定量社会科学的因果推断

Causal Inference in Quantitative Social Sciences

江　艇
中国人民大学经济学院

Last updated: March 7, 2021

Lecture 2　潜在结果分析框架

- 潜在结果分析框架 (potential outcomes framework) 是对因果推断问题进行规范化表述的语言。可以追溯到统计学先驱 Neyman, Pearson, Fisher，但主要由统计学家 Donald Rubin 在 1974-1980 年的一系列研究所奠定的，因此也称为鲁宾因果模型 (Rubin causal model)。

- 定义潜在结果

$$\begin{aligned} Y_i^0 &= \text{不对个体 } i \text{ 实施某种处理的潜在结果} \\ Y_i^1 &= \text{对个体 } i \text{ 实施某种处理的潜在结果} \end{aligned}$$

- 定义个体处理效应（因果效应）

$$\tau_i \triangleq Y_i^1 - Y_i^0$$

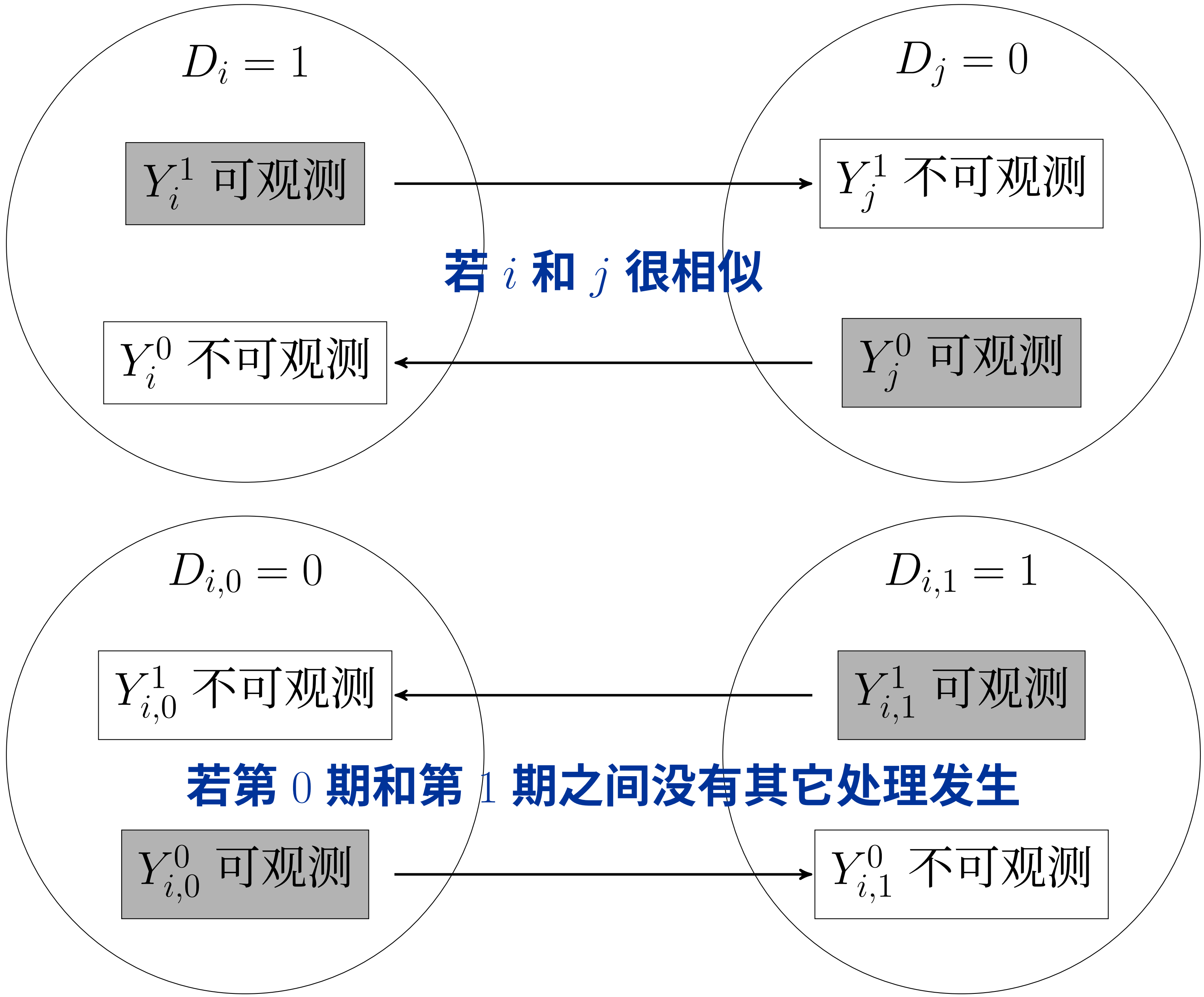
- 观测到的处理状态

$$D_i = \begin{cases} 1 & \text{对个体 } i \text{ 实施一项处理（进入处理组、实验组）} \\ 0 & \text{不对个体 } i \text{ 实施一项处理（进入控制组、对照组）} \end{cases}$$

