观察性研究中的因果推断

observational study

Chapter 10

1. 目的：因果作用

2. 特点：无随机性

Cochran (1965)

Part III

"理想化"

依赖于协变量 X

假定 —— 可忽略性 / 无混杂

$Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X$ —— 和 SRE 一样

本质不同：① 不可验证的假定

某种意义，只是因果推断的一个充分条件

Part IV

"不理想化"

② $\Pr(Z=1 \mid Y(1), Y(0), X)$

$= \Pr(Z=1 \mid X)$

$= e(X)$ 倾向得分

propensity score

即使相信可忽略，也不知道倾向得分.

假定: 可忽略性 / 无混杂

强可忽略性: $Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X$

$\Downarrow$

可忽略性: $\left\{\begin{array}{l} Z \perp\!\!\!\perp Y(1) \mid X \\ Z \perp\!\!\!\perp Y(0) \mid X \end{array}\right\}$ 弱了

$= SRE$

直观上　无混杂　Confounding　有混杂

$$\begin{cases} Y(1) = f_1(X, V_1) \\ Y(0) = f_0(X, V_0) \\ Z = 1\big(g(X, V) \geq 0\big) \\ (V_1, V_0) \perp\!\!\!\perp V \end{cases} \qquad \begin{cases} Y(1) = f_1(X, U, V_1) \\ Y(0) = f_0(X, U, V_0) \\ Z = 1\big(g(X, U, V) \geq 0\big) \\ (V_1, V_0) \perp\!\!\!\perp V \end{cases}$$

$$\Rightarrow Z \perp\!\!\!\perp \big(Y(1), Y(0)\big) \mid X \qquad \Rightarrow Z \not\!\perp\!\!\!\perp \big(Y(1), Y(0)\big) \mid X$$

定理10.1　如果 $Z \perp\!\!\!\perp \big(Y(1), Y(0)\big) \mid X$

那么 $\tau = E\big(Y(1) - Y(0)\big)$ 可识别.

来自计量经济学 —— identifiable (Bickel & Doksum's book)

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

X 离散 $= \sum_{k=1}^{K} \Pr(X=k)\left[E(Y|Z=1, X=k) - E(Y|Z=0, X=k)\right]$

standardization

$=$ SRE 公式一样

一般 $= \int \left[E(Y|Z=1, X=x) - E(Y|Z=0, X=x)\right] F(dx)$

Naive: $E(Y|Z=1) - E(Y|Z=0)$

X 离散 $= \sum_{k=1}^{K} \Pr(X=k|Z=1)\, \underline{E(Y|Z=1, X=k)}$

$- \sum_{k=1}^{K} \Pr(X=k|Z=0)\, \underline{E(Y|Z=0, X=k)}$

一样

证明: $\tau = E(Y(1) - Y(0))$

tower property $= E\left[E(Y(1) - Y(0)|X)\right]$

$= E\left[E(Y(1)|X) - E(Y(0)|X)\right]$

可忽略性 $= E\left[E(Y\cancel{(1)} \mid Z=1, X) - E(Y\cancel{(0)} \mid Z=0, X) \right]$

g-formula $= E\left[E(Y \mid Z=1, X) - E(Y \mid Z=0, X) \right]$ $\square$

估计 τ:

$$\mu_1(X) = E(Y \mid Z=1, X)$$

$$\mu_0(X) = E(Y \mid Z=0, X)$$

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i) \right]$$

e.g. $E(Y \mid Z, X) = \beta_0 + \beta_z Z + \beta_x^T X$

模型对

$\Rightarrow \tau = \beta_z$

e.g. $E(Y \mid Z, X) = \beta_0 + \beta_z Z + \beta_x^T X + \beta_{zx}^T X Z$

$$\Rightarrow \tau = \beta_Z + \beta_{Zx}^T E(X)$$

零中心化 X_i's

$\Rightarrow \hat{\beta}_Z$ 即可

e.g. Y 二值

$$Pr(Y=1 \mid Z, X) = \frac{e^{\beta_0 + \beta_Z Z + \beta_x^T X}}{1 + e^{\beta_0 + \beta_0 Z + \beta_x^T X}}$$

