$Y_i = Y_i(W_i)$ SUTVA

$W_i \perp\!\!\!\perp \{Y_i(0), Y_i(1)\}$ random treatment assignment

SUTVA: stable unit treatment value assumption

1. 单位稳定性: unit stability 受试单位相互独立

单一治疗情况

2. single treatment level

Lecture 1 Randomized Controlled Trials

随机对照实验

回归在模型分析 RCT 平均治疗效果中的作用

减小方差； 但在确定 ATE 中不起作用

$$\Delta_i = Y_i(1) - Y_i(0)$$

ATE： $\tau = E[Y_i(1) - Y_i(0)]$

↓ 下面对 τ 进行估计，建模

difference-in-means

均值差异估计

$$\hat{\tau}_{DM} = \frac{1}{n_1}\sum_{W_i=1} Y_i - \frac{1}{n_0}\sum_{W_i=0} Y_i \quad , \quad n_w = |\{i : W_i = w\}|$$

一致无偏

当使用线性模型时：

拟

$$Y_i(w) = c(w) + X_i \beta(w) + \varepsilon_i(w)$$

$$E[\varepsilon_i(w) | X_i] = 0 \quad , \quad Var[\varepsilon_i(w) | X_i] = \sigma^2$$

令 $A = Var[X]$ $\qquad \|v\|_A^2 = v'Av$

则方差 $V_{DM} = 4\sigma^2 + \|\beta(0) + \beta(1)\|_A^2 + \|\beta(0) - \beta(1)\|_A^2$

⇒ 结果： $\tau = E[Y(1) - Y(0)] = c(1) - c(0) + E[X](\beta(1) - \beta(0))$

当用了 OLS： $\hat{\tau}_{OLS} = \hat{c}(1) - \hat{c}(0) + \bar{X}(\hat{\beta}(1) - \hat{\beta}(0))$

$$\hat{\tau}_{OLS} - \tau = \underbrace{\hat{c}(1) - c(1)}_{N(0,\, \sigma^2/n_1)} - \underbrace{(\hat{c}(0) - c(0))}_{N(0,\, \sigma^2/n_0)} + \underbrace{\bar{X}(\beta(1) - \beta(0))}_{N(0,\, \|\beta(1) - \beta(0)\|_A^2/n)} + \underbrace{\bar{X}(\hat{\beta}(1) - \hat{\beta}(0) - \beta(1) + \beta(0))}_{O_p(1/n)}$$

以 $\frac{1}{n}$ 的速度 → 0

可得

$$\sqrt{n}(\hat{\tau}_{OLS} - \tau) \Rightarrow N(0, V_{OLS}) \quad , \quad V_{OLS} = 4\sigma^2 + \|\beta(0) - \beta(1)\|_A^2$$

不确定性减小。

$$\therefore \quad V_{DM} = V_{OLS} + \|\beta(0) + \beta(1)\|_A^2$$

线 一元

No.

Date

OLS 在 ∃模型中比 difference-in-means 好

证：若模型为：

$$Y_i(w) = \mu_{(w)}(X_i) + \varepsilon_i(w)$$

$$E[\varepsilon_i(w) \mid X_i] = 0 \quad , \quad Var[\varepsilon_i(w) \mid X_i] = \sigma^2$$

DM有：

$$\Rightarrow \sqrt{n}\,(\hat{\tau}_{DM} - \tau) \Rightarrow N(0, V_{DM}) = 4\sigma^2 + 2Var[\mu_{(0)}(X_i)] + 2Var[\mu_{(1)}(X_i)]$$

OLS有： 求 Huber-white

$$(c^*_{(w)}, \beta^*_{(w)}) = \operatorname{argmin}_{c,\beta} \{E[(Y_i(w) - X_i\beta - c)^2]\}$$

有：

$$* \begin{cases} \sqrt{n_w}\left(\begin{pmatrix} \hat{c}_{(w)} \\ \hat{\beta}_{(w)} \end{pmatrix} - \begin{pmatrix} c^*_{(w)} \\ \beta^*_{(w)} \end{pmatrix}\right) \Rightarrow N\left(0, \begin{pmatrix} MSE^*_{(w)} & 0 \\ 0 & \cdots \end{pmatrix}\right) \\ c^*_{(w)} = E[Y_i(w)] \qquad MSE^*_w = E[(Y_i(w) - X_i\beta^*_{(w)} - \hat{c}^*_{(w)})^2] \end{cases}$$

$$\sqrt{n}\,(\hat{\tau}_{OLS} - \tau) \Rightarrow N(0, V_{OLS}) \quad , \quad V_{OLS} = V_{DM} - \|\beta^*_{(0)} + \beta^*_{(1)}\|^2_A$$

$V_{OLS} < V_{DM}$ 恒成立

小结：OLS 可看作：

$$\hat{\tau}_{OLS} = \frac{1}{n}\sum_{i=1}^{n}\left(\underbrace{(\hat{c}_{(1)} + X_i\hat{\beta}_{(1)})}_{\hat{\mu}_{(1)}(X_i)} - \underbrace{(\hat{c}_{(0)} + X_i\hat{\beta}_{(0)})}_{\hat{\mu}_{(0)}(X_i)}\right)$$

$\hat{\mu}_{(w)}(x)$ 可看作在 x 处的 OLS 值；

可以用其它方法

No.

Date

则 $\sqrt{n}\,(\hat{\tau}_{AGG} - \tau) \Rightarrow N(0, V_{AGG})$

其中:

$$V_{AGG} = Var[\tau(X_i)] + \sum_{x \in \mathcal{X}} \pi^2(x) \frac{1}{\pi(x)} \frac{\sigma^2(x)}{e(x)(1-e(x))}$$

$$= Var[\tau(X_i)] + E\left[\frac{\sigma^2(X_i)}{e(X_i)(1-e(X_i))} \right]$$

No.

Date

Lecture 2 Unconfoundedness and the Propensity Score

unconfoundedness 未混杂性：希望去排除混杂因素的影响

propensity score：倾向得分

aggregating difference-in-means estimators 聚合均值差异估计

简单根据

样本量加权

$\{Y_i(0), Y_i(1)\} \perp W_i \mid X_i = x$, for all $x \in \mathcal{X}$

→ X_i 的值为数值，X_i 是离散的

x 为分组变量。

定义 group-wise ATE：

$$\tau(x) = E[Y_i(1) - Y_i(0) \mid X_i = x]$$

$$\hat{\tau}_{AGG} = \sum_{x \in \mathcal{X}} \frac{n_x}{n} \hat{\tau}(x)$$

其中 $\hat{\tau}(x) = \frac{1}{n_{x1}} \sum_{\{X_i = x, W_i = 1\}} Y_i - \frac{1}{n_{x0}} \sum_{\{X_i = x, W_i = 0\}} Y_i$

$$n_x = |\{i : X_i = x\}|$$

$$n_{xw} = |\{i : X_i = x, W_i = w\}|$$

再定 $e(x) = P[W_i = 1 \mid X_i = x]$

$$\sqrt{n_x}(\hat{\tau}(x) - \tau(x)) \Rightarrow N\left(0, \frac{Var[Y_i(0) \mid X_i = x]}{1 - e(x)} + \frac{Var[Y_i(1) \mid X_i = x]}{e(x)}\right)$$

$$= N\left(0, \frac{\sigma^2(x)}{e(x)(1 - e(x))}\right)$$

再定 $\hat{\pi}(x) = P[X_i = x] = n_x / n$

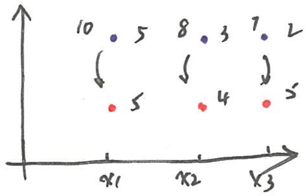
$$\hat{\tau}_{AGG} = \sum_{x \in \mathcal{X}} \hat{\pi}(x)\hat{\tau}(x) = \underbrace{\sum_{x \in \mathcal{X}} \pi(x)\tau(x)}_{\tau} + \underbrace{\sum \pi(x)(\hat{\tau}(x) - \tau(x))}_{N(0, \sum \pi^2(x) Var[\hat{\tau}(x)])}$$

$$+ \underbrace{\sum (\hat{\pi}(x) - \pi(x))\tau(x)}_{N(0, n^{-1} Var[\tau(X_i)])} + \underbrace{\sum (\hat{\pi}(x) - \pi(x))(\hat{\tau}(x) - \tau(x))}_{O_P(1/n)}$$

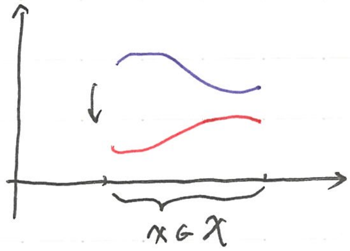
No.

Date

离散：



连续



对 $\forall x \in \mathcal{X} \longleftrightarrow e(x)$

$$P[W_i=w \mid \{Y_i(0), Y_i(1)\}, e(X_i)]$$

$$= \int_{\mathcal{X}} P[W_i=w \mid \{Y_i(0), Y_i(1)\}, X_i=x]\, P[X_i=x \mid e(X_i)]\, dx$$

$$= \int_{\mathcal{X}} P[W_i=w \mid X_i=x]\, P[X_i=x \mid e(X_i)]\, dx$$

Lebesgue integration

$$= e(X_i)\, \mathbb{1}_{w=1} + (1-e(X_i))\, \mathbb{1}_{w=0}$$

连续的 X 和 倾向得分：

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid X_i$$ 在 X_i 下 W_i 无法 peek Y_i；unconfoundedness

当 $e(x) = P[W_i = 1 \mid X_i = x]$

有

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid e(X_i)$$

Propensity stratification 倾向分层

1. 将 X_i 排序 s.t.
$$\hat{e}(X_{i1}) \le \hat{e}(X_{i2}) \le \cdots \le \hat{e}(X_{in})$$

分层使对照和实验组更相似

2. 用分层 J 分层，分别计算 J 中的 $\hat{\tau}$.
$$\hat{\tau}_j = \frac{\sum_{j=\lfloor (j-1)n/J \rfloor+1}^{\lfloor jn/J \rfloor} W_i Y_i}{\sum_{j=\lfloor (j-1)n/J \rfloor+1}^{\lfloor jn/J \rfloor} W_i} - \frac{(1-W_i) Y_i}{(1-W_i)}.$$

3.
$$\hat{\tau}_{STRAT} = \frac{1}{J} \sum_{j=1}^{J} \hat{\tau}_j$$ 一致的 consistent

No.

Date . .

$$\begin{cases} Y_i(0) = c(X_i) - (1 - e(X_i))\tau(X_i) + \varepsilon_i(0), & E[\varepsilon_i(0) \mid X_i] = 0 \\ Y_i(1) = c(X_i) + e(X_i)\tau(X_i) + \varepsilon_i(1), & E[\varepsilon_i(1) \mid X_i] = 0 \end{cases}$$

令 $\mathrm{Var}[\varepsilon_i(w) \mid X_i = x] = \sigma^2(x)$

No.

Date

☆ Inverse-propensity weighting 逆概率加权

IPW:

$$\hat{\tau}_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{W_iY_i}{\hat{e}(X_i)} - \frac{(1-W_i)Y_i}{1-\hat{e}(X_i)}\right)$$

oracle IPW: 完美已知 e(x)

$$\hat{\tau}^*_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{W_iY_i}{e(X_i)} - \frac{(1-W_i)Y_i}{1-e(X_i)}\right)$$

*: 若有 overlap $\eta \le e(x) \le 1-\eta$ for all $x \in \mathcal{X}$

也有 $|Y_i| \le M$:

$$\sup_{x\in\mathcal{X}} |e(x) - \hat{e}(x)| = O_p(a_n) \to 0$$

有:

$$|\hat{\tau}_{IPW} - \hat{\tau}^*_{IPW}| = O_p\left(\frac{a_n M}{\eta}\right)$$

若 $\hat{\tau}^*_{IPW}$ 是一致, 则 $\hat{\tau}_{IPW}$ 也是

而 $E[\hat{\tau}^*_{IPW}] = E[Y_i(1) - Y_i(0)]$ unbiased τ.

