主分层与中介分析

罗姗姗

March, 2023

目录

目录

后处理混杂变量

到目前为止，我们讨论的大多数问题都是调整前处理混杂变量，即协变量。

在治疗后（但在最终结果之前）发生混杂会对因果推断产生不同的挑战。

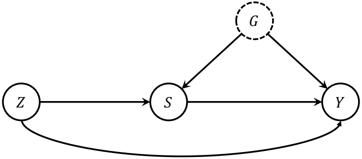
后处理混杂：一个后处理中间变量 S 位于 Z 和 Y 之间的因果路径上：

$$Z \longrightarrow S \longrightarrow Y$$

在经济学中被称为“内生选择问题”。

Rosenbaum（1984）表明：以与前处理协变量 X 相同的方式调整后处理变量 D 会导致偏倚的因果作用。包括一系列（看似不同的）问题。

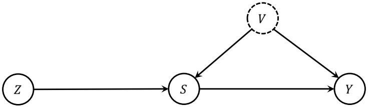
主分层框架



S_1	S_0	G	Description
1	1	ss	Always-taker
1	0	$s\bar{s}$	Complier
0	1	$\bar{s}s$	Defier
0	0	$\bar{s}\bar{s}$	Never-taker

- Z: 二值处理变量,$Z=1$ 表示接受处理, $Z=0$ 表示接受对照
- S: 二值中间变量
- Y: 表示感兴趣的结果变量
- S_z: 表示在接受处理分配 $Z=z$ 后的中间变量的潜在结果
- G: 主分层变量, 定义为 $G=(S_1,S_0)$.

不依从问题



S_1	S_0	G	Description
1	1	ss	Always-taker
1	0	$s\bar{s}$	Complier
0	1	$\bar{s}s$	Defier
0	0	$\bar{s}\bar{s}$	Never-taker

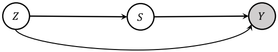
令 $Z = 1$ 表示成长于大学附近, $Z = 0$ 表示没有成长于大学附近; $S = 1$ 表示完成了高中学业, $S = 0$ 表示没有. 令 Y 表示对数收入. Angrist et al. (1996) 讨论了依从组上的因果作用的识别性:

$$\mathbb{E}(Y_{S=1} - Y_{S=0} \mid S_1 = S_0 = 1).$$

关于上述参数的识别和估计参考工具变量在单调性假设下的识别性。

目录

死亡删失



S_1	S_0	G	Description	Y_1	Y_0
1	1	ss	Always-survivors	✓	✓
1	0	$s\bar{s}$	Protected	✓	×
0	1	$\bar{s}s$	Harmed	×	✓
0	0	$\bar{s}\bar{s}$	Doomed	×	×

处理 $Z=1$ 表示接受安慰剂，$Z=0$ 表示接受毒性处理；$S=1$ 表示存活，$S=0$ 表示死亡；Y 是感兴趣的结果变量 (e.g., 体重，注意 $S=0$ 时所对应的 Y 没有定义).

定义及记号

- Z：表示二值的处理变量:$Z \in \{1, \ldots, m\}$
- X：表示处理前协变量
- S：表示生存状态，其中 $S = 1$ 表示生存，$S = 0$ 表示死亡
- Y：表示感兴趣的结果变量
- S_z：表示在接受处理分配 $Z = z$ 后的生存状况的潜在结果
- Y_z：表示在接受处理分配 $Z = z$ 后的潜在结果
- G：主分层变量，定义为 $G = (S_0\ S_1)$

Assumption 1 (可忽略性)

$(S_0, S_1, Y_0, Y_1) \perp\!\!\!\perp Z \mid X.$

Assumption 2 (单调性)

$S_1 \geq S_0$.

