第三章　观察性研究

罗姗姗

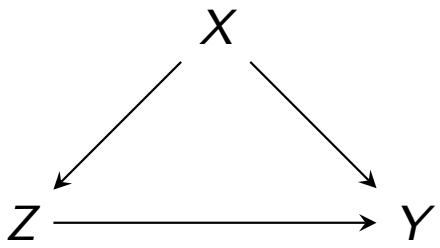
北京工商大学 数学与统计学院

因果推断课题组

目录

目录

观察性研究



实际研究中的数据经常并不是随机化的.

- 在观察性研究中，处理分配机制是未知的.
- 在观察性研究中，Z 不再随机化时，可能会：
  - $Z \not\perp\!\!\!\perp X$.

    这意味着处理组和对照组的协变量分布不再均匀.
  - $Z \not\perp\!\!\!\perp \{Y(0), Y(1)\}$.

    一些不可检验的假设将被引入用于 ATE 的识别.
- 在观察性研究中，相关一般不能表示因果.

记号

- 我们将所有的受试者记为 $i(i = 1, \ldots, n)$，并用 X_i 表示个体 i 的基线协变量、Z_i 表示个体 i 的原因变量（二值）、Y_i 表示个体 i 的结果变量。
- 我们仍用记号 $Y_i(1)$ 和 $Y_i(0)$ 表示个体 i 在处理组和对照组下的潜在结果。
- 为简单起见，我们假设每个个体是独立同分布的（independent and identically distributed, IID）：

$$\{X_i, Z_i, Y_i(1), Y_i(0)\}_{i=1}^{n} \overset{\text{IID}}{\sim} \{X, Z, Y(1), Y(0)\}$$

因此，我们可以省略下标 i。

感兴趣的因果作用 I

我们感兴趣的因果作用包括：

- 平均因果作用（总体因果作用）：

$$\tau = E\{Y(1) - Y(0)\}$$

- 处理组平均因果作用（average causal effect of on the treated ）：

$$\tau_{\mathrm{T}} = E\{Y(1) - Y(0) \mid Z = 1\}$$

- 对照组平均因果作用（average causal effect on the control）：

$$\tau_{\mathrm{C}} = E\{Y(1) - Y(0) \mid Z = 0\}$$

因果作用和随机化的重要性 I

- 通过期望的线性性质，我们有：

$$\begin{aligned}\tau_{\mathrm{T}} &= E\{Y(1) \mid Z=1\} - E\{Y(0) \mid Z=1\} \\ &= E(Y \mid Z=1) - E\{Y(0) \mid Z=1\}. \\ \tau_{\mathrm{C}} &= E\{Y(1) \mid Z=0\} - E\{Y(0) \mid Z=0\} \\ &= E\{Y(1) \mid Z=0\} - E(Y \mid Z=0).\end{aligned}$$

- 在上述 τ_{T} 和 τ_{C} 的两个公式中，$E(Y \mid Z=1)$ 和 $E(Y \mid Z=0)$ 可以直接从数据中观测得到
- 但 $E\{Y(0) \mid Z=1\}$ 和 $E\{Y(1) \mid Z=0\}$ 不能
- 后两者是反事实的，因为它们是与实际接受的处理相反的潜在结果的均值。

因果作用和随机化的重要性 II

- 简单的均值差异可以如下表示：

$$\begin{aligned}\tau_{\mathrm{PF}} &= E(Y \mid Z=1) - E(Y \mid Z=0) \\ &= E\{Y(1) \mid Z=1\} - E\{Y(0) \mid Z=0\}\end{aligned}$$

通常对上述定义的因果作用具有偏差。例如，

$$\tau_{\mathrm{PF}} - \tau_{\mathrm{T}} = \underbrace{E\{Y(0) \mid Z=1\} - E\{Y(0) \mid Z=0\}}_{\textit{selection bias}},$$

$$\tau_{\mathrm{PF}} - \tau_{\mathrm{C}} = \underbrace{E\{Y(1) \mid Z=1\} - E\{Y(1) \mid Z=0\}}_{\textit{selection bias}}.$$

