Chapter 12 双稳健估计

doubly robust estimation

观察性研究 $\begin{cases} Z \perp\!\!\!\perp (Y(1), Y(0)) \mid X \\ 0 < e(x) < 1 \end{cases}$ 基本假设

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

倾向得分

$$\tau \overset{\text{定义}}{=} E\big(Y(1) - Y(0)\big)$$

$$= E\Big\{ \underbrace{E(Y \mid Z=1, X)}_{\mu_1(X)} - \underbrace{E(Y \mid Z=0, X)}_{\mu_0(X)} \Big\}$$

可识别

identifiable

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

逆概加权

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

regression

$(Z_i=1, Y_i, X_i)$ 回归　　$(Z_i=0, Y_i, X_i)$ 回归

$\xrightarrow{P} \tau$　　$\hat{\mu}_1(\cdot)$ $\hat{\mu}_0(\cdot)$ 模型正确

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

($\hat{\tau}^{hajek}$ 呢)

估计的倾向得分

glm($Z_i \sim X_i$)

$\xrightarrow{P} \tau$　　倾向得分 $\hat{e}(\cdot)$ 对

如何结合 $\hat{\tau}^{reg}$ 和 $\hat{\tau}^{ht}$?

简单的组合 $f(\hat{\tau}^{reg}, \hat{\tau}^{ht})$?

也许可提高精度

但是仍需要求结果和倾向得分模型都要对.

允许模型错误

α, β_1, β_0 参数

一个巧妙的组合:

$$\tilde{\mu}_1^{dr} = E\left\{ \frac{Z\left(Y - \mu_1(X,\beta_1)\right)}{e(X,\alpha)} + \mu_1(X,\beta_1) \right\}$$

$$\tilde{\mu}_0^{dr} = E\left\{ \frac{(1-Z)\left(Y - \mu_0(X,\beta_0)\right)}{1-e(X,\alpha)} + \mu_0(X,\beta_0) \right\}$$

需要构造 $\hat{\mu}_1^{dr}$, $\hat{\mu}_0^{dr}$ 和 $\hat{\tau}^{dr} = \hat{\mu}_1^{dr} - \hat{\mu}_0^{dr}$

必须要 $e(X,\hat{\alpha})$ 和 $\begin{pmatrix} \mu_1(X,\hat{\beta}_1) \\ \mu_0(X,\hat{\beta}_0) \end{pmatrix}$

定理 12.1 假设 $\begin{cases} Z \perp\!\!\!\perp \{Y(1), Y(0)\} \mid X \\ 0 < e(X) < 1 \end{cases}$

1. 若 $\underbrace{e(X,\alpha) = e(X)}_{e\text{ 对}}$ 或 $\underbrace{\mu_1(X,\beta_1) = \mu_1(X)}_{\mu_1\text{ 对}}$

则 $\tilde{\mu}_1^{dr} = E(Y(1))$

2. 若 $e(X,\alpha) = e(X)$ 或 $\mu_0(X,\beta_0) = \mu_0(X)$

则 $\tilde{\mu}_0^{dr} = E(Y(0))$

3. 若 $e(X,\alpha) = e(X)$ 或 $\begin{pmatrix} \mu_1(X,\beta_1) = \mu_1(X) \\ \mu_0(X,\beta_0) = \mu_0(X) \end{pmatrix}$

则 $\tilde{\tau}^{dr} = \tilde{\mu}_1^{dr} - \tilde{\mu}_0^{dr} = \tau$

证明：只证 1.

$$\tilde{\mu}_1^{dr} - E(Y(1))$$

定义

$$= E\left\{ \frac{Z\left(Y(1) - \mu_1(X, \beta_1)\right)}{e(X, \alpha)} + \mu_1(X, \beta_1) \right\} - E\{Y(1)\}$$

结合

$$= E\left\{ \frac{Z\left(Y(1) - \mu_1(X, \beta_1)\right)}{e(X, \alpha)} - \left(Y(1) - \mu_1(X, \beta_1)\right) \right\}$$

结合

$$= E\left\{ \frac{Z - e(X, \alpha)}{e(X, \alpha)} \left(Y(1) - \mu_1(X, \beta_1)\right) \right\}$$

tower property
ignorability

$$= E\left\{ E\left(\frac{Z - e(X, \alpha)}{e(X, \alpha)} \left(Y(1) - \mu_1(X, \beta_1)\right) \Big| X \right) \right\}$$