- 观测到的结果

$$Y_i = \begin{cases} Y_i^1 & D_i = 1 \\ Y_i^0 & D_i = 0 \end{cases}$$

- 关于这一定义的几点评论：
  - 因果效应的定义只取决于潜在结果，而不取决于究竟哪个潜在结果被实际观测到（实际发生）。
  - 因果效应永远是同一个体在同一时点上的不同潜在结果之间的比较。
  - 因果推断的基本难题就是数据缺失：我们永远只能至多观察到一个潜在结果。没有观察到的那个潜在结果叫做反事实 (counterfactual)。
  - 因果分析的关键就是构造反事实，也就是个插值问题。而且反事实只能从实际观测到的、处理状态相反的其他结果中去寻找，要么是其他个体的潜在结果，要么是同一个体其他时期的潜在结果。
  - 我们实际衡量因果效应时必须涉及多个个体或同一个体的不同时期，但无论是横截面比较还是事前事后比较，都不是因果效应的定义，尽管它们对于衡量因果效应很重要。


$D_i = 1$
$D_j = 0$
Y_i^1 可观测
Y_j^1 不可观测
若 i 和 j 很相似
Y_i^0 不可观测
Y_j^0 可观测
$D_{i,0} = 0$
$D_{i,1} = 1$
$Y_{i,0}^1$ 不可观测
$Y_{i,1}^1$ 可观测
若第 0 期和第 1 期之间没有其它处理发生
$Y_{i,0}^0$ 可观测
$Y_{i,1}^0$ 不可观测

被估计量

- 平均处理效应 (average treatment effect, ATE)

$$\tau \triangleq \mathbb{E}\left(Y_i^1 - Y_i^0\right)$$

- 处理组的平均处理效应 (average treatment effect on the treated, ATT)

$$\tau_1 \triangleq \mathbb{E}\left(Y_i^1 - Y_i^0 | D_i = 1\right)$$

- 控制组的平均处理效应 (average treatment effect on the untreated, ATU)

$$\tau_0 \triangleq \mathbb{E}\left(Y_i^1 - Y_i^0 | D_i = 0\right)$$

- 除非随机分组，一般来说以上三者并不相等。
- 三者存在如下关系：

$$\tau = \tau_1 \cdot \mathrm{Pr}(D_i = 1) + \tau_0 \cdot \mathrm{Pr}(D_i = 0)$$

- 定义在协变量上的条件平均处理效应 (Conditional ATE, CATE)

$$
\begin{aligned}
\tau(x) &\triangleq \mathbb{E}\big(Y_i^1 - Y_i^0 | X_i = x\big) \\
\tau_1(x) &\triangleq \mathbb{E}\big(Y_i^1 - Y_i^0 | X_i = x, D_i = 1\big) \\
\tau_0(x) &\triangleq \mathbb{E}\big(Y_i^1 - Y_i^0 | X_i = x, D_i = 0\big)
\end{aligned}
$$

根据期望迭代定律,

$$
\begin{aligned}
\tau &= \mathbb{E}\left\{\mathbb{E}\big(Y_i^1 - Y_i^0 | X_i = x\big)\right\} && = \int \tau(x) dF_X \\
\tau_1 &= \mathbb{E}\left\{\mathbb{E}\big(Y_i^1 - Y_i^0 | X_i = x, D_i = 1\big)\right\} && = \int \tau_1(x) dF_{X|D=1} \\
\tau_0 &= \mathbb{E}\left\{\mathbb{E}\big(Y_i^1 - Y_i^0 | X_i = x, D_i = 0\big)\right\} && = \int \tau_0(x) dF_{X|D=0}
\end{aligned}
$$

第一类识别假设

- 显然，分配机制对于平均处理效应的识别至关重要。下面这个例子说明，不对分配机制进行正式建模，可能使得平均处理效应无法被正确识别。

Unit	Treatment	Observed Outcome
Patient #1	Surgery	7
Patient #2	Drug	6
Patient #3	Surgery	5
Patient #4	Drug	8
Average Difference		-1

Unit	Potential Outcomes		Causal Effects
	Surgery	Drug	S-D
Patient #1	7	1	6
Patient #2	5	6	-1
Patient #3	5	1	4
Patient #4	7	8	-1
Average Effect			2

