第三章　观察性研究

罗姗姗

北京工商大学 数学与统计学院

因果推断课题组

目录

目录

倾向得分

$$
\begin{tikzcd}
 & X \arrow[dl] \arrow[dr] & & U \arrow[ll] \arrow[dl] \\
Z \arrow[rr] & & Y &
\end{tikzcd}
$$

Rosenbaum and Rubin (1983) 提出了“倾向得分”概念，并讨论了其在观测性研究中的因果推断中的作用。这篇论文被广泛引用，被认为是统计学领域中的经典之作。

Assumption 1 (positivity, or overlap)

$$0 < \mathrm{pr}(Z = 1 \mid X) < 1.$$

- 上述概率 $e(x) = \mathrm{pr}(Z = 1 \mid X = x)$ 常被称作倾向得分 (propensitity score).
- 在协变量 X 的每一层里，该假设都要求存在接受处理或接受对照的个体.
- 该假设可以直接检验.

逆概加权 I

定理 1

在给定可忽略性假定下，即 $Z \perp\!\!\!\perp \{Y(1), Y(0)\} \mid X$ 及 $0 < e(X) < 1$，我们有

$$\mathbb{E}\{Y(1)\} = \mathbb{E}\left\{\frac{ZY}{e(X)}\right\}, \quad \mathbb{E}\{Y(0)\} = \mathbb{E}\left\{\frac{(1-Z)Y}{1-e(X)}\right\},$$

$$\tau = \mathbb{E}\{Y(1) - Y(0)\} = \mathbb{E}\left\{\frac{ZY}{e(X)} - \frac{(1-Z)Y}{1-e(X)}\right\}$$

逆概加权 II

证明.

$$
\begin{aligned}
\mathbb{E}\left\{\frac{ZY}{e(X)}\right\} =& \mathbb{E}\left\{\frac{ZY(1)}{e(X)}\right\} \\
=& \mathbb{E}\left[\mathbb{E}\left\{\frac{ZY(1)}{e(X)} \mid X\right\}\right] \\
=& \mathbb{E}\left[\frac{1}{e(X)}\mathbb{E}\{ZY(1) \mid X\}\right] \\
=& \mathbb{E}\left[\frac{1}{e(X)}\mathbb{E}(Z \mid X)\mathbb{E}\{Y(1) \mid X\}\right] \\
=& \mathbb{E}\left[\frac{1}{e(X)}e(X)\mathbb{E}\{Y(1) \mid X\}\right] \\
=& \mathbb{E}[\mathbb{E}\{Y(1) \mid X\}] \\
=& \mathbb{E}\{Y(1)\}
\end{aligned}
$$

$\square$

逆概加权估计 (Inverse probability weighting, IPW) I

- 上述定理启发我们考虑如下矩估计量：

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

  其中 $\hat{e}(X_i)$ 是估计的倾向得分。

- 上述估计量被称为逆概加权估计估计量，也称为 Horvitz-Thompson（HT）估计量。
- Horvitz and Thompson (1952) 在调查抽样中提出了这一方法，Rosenbaum (1987) 则在观测性研究中使用了它。

逆概加权估计 (Inverse probability weighting, IPW) II

IPW 估计相当于对样本做了权重调整,

$$\widehat{\tau}^{\mathrm{ht}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Z_i Y_i}{\widehat{e}(X_i)} - \frac{(1-Z_i)Y_i}{1-\widehat{e}(X_i)}\right\},$$

对于处理组的个体 i, 我们采用权重 $\widehat{e}(X_i)^{-1}$ 进行调整.

对于对照组的个体 i, 我们采用权重 $\{1-\widehat{e}(X_i)\}^{-1}$ 进行调整.

- 当 X 是离散或维数较低时, 我们可以采用非参数估计等方法估计 $\widehat{e}(X)$.
- 当 X 维数较高时, 我们可以采用对 $e(X) = \mathrm{pr}(Z=1 \mid X)$ 建立参数模型 $e(X;\beta)$, 并通过极大似然估计或矩估计等方法求解参数 $\widehat{\beta}$.

值得注意的是, 当 $e(X;\beta)$ 被错误指定时, $\widehat{\tau}_{\mathrm{ipw}}$ 并不能相合估计 τ.

