Lecture 6

类似于 CRE，在 observational study 中，除了构造 estimator（IPW）来估计和检验 ATE τ，还可以通过 model-based estimation procedure 来估计 ATE.

§1 Observational study 下计算 ATE 的 regression method

1. Definition: Observational study 下计算 ATE 的 model

① 考虑以下 model：

$$Y_i(0) = X_i^T\beta_0 + \varepsilon_i(0) \quad (*) \quad \Longleftrightarrow \quad W_iY_i = (W_iX_i^T)\beta_0 + W_i\varepsilon_i(0)$$

$$Y_i(1) = X_i^T\beta_1 + \varepsilon_i(1) \quad (\#) \quad \Longleftrightarrow \quad (1-W_i)Y_i = ((1-W_i)X_i^T)\beta_1 + (1-W_i)\varepsilon_i(1)$$

$$P(W_i=1|X_i) = p(X_i)$$

其中 $\varepsilon_i(0), \varepsilon_i(1)$ 与 X_i, W_i independent

② 若 unconfoundedness condition 成立，则 ATE 可被表示为：

$$\begin{aligned}\tau &= E[Y_i(1)] - E[Y_i(0)] \\ &= E[X_i]^T\cdot(\beta_1-\beta_0) \\ &= \mu_X^T(\beta_1-\beta_0)\end{aligned}$$

注：此处采用的是 linear model，现实中可以采用更复杂的 models（如 non-parametric regression）：

$$E[Y_i(w)|X_i] = g_w(X_i) \quad \text{for } w=0 \text{ 或 } 1$$

2. Definition: model-based estimation

① μ_X 可以使用 sample mean 估计：

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^n X_i$$

② β_0, β_1 可以分别对 $(*)$，$(\#)$ 进行 linear regression 得到 LSE：

$$\hat{\beta}_1 = \left(\sum_{i=1}^n W_iX_iX_i^T\right)^{-1}\cdot\sum_{i=1}^n W_iX_iY_i \qquad (\text{利用 } W_i^2=W_i)$$

$$\hat{\beta}_0 = \left(\sum_{i=1}^n(1-W_i)X_iX_i^T\right)^{-1}\cdot\sum_{i=1}^n(1-W_i)X_iY_i \qquad (\text{利用 } (1-W_i)^2=1-W_i)$$

③ τ 可以被估计为

$$\hat{\tau} = \hat{\mu}_X^T(\hat{\beta}_1-\hat{\beta}_0)$$

3. Theorem: $\hat{\beta}_1$ 和 $\hat{\beta}_0$ 的 unbiasedness

$\hat{\beta}_1$ 和 $\hat{\beta}_0$ 可被改写为：

① $\hat{\beta}_1 = \beta_1 + \left(\sum_{i=1}^n W_iX_iX_i^T\right)^{-1}\cdot\sum_{i=1}^n W_iX_i\varepsilon_i(1)$

② $\hat{\beta}_0 = \beta_0 + \left(\sum_{i=1}^n(1-W_i)X_iX_i^T\right)^{-1}\cdot\sum_{i=1}^n(1-W_i)X_i\varepsilon_i(0)$

且有

① $E[\hat{\beta}_1|X] = \beta_1$ （若默认 fixed design，则无需 condition on X）

② $E[\hat{\beta}_0|X] = \beta_0$

证明: β_1 相关的结论

① $$\begin{aligned}\sum_{i=1}^n W_iX_iY_i &= \sum_{i=1}^n W_iX_iY_i(1) \\ &= \sum_{i=1}^n W_iX_i(X_i^T\beta_1 + \varepsilon_i(1))\end{aligned}$$

$$= (\sum_{i=1}^n W_i X_i X_i^T) \beta_1 + \sum_{i=1}^n W_i X_i \varepsilon_i(1)$$

$$\Rightarrow \hat{\beta}_1 = (\sum_{i=1}^n W_i X_i X_i^T)^{-1} \cdot \sum_{i=1}^n W_i X_i Y_i$$

$$= \beta_1 + (\sum_{i=1}^n W_i X_i X_i^T)^{-1} \cdot \sum_{i=1}^n W_i X_i \varepsilon_i(1)$$

