第二章　随机化试验和观察性研究

罗姗姗

February, 2023

目录

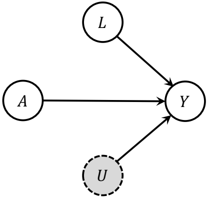
识别性

$$\boxed{\text{Causal Estimand}} \xrightarrow{\text{Identification}} \boxed{\text{Statistical Estimand}} \xrightarrow{\text{Estimation}} \boxed{\text{Estimate}}$$

- 平均因果作用 ATE 可识别：观测变量的分布 $\mathrm{pr}(L, A, Y)$ 可以唯一确定 ATE 的取值.

- 如果 ATE 不可识别，这意味着至少存在两个不相等的值 $\mathrm{ATE} = \mathrm{ATE}'$ 都满足观测到的数据.

随机化试验

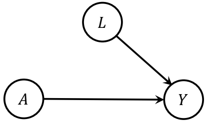


(i) 随机化

Fisher (1925, 1935) 通过随机化试验将反事实理论应用于评估因果作用.

- 在随机化试验中, 处理分配机制是已知的.
- 在随机化试验中, 我们会有:
  - $A \perp\!\!\!\perp L$.
  - $A \perp\!\!\!\perp U$.
  - $A \perp\!\!\!\perp (Y_0, Y_1)$.
- 随机化并不需要引入关于 Y 的任何模型假设.

随机化试验



(i) 随机化

Fisher (1925, 1935) 通过随机化试验将反事实理论应用于评估因果作用.

Assumption 1 (随机化)

$$A \perp\!\!\!\perp (Y_0, Y_1)$$

在随机化分配下, ATE 可以如下识别:

$$\begin{aligned}\tau &= E(Y_1 - Y_0)\\ &= E(Y_1) - E(Y_0)\\ &= E(Y_1 \mid A=1) - E(Y_0 \mid A=0)\\ &= E(Y \mid A=1) - E(Y \mid A=0)\,.\end{aligned}$$

在理想的随机化下, 因果可以由相关表示.

随机化试验

Table 2.1

	A	Y	Y^0	Y^1
Rheia	0	0	0	?
Kronos	0	1	1	?
Demeter	0	0	0	?
Hades	0	0	0	?
Hestia	1	0	?	0
Poseidon	1	0	?	0
Hera	1	0	?	0
Zeus	1	1	?	1
Artemis	0	1	1	?
Apollo	0	1	1	?
Leto	0	0	0	?
Ares	1	1	?	1
Athena	1	1	?	1
Hephaestus	1	1	?	1
Aphrodite	1	1	?	1
Cyclope	1	1	?	1
Persephone	1	1	?	1
Hermes	1	0	?	0
Hebe	1	0	?	0
Dionysus	1	0	?	0

(Hernán & Robins, 2020)

随机化下，ATE 可以如下估计：

$$\widehat{\tau} = \frac{\sum_{i=1}^{n} A_i Y_i}{\sum_{i=1}^{n} A_i} - \frac{\sum_{i=1}^{n} (1 - A_i) Y_i}{\sum_{i=1}^{n} (1 - A_i)}.$$

- A: 二值处理变量，$A = 1$ 代表接受心脏移植，$A = 0$ 代表未接受.
- Y: 二值结果变量，$Y = 1$ 代表死亡，$Y = 0$ 代表存活.
- ATE 可以如下估计：

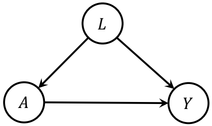
$$\widehat{\tau} = \widehat{\text{pr}}\,(Y = 1 \mid A = 1) - \widehat{\text{pr}}\,(Y = 1 \mid A = 0)$$
$$= \frac{7}{13} - \frac{3}{7} > 0.$$

随机化下其他重要的问题

- Interference, SUTVA 假设不成立时
- Noncompliance, 非依从
- Multiarm (more than 2) experiments, 多处理的随机化试验
- Covariate adjustment in RCT, 利用协变量信息提高对 ATE 的估计
- Rerandomization, 重随机化以平衡处理组和对照组协变量分布

...

观察性研究



(ii) 观察性研究

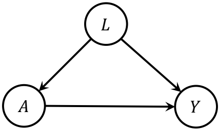
实际研究中的数据经常并不是随机化的.

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

    这意味着处理组和对照组的协变量分布不再均匀.
  - $A \not\perp\!\!\!\perp (Y_0, Y_1)$.