单调性假设下的主分层变量

S_1	S_0	G	Description	Y_1	Y_0
1	1	ss	Always-survivors	✓	✓
1	0	$s\bar{s}$	Protected	✓	$\times$
0	0	$\bar{s}\bar{s}$	Doomed	$\times$	$\times$

我们将感兴趣的参数定义为:

$$\Delta = E(Y_{Z=1} - Y_{Z=0} \mid G = ss).$$

- 在可忽略性假设及单调性假设下: 主分层的权重 $\pi_g(X) = \mathrm{pr}(G = g \mid X)$ 可识别.

$$
\begin{aligned}
\mathrm{pr}(S=1 \mid Z=1, X) &= \mathrm{pr}(S_{Z=1}=1 \mid X) \\
&= \mathrm{pr}(S_{Z=1}=1, S_{Z=0}=1 \mid X) + \mathrm{pr}(S_{Z=1}=1, S_{Z=0}=0 \mid X) \\
&= \mathrm{pr}(G = ss \mid X) + \mathrm{pr}(G = s\bar{s} \mid X), \\
\mathrm{pr}(S=1 \mid Z=0, X) &= \mathrm{pr}(S_{Z=0}=1 \mid X) \\
&= \mathrm{pr}(S_{Z=0}=1, S_{Z=1}=1 \mid X) + \mathrm{pr}(S_{Z=0}=1, S_{Z=1}=0 \mid X) \\
&= \mathrm{pr}(G = ss \mid X) + 0, \\
\mathrm{pr}(S=0 \mid Z=1, X) &= \mathrm{pr}(S_{Z=1}=0 \mid X) \\
&= \mathrm{pr}(S_{Z=0}=1, S_{Z=1}=0 \mid X) + \mathrm{pr}(S_{Z=0}=0, S_{Z=1}=0 \mid X) \\
&= 0 + \mathrm{pr}(G = \bar{s}\bar{s} \mid X).
\end{aligned}
$$

在 X 的每一层上，等式左边均可观测，等式右边有三个未知数，可识别。

- 在可忽略性假设下:

$$
\begin{aligned}
\mu_z =& E(Y_z \mid G = ss) \\
=& E\{\underbrace{E(Y_z \mid G = ss, X)}_{\mu_z(X)=E(Y|Z=z,G=ss)} \mid G = ss\} \\
=& E\{\mu_z(X) \mid G = ss\} \\
=& \frac{E\{\mu_z(X)\mathbb{I}(G = ss)\}}{\text{pr}(G = ss)} \\
=& \frac{E[E\{\mu_z(X)\mathbb{I}(G = ss) \mid X\}]}{\text{pr}(G = ss)} \\
=& \frac{E[\mu_z(X)E\{\mathbb{I}(G = ss) \mid X\}]}{\text{pr}(G = ss)} \\
=& \frac{E\left\{\mu_z(X)\pi_{ss}(X)\right\}}{E\{\pi_{ss}(X)\}},
\end{aligned}
$$

其中 $\pi_g(X) = \text{pr}(G = g \mid X)$ 是可识别的，我们只需要证明 $\mu_z(X)$ 的可识别性.

我们将感兴趣的参数定义为:

$$\begin{aligned}\Delta &= E(Y_{Z=1} - Y_{Z=0} \mid G = ss) \\ &= \underbrace{E(Y_{Z=1} \mid G = ss)}_{\mu_1} - \underbrace{E(Y_{Z=0} \mid G = ss)}_{\mu_0}.\end{aligned}$$

我们首先说明 μ_0 是可识别的，只需要 $E(Y \mid Z = 0, G = ss, X)$ 的识别性即可

$$\begin{aligned}\mu_0 &= \frac{E\{\mu_0(X)\pi_{ss}(X)\}}{E\{\pi_{ss}(X)\}} \\ &= \frac{E\{E(Y \mid Z = 0, G = ss, X)\pi_{ss}(X)\}}{E\{\pi_{ss}(X)\}} \\ &= \frac{E\{E(Y \mid Z = 0, S = 1, X)\pi_{ss}(X)\}}{E\{\pi_{ss}(X)\}}\end{aligned}$$

类似的，关于参数 μ_1, 我们也只需要证明 $\mu_1(X) = E(Y \mid Z = 1, G = ss, X)$ 的识别性.

Outline

关键识别假设

Assumption 3 (辅助变量)

我们假设辅助变量 X 满足: $Y \perp\!\!\!\perp X \mid (Z, G)$.

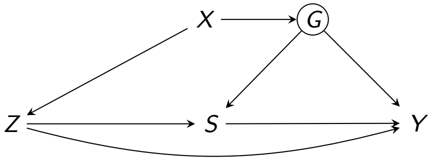


图: 辅助变量

$$E(Y \mid Z = 1, G = ss, X) = E(Y \mid Z = 1, G = ss)$$

识别性 I

定理 1 (非参识别性)

给定可忽略性、单调性及替代变量假设，因果参数 $\mu_1(X) = E(Y \mid Z = 1, G = ss, X)$ 在一些正则条件下是可识别的.