逆概加权估计 (Inverse probability weighting, IPW) III

- Hirano et al. (2003) 指出在 IPW 估计时, 采用非参数估计的倾向得分 $\widehat{e}(X)$ 将比使用正确的倾向得分的 $e(X)$ 具有更小的渐进方差.
- 倾向评分一个众所周知的结果: 使用估计的倾向评分通常比真实的倾向评分有更好的 ATE 的估计 (Rosenbaum, 1987).
- Rosenbaum (1987, pp 391) 给出了一些直观的解释:

  " the same reason that covariate adjustment in RCT outperforms the unadjusted difference-in-means estimator –estimated PS corrects for chance imbalance in the sample, but true PS does not. "

- Hirano et al. (2003) 指出在 IPW 估计时, 采用非参数估计的倾向得分 $\widehat{e}(X)$ 将比使用正确的倾向得分的 $e(X)$ 具有更小的渐进方差.

逆概加权估计 (Inverse probability weighting, IPW) IV

- 然而，估计量 $\hat{\tau}^{\mathrm{ht}}$ 存在许多问题，经常会由于极端权重使得估计非常不稳定。
- Hájek (1971) 基于倾向得分也提出一类权重估计量，

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

- 上述估计量被称为 Hájek 估计量，与 HT 估计量相比，可以发现权重被重新调整。
- 许多数值研究发现，$\hat{\tau}^{\mathrm{hajek}}$ 比 $\hat{\tau}^{\mathrm{ht}}$ 有着更加稳健的估计。

目录

因果作业的两个识别表达式 I

在无混杂性（unconfoundedness）假设 $Z \perp\!\!\!\perp \{Y(1), Y(0)\} \mid X$ 和正数性（positivity、overlap）假设 $0 < e(X) < 1$ 下，我们曾讨论过平均因果作用 $\tau = \mathbb{E}\{Y(1) - Y(0)\}$ 的两个识别公式。

- 第一个是基于回归模型的公式：

$$\tau = \mathbb{E}\{\mu_1(X)\} - \mathbb{E}\{\mu_0(X)\}$$

其中

$$\mu_z(X) = \mathbb{E}\{Y(z) \mid X\} = \mathbb{E}(Y \mid Z = z, X)$$

是给定协变量时的两个条件均值函数。

因果作业的两个识别表达式 II

- 第二个是基于倾向得分的逆概加权（IPW）公式：

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

其中

$$e(X) = \mathrm{pr}(Z = 1 \mid X)$$

是倾向得分。

因果作业的两个识别表达式 III

- 基于回归方法的模型需要 $\mu_z(X) = \mathbb{E}(Y \mid Z = z, X)$ 被正确指定，当 $\mu_z(X)$ 被正确指定时，平均因果作用可以被相合估计。
- 基于倾向得分的逆概加权估计需要 $e(X) = \mathrm{pr}(Z = 1 \mid X)$ 被正确指定，当 $e(X)$ 被正确指定时，平均因果作用可以被相合估计。
- 理论上，基于上述两种估计方法我们可以构造无数多种平均因果作用的估计式。
- 这驱使我们去构造一种更有原则的估计方法，使得在回归模型或倾向得分模型任一被正确指定时都能达到稳健估计，这就是所谓的双稳健估计量。

双稳健估计 (Doubly robust estimator) I

- 我们定义如下估计量：

$$
\begin{aligned}
\tilde{\mu}_1^{\mathrm{dr}} &= \mathbb{E}\left[\frac{Z\{Y-\mu_1(X,\beta_1)\}}{e(X,\alpha)}+\mu_1(X,\beta_1)\right], \\
\tilde{\mu}_0^{\mathrm{dr}} &= \mathbb{E}\left[\frac{(1-Z)\{Y-\mu_0(X,\beta_0)\}}{1-e(X,\alpha)}+\mu_0(X,\beta_0)\right].
\end{aligned} \tag{1}
$$

  这也可以写成以下形式：

$$
\begin{aligned}
\tilde{\mu}_1^{\mathrm{dr}} &= \mathbb{E}\left[\frac{ZY}{e(X,\alpha)}-\frac{Z-e(X,\alpha)}{e(X,\alpha)}\mu_1(X,\beta_1)\right], \\
\tilde{\mu}_0^{\mathrm{dr}} &= \mathbb{E}\left[\frac{(1-Z)Y}{1-e(X,\alpha)}-\frac{e(X,\alpha)-Z}{1-e(X,\alpha)}\mu_0(X,\beta_0)\right].
\end{aligned} \tag{2}
$$