一般不为零，它们量化了选择偏差。它们衡量了处理组和对照组之间潜在结果均值的差异。

因果作用和随机化的重要性 III

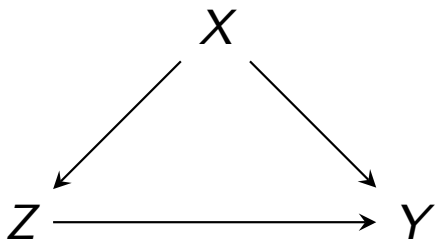
- 但在随机化下，由于 $Z \perp\!\!\!\perp \{Y(0), Y(1)\}$，会有

$$\tau_{\mathrm{PF}} - \tau_{\mathrm{T}} = E\{Y(0) \mid Z = 1\} - E\{Y(0) \mid Z = 0\} = 0,$$
$$\tau_{\mathrm{PF}} - \tau_{\mathrm{C}} = E\{Y(1) \mid Z = 1\} - E\{Y(1) \mid Z = 0\} = 0,$$

  也就是 $\tau_{\mathrm{PF}} = \tau_{\mathrm{T}} = \tau_{\mathrm{C}} = \tau$.

- 从上面的讨论可以看出，随机化的基本好处在于平衡处理组和对照组的潜在结果分布，这比平衡观测到的协变量分布更为重要。
- 没有随机化，选择偏差项可以特别大，尤其是对于无界的结果变量。
- 这凸显了在观测性研究中进行因果推断的根本困难。

经典案例 – Berkeley 录取率 I



实际研究中的数据经常

并不是随机化的.

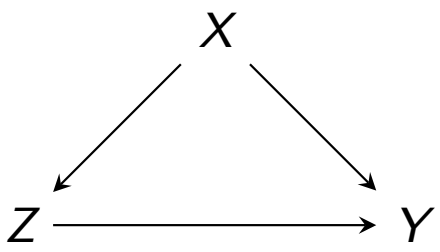
	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
Total	12,763	41%	8,442	44%	4,321	35%

- Z: 二值处理变量, $Z=1$ 代表男性, $Z=0$ 代表女性.
- Y: 二值结果变量, $Y=1$ 代表录取, $Y=0$ 代表未录取.
- L: 协变量, 表示专业.

ATE 可以如下估计吗？

$$\widehat{\tau} = \widehat{E}(Y \mid Z=1) - \widehat{E}(Y \mid Z=0) = \widehat{\text{pr}}(Y=1 \mid Z=1) - \widehat{\text{pr}}(Y=1 \mid Z=0)$$
$$= 9\%.$$

经典案例 – Berkeley 录取率 II



实际研究中的数据经常

并不是随机化的.

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
A	933	64%	**825**	62%	108	82%
B	585	63%	**560**	63%	25	68%
C	918	35%	325	37%	**593**	34%
D	792	34%	417	33%	375	35%
E	584	25%	191	28%	**393**	24%
F	714	6%	373	6%	341	7%
Total	**4526**	**39%**	**2691**	**45%**	**1835**	**30%**

Legend:

greater percentage of successful applicants than the other gender

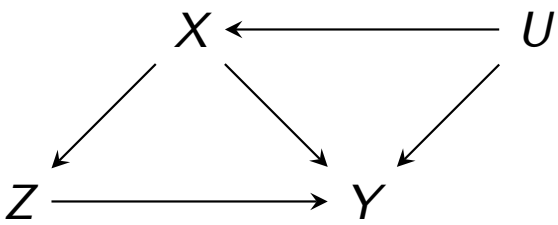
greater number of applicants than the other gender

bold – the two 'most applied for' departments for each gender

(部分专业, 图源来自 Wikipedia)

目录

观察性研究 I



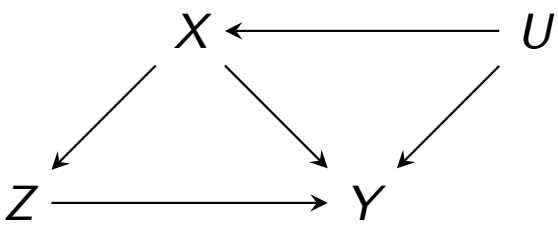
Rosenbaum and Rubin (1983) 提出的可忽略性假定是在观察性研究里评估因果作用最重要的假定.

Assumption 1 (可忽略性, ignorability)

$$Z \perp\!\!\!\perp \{Y(0), Y(1)\} \mid X.$$

- $\text{pr}\{Z \mid X, Y(0), Y(1)\} = \text{pr}(Z \mid X)$.
- 在协变量的每个分层下，处理分配机制可以被视为随机化.
- 随机化试验也满足可忽略性假设.
- 可忽略性又被称作无混杂 (unconfoundedness) 假设，所有未观测的混杂不同时指向处理变量和结果变量.
- 该假设无法直接检验.