    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.

Yule-Simpson 悖论



(ii) 观察性研究

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

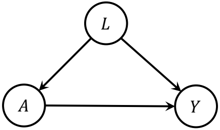
- A: 二值处理变量, $A = 1$ 代表男性, $A = 0$ 代表女性.
- Y: 二值结果变量, $Y = 1$ 代表录取, $Y = 0$ 代表未录取.
- L: 协变量, 表示专业.

实际研究中的数据经常并不是随机化的.

ATE 可以如下估计吗？

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

Yule-Simpson 悖论



(ii) 观察性研究

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:

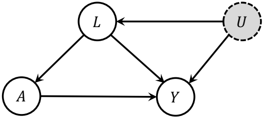
greater percentage of successful applicants than the other gender

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）

观察性研究



(ii) 观察性研究

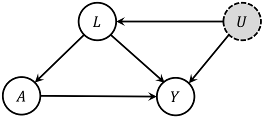
Rosenbaum & Rubin (1983) 提出的可忽略性假定是在观察性研究里评估因果作用最重要的假定.

Assumption 2 (可忽略性, ignorability)

$$A \perp\!\!\!\perp (Y_0, Y_1) \mid L.$$

- $\text{pr}(A \mid L, Y_0, Y_1) = \text{pr}(A \mid L)$.
- 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) 观察性研究

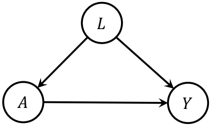
Rosenbaum & Rubin (1983) 提出的可忽略性假定是在观察性研究里评估因果作用最重要的假定.

Assumption 3 (positivity, or overlap)

$$0 < \text{pr}(A = 1 \mid L) < 1.$$

- 上述概率 $e(l) = \text{pr}(A = 1 \mid L = l)$ 常被称作倾向得分 (propensitity score).
- 在协变量 L 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.

可忽略性假设



(ii) 观察性研究

Rosenbaum & 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 4 (强可忽略性, strong ignorability)

(i) $A \perp\!\!\!\perp (Y_0, Y_1) \mid L$; (ii) $0 < \mathrm{pr}(A = 1 \mid L) < 1$.

在强可忽略性假定下, ATE 可以如下识别:

$$
\begin{aligned}
\tau &= E(Y_1 - Y_0) \\
&= E\{E(Y_1 - Y_0 \mid L)\} \\
&= E\{E(Y_1 \mid L)\} - E\{E(Y_0 \mid L)\} \\
&= E\{E(Y_1 \mid A = 1, L)\} - E\{E(Y_0 \mid A = 0, L)\} \\
&= E\{E(Y \mid A = 1, L)\} - E\{E(Y \mid A = 0, L)\}.
\end{aligned}
$$

如果不对混杂因素 L 进行调整, 将会导致有偏估计.

观察性研究

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

$$\begin{aligned}
\tau &= E(Y_1 - Y_0) \\
&= E\{E(Y \mid A = 1, L)\} - E\{E(Y \mid A = 0, L)\} \\
&= \sum_{k=1}^{K} \underbrace{E(Y \mid A = 1, L = k)}_{\mu(1,k)} \mathrm{pr}(L = k) - \sum_{k=1}^{K} \underbrace{E(Y \mid A = 0, L = k)}_{\mu(0,k)} \mathrm{pr}(L = k).
\end{aligned}$$

- 首先, $\mu(a, k)$ 可以如下估计: $\widehat{\mu}(a, k) = \dfrac{\sum_{i=1}^{n} Y_i \delta(A_i = a, L_i = k)}{\sum_{i=1}^{n} \delta(A_i = a, L_i = k)}$, 其中 $\delta(\cdot)$ 表示示性函数.
- 其次, $\mathrm{pr}(L = k)$ 可以如下估计: $\widehat{\mathrm{pr}}(L = k) = \dfrac{\sum_{i=1}^{n} \delta(L_i = k)}{n}$.
- 最后, 我们估计 ATE 如下,

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

当 L 是维度很高或者连续时, 我们可以建立参数模型来估计因果作用.