若有 overlap. $\hat{\tau}^*_{IPW}$ 以 $\frac{1}{\sqrt{n}}$ 一致于 τ

方差 P15

$$n\,Var[\hat{\tau}^*_{IPW}] = E\left[\frac{c^2(X_i)}{e(X_i)(1-e(X_i))}\right] + Var[\tau(X_i)] + E\left[\frac{\sigma^2(X_i)}{e(X_i)(1-e(X_i))}\right]$$

有 $\sqrt{n}(\hat{\tau}^*_{IPW} - \tau) \Rightarrow N(0, V_{IPW^*})$,

V_{IPW^*}

$$= E\left[\frac{c^2(X_i)}{e(X_i)(1-e(X_i))}\right] + Var[\tau(X_i)] + E\left[\frac{\sigma^2(X_i)}{e(X_i)(1-e(X_i))}\right]$$

KOKUYO

$c(X)$ 是 X 与 Y 的关系 / 非 X 与 △Y 的关系

对比AGG $$V_{IPW*} = V_{AGG} + E\left[\frac{c^2(X_i)}{e(X_i)(1-e(X_i))}\right]$$

→ 除非 $c(X_i)$ 处处为0，V_{IPW*} 比 V_{AGG} 差.

对比 linear modeling IPW 基以 $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid X_i$

LM 依赖线性性.

Lecture 3 Efficient Treatment Effect Estimation via Augmented IPW

原本的普通 IPW 大样本特性不好，例当X为离散时，IPW不如分组求τ.

ATE两种特征：

① $e(x) = P[W_i=1 \mid X_i = x]$

$$\tau = E[\hat{\tau}_{IPW}]$$

$$\hat{\tau}^*_{IPW} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{W_i Y_i}{e(X_i)} - \frac{(1-W_i)Y_i}{1-e(X_i)}\right)$$

② $\mu_{(w)}(x) = E[Y_i(w) \mid X_i = x]$

$$\tau(x) = \mu_{(1)}(x) - \mu_{(0)}(x)$$

so: $\tau = E[\mu_{(1)}(X) - \mu_{(0)}(X)]$

$$\hat{\tau}_{REG} = n^{-1}\sum_{i=1}^{n}\left(\hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i)\right)$$

Augmented IPW — AIPW estimator

$$\hat{\tau}_{AIPW} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i) + W_i\frac{Y_i - \hat{\mu}_{(1)}(X_i)}{\hat{e}(X_i)} - (1-W_i)\frac{Y_i - \hat{\mu}_{(0)}(X_i)}{1-\hat{e}(X_i)}\right)$$

先用 $\mu(x)$ 估计，再用 IPW 处理回归残差 ：IPW使数据中过多的特征被公平化

AIPW的双重稳健性： AIPW consistent if either $\hat{\mu}_{(w)}(x)$ are consistent or $\hat{e}(x)$ is consistent

当 $\hat{\mu}_{(w)}(x) \approx \mu_{(w)}(x)$

$$\hat{\tau}_{AIPW} = \underbrace{\frac{1}{n}\sum_i\left(\hat{\mu}_{(1)}(X_i) - \hat{\mu}_{(0)}(X_i)\right)}_{\text{consistent}} + \underbrace{\frac{1}{n}\sum_i\left(\frac{W_i}{\hat{e}(X_i)}\left(Y_i - \hat{\mu}_{(1)}(X_i)\right) - \frac{1-W_i}{1-\hat{e}(W_i)}\left(Y_i - \hat{\mu}_{(0)}(X_i)\right)\right)}_{\text{mean}=0}$$

因为 $E[Y_i - \hat{\mu}_{(W_i)}(X_i) \mid X_i, W_i] \approx 0$

把 $e(x)$ 来抵了

No.

Date

当 $\hat{e}(x)$ is consistent $\hat{e}(x) \approx e(x)$

$$\hat{\tau}_{AIPW} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{W_i Y_i}{\hat{e}(X_i)} - \frac{(1-W_i)Y_i}{1-\hat{e}(X_i)}\right)}_{IPW} + \underbrace{\frac{1}{n}\sum\left(\hat{\mu}_{(1)}(X_i)\left(1-\frac{W_i}{\hat{e}(X_i)}\right) - \hat{\mu}_{(0)}(X_i)\left(1-\frac{1-W_i}{1-\hat{e}(X_i)}\right)\right)}_{mean=0}$$

因为 $E[1 - W_i/\hat{e}(X_i) \mid X_i] \approx 0$

把 $\hat{\mu}_{(w)}(X_i)$ 乘走了

半参数效率： 大样本下

AIPW asymptotically optimal 所有非参数估计器 最优

当 $\mu_{(w)}(x)$ 和 $e(x)$ 足够准： $\sqrt{n}\,(\hat{\tau}_{AIPW} - \underbrace{\hat{\tau}^{*}_{AIPW}}_{\text{oracle AIPW}}) \longrightarrow_p 0$

若该成立

则下列 CLT 也成立，

中心极限定理

$\hat{\tau}^{*}_{AIPW}$ is just an IID average.

$$\sqrt{n}\,(\hat{\tau}_{AIPW} - \tau) \Rightarrow N(0, V^{*})$$

$$V^{*} = Var[\tau(X_i)] + E\left[\frac{\sigma_0^2(X_i)}{1-e(X_i)}\right] + E\left[\frac{\sigma_1^2(X_i)}{e(X_i)}\right]$$

证明了其效果最好，

Cramer-Rao 型约束.

KOKUYO

No.

Date

AIPW 和 交叉拟合 —— 分数据，随机选数据 M or 更.

分数据 random → two halves I_1 and I_2

$$\hat{\tau}_{AIPW} = \frac{|I_1|}{n}\hat{\tau}^{I_1} + \frac{|I_2|}{n}\hat{\tau}^{I_2}$$

$$\hat{\tau}^{I_1} = \frac{1}{|I_1|}\sum_{i\in I_1}\left(\hat{\mu}_{(1)}^{I_2}(X_i) - \hat{\mu}_{(0)}^{I_2}(X_i) + W_i\frac{Y_i - \hat{\mu}_{(1)}^{I_2}(X_i)}{\hat{e}^{I_2}(X_i)} - (1-W_i)\frac{Y_i - \hat{\mu}_{(0)}^{I_2}(X_i)}{1-\hat{e}^{I_2}(X_i)}\right),$$

可忽略，asymptotically，我们所选的具体的机器学习调整的特殊性.

只依赖以下条件：

1. Overlap : $\eta < e(x) < 1-\eta$ for all $x \in \mathcal{X}$ $(\eta > 0)$

2. Consistency :

所有机器学习调整方法都有大样本一致性

$$\sup_{x\in\mathcal{X}}\left|\hat{\mu}_{(w)}^{I_2}(x) - \mu_{(w)}(x)\right|, \quad \sup_{x\in\mathcal{X}}\left|\hat{e}^{I_2}(x) - e(x)\right| \xrightarrow{p} 0$$

3. Risk decay : outcome 和 propensity model 的衰减 :

$$E\left[\left(\hat{\mu}_{(w)}^{I_2}(X_i) - \mu_{(w)}(X_i)\right)^2\right]E\left[\left(\hat{e}^{I_2}(X_i) - e(X_i)\right)^2\right] = o\left(\frac{1}{n}\right)$$

若 $\hat{\mu}_{(w)}$ 和 $\hat{e}$ 都 $\sqrt{n}$-consistent ∴ error → $O(1/n^2)$ 均方根误差

要满足，一种简单的方式是确保所有回归调整 $o(n^{-1/4})$，即以 RMSE 方面较快速度减小

No.

Date

注：叉 $\hat{\tau}_{AIPW} = \frac{|\mathcal{I}_1|}{n}\hat{\tau}^{\mathcal{I}_1} + \frac{|\mathcal{I}_2|}{n}\hat{\tau}^{\mathcal{I}_2}$；$\hat{\tau}^{\mathcal{I}_1} = \frac{1}{|\mathcal{I}_1|}\sum_{i\in\mathcal{I}_1}\left(\hat{\mu}_{(1)}^{\mathcal{I}_2}(X_i) - \hat{\mu}_{(0)}^{\mathcal{I}_2}(X_i) + W_i\frac{Y_i-\hat{\mu}_{(1)}^{\mathcal{I}_2}(X_i)}{\hat{e}^{\mathcal{I}_2}(X_i)} - (1-W_i)\frac{Y_i-\hat{\mu}_{(0)}^{\mathcal{I}_2}(X_i)}{1-\hat{e}^{\mathcal{I}_2}(X_i)}\right)$

叉 $\hat{\tau}_{AIPW}$ 的特征：

理想下非叉 AIPW

$$\sqrt{n}\,(\hat{\tau}_{AIPW} - \hat{\tau}^*) \to_p 0$$

证明：

$$\hat{\tau}^* = \frac{|\mathcal{I}_1|}{n}\hat{\tau}^{\mathcal{I}_1,*} + \frac{|\mathcal{I}_2|}{n}\hat{\tau}^{\mathcal{I}_2,*}$$

$$\begin{cases} \hat{\tau}^{\mathcal{I}_1} = \hat{\mu}_{(1)}^{\mathcal{I}_1} - \hat{\mu}_{(0)}^{\mathcal{I}_1} \\ \hat{\mu}_{(1)}^{\mathcal{I}_1} = \frac{1}{|\mathcal{I}_1|}\sum_{i\in\mathcal{I}_1}\left(\hat{\mu}_{(1)}^{\mathcal{I}_2}(X_i) + W_i\frac{Y_i-\hat{\mu}_{(1)}^{\mathcal{I}_2}(X_i)}{\hat{e}^{\mathcal{I}_2}(X_i)}\right). \end{cases}$$

show

$\Rightarrow$ 只需证：$\sqrt{n}\,(\hat{\mu}_{(1)}^{\mathcal{I}_1} - \hat{\mu}_{(1)}^{\mathcal{I}_1,*}) \to_p 0$，详P23.

$$\hat{\mu}_{(1)}^{\mathcal{I}_1} - \hat{\mu}_{(1)}^{\mathcal{I}_1,*} = \frac{1}{|\mathcal{I}_1|}\sum_{i\in\mathcal{I}_1}\left(\hat{\mu}_{(1)}^{\mathcal{I}_2}(X_i) + W_i\frac{Y_i-\hat{\mu}_{(1)}^{\mathcal{I}_2}(X_i)}{\hat{e}^{\mathcal{I}_2}(X_i)} - \mu_{(1)}(X_i) - W_i\frac{Y_i-\mu_{(1)}(X_i)}{e(X_i)}\right)$$

$$= \frac{1}{|\mathcal{I}_1|}\sum_{i\in\mathcal{I}_1}\left(\left(\hat{\mu}_{(1)}^{\mathcal{I}_2}(X_i) - \mu_{(1)}(X_i)\right)\left(1-\frac{W_i}{e(X_i)}\right)\right) \leq \frac{o_p(1)}{n}$$

$$+ \frac{1}{|\mathcal{I}_1|}\sum_{i\in\mathcal{I}_1}W_i\left(\left(Y_i-\mu_{(1)}(X_i)\right)\left(\frac{1}{\hat{e}^{\mathcal{I}_2}(X_i)} - \frac{1}{e(X_i)}\right)\right)$$

$$- \frac{1}{|\mathcal{I}_1|}\sum_{i\in\mathcal{I}_1}W_i\left(\left(\hat{\mu}_{(1)}^{\mathcal{I}_2}(X_i) - \mu_{(1)}(X_i)\right)\left(\frac{1}{\hat{e}^{\mathcal{I}_2}(X_i)} - \frac{1}{e(X_i)}\right)\right) \leq o_p\left(\frac{1}{\sqrt{n}}\right)$$

K个折叠：

$$\hat{\tau}_{AIPW} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\mu}_{(1)}^{(-k(i))}(X_i) - \hat{\mu}_{(0)}^{(-k(i))}(X_i) + W_i\frac{Y_i-\hat{\mu}_{(1)}^{(-k(i))}(X_i)}{\hat{e}^{(-k(i))}(X_i)} - (1-W_i)\frac{Y_i-\hat{\mu}_{(0)}^{(-k(i))}(X_i)}{1-\hat{e}^{(-k(i))}(X_i)}\right).$$

No.