观察性研究 II



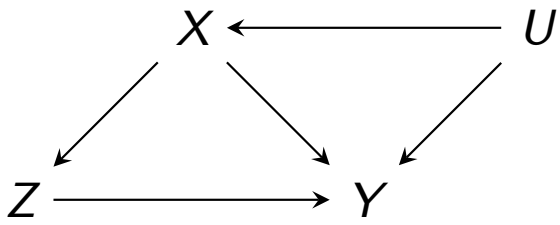
Rosenbaum and Rubin (1983) 提出的可忽略性假定是在观察性研究里评估因果作用最重要的假定.

Assumption 2 (positivity, or overlap)

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

- 上述概率 $e(x) = \text{pr}(Z = 1 \mid X = x)$ 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.

识别性



Rosenbaum and Rubin (1983) 提出的可忽略性假定是在观察性研究里评估因果作用最重要的假定.

Assumption 3 (强可忽略性, strong ignorability)

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

在强可忽略性假定下, ATE 可以如下识别:

$$
\begin{aligned}
\tau &= E\{Y(1) - Y(0)\} \\
&= E\{E(Y(1) - Y(0) \mid X)\} \\
&= E\{E(Y(1) \mid X)\} - E\{E(Y(0) \mid X)\} \\
&= E\{E(Y(1) \mid Z = 1, X)\} - E\{E(Y(0) \mid Z = 0, X)\} \\
&= E\{E(Y \mid Z = 1, X)\} - E\{E(Y \mid Z = 0, X)\}.
\end{aligned}
$$

如果不对混杂因素 X 进行调整, 将会导致有偏估计.