$\perp\!\!\!\perp$ given X

$$= E\left[E\left(\frac{Z - e(X, \alpha)}{e(X, \alpha)} \Big| X \right) \cdot E\left(Y(1) - \mu_1(X, \beta_1) \Big| X \right) \right]$$

E(·)乘积

$$= E\left\{ \frac{e(X) - e(X,\alpha)}{e(X,\alpha)} \cdot \left(\mu_1(X) - \mu_1(X,\beta_1) \right) \right\}$$

可能是 doubly fragile

$$= 0 \quad \text{若} \quad e(X) = e(X,\alpha)$$

$$\text{或} \quad \mu_1(X) = \mu_1(X,\beta_1) \qquad \square$$

$\hat{\mu}_1^{dr}$, $\hat{\mu}_0^{dr}$ 从何而来?

$$\hat{\mu}_1^{dr} = \frac{1}{n}\sum_{i=1}^{n} \frac{Z_i\left(Y_i - \mu_1(X_i,\hat{\beta}_1)\right)}{e(X_i,\hat{\alpha})} + \frac{1}{n}\sum_{i=1}^{n} \mu_1(X_i,\hat{\beta}_1)$$

$$\hat{\mu}_0^{dr} = \frac{1}{n}\sum_{i=1}^{n} \frac{(1-Z_i)\left(Y_i - \mu_0(X_i,\hat{\beta}_0)\right)}{1-e(X_i,\hat{\alpha})} + \frac{1}{n}\sum_{i=1}^{n} \mu_0(X_i,\hat{\beta}_0)$$

Robins : 半参数理论

估计量本身出现在抽样调查

12.2.1 抽样调查的想法

$$\mu_1 = E\left(\frac{ZY}{e(X)}\right)$$

也可以 $\mu_1 = E(Y(1))$

$$= E\left(Y(1) - \mu_1(X, \beta_1)\right) + E\left(\mu_1(X, \beta_1)\right)$$

$\approx E(Y(1) \mid X)$

但允许错

$$\overset{IPW}{=} E\left(\frac{Z\left(Y - \mu_1(X, \beta_1)\right)}{e(X)}\right) + E\left(\mu_1(X, \beta_1)\right)$$

目标：提高 IPW 的估计精度

特殊性：$e(X)$ 对

$\mu_1(X, \beta_1)$ 可能错

12.2.2 纠偏

$\tilde{\mu}_1 = E(\mu_1(X, \beta_1))$ 可能 $\neq \mu_1$

偏 $= \tilde{\mu}_1 - \mu_1 = E(\mu_1(X, \beta_1) - Y(1))$

$$\xrightarrow[\text{估计}]{\text{IPW}} E\left(\frac{Z(\mu_1(X, \beta_1) - Y)}{e(X, \alpha)} \right)$$

$\Rightarrow \tilde{\mu}_1^{dr} = \tilde{\mu}_1 -$ 偏

Chapter 13 ATT ATC

$$\tau_T = E\left(Y(1) - Y(0) \mid Z=1\right)$$

$$\tau_C = E\left(Y(1) - Y(0) \mid Z=0\right)$$

回顾：无新概念

Chapter 14 倾向得分好用

用 $e(X_i)$ 于回归中

CRE
$$\begin{cases} lm(Y_i \sim 1 + Z_i) \\ lm(Y_i \sim 1 + Z_i + X_i + Z_i \cdot X_i) \end{cases}$$

观察性研究
$$\left\{ lm\left(Y_i \sim 1 + Z_i,\ weights = \begin{cases} \frac{1}{e(X_i)} & Z_i=1 \\ \frac{1}{1-e(X_i)} & Z_i=0 \end{cases}\right)\right.$$

Basu's elephant

HW $\Rightarrow$ $\text{coef}(Z_i) = \hat{\tau}^{\text{hajek}}$

$\text{lm}(Y_i \sim 1 + Z_i + X_i + Z_i \cdot X_i,\ \text{weights} = ($[illegible]$)\ )$

$$\Rightarrow \hat{\tau}^{\text{reg}}_{\text{wls}} = \frac{1}{n}\sum_{i=1}^{n} \mu_1(X_i, \hat{\beta}_1) - \frac{1}{n}\sum_{i=1}^{n} \mu_0(X_i, \hat{\beta}_0)$$