Date

效率评分: Efficient Score:

$$S(\theta) = \frac{d(\log L(\theta))}{d\theta}$$ $L(\theta)$为似然函数

经验方差: Empirical Variance $Var(S(\theta))$

n个样本集合, $$var(S(\theta)) = \frac{1}{n-1}\sum\left(S(\theta_i) - \overline{S(\theta)}\right)^2$$

有效方差: 给定问题下可达最小方差 (C-R下界有关)

Confidence intervals 置信区间

有效得分的经验方差收敛于有效方差 V_*:

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(\mu_{(1)}(X_i)-\mu_{(0)}(X_i)+W_i\frac{Y_i-\mu_{(1)}(X_i)}{e(X_i)}-(1-W_i)\frac{Y_i-\mu_{(0)}(X_i)}{1-e(X_i)}-\hat{\tau}^*\right)^2$$

$$\to_p V^*,$$

$\hat{\tau}^*$ 为非交叉完美AIPW:

交叉AIPW也是 $\hat{V}_{AIPW}\to_p V^*$:

$$\hat{V}_{AIPW}=\frac{1}{n-1}\sum_{i=1}^{n}\Bigg(\hat{\mu}_{(1)}^{(-k(i))}(X_i)-\hat{\mu}_{(0)}^{(-k(i))}(X_i)$$

$$+W_i\frac{Y_i-\hat{\mu}_{(1)}^{(-k(i))}(X_i)}{\hat{e}^{(-k(i))}(X_i)}-(1-W_i)\frac{Y_i-\hat{\mu}_{(0)}^{(-k(i))}(X_i)}{1-\hat{e}^{(-k(i))}(X_i)}-\hat{\tau}_{AIPW}\Bigg)^2$$

置信 α-level:

$$\tau\in\left(\hat{\tau}_{AIPW}\pm\frac{1}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\hat{V}_{AIPW}}\right),$$

No.

Date

Lecture 4 Estimating Treatment Heterogeneity

估计治疗异质性　对于 (X_i, Y_i, W_i)　$X_i \in \mathcal{X}$

之前我们注意 ATE：$E[Y_i(1) - Y_i(0)]$

异质性分析：conditional average treatment effect.

CATE:

$$\tau(x) = E[Y_i(1) - Y_i(0) \mid \underline{X_i = x}]$$

一类 X_i 的均值

Regularization bias 正则化偏差

对于 $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid X_i$:

$$\tau(x) = \mu_{(1)}(x) - \mu_{(0)}(x), \quad \mu_{(w)}(x) = E[Y_i \mid X_i = x, W_i = w]$$

problem:

CATE 的估计：$\hat{\tau}_T(x) = \hat{\mu}_{(1)}(x) - \hat{\mu}_{(0)}(x)$.

该方法的问题：若 $\hat{\mu}_1$ 样本过少，$\hat{\mu}_1$ 被严重正则化（线性）

而 $\hat{\mu}_0$ 样本多，$\hat{\mu}_0$ 正确，则 $\hat{\tau}_T(x)$ 会用 $\hat{\mu}_1$ 出错.

Semiparametric modeling 半参数建模

$$Y_i(w) = \underbrace{f(X_i)}_{\text{与}X_i\text{的关}} + w\,\tau(X_i) + \varepsilon_i(w), \quad P[W_i = 1 \mid X_i] = e(x)$$

where $\tau(X_i) = \underbrace{\psi(x)}_{\text{效应与}X\text{的非参数关系}} \cdot \beta$ — 治疗效果的参数

$\varepsilon_i = \varepsilon_i(w)$

在 under confoundness T.. rewrite

$$Y_i - m(X_i) = (W_i - e(X_i))\psi(X_i)\beta + \varepsilon_i$$

where. $m(x) = E[Y_i \mid X_i = x] = f(X_i) + e(X_i)\tau(X_i)$

Robinson

No.

Date

注: $Y_i - m(X_i) = \underline{(W_i - e(X_i))}\psi(X_i)\cdot\beta + \varepsilon_i$ where

$m(x) = E[Y_i|X_i=x] = f(X_i) + e(X_i)\tau(X_i)$

通过上面的式子, 设计 residual-on-residual OLS 模型 (残差回归)

$$\begin{cases} \tilde{Y}_i^* = Y_i - m(X_i) \\ \tilde{Z}_i^* = (W_i - e(X_i))\psi(X_i) \end{cases}$$

只与 X_i Y_i W_i $e(X_i)$ 有关, 数据预处理.

then:

estimate $\hat{\zeta}_R^*$ by $\tilde{Y}_i^* \sim \tilde{Z}_i^*$

则有:

$$\sqrt{n}(\hat{\zeta}^* - \beta) \Rightarrow N(0, V_R), \quad V_R = Var[\tilde{Z}_i^*]^{-1} Var[\tilde{Z}_i^* \tilde{Y}_i^*] Var[\tilde{Z}_i^*]^{-1}$$

半参数有效方差

当同方差下: $Var(\varepsilon_i|X_i, W_i) = \sigma^2$ 恒定, V_R 为 β 的 semiparametrically efficient varianc

不同方差, V_R 仍成立, 但不是半参数有效方差

上述残差回归为理想化的, 因为 $m(X_i)$ 与 $e(X_i)$ 不知;

用交叉验证模拟:

非参数回归

1. non-parametric regression $Y \sim X$ $W \sim X$

   $\Rightarrow$ 得 $\hat{m}(x)$ $\hat{e}(x)$

2. define: $\tilde{Y}_i = Y_i - \hat{m}^{(-k(i))}(X_i)$ 删去i 估出的 m

   $\tilde{Z}_i = (W_i - \hat{e}^{(-k(i))}(X_i))\psi(X_i)$

3. estimate $\hat{\zeta}_R$ $\tilde{Y}_i \sim \tilde{Z}_i$

其中有: $E[(\hat{m}(X) - m(X))^2]^{\frac{1}{2}}, E[(\hat{e}(X) - e(X))^2]^{\frac{1}{2}} = o_p(\frac{1}{n^{1/4}})$

则 $\sqrt{n}(\hat{\zeta}_R^* - \hat{\zeta}_R) \longrightarrow_p 0$

无差

KOKUYO

No.

Date

附总结:

当治疗效果是常数: $Y_i(w) = f(X_i) + \tau w + \varepsilon_i$ and $E[\varepsilon_i \mid X_i, W_i]$

有:

$$\hat{\tau}_R = \frac{\sum_{i=1}^{n} \left(Y_i - \hat{m}^{(-k(i))}(X_i)\right)\left(W_i - \hat{e}^{(-k(i))}(X_i)\right)}{\sum_{i=1}^{n} \left(W_i - \hat{e}^{(-k(i))}(X_i)\right)^2} \quad (4.15)$$

前提是干扰组件收敛足够快

即使 τ 非常数,

$\overline{Y(1)} - \overline{Y(0)}$

$\hat{\tau}$ 也会收敛到 $\tau(x)$ 的一个非负权重的加权平均数

当

R-loss

$\tau(x)$ 异质性较低且在重叠方面有困难的情况下，用(4.15)替代

No.

Date

A loss function for treatment heterogeneity

之前的 estimator : $\tau(x) = \psi(x) \cdot \beta$

大多不是线性，用ML方法:

$$\mu_{(w)}(x) = E[Y_{(w)} \mid X = x] \quad \text{for } w \in \{0,1\}$$

under confoundness $E[\varepsilon_i(w_i) \mid X_i, w_i] = 0$, where $\varepsilon_i(w) := Y_i(w) - (\underbrace{\mu_{(0)}(X_i)}_{\text{类似之前的 } f(X_i)} + w\,\tau(X_i))$

参照之前的重写上式:

$$Y_i - m(X_i) = (W_i - e(X_i))\,\tau(X_i) + \varepsilon_i$$

where $m(x) = E[Y \mid X = x] = \mu_{(0)}(X_i) + e(X_i)\,\tau(X_i)$

$\varepsilon_i := \varepsilon_i(W_i)$

上式也可等价写成:

$$\tau(\cdot) = \operatorname{argmin}_{\tau'} \left\{ E\left[\left((Y_i - m(X_i)) - (W_i - e(X_i))\,\tau'(X_i) \right)^2 \right] \right\}$$

已知 $m(X_i)$ $e(X_i)$ 可用 argmin 算出 $\tau(\cdot)$

$$\hat{\tau}_R^*(\cdot) = \operatorname{argmin}_{\tau'} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left((Y_i - m(X_i)) - (W_i - e(X_i))\,\tau'(X_i) \right)^2 + \Lambda_n(\tau'(\cdot)) \right\}$$

对 $\tau(\cdot)$ 函数复杂性的正则化

$m(x)$、$e(x)$ 不可知 (除非 RCT)

考虑交叉匹配

$$\hat{\tau}_R(\cdot) = \operatorname{argmin}_\tau \left\{ \hat{L}_n(\tau(\cdot)) + \Lambda_n(\tau(\cdot)) \right\}.$$

$$\hat{L}_n(\tau(\cdot)) = \frac{1}{n} \sum_{i=1}^{n} \left((Y_i - \hat{m}^{(-k(i))}(X_i)) - (W_i - \hat{e}^{(-k(i))}(X_i))\,\tau(X_i) \right)^2$$

是 lasso 或核回归…

像 CNN

lasso: $\hat{\tau}(x) = x \cdot \hat{\beta}$ with:

$$\hat{\beta} = \operatorname{argmin}_\beta \left\{ \frac{1}{n} \sum_{i=1}^{n} \left((Y_i - \hat{m}^{(-k(i))}(X_i)) - (W_i - \hat{e}^{(-k(i))}(X_i))\, X_i \beta \right)^2 + \lambda \|\beta\|_1 \right\}$$

No.

Date

为什么传统IPW方法不适用？

之前假设：

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i | Z_i \qquad \text{unconfoundness.}$$

$$\eta \le P\{W_i=1 | Z_i\} \le 1-\eta \qquad \text{overlap.} \qquad \Rightarrow$$

在RDD中，unfoundness 成立，而 overlap 不成立 $P\{W_i=1|Z_i=z\}$ is 0 or 1；

所以只要涉及除 $P\{W_i=1|Z_i=z\}$ 都不成立

我们需要比较在c附近的Z_i，相似但不连续：

overlap 失效 ⇒ 不能对 τ_c 进行$\sqrt{n}$一致估计，τ_c 的最小最大误差将以差分数速率递减，具体速率取决于条件响应函数 $\mu_{(w)}(z)$ 的平滑程度：

例：

若 $\mu_{(w)}(z)$ 在 c 有一个均匀有界的二阶导

则可以实现 $n^{-2/5}$ 的收敛率 $\to 0$

Campus

No.

Date

Lecture 5 Regression Discontinuity Designs

回归不连续设计（临界值）

引入断点的符号:

运行变量 $Z_i \in \mathbb{R}$: $W_i = 1(\{Z_i \geq c\})$ c 为阈值

当 $Z_i \to c$ ，c 邻域的样本特征相似，仅有干预与否不同

令 $\mu_{(w)}(z) = E[Y_i(w) \mid Z_i]$

若 $\mu_{(0)}(z)$ 与 $\mu_{(1)}(z)$ 连续，定义 CATE: at $z=c$ $\tau_c = \mu_{(1)}(c) - \mu_{(0)}(c)$

via:

$$\tau_c = \lim_{z \to c^+} E[Y_i \mid Z_i = z] - \lim_{z \to c^-} E[Y_i \mid Z_i = z]$$

Estimation via local linear regression 局部线性回归

取较小的带宽 $h_n \to 0$ ；一个对称的加权函数 $K(\cdot)$ ，拟合 $\mu_{(w)}(z)$

$$\hat{\tau}_c = \arg\min \left\{ \underbrace{\sum_{i=1}^{n} K\left(\frac{|Z_i - c|}{h_n}\right)}_{权} \times \left(Y_i - a - \tau W_i - \beta_{(0)} \underbrace{(Z_i - c)_-}_{负部} - \beta_{(1)} \underbrace{(Z_i - c)_+}_{正部}\right)^2 \right\}$$

截距 a ；斜率参数 $\beta_{(w)}$ 为干扰参数.