其他因果量的可识别性 I

在强可忽略性假设下，我们可以识别处理组平均因果作用 τ_{T} 以及对照组平均因果作用 τ_{C}。此外，我们也可以识别如下因果量：

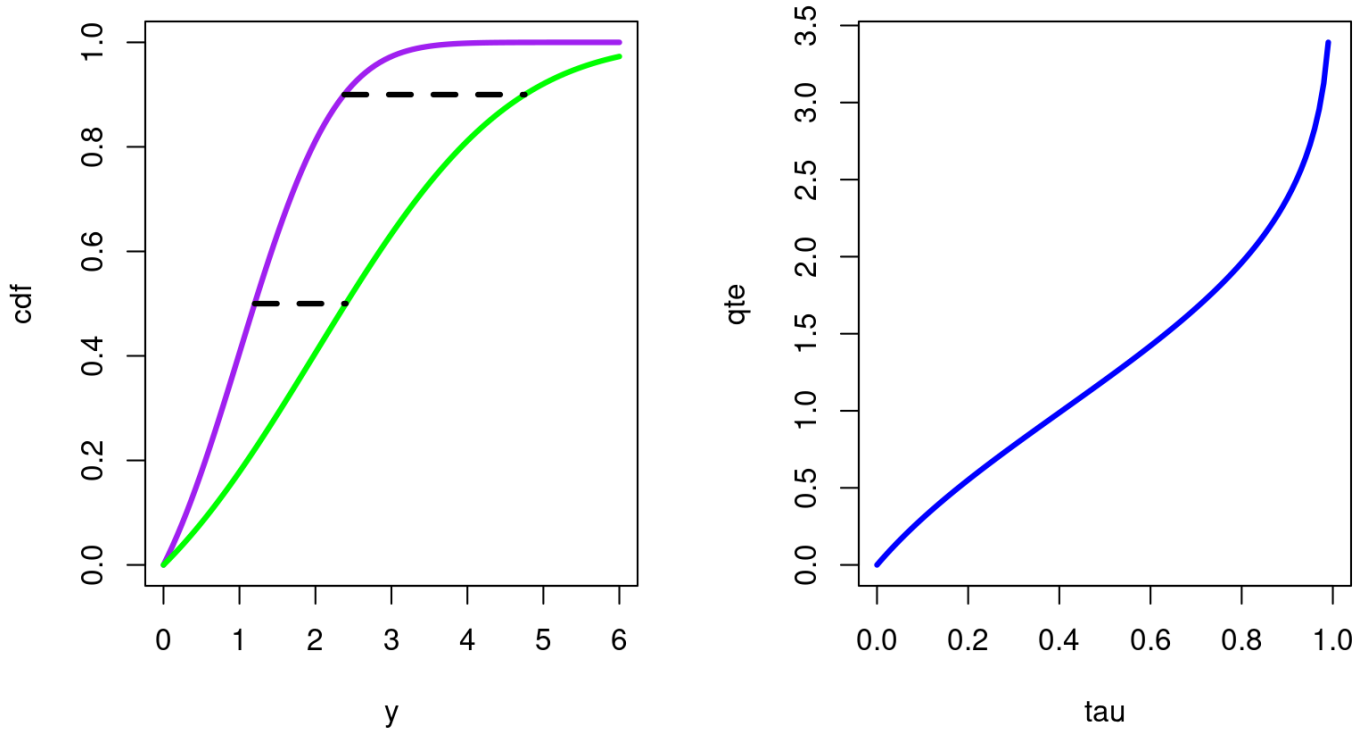
- 分布因果作用（distributional causal effect）：

$$\mathrm{DCE}_y = \mathrm{pr}\{Y(1) > y\} - \mathrm{pr}\{Y(0) > y\}$$

以 $\mathrm{pr}\{Y(1) > y\}$ 为例，

$$\mathrm{pr}\{Y(1) > y \mid X\} = \mathrm{pr}(Y > y \mid Z = 1, X),$$
$$\mathrm{pr}\{Y(1) > y\} = E_X\{\mathrm{pr}(Y > y \mid Z = 1, X)\}.$$

其他因果量的可识别性 II



- 分位数因果作用 (quantile causal effect)：

$$\mathrm{QCE}_q = \mathrm{quantile}_q\{Y(1)\} - \mathrm{quantile}_q\{Y(0)\}$$

其中 $\mathrm{quantile}_q(U)$ 表示随机变量 U 对应的分布的第 q 个分位数点。

目录

离散时的分层估计

当 X 是离散取值时，假设取值范围为 $X = 1, \ldots, K$.

$$\begin{aligned}\tau &= E\{Y(1) - Y(0)\} \\ &= E\{E(Y \mid Z = 1, X)\} - E\{E(Y \mid Z = 0, X)\} \\ &= \sum_{k=1}^{K} \underbrace{E(Y \mid Z = 1, X = k)}_{\mu(1,k)} \mathrm{pr}(X = k) - \sum_{k=1}^{K} \underbrace{E(Y \mid Z = 0, X = k)}_{\mu(0,k)} \mathrm{pr}(X = k).\end{aligned}$$

1. $\mu(a, k)$ 可以如下估计: $\widehat{\mu}(a, k) = \dfrac{\sum_{i=1}^{n} Y_i \mathbb{I}(Z_i = a, X_i = k)}{\sum_{i=1}^{n} \mathbb{I}(Z_i = a, X_i = k)}$，其中 $\mathbb{I}(\cdot)$ 表示示性函数.
2. $\mathrm{pr}(X = k)$ 可以如下估计: $\widehat{\mathrm{pr}}(X = k) = \dfrac{\sum_{i=1}^{n} \delta(X_i = k)}{n}$.
3. 我们估计 ATE 如下，

$$\widehat{\tau} = \sum_{k=1}^{K} \widehat{\mu}(1, k)\widehat{\mathrm{pr}}(X = k) - \sum_{k=1}^{K} \widehat{\mu}(0, k)\widehat{\mathrm{pr}}(X = k).$$

当 X 是维度很高或者连续时，上述估计方法将不再适用

经典案例 – Berkeley 录取率 I

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
A	933	64%	**825**	62%	108	82%
B	585	63%	**560**	63%	25	68%
C	918	35%	325	37%	**593**	34%
D	792	34%	417	33%	375	35%
E	584	25%	191	28%	**393**	24%
F	714	6%	373	6%	341	7%
Total	**4526**	**39%**	**2691**	**45%**	**1835**	**30%**

$$\widehat{E}\{Y(1)\} = \underbrace{0.62}_{\widehat{\mu}(1,1)} \times \underbrace{\frac{933}{4526}}_{\widehat{\text{pr}}(X=1)} + \underbrace{0.63}_{\widehat{\mu}(1,2)} \times \underbrace{\frac{585}{4526}}_{\widehat{\text{pr}}(X=2)} + \ldots + \underbrace{0.06}_{\widehat{\mu}(1,6)} \times \underbrace{\frac{714}{4526}}_{\widehat{\text{pr}}(X=6)} = 0.39.$$

