# Topics in Causal Inference

STAT41530

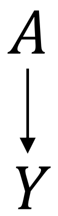
Jingshu Wang

# Lecture 6

Topic:
Estimation and statistical inference

- Randomization inference
- Point estimation for observational data
- Statistical inference
  - Bootstrap
    - Regular and asymptotically linear (RAL) estimator

# Estimation for Completely randomized experiment




$A \perp Y(a)$ for all $a$

Joint distribution $(Y(0), Y(1))$ is unidentifiable

- Average treatment effect: $\tau = E[Y(1)] - E[Y(0)]$
- Point estimator for $\tau$ :

$$\hat{\tau} = \frac{1}{n_1}\sum_{i=1}^{n} Y_i 1_{A_i=1} - \frac{1}{n_0}\sum_{i=1}^{n} Y_i 1_{A_i=0}$$

- Statistical inference for $\tau$
  - Assume that $(Y_i(0), Y_i(1), A_i)$ are i.i.d across $i$
  - Randomization inference: perform statistical test without the i.i.d. assumption
    - View all potential outcomes $\{Y_i(0), Y_i(1)\}_{i=1}^{n}$ as fixed constants
    - Randomness in data comes solely from random treatment assignment
    - $Y_i = A_i Y_i(1) + (1 - A_i) Y_i(0)$: random and is either $Y_i(0)$ or $Y_i(1)$

# Fisher randomization test

- Test for sharp null hypothesis: $H_0$: $Y_i(0) = Y_i(1)$ for $i = 1, 2, \cdots, n$
- All potential outcomes are known under $H_0$: $Y_i(0) = Y_i(1) = Y_i$ → fixed
- Distribution of $\{A_i\}$ is known → distribution of $\hat{\tau} = \frac{1}{n_1}\sum_{i=1}^{n} Y_i 1_{A_i=1} - \frac{1}{n_0}\sum_{i=1}^{n} Y_i 1_{A_i=0}$ is known under $H_0$
- Procedure:
  - Randomly draw treatment assignments $\{A_i^{(b)}\}$ for $B$ times
  - Each time compute the corresponding observed outcomes $Y_i^{(b)} = A_i^{(b)} Y_i(1) + (1 - A_i^{(b)}) Y_i(0)$ and test statistics $\hat{\tau}^{(b)} = \frac{1}{n_1}\sum_{i=1}^{n} Y_i^{(b)} 1_{A_i^{(b)}=1} - \frac{1}{n_0}\sum_{i=1}^{n} Y_i^{(b)} 1_{A_i^{(b)}=0}$
  - $\{\hat{\tau}^{(b)}, b = 1, \cdots, B\}$ form the null distribution of $\hat{\tau}$, and we compute p-value by comparing the observed $\hat{\tau}$ with its null distribution
- The idea work for both completely randomized experiment / conditional randomized experiment

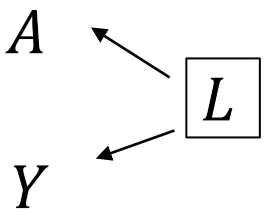
# Neyman repeated sampling inference

- Provide a conservative CI for $\tau$ with randomization inference
- Variance of $\hat{\tau}$ satisfy

$$V = Var(\hat{\tau}) = \frac{1}{n_1} S_1^2 + \frac{1}{n_0} S_0^2 - \frac{1}{n} S_\tau^2$$

- $S_a^2 = \frac{\sum_{i=1}^{n} [Y_i(a) - \bar{Y}(a)]^2}{n-1}, a = 0,1; S_\tau^2 = \frac{\sum_{i=1}^{n} [\tau_i - \tau]^2}{n-1}$ unknown fixed parameters
    - Sample variances of $Y_i$ for the treatment / control group ($s_a^2$) provides unbiased estimates of $S_a^2$
    - $S_\tau^2$ is not identifiable
- A conservative estimate of $Var(\hat{\tau})$: $\hat{V} = \frac{1}{n_1} s_1^2 + \frac{1}{n_2} s_2^2$
- Finite sample distribution of $\hat{\tau}$ is complicated
- Asymptotic normality: under proper assumptions $\sqrt{n}(\hat{\tau} - \tau) \rightarrow N(0, nV)$

# Estimation in observational studies




$$A \perp Y(a)\ |L \ \text{ for all } a$$

- IPW
- Standardization (outcome regression)
- Doubly robust estimator
- Matching

# Inverse probability weighting (IPW) estimator

- $\mathbb{E}\left[Y(a)\right] = \mathbb{E}\left[\frac{Y1_{A=a}}{P(A=a \mid L)}\right]$
- Weights create a "pseudo-population" where covariates between two groups are balanced:

$$E\left[\frac{AL}{e(L)}\right] = E\left[\frac{(1-A)L}{1-e(L)}\right]$$

We should check for covariance balancing after weighting to evaluate the estimate of $e(L)$

- IPW estimator

$$\widehat{\tau}_{IPW} = \frac{1}{n}\sum_i \frac{Y_i 1_{A_i=1}}{\widehat{e}(L_i)} - \frac{1}{n}\sum_i \frac{Y_i 1_{A_i=0}}{1-\widehat{e}(L_i)}$$