证明

$$
\begin{aligned}
E(Y \mid Z = 1, S = 1, X) &= E(Y \mid Z = 1, S = 1, X, G = ss)\mathrm{pr}(G = ss \mid X) \\
&\quad + E(Y \mid Z = 1, S = 1, X, G = s\bar{s})\mathrm{pr}(G = s\bar{s} \mid X) \\
&= E(Y \mid Z = 1, G = ss)\mathrm{pr}(G = ss \mid X) \\
&\quad + E(Y \mid Z = 1, G = s\bar{s})\mathrm{pr}(G = s\bar{s} \mid X)
\end{aligned}
$$

识别性 II

证明

我们取 $X=1,0$,

$$
\begin{aligned}
E(Y \mid Z=1, S=1, X=1) = {} & E(Y \mid Z=1, G=ss)\mathrm{pr}(G=ss \mid X=1) \\
& + E(Y \mid Z=1, G=s\bar{s})\mathrm{pr}(G=s\bar{s} \mid X=1) \\
E(Y \mid Z=1, S=1, X=0) = {} & E(Y \mid Z=1, G=ss)\mathrm{pr}(G=ss \mid X=0) \\
& + E(Y \mid Z=1, G=s\bar{s})\mathrm{pr}(G=s\bar{s} \mid X=0)
\end{aligned}
$$

由于 $\mathrm{pr}(G=s\bar{s} \mid X)$ 是可识别的，上述两个方程，两个未知数，在正则性条件得到保证的基础上，参数 $\mu_1(X)=E(Y \mid Z=1, G=ss, X)=E(Y \mid Z=1, G=ss)$和 $E(Y \mid Z=1, G=s\bar{s})$ 是可识别的.

非参数估计 I

对于 X 的每个可能取值 x，简便起见，我们假设 $x = 0, 1$

1. 我们如下估计 $\hat{\mathrm{pr}}(G = ss \mid X = x) = \frac{\sum_{i=1}^N \mathbb{I}(S_i=1, Z_i=0, X_i=x)}{\sum_{i=1}^N \mathbb{I}(Z_i=0, X_i=x)}$.
2. 我们如下估计 $\hat{\mathrm{pr}}(G = \bar{s}\bar{s} \mid X = x) = \frac{\sum_{i=1}^N \mathbb{I}(S_i=0, Z_i=1, X_i=x)}{\sum_{i=1}^N \mathbb{I}(Z_i=1, X_i=x)}$.
3. 我们如下估计 $\hat{\mathrm{pr}}(G = s\bar{s} \mid X = x) = 1 - \hat{\mathrm{pr}}(G_i = ss \mid X_i = x) - \hat{\mathrm{pr}}(G_i = \bar{s}\bar{s} \mid X_i = x)$.
4. 我们如下估计 $\hat{E}(Y \mid Z = 0, S = 1, X = x) = \frac{\sum_{i=1}^N \mathbb{I}(S_i=1, Z_i=0, X_i=x, Y_i)}{\sum_{i=1}^N \mathbb{I}(Z_i=0, S_i=1, X_i=x)}$.
5. 我们如下估计 $\mu_0 = E(Y_0 \mid G = ss)$,

$$\hat{\mu}_0 = \frac{\sum_{i=1}^N \hat{E}(Y \mid Z = 0, S = 1, X_i)\hat{\mathrm{pr}}(G_i = ss \mid X_i)}{\sum_{i=1}^N \hat{\mathrm{pr}}(G_i = ss \mid X_i)}$$

6. 我们如下估计 $\hat{E}(Y \mid Z = 1, S = 1, X = x) = \frac{\sum_{i=1}^N \mathbb{I}(S_i=1, Z_i=1, X_i=x, Y_i)}{\sum_{i=1}^N \mathbb{I}(Z_i=1, S_i=1, X_i=x)}$.