Yule-Simpson 悖论

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%**

$$\widehat{E}(Y_1) = \underbrace{0.62}_{\widehat{\mu}(1,1)} \times \underbrace{\frac{933}{4526}}_{\widehat{\text{pr}}(L=1)} + \underbrace{0.63}_{\widehat{\mu}(1,2)} \times \underbrace{\frac{585}{4526}}_{\widehat{\text{pr}}(L=2)} + \ldots + \underbrace{0.06}_{\widehat{\mu}(1,6)} \times \underbrace{\frac{714}{4526}}_{\widehat{\text{pr}}(L=6)} = 0.39.$$

Association Does Not Imply Causation!

$$\widehat{E}(Y_0) = \underbrace{0.82}_{\widehat{\mu}(0,1)} \times \underbrace{\frac{933}{4526}}_{\widehat{\text{pr}}(L=1)} + \underbrace{0.68}_{\widehat{\mu}(0,2)} \times \underbrace{\frac{585}{4526}}_{\widehat{\text{pr}}(L=2)} + \ldots + \underbrace{0.07}_{\widehat{\mu}(0,6)} \times \underbrace{\frac{714}{4526}}_{\widehat{\text{pr}}(L=6)} = 0.43.$$

$$\widehat{\tau} = \widehat{E}(Y_1) - \widehat{E}(Y_0) = -0.04$$

回归估计 (Outcome regression model)

- 在给定可忽略性假定下, ATE 可以如下识别:

$$\begin{aligned}\tau &= E(Y_1 - Y_0) \\ &= E\{\underbrace{E(Y \mid A=1, L)}_{\mu(1,L)}\} - E\{\underbrace{E(Y \mid A=0, L)}_{\mu(0,L)}\}.\end{aligned}$$

- 上述识别表达式启发我们可以如下估计 ATE:

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

上式常被称为回归估计量, 其中 $\widehat{\mu}(A, L)$ 是结果变量模型 $\mu(A, L) = E(Y \mid A, L)$ 的估计. 我们可以对 $\mu(A, L)$ 建立回归模型 $\mu(A, L; \alpha)$, 并通过如下矩约束条件求解参数 $\widehat{\alpha}$,

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

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

逆概加权估计 (Inverse probability weighted, IPW)

- 在给定可忽略性假定下, ATE 也可以如下识别:

$$\begin{aligned}\tau &= E(Y_1 - Y_0)\\ &= E\{E(Y \mid A=1, L)\} - E\{E(Y \mid A=0, L)\}\\ &= E\left\{\frac{AY}{e(L)}\right\} - E\left\{\frac{(1-A)Y}{1-e(L)}\right\},\end{aligned}$$

  其中 $e(L) = \mathrm{pr}(A=1 \mid L)$ 是倾向得分 (propensity score).

- 我们可以如下估计 ATE:

$$\widehat{\tau}_{\mathrm{ipw}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i Y_i}{\widehat{e}(L_i)} - \frac{(1-A_i)Y_i}{1-\widehat{e}(L_i)}\right\},$$

  上式被称为逆概加权估计量, 其中 $\widehat{e}(L_i)$ 是倾向得分 $e(L_i) = \mathrm{pr}(A_i = 1 \mid L_i)$ 的估计.

逆概加权估计 (Inverse probability weighted, IPW)

IPW 估计相当于对样本做了权重调整,

$$\hat{\tau}_{\mathrm{ipw}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i Y_i}{\hat{e}(L_i)} - \frac{(1-A_i)Y_i}{1-\hat{e}(L_i)}\right\},$$

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

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

- 当 L 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 $\hat{e}(L)$.
- 当 L 维数较高时, 我们可以采用对 $e(L) = \mathrm{pr}(A = 1 \mid L)$ 建立参数模型 $e(L;\beta)$, 并通过极大似然估计或矩估计等方法求解参数 $\hat{\beta}$.

  值得注意的是, 当 $e(L;\beta)$ 被错误指定时, $\hat{\tau}_{\mathrm{ipw}}$ 并不能相合估计 τ.
- Hirano et al. (2003) 指出在 IPW 估计时, 采用非参数估计的倾向得分 $\hat{e}(L)$ 将比使用正确的倾向得分的 $e(L)$ 具有更小的渐进方差.

逆概加权估计 (Inverse probability weighted, IPW)

- 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 $\widehat{e}(L)$ 将比使用正确的倾向得分的 $e(L)$ 具有更小的渐进方差.

