IDS 702: Module 6.6

Propensity scores

Dr. Olanrewaju Michael Akande

Propensity scores

- The propensity score (ps) is defined as the conditional probability of receiving a treatment given pre-treatment covariates X.
- That is,

  $$e(X) = \mathbb{P}\mathrm{r}[W = 1|X] = \mathbb{E}[W|X],$$

  where $X = (X_1, \ldots, X_p)$ is the vector of p covariates/predictors.
- Propensity score is a probability, analogous to a summary statistic.
- Propensity score has really nice properties which makes it desirable to use within our causal inference framework.

IDS 702

Balancing property of propensity score

- Property 1. The propensity score e(X) balances the distribution of all X between the treatment groups:

$$W \perp X | e(X)$$

- Equivalently,

$$\mathbb{P}\mathrm{r}[W_i = 1 | X_i, e(X_i)] = \mathbb{P}\mathrm{r}[W_i = 1 | e(X_i)].$$

- The propensity score is NOT the only balancing score. Generally, a balancing score $b(x)$ is a function of the covariates such that:

$$W \perp X | b(X)$$

Remarks on the balancing property

- Rosenbaum and Rubin (1983) show that all balancing scores are a function of $e(X)$.
- If a subclass of units or a matched treatment-control pair are homogeneous in $e(X)$, then the treatment and control units have the same distribution of X.
- The balancing property is a statement on the distribution of X, NOT on assignment mechanism or potential outcomes.

Propensity score: unconfoundedness

- Property 2. If W is unconfounded given X, then W is unconfounded given $e(X)$, i.e.,
- That is, if

  $$Y_i(0), Y_i(1) \perp W_i | X_i$$

  holds, then

  $$Y_i(0), Y_i(1) \perp W_i | e(X_i),$$

  also holds.
- Given a vector of covariates that ensure unconfoundedness, adjustment for differences in propensity scores removes all biases associated with differences in the covariates.

IDS 702

Propensity score: unconfoundedness

- $e(X)$ can be viewed as a summary score of the observed covariates.
- This is great because causal inference can then be drawn through stratification, matching, regression, etc. using the scalar $e(X)$ instead of the high dimensional covariates.
- The propensity score balances the **observed covariates**, but does not generally balance **unobserved covariates**.
- In most observational studies, the propensity score e(X) is unknown and thus needs to be estimated.
- However, since we always observe X and W, estimation can be done using models for binary outcomes.

IDS 702

What's next?

Move on to the readings for the next module!

IDS 702