$$\widehat{E}\{Y(0)\} = \underbrace{0.82}_{\widehat{\mu}(0,1)} \times \underbrace{\frac{933}{4526}}_{\widehat{\text{pr}}(X=1)} + \underbrace{0.68}_{\widehat{\mu}(0,2)} \times \underbrace{\frac{585}{4526}}_{\widehat{\text{pr}}(X=2)} + \ldots + \underbrace{0.07}_{\widehat{\mu}(0,6)} \times \underbrace{\frac{714}{4526}}_{\widehat{\text{pr}}(X=6)} = 0.43.$$

$$\widehat{\tau} = \widehat{E}\{Y(1)\} - \widehat{E}\{Y(0)\} = -0.04$$

Association Does Not Imply Causation!

回归估计 I

- 除了分层估计以外，在实际研究中，最常用的分析方法是对观测到的结果运行 OLS（最小二乘）回归，即假设结果变量符合以下模型：

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

- 如果上述线性模型是正确的，那么我们有：

$$\begin{aligned}\tau(X) &= E(Y \mid Z = 1, X) - E(Y \mid Z = 0, X) \\ &= \left(\beta_0 + \beta_z + \beta_x^{\mathrm{T}} X\right) - \left(\beta_0 + \beta_x^{\mathrm{T}} X\right) \\ &= \beta_z,\end{aligned}$$

这意味着因果作用在协变量方面是均匀的，即不存在异质性。

回归估计 II

- 结合可忽略性假设，这意味着：

$$\tau = E\{\tau(X)\} = \beta_z,$$

从而线性模型 Z 前面的系数具有因果含义。这是线性模型最重要的应用之一。

回归估计 III

- 此外，我们可以进一步考虑由协变量引起的因果作用异质性，即考虑以下存在交互项的线性模型

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

我们有：

$$\begin{aligned}
\tau(X) &= E(Y \mid Z = 1, X) - E(Y \mid Z = 0, X) \\
&= \left(\beta_0 + \beta_z + \beta_x^{\mathrm{T}} X + \beta_{zx}^{\mathrm{T}} X\right) - \left(\beta_0 + \beta_x^{\mathrm{T}} X\right) \\
&= \beta_z + \beta_{zx}^{\mathrm{T}} X,
\end{aligned}$$

回归估计 IV

- 由可忽略性假设，我们有：

$$\tau = E\{\tau(X)\} = E\left(\beta_z + \beta_{zx}^{\mathrm{T}} X\right) = \beta_z + \beta_{zx}^{\mathrm{T}} E(X).$$

- 因此，τ 的估计量是 $\widehat{\beta}_z + \widehat{\beta}_{zx}^{\mathrm{T}} \bar{X}$，其中 $\widehat{\beta}_z$ 是回归系数，$\bar{X}$ 是 X 的样本均值。

- 如果我们对协变量进行中心化 $\bar{X} = 0$，那么估计量就是 Z 的回归系数。

回归估计 V

- 更为一般地，我们可以使用其他更复杂的模型来估计因果作用。
- 基于处理组和对照组的数据，我们可以分别构造两个估计量 $\widehat{\mu}(1, X)$ 和 $\widehat{\mu}(0, X)$，我们可以通过如下矩约束条件求解参数 $\widehat{\alpha}$,

$$E\{Y - \mu(Z, X; \widehat{\alpha}) \mid Z, X\} = 0.$$

并构建以下估计量

$$\widehat{\tau}^{\text{reg}} = \frac{1}{n} \sum_{i=1}^{n} \{\widehat{\mu}(1, X_i) - \widehat{\mu}(0, X_i)\},$$

值得注意的是，当 $\mu(Z, X; \alpha)$ 被错误指定时，$\widehat{\tau}^{\text{reg}}$ 并不能相合估计 τ.