β_Z = Conditional odds ratio of Z on Y

条件优势比

$$\Rightarrow \hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{e^{\hat{\beta}_0 + \hat{\beta}_Z + \hat{\beta}_x^T X_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_Z + \hat{\beta}_x^T X_i}} - \frac{e^{\hat{\beta}_0 + \hat{\beta}_x^T X_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_x^T X_i}} \right)$$

APE: Average Partial Effect } 计算

AME: Average marginal effect } 结构

Chapter 11

$$\Pr(X, Z, Y(1), Y(0))$$

$$= \Pr(X)$$ → 协变量：背景

$$\cdot \Pr(Y(1), Y(0) \mid X)$$ → 结果模型

$$\cdot \Pr(Z \mid Y(1), Y(0), X)$$ → 唯一可思考的

定义 $e(Y(1), Y(0), X) = \Pr(Z=1 \mid Y(1), Y(0), X)$

为倾向得分 (propensity score)

强可忽略性: $e(X) = \Pr(Z=1 \mid X)$

Rosenbaum & Rubin (1983 Biometrika)

定理11.1 若 $Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X$

那么 $Z \perp\!\!\!\perp (Y(1), Y(0)) \mid e(X)$.

（降维的工具）

证明: 目的要证,

$$\Pr(Z=1 \mid Y(1), Y(0), e(X)) = \Pr(Z=1 \mid e(X))$$

$$\text{左边} = E\left(Z \mid Y(1), Y(0), e(X)\right)$$

$$= E\left[\underbrace{E(Z \mid Y(1), Y(0), \cancel{e(X)}, X)}_{E(Z \mid X) \,=\, \Pr(Z=1 \mid X) = e(X)} \,\middle|\, Y(1), Y(0), e(X)\right]$$

可忽略性

$$= E\left[e(X) \mid Y(1), Y(0), e(X)\right] = e(X).$$

右边 $= E\left(Z \mid e(X)\right)$

$$= E\left[\underbrace{E\left(Z \mid \cancel{e(X)}, X\right)}_{= e(X)} \,\middle|\, e(X)\right]$$

$$= e(X) \qquad \square$$

倾向得分分层估计:

① 估计 $e(X)$: e.g. $glm(Z \sim X)$

$\Longrightarrow \hat{e}(X)$

② 离散化 $\hat{e}(X) \Longrightarrow \hat{e}'(X) \in \{1 \cdots K\}$, $K = 5$

③ $SRE\left[Y \sim Z \mid \hat{e}'(X)\right]$

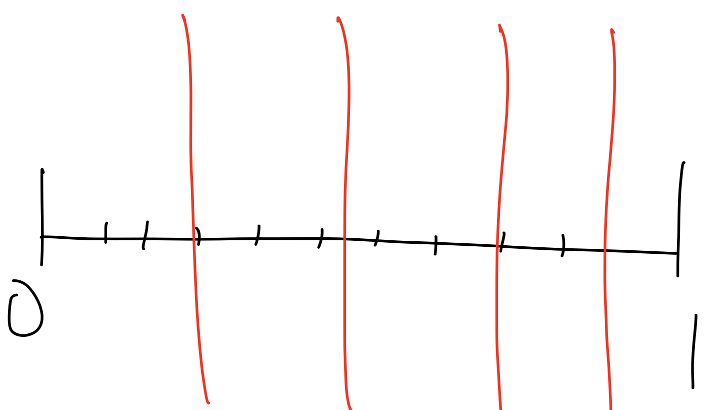
$\Longrightarrow \hat{\tau}_S$ 分层估计

问题：

① $\hat{e}(x)$ 模型对与否？

一般地，$\hat{e}(x) \not\to e(x)$：有偏估计

稳健性



只依赖于 $\hat{e}(x)$ 的排序

② K的选择

$K=5$ Cochran 推荐

$n\to\infty$，应该 $K\to\infty$

Wang et al. (2020)：贪婪算法

$K \nearrow$ 直到 $\hat{C}_s$，$\hat{V}_s$ 无法计算

② 方差估计

分层 $\Longrightarrow$ 假想为 SRE

忽略了 $\hat{e}(X)$ 的随机性

$\because$ 估计倾向得分比真实倾向得分好

Su et al (2023)