PS：常用的加权函数:

窗函数 $K(x) = 1(\{|x| \leq 1\})$ window function

三角核 $K(x) = (1 - |x|)_+$ triangle kernel

$$\hat{\tau}_c = \arg\min \left\{ \sum_{i=1}^{n} K\left(\frac{|Z_i - c|}{h_n}\right) \times (Y_i - a - \tau W_i - \beta_{(0)}(Z_i - c)_- - \beta_{(1)}(Z_i - c)_+)^2 \right\}$$

估计器性质

说明：$\mu_{(0)}(z)$ 和 $\mu_{(1)}(z)$ 的连续性假设，量化平滑度：$\mu_{(w)}(z)$ 两次可微，二次导均匀有界：

$$\left| \frac{d^2}{dz^2} \mu_{(w)}(z) \right| \le B \quad \text{for all } z \in \mathbb{R} \text{ and } w \in \{0,1\}.$$

若连续性较差，仅有 Lipschitz 连续（有界不可微），局部平均可能更好，因为函数在局部无明显线性趋势

若连续性更好，$k \ge 3$ 阶导也有界，用更高阶的局部回归有利于改善收敛速度

在连续性假设下，估计器的误差：

Taylor expansion： $\mu_{(w)}(z)$ 在 c 处展开

$$\mu_{(w)}(z) = a_{(w)} + \beta_{(w)}(z-c) + \underbrace{\frac{1}{2}\rho_{(w)}(z-c)}_{\text{误差项}} \qquad |\rho_{(w)}(x)| \le B \underbrace{x^2}_{z-c}$$

while nothing that $\tau_c = a_{(1)} - a_{(0)}$

$\hat{\tau}_c$ (5.4) 可分为处理样本和对照样本的回归：

$$\hat{a}_{(1)}, \hat{\beta}_{(1)} = \arg\min_{a,\beta} \left\{ \sum_{Z_i \ge c} K\left(\frac{|Z_i - c|}{h_n}\right)(Y_i - a - \beta(Z_i - c))^2 \right\}$$

同 $\hat{a}_{(0)}, \hat{\beta}_{(0)} \Rightarrow \hat{\tau} = \hat{a}_{(1)} - \hat{a}_{(0)}$

No.

Date

$\hat{a}_{(1)}$的闭合解

之后讨论，窗核 $K(x)=1\ (\{|x|\le 1\})$ 下：

$$\hat{a}_{(1)} = \sum_{c\le Z_i\le c+h_n} \gamma_i Y_i, \quad \gamma_i = \frac{\hat{E}_{(1)}[(Z_i-c)^2] - \hat{E}_{(1)}[Z_i-c]\cdot(Z_i-c)}{\hat{E}_{(1)}[(Z_i-c)^2] - \hat{E}_{(1)}[Z_i-c]^2}$$

where. $\hat{E}_{(1)}[Z_i-c] = \sum_{c\le Z_i\le c+h_n} (Z_i-c)\Big/|\{i: c\le Z_i\le c+h_n\}|$

在带内的E

也可算出：$\sum_{c\le Z_i\le c+h_n} \gamma_i = 1$ and $\sum_{c\le Z_i\le c+h_n} \gamma_i(Z_i-c)=0$

$\Downarrow$.

$$\hat{a}_{(1)} = a_{(1)} + \underbrace{\sum_{c\le Z_i\le c+h_n} \gamma_i \rho_{(1)}(Z_i-c)}_{\text{curvature bias 曲率偏差}} + \underbrace{\sum_{c\le Z_i\le c+h_n} \gamma_i (Y_i-\mu_{(1)}(Z_i))}_{\text{sampling noise 抽样噪声}}.$$

$\hat{a}_{(0)}$类似、

需要约束二者：

1. 曲率偏差：受 Bh_n^2 的上界约束，B为关于 $\mu_{(w)}(z)$ 的二阶导的界，h_n 为带宽
2. 抽样噪声：均值为0，在 $Var[Y_i|Z_i]\le\sigma^2$ 的假设下，
   其方差可被约束至 $\sigma^2\sum_{c\le Z_i\le c+h_n}\gamma_i^2$

若 Z_i 在z的一个邻域内具有连续的，非零密度函数 $f(z)$

则

$$\sigma^2\sum_{c\le Z_i\le c+h_n}\gamma_i^2 \approx \frac{4\sigma^2}{|\{i: c\le Z_i\le c+h_n\}|} \approx \frac{4\sigma^2}{f(c)}\frac{1}{nh_n}$$

$\Downarrow$

偏差$^2 \sim h_n^4$，方差 $\sim \frac{1}{h_n n}$

$$\hat{\tau}_c = \tau_c + O_p(n^{-2/5}) \quad \text{with} \quad h_n\sim n^{-1/5}$$

标准误

若有k阶导平滑 $h_n\sim n^{-1/2k+1}$

KOKUYO

No.

Date

← 可能含可能不含

运行变量噪声 (Z_i) $\qquad u_i \sim G$

定义两个条件：(1) Z_i 噪声： $Z_i \mid u_i \sim N(0, \nu^2)$ ， $\nu > 0$

(2) Z_i unconfounded： $[\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp Z_i] \mid u_i$

$$Y_i = Y_i(W_i) \quad \text{with} \quad \{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid u_i$$

高斯正态分布 $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$

$$e(u) := P[W_i = 1 \mid u_i] = P[Z_i \geq c \mid u_i] = 1 - \Phi\left(\frac{c - u_i}{\nu}\right)$$

$$\alpha_{(w)}(u) := E[Y_i(w) \mid u_i = u] \qquad \tau(u) = E[Y_i(1) - Y_i(0) \mid u_i = u]$$

u_i 观测不到，无法算倾向得分

一种去卷积 deconvolution 类型的估计器：

设置：
$$\hat{\tau}_\gamma = \frac{1}{n} \sum_{\{i: Z_i \geq c\}} \gamma_+(Z_i) Y_i - \frac{1}{n} \sum_{\{i: Z_i < c\}} \gamma_-(Z_i) Y_i$$

并使加权函数满足：

$$E[\mathbb{1}\{Z_i \geq c\} \gamma_+(Z_i)] = E[1\{Z_i < c\} \gamma_-(Z_i)] = 1$$

则可以验证：

$$E[\hat{\tau}_\gamma] = \underbrace{\int h_+(u) \tau(u) \, dG(u)}_{\text{加权处理效应}} + \int \underbrace{(\overset{\text{权}}{h_+(u) - h_-(u)}) \alpha_{(0)}(u) \, dG(u)}_{\text{混杂偏见}}$$

$$\underbrace{h_+(u)}_{\text{加权函数}} = \int_c^{\infty} \gamma_+(z) \underline{\varphi_\nu(z-u)} \, dz \qquad h_-(u) = \int_{-\infty}^{c} \gamma_-(z) \varphi_\nu(z-u) \, dz$$

$$\underset{\text{正态}}{\varphi_\nu(z-u)} = \frac{1}{\sqrt{2\pi\nu^2}} e^{-\frac{1}{2}\left(\frac{z-u}{\nu}\right)^2}$$

所以，在 $|\mu''_{(w)}(z)| \le B$，已知 $\{Z_1, \cdots, Z_n\}$ 时：

minimax linear estimator：

$$\hat{\tau}_c(\gamma^B) = \sum_{i=1}^{n} \gamma_i^B Y_i,$$

$$\gamma^B = \operatorname{argmin}\{\sigma^2 \|\gamma\|_2^2 + I_B^2(\gamma)\}$$

对比局部线性估计

在 $I_B(\gamma) =$ 成立下

MSE 较小，更好

Lecture 6. Finite Sample Inference in RDDs

有限样本RDD：Z_i不连续，或截断函数更复杂

RDD 使用线性回归的性质：

$$\hat{\tau}_c = \operatorname{argmin}\left\{\sum_{i=1}^{n} K\left(\frac{|Z_i-c|}{h_n}\right)\times\left(Y_i - a - \tau W_i - \beta_{(0)}(Z_i-c)_- - \beta_{(1)}(Z_i-c)_+\right)^2\right\} \quad (6.2)$$

$$\hat{\tau}_c(\gamma) = \sum_{i=1}^{n}\gamma_i Y_i \quad (6.3)$$

6.2的解可写为6.3

6.3性质为线性估计器的性质

$\Downarrow$

$Y_i(w) = \mu_{(w)}(Z_i) + \varepsilon_i(w)$，$\varepsilon_i(w)\,|\,Z_i \sim N(0, \sigma^2)$；$\gamma_i$ 仅为 Z_i 的函数

有 $\hat{\tau}_c(\gamma)\,|\,\{Z_1,\cdots,Z_n\} \sim N\left(\hat{\tau}_c^*(\gamma), \sigma^2\|\gamma\|_2^2\right)$　$\|\gamma\|_2$ 2-范数

$\hat{\tau}_c^*(\gamma) = \sum_{i=1}^{n}\gamma_i \mu_{W_i}(Z_i)$　其中 $W_i = 1(|Z_i \geq c|)$

$\gamma = \{\gamma_1, \cdots, \gamma_n\}$

$\|\gamma\|_2 = \sqrt{\gamma_1^2 + \cdots + \gamma_n^2}$

Minimax linear estimation：最小化最坏情况下的MSE

任何线性估计器的条件方差可以直接观察到 $\sigma^2\|\gamma\|_2^2$　（同方差误差）

而其 Bias，依赖于未知函数 $\mu_{(w)}(Z)$，无法观察：

$$\text{Bias}\left(\hat{\tau}_c(\gamma)\,|\,\{Z_1,\cdots,Z_n\}\right) = \sum_{i=1}^{n}\gamma_i\mu_{W_i}(Z_i) - \left(\mu_{(1)}(c) - \mu_{(0)}(c)\right)$$

但当 $\mu_{(w)}(Z)$ 平滑的假定下，其仍可被界定：$|\mu''_{(w)}(Z)| \leq B$

即：

$$\underline{\left|\text{Bias}\left(\hat{\tau}_c(\gamma)\,|\,\{Z_1,\cdots,Z_n\}\right)\right| \leq I_B(\gamma)}$$

$$I_B(\gamma) = \sup\left\{\sum_{i=1}^{n}\gamma_i\mu_{W_i}(Z_i) - \left(\mu_{(1)}(c) - \mu_{(0)}(c)\right) : |\mu''_{(w)}(Z)| \leq B\right\}$$

sup内不带ABS.
二阶导在零点附近对称

则 $\text{MSE}\left(\hat{\tau}_c(\gamma)\,|\,\{Z_1,\cdots,Z_n\}\right) \leq \underline{\sigma^2\|\gamma\|_2^2} + \underline{I_B^2(\gamma)}$

均方误差 = 方差 + 偏差^2　　与μ无关

$\Longrightarrow$

注：最终解 $\hat{f}_c(\gamma) = \sum_{i=1}^{n} \gamma_i Y_i$

$\hat{f}_c(\gamma)$ 为差，即：$\hat{f}_c(\gamma) = \underbrace{\sum \gamma_i^{(1)} Y_i}_{Y=1} + \underbrace{\sum \gamma_i^{(2)} Y_i}_{Y=0}$

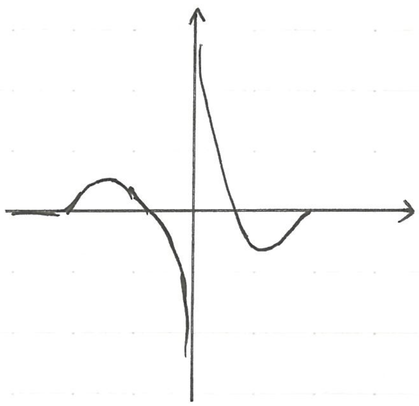
γ Y_1部 和为1的权

γ Y_0部 和为-1权

故有：$\sum \gamma_i = 1 + (-1) = 0$

$\sum \gamma^{(1)} = \sum \gamma w_i = 1$

$\sum \gamma^{(2)} = \sum \gamma (1 - w_i) = -1$



$|\mu''_{(w)}(z)| \le B$，已知 $\{z_1, \cdots, z_n\}$

No.

minimax： $\hat{\tau}_c(\gamma^B) = \sum_{i=1}^{n} \gamma_i^B Y_i$， $\gamma^B = \operatorname{argmin}\{\sigma^2\|\gamma\|_2^2 + I_B^2(\gamma)\}$ (6.8)

Deriving the minimax linear weights

求解 6.8 中的 γ^B

$$\mu_{(w)}(z) = a_{(w)} + \beta_{(w)}(z-c) + \rho_{(w)}(z)$$

其中 $\rho_{(w)}(z)$ 为一个函数：$\rho_{(w)}(c) = \rho'_{(w)}(c) = 0$

其二阶导受 B 限制

$a_{(w)}, \beta_{(w)}$ 不受限制，可使 $I_B(\gamma) \to \infty$；除非加以限制：

$$\sum_{i=1}^{n} \gamma_i W_i = 1 \qquad \sum_{i=1}^{n} \gamma_i = 0 \qquad \sum_{i=1}^{n} \gamma_i (z_i - c)_+ = 0 \qquad \sum_{i=1}^{n} \gamma_i (z_i - c)_- = 0$$

接下来只需控制 $\rho_{(w)}(z)$：重写 6.8：

$$\{\gamma^B, t\} = \operatorname{argmin}\{\underbrace{\sigma^2\|\gamma\|_2^2}_{\text{方差}} + \underbrace{B^2 t^2}_{\text{偏差}}\}$$

subject to： $\sum_{i=1}^{n} \gamma_i W_i \rho_{(1)}(z_i) + \sum_{i=1}^{n} \gamma_i (1-W_i) \rho_{(0)}(z_i) \le t$

for all $\rho_{(w)}(\cdot)$ with $\rho_{(w)}(c) = \rho'_{(w)}(c) = 0$

and $|\rho''_{(w)}(z)| \le 1$

$$\sum_{i=1}^{n} \gamma_i W_i = 1 \qquad \sum_{i=1}^{n} \gamma_i = 0$$

$$\sum_{i=1}^{n} \gamma_i W_i (z_i - c) = 0 \qquad \sum_{i=1}^{n} \gamma_i (1-W_i)(z_i - c) = 0$$

→ 化对偶问题求解

(6.10)

No.

