hat{\tau}_{no-adjust}) - \tau = E[Y_i(1) | W_i = 1] - E[Y_i(1) | W_i = 1] \cdot \Pr[W_i = 1] \\ &\quad - E[Y_i(1) | W_i = 0] \cdot \Pr[W_i = 0] \\ &\quad - E[Y_i(0) | W_i = 0] + E[Y_i(0) | W_i = 0] \cdot \Pr[W_i = 0] \end{aligned}$$

$$\begin{aligned}
& + E[Y_i(0) | W_i = 1] \cdot \Pr[W_i = 1] \\
= & \Pr[W_i = 0] \cdot [E[Y_i(1) | W_i = 1] - E[Y_i(1) | W_i = 0]] \\
& + \Pr[W_i = 1] \cdot [E[Y_i(0) | W_i = 1] - E[Y_i(0) | W_i = 0]]
\end{aligned}$$

Now, assume a linear model:

$$Y_i(1) = \alpha_1 + \beta' \cdot X$$

$$Y_i(0) = \alpha_0 + \beta' \cdot X$$

Then each term above in square brackets is  $\beta(\bar{X}_T - \bar{X}_C)$  and thus the sum of the two terms is  $\beta(\bar{X}_T - \bar{X}_C)$ . Thus this is the bias in the naïve estimate.

**Now, let's finish the discussion of the observational study with regular assignment mechanism, the unconfounded case.**

### **Topics still to be covered:**

Overlap/ Robustness (Imbens, 2004) – matching/ propensity scores “do the right thing”

Variance Estimation

Assessing Unconfoundedness

Evaluation Literature

### **Variance Estimation (Chapter 19, I & R)**

$$\text{ATE: } \tau_{FS} = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$$

$$\tau_{SP} = E[Y_i(1) - Y_i(0)]$$

Estimate is the same in either FS or SP case but the variance can depend on which estimand is being used. The same comment applied for ATT.

=> Imbens and Rubin recommend focusing on the finite sample estimand for talking about variance estimation.

General approach for variance estimation:

Consider an estimator which is a linear weighted average of the observed data.

$$\hat{\tau} = \sum_{i=1}^N \lambda_i Y_i^{obs} \text{ with } \lambda_i = \lambda(W_i, X_i, W_{(-i)}, X_{(-i)}), \sum_{\substack{i=1 \\ W_i=1}}^N \lambda_i = +1 \text{ and } \sum_{\substack{i=1 \\ W_i=0}}^N \lambda_i = -1.$$

Many of our estimators are of this type:

Example:

Estimator	$W_i = 1$	$W_i = 0$
Naïve sample mean	$\lambda_i = \frac{1}{N_T}$	$\lambda_i = -\frac{1}{N_C}$
Weighted propensity score	$\lambda_i = \frac{1}{\sum_{i=1}^N \frac{W_i}{e(x_i)}}$	$\lambda_i = -\frac{1}{\sum_{i=1}^N \frac{(1-W_i)}{(1-e(x_i))}}$

To build the variance estimator we first need some notation.

$$\mu_c(x) = E(Y_i(0) | X_i = x) = E(Y_i^{obs} | W_i = 0, X_i = x)$$

$$\mu_t(x) = E(Y_i(1) | X_i = x) = E(Y_i^{obs} | W_i = 1, X_i = x)$$

$$\sigma_c^2(x) = Var(Y_i(0) | X_i = x) = Var(Y_i^{obs} | W_i = 0, X_i = x)$$

$$\sigma_t^2(x) = Var(Y_i(1) | X_i = x) = Var(Y_i^{obs} | W_i = 1, X_i = x)$$

$$\mu_i = \begin{cases} \mu_c(x_i) & \text{if } W_i = 0 \\ \mu_t(x_i) & \text{if } W_i = 1 \end{cases}$$

$$\sigma_i^2 = \begin{cases} \sigma_c^2(x_i) & \text{if } W_i = 0 \\ \sigma_t^2(x_i) & \text{if } W_i = 1 \end{cases}$$

Note that we can write our estimator as

$$\hat{\tau} = \sum_{i=1}^N \lambda_i \mu_i + \sum_{i=1}^N \lambda_i [Y_i^{obs} - \mu_i]$$

The 1<sup>st</sup> term on the right hand side (RHS) of the above expression for  $\hat{\tau}$  is  $\tau + bias$  and the 2<sup>nd</sup> term of RHS has the expected value of zero under unconfoundedness and variance =  $\sigma_i^2$ . Then the variance just comes from the second term (first is constant) and can be written:

$$Var(\hat{\tau} | X, W) = \sum_{i=1}^N \lambda_i^2 \sigma_i^2$$

$$\hat{V}(\hat{\tau} | X, W) = \sum_{i=1}^N \lambda_i^2 \hat{\sigma}_i^2, \text{ where we need to estimate } \sigma_i^2.$$

Idea: Suppose we could match on  $X, W$  exactly. If case  $i$  and case  $j$  match exactly (say both are treatment cases), then

$$E((Y_i^{obs} - Y_j^{obs})^2 | W_i = W_j = 1, X_i = X_j = x) = 2 \cdot \sigma_i^2(x)$$

This means that

$$\hat{\sigma}_i^2 = \frac{(Y_i^{obs} - Y_j^{obs})^2}{2} \text{ is an unbiased estimator for } \sigma_i^2.$$

- If  $> 2$  matching units  $\Rightarrow$  can pool information to get a variance estimate.
- $\hat{\sigma}_i^2$  is unbiased, not consistent (it is based only 2 units). But still it turns out that  $\hat{V} \rightarrow Var$  if we plug this in for each  $i$ .
- Problem with the approach: Can't find perfect matches. Instead use nearest neighbor to  $i$  within its group (treatment/ control). Call the neighbor  $k(i)$  and use

$$\hat{\sigma}_i^2 = \frac{(Y_i^{obs} - Y_{k(i)}^{obs})^2}{2}$$

Note that we can use more than one nearest neighbor. We could also use non-parametric variance estimate.

Alternative to the general approach – variance estimates based on specific estimators. For example, remember the paired estimator in matching.

$$\hat{\tau} = \frac{1}{N_T} \sum_{i=1}^N W_i (Y_i^{obs} - Y_{k(i)})$$

For it we can use the variance among the individual paired estimates

$$\text{Var}(\hat{\tau}) = \frac{1}{N_T} \left[ \frac{1}{(N_T - 1)} \sum_{i=1}^N W_i (Y_i^{obs} - Y_{k(i)} - \hat{\tau})^2 \right]$$

Similarly for approaches based on regression we can use regression coefficient variance formulas to get variance estimates.

Imbens and Rubin argue that the general approach is better in that the other approaches tend to rely on stronger assumptions (e.g., exact matches or a linear model).

Some have suggested that we could also use bootstrap to estimate variance but the theoretical justification for the bootstrap has not been worked out in this case.