- 公式(1)以回归模型为基础，增加了回归模型估计的残差部分。公式(2)以逆概加权估计量为基础，增加了倾向得分估计残差的部分。
- 因此，双稳健估计量也被称为增强逆概加权估计量（augmented inverse propensity score weighting, AIPW）。

性质 I

定理 2

假设无混杂性 $Z \perp\!\!\!\perp \{Y(1), Y(0)\} \mid X$ 及 $0 < e(X) < 1$。

1. 如果 $e(X, \alpha) = e(X)$ 或 $\mu_1(X, \beta_1) = \mu_1(X)$，则 $\tilde{\mu}_1^{\mathrm{dr}} = \mathbb{E}\{Y(1)\}$.
2. 如果 $e(X, \alpha) = e(X)$ 或 $\mu_0(X, \beta_0) = \mu_0(X)$，则 $\tilde{\mu}_0^{\mathrm{dr}} = \mathbb{E}\{Y(0)\}$.
3. 如果 $e(X, \alpha) = e(X)$ 或 $\{\mu_1(X, \beta_1) = \mu_1(X), \mu_0(X, \beta_0) = \mu_0(X)\}$，则 $\tilde{\mu}_1^{\mathrm{dr}} - \tilde{\mu}_0^{\mathrm{dr}} = \tau$.

根据上述定理，如果倾向得分模型或结果模型任一被正确指定，那么 $\tilde{\mu}_1^{\mathrm{dr}} - \tilde{\mu}_0^{\mathrm{dr}}$ 等于 τ。这就是它被称为双稳健估计量的原因。

性质 II

（仅证明 $\mu_1 = \mathbb{E}\{Y(1)\}$）：

证明.

$$
\begin{aligned}
\tilde{\mu}_1^{\mathrm{dr}} - \mathbb{E}\{Y(1)\} &= \mathbb{E}\left[\frac{Z\{Y(1)-\mu_1(X,\beta_1)\}}{e(X,\alpha)} - \{Y(1)-\mu_1(X,\beta_1)\}\right] \\
&= \mathbb{E}\left[\frac{Z-e(X,\alpha)}{e(X,\alpha)}\{Y(1)-\mu_1(X,\beta_1)\}\right] \\
&= \mathbb{E}\left(\mathbb{E}\left[\frac{Z-e(X,\alpha)}{e(X,\alpha)}\{Y(1)-\mu_1(X,\beta_1)\}\,\middle|\,X\right]\right) \\
&= \mathbb{E}\left[\mathbb{E}\left\{\frac{Z-e(X,\alpha)}{e(X,\alpha)}\,\middle|\,X\right\} \times \mathbb{E}\left\{Y(1)-\mu_1(X,\beta_1) \mid X\right\}\right] \\
&= \mathbb{E}\left[\frac{e(X)-e(X,\alpha)}{e(X,\alpha)} \times \{\mu_1(X)-\mu_1(X,\beta_1)\}\right].
\end{aligned}
$$

于是当 $e(X) = e(X,\alpha)$ 或 $\mu_1(X) = \mu_1(X,\beta_1)$ 时，我们有 $\tilde{\mu}_1^{\mathrm{dr}} - \mathbb{E}\{Y(1)\} = 0$. $\square$

双稳健估计量 I

通过以上定理，我们可以如下构建样本版本：

1. 拟合倾向得分：$e(X,\hat{\alpha})$；
2. 拟合结果均值：$\mu_1(X,\hat{\beta}_1)$ 和 $\mu_0(X,\hat{\beta}_0)$；
3. 构建双重稳健估计量：$\hat{\tau}^{\mathrm{dr}}=\hat{\mu}_1^{\mathrm{dr}}-\hat{\mu}_0^{\mathrm{dr}}$，其中

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

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

4. 使用 Bootstrap 方法非参数估计 $\hat{\tau}^{\mathrm{dr}}$ 的方差。

稳健估计

- 当任一模型正确指定时，双稳健估计仍具有相合估计.
- 当两个模型都不正确时，双稳健估计可能会带来比回归和逆概加权估计更大的偏差 (Kang and Schafer, 2007).
- 双稳健估计量在近些年得到了广泛的应用.
- 在可忽略性假定不满足的时候，许多文献也建立 ATE 的识别性，并在不同的识别假设下考虑了 ATE 的稳健估计.
- 值得注意的是，许多稳健估计量需要的模型通常难以正确指定.
- 除了参数模型之外，许多非参数的方法也被用于稳健估计.