Date

Inference with linear estimators

构造 $\hat{\tau}_c(\gamma^B)$ 的置信区间

$$err := \hat{\tau}_c - \tau_c$$

$$err \mid \{z_1, \cdots, z_n\} \sim N(bias, \sigma^2 \|\gamma^B\|_2^2).$$

同时在 (6.10) 中 $|bias| \le Bt$

又 Gaussian 分布是单峰的:

$$P[|err| \ge \zeta] \le P[|Bt + \sigma\|\gamma^B\|_2 S| \ge \zeta] \qquad S \sim N(0,1)$$

$$\Downarrow$$

level-α 置信区间:

$$P[\tau_c \in I_\alpha \mid \{z_1, \dots, z_n\}] \ge 1-\alpha$$

$$I_\alpha = (\hat{\tau}_c(\gamma^B) - \zeta_\alpha^B, \hat{\tau}_c(\gamma^B) + \zeta_\alpha^B)$$

$$\zeta_\alpha^B = \inf\{\zeta : P[|Bt + \sigma\|\gamma^B\|_2 S| > \zeta] \le \alpha, \quad S \sim N(0,1)\}$$

(6.13)

在 z_i 上有条件：对任何分布的运行变量成立，

$\Rightarrow$ 非标准设置中的回归不连续性

Example: Discrete running variable 离散运行变量 z_i

τ_c 仅通过 $|\mu''_{(w)}(z)| \le B$ 无法确定，但 6.13 仍有效

注: 当 z_i 连续，(6.13) 以 $n^{-2/5}$ 渐近缩小

z_i 离散，置信区间长度不会 $\to 0$

当 z 不在运行变量的支持范围内时，需要一个插值函数代表 $\mu_{(w)}(z)$ 且 st. $|\mu''_{(w)}(z)| \le B$.

Example: Multivariate running variable 多元运行变量　　2/3 测试 or 地理边界 ---

RDD 以 c 为阈，用不了　　(也有人换为一维距离 ρ, 但会失信息)

而 minimax 可以直接拓展为多元，

No.

Date

Beyond homoskedaticity 　等方差性不成立时

→ $\varepsilon_i = Y_i - \mu_{(w_i)}(Z_i) \sim$ Gaussian

$var = \sigma^2$.

当等方差不成立:

$err \mid \{Z_1, \dots Z_n\} \sim N(bias, \sigma^2 \|\hat{\gamma}^B\|_2^2)$ 　(6.11) 不成立

则: 需用中心极限定理证明:

$$\hat{\tau}_c(\gamma) \mid \{Z_1, \dots, Z_n\} \approx N\left(\hat{\tau}_c^*(\gamma), \sum_{i=1}^{n} \gamma_i^2 Var[Y_i \mid Z_i, W_i]\right) \quad (6.14)$$

高斯有效近似后, 置信区间仍可算,

估计条件方差 (6.14) : 　$\hat{V}_n = \sum_{i=1}^{n} \gamma_i^2 (Y_i - \underline{\hat{\mu}_{(w_i)}(Z_i)})^2$

局部线性回归所得,

若 $\hat{\mu}_{(w_i)}(Z_i)$ 是错的, 该界过于保守,

残差膨胀

之前的 6.8 　$\hat{\tau}_c(\gamma^B) = \sum_{i=1}^{n} \gamma_i^B Y_i$, $\gamma^B = \arg\min \{\sigma^2 \|\gamma\|_2^2 + I_B^2(\gamma)\}$

在同方差下 minimax

若在异方差下, 需在 6.10 引入 σ_i^2 (对偶的那个)

KOKUYO

Balance

$\hat{e}(X_i)$ 设计原则 → 应用1：改进 $\hat{e}(x)$ · 损失函数

2：高维 ATE 估计

Lecture 7 Balancing Estimators 平衡估计器

IPW. AIPW 用到了倾向性得分 $e(x)$，本Lecture 用凸优化导出如何估计 $e(x)$

Review:

IID samples : $\{X_i, Y_i, W_i\} \in \mathcal{X} \times \mathbb{R} \times \{0,1\}$

goal : estimate $\tau = \mu(1) - \mu(0)$, where $\mu(w) = E[Y_i(w)]$

Unconfoundedness : $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp W_i \mid X_i$ (7.1)

Overlap : $0 < \eta \le e(X_i) \le 1-\eta < 1$ (7.2)

where $e(x) = P[W_i = 1 \mid X_i = x]$

η 为正常数

⇓

估计 $\mu(1)$: IPW :

$$\hat{\mu}_{IPW}(1) = \frac{1}{n}\sum_{i=1}^{n} \frac{W_i Y_i}{\hat{e}(X_i)} \tag{7.3}$$

下证 IPW 的连续性：(简要，不严谨)

1. Population balance : unconfoundedness 下，$\mu(1) = E\left[\frac{W_i Y_i}{e(X_i)}\right]$
2. oracle estimator : 当 $e(x)$ 为真下，$\hat{\mu}_{IPW}(1)$ 无偏

   overlap 下，$\hat{\mu}_{IPW}(1)$ 方差有界
3. Feasible approximation : Cauchy-Schwartz 不等式 $|a \cdot b|^2 \le \|a\|^2 \cdot \|b\|^2$

   若 $\hat{e}(X_i)$ 满足 overlap :

$$|\hat{\mu}_{IPW}(1) - \hat{\mu}^{*}_{IPW}(1)| \le \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\hat{e}(X_i)} - \frac{1}{e(X_i)}\right)^2} \sqrt{\frac{1}{n}\sum_{i=1}^{n}(W_i Y_i)^2}$$

$$\le \frac{1}{\eta^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{e}(X_i) - e(X_i))^2} \sqrt{\frac{1}{n}\sum_{i=1}^{n}(W_i Y_i)^2} \tag{7.4}$$

在 AIPW 中，$n \to \infty$，7.4 的误差相较 $1/\sqrt{n}$ 为一个极小项

No.

Date

Population vs. sample balance

将条件响应函数用基础函数展开 basis expansion.

$$\mu_{(w)}(x) = \sum_{j=1}^{\infty} \beta_j(w) \underline{\psi_j(x)}$$ ← 基函数 (多项式基、傅里叶基，…)

其中 $\psi_j(\cdot)$ 是一组预定义的基函数

在合理的正则性条件下：有

? 连续可积有界平滑…

$$\mu(w) = \sum_{j=1}^{\infty} \beta_j(w) E[\psi_j(X_i)]$$ （左右求E）

揭示性证，IPW 在总体 population 上有效：

已有 unconfoundedness 无混杂性

$Y_i = \mu_{(w_i)}(X_i) + \varepsilon_i$ with $E[\varepsilon_i | X_i, w_i] = 0$

$$E\left[\frac{W_iY_i}{e(X_i)}\right] = E\left[\frac{W_i}{e(X_i)} \sum_{j=1}^{\infty} \beta_j(w) \psi_j(X_i)\right]$$

$$= \sum_{j=1}^{\infty} \beta_j(w) E\left[\frac{W_i \psi_j(X_i)}{e(X_i)}\right]^{w_i=1}$$ ← 本质 $E[W_i|X_i] = e(X_i)$

$$\Downarrow$$

$$= \sum_{j=1}^{\infty} \beta_j(w) E[\psi_j(X_i)] = \mu(1)$$

所以，我们在选择权重的时候：

$$\frac{1}{n}\sum_{i=1}^{n} \frac{W_i \psi_j(X_i)}{\hat{e}(X_i)} \approx \frac{1}{n}\sum_{i=1}^{n} \psi_j(X_i), \quad \text{for all } j = 1, 2, \cdots$$

(7.8)

范数：

0范数，$\sum (x_i)^0$

1范数，$\sum |x_i|$

2范数 $\sqrt{\sum x_i^2}$

∞范数 $\max(|x_i|)$

$$L_p = \|x\|_p = \sqrt[p]{\sum_{i=1}^{n} x_i^p}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{W_i\psi_j(X_i)}{\hat{e}(X_i)} \approx \frac{1}{n}\sum_{i=1}^{n}\psi_j(X_i), \text{ for all } j=1,2,\cdots \quad (7.8)$$

①

Balancing loss function for propensity estimation

假设，有一个线性的结果模型：$\mu_{(w)}(x) = x\cdot\beta(w)$ ← 实现样本平衡仅涉及到平衡协变量 X_i

一个 logistic 倾向模型：$e(x) = 1/(1+e^{-x\theta})$

改写 7.8.

$$\frac{1}{n}\sum_{i=1}^{n}(1+e^{-X_i\hat{\theta}})W_iX_i = \frac{1}{n}\sum_{i=1}^{n}X_i \quad (7.9)$$

更好的平衡，(仅限估值)

$$\Downarrow$$

#

7.9 为以下凸问题的 KKT 条件： ⟶ 因此可用一个 Newton descent 寻出唯一解

$$\hat{\theta} = \operatorname{argmin}_\theta\left\{\frac{1}{n}\sum_{i=1}^{n} l_\theta(X_i, Y_i, W_i)\right\}$$

$$l_\theta(X_i, Y_i, W_i) = W_ie^{-X_i\theta} + (1-W_i)X_i\theta \quad (7.10)$$

作为损失函数更好.

PS.

估计 $\mu_{(0)}, \mu_{(1)}$ 得用 2 个不同倾向模型

②

ATE estimation with high-dimensional confounders 用高维混杂变量处理 ATE

例 $X_i \in R^p$ 可能 $p \gg n$

而 $\mu_{(w)}(x) = x\cdot\beta(w)$ 对某个 $\beta(w)\in R^p$ 高维情况下，有足够稳定的 $\hat{e}(x)$ 很难.

找精准权重不可能，只求近似平衡：

Hölder 不等式：
$$\left(\sum_{i=1}^{n}|a_ib_i|\right) \le \left(\sum_{i=1}^{n}|a_i|^p\right)^{1/p}\left(\sum_{i=1}^{n}|b_i|^q\right)^{1/q}$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{W_i\mu_{(1)}(X_i)}{\hat{e}(X_i)} - \frac{1}{n}\sum_{i=1}^{n}\mu_{(1)}(X_i)$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{W_iX_i}{\hat{e}(X_i)} - \frac{1}{n}\sum_{i=1}^{n}X_i\right)\beta_{(1)} \le \underbrace{\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{W_iX_j}{\hat{e}(X_i)} - \frac{1}{n}\sum_{i=1}^{n}X_i\right\|_\infty}_{\text{不平衡度}} \|\beta_{(1)}\|_1 \quad (7.11)$$

? ⇒ 暗示：好的逆倾向性权重应使平衡程度最差的情况下最好

problems： ① 如何参数化 $\hat{e}(X_i)$，以实现 7.11 supnorm approximate balance

② 上述界限只有在 $\|\beta_{(1)}\|_1$ 有界时才有意义

高维统计中，通常设 $\beta_{(1)}$ 应稀疏，却未限制其 1-范

No.
Date

井、
对于某个 well chosen ζ:

调节权重.

$$\hat{\gamma} = \operatorname{argmin}_{\gamma}\left\{ \underbrace{\frac{1}{n^2}\|\gamma\|_2^2}_{L_2范} + \zeta \Big\| \underbrace{\frac{1}{n}\sum_{i=1}^{n}(\gamma_i W_i - 1)X_i}_{不平衡度} \Big\|_\infty^2 : \gamma_i \geq 1 \right\} \quad (7.12)$$