非参数估计 II

7. 我们如下估计 $\hat{E}(Y \mid Z=1, G=ss)$:

$$\hat{E}(Y \mid Z=1, S=1, X=1) = \gamma_{ss}\hat{\text{pr}}(G=ss \mid X=1) + \gamma_{s\bar{s}}\hat{\text{pr}}(G=s\bar{s} \mid X=1),$$
$$\hat{E}(Y \mid Z=1, S=1, X=0) = \gamma_{ss}\hat{\text{pr}}(G=ss \mid X=0) + \gamma_{s\bar{s}}\hat{\text{pr}}(G=s\bar{s} \mid X=0),$$

   其中 $\gamma_{ss} = E(Y \mid Z=1, G=ss)$, $\gamma_{s\bar{s}} = E(Y \mid Z=1, G=s\bar{s})$.

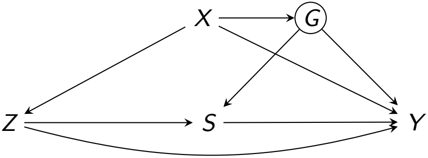
8. 我们如下估计 $\mu_1 = E(Y_1 \mid G=ss)$,

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{N} \gamma_{ss}\hat{\text{pr}}(G_i=ss \mid X_i)}{\sum_{i=1}^{N} \hat{\text{pr}}(G_i=ss \mid X_i)} = \gamma_{ss}.$$

Outline

识别性 I

在现实中，当难以找到辅助变量满足图 1 时，考虑参数模型是更为合适的。



识别性 II

不失一般性，我们考虑以下线性模型:

$$E(Y \mid Z=1, G=g, X=x) = \beta_0 + \beta_{ss}\mathbb{I}(G=ss) + \beta_x X.$$

$$\Rightarrow E(Y \mid Z=1, G=ss, X=x) = \beta_0 + \beta_{ss} + \beta_x X.$$

$$\Rightarrow E(Y \mid Z=1, G=s\bar{s}, X=x) = \beta_0 + \beta_x X.$$

定理 2 (线性识别)

给定可忽略性、单调性及线性模型假设，因果参数 $\mu_1(X) = E(Y \mid Z=1, G=ss, X)$ 在一些正则条件下是可识别的.

识别性 III

证明

$$
\begin{aligned}
E(Y \mid Z=1, S=1, X) &= E(Y \mid Z=1, S=1, X, G=ss)\mathrm{pr}(G=ss \mid X) \\
&\quad + E(Y \mid Z=1, S=1, X, G=s\bar{s})\mathrm{pr}(G=s\bar{s} \mid X) \\
&= (\beta_0 + \beta_{ss} + \beta_x X)\mathrm{pr}(G=ss \mid X) + (\beta_0 + \beta_x X)\mathrm{pr}(G=s\bar{s} \mid X) \\
&= \beta_0 + \beta_{ss}\mathrm{pr}(G=ss \mid X) + \beta_x X
\end{aligned}
$$

$$
\begin{aligned}
E(Y \mid Z=1, S=1, X=0) &= \beta_0 + \beta_{ss}\mathrm{pr}(G=ss \mid X=0) + 0\beta_x \\
E(Y \mid Z=1, S=1, X=1) &= \beta_0 + \beta_{ss}\mathrm{pr}(G=ss \mid X=1) + \beta_x, \\
E(Y \mid Z=1, S=1, X=2) &= \beta_0 + \beta_{ss}\mathrm{pr}(G=ss \mid X=2) + 2\beta_x.
\end{aligned}
$$

三个方程，三个未知数，上述参数 $(\beta_0, \beta_{ss}, \beta_x)$ 是可识别的.

非参数估计 I

对于 X 的每个可能取值 x，简便起见，我们假设 $x=0,1,2$

1. 我们如下估计 $\hat{\text{pr}}(G=ss \mid X=x)=\frac{\sum_{i=1}^N \mathbb{I}(S_i=1,Z_i=0,X_i=x)}{\sum_{i=1}^N \mathbb{I}(Z_i=0,X_i=x)}$.
2. 我们如下估计 $\hat{\text{pr}}(G=\bar{s}\bar{s} \mid X=x)=\frac{\sum_{i=1}^N \mathbb{I}(S_i=0,Z_i=1,X_i=x)}{\sum_{i=1}^N \mathbb{I}(Z_i=1,X_i=x)}$.
3. 我们如下估计 $\hat{\text{pr}}(G=s\bar{s} \mid X=x)=1-\hat{\text{pr}}(G_i=ss \mid X_i=x)-\hat{\text{pr}}(G_i=\bar{s}\bar{s} \mid X_i=x)$.
4. 我们如下估计 $\hat{E}(Y \mid Z=0,S=1,X=x)=\frac{\sum_{i=1}^N \mathbb{I}(S_i=1,Z_i=0,X_i=x,Y_i)}{\sum_{i=1}^N \mathbb{I}(Z_i=0,S_i=1,X_i=x)}$.
5. 我们如下估计 $\mu_0=E(Y_0 \mid G=ss)$,

