Chapter 26 书中最后一章

2022 成稿

随时间变化的处理和混杂

J. Robins 的贡献：尤其是随时间变化的混杂

例子： HIV 病人

吃药与否随时间变化

例子： 教育学

不同的教育/课程方式随时间变化

例子： 政治科学 (political science)

竞选策略变化

论文：$K=2$ 两个时间点

讲概念

问题：一般 $K \geq 2$

有技术推广

<u>$K=2$</u>：变量的时间先后顺序

$$X_0 \longrightarrow Z_1 \longrightarrow X_1 \longrightarrow Z_2 \longrightarrow Y$$

X_0	Z_1	X_1	Z_2	Y
基线协变量 baseline covariates	时间点1的处理 $\in \{0,1\}$	随时间变化的协变量	时间点2的处理 $\in \{0,1\}$	结果

潜在结果 $Y(z_1, z_2)$

2×2 因子实验

观察的结果 $Y = Y(Z_1, Z_2)$

$$= \sum_{z_1=0,1}\sum_{z_2=0,1} 1(Z_1=z_1)\,1(Z_2=z_2)\,Y(z_1,z_2)$$

可忽略性的推广：

$$Z \perp\!\!\!\perp Y(z) \mid X$$

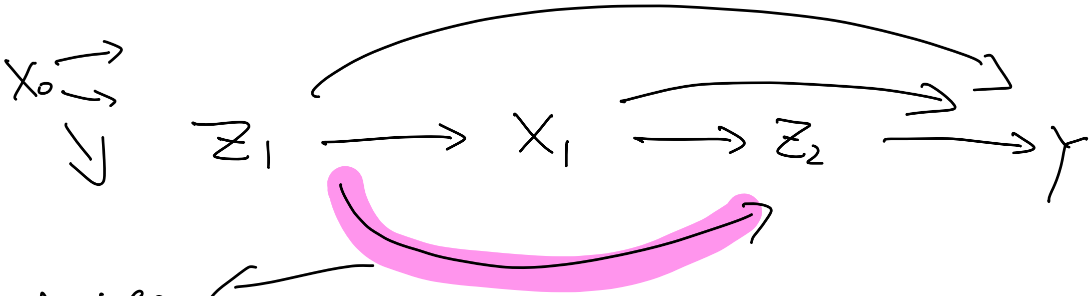
$\rightsquigarrow$ 序贯可忽略性

(SI) sequential ignorability

$$\begin{cases} Z_1 \perp\!\!\!\perp Y(z_1,z_2) \mid X_0 \\ Z_2 \perp\!\!\!\perp Y(z_1,z_2) \mid (Z_1, X_1, X_0) \end{cases}$$

