Causal Inference (Part 2)

Jae-kwang Kim [1]

[1] Adopted from the lecture notes of professor Marie Davidian

Observational Studies

- We now consider observational studies.
- **Complication**: Individuals may receive treatment according to physician discretion or their own choice.
- Therefore, individuals who receive treatment 1 may have different characteristics than those who receive treatment 0.
- Thus, the Difference-in-Means estimator $\hat{\tau}$ is no longer unbiased.
- We may want to utilize the auxiliary variable X to control the systematic difference.

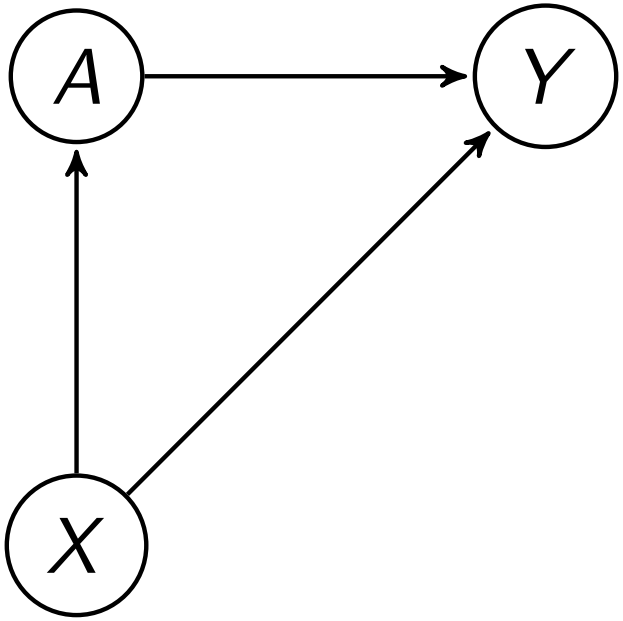
Confounders

- If a variable is related to both A and outcome variable $\{Y(1), Y(0)\}$, then it is called a confounder.
- There could be several confounders (e.g. gender, race, age group, education level).
- Suppose that the confounding variables are known. Among individuals who share the same confounders, all factors associated with the selection and outcome are taken into account.
- That is, the assignment of treatment is effectively *at random* for individuals sharing the same confounders.
- Formally expressed as

$$\{Y(1), Y(0)\} \perp A \mid X^*$$

where X^* is the set of all confounders.

Confounding factor example



- A: smoking
- Y: lung cancer
- X: stress

No Unmeasured confounders assumption

- The auxiliary variable X contains the individual characteristics recorded in the observed data and available to the data analyst.
- In general, there is no guarantee that all the confounders are included in X.
- So, we assume that the confounders are all included in X so that

$$\{Y(1), Y(0)\} \perp A \mid X \tag{1}$$

holds. This is *No Unmeasured Confounders* (NUC) assumption. This is also called the *ignorability* assumption.

- **Fundamental difficulty**: It is impossible to verify from the observed data that there are no unmeasured confounders and that (1) is valid.

Main Result

- Three assumptions
  1. SUTVA
  $$Y = Y(1) \cdot A + Y(0) \cdot (1 - A)$$
  2. NUC assumption in (1).
  3. Positivity assumption:
  $$P(A = 1 \mid X) > 0$$
  for all $x \in \mathcal{X}$ with $p_X(x) > 0$.
- Under the three assumptions, we obtain the following.

$$E\{E(Y \mid X, A = a)\} = E\{Y(a)\} \tag{2}$$

for $a = 0, 1$.

Justification

- SUTVA leads to

$$E\{Y \mid X, A = a\} = E\{Y(a) \mid X, A = a\}$$

- NUC assumption leads to

$$E\{Y(a) \mid X, A = a\} = E\{Y(a) \mid X\} \tag{3}$$

- Now, note that

$$E\left[E\{Y(a) \mid X\}\right] = \int_{\mathcal{X}} E\{Y(a) \mid X = x\} p_x(x) d\nu(x)$$

$$E\left[E\{Y(a) \mid X, A = a\}\right] = \int_{\mathcal{X}} E\{Y(a) \mid X = x, A = a\} p_x(x, a) d\nu(x)$$

Thus, given (3) holds, two terms are equal if $p(x, a) > 0$ for all $x \in \mathcal{X}$ with $p_x(x) > 0$.

Outcome regression

- Implication of (2) :

$$\tau = E\{E(Y \mid X, A = 1)\} - E\{E(Y \mid X, A = 0)\}$$

  depends on $E(Y \mid X, A)$, the regression function of Y on X and A.
- If

$$E(Y \mid X = x, A = a) = Q(x, a)$$

  is known, we can obtain

$$\tau = E\{Q(X, 1)\} - E\{Q(X, 0)\}.$$

- In practice, $Q(x, a)$ is unknown.