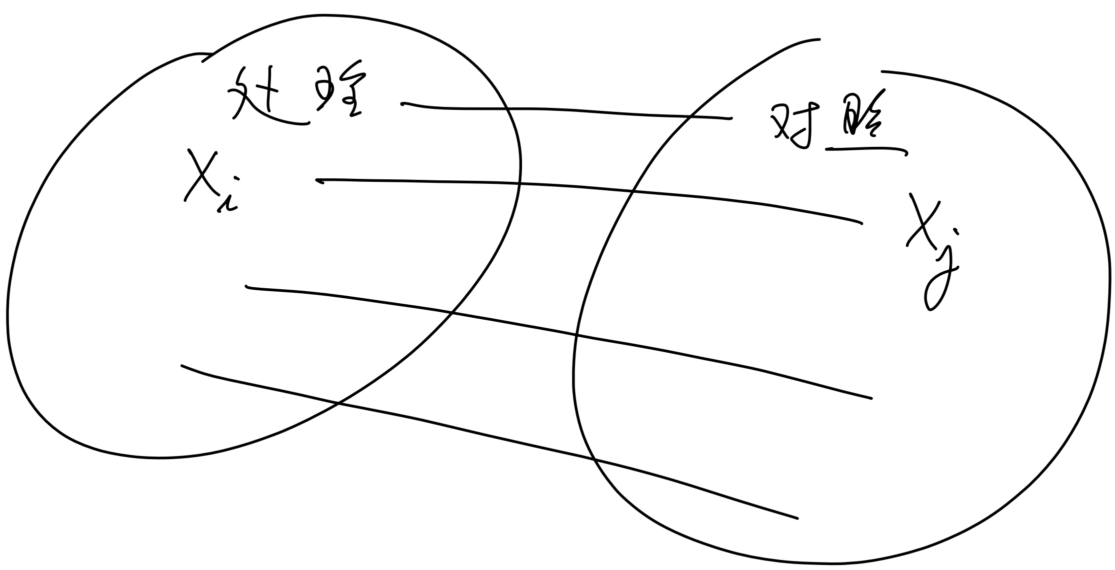
$$\hat{\tau}^{\text{dr}}_{\text{wls}} = \hat{\tau}^{\text{reg}}_{\text{wls}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} \frac{Z_i\left(Y_i - \mu_1(X_i, \hat{\beta}_1)\right)}{\hat{e}(X_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{(1-Z_i)\left(Y_i - \mu_0(X_i, \hat{\beta}_0)\right)}{1-\hat{e}(X_i)}}_{=0}$$

套用 WLS

$= \text{coef of } Z_i$

Chapter 15 匹配之估计

Cochran
Rubin
} 早期研究
甚至没用潜在结果



最简单情况: 1-1 匹配

精确匹配: $X_i = X_{m(i)}$

$\Rightarrow$ MPE

现实中: $X_i \neq X_{m(i)}$

依然用 FRT (假装 MPE)

Rosenbaum 书中推荐

Guo & Rosenhausler (2022 Biometrika)

Abadie - Imbens (AI) 的理论:

1—M 匹配

数据: $(X_i, Z_i, Y_i(1), Y_i(0))_{i=1}^{n}$ iid

对 $Z_i = 1$: $\hat{Y}_i(1) = Y_i$

$$\hat{Y}_i(0) = \frac{1}{M} \sum_{k \in J_i} Y_k$$

?

对照组中与 i 最近的 M 个人

(有放回的匹配)

对 $Z_i=0$, $\hat{Y}_i(1) = \frac{1}{M}\sum_{k\in J_i} Y_k$?

在对照组中与 i 最佳匹配的 M 个人

$\hat{Y}_i(0) = Y_i$

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

构造上看与 $\hat{\tau}^{reg}$ 接近

某种意义：$\hat{\tau}^{m}$ ~ 最近邻回归

nearest neighbor regression

$\Rightarrow$ 非参数理论：$\hat{\tau}^{m}$ 可能有大偏差 (dim(x) 高)

纠偏: $\hat{\tau}^{mbc} = \hat{\tau}^{m} - \hat{B}$,

其中 $\hat{B} = \frac{1}{n}\sum_{i=1}^{n} \hat{B}_i$.