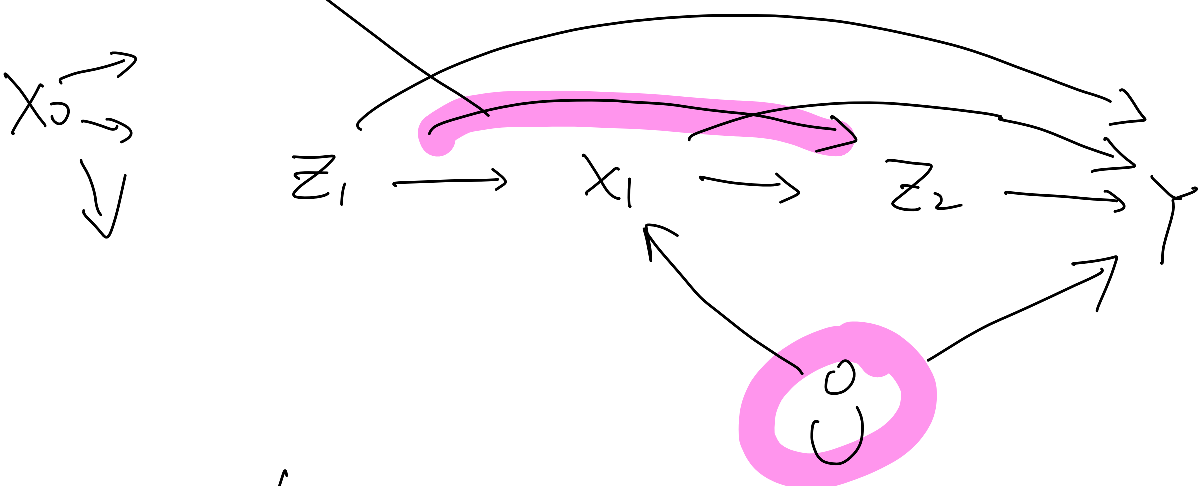
假设在序贯的随机化实验中可以自然成立。

一般地，它无法被证伪。



书中暂时没有

无混杂 ⇒ SI 成立



有 X_1-Y 混杂 ⇒ SI 也成立

29.2 g-formula 和回归

一个时间点: $E(Y(z)) = E\{E(Y \mid Z=z, X)\}$

$$= \sum_x E(Y \mid Z=z, x)\Pr(x)$$

定理 29.1 SI $\Rightarrow$ g-formula

$$E\{Y(z_1,z_2)\} = E\left[E\left(E(Y \mid z_2, z_1, X_1, X_0) \mid z_1, X_0 \right) \right]$$

$$= \sum_{x_0} \sum_{x_1} E(Y \mid z_2, z_1, x_1, x_0)\, Pr(x_1 \mid z_1, x_0)\, Pr(x_0)$$

解释: $E(Y)$

$$= \sum_{x_0} \cancel{\sum_{z_1}} \sum_{x_1} \cancel{\sum_{z_2}} E(Y \mid z_2, z_1, x_1, x_0)$$

$$\cdot \cancel{Pr(z_2 \mid z_1, x_1, x_0)}$$

$$\cdot Pr(x_1 \mid z_1, x_0)$$

$$\cdot \cancel{Pr(z_1 \mid x_0)}$$

$$\cdot Pr(x_0) \Rightarrow \text{g-formula}$$

固定

$Z_1 = z_1$

$Z_2 = z_2$

<u>证明</u> $E\{Y(z_1, z_2)\}$

$$= E\{\underbrace{E(Y(z_1, z_2) \mid X_0)}\}$$

$$\overset{SI(i)}{=} E\{Y(z_1, z_2) \mid z_1, X_0\}$$

$$= E\{E(Y(z_1, z_2) \mid z_1, X_1, X_0) \mid z_1, X_0\}$$

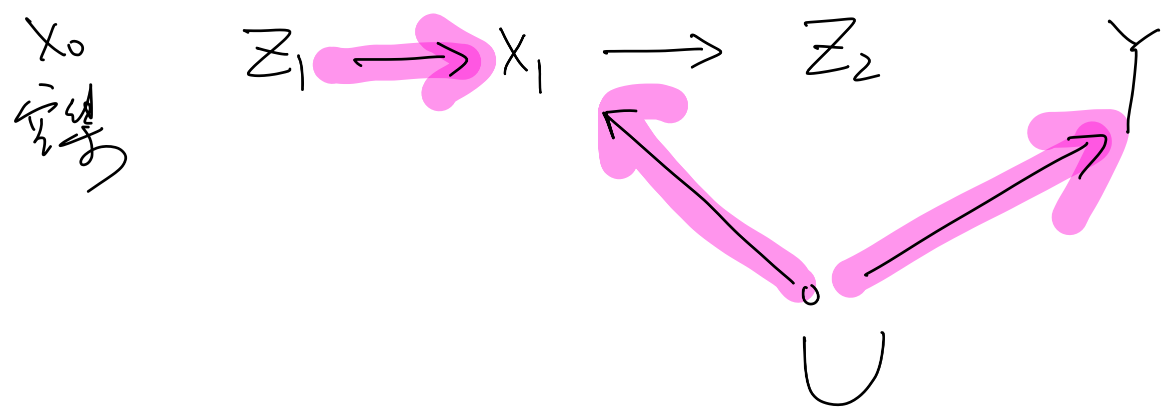
$$\overset{SI(ii)}{=} E\{E(Y(\cancel{z_1}, \cancel{z_2}) \mid z_1, z_2, X_1, X_0) \mid z_1, X_0\}$$

□

<u>g-null paradox</u> HW 29.1

$n \to \infty$ 永远无法拒绝零假设

即使 z_1, z_2 对 Y 无作用.