Outcome regression model

- Idea: Let's use a statistical model for $Q(x, a) = Q(x, a; \beta)$ with parameter β.
- For example, if Y is continuous, we can consider a linear regression model such as

$$Q(\mathbf{x}, a; \beta) = \beta_1 + \beta_2 a + \beta_3' \mathbf{x} + \beta_4' \mathbf{x} \cdot a$$

- Model parameter can be estimated from each sample.
- We can estimate τ by

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} \left\{ Q(x_i, 1; \hat{\beta}) - Q(x_i, 0; \hat{\beta}) \right\}$$

- Thus, the theory of imputation (= prediction) can be applied directly.

Propensity score approach

- The treatment indicator function A takes values in $\{0, 1\}$.
- The propensity score function $\pi(X) = P(A = 1 \mid X)$ reflects the "propensity" for choosing $A = 1$ for individuals with characteristics X.
- Originally proposed by Rosenbaum and Rubin (1983).

Inverse propensity score weighting

- For now, suppose that $\pi(X)$ is known.
- Estimation of $\mu_1 = E\{Y(1)\}$: Use

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i}{\pi(X_i)}$$

- Estimation of $\mu_0 = E\{Y(0)\}$: Use

$$\hat{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n} \frac{(1-A_i) Y_i}{1-\pi(X_i)}$$

- The unbiasedness requires the same three assumptions (SUTVA, NUC, and Positivity) in addition to $\pi(x)$ known.

Justification

Estimation of τ

- If $\pi(x)$ known

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i Y_i}{\pi(X_i)} - \frac{(1-A_i)Y_i}{1-\pi(X_i)}\right\}$$

- Observational studies: Posit and fit propensity model $\pi(x;\gamma)$

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i Y_i}{\pi(X_i;\hat{\gamma})} - \frac{(1-A_i)Y_i}{1-\pi(X_i;\hat{\gamma})}\right\}$$

- It is based on the assumption that the propensity model $\pi(x;\gamma)$ is correctly specified.

Outcome regression vs inverse propensity weighted estimator

- Both require SUTVA, NUC, and the positivity assumption.
- Outcome regression estimator: requires a model assumption on $E(Y \mid x, a) = Q(x, a; \beta)$.
- IPW estimator: requires a model assumption on $P(A = 1 \mid x) = \pi(x; \gamma)$
- Maybe a doubly robust estimator can be developed to protect against model misspecification.

Double robust estimation of τ

- Given fitted models $\pi(x;\hat{\gamma})$ and $Q(x,a;\hat{\beta})$, we can consider

$$\hat{\tau}_{\mathrm{DR}} = \hat{\mu}_{1,\mathrm{DR}} - \hat{\mu}_{0,\mathrm{DR}} \tag{4}$$

where

$$\hat{\mu}_{1,\mathrm{DR}} = \frac{1}{n}\sum_{i=1}^{n}\left[Q(x_i,1;\hat{\beta}) + \frac{A_i}{\pi(x_i;\hat{\gamma})}\left\{Y_i - Q(x_i,1;\hat{\beta})\right\}\right]$$

and

$$\hat{\mu}_{0,\mathrm{DR}} = \frac{1}{n}\sum_{i=1}^{n}\left[Q(x_i,0;\hat{\beta}) + \frac{1-A_i}{1-\pi(x_i;\hat{\gamma})}\left\{Y_i - Q(x_i,0;\hat{\beta})\right\}\right]$$

Justification

Discussion

- $\hat{\tau}_{\mathrm{DR}}$ is doubly robust.
- It is efficient (i.e. optimal) when both the propensity and the outcome regression models are correctly specified.
  - It achieves the smallest variance among the class of estimators of the form

$$\hat{\tau}_h = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{A_i Y_i}{\pi(X_i;\hat{\gamma})} - \frac{(1-A_i)Y_i}{1-\pi(X_i;\hat{\gamma})} - \{A_i - \pi(X_i;\hat{\gamma})\}\, h(X_i)\right]$$

  - That is, the optimal $h(x)$ is

$$h^*(X) = \frac{E(Y \mid X, A=1)}{\pi(X)} + \frac{E(Y \mid X, A=0)}{1-\pi(X)}$$

  - Based on the semiparametric theory (Tsiatis, 2006).

REFERENCES

Rosenbaum, P. R. and D. B. Rubin (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**, 41–55.

Tsiatis, A. A. (2006), *Semiparametric Theory and Missing Data*, Springer.