$$\hat{\mu}_0=\frac{\sum_{i=1}^N \hat{E}(Y \mid Z=0,S=1,X_i)\hat{\text{pr}}(G_i=ss \mid X_i)}{\sum_{i=1}^N \hat{\text{pr}}(G_i=ss \mid X_i)}.$$

6. 我们如下估计 $\hat{E}(Y \mid Z=1,S=1,X=x)=\frac{\sum_{i=1}^N \mathbb{I}(S_i=1,Z_i=1,X_i=x,Y_i)}{\sum_{i=1}^N \mathbb{I}(Z_i=1,S_i=1,X_i=x)}$.

非参数估计 II

7. 我们通过求解下述方程获得参数 $\hat{\beta}_0, \hat{\beta}_{ss}, \hat{\beta}_x$:

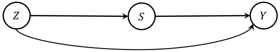
$$\hat{E}(Y \mid Z=1, S=1, X=0) = \beta_0 + \beta_{ss}\hat{\text{pr}}(G=ss \mid X=0) + 0\beta_x$$
$$\hat{E}(Y \mid Z=1, S=1, X=1) = \beta_0 + \beta_{ss}\hat{\text{pr}}(G=ss \mid X=1) + \beta_x,$$
$$\hat{E}(Y \mid Z=1, S=1, X=2) = \beta_0 + \beta_{ss}\hat{\text{pr}}(G=ss \mid X=2) + 2\beta_x.$$

8. 我们如下估计 $\hat{\mu}_1(X)$: $\hat{\mu}_1(X) = \beta_0 + \beta_{ss}\mathbb{I}(G=ss) + \beta_x X$.
9. 我们如下估计 $\mu_1 = E(Y_1 \mid G=ss)$,

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{N} \hat{\mu}_1(X)\hat{\text{pr}}(G_i = ss \mid X_i)}{\sum_{i=1}^{N} \hat{\text{pr}}(G_i = ss \mid X_i)}.$$

目录

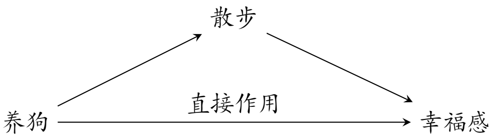
替代指标评估



S_1	S_0	G	描述	$E(Y_1 - Y_0 \mid G)$
1	1	ss	causal necessity	0
1	0	$s\bar{s}$	causal sufficiency	$\neq 0$
0	1	$\bar{s}s$	causal sufficiency	$\neq 0$
0	0	$\bar{s}\bar{s}$	causal necessity	0

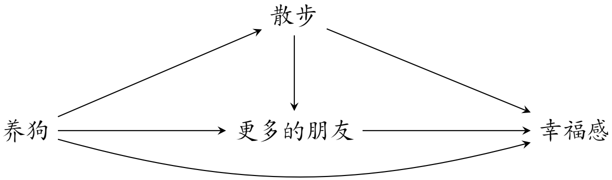
处理 $Z=1$ 表示接受处理, $Z=0$ 表示接受对照; $S=1$ 表示三年内未复发癌症, $S=0$ 表示复发; $Y=1$ 表示存活, $Y=0$ 表示死亡.

Toy Example I



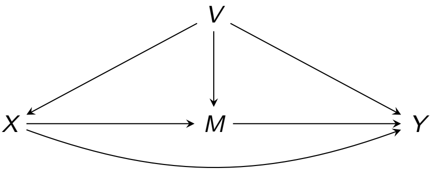
- 这个图表展示了散步和养狗对幸福感的影响。
- 箭头从养狗直接指向幸福感，表示养狗对幸福感的影响是直接的。
- 具体来说，养狗通过对散步的促进，间接地提高了幸福感。这可以解释为养狗让人们更有动力去散步，而散步本身对幸福感有积极影响。从而，表示养狗间接地促进了散步，从而提升幸福感。

