Segment 4: Observational Studies

Section 05: Propensity Scores: A First Look

Propensity Scores: Estimating the Assignment Mechanism

In observational studies, the **key assumption** about whether the observed data can be used to estimate causal effects will be whether the assignment mechanism is (conditionally) ignorable/unconfounded:

$$Pr(\mathbf{Z}|\mathbf{X}, \mathbf{Y}^t, \mathbf{Y}^c) = Pr(\mathbf{Z}|\mathbf{X})$$

For the cases considered in this course, we further assume that:

$$Pr(\mathbf{Z}|\mathbf{X}) = c \prod_{i=1}^{n} \left(Pr(Z_i = 1|\mathbf{X}_i)^{Z_i} (1 - Pr(Z_i = 1|\mathbf{X}_i))^{1-Z_i} \right)$$

where $Pr(Z_i = 1|\mathbf{X}_i) = e_i(\mathbf{X}_i)$ (or just e_i) is called the **propensity score**, i.e., the propensity of the i^{th} unit to receive treatment $Z = 1$

Propensity Scores in Observational Studies

$$Pr(\mathbf{Z}|\mathbf{X}) = c\prod_{i=1}^{n} e_i(\mathbf{X}_i)^{Z_i}(1-e_i)^{1-Z_i}$$

- $e_i(\mathbf{X}_i) = e_i = Pr(Z_i = 1|\mathbf{X}_i)$
- Are *unknown* and must be *estimated*
- Represent a "one number" summary of all of the covariate information in $\mathbf{X}_i$
- Can be used for confounding adjustment
  - E.g., rather than block on $\mathbf{X}_i$, block on the scalar summary e_i
- Can be used to assess overlap/balance
  - E.g., rather than assess overlap across multidimensional $\mathbf{X}_i$, assess overlap on e_i

Some Theory: Balancing Scores

from Rosenbaum and Rubin (1983)

A *balancing score*, $b(\mathbf{X}_i)$ is a lower-dimensional summary of $\mathbf{X}_i$ such that

Y_i^0 Y_i^1

$$Z_i \perp\!\!\!\perp X_i|b(\mathbf{X}_i)$$

$$Z_i \perp\!\!\!\perp Y_i(0), Y_i(1)|\mathbf{X}_i \Rightarrow Z_i \perp\!\!\!\perp Y_i(0), Y_i(1)|b(\mathbf{X}_i)$$

- Treated/control units with the same value of $b(\mathbf{X})$ have the same distribution of $\mathbf{X}$
- Points towards a vast set of methods that adjust for confounding via adjustment for *balancing scores*
- The propensity score is one (common) example of a *balancing score*

Main Theoretical Propensity Score Results

from Rosenbaum and Rubin (1983)

1. The propensity score is a balancing score
2. The propensity score is the *coarsest* balancing score
3. Strong ignorability $|\mathbf{X} \Rightarrow$ Strong ignorability $|e(\mathbf{X})$
4. Sample estimates of $e(\mathbf{X})$ can produce sample balance on $\mathbf{X}$
5. $e_i(\mathbf{X}_i) - e_j(\mathbf{X}_j)$ can be viewed as a *distance* between units i, j
   - In terms of their propensity to have been assigned $Z = 1$
6. **At any value of $e(\mathbf{X})$, $\bar{Y}_t^{obs} - \bar{Y}_c^{obs}$ is an unbiased estimate of the average causal effect at that value of $e(\mathbf{X})$.**
   - Pair matching, subclassification, and covariance adjustment based on $e(X_i)$ can produce unbiased causal estiamtes
   - Under ignorability!

Relevance for Estimating Causal Effects

$$E[Y^0 | X, Z=0] \underset{\text{ignorability}}{=} E[Y^0 | X] = E[Y^0 | e(X)]$$

$$E[Y^1 | X, Z=1] = E[Y^1 | X] = E[Y^1 | e(X)]$$

$$E_{e(X)}\Big[\underbrace{E\big[Y^1 | e(X) - Y^0 | e(X)\big]}_{\tau_{CATE|e(X)}}\Big] = \tau_{SATE}$$

Relevance for Estimating Causal Effects

Corollaries: Under ignorability and $e(\mathbf{X})$ a balancing score:

- Expected difference in Y^{obs} among treated/untreated pairs (or groups) of units with the same value of $e(\mathbf{X})$ is an estimate of the CATE$_{e(\mathbf{X})}$ at that value of $e(\mathbf{X})$ and the average of these CATEs across the values of $e(\mathbf{X})$ is the overall average ATE.
- $\rightarrow$ Pair matching, subclassifying, and covariance adjustment for $e(\mathbf{X})$ will adjust for confounding
- Essentially, $e(\mathbf{X})$ can serve as a good "blocking factor" because it ensures (in theory) balance on $\mathbf{X}$ across treatment groups

Estimating the Propensity Score (PS)

$$e(\mathbf{X}_i) = Pr(Z_i = 1|\mathbf{X}_i)$$

- $Pr(Z = 1|\mathbf{X}) = E[Z|\mathbf{X}] \rightarrow$ regression for a binary "outocome"
  - Simplest method: Logistic regression
    - `stan_glm(treat ~ x1 + x2 + ..., family=binomial(link = "logit"))`
  - Could use any model for binary outcome (GLM, boosted regression, etc.)
  - "Estimated PS": $\hat{e}_i$
- Goal is to estimate $\hat{e}_i$ *that balances covariates,* X_i
  - **Not** necessarily the same as the "usual" goal of obtaining the best estimate in terms of mean squared error, maximum discrimination, predictive power, etc.
- Estimating the PS is just a "means to an end," we don't generally care about interpreting the PS

Not Just Any Dimension Reduction

- There are *many* tools for dimension reduction to reduce $\mathbf{X} \rightarrow$ scalar
  - But not all dimension reduction tools are designed to adjust for *confounding!*

- The *propensity score* is a very specific type of dimension reduction
  - Explicitly defined based on the *assignment mechanism*
  - Is a *balancing score*
  - Specifically defined to adjust for confounding

Practical Considerations

In practice, we need to make (at least) the following decisions about estimating the PS

- Which covariates to include
  - We *need* all of the *confounders*
  - We *want* predictors of Y (even if unrelated to W)
  - We *don't want* predictors of W that are unrelated to Y
  - (only confounders have implications for bias, other variables have implications for precision)
- How to empirically select variables?
  - Subject for a later lecture...
  - Common recommendation: Include everything to be safe (possibly including interactions)!
- Functional form of the model specification

Uses of Propensity Scores

Goal: Analyze the observational study to ensure that treated/control comparisons are "like vs. like" in terms of X

(assuming that all relevant X are measured)

AKA: "Adjust for confounding"

Strategies:

1. Stratify on X directly
2. Matching/blocking
3. Fit a regression model
4. Overlap checking/trimming observations

In observational studies, all of the above are frequently accomplished with *propensity scores*