第三章　观察性研究

罗姗姗

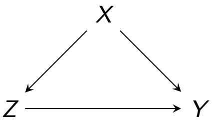
北京工商大学 数学与统计学院

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，

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

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

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

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

实际研究中的数据经常并不是随机化的.

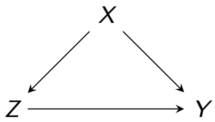
	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
Total	12,763	41%	8,442	44%	4,321	35%

- Z: 二值处理变量, $A=1$ 代表男性, $A=0$ 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？

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

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

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

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

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

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)

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

- 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.

识别性

Assumption 3 (强可忽略性, strong ignorability)

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

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

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.

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：

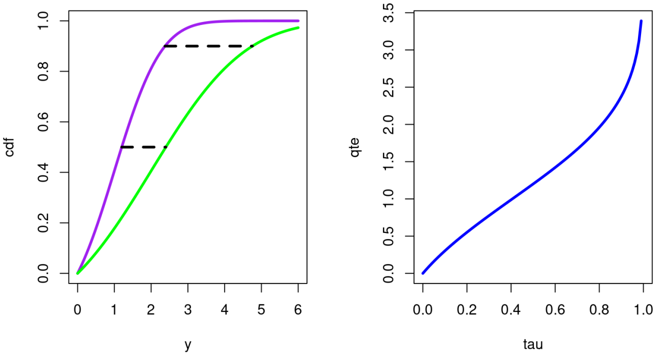
- 分布因果作用（distributional causal effect）：

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

以 $\mathrm{pr}\{Y(1) > y\}$ 为例，

$$\mathrm{pr}\{Y(1) > y \mid X\} = \mathrm{pr}(Y > y \mid Z = 1, X),$$
$$\mathrm{pr}\{Y(1) > y\} = \mathbb{E}_X\{\mathsf{pr}(Y > y \mid Z = 1, X)\}.$$

其他因果量的可识别性 II



- 分位数因果作用 (quantile causal effect)：

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

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

目录

目录

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!

目录

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。

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。

References I

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

证明题：

1. 证明处理组因果作用可识别
2. 证明分布因果作用可识别或分位数因果作用可识别

上机作业：

1. 计算所给数据的分层估计、或回归估计量

作业扫描成 pdf 发邮箱:smengchen@163.com