中介分析 I



- 在这个新的图表中，养狗对于更多的朋友有直接作用，更多的朋友对于幸福感有间接作用。散步与更多的朋友之间也有一条边，表示散步可能帮助人们交到更多的朋友。
- 此时有两个中间变量，散步和更多的朋友
- 在本课程中，我们只考虑一个中间变量

两种经典的线性建模方法 I



❶ (做差法) 考虑对结果变量建模, 不纳入中介变量:

$$E(Y \mid X = x, V = v) = \theta_0 + \theta_1 x + \theta_2 v$$

考虑对结果变量建模, 纳入中介变量:

$$E(Y \mid X = x, M = m, V = v) = \beta_0 + \beta_1 x + \beta_2 m + \beta_3 v \tag{1}$$

其中 β_1 为直接作用; $\theta_1 - \beta_1$ 为间接作用。

两种经典的线性建模方法 II

2. (做乘法) 考虑对中介变量建立线性模型:

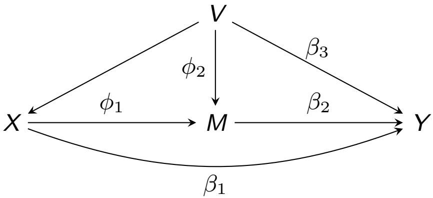
$$E(M \mid X = x, V = v) = \phi_0 + \phi_1 x + \phi_2 v$$

考虑对结果变量建模，纳入中介变量:

$$E(Y \mid X = x, M = m, V = v) = \beta_0 + \beta_1 x + \beta_2 m + \beta_3 v \tag{2}$$

其中 β_1 直接作用; $\phi_1 \times \beta_2$ 间接作用。

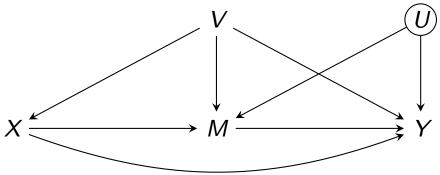
两种经典的线性建模方法 III



其中 $\phi_1, \phi_2, \beta_1, \beta_2$ 和 β_3 分别标识对应的系数。注意到间接作用的系数 $\phi_1\beta_2$ 没有单独的箭头，但可以从路径 $X \rightarrow M \rightarrow Y$ 上的系数相乘得到。

- 不能处理其它类型的结果、中介变量以及非线性模型。
- 需要提出独立于函数形式的直接和间接作用的定义

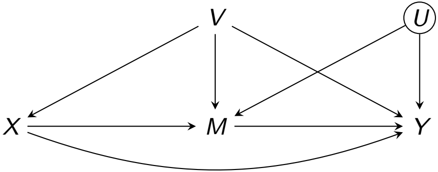
传统方法的限制



问题 1: 中介-结果的混杂变量

即使原因是随机的，或者所有原因-结果 ($X - Y$) 的混杂变量都包含在模型中，仍然可能存在中介-结果 ($M - Y$) 的未观测混杂变量 U。如果不控制中介-结果的混杂变量 U，则标准方法的结果可能高度偏倚。

Limitation 1–忽略 $X-Y$ 之间的混杂变量



- 即使原因是随机的，或者所有原因-结果 $(X-Y)$ 的混杂变量都包含在模型中，仍然可能存在中介-结果 $(M-Y)$ 的未观测混杂变量 U。
- 此时(1)和(2) 都不能相合估计
- 如果不控制中介-结果的混杂变量 U，则可能得到有偏的结论。

Limitation 2: 交互作用

当回归模型中包含交互项时：

$$E(Y \mid X = x, V = v) = \theta_0 + \theta_1 x + \theta_2 v$$

$$E(Y \mid X = x, M = m, V = v) = \beta_0 + \beta_1 x + \beta_2 m + \beta_3 v + \beta_4 xm$$