双稳健估计 (Doubly robust estimation)

- 当所需要的回归模型不正确时, $\widehat{\tau}_{\mathrm{reg}}$ 将产生较大的偏差.
- 当所需要的倾向得分模型不正确时, $\widehat{\tau}_{\mathrm{ipw}}$ 将产生较大的偏差.
- 双稳健估计结合了回归估计与逆概加权估计的优点, 能有效缓解任一模型误设导致的偏差.

$$\begin{aligned}\widehat{\tau}_{\mathrm{dr}} =& \frac{1}{n}\sum_{i=1}^{n}\left[\frac{A_i\left\{Y_i-\mu(1,L_i;\widehat{\alpha})\right\}}{e(L_i;\widehat{\beta})}+\mu(1,L_i;\widehat{\alpha})\right] \\ & -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(1-A_i)\left\{Y_i-\mu(0,L_i;\widehat{\alpha})\right\}}{1-e(L_i;\widehat{\beta})}+\mu(0,L_i;\widehat{\alpha})\right].\end{aligned}$$

上述 $\widehat{\tau}_{\mathrm{dr}}$ 估计量有两个等价的数学表示，进而揭示了 $\widehat{\tau}_{\mathrm{dr}}$ 的双稳健性质.

双稳健估计 (Doubly robust estimation)

当回归模型 $\mu(A_i, L_i; \alpha)$ 被正确指定时, 我们可以得到 α 的相合估计量 $\widehat{\alpha}$.

$$\widehat{\tau}_{\mathrm{dr}} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\mu(1, L_i; \widehat{\alpha}) - \frac{1}{n}\sum_{i=1}^{n}\mu(0, L_i; \widehat{\alpha})}_{\widehat{\tau}_{\mathrm{reg}} \xrightarrow{p} \tau}$$

$$+ \underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{A_i}{e(L_i; \widehat{\beta})}\{\mu(1, L_i; \widehat{\alpha}) - Y_i\}}_{\xrightarrow{p} E\left[\frac{A_i}{e\left(L_i;\beta^*\right)}\{\mu(1,L_i;\alpha^*)-Y_i\}\right]=0} - \underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{1 - A_i}{1 - e(L_i; \widehat{\beta})}\{\mu(0, L_i; \widehat{\alpha}) - Y_i\}}_{\xrightarrow{p} E\left[\left\{\frac{1-A_i}{1-e\left(L_i;\beta^*\right)}\right\}\{\mu(0,L_i;\alpha^*)-Y_i\}\right]=0},$$

其中 α^* 和 β^* 分别表示当样本量趋于无穷时 $\widehat{\alpha}$ 和 $\widehat{\beta}$ 依概率收敛的极限值. 注意到，当回归模型 $\mu\left(A_i, L_i; \alpha\right)$ 被正确指定时，无论倾向得分模型 $e\left(L_i; \beta\right)$ 正确指定与否，我们都有

$$E\left[\frac{A_i}{e\left(L_i; \beta^*\right)}\left\{\mu\left(1, L_i; \alpha^*\right) - Y_i\right\}\right] = E\left[\underbrace{E\left\{\mu\left(1, L_i; \alpha^*\right) - Y_i \mid A_i = 1, L_i\right\}}_{=0}\frac{A_i}{e\left(L_i; \beta^*\right)}\right] = 0.$$

双稳健估计 (Doubly robust estimation)

当倾向得分模型 $e(L_i;\beta)$ 被正确指定时, 我们可以得到 β 的相合估计量 $\widehat{\beta}$.

$$
\begin{aligned}
\widehat{\tau}_{\mathrm{dr}} = & \underbrace{\frac{1}{n}\sum_{i=1}^{n}\frac{A_i}{e(L_i;\widehat{\beta})}Y_i - \frac{1}{n}\sum_{i=1}^{n}\frac{1-A_i}{1-e(L_i;\widehat{\beta})}Y_i}_{\widehat{\tau}_{\mathrm{ipw}}\xrightarrow{p}\tau} \\
& + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left\{1-\frac{A_i}{e(L_i;\widehat{\beta})}\right\}\mu(1,L_i;\widehat{\alpha})}_{\xrightarrow{p}E\left[\left\{1-\frac{A_i}{e(L_i;\beta^*)}\right\}\mu(1,L_i;\alpha^*)\right]=0} - \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left\{1-\frac{1-A_i}{1-e(L_i;\widehat{\beta})}\right\}\mu(0,L_i;\widehat{\alpha})}_{\xrightarrow{p}E\left[\left\{1-\frac{1-A_i}{1-e(L_i;\beta^*)}\right\}\mu(0,L_i;\alpha^*)\right]=0}.
\end{aligned}
$$