用 g-formula: $E(Y \mid Z_2, Z_1, X_1) = \beta_0 + \beta_1 Z_2 + \beta_2 Z_1 + \beta_3 X_1$

最简单：线性模型

$$\Rightarrow E\{Y(Z_1, Z_2)\} = \sum_{x_1} E(Y \mid Z_2, Z_1, x_1) \Pr(x_1 \mid Z_1)$$

$$= \sum_{x_1} \left(\beta_0 + \beta_1 Z_2 + \beta_2 Z_1 + \beta_3 x_1 \right) \Pr(x_1 \mid Z_1)$$

$$= \beta_0 + \beta_1 Z_2 + \beta_2 Z_1 + \beta_3 E(X_1 \mid Z_1)$$

与 (Z_1, Z_2) 无关 的 充分条件 是：

(1) $\beta_1 = 0, \quad \beta_2 = 0, \quad \beta_3 = 0$

或 (2) $\beta_1=0$, $\beta_2=0$, $E(X_1|Z_1)$ 不依赖于 Z_1

解决方案: 29.2.2

IPW

$$E(Y(z)) = E\left\{\frac{1(Z=z)Y}{Pr(Z=z|X)}\right\}$$

推广: $e(z_1, X_0) = Pr(Z_1 = z_1 | X_0)$

$e(z_2, Z_1, X_1, X_0) = Pr(Z_2 = z_2 | Z_1, X_1, X_0)$

定理 29.2 SI

$$E(Y(z_1,z_2)) = E\left\{\frac{1(Z_1=z_1)1(Z_2=z_2)Y}{e(z_1,X_0)\,e(z_2,Z_1,X_1,X_0)}\right\}$$

要求 overlap $0 <$ < 1

$\Rightarrow$ 估计不稳定

一般 K： 2^K 个潜在结果

2^K 因子实验

参数过多！

Marginal Structural Model (MSM)

$$E\{Y(z_1, z_2)\} = f(z_1, z_2; \beta)$$

非饱和模型

$$E\{Y(z_1, z_2) \mid X_0\} = f(z_1, z_2, X_0; \beta)$$

只能包含基线协变量

$\Uparrow$ 需要选择 f

$$\beta = \arg\min_{b} \sum_{z_2} \sum_{z_1} E\left(Y(z_1, z_2) - f(z_1, z_2, X_0; b) \right)^2$$

定理 29.3

$$= \arg\min_{b} \sum_{z_2} \sum_{z_1} E\left[\frac{1(Z_1 = z_1) 1(Z_2 = z_2)}{e(z_1, X_0)\, e(z_2, z_1, X_1, X_0)} \left(Y - f(z_1, z_2, X_0; b) \right)^2 \right]$$

(SNM)

29.4.2 Structural nested model

$$E\left(Y(z_1, 0) - Y(0, 0) \mid Z_1 = z_1, X_0 \right) = g_1(z_1, X_0; \beta)$$

逻辑 $g_1(0, X_0; \beta) = 0$

$$E\left(Y(z_1, z_2) - Y(z_1, 0) \mid Z_2 = z_2, Z_1 = z_1, X_1, X_0 \right) = g_2(z_2, z_1, X_1, X_0; \beta)$$

