Causal Inference (What If)

Ch15 Outcome Regression and Propensity Scores

이종진

Seoul National University

ga0408@snu.ac.kr

August 12, 2021

Ch 15, Outcome Regression and Propensity Scores

- (IP weighting, standardization and g-estimation - the g-methods.)
- Outcome regression and propensity scores
  - Commonly used parametric methods
  - But, these methods do not work in general
- These methods are commonly used but have limited applicability for complex longitudinal data.

Outcome regression

- Outcome regression is widely used for purely predictive purposes
  - It is all about association, not causation
- Present outcome regression as a method to estimate the parameters of structural models for causal inference
  - Under exchangeability, positivity, and well-defined interventions.

Outcome regression

- The structural model

$$E[Y^{a,c=0}|L] = \beta_0 + \beta_1 a + \beta_2 aL + \beta_3 L$$

- Outcome regression

$$E[Y|A, C=0, L] = \alpha_0 + \alpha_1 A + \alpha_2 AL + \alpha_3 L$$

- Under exchangeability, positivity, and well-defined interventions, both means are equal
- That is, $\alpha = \beta$
- (L are sufficient to adjust for confounding / outcome model is correctly specified)

Propensity scores

- Propensity scores

$$\pi(L) = Pr[A = 1|L]$$

- IP weighting and g-estimation, we already estimated the propensity scores.
- In an ideal randomized trial, $\pi(L) = 0.5$
- If $\pi(L)$ is unknown, needs to be estimated from the data.
- Generally, fit a logistic model for the probability of A conditional on the L

Propensity scores as balancing scores

- Propensity score balances the covariates between the treated and the untreated

$$A \perp\!\!\!\perp L \mid \pi(L)$$

- That is, the distribution of L will be the same in the treated and the untreated.
- Propensity scores only balances the measured covariates.
- Randomization balances both the measured and the unmeasured covariates

Causal inference using propensity score

- Causal inference using propensity score requires exchangeablity, positivity, and consistency, too.
- $Y^a \perp\!\!\!\perp A \mid L$, exchangeability within levels of covariates L implies

$$Y^a \perp\!\!\!\perp A \mid \pi(L)$$

- Rubin(1983) proved that exchangeability and positivity based on the variables L implies exchangeability and positivity based on a balancing scores.
- Propensity scores can be used to estimate causal effects using stratification, standardization and matching.

Propensity stratification and standardization

- Under identifying conditions,

$$E[Y^{a=1,c=0}|\pi(L) = s] - E[Y^{a=0,c=0}|\pi(L) = s] =$$

$$E[Y|A = 1, C = 0, \pi(L) = s] - E[Y|A = 0, C = 0, \pi(L) = s]$$

- The conditional effect might be estimated by restricting the analysis to individiuals with the value of s of the propensity score
- The propensity score $\pi(L)$ is generally a continuous variable
  - Create strata that contain individuals with similar value of $\pi(L)$
  - The deciles of the estimated $\pi(L)$ is popular choice
  - Outcome regression with covariates are treatment A and 9 indicators for the deciles.

Matching

- Form a matched population in which the treated and the untreated are exchangeable because they have the same distribution of $\pi(L)$.
- Under exchangeability and positivity given $\pi(L)$, the associational measures in the matched population are consistent estimates of effect measures.
- Match the treated and the untreated individuals with a close value of $\pi(L)$
- s with $s \pm 0.05$

Matching

- Defining closeness in propensity matching entails a bias-variance trade-off
  - If the closenss criteria are too loose, the distribution of $\pi(L)$ will differ. (Exchangeability will not hold)
  - If the criteria are too tight, many individuals are excluded by the matching population and the effect estimate have wider CI. (Exchangeability holds)