Unconfounded Treatment Assignment

Hwichang Jeong

July 26, 2022

Seoul National University

Introduction

- Move toward the setting of observational studies.
- Relax the classical randomized experiment assumption.

Assignment Assumption

- Probabilistic assignment
- Individualistic assignment
- Unconfounded assignment
  - $Pr(W|X, Y(0), Y(1))$ are free from dependence on the potential outcomes.
  - In combination with individualistic assignment,

$$Pr(W|X, Y(0), Y(1)) = c \cdot \prod_{i=1}^{N} e(X_i)^{W_i}(1 - e(X_i))^{1-W_i}$$

where $e(x)$ is propensity score.

Why Is Unconfoundedness an Important Assumption?

- The assumption is extremely widely used.

$$Y_i^{obs} = \alpha + \tau_{sp} \cdot W_i + X_i \beta + \epsilon_i$$

- It is assumed that $\epsilon_i \perp\!\!\!\perp W_i, X_i$

Why Is Unconfoundedness an Important Assumption?

- In the potential outcome formulation, we have

$$Y_i(0) = \alpha + X_i\beta + \epsilon_i$$

$$Y_i(1) = Y_i(0) + \tau_{sp}$$

- Then ϵ_i is a function of $Y_i(0)$ and X_i given the parameters

$$Pr(W_i = 1|X_i, Y_i(0), Y_i(1)) = Pr(W_i|\epsilon_i, X_i) = Pr(W_i|X_i)$$

Why Is Unconfoundedness an Important Assumption?

- By assuming unconfoundedness we can compare the particular treated units with control units.
- If there is an alternative of unconfoundedness, it must involve looking for a comparison which is different in terms of observed pre-treatment variables.
- In many cases it would appear implausible.

Balancing Scores And the Propensity Score

- Assuming individualistic assignment and unconfounded assignment,

$$Pr(W|X, Y(0), Y(1)) = c \cdot \prod_{i=1}^{N} e(X_i)^{W_i} (1 - e(X_i))^{1-W_i}$$

- A balancing score $b(x)$ is a function of the covariates such that

$$W_i \perp\!\!\!\perp X_i | b(X_i)$$

- The probability of receiving the treatment given the covariates is free of dependence on the covariates given the balancing score.

Balancing Scores And the Propensity Score

Lemma
The propensity score is a balancing score.

Lemma
The propensity score is the coarsest balancing score. That is, the propensity score is a function of every balancing score.

Design phase

- Prior to estimating causal effects it is important to conduct design phase of an observational study.
  - Assessing Balance - CH13, CH14
  - Subsample selection using matching on the propensity score - CH15
  - Subsample selection through trimming using propensity score - CH16

Estimation

- Discuss five broad classes of strategies for estimation
  - Model-Based imputation
  - Regression estimators
  - Weighting estimators that use the propensity score
  - Blocking estimators that use the propensity score (CH17)
  - Matching Estimators (CH18)
- Blocking and matching are relatively attractive because of the robustness properties that stem form the combination of methods.

Weighting Estimators that use the propensity score

- Weighting exploits the two equalties

$$\mathbb{E}\left[\frac{Y_i^{obs} \cdot W_i}{e(X_i)}\right] = \mathbb{E}_{sp}[Y_i(1)]$$

$$\mathbb{E}\left[\frac{Y_i^{obs} \cdot (1 - W_i)}{1 - e(X_i)}\right] = \mathbb{E}_{sp}[Y_i(0)]$$

Weighting Estimators that use the propensity score

- Horvitz and Thompson introduced estimator of ATE as

$$\hat{\tau}^{ht} = \frac{1}{N} \sum_{i:W_i=1} \lambda_i \cdot Y_i^{obs} - \frac{1}{N} \sum_{i:W_i=0} \lambda_i \cdot Y_i^{obs}$$

where

$$\lambda_i = \frac{1}{e(X_i)^{W_i} \cdot (1 - e(X_i))^{1-W_i}}$$

Blocking Estimators That Use the Propensity Score

- Let b_j, $j = 0, 1, \cdots, J$ denote the subclass boundaries, with $b_0 = 0$, $b_J = 1$.
- Let $B_i(j)$ be a binary indicator, equal to 1 if $b_{j-1} < e(X_i) < b_j$ and zero otherwise.

Blocking Estimators That Use the Propensity Score

- Estimate the finite-sample average effect in subclass j by

$$\hat{\tau}^{dif}(j) = \frac{\sum_{i:B_i(j)=1} Y_i \cdot W_i}{\sum_{i:B_i(j)=1} W_i} - \frac{\sum_{i:B_i(j)=1} Y_i \cdot (1 - W_i)}{\sum_{i:B_i(j)=1} (1 - W_i)}$$

- Estimate the overall finite-sample average effect of the treatment by

$$\hat{\tau}^{strat} = \sum_{j=1}^{J} \frac{N(j)}{N} \cdot \hat{\tau}^{dif}(j)$$

where $N(j) = \sum_{i=1}^{N} B_i(j)$

Matching Estimator

- For a given treated unit with a particular set of values for the covariates, one looks for a control unit with as similar a set of covariates as possible.

Efficiency Bounds

- Define

$$\mu_c(x) = \mathbb{E}_{sp}[Y_i(0)|X_i = x],\ \mu_t(x) = \mathbb{E}_{sp}[Y_i(1)|X_i = x]$$

$$\sigma_c^2(x) = \mathbb{V}_{sp}[Y_i(0)|X_i = x],\ \text{and } \sigma_t^2(x) = \mathbb{V}_{sp}[Y_i(1)|X_i = x]$$

- The super-population average treatment effect defined as

$$\tau_{sp} = \mathbb{E}_{sp}[Y_i(0) - Y_i(1)] = \mathbb{E}_{sp}[\tau_{sp}(X_i)]$$

where

$$\tau_{sp}(x) = \mu_t(x) - \mu_c(x) = \mathbb{E}_{sp}[Y_i(1) - Y_i(0)|X_i = x]$$

- The finite-sample average effect conditional on the values of the pre-treatment variables is defined as

$$\tau_{cond} = \frac{1}{N}\sum_{i=1}^{N} \tau_{sp}(X_i)$$

Efficiency Bounds

- Under unconfoundedness and probabilistic assignment, and without additional functional form restrictions beyond smoothness, the sampling variance bound for estimators for τ_{sp}, normalized by the sample size is

$$\mathbb{V}_{sp}^{eff} = \mathbb{E}_{sp}\left[\frac{\sigma_c^2(X_i)}{1-e(X_i)} + \frac{\sigma_t^2(X_i)}{e(X_i)} + \left(\tau_{sp}(X_i) - \tau_{sp}\right)^2\right]$$

- The sampling variance bound for estimators for τ_{cond} is

$$\mathbb{V}_{cond}^{eff} = \mathbb{E}_{sp}\left[\frac{\sigma_c^2(X_i)}{1-e(X_i)} + \frac{\sigma_t^2(X_i)}{e(X_i)}\right]$$