机器学习的一些方法

- 机器学习模型经常用于预测，并不能直接用于预测反事实结果 (这是因果推断关心的).
- 机器学习的一些重要思想:
  - Sample splitting, 以便构建模型和估计因果作用.
  - Double learning, 分别学习倾向得分和结果变量模型，并结合两个模型的优势用于因果推断.
  - Cross-fitting, 数据交叉验证来控制偏差和方差.
- 一些有代表性的工作:
  - 因果树和因果森林 (Causal trees and forest, Wager & Athey, 2018).
  - 双重机器学习 (Double/debiased machine learning, Chenozhukov et al., 2018).
  - BART (Bayesian additive regression trees, Chipman et al., 2010).
- 机器学习的方法仍不能解决因果推断的一些基本问题，例如可忽略性假定.

数值模拟

我们评估 $\hat{\tau}^{\mathrm{reg}}$, $\hat{\tau}^{\mathrm{ht}}$ 以及 $\hat{\tau}^{\mathrm{dr}}$ 的有限样本表现.

- $X \sim N(0,1)$
- $\mathrm{pr}(Z=1 \mid X) = \mathrm{expit}(0.5X - 0.2X^2) \quad \Rightarrow \quad e(X) = \mathrm{expit}(\beta_0 + \beta_1 X + \beta_2 X^2)$
- $Y = 1.5Z + 2X + X^2 + \varepsilon,\ \varepsilon \sim N(0,1) \quad \Rightarrow \quad \mu_z(X) = \alpha_1 + \alpha_2 X + \alpha_3 X^2$

Case 1: 所有模型都被正确指定, 即

$$e(X;\beta) = e(X); \quad \mu_z(X;\alpha) = \mu_z(X).$$

Case 2: 回归模型被正确指定, 倾向得分被错误指定, 即

$$e(X;\beta) = \mathrm{expit}(\beta_0 + \beta_1 X); \quad \mu_z(X;\alpha) = \mu_z(X).$$

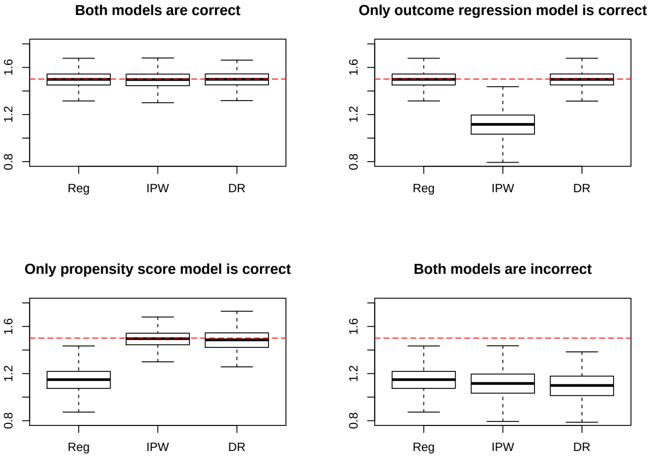
Case 3: 倾向得分模型被正确指定, 回归模型被错误指定, 即

$$e(X;\beta) = \mathrm{expit}(X); \quad \mu_z(X;\alpha) = \alpha_1 + \alpha_2 X.$$

Case 4: 所有模型都被错误指定, 即

$$e(X;\beta) = \mathrm{expit}(\beta_0 + \beta_1 X); \ \mu_z(X;\alpha) = \alpha_1 + \alpha_2 X.$$

稳健估计


Both models are correct
Only outcome regression model is correct
Only propensity score model is correct
Both models are incorrect
0.8
1.2
1.6
Reg
IPW
DR

目录

倾向得分重要的性质

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ Z \arrow[rr] & & Y \end{tikzcd}$$

$$\begin{tikzcd} X \arrow[d] \arrow[ddr] & \\ e(X) \arrow[d] & \\ Z \arrow[r] & Y \end{tikzcd}$$

定理 3 (倾向得分定理, Rosenbaum and Rubin (1983))

在可忽略性假设下, 即

$$\text{(i) } Z \perp\!\!\!\perp \{Y(0), Y(1)\} \mid X; \quad 0 < \mathrm{pr}(Z = 1 \mid X) < 1,$$

$$\Longrightarrow \text{(ii) } Z \perp\!\!\!\perp \{Y(0), Y(1)\} \mid e(X); \quad 0 < \mathrm{pr}\{Z = 1 \mid e(X)\} < 1.$$

平衡得分 (balancing score), 降维 (dimension reduction), 分层 (stratification), 匹配 (matching)...