由于存在交互项 XM (系数 θ_4)，此时如果直接使用做差法得到的因果作用，

$$\text{间接作用} = \theta_1 - \beta_1$$

$$\text{直接作用} = \beta_1$$

的意义并不明确

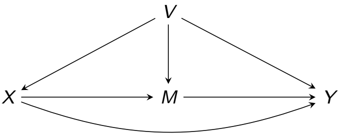
反事实框架

- Y：每个个体的感兴趣的结果变量
- X：每个个体的原因变量或感兴趣的处理变量
- M：每个个体的后处理中间变量（可能在 X 和 Y 之间的路径上）
- V：每个个体的协变量集合
- Y_x：当干预将 X 设为 x 时，每个个体的反事实结果变量 Y
- M_x：当干预将 X 设为 x 时，每个个体的反事实中间变量 M
- $Y_{x,m}$：当干预将 X 设为 x 且将 M 设为 m 时，每个个体的反事实结果变量 Y

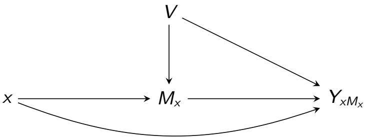
反事实框架

i	X	M	Y	M_0	M_1	Y_{00}	Y_{01}	Y_{10}	Y_{11}
1	1	1	25	0	1	18	18	20	25
2	1	0	40	1	0	45	43	40	48
3	0	1	40	1	0	34	40	41	38
4	0	0	30	0	0	30	20	23	25
...	...	...	...	...	...	...	...	...	...

总因果作用



(a) 原始的有向无环图



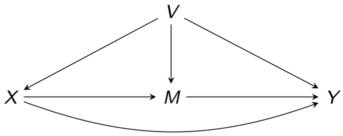
(b) 对节点 X 进行干预的有向无环图

总因果作用（Total Causal Effect）的反事实定义为：

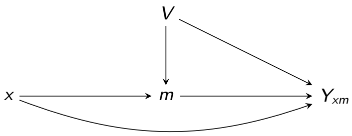
$$\mathrm{ATE} = \mathbb{E}\left(Y_{1M_1}\right) - \mathbb{E}\left(Y_{0M_0}\right) = \mathbb{E}\left(Y_1\right) - \mathbb{E}\left(Y_0\right)$$

其中，Y_x 代表未干预时的结果，Y_{1M_1} 代表在干预后所有可能的结果，Y_x 代表干预后的结果。

平均控制直接作用



(a) 原始的有向无环图



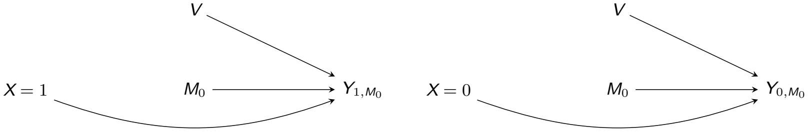
(b) 对节点 X 进行干预的有向无环图

平均控制直接作用（Average Controlled Direct Effect）的反事实定义为：

$$\mathrm{CDE}(m) = \mathbb{E}\left(Y_{1,m}\right) - \mathbb{E}\left(Y_{0,m}\right)$$

- 即在控制散步 $M = m$ 的情况下，养狗直接对幸福感的作用。这是外部可以施加的干预，可以帮助政策制定或评估。
- $\mathrm{CDE}(m)$ 受 m 的任意选择的影响。
- 我们如何获得间接作用？不能简单地从总作用中减去 CDE。

平均自然直接作用



(a) (b)

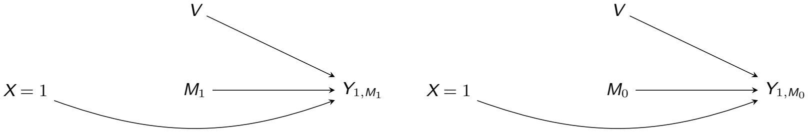
平均自然直接作用（Average Natural Direct Effect）的反事实定义为：

$$\mathrm{NDE}(0) = \mathbb{E}\left(Y_{1,M_0}\right) - \mathbb{E}\left(Y_{0,M_0}\right)$$

$$\mathrm{NDE}(1) = \mathbb{E}\left(Y_{1,M_1}\right) - \mathbb{E}\left(Y_{0,M_1}\right)$$

- 自然（natural）是相较于控制（controlled）而言的，此时 M_0 是在 $Z=0$ 下散步状态的潜在结果，外部无法施加干预。
- $\mathrm{NDE}(0)$: 所有人 M 的取值都是没养狗下的散步状态。在随机分配养狗与不养狗时，处理组和对照组的幸福感的差异。

平均自然间接作用



(a) (b)