之后，使用 $1/\hat{e}(X_i) = \hat{\gamma}_i$ 作为逆倾向性加权

用 sub-Gaussian concentration inequalities 验证 $e(x)$ 满足:

$$\frac{1}{n^2}\Big\|\frac{1}{e(X_i)}\Big\|_2^2 \leq \frac{\eta^{-2}}{n}, \quad \Big\|\frac{1}{n}\sum_{i=1}^{n}(\gamma_i W_i - 1)X_i\Big\|_\infty^2 = O_p\left(\frac{\eta^{-2}\log(p)}{n}\right)$$

7.13

这里 $\hat{e}(X_i)$ 可能不是 $e(x)$ 的好估计，但 $1/\hat{e}(x) = \hat{\gamma}_i$ 在 ζ 好的情况下满足 缩放界限

二、
用高维回归增强加权估计器，解决对 $\|\beta_{(1)}\|_1$ 的依赖

增强平衡估计器:

$$\hat{\mu}_{AB(1)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_{(1)}(X_i) + \hat{\gamma}_i W_i\left(Y_i - \hat{\mu}_{(1)}(X_i)\right), \quad \hat{\mu}_{(1)}(X_i) = X_i\hat{\beta}_{(1)}$$

(7.14)

$$\Downarrow$$

$$\hat{\mu}_{AB(1)} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}X_i\beta_{(1)}}_{均: \mu_{(1)}(X_i)} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_i W_i(Y_i - X_i\beta_{(1)})}_{noise,\ mean\text{-}0}$$

$$+ \underbrace{\left(\frac{1}{n}(1-\hat{\gamma}_i W_i)X_i\right)\left(\hat{\beta}_{(1)} - \beta_{(1)}\right)}_{偏差\ \leq\ \|\frac{1}{n}(1-\hat{\gamma}_i W_i)X_i\|_\infty \|\hat{\beta}_{(1)} - \beta_{(1)}\|_1} \quad (7.15)$$

$\hat{\gamma}_i$ 只needs纠正 $\beta_{(1)} - \hat{\beta}_{(1)}$，而非 $\hat{\beta}_{(1)}$

而在高维回归中，我们知道如何获取 $\|\hat{\beta}_{(1)} - \beta_{(1)}\|_1$ 的强界限

尤其在稀疏性 $\|\beta_{(1)}\|_0 \leq k$ 和受限特征条件下，lasso 可以实现 1-范数误差:

即

$$\|\hat{\beta}_{(1)} - \beta_{(1)}\|_1 = O_p\left(k\sqrt{\frac{\log(p)}{n}}\right)$$

No.

Date

整合：

(7.16)

X_i 满足 $\|\hat\beta(1)-\beta(1)\|_1 = O_p\left(k\sqrt{\frac{\log(p)}{n}}\right)$ 成立的条件.

$\beta(1)$ 是 k-sparse（k稀疏），$k \ll \sqrt{n}/\log(p)$

Then:

① 在 treated units 上运行 lasso 回归

⇒ find $\beta(1)$ 的 1-范数误差衰减为 $o_p(1/\sqrt{\log(p)})$

② 用 (7.12) 拟合 $\hat\gamma$，其中无穷范数的不平衡为 $O_p(\sqrt{\log(p)/n})$

③ 用 (7.14) 估计 $\mu_{AB}(1)$，将两个界限代入 (7.15)

$$\hat\mu_{AB}(1) - \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i\beta(1)}_{\mu_{(1)}(X_i)\ \text{mean}} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\hat\gamma_i W_i(Y_i - X_i\beta(1))}_{\text{noise}\ \sim O_p(1/\sqrt{n})} + o_p\left(\frac{1}{\sqrt{n}}\right) \quad (7.17)$$

也有中心极限：

$$\frac{\sum_{i=1}^{n}\hat\gamma_i W_i(Y_i - X_i\beta(1))}{\sqrt{\sum_{i=1}^{n}\hat\gamma_i^2 W_i^2(Y_i - X_i\beta(1))^2}} \Rightarrow N(0,1)$$

比预期弱，因为我们只用了 $\mu_{(1)}(X_i)$ 样本平均值的置信区间，$\bar\mu(1) := n^{-1}\sum_{i=1}^{n}X_i\beta(1)$ 而不是 $\mu(1)$ 本身

要得到关于 $\mu(1)$ 的结果，需要展示 $\hat\gamma_i$ 的收敛性

KOKUYO

	$t=t_1$	$t=t_2$	$\cdots$	$t=T$
$i=1$				
$i=2$				
$\vdots$				
$i=n$				

附，若对 ε_{it} 建模：视行间噪声 IID. 即 $Var[\varepsilon_i] = \Sigma$

（将每个 ε_{it} 视为独立难以证明且不合理）

当 $T=2$：

直接考虑差异的方差：

$$\widehat{Var}[\hat{\tau}] = \frac{\widehat{Var}[Y_{i2}-Y_{i1} \mid W_{i2}=1]}{|\{i: W_{i2}=1\}|} + \frac{\widehat{Var}[Y_{i2}-Y_{i1} \mid W_{i2}=0]}{|\{i: W_{i2}=0\}|} \quad (8.4)$$

该式在 ε_i 行内相关时，也 Robust

Lecture 8 Methods for Panel Data 面板数据

面板数据（纵向数据）

$i=1,\dots,n$ 个单位，在 $t=1,\dots,T$ 个时间段的数据

A constant treatment effect model 常量治疗效应

结果：$Y_{it}\in R$ ；治疗分配：$W_{it}\in\{0,1\}$

$$Y_{it}=Y_{it}(0)+W_{it}\tau, \quad \text{for all } i=1,\dots,n,\ t=1,\dots,T \qquad (8.1)$$

PS：这是最基础的模型，其未考虑

- treatment heterogeneity 治疗异质性：治疗在所有t内以相同方式影响所有i （τ与t无关）（之前讨论过）
- treatment dynamics 治疗动态性：单位i在t的结果仅受其在t接受的治疗影响（与t前无关）

The two-way model

$$Y_{it}=\alpha_i+\beta_t+W_{it}\tau+\varepsilon_{it}, \quad E[\varepsilon|\alpha,\beta,W]=0 \qquad (8.2)$$

单位和时间上的偏移

治疗严格外生，强于无偏混淆性

如：

$i\in N_1 \quad W_{i\cdot}=(0,0)$

$i\in N_2 \quad W_{i\cdot}=(0,1)$

闭合解：

$$\hat{\tau}=\frac{1}{|\{i:W_{i2}=1\}|}\cdot\sum_{\{i:W_{i2}=1\}}(Y_{i2}-Y_{i1})-\frac{1}{|\{i:W_{i2}=0\}|}\sum_{\{i:W_{i2}=0\}}(Y_{i2}-Y_{i1})$$

（DID法）

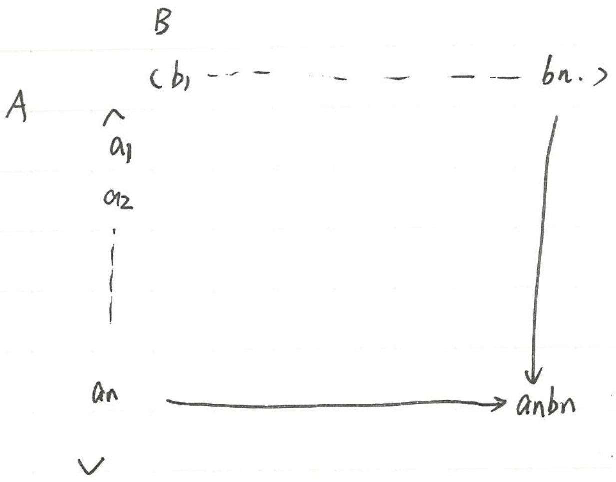
缺 在n, T很大，且W_{it}有通用分布下，(8.2)不适合；

类似地，只有2个时间段，用OLS处理出的有时也称为DID，（但OLS无闭合解）

优点：平行趋势（单位仅由α_i不同），下很直观，无平行则不可用

No.

Date



若$k\neq1$

$a_{n1}b_{n1}+a_{n2}b_{n2}\cdots\cdots$

PS:

合成控制并非处理多元式固定效应的唯一方法:

例:

通过k-均值聚类矩阵Y的行,然后分割为每个群组拟合时变基线模型 P66.文献

Interactive panel models 交互式面板模型

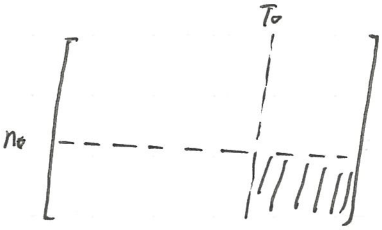
$$Y_{it} = A_{i\cdot} B_{t\cdot}^{\top} + W_{it}\tau + \varepsilon_{it}, \quad E[\varepsilon | A, B, W] = 0, \quad A \in \mathbb{R}^{n\times k},\ B \in \mathbb{R}^{T\times k} \quad (8.5)$$

其中k是某个秩参数

Equivalently:

$$Y_{it} = L_{it} + W_{it}\tau + \varepsilon_{it} \quad \text{for some rank-}k\text{ matrix } L$$

处理队模型：合成控制 and 合成差分



设，当且仅当 $i>n_0$ 且 $t>T_0$：$W_{it}=1$

$$\underbrace{\sum_{i=1}^{n_0}\gamma_i Y_{it}}_{T_0\text{未受T}} \approx \alpha + \underbrace{\frac{1}{n-n_0}\sum_{i=n_0+1}^{n} Y_{it}}_{T_0\text{后受T}}, \quad t=1,\dots,T_0$$

α是一个类似于固定效应相似的参数

其中 γ_i 的选择：

$$\gamma = \arg\min_{\gamma',\alpha}\left\{ \left\| \sum_{i=1}^{n_0}\gamma_i' Y_{i(1:T_0)} - \frac{1}{n-n_0}\sum_{i=n_0+1}^{n} Y_{i(1:T_0)} - \alpha\right\|_2^2 : \sum_{i=1}^{n_0}\gamma_i'=1,\ \gamma_i'\ge 0\right\} \quad (8.7)$$

γ_i'的目的是创建平行趋势，其应可平衡 A_i

最后：

$$\hat\tau = \arg\min_{\tau',\alpha',\beta'}\left\{\sum_{i,t}\gamma_i (Y_{it}-\alpha_i-\beta_t-W_{it}\tau)^2\right\}, \quad (8.8)$$

$\gamma_i = \frac{1}{n-n_0}$, for $i>n_0$

其有封闭解：

$$\hat\tau = \frac{1}{n-n_0}\sum_{i=n_0+1}^{n}\left(\frac{1}{T-T_0}\sum_{t=T_0+1}^{T}Y_{it} - \frac{1}{T_0}\sum_{t=1}^{T_0}Y_{it}\right) - \sum_{i=1}^{n_0}\gamma_i\left(\frac{1}{T-T_0}\sum_{t=T_0+1}^{T}Y_{it}-\frac{1}{T_0}\sum_{t=1}^{T_0}Y_{it}\right)$$

与 T_0 与 n 有关

且该解，在适当的大面板渐近下 $(n_0, n, T\to\infty)$ 为估计量一致

并允许对低秩规范(8.5)中的τ进行渐近正态推断

(8.2): $Y_{it} = \alpha_i + \beta_t + W_{it}\tau + \varepsilon_{it}$, $E[\varepsilon \mid \alpha, \beta, W] = 0$

P67, [2019]

附: (8.11) $\hat{\tau} = \sum_{it} \gamma_{it} Y_{it}$ 在 (8.2) 下也无偏

只要满足:

$$\begin{cases} \sum_t \gamma_{it} = 0 \quad \text{for all } i = 1, \dots, n \\ \sum_i \gamma_{it} = 0 \quad \text{for all } t = 1, \dots, T \\ \sum_{it} \gamma_{it} W_{it} = 1 \end{cases}$$

三个等式排除了固定效应 α_i, β_t

基于此, 得一双重稳健估计器

$$\gamma = \arg\min_{\gamma'} \sum_{it} \gamma_{it}'^2$$

subject to: $\sum_{it} \gamma_{it} W_{it} = 1$, $\sum_{\{i: S_i = s\}} \gamma_{it} = 0$ for all s, t,

$\sum_t \gamma_{it} = 0$ for all i ; $\sum_i \gamma_{it} = 0$ for all t,

只要(8.2) 或 (8.10) 成立, 其无偏

$$Y_{it} = Y_{it}(0) + W_{it}\tau$$

Identification via exchangeability

"design-based" assumptions.

$$Y_{i\cdot}(0) \perp\!\!\!\perp W_{i\cdot} \mid S_i \quad , \quad S_i = \sum_{t=1}^{T} W_{it} \qquad (8.10)$$

其一个推论为:

形如: $\hat{\tau} = \sum_{i,t} \gamma_{it} Y_{it}$ 的估计器，实现了τ的无偏估计. (8.11)

其中 γ-matrix 仅依赖 W_{it}

在(8.10)下，$Y_{it}(0)$ (对t) 可交换:

$$E[\hat{\tau} \mid W, \gamma] = \sum_{i,t} \gamma_{it} E[Y_{it}(0) \mid S_i] + \tau \sum_{i,t} \gamma_{it} W_{it}$$

对 i,t, $\underbrace{\sum \gamma_{it}}_{0} \underbrace{E[Y_{it}(0) \mid S_i]}_{\text{constant}} = 0$

因此，当 $\sum_{i,t} \gamma_{it} W_{it} = 1$ ，且

$$\sum_{\{i: S_i = s\}} \gamma_{it} = 0 \quad \text{for all } t = 1, \cdots, \text{ and } s \in \mathcal{S}$$

$\mathcal{S}$ 表示 S_i 的支持

(8.11)无偏

通过最小化这些无偏性约束下的方差来选择γ:

$$\gamma = \arg\min_{\gamma'} \left\{ \sum_{i,t} \gamma_{it}^2 : \sum_{i,t} \gamma_{it} W_{it} = 1 , \sum_{i,t} \gamma_{it} = 0 \text{ for all } s, t \right\} \quad (8.14)$$

一般来说，只要 W_{it} 在相同 S_i 下有变化

Z_i 作为工具变量的三个条件:

① Z_i 必须是 exogenous (外生的) $\varepsilon_i \perp Z_i$

② Z_i relevant $Cov[W_i, Z_i] \neq 0$

③ Z_i exclusion restriction 排除性限制

$\Downarrow$

Z_i对Y_i的所有影响必须通过W_i

No.