匹配 (Matching)

- 通过回归估计模型，

  对于处理组的个体 i, 我们使用 $\widehat{\mu}_0(X_i)$ "填补" 个体 i 在接受对照时的潜在结果 $Y_i(0)$.

  对于对照组的个体 i, 我们使用 $\widehat{\mu}_1(X_i)$ "填补" 个体 i 在接受治疗时的潜在结果 $Y_i(1)$.

- 我们考虑匹配估计，

  对于处理组的个体 i, 我们可以在对照组里找与 X_i 最接近的个体的结果用于 "填补" 个体 i 在接受对照时的潜在结果 $Y_i(0)$.

  对于对照组的个体 i, 我们可以在处理组里找与 X_i 最接近的个体的结果用于 "填补" 个体 i 在接受治疗时的潜在结果 $Y_i(0)$.

- 匹配的目的在于平衡处理组和对照组协变量的分布.

- 匹配的思想也可用于估计 ATT 和 ATC.

最近邻匹配

个体 i 根据 L_2 范数匹配的集合定义为:

$$J_M(i) = \left\{ j = 1, \ldots, n : A_j = 1 - Z_i \text{ and } \sum_{k:A_k=1-Z_i} \delta\left(\|X_i - X_k\| \leqslant \|X_i - X_j\|\right) \leqslant M \right\},$$

- 其中 M 为整数, 代表每一个个体的匹配数据的个数.
- $J_M(i)$ 的定义允许在构造匹配集合过程中放回已被使用的个体，不同的个体可以选择相同的协变量进行匹配.

我们可以如下估计 ATE:

$$\widehat{\tau}_{\mathrm{mat}} = \sum_{i=1}^{n} \frac{\widehat{Y}_i(1)}{n} - \sum_{i=1}^{n} \frac{\widehat{Y}_i(0)}{n},$$

上述估计量被称为匹配估计量，其中

$$\widehat{Y}_i(0) = \begin{cases} \frac{1}{M} \sum_{j \in J_M(i)} Y_j, & Z_i = 1, \\ Y_i, & Z_i = 0. \end{cases} \quad \widehat{Y}_i(1) = \begin{cases} Y_i, & Z_i = 1, \\ \frac{1}{M} \sum_{j \in J_M(i)} Y_j, & Z_i = 0. \end{cases}$$

倾向得分重要的性质

证明.

我们只要证明

$$\mathrm{pr}(Z=1 \mid Y(1), e(X)) = \mathrm{pr}(Z=1 \mid e(X))$$

$$\begin{aligned}
\mathsf{pr}(Z=1 \mid Y(1), e(X)) &= \mathbb{E}\{Z \mid Y(1), e(X)\} \\
&= \mathbb{E}[\mathbb{E}\{Z \mid Y(1), e(X), X\} \mid Y(1), e(X)] \\
&= \mathbb{E}\{\mathbb{E}(Z \mid Y(1), X) \mid Y(1), e(X)\} \\
&= \mathbb{E}\{\mathbb{E}(Z \mid X) \mid Y(1), e(X)\} \\
&= \mathbb{E}\{e(X) \mid Y(1), e(X)\} \\
&= e(X) \\
&= \mathsf{pr}(Z=1 \mid e(X))
\end{aligned}$$

其中最后一个等式是由于 □

倾向得分重要的性质

证明.

