ESTIMATING THE SATE

In the previous lesson, we introduced the sharp null hypothesis of no effect in uniform randomized experiment

This lesson we introduce the following estimate for the SATE:

$$n^{-1} \sum_{s=1}^{S} \sum_{i'=1}^{n_s} y_{si'}(1) - y_{si'}(0) \tag{1}$$

where effects now vary over subjects.

ESTIMATING THE SATE

For simplicity, first consider the case of a completely randomized experiment: $S = 1$.

$$\bar{Y}_1 - \bar{Y}_0 = \frac{\sum_{i=1}^{n} Z_i Y_i}{\sum_{i=1}^{n} Z_i} - \frac{\sum_{i=1}^{n}(1 - Z_i) Y_i}{\sum_{i=1}^{n}(1 - Z_i)} \tag{2}$$

where $Y_i = Z_i y_i(1) + (1 - Z_i) y_i(0))$ and $\sum_{i=1}^{n} Z_i = n_1$ for all assignment vectors in Ω

ESTIMATING THE SATE

We want to show that (**??**) is an unbiased estimate of the SATE. Remember: the potential outcomes are fixed constants, randomization is through the assignment mechanism

The expectation is taken over the randomization set Ω for the completely randomized experiment. Remember; all $n!/n_1!n_0!$ assignment vectors are equally likely, and thus each subject appears in n_1/n of the assignments. Thus,

$$E\left(\frac{\sum_{i=1}^n Z_i Y_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n (1-Z_i) Y_i}{\sum_{i=1}^n (1-Z_i)}\right) = \left(\sum_{i=1}^n E\left(\frac{Z_i y_i(1)}{n_1} - \frac{(1-Z_i) y_i(0)}{n_0}\right)\right)$$
$$= \sum_{i=1}^n \frac{(n_1/n) y_i(1)}{n_1} - \frac{(n_0) y_i(0)}{n_0/n} = n^{-1} \sum_{i=1}^n (y_i(1) - y_i(0)). \quad (3$$

ESTIMATING THE SATE

The result extends readily to block randomized experiment—-remember a block randomized consists of S independent completely randomized experiments. Now rewrite the SATE (**??**) as

$$\sum_{s=1}^{S} \frac{n_s}{n} \frac{\sum_{i'=1}^{n_s} (y_{si'}(1) - y_{si'}(0))}{n_s} \tag{4}$$

which is a weighted average of the within stratum SATEs, with weights equal to the proportion of units within the stratum,

ESTIMATING THE SATE

$$\bar{Y}_{s1} - \bar{Y}_{s0} = \frac{\sum_{i'=1}^{n_s} Z_{si} Y_{si'}}{\sum_{i'=1}^{n_s} Z_{si'}} - \frac{\sum_{i'=1}^{n_s} (1 - Z_{si'}) Y_{si'}}{\sum_{i'=1}^{n_s} (1 - Z_{si'})} \tag{5}$$

is unbiased for the SATE within stratum s, and

$\sum_{s=1}^{S} (n_s/n)(\bar{Y}_{s1} - \bar{Y}_{s0})$ is unbiased for $\sum_{s=1}^{S} \frac{n_s}{n} \frac{\sum_{i'=1}^{n_s} (y_{si'}(1) - y_{si'}(0))}{n_s}$

ESTIMATING THE SATE

Test hypotheses and construct interval estimates: Need the variance of the estimated SATE and and an estimate of the variance. Only give results here—-the derivation is tedious! only for the case of the completely randomized experiment.

$$V(\bar{Y}_1 - \bar{Y}_0) = E([(\sum_{i=1}^{n}(\frac{Z_i y_i(1)}{n_1} - \frac{(1-Z_i)y_i(0)}{n_0})) - n^{-1}\sum_{i=1}^{n}(y_i(1) - y_i(0)]^2) \tag{6}$$

ESTIMATING THE SATE

$$V(\bar{Y}_1 - \bar{Y}_0) = \frac{1}{n_0(n-1)}\sum_{i=1}^{n}(y_i(0) - \bar{Y}(0))^2$$
$$+\frac{1}{n_1(n-1)}\sum_{i=1}^{n}(y_i(1) - \bar{Y}(1))^2 -$$
$$\frac{1}{n(n-1)}\sum_{i=1}^{n}(y_i(1) - y_i(0) - (\bar{Y}(1) - \bar{Y}(0)))^2. \tag{7}$$

where $\bar{Y}(0) = \sum_{i=1}^{n} y_i(0)$, $\bar{Y}(1) = \sum_{i=1}^{n} y_i(1)$.
First term is the variance of the $y_i(0)$ terms, divided by the number of observations in the control group

Second term is the variance of the $y_i(1)$ terms, divided by the number of treatment units

Third term is the variance of the unit effects divided by the

ESTIMATING THE SATE

Recall the third term is the variance of the unit effects, divided by n. Generally no information on 3rd term– we see $y_i(0)$ or $y_i(1)$, not both

But–when unit treatment effects $y_i(1) - y_i(0)$ are constant, this term vanishes.

ESTIMATING THE SATE

To estimate the variance (**??**), let $s_0^2 = (n_0 - 1)^{-1} \sum_{i:Z_i=0}(y_i(0) - \bar{Y}_0)^2$, i.e., s_0^2 is the "sample" variance of the control observations, and define s_1^2 analogously.

A conservative (because it will generally be larger) estimate of the variance (**??**) is then

$$\hat{V}(\bar{Y}_1 - \bar{Y}_0) = s_0^2/n_0 + s_1^2/n_1. \tag{8}$$

(**??**) is unbiased for (**??**) when the unit treatment effects are constant.

ESTIMATING THE SATE

Hypothesis tests and interval estimates of the SATE may be obtained using a normal approximation for the randomization distribution of the estimated SATE (**??**):

$$\frac{\bar{Y}_1 - \bar{Y}_0}{(s_0^2/n_0 + s_1^2/n_1)^{1/2}} \text{ is approximately } N(0,1)$$

Sometimes the n units are of primary interest, often not.