回归模型

$$\hat{B}_i = (2Z_i - 1)\frac{1}{M}\sum_{k\in J_i}\left(\hat{\mu}_{1-Z_i}(X_i) - \hat{\mu}_{1-Z_i}(X_k)\right)$$

匹配的质量

整理以上公式:

K_i = 个体i 用于匹配 次数

推论 15.2 $\hat{\tau}^{mbc} = \hat{\tau}^{reg}$

$$+\frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{K_i}{M}\right)Z_i\left(Y_i - \hat{\mu}_1(X_i)\right) - \frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{K_i}{M}\right)(1-Z_i)\left(Y_i - \hat{\mu}_0(X_i)\right)$$

回归: $\hat{\tau}^{reg} = \frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_1(X_i) - \frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_0(X_i)$

$$\hat{\tau}^{dr} = \hat{\tau}^{reg} + \frac{1}{n}\sum_{i=1}^{n}\frac{Z_i\left(Y_i - \hat{\mu}_1(X_i)\right)}{\hat{e}(X_i)} - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-Z_i)\left(Y_i - \hat{\mu}_0(X_i)\right)}{1-\hat{e}(X_i)}$$

$$\hat{\tau}^{mbc} = \hat{\tau}^{dr}$$ 如果

$$\begin{cases} \hat{e}(X_i) = \dfrac{1}{1+\frac{K_i}{M}} & (Z_i = 1) \\ 1-\hat{e}(X_i) = \dfrac{1}{1+\frac{K_i}{M}} & (Z_i = 0) \end{cases}$$

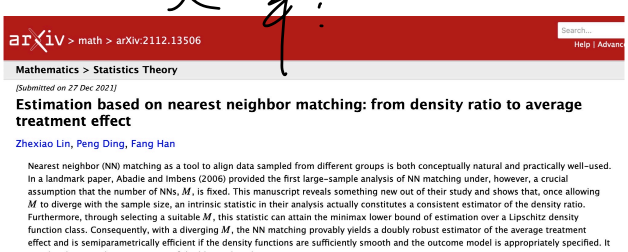
学到了如下知识:

1. $\hat{\tau}^{mbc}$ 可能具有双稳健性

2. 应该允许 $M \longrightarrow \infty$

$\Longrightarrow$ $\hat{e}(X_i)$ 相合性

文章: Lin, Ding, Han (2021) Econometrica



Mathematics > Statistics Theory

[Submitted on 27 Dec 2021]

Estimation based on nearest neighbor matching: from density ratio to average treatment effect

Zhexiao Lin, Peng Ding, Fang Han

Nearest neighbor (NN) matching as a tool to align data sampled from different groups is both conceptually natural and practically well-used. In a landmark paper, Abadie and Imbens (2006) provided the first large-sample analysis of NN matching under, however, a crucial assumption that the number of NNs, M, is fixed. This manuscript reveals something new out of their study and shows that, once allowing M to diverge with the sample size, an intrinsic statistic in their analysis actually constitutes a consistent estimator of the density ratio. Furthermore, through selecting a suitable M, this statistic can attain the minimax lower bound of estimation over a Lipschitz density function class. Consequently, with a diverging M, the NN matching provably yields a doubly robust estimator of the average treatment effect and is semiparametrically efficient if the density functions are sufficiently smooth and the outcome model is appropriately specified. It can thus be viewed as a precursor of double machine learning estimators.

Cochran

Rubin's PhD 论文：关于 $\hat{\tau}^{mbc}$

没有理论

Abadie & Imbens: Econometrica
- 2006
- 2008 Bootstrap fails

Rubin PhD 论文 $\Leftarrow$ JBES 2011 bias correction $\hat{\tau}^{mbc}$

Rosenbaum & Rubin (1983) Biometrika $\Leftarrow$ Econometrica 2016 用 $\hat{e}(X_i)$ 进行匹配

观察性研究：$\begin{cases} Z \perp\!\!\!\perp \big(Y(1), Y(0)\big) \mid X \\ 0 < e(X) < 1 \end{cases}$

估计量：$\begin{cases} \hat{\tau}^{reg} \\ \hat{\tau}^{ht} \,/\, \hat{\tau}^{hajek} : \hat{\tau}^{ipw} \\ \hat{\tau}^{dr} \\ \hat{\tau}^{m} \,/\, \hat{\tau}^{mbc} \end{cases}$