我们有 $\mathrm{pr}\{Z=1 \mid X, e(X)\} = \mathrm{pr}(Z=1 \mid X) = e(X)$, 及

$$\mathrm{pr}\{Z=1 \mid e(X)\} = \mathbb{E}\left[\mathrm{pr}\{Z=1 \mid X, e(X)\} \mid e(X)\right] = \mathbb{E}\{e(X) \mid e(X)\} = e(X)$$

因此, $\mathrm{pr}\{Z=1 \mid X, e(X)\} = \mathrm{pr}\{Z=1 \mid e(X)\}$, 即 $X \perp\!\!\!\perp Z \mid e(X)$. □

再谈匹配

- 当 X 维度比较高的时候，使用高维协变量进行匹配会比较困难. 根据倾向得分定理，我们可以使用倾向得分进行匹配，

$$J_M(i) = \left\{ j = 1, \dots, n : A_j = 1 - Z_i \text{ and } \sum_{k: A_k = 1 - Z_i} \delta\left(|e(X_i) - e(X_k)| \leqslant |e(X_i) - e(X_j)|\right) \leqslant M \right\}.$$

- 我们可以如下估计 ATE: $$\widehat{\tau}_{\mathrm{mat}} = \sum_{i=1}^{n} \frac{\widehat{Y}_i(1)}{n} - \sum_{i=1}^{n} \frac{\widehat{Y}_i(0)}{n},$$

  其中
$$\widehat{Y}_i(0) = \begin{cases} \frac{1}{M} \sum_{j \in J_M(i)} Y_j, & Z_i = 1, \\ Y_i, & Z_i = 0. \end{cases} \quad \widehat{Y}_i(1) = \begin{cases} Y_i, & Z_i = 1, \\ \frac{1}{M} \sum_{j \in J_M(i)} Y_j, & Z_i = 0. \end{cases}$$

- Abadie and Imbens (2006, 2016) 讨论了上述估计量的渐进性质.

再谈处理组的因果作用

- 由定义，ATE, ATT 和 ATC 三者的关系如下：

$$\tau = \mathrm{pr}\,(Z=1)\,\tau_{\mathrm{ATT}} + \mathrm{pr}\,(Z=0)\,\tau_{\mathrm{ATC}}$$

- 在随机化下，我们有 ATE = ATT = ATC, 因为

$$\underbrace{E\,(Y(1)-Y(0))}_{\tau} = \underbrace{E\,(Y(1)-Y(0)\mid Z=1)}_{\tau_{\mathrm{ATT}}} = \underbrace{E\,(Y(1)-Y(0)\mid Z=0)}_{\tau_{\mathrm{ATC}}}$$

- 在观察性研究中，ATE 一般和 ATT 与 ATC 是不同的.
- 我们接下来讨论 ATT 的识别条件，ATC 类似.

处理组的因果作用

Assumption 2 (可忽略性, ignorability)

(i) $Z \perp\!\!\!\perp Y(0) \mid X$; (ii) $\mathrm{pr}(Z=1 \mid X) < 1$.

$$
\begin{tikzcd}
 & X \arrow[dl] \arrow[dr] & \\
Z \arrow[rr] & & Y
\end{tikzcd}
$$

当我们在讨论 ATT 的识别性时，可忽略性和 positivity 假设都可以被放松.

在上述可忽略性假定下，ATT 可以如下识别：

$$
\begin{aligned}
\tau_{\mathrm{ATT}} &= \mathbb{E}\left(Y(1) - Y(0) \mid Z=1\right) \\
&= \mathbb{E}\left(Y(1) \mid Z=1\right) - \mathbb{E}\left\{E\left(Y(0) \mid Z=1, X\right) \mid Z=1\right\} \\
&= \mathbb{E}(Y \mid Z=1) - \mathbb{E}\left\{E\left(Y(0) \mid Z=1, X\right) \mid Z=1\right\} \\
&= \mathbb{E}(Y \mid Z=1) - \mathbb{E}\left\{E\left(Y(0) \mid Z=0, X\right) \mid Z=1\right\} \\
&= \mathbb{E}(Y \mid Z=1) - \mathbb{E}\{E(Y \mid Z=0, X) \mid Z=1\}.
\end{aligned}
$$

References I

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267.

Abadie, A. and Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica*, 84(2):781–807.

Hájek, J. (1971). Comment on a paper by d. basu. in foundations of statistical inference. toronto: Holt, rinehart and winston.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685.

Kang, J. D. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4):523–539.

References II

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American statistical Association*, 82(398):387–394.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

本周作业

1. 证明逆概加权识别表达式
2. 习题 13.3 或习题 13.9
3. 上机作业：运行代码 13.3 及习题 13.2

   作业扫描成 pdf 发邮箱:smengchen@163.com