逻辑 $g_2(0, z_1, X_1, X_0; \beta) = 0$

定义：$U_2(\beta) = Y - g_2(Z_2, Z_1, X_1, X_0; \beta)$

$U_1(\beta) = U_2(\beta) - g_1(Z_1, X_0; \beta)$

直觉：$U_2(\beta) \approx Y(Z_1, 0)$

$U_1(\beta) \approx Y(0,0)$

引理 29.1　SI，SNM

① $E(U_2(\beta) \mid Z_2, Z_1, X_1, X_0) = E(U_2(\beta) \mid Z_1, X_1, X_0)$

$= E(Y(Z_1, 0) \mid Z_1, X_1, X_0)$

$\| ?$

定义 $Z_1 Y(1,0) + (1 - Z_1) Y(0,0)$

② $E(U_1(\beta) \mid Z_1, X_0) = E(U_1(\beta) \mid X_0)$

$= E(Y(0,0) \mid X_0)$

$\Longrightarrow$ 构造估计方程

estimating equation

<u>定理29.4</u>　SI, SNM

$$
\begin{cases}
E\left\{ h_2(Z_1, X_1, X_0)\big(Z_2 - e(1, X_1, X_0)\big) U_2(\beta) \right\} = 0 \\
E\left\{ h_1(X_0)\big(Z_1 - e(1, X_0)\big) U_1(\beta) \right\} = 0
\end{cases}
$$

任意 h_2

任意 h_1

或 $\frac{1}{e}$

证明: 自阅读

例2
$$\begin{cases} g_1(Z_1, X_0; \beta) = \beta_1 Z_1 \\ g_2(Z_2, Z_1, X_1, X_0; \beta) = (\beta_2 + \beta_3 Z_1) Z_2 \end{cases}$$

选 $h_1 = 1$, $h_2 = \begin{pmatrix} 1 \\ Z_1 \end{pmatrix}$

$$\Rightarrow \begin{cases} E\left(\left[Z_2 - e(1, Z_1, X_1, X_0)\right]\left[Y - (\beta_2 + \beta_3 Z_1) Z_2\right]\right) = 0 \\ E\left(Z_1\left[Z_2 - e(1, Z_1, X_1, X_0)\right]\left[Y - (\beta_2 + \beta_3 Z_1) Z_2\right]\right) = 0 \\ E\left(\left[Z_1 - e(1, X_0)\right]\left[Y - (\beta_2 + \beta_3 Z_1) Z_2 - \beta_1 Z_1\right]\right) = 0 \end{cases}$$

习题

$K=2 \Rightarrow K\geq 2$ 推广

HW 29.6, 29.7, 29.8, 29.9

$K=2 \Rightarrow K=1$

问 SNM 是什么?

HW 29.4

$$E\left(Y(z)-Y(0) \mid Z=z, X\right)=g(z,X;\beta)$$

$$\Rightarrow E\left\{h(X)\left(Z-e(X)\right)\left(Y-g(Z,X;\beta)\right)\right\}=0$$

特例: $g(z,X;\beta)=\beta z$

用 $$E\left\{\left(Z-e(X)\right)\left(Y-\beta Z\right)\right\}=0$$

$$\Rightarrow \beta=\frac{E\left\{\left(Z-e(X)\right)Y\right\}}{E\left\{\left(Z-e(X)\right)Z\right\}}$$

$$= OLS(Y \sim \check{Z})$$

其中 $\check{Z} = Z - e(X)$.

James M. Robins

FOLLOW

professor of epidemiology and biostatistics harvard school of public health
Verified email at hsph.harvard.edu

statistics epidemiology biostatistics causal inference

TITLE	CITED BY	YEAR
Marginal structural models and causal inference in epidemiology JM Robins, MA Hernan, B Brumback Epidemiology, 550-560	5261	2000
Causal diagrams for epidemiologic research S Greenland, J Pearl, JM Robins Epidemiology, 37-48	3915	1999
Estimation of regression coefficients when some regressors are not always observed JM Robins, A Rotnitzky, LP Zhao Journal of the American statistical Association 89 (427), 846-866	3128	1994
Causal inference MA Hernán, JM Robins CRC	3073	2010
Causal diagrams for empirical research J Pearl Biometrika 82 (4), 669-688	2834	1995
A structural approach to selection bias MA Hernán, S Hernández-Díaz, JM Robins Epidemiology, 615-625	2603	2004