- 因此，因果识别的关键假设往往就是关于分配机制的假设。
- 在潜在结果分析框架下，分配机制（倾向得分）可以表示为

$$\pi \triangleq \mathsf{Pr}\left(D = 1|X, Y^0, Y^1\right)$$

比如在前面的例子中，

$$\pi = \mathbb{1}\left(Y^1 - Y^0 > 0\right)$$

- 第一类识别假设：分配机制不取决于潜在结果和可观测变量。

$$\mathsf{Pr}(D = 1|X, Y^0, Y^1) = \mathsf{Pr}(D = 1)$$

- 第一类识别假设（重新表述）：潜在结果均值独立于处理状态。

Assumption ID.1: $\mathbb{E}\left(Y^d|D\right) = \mathbb{E}\left(Y^d\right),\ d = 0, 1$

等价地，

$$\mathbb{E}(Y^0|D = 1) = \mathbb{E}(Y^0|D = 0)$$
$$\mathbb{E}(Y^1|D = 1) = \mathbb{E}(Y^1|D = 0)$$

- 识别：此时**组间均值差异**能够直接识别 ATE (τ), ATT (τ_1) 和 ATU (τ_0).

$$
\begin{aligned}
&\mathbb{E}(Y|D=1)-\mathbb{E}(Y|D=0)\\
&=\mathbb{E}\big(Y^1|D=1\big)-\mathbb{E}\big(Y^0|D=0\big)\\
&=\mathbb{E}\big(Y^1\big)-\mathbb{E}\big(Y^0\big)\\
&=\mathbb{E}\big(Y^1-Y^0\big)\\
&=\tau
\end{aligned}
$$

  要想识别 τ_1，需要用到 $\mathbb{E}\big(Y^0|D\big)=\mathbb{E}\big(Y^0\big)$.
  要想识别 τ_0，需要用到 $\mathbb{E}\big(Y^1|D\big)=\mathbb{E}\big(Y^1\big)$.
  证明过程类似。因此有

$$\tau=\tau_1=\tau_0$$

- 估计：样本组间均值差异是总体组间均值差异的一致估计。

$$\bar{Y}_{D=1}-\bar{Y}_{D=0}\rightarrow_p \mathbb{E}(Y|D=1)-\mathbb{E}(Y|D=0)$$

- 第一类识别假设就是随机实验 (也称随机控制试验, randomized controlled trial, RCT) 的识别假设。

重新理解线性结构模型

- 在假设**ID.1**下，样本组间均值差异是 τ 的一致估计；在假设**LS.1**下，样本组间均值差异是线性结构模型的结构参数 β_1 的一致估计。一个自然的问题是：τ 和 β_1 是什么关系？以及，假设**ID.1**和假设**LS.1**是什么关系？
- 定义潜在结果：

$$
\begin{aligned}
Y_i^1 &= \mathbb{E}(Y_i^1) + U_i^1 \triangleq \alpha_1 + U_i^1 \\
Y_i^0 &= \mathbb{E}(Y_i^0) + U_i^0 \triangleq \alpha_0 + U_i^0
\end{aligned}
$$

- 线性结构模型的结构参数 β_1 等价于潜在结果框架下定义的平均处理效应 τ。

$$
\begin{aligned}
Y_i =& (1 - D_i)Y_i^0 + D_i Y_i^1 \\
=& (1 - D_i)(\alpha_0 + U_i^0) + D_i(\alpha_1 + U_i^1) \\
=& \alpha_0 + \underbrace{(\alpha_1 - \alpha_0)}_{\text{ATE}} D_i + \left[D_i U_i^1 + (1 - D_i)U_i^0\right]
\end{aligned}
$$

- 并且线性结构模型并没有 $\beta_{1i} \equiv \beta_1$ 这样的限制性假设，允许处理效应存在异质性，β_1 衡量的是其平均值。
- 我们还能方便地证明，假设**ID.1**和假设**LS.1**是等价的。

证明：假设 ID.1 与假设 LS.1 等价。

假设**LS.1**要求

$$\text{Cov}(D, DU^1 + (1-D)U^0) = 0$$

$$\begin{aligned}
& LHS \\
&= \mathbb{E}(DU^1) - \mathbb{E}(D)\mathbb{E}(DU^1 + (1-D)U^0) \\
&= (1-\mathbb{E}(D))\mathbb{E}(DU^1) - \mathbb{E}(D)\mathbb{E}((1-D)U^0) \\
&= (1-\mathbb{E}(D))\mathbb{E}(U^1|D=1)P(D=1) - \mathbb{E}(D)\mathbb{E}(U^0|D=0)P(D=0) \\
&= \mathbb{E}(D)\big(1-\mathbb{E}(D)\big)\Big[\mathbb{E}\big(U^1|D=1\big) - \mathbb{E}\big(U^0|D=0\big)\Big]
\end{aligned}$$

若假设**ID.1**成立,

$$\mathbb{E}\big(Y^d|D\big) = \mathbb{E}\big(Y^d\big)$$

则

$$\mathbb{E}(U^d|D) = 0$$

$$\mathbb{E}\big(U^1|D=1\big) - \mathbb{E}\big(U^0|D=0\big) = 0$$