回归估计 VI

- 针对二值的结果变量，我们可以使用 logistic 模型来拟合：

$$E(Y \mid Z, X) = \mathrm{pr}(Y = 1 \mid Z, X) = \frac{e^{\beta_0 + \beta_z Z + \beta_x^{\mathrm{T}} X}}{1 + e^{\beta_0 + \beta_z Z + \beta_x^{\mathrm{T}} X}}$$

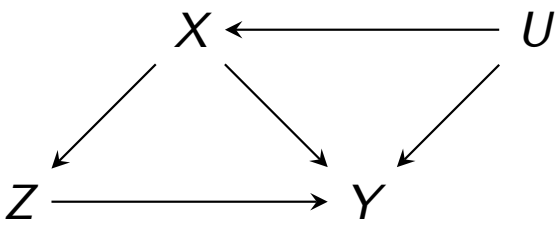
- 基于系数估计 $\widehat{\beta}_0, \widehat{\beta}_z, \widehat{\beta}_x$，我们可以得到以下的平均因果作用估计量：

$$\widehat{\tau} = n^{-1} \sum_{i=1}^{n} \left\{ \frac{e^{\widehat{\beta}_0 + \widehat{\beta}_z + \widehat{\beta}_x^{\mathrm{T}} X_i}}{1 + e^{\widehat{\beta}_0 + \widehat{\beta}_z + \widehat{\beta}_x^{\mathrm{T}} X_i}} - \frac{e^{\widehat{\beta}_0 + \widehat{\beta}_x^{\mathrm{T}} X_i}}{1 + e^{\widehat{\beta}_0 + \widehat{\beta}_x^{\mathrm{T}} X_i}} \right\}.$$

- 这个估计量不仅仅是 logistic 模型中处理变量的系数。它是所有系数以及协变量的经验分布的非线性函数。

目录

倾向得分



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 4 (positivity, or overlap)

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

- 上述概率 $e(x) = \text{pr}(Z = 1 \mid X = x)$ 常被称作倾向得分 (propensity 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

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

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

逆概加权 II

证明.

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

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：

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

  其中 $\widehat{e}(X_i)$ 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,

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

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 $e(X) = \mathsf{pr}(Z=1 \mid X)$ 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, $\widehat{\tau}_{\mathsf{ipw}}$ 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 $\widehat{\tau}^{\text{ht}}$ 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，

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

- 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，$\widehat{\tau}^{\text{hajek}}$ 比 $\widehat{\tau}^{\text{ht}}$ 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 $\tau = E\{Y(1) - Y(0)\}$ 的两个识别公式。

- 第一个是基于回归模型的公式：

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

其中

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

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：

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

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 $\mu_z(X) = E(Y \mid Z = z, X)$ 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 $e(X) = \text{pr}(Z = 1 \mid X)$ 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：

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

这也可以写成以下形式：

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

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

证明.

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

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

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

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

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

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 $\widehat{\tau}^{\text{reg}}$, $\widehat{\tau}^{\text{ht}}$ 以及 $\widehat{\tau}^{\text{dr}}$ 的有限样本表现.

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

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

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

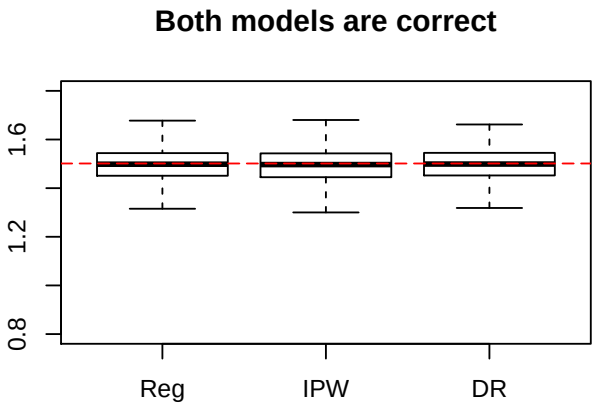
Case 3: 倾向得分模型被正确指定, 回归模型被错误指定, 即

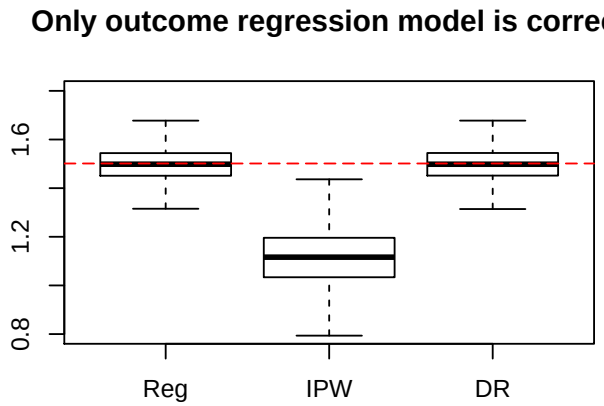
$$e(X;\beta) = \text{expit}(X); \quad \mu_z(X;\alpha) = \alpha_1 + \alpha_2 X.$$