其中 α^* 和 β^* 分别表示当样本量趋于无穷时 $\widehat{\alpha}$ 和 $\widehat{\beta}$ 依概率收敛的极限值. 注意到，当倾向得分模型 $e(L_i;\beta)$ 被正确指定时，回归模型 $\mu(A_i,L_i;\alpha)$ 无论正确指定与否，我们都有

$$
E\left[\left\{1-\frac{A_i}{e(L_i;\beta^*)}\right\}\mu(1,L_i;\alpha^*)\right] = E\left[\underbrace{E\left\{\left.\frac{e(L_i;\beta^*)-A_i}{e(L_i;\beta^*)}\right| L_i\right\}}_{=0}\mu(1,L_i;\alpha^*)\right] = 0.
$$

双稳健估计

- 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 & Schafer, 2007).
- 双稳健估计量在近些年得到了广泛的应用.

稳健估计

- 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 $\widehat{\tau}_{\text{reg}}$, $\widehat{\tau}_{\text{ipw}}$ 以及 $\widehat{\tau}_{\text{dr}}$ 的有限样本表现.

- $L \sim N(0,1)$
- $\text{pr}(A=1 \mid L) = \text{expit}(0.5L - 0.2L^2) \quad \Rightarrow \quad e(L) = \text{expit}(\beta_0 + \beta_l L + \beta_2 L^2)$
- $Y = 1.5A + 2L + L^2 + \varepsilon,\ \varepsilon \sim N(0,1) \quad \Rightarrow \quad \mu(A,L) = \alpha_0 + \alpha_1 A + \alpha_2 L + \alpha_3 L^2$

Case 1: 所有模型都被正确指定, 即

$$e(L;\beta) = e(L); \quad \mu(A,L;\alpha) = \mu(A,L).$$

Case 2: 回归模型被正确指定, 倾向得分被错误指定, 即

$$e(L;\beta) = \text{expit}(\beta_0 + \beta_1 L); \quad \mu(A,L;\alpha) = \mu(A,L).$$

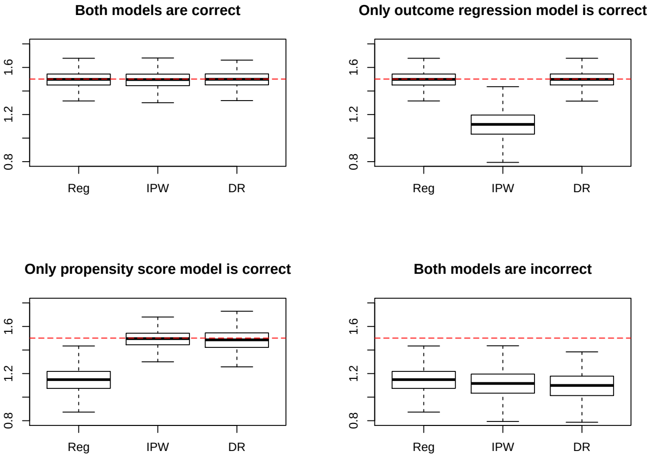
Case 3: 倾向得分模型被正确指定, 回归模型被错误指定, 即

$$e(L;\beta) = \text{expit}(L); \quad \mu(A,L;\alpha) = \alpha_0 + \alpha_1 A + \alpha_2 L.$$

Case 4: 所有模型都被错误指定, 即

$$e(L;\beta) = \text{expit}(\beta_0 + \beta_1 L); \ \mu(A,L;\alpha) = \alpha_0 + \alpha_1 A + \alpha_2 L.$$

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

倾向得分重要的性质

(i) (ii)

定理 1 (倾向得分定理, Rosenbaum & Rubin (1983))

在可忽略性假设下, 即

$$(\text{i})\ A \perp\!\!\!\perp (Y_0, Y_1) \mid L; \quad 0 < \text{pr}(A = 1 \mid L) < 1,$$

$$\Longrightarrow (\text{ii})\ A \perp\!\!\!\perp (Y_0, Y_1) \mid e(L); \quad 0 < \text{pr}\{A = 1 \mid e(L)\} < 1.$$

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

倾向得分重要的性质

证明.