Date

Lecture 9 Instrumental Variables Regression

工具变量回归，识别内生处理效果（有偏 $Y_i(0) \not\perp W_i$）

A structural model 结构模型

无偏模型：

$$Y_i(w) = Y_i(0) + w\tau \quad , \qquad Y_i = Y_i(W_i) \qquad (9.1)$$

一般无偏指：$\{Y_i(w)\} \perp W_i$，(9.1)中指 $Y_i(0) \perp W_i$

$$\Downarrow$$

$$Y_i = \alpha + W_i\tau + \varepsilon_i \qquad (9.2)$$

where: $\alpha = E[Y_i(0)]$ $\quad \varepsilon_i = Y_i(0) - E[Y_i(0)]$

$E[\varepsilon_i | W_i] = 0$ ，用 OLS $Y_i \sim W_i$ 求 $\hat{\tau}$

$$\Downarrow$$

当无偏：$Y_i(0) \not\perp W_i$，(9.2)仍成立，但 $E[\varepsilon_i | W_i] \neq 0$

若 9.2 用 OLS：

$$\tau_{OLS} = \frac{Cov[Y_i, W_i]}{Var[W_i]} = \frac{Cov[\tau W_i + \varepsilon_i, W_i]}{Var[W_i]} = \tau + \frac{Cov[\varepsilon_i, W_i]}{Var[W_i]} \neq \tau$$

OLS的估计

Indentification using instrumental variables. Z_i

$$Y_i = \alpha + W_i\tau + \varepsilon_i \quad , \quad \varepsilon_i \perp Z_i$$

$$W_i = Z_i\gamma + \eta_i \qquad (9.4)$$

Z_i 与 ε_i 不相关：

$$Cov[Y_i, Z_i] = Cov[\tau W_i + \varepsilon_i, Z_i] = \tau Cov[W_i, Z_i]$$

即：

$$\tau = Cov[Y_i, Z_i] / Cov[W_i, Z_i]$$

则在估计时：

$$\hat{\tau} = \widehat{Cov}[Y_i, Z_i] / \widehat{Cov}[W_i, Z_i]$$

KOKUYO

No.

Date

可拓展为: $\varepsilon_i \perp\!\!\!\perp Z_i \mid X_i$

$$\tau(x) = \frac{\mathrm{Cov}[Y_i, w(Z_i) \mid X_i = x]}{\mathrm{Cov}[W_i, w(Z_i) \mid X_i = x]} \Rightarrow$$

$\Downarrow$ 有 构建平均效应 $\tau = E[\tau(X)]$ 的 双重稳健估计量.

$\tau(\cdot)$ 随机森林估计量

当 x_1, x_2 独立.

$$\mathrm{Var}(X_1 \cdot X_2) = \mathrm{Var}(X_1)\mathrm{Var}(X_2) + E^2(X_2)\mathrm{Var}(X_1) + E^2(X_1)\mathrm{Var}(X_2)$$

No.
Date

Optimal instruments 最优工具变量

将 Z_i 拓展为 $\mathcal{Z}$ 高维空间:

$$Y_i = \tau W_i + \varepsilon_i \quad , \quad \varepsilon_i \perp\!\!\!\perp Z_i \quad , Y_i, W_i \in \mathbb{R} \quad , \quad Z_i \in \mathcal{Z}$$

$\mathcal{Z}$ 是 high-dimensional space.

w 为函数: $w: \mathcal{Z} \to \mathbb{R}$:

$$\tau = \frac{Cov[Y_i, w(Z_i)]}{\underline{Cov[W_i, w(Z_i)]}} \quad (9.8)$$

必须为0

问题: $w(\cdot)$ 什么时候最优? 最优转换:

$$\hat{\tau}_w = \frac{\widehat{Cov}[Y_i, w(Z_i)]}{\widehat{Cov}[W_i, w(Z_i)]} = \frac{\frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})(w(Z_i) - \overline{w(Z)})}{\frac{1}{n}\sum_{i=1}^n (W_i - \bar{W})(w(Z_i) - \overline{w(Z)})} \quad (9.8)$$

: with: $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$

注意到, $\hat{\tau}_w$ (9.8) 的值为估计方程的解

$$\frac{1}{n}\sum_{i=1}^n \left(w(Z_i) - \overline{w(Z)}\right)\left(Y_i - \bar{Y} - \hat{\tau}_w (W_i - \bar{W})\right) = 0 \quad (9.10)$$

⇓ 可推

$$n(\hat{\tau}_w - \tau) \Rightarrow \mathcal{N}(0, V_w) \quad , \quad V_w = \frac{Var[\varepsilon_i] Var[w(Z_i)]}{\underline{Cov[W_i, w(Z_i)]^2}} \quad (9.11)$$

$E(\varepsilon_i) = 0$

由于 $Z_i \perp\!\!\!\perp \varepsilon_i \Rightarrow Var[\varepsilon_i (w(Z_i) - E[w(Z_i)])] = Var[\varepsilon_i] Var[w(Z_i)]$

求 Z_i 项.

最优:

$$w^*(\cdot) \in \operatorname{argmax}_{w'} \left\{ Cov[W_i, w'(Z_i)]^2 / Var[w'(Z_i)] \right\}.$$

解为:

$$w^*(z) \propto E[W_i \mid Z_i = z]$$ ⇒ $w^*(Z)$ 是 W_i 从 Z_i 的最佳预测

KOKUYO

No.

Date

注:

若对一些函数 ψ_i 满足 $E[\psi_i(\theta)]=0$

当通过

$\frac{1}{n}\sum_{i=1}^{n}\psi_i(\hat{\theta})=0$ 去估计 $\hat{\theta}$,

由非参数 delta 方法:

一般来说: $\sqrt{n}(\hat{\theta}-\theta) \Rightarrow N(0,V)$

其中 $V=\frac{Var[\psi_i(\theta)]}{E[\psi_i'(\theta)]^2}$

Cross-fitting and feasible estimation 交叉拟合和可行估计

最优工具变量解:

$$w^*(z) = E[W_i \mid Z_i = z]$$

一般 step: ① 拟合一个非参数的第一阶段回归,得 $E[W_i \mid Z_i = z]$ 的估计 $\hat{w}(\cdot)$
② 用 $\hat{w}(\cdot)$ 作为 IV 运行 (9.9) $\Rightarrow$ $\hat{\tau}$

但当 IV 较弱时,会有过拟合偏见:

当 $\hat{w}(Z_i)$ 由训练数据拟合的,则 $\hat{w}(Z_i) \not\perp \varepsilon_i$ ∵因为 $\hat{w}(Z_i) \Leftarrow W_i \Leftarrow \varepsilon_i$

$\Downarrow$

例 Z_i 为 noise $W_i = \alpha + \beta Z_i + \eta_i$

$Y_i = \gamma + \tau \hat{W_i} + \varepsilon_i$

$= \gamma + \tau(\hat{\alpha} + \hat{\beta} Z_i) + \varepsilon_i$

偏

(9.13)

$\Downarrow$

与 $Y_i = \gamma + \tau W_i + \varepsilon_i$

等价

用交叉拟合:

K个折叠,对每个k:

$$\hat{\tau} = \frac{\widehat{\mathrm{Cov}}[Y_i, \hat{w}^{(-k(i))}(Z_i)]}{\widehat{\mathrm{Cov}}[W_i, \hat{w}^{(-k(i))}(Z_i)]}$$

则有: $\hat{w}^{(-k(i))}(Z_i) \perp\!\!\!\perp \varepsilon_i$

特别是,当 $\hat{w}^{(-k(i))}(z)$ 在MSE中对 $E[W_i \mid Z_i = z]$ 是一致的,则 9.13 与 9.9 等价

$\Downarrow$

$\lim_{n \to \infty} E[(\hat{w}_n - E)^2] = 0$ n为样本量。

Non-parametric instrumental variables regression 非参数 IV 回归

本章假设 $Y_i = W_i \tau + \varepsilon_i$ （即 Z_i 线性作用于 Y_i）

推广：

$$Y_i = g(W_i) + \varepsilon_i \quad Z_i \perp\!\!\!\perp \varepsilon_i, \quad Y_i, W_i \in \mathbb{R}, \quad Z_i \in \mathcal{Z}$$

$g(\cdot)$ 为一般平滑函数

不能简单进行 $Y_i \sim W_i$ 非参回归，即 $g(\cdot) \neq E[Y_i \mid W_i = w]$

则应改：

$\Rightarrow$ $Z_i \perp\!\!\!\perp \varepsilon_i$, $E[\varepsilon_i] = 0$:

$$\begin{aligned} E[Y_i \mid Z_i = z] &= E[g(W_i) + \varepsilon_i \mid Z_i = z] \\ &= E[g(W_i) \mid Z_i = z] \\ &= \int_{\mathbb{R}} g(w) f(w \mid z)\, dw \end{aligned} \tag{9.15}$$

其中 $f(w \mid z)$ 表示给定 $Z_i = z$ 时 W_i 的条件密度

则学习方案：(1) 拟合 $\hat{f}(w \mid z)$，且用交叉拟合.

(2) 最小化估求 $g(w)$，选择一个合适的函数类 $\mathcal{G}$.

$$\hat{g}(\cdot) = \operatorname{argmin}_{g \in \mathcal{G}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \int_{\mathbb{R}} g(w) \hat{f}^{(-k(i))}(w \mid Z_i)\, dw \right)^2 \right\} \tag{9.16}$$

为了在实践中解决 (9.16)，一种方法是用基函数展开近似 $g(w)$，即：

$g_K(w) = \sum_{k=1}^{K} \beta_k \psi_k(w)$ 其中 $\psi_k(\cdot)$ 是一组预先确定的基函数，$g_K(w)$ 随 $K \uparrow$ 更好逼近 $g(w)$.

则有：

$$\hat{\beta} = \operatorname{argmin}_{\beta} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{m}^{(-k(i))}(Z_i)\beta \right)^2 \right\} \tag{9.17}$$

其中 $\hat{m}^{(-k(i))}(Z_i) = \int_{\mathbb{R}} \psi_k(w) \hat{f}^{(-k(i))}(w \mid Z_i)\, dw$

类比 9.13

其为 $g(\cdot)$ 一致估计需要条件，(9.15) 很复杂，用 (9.17) 需仔细正则化，即使如此，

4 types:

	$W_i(1)=0$	$W_i(1)=1$
$W_i(0)=0$	never taker	complier 依从者
$W_i(0)=1$	defier 反抗者	always taker

对 $\forall i$: $W_i(1) \geq W_i(0)$.

