Adjusting for confounders

Tyler Tuan

Without the needs for intervention, we can infer causality by adjusting for confounders.

Propensity-based methods

- Assignment Mechanism
  - Randomized (intervention)
  - Propensity (observation)
- Propensity grouping
- IP-weighting
- G-formula

Assignment Mechanism

1. Intervention
   a. Randomized trials
   b. Independent from confounders
2. Observation
   a. Propensity
   b. Depend on confounders

Data of choice:

1) Set of treatments A
2) Set of responses Y
3) Find all possible causal relationships

$$E[Y^{A=a}]$$

Propensity grouping

- Treatment is independent from confounders given propensity score $\mathrm{e(L) = Pr(A = 1|L)}$
- By stratifying on propensity score, we are approximating a randomized trial within the subpopulation
- $A \perp L \mid e(L)$, confounders are adjusted within the strata of propensity
- $E[Y^{A=a}] = \mathrm{E[Y|A = a, e(L)]}$
- Problem: Continuous treatments... Move to IP-weighting and G-formula, which use the propensity score as numerical value instead of stratification

IP-weighting

- Create pseudo-population, which adjusted for confounders, through the use of numerical value of propensity score instead of stratification
- $E[Y^{A=a}] = E[\frac{I(A=a)Y}{f(A=a|L)}]$
- Mean of Y reweighted by the IP-weight $W^A = {}^1/_{f(A|L)}$
- Fit $E[Y|A=a] = \theta_0 + \theta_1 a$ with weighted linear regression with weight being IP weight
- Such model is estimation for $\hat{E}[Y^{A=a}]$
- Weighted non-linear for $E[Y|A=a]$?

G-formula

- Standardization: Counterfactual distribution given confounders is the same as response distribution given confounders and treatment
- $E[Y^{A=a}]$

$$= \sum_{l} E[Y^{A=a}|L = l]\Pr[L = l]$$

$$= \sum_{l} E[Y|A = a, L = l]\Pr[L = l]$$

- The real work here is to estimate $\hat{E}[Y|A = a, L = l]$ through a ML model of choice
- $\hat{E}[Y^{A=a}] = \frac{1}{N}\sum_i \hat{E}[Y|A = a, L_i]$

How Machine Learning is incorporated?

- Estimate Propensity Score: $\Pr[A = a|L]$
- Estimate IP-weighted response: $E[Y|A = a]$ along with IP weights info
- Estimate G-formula: $\hat{E}[Y|A = a, L = l]$
- Note that, IP-weighting requires two ML models while G-formula requires only one
- Does overfitting and underfitting matter? It might be true that overfitting is good...