- We can estimate $e(L)$ by logistic regression
- IPW with normalized weights (Fact: $E\left[\frac{1_{A=1}}{e(L)}\right] = 1$):

$$\widehat{\tau}_{IPW,2} = \frac{\sum_i \frac{Y_i 1_{A_i=1}}{\widehat{e}(L_i)}}{\sum_i \frac{1_{A_i=1}}{\widehat{e}(L_i)}} - \frac{\sum_i \frac{Y_i 1_{A_i=0}}{1-\widehat{e}(L_i)}}{\sum_i \frac{1_{A_i=0}}{1-\widehat{e}(L_i)}}$$

Typically reduce variance and lead to more stable estimates (Hirano, Imbens, Ridder 2003 Ecnometrica)

# Standardization (outcome regression) estimator

- Put a model for the conditional expectation
- Linear model: $\mu_a(L) = E[Y \mid A = a, L] = \beta_0 + \beta_1 a + \beta_2 L$
- This is essentially a model on the potential outcomes:

$$E[Y(a) \mid L] = E[Y \mid A = a, L] = \beta_0 + \beta_1 a + \beta_2 L$$

- Estimator

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_1(L_i) - \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_0(L_i)$$

What if we incorrectly specify the parametric models for $e(L)$ or $\mu_a(L)$?
$[e(L) = e_1(L) = 1 - e_0(L)]$

# Doubly robust estimator estimator

- Let

$$f(a, L, Y) = \frac{Y 1_{A=a}}{\tilde{e}_a(L)} - \frac{1_{A=a} - \tilde{e}_a(L)}{\tilde{e}_a(L)} \tilde{\mu}_a(L)$$

- If we correctly specify the propensity score model, then $\tilde{e}_a(L) = e_a(L)$
- If we correctly specify the outcome model, then $\tilde{\mu}_a(L) = \mu_a(L)$

$$\begin{aligned} \mathbb{E}\left[f(a, L, Y) \mid L\right] &= \frac{\mathbb{E}\left[Y \mid A = a, L\right] \mathbb{P}\left[A = a \mid L\right] - \left(\mathbb{P}\left[A = a \mid L\right] - \tilde{e}_a(L)\right) \tilde{\mu}_a(L)}{\tilde{e}_a(L)} \\ &= \frac{\mu_a(L) e_a(L) - e_a(L) \tilde{\mu}_a(L) + \tilde{e}_a(L) \tilde{\mu}_a(L)}{\tilde{e}_a(L)} \\ &= \frac{\left(\mu_a(L) - \tilde{\mu}_a(L)\right) \left(e_a(L) -_a (L)\right)}{\tilde{e}_a(L)} + \mu_a(L) \end{aligned}$$

- Doubly robust property: if either model is correct, we have $E[Y(a) | L] = E[f(a, L, Y) | L]$
- Estimator:

$$\widehat{\tau} = \frac{1}{n} \sum_i \left[\frac{Y_i 1_{A_i=1}}{\widehat{e}(L_i)} - \frac{1_{A_i=1} - \widehat{e}(L_i)}{\widehat{e}(L)} \widehat{\mu}_1(L_i)\right] - \frac{1}{n} \sum_i \left[\frac{Y_i 1_{A_i=0}}{1 - \widehat{e}(L_i)} - \frac{1_{A_i=0} - (1 - \widehat{e}(L_i))}{1 - \widehat{e}(L)} \widehat{\mu}_0(L_i)\right]$$

# Matching estimator

- $J_i$: $M_i$ closest units to unit j under alternative treatment
- Define

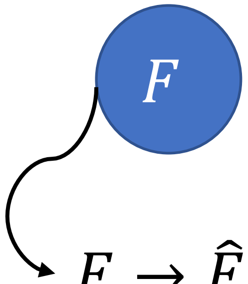
$$\hat{Y}_i(1) = \begin{cases} \frac{1}{M_i}\sum_{j \in J_i} Y_j & \text{if } A_i = 0 \\ Y_i & \text{if } A_i = 1 \end{cases}, \qquad \hat{Y}_i(0) = \begin{cases} Y_i & \text{if } A_i = 0 \\ \frac{1}{M_i}\sum_{j \in J_i} Y_j & \text{if } A_i = 1 \end{cases}$$

- Matching estimator:

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i(1) - \frac{1}{n}\sum_{i=1}^{n} \hat{Y}_i(0)$$

- Reference review paper by Stuart (2008, Stat Sci)
- R package: Matching, Matchit

# Bootstrap




| | Parameter | Estimator | |
|---|---|---|---|
| | $t(F)$ | $\hat{t}$ | |
| $F \to \hat{F}$ | $t(\hat{F})$ | $\hat{t}_1^*, \cdots, \hat{t}_B^*$ | ← B bootstrap samples |

- Use bootstrap samples to approximate both bias($\hat{t}$) and var($\hat{t}$)

- Bootstrap by sampling with replacement can be used for statistical inference of IPW, standardization and doubly robust estimators
- A more complicated bootstrap is needed for matching estimator (Otsu and Rai, 2017 JASA)
- Need large $n$