假设没有 defier: 单调性 monotonicity

有: $E[Y_i|Z_i=1] - E[Y_i|Z_i=0]$

$$= E[Y_i(W_i(1))|Z_i=1] - E[Y_i(W_i(0))|Z_i=0]$$

$$= E[Y_i(W_i(1)) - Y_i(W_i(0))]$$

$$= E[\mathbb{1}(\{C_i = \text{complier}\})(Y_i(1) - Y_i(0))]$$

若 compliers 一定存在，由 Bayes' rule:

$$\tau_{LATE} = \frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[W_i|Z_i=1] - E[W_i|Z_i=0]}$$

$$= E[Y_i(1) - Y_i(0)|C_i = \text{complier}]. \quad (10.3)$$

LATE (complier average treatment)

PS: 当 (10.1) 不成立，有 nt, at, defier...

$\tau_{ATE} = E[Y_i(1) - Y_i(0)]$

nevertaker 的 $W_i=1$ 观测不到

No.
Date

Lecture 10 Local Average Treatment Effect. 局部平均处理效应

由L9的IV模型：

$$Y_i = \alpha + W_i \tau + \varepsilon_i \quad , \quad \varepsilon_i \perp\!\!\!\perp Z_i$$
$$W_i = Z_i \gamma + \eta_i \qquad (10.1)$$

当 Z_i 为二元：

$$\tau = Cov[Y_i, Z_i] / Cov[W_i, Z_i]$$
$$= \frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[W_i|Z_i=1] - E[W_i|Z_i=0]} \qquad (10.2)$$

Treatment effect estimation under non-compliance 非依从性

非依从：RCT下：

$Y_i \in \mathbb{R}$，通常

实际接受治疗 $W_i \in \{0,1\}$ 非随机，因为有非依从性

分配的治疗 $Z_i \in \{0,1\}$ 随机

实际上，(10.1)中 τ 为常数也许不对，W_i 也许与 τ 相关，等，

需要定义：$\{W_i(0), W_i(1)\} \Rightarrow W_i(Z_i)$ ， $Y_i = Y_i(W_i, Z_i)$

综上：IV有效假设如下：

- Exclusion restriction $Y_i(w,z) = Y_i(w)$ 治疗分配仅通过接受治疗来影响结果
- Exogeneity $\{Y_i(0), Y_i(1), W_i(0), W_i(1)\} \perp\!\!\!\perp Z_i$ 治疗分配是随机的
- Relevance $E[W_i(1) - W_i(0)] \neq 0$ 治疗分配影响到治疗的接受

KOKUYO

No.

Date

Supply and demand. 价格 — 需求

对每个市场 $i=1,\cdots,n$，有一个供应曲线 $S_i(p,z)$ 和一个需求曲线 $Q_i(p,z)$

价格 $p\in\mathbb{R}$，IV: $z\in\{0,1\}$

（假设 $S_i(\cdot,z)$ 连续且单增，$Q_i(\cdot,z)$ 连续且单减）

该假设下：P_i 是 $S_i(P_i,Z_i)=Q_i(P_i,Z_i)$ 的唯一解

即：Z_i 为 IV；P_i 为 treatment；$Q_i(P_i,Z_i)$ 为 outcome (Y_i)

$Z_i \to S \to P_i \to Q_i(p)$

则约束条件：

- 排除限制 Exclusion restriction: $Q_i(p,z)=Q_i(p)$ for all p and z
- 外生 Exogeneity: $\{Q_i(p), S_i(p,z)\} \perp\!\!\!\perp Z_i$
- 相关 Relevance: $Cov[P_i,Z_i]\neq 0$
- 单调 Monotonicity: $S_i(P_i,1)\leq S_i(P_i,0)$ 以概率处处成立

设 $Q_i(p)$ 可微，$Q_i'(p)$ 为其导数（不必须，因为 $Q_i(\cdot)$ 已↑，因此分布导数必存在）

$$\text{ILATE} = \frac{E[Q_i|Z_i=1]-E[Q_i|Z_i=0]}{E[P_i|Z_i=1]-E[P_i|Z_i=0]}$$

$$= \frac{\int E[Q_i'(p)\,|\,P_i(0)\leq p\leq P_i(1)]\,P[P_i(0)\leq p\leq P_i(1)]\,dp}{\int P[P_i(0)\leq p\leq P_i(1)]\,dp}$$

测度积分

推导：

$$E[Q_i|Z_i=1]-E[Q_i|Z_i=0]$$

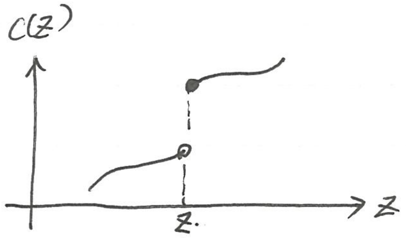
$$= E[Q_i(P_i(1))|Z_i=1]-E[Q_i(P_i(0))|Z_i=0]$$

$$= E[Q_i(P_i(1))-Q_i(P_i(0))]$$

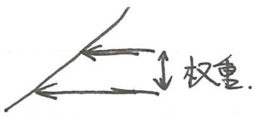
$$= E\left[\int_{P_i(0)}^{P_i(1)} Q_i'(p)\,dp\right] = \underline{\int E[Q_i'(p)\,|\,P_i(0)\leq p\leq P_i(1)]\,P[P_i(0)\leq p\leq P_i(1)]\,dp},$$

交换积分次序

Fubini 定理 ~ 关系



$z \in S$， S为间断点



$$(10.2): \quad \tau LATE = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[W_i|Z_i=1]-E[W_i|Z_i=0]}$$

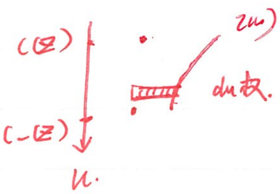
改写为积分：

分子：

$z \in S$的概率

权重.

$$\sum_{z\in S} \left(\underbrace{\int_{c_-(z)}^{c(z)} \tau(u)\,du}\right)\varphi(z) + \int_{R\setminus S} \underbrace{\tau(c(z))}_{u=c(z)\text{处理期望}}\, c'(z)\, \underbrace{\varphi(z)}_{\text{概率}}\, dz$$



$\Rightarrow$

分母.

$$\sum_{z\in S} (c(z)-c_-(z))\varphi(z) + \int_{R\setminus S} c'(z)\varphi(z)\,dz$$ 总权

No.

Date

Threshold crossing and willingness to pay
阈值穿越　　　　支付意愿

假设每个受试者有一个潜在的和内源性的变量 U_i，使得

$$W_i = 1(\{U_i \geq c(Z_i)\})$$

$c(z)$ 为一个依赖于 z 的截止函数

不失一般性，取 $U_i \sim Unif([0,1])$ 均匀分布，可得 $c(z) = 1 - P[W_i=1 | Z_i = z]$

则可产生 IV：

① $\{Y_i(0), Y_i(1)\}$ such that. $Y_i = Y_i(W_i)$　　影响存在且通路唯一

② $\{Y_i(0), Y_i(1), U_i\} \perp Z_i$　　外生

③ $c(Z_i)$ 有一个非平凡的分布　　相关

④ $c(z)$，cadlag（右侧连续），非减的　　$W_i \cdot Z_i$ 非减

定义，marginal treatment effect 边际处理效应。

$$\tau(u) = E[Y_i(1) - Y_i(0) | U_i = u]$$

下证，IV 可找到 $\tau(u)$ 的加权平均，在这里，设 $Z_i \sim N(0,1)$，(10.2) 改写为

$$\tau_{LATE} = \frac{\sum_{z\in S}\left(\int_{c_-(z)}^{c(z)} \tau(u)du\right)\varphi(z) + \int_{R\backslash S} \tau(c(z))c'(z)\varphi(z)dz}{\sum_{z\in S}(c(z) - c_-(z))\varphi(z) + \int_{R\backslash S} c'(z)\varphi(z)dz}$$

其中 $S\subset R$ 是 $c(\cdot)$ 的不连续点集，$c_-(z) = \lim_{a\uparrow z} c(a)$

可看出 τ_{LATE} 是 $\tau(u)$ 的凸平均

注：$c(z)$ 为单调递增，其也必须具有有界的变化。

所以写：$c(z) = c_0 + \int_{-\infty}^{z} c'(a)da$　其中 $c'(z)$ 是某个非负的 Lebesgue 可测函数

KOKUYO

#. $H'(z)$ 表示 $H(\cdot)$ 的分布导数：?

$$-H'(z) = \begin{cases} \left(\int_{c_-(z)}^{c(z)} \tau(u)\,du\right)\delta_z, & \text{for } z \in S, \\ \tau(c(z))\,c'(z) & \text{else.} \end{cases}$$

δ_z 是在 z 处的 Dirac δ 函数.

满足 1. $\delta(z) = 0$ when $z \neq 0$

2. $\int_{-\infty}^{\infty} \delta(z)\,dz = 1$

No.

Date

Example 1. Single jump and Multiple jumps

当 $c(z)$ 为常数，仅有单一跳跃： $c(z) = c_0 + \delta \mathbb{1}\{z \geq z_1\}$

则：

$$\begin{cases} \text{Never taker}: & U_i < c_0 \\ \text{compliers}: & c_0 \leq U_i < c_0 + \delta_1 \\ \text{always takers}: & U_i \geq c_0 + \delta_1 \end{cases}$$

当有K个跳跃： $c(z) = c_0 + \sum_{k=1}^{K} \delta_k \mathbb{1}\{z \geq z_k\}$

则：

$$\tau_{LATE} = \frac{\sum_{k=1}^{K} E[\tau(U_i) \mid c_-(z_k) \leq U_i < c(z_k)] \, (z_k - c_-(z_k))\varphi(z_k)}{\sum_{k=1}^{K} (c(z_k) - c_-(z_k))\varphi(z_k)}$$

在 $c(\cdot)$ 上的 compliers 的 ATE 凸组合，

权重依赖于 u 空间中 stratum 的大小和 $\varphi(z_k)$

Example 2: Continuous cutoff 连续阈值.

若 $c(z)$ 无跳跃：

$$\tau_{LATE} = \int_{\mathbb{R}} \tau(c(z))\, c'(z)\varphi(z)dz \Big/ \int_{\mathbb{R}} c'(z)\varphi(z)dz$$

协方差：

$$\begin{aligned} Cov[Y_i, Z_i] &= Cov[Y_i(0) + (Y_i(1) - Y_i(0))W_i, Z_i] \\ &= Cov[(Y_i(1) - Y_i(0))W_i, Z_i] \\ &= Cov[(Y_i(1) - Y_i(0))\mathbb{1}(\{U_i \geq c(Z_i)\}), Z_i] \\ &= Cov[\tau(U_i)\mathbb{1}(\{U_i \geq c(Z_i)\}), Z_i] \end{aligned}$$

令 $H(z) = E[\tau(U_i)\mathbb{1}(\{U_i \geq c(z)\})]$ ， Z_i 为标准 Gaussian.

$$Cov[H(Z_i), Z_i] = E[H'(Z_i)]$$

$$\tau = \frac{Cov[Y_i, Z_i]}{Cov[W_i, Z_i]} = \frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[W_i|Z_i=1] - E[W_i|Z_i=0]} \quad (10.2)$$

$$\tau(u) = E[Y_i(1) - Y_i(0) | U_i = u] \quad (10.6)$$

Estimating the marginal treatment effect 估计边际处理效应

在 $c(z)$ 的连续点处：

$$\tau(c(z)) = \frac{\frac{d}{dz} E[Y_i|Z_i=z]}{\frac{d}{dz} P[W_i=1|Z_i=z]} \quad (10.12)$$

假设比率导数良好
(分子分母可微，分母非0)

下证 (10.12).

$$E[Y_i|c(Z_i)=c] = E[Y_i(0) + \mathbb{1}(U_i \geq c)(Y_i(1) - Y_i(0)) | c(Z_i) = c]$$

$$= E[Y_i(0) + \mathbb{1}(U_i \geq c)(Y_i(1) - Y_i(0))]$$

$$= E[Y_i(0)] + \int_c^1 \tau(u)du \quad (Fabini)$$

$$\Rightarrow \tau(c) = -\frac{d}{dc} E[Y_i|c(Z_i)=c] \checkmark$$

链式法则：

$$\frac{d}{dz} E[Y_i|Z_i=z] = \frac{d}{dc} E[Y_i|c(Z_i)=c] c'(z) \checkmark$$

又假设 $U_i \sim U(0,1) \perp Z_i$ $\quad c'(z) = -\frac{d}{dz} P[W_i=1|Z_i=z] \checkmark$

联立. (10.12) 得 ✓