② $E[(\sum_{i=1}^n W_i X_i X_i^T)^{-1} \cdot \sum_{i=1}^n W_i X_i \varepsilon_i(1) \mid X]$

$$= \sum_{i=1}^n E[(\sum_{i=1}^n W_i X_i X_i^T)^{-1} W_i X_i \mid X_i] \cdot \underbrace{E[\varepsilon_i(1)]}_{=0} \quad (\varepsilon_i(1) \perp W_i, X_i)$$

$$= 0$$

$$\Rightarrow E[\hat{\beta}_1 \mid X] = \beta_1$$

4. **Theorem**: $\hat{\beta}_1$, $\hat{\beta}_0$ 和 $\hat{\tau}$ 的 consistency

若 model assumption 成立

则 ① $\hat{\beta}_1 \xrightarrow{P} \beta_1$

② $\hat{\beta}_0 \xrightarrow{P} \beta_0$

③ $\hat{\tau} \xrightarrow{P} \tau$

现在我们有两种估计 ATE τ 的方法：

① nonparametric IPW estimator：要求 $p(\cdot)$ 的估计 consistent

② parametric model-based estimation：要求 model assumption 成立

我们希望结合两种方法的优势.

§2 Doubly robust estimation

1. **Definition**: Doubly robust estimation

Doubly robust estimator 的形式为：

$$\begin{cases} \hat{\mu}_1^{+} = \hat{\mu}_X \hat{\beta}_1 + \frac{1}{n} \sum_{i=1}^n \frac{W_i}{\hat{p}(X_i)} (Y_i - X_i^T \hat{\beta}_1) \\ \hat{\mu}_0^{+} = \underbrace{\hat{\mu}_X \hat{\beta}_0}_{\text{linear model}} + \underbrace{\frac{1}{n} \sum_{i=1}^n \frac{1-W_i}{1-\hat{p}(X_i)} (Y_i - X_i^T \hat{\beta}_0)}_{\text{IPW for model residual}} \end{cases}$$

$$\Rightarrow \hat{\tau}^{+} = \hat{\mu}_1^{+} - \hat{\mu}_0^{+}$$

注：① linear model 也可以被替换为 non-linear model

② 若 $\hat{p} \approx p$，则无论 model assumption 是否成立，estimation bias 都很小：

$$\frac{1}{n} \sum_{i=1}^n \frac{W_i}{\hat{p}(X_i)} (Y_i - X_i^T \hat{\beta}_1)$$

$$\approx \frac{1}{n} \sum_{i=1}^n \frac{W_i}{p(X_i)} (Y_i - X_i^T \beta_1) \quad (\hat{p}(X_i) \approx p(X_i),\ \hat{\beta}_1 \approx \beta_1)$$

$$\approx E[\frac{W_i}{p(X_i)} (Y_i - X_i^T \beta_1)]$$

$$= E[E[\frac{W_i}{p(X_i)} (Y_i(1) - X_i^T \beta_1) \mid X]]$$

$$= E[Y_i(1)] - \mu_X^T \beta_1$$

$$\Rightarrow \hat{\mu}_1^{+} \approx \hat{\mu}_X \hat{\beta}_1 + E[Y_i(1)] - \mu_X^T \beta_1 \approx E[Y_i(1)]$$

③ 若 model assumption 成立，则无论 $\hat{p}$ 是否 $\approx p$，estimation bias 都很小

$$\frac{1}{n} \sum_{i=1}^n \frac{W_i}{\hat{p}(X_i)} (Y_i - X_i^T \hat{\beta}_1)$$

$$\approx \frac{1}{n}\sum_{i=1}^{n}\frac{W_i}{p(X_i)}(Y_i(1)-X_i^T\beta_1)$$

$$\approx \frac{1}{n}\sum_{i=1}^{n}\frac{W_i}{p(X_i)}\varepsilon_i(1)$$

$$\approx 0$$

$$\Rightarrow \hat{\mu}_1^{+} \approx \hat{\mu}_X\hat{\beta}_1 + 0 \approx Y_i(1) \quad \text{(model assumption)}$$