$\Longrightarrow$ 高估方差

一个方案： bootstrap 整个估计量

<u>定理 11.2</u> 若 $Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X$

那么 IPW

$$\begin{cases} E(Y(1)) = E\left(\dfrac{ZY}{e(X)}\right) \\ E(Y(0)) = E\left(\dfrac{(1-Z)Y}{1-e(X)}\right) \end{cases}$$

$$\Rightarrow \tau = E\big(Y(1) - Y(0)\big)$$

$$= E\left(\frac{ZY}{e(X)}\right) - E\left(\frac{(1-Z)Y}{1-e(X)}\right)$$

估计量: $\hat{\tau}^{ht} = \frac{1}{n}\sum_{i=1}^{n} \frac{Z_i Y_i}{\hat{e}(X_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{(1-Z_i)Y_i}{1-\hat{e}(X_i)}$

Horvitz-Thompson

$$\hat{\tau}^{haj} = \frac{\sum_{i=1}^{n} \frac{Z_i Y_i}{\hat{e}(X_i)}}{\sum_{i=1}^{n} \frac{Z_i}{\hat{e}(X_i)}} - \frac{\sum_{i=1}^{n} \frac{(1-Z_i)Y_i}{1-\hat{e}(X_i)}}{\sum_{i=1}^{n} \frac{1-Z_i}{1-\hat{e}(X_i)}}$$

Hájek

本质问题: $Y_i \longrightarrow Y_i + C$

ht 变

haj 不变

证明：只证 $Y(1)$ 部分:

$$E\left(\frac{ZY}{e(X)}\right)$$

$$= E\left(\frac{ZY(1)}{e(X)}\right)$$

tower $$= E\left(E\left(\frac{ZY(1)}{e(X)} \,\middle|\, X\right)\right)$$

$$= E\left(\cancel{\frac{1}{e(X)}}\ \cancel{E(Z|X)}\ E(Y(1)|X)\right)$$

tower $$= E\left(E\left(Y(1)|X\right)\right)$$

$$= E\left(Y(1)\right) \qquad \square$$

① $\hat{e}(x) \longrightarrow e(x)$

倾向得分模型正确

② 条件概念的：$0 < e(x) < 1$

overlap

重叠

"无概论"个体：$Y(1), Y(0)$ 都有

正概率存在！

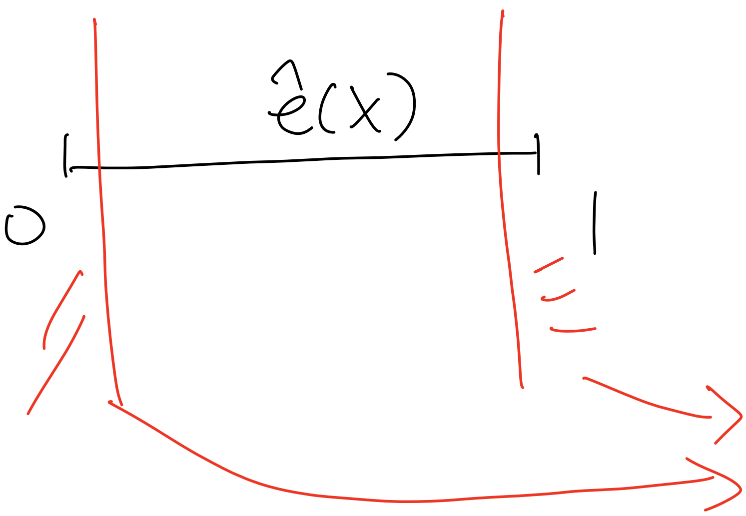
③ 致命问题：IPW 估计

都是极其糟糕的

$$\frac{*}{0} = \infty$$

$$\frac{*}{\varepsilon} = 大$$

估计量不稳定



trimming

丢弃极端 $\hat{e}(x_i)$

[0.1, 0.9]

[0.05, 0.95]

书: trunation

定理 11.3 $Z \perp\!\!\!\perp X \mid e(X)$

结论只涉及 (Z, X)，无 Y.

Ker-Chau Li: SIR 与之有关

可用于评估 $e(x)$ 模型

证明：要证

$$\Pr(Z=1 \mid X, e(X)) = \Pr(Z=1 \mid e(X))$$

左边 $= \Pr(Z=1 \mid X) = e(X)$

右边 $= e(X)$ $\square$

对数据分析的指导：检测协变量分布是否平衡（Section 11.3.2）