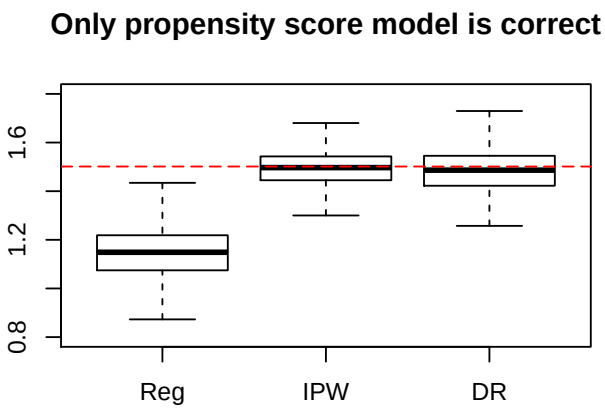
Case 4: 所有模型都被错误指定, 即

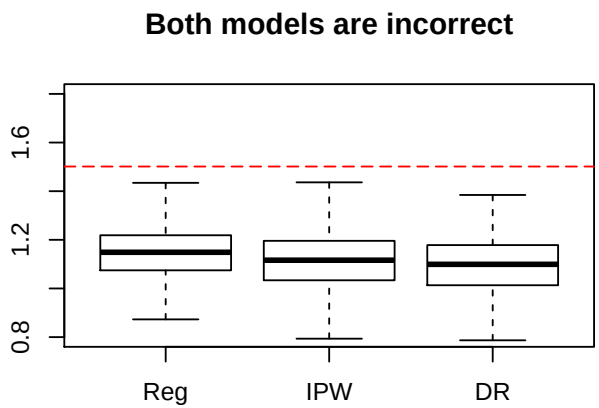
$$e(X;\beta) = \text{expit}(\beta_0 + \beta_1 X); \ \mu_z(X;\alpha) = \alpha_1 + \alpha_2 X.$$

稳健估计


Both models are correct
0.8
1.2
1.6
Reg
IPW
DR

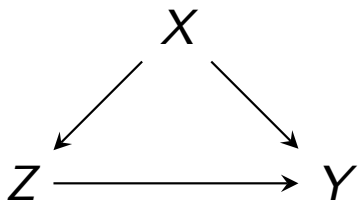

Only outcome regression model is correct
0.8
1.2
1.6
Reg
IPW
DR

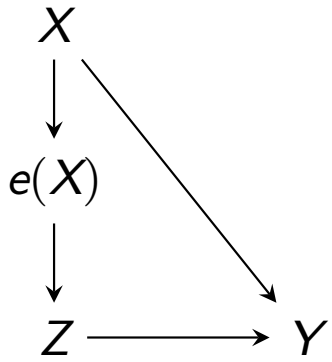

Only propensity score model is correct
0.8
1.2
1.6
Reg
IPW
DR


Both models are incorrect
0.8
1.2
1.6
Reg
IPW
DR

目录

倾向得分重要的性质





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

在可忽略性假设下, 即

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

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

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:

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

- 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:

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

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

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

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

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 $\text{pr}\{Z=1 \mid X, e(X)\} = \text{pr}(Z=1 \mid X) = e(X)$, 及

$$\text{pr}\{Z=1 \mid e(X)\} = E[\text{pr}\{Z=1 \mid X, e(X)\} \mid e(X)] = E\{e(X) \mid e(X)\} = e(X)$$

因此, $\text{pr}\{Z=1 \mid X, e(X)\} = \text{pr}\{Z=1 \mid e(X)\}$, 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,

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

- 我们可以如下估计 ATE:

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

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：

$$\tau = \text{pr}\left(Z=1\right)\tau_{\text{ATT}} + \text{pr}\left(Z=0\right)\tau_{\text{ATC}}$$

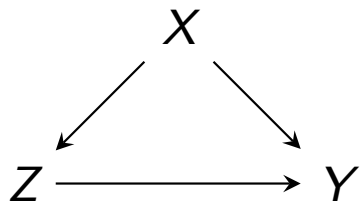
- 在随机化下，我们有 ATE = ATT = ATC, 因为

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

- 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



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.

Assumption 5 (可忽略性, ignorability)

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

在上述可忽略性假定下，ATT 可以如下识别：

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

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.