接下来我们希望研究 $\hat{\tau}^{+}$ 的 stochastic property，但由于 $\hat{\tau}^{+}$ 中提供 randomness 的变量太多，我们先研究 "oracle double-debiased estimator" 的 stochastic property

2. **Definition:** oracle double-debiased estimator

Oracle double-debiased estimator 为：

$$\begin{cases} \hat{\mu}_1^{*} = \hat{\mu}_X\beta_1 + \frac{1}{n}\sum_{i=1}^{n}\frac{W_i}{p(X_i)}(Y_i-X_i^T\beta_1) \\ \hat{\mu}_0^{*} = \hat{\mu}_X\beta_0 + \frac{1}{n}\sum_{i=1}^{n}\frac{1-W_i}{1-p(X_i)}(Y_i-X_i^T\beta_0) \end{cases}$$

(把 $\hat{\beta}_0, \hat{\beta}_1, \hat{p}(X_i)$ 替换为 underlying true values $\beta_0, \beta_1, p(X_i)$)

$$\Rightarrow \hat{\tau}^{*} = \hat{\mu}_1^{*} - \hat{\mu}_0^{*}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^T(\beta_1-\beta_0) + \frac{W_i}{p(X_i)}(Y_i(1)-X_i^T\beta_1) - \frac{1-W_i}{1-p(X_i)}(Y_i(0)-X_i^T\beta_0)\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\underbrace{\left\{\left(1-\frac{W_i}{p(X_i)}\right)X_i^T\beta_1 - \left(1-\frac{1-W_i}{1-p(X_i)}\right)X_i^T\beta_0 + \frac{W_iY_i(1)}{p(X_i)} - \frac{(1-W_i)Y_i(0)}{1-p(X_i)}\right\}}_{:=Z}$$

3. **Definition:** oracle double-debiased estimator 的 variance 的 estimator

可以使用 method of moment 得到 oracle double-debiased estimator 的 variance 的 estimator：

$$n\cdot Var(\hat{\tau}^{*}) = Var(Z)$$

$$= E[Z^2] - E[Z]^2$$

$$\Rightarrow \hat{\Upsilon}^{*2} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left(1-\frac{W_i}{p(X_i)}\right)X_i^T\beta_1 - \left(1-\frac{1-W_i}{1-p(X_i)}\right)X_i^T\beta_0 + \frac{W_iY_i(1)}{p(X_i)} - \frac{(1-W_i)Y_i(0)}{1-p(X_i)}\right\}^2 - \hat{\tau}^{*2}$$

是 $n\cdot Var(\hat{\tau}^{*})$ 的 estimator

4. 使用 doubly robust estimator 对 ATE 进行 test

令 hypothesis 为：

$$H_0: \tau = 0 \quad v.s. \quad H_1: \tau > 0$$

则 ① 若 oracle double-debiased estimator 可以求出，则 test statistics 和 p-value 为：

$$\hat{L}^{*} = \frac{\sqrt{n}\,\hat{\tau}^{*}}{\hat{\Upsilon}^{*}} \quad 和 \quad \hat{p}^{*} = 1-\Phi(\hat{L}^{*})$$

② 若 oracle double-debiased estimator 不可求出，则先用 $\hat{\beta}_0, \hat{\beta}_1, \hat{p}(X_i)$ 估计 $\hat{\Upsilon}^{*2}$：

$$\hat{\Upsilon}^{+2} = \frac{1}{n}\sum_{i=1}^{n}\left\{\left(1-\frac{W_i}{\hat{p}(X_i)}\right)X_i^T\hat{\beta}_1 - \left(1-\frac{1-W_i}{1-\hat{p}(X_i)}\right)X_i^T\hat{\beta}_0 + \frac{W_iY_i}{\hat{p}(X_i)} - \frac{(1-W_i)Y_i}{1-\hat{p}(X_i)}\right\}^2 - \hat{\tau}^{+2}$$

test statistics 和 p-value 为：

$$\hat{L}^{+} = \frac{\sqrt{n}\,\hat{\tau}^{+}}{\hat{\Upsilon}^{+}} \quad 和 \quad \hat{p}^{+} = 1-\Phi(\hat{L}^{+})$$