我们只要证明

$$\mathrm{pr}(A=1 \mid Y_1, e(L)) = \mathrm{pr}(A=1 \mid e(L))$$

$$\begin{aligned}
\mathrm{pr}\left(A=1 \mid Y_1, e(L)\right) &= \mathbb{E}\left\{A \mid Y_1, e(L)\right\} \\
&= \mathbb{E}\left[\mathbb{E}\left\{A \mid Y_1, e(L), L\right\} \mid Y_1, e(L)\right] \\
&= \mathbb{E}\left\{\mathbb{E}(A \mid Y_1, L) \mid Y_1, e(L)\right\} \\
&= \mathbb{E}\left\{\mathbb{E}(A \mid L) \mid Y_1, e(L)\right\} \\
&= \mathbb{E}\left\{e(L) \mid Y_1, e(L)\right\} \\
&= e(L) \\
&= \mathrm{pr}(A=1 \mid e(L))
\end{aligned}$$

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 $\mathrm{pr}\{A=1 \mid L, e(L)\} = \mathrm{pr}(A=1 \mid L) = e(L)$, 及

$$\mathrm{pr}\{A=1 \mid e(L)\} = E[\mathrm{pr}\{A=1 \mid L, e(L)\} \mid e(L)] = E\{e(L) \mid e(L)\} = e(L)$$

因此, $\mathrm{pr}\{A=1 \mid L, e(L)\} = \mathrm{pr}\{A=1 \mid e(L)\}$, 即 $L \perp\!\!\!\perp A \mid e(L)$. $\square$

匹配 (Matching)

- 通过回归估计模型,

  对于处理组的个体 i, 我们使用 $\widehat{\mu}(0, L_i)$ "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 $\widehat{\mu}(1, L_i)$ "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 L_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 L_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:

$$J_M(i)=\left\{j=1,\dots,n: A_j=1-A_i \text{ and } \sum_{k:A_k=1-A_i}\delta\left(\|L_i-L_k\|\leqslant\|L_i-L_j\|\right)\leqslant M\right\},$$

- 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

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

再谈匹配

- 当 L 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，

$$J_M(i) = \left\{ j = 1, \ldots, n : A_j = 1 - A_i \text{ and } \sum_{k: A_k = 1 - A_i} \delta\left(|e(L_i) - e(L_k)| \leqslant |e(L_i) - e(L_j)|\right) \leqslant M \right\}.$$

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

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

再谈处理组的因果效应

- 由定义，ATE, ATT 和 ATC 三者的关系如下:

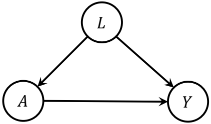
$$\tau = \mathsf{pr}(A=1)\,\tau_{\mathrm{ATT}} + \mathsf{pr}(A=0)\,\tau_{\mathrm{ATC}}$$

- 在随机化下，我们有 ATE = ATT = ATC, 因为

$$\underbrace{E(Y_1 - Y_0)}_{\tau} = \underbrace{E(Y_1 - Y_0 \mid A=1)}_{\tau_{\mathrm{ATT}}} = \underbrace{E(Y_1 - Y_0 \mid A=0)}_{\tau_{\mathrm{ATC}}}$$

- 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.

处理组的因果效应



(ii) 观察性研究

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) $A \perp\!\!\!\perp Y_0 \mid L$; (ii) $\mathrm{pr}(A=1 \mid L) < 1$.

在上述可忽略性假定下, ATT 可以如下识别:

$$
\begin{aligned}
\tau_{\mathrm{ATT}} &= E(Y_1 - Y_0 \mid A=1) \\
&= E(Y_1 \mid A=1) - E\{E(Y_0 \mid A=1, L) \mid A=1\} \\
&= E(Y \mid A=1) - E\{E(Y_0 \mid A=1, L) \mid A=1\} \\
&= E(Y \mid A=1) - E\{E(Y_0 \mid A=0, L) \mid A=1\} \\
&= E(Y \mid A=1) - E\{E(Y \mid A=0, L) \mid A=1\}.
\end{aligned}
$$

本周作业

1. 证明逆概加权识别表达式
2. 上机作业：计算所给数据的回归估计量，逆概加权估计量，双稳健估计量
3. 附加题：验证双稳健估计量的双稳健性质，可参考文献：

   Tsiatis, A. A. (2006). Semiparametric theory and missing data. Springer. P147-150.
4. 作业扫描成 pdf 发邮箱：shan3_luo@163.com
5. 上机作业要求用 Rmarkdown 输出 pdf，注意注释及格式
6. 截止日期：下周三，2023.03.09.