第一周作业 I

证明题（必选）：

1. 证明处理组因果作用可识别

上机作业（三选二），要求提交数据生成过程和估计过程：

- 用 R 语言实现以下例子：
  1. 基线协变量 $X = 1$ 表示数学专业，$X = 2$ 表示物理专业，$X = 3$ 表示文学专业，满足

$$\mathrm{pr}(X = 1) = 0.3,\ \mathrm{pr}(X = 2) = 0.4,\ \mathrm{pr}(X = 3) = 0.3.$$

  2. 潜在结果 $Y(0)$ 表示女性下的录取状况，$Y(1)$ 表示男性下的录取状况，满足

$$\mathrm{pr}\{Y(0) = 1 \mid X = 1\} = 0.5,\ \mathrm{pr}\{Y(1) = 1 \mid X = 1\} = 0.65$$

$$\mathrm{pr}\{Y(0) = 1 \mid X = 2\} = 0.4,\ \mathrm{pr}\{Y(1) = 1 \mid X = 2\} = 0.55$$

$$\mathrm{pr}\{Y(0) = 1 \mid X = 3\} = 0.3,\ \mathrm{pr}\{Y(1) = 1 \mid X = 3\} = 0.45$$

第一周作业 II

3. 每个专业的男生的比例为，$Z = 1$ 表示男性，$Z = 0$ 表示女性：

$$\text{pr}(Z = 1 \mid X = 1) = 0.50$$

$$\text{pr}(Z = 1 \mid X = 2) = 0.65$$

$$\text{pr}(Z = 1 \mid X = 3) = 0.30$$

如何评估 $\mathbb{E}\{Y(1) - Y(0)\}$?

- 用 R 语言实现以下例子：

1. 基线协变量 X 表示体重，服从均值为 110，标准差为 20 的正态分布，即 $X \sim N(110, 20^2)$.
2. 潜在结果 $Y(0)$ 表示不吃药下的血压，$Y(1)$ 表示吃药下的血压，满足

$$Y(0) = 120 + 0.1X + \epsilon_0,\ \epsilon_0 \sim N(0, 1)$$

$$Y(1) = 100 + 0.1X + \epsilon_1,\ \epsilon_1 \sim N(0, 1)$$

3. 用 Z 表示是否服药，$Z = 1$ 表示吃药满足

$$\text{pr}(Z = 1 \mid X) = \exp(-1 + 0.01X)/\{1 + \exp(-1 + 0.01X)\}$$

如何评估 $\mathbb{E}\{Y(1) - Y(0)\}$?

第一周作业 III

- 用 R 语言实现以下例子：
  1. 基线协变量 X 表示体重，服从均值为 110，标准差为 20 的正态分布，即 $X \sim N(110, 20^2)$.
  2. 潜在结果 $Y(0)$ 表示不吃药下的血压，$Y(1)$ 表示吃药下的血压，满足

$$Y(0) = 120 + 0.2X + \epsilon_0,\ \epsilon_0 \sim N(0,1)$$

$$Y(1) = 100 + 0.1X + \epsilon_1,\ \epsilon_1 \sim N(0,1)$$

  3. 用 Z 表示是否服药，$Z = 1$ 表示吃药满足

$$\text{pr}(Z = 1 \mid X) = \exp(-1 + 0.01X)/\{1 + \exp(-1 + 0.01X)\}$$

  如何评估 $\mathbb{E}\{Y(1) - Y(0)\}$?

作业扫描成 pdf 发邮箱:smengchen@163.com

第二周作业

1. 阅读教材 Page 139 - 162，并运行课堂所给代码（不提交）
2. 验证教材定理 11.1、11.2，并提交读书笔记
3. 上机作业：利用回归估计、逆概加权估计、双稳健估计分析教材 Page 128 实际数据 Example 10.3（要求提交，参考教材 Section 12.3.3）

   作业扫描成 pdf 发邮箱:smengchen@163.com

   抄送邮箱:shan3_luo@163.com

附加作业

1. 验证双稳健估计量的双稳健性质。
2. 假设结果变量模型 $\mu(Z, X; \alpha)$ 正确指定，真值 α_0 满足 $\mu(Z, X; \alpha_0) = \mathbb{E}(Y \mid Z, X)$，试求以下回归估计的渐近方差，

$$\frac{1}{n} \sum_{i=1}^{n} \{\mu(1, X_i; \widehat{\alpha}) - \mu(0, X_i; \widehat{\alpha})\}, \tag{3}$$

3. 假设倾向得分模型正确指定，试求逆概加权估计量渐近方差。