平均自然间接作用（Average Natural Indirect Effect）的反事实定义为：

$$\mathrm{NIE}(1) = \mathbb{E}\left(Y_{1,M_1}\right) - \mathbb{E}\left(Y_{1,M_0}\right)$$

$$\mathrm{NIE}(0) = \mathbb{E}\left(Y_{0,M_1}\right) - \mathbb{E}\left(Y_{0,M_0}\right)$$

定义整理 I

- 控制直接作用 (CDE)：比较干预 $X = 1$ 和 $X = 0$ 时，在 $M = m$ 时的结果变化

$$\mathrm{CDE}(m) = \mathbb{E}\left(Y_{1m} - Y_{0m}\right)$$

- 自然直接作用 (NDE)：比较干预 $X = 1$ 和 $X = 0$ 时，在 $M = M_0$ 时的结果变化

$$\mathrm{NDE}(0) = \mathbb{E}\left(Y_{1M_0} - Y_{0M_0}\right)$$

- 自然间接作用 (NIE)：比较在 $X = 1$ 时，干预 $M = M_1$ 和 $M = M_0$ 时的结果变化

$$\mathrm{NIE}(1) = \mathbb{E}\left(Y_{1M_1} - Y_{1M_0}\right)$$

定义整理 II

- 总作用可以分解为直接作用和间接作用：

$$
\begin{aligned}
Y_1 - Y_0 = Y_{1M_1} - Y_{0M_0} \\
&= (Y_{1M_1} - Y_{1M_0}) + (Y_{1M_0} - Y_{0M_0}) \\
&= \mathrm{NIE}(1) + \mathrm{NDE}(0) \\
&= \mathrm{NIE}(0) + \mathrm{NDE}(1)
\end{aligned}
$$

- 整体效应的贡献主要来自于哪个因素？
  - 直接因果机制?
  $$\frac{\mathrm{NDE}(0)}{\mathrm{ATE}}$$
  - 间接因果机制?
  $$\frac{\mathrm{NIE}(1)}{\mathrm{ATE}}$$
- 自然直接作用和间接作用的定义不预设原因和中介变量对结果变量没有交互作用。

定义整理 III

- 将总作用分解为自然直接作用和间接作用的作用分解也不预设原因和中介变量对结果变量没有交互作用。
- 自然直接作用和间接作用对作用分解很有用；一般来说，控制直接作用并不是很有用。

反事实框架

i	X	M	Y	M_0	M_1	$Y_{0,0}$	$Y_{0,1}$	$Y_{1,0}$	$Y_{1,1}$
1	1	1	25	0	1	18	18	20	25
2	1	0	40	1	0	45	43	40	48
3	0	1	40	1	0	34	40	41	38
4	0	0	30	0	0	30	20	23	25
...	...	...	...	...	...	...	...	...	...

对于个体 2:

$$\mathrm{TE} = Y_1 - Y_0 = Y_{1,M_1} - Y_{0,M_0} = Y_{1,0} - Y_{0,1} = 40 - 43 = -3$$

$$\mathrm{CDE}(1) = Y_{1,1} - Y_{0,1} = 48 - 43 = 5$$

$$\mathrm{CDE}(0) = Y_{1,0} - Y_{0,0} = 40 - 45 = -5$$

$$\mathrm{NDE}(1) = Y_{1,M_0} - Y_{0,M_0} = Y_{1,1} - Y_{0,1} = 48 - 43 = 5$$

$$\mathrm{NIE}(1) = Y_{1,M_1} - Y_{1,M_0} = Y_{1,0} - Y_{1,1} = 40 - 48 = -8$$

在线性模型下的中介分析 I

- 我们考虑如下回归模型：

$$
\begin{aligned}
Y &= \theta_0 + \theta_1 X + \theta_2 M + \theta_3 XM + \theta_4^{\mathrm{T}} A + \epsilon_1, \\
M &= \beta_0 + \beta_1 X + \beta_2^{\mathrm{T}} A + \epsilon_2.
\end{aligned}
$$

- 首先考虑总作用 $\mathrm{ATE}(x, x^*)$,

$$
\begin{aligned}
M_x &= \beta_0 + \beta_1 x + \beta_2^{\mathrm{T}} A + \epsilon_2, \\
Y_{x, M_x} &= \theta_1 x + \theta_2 M_x + \theta_3 x M_x + \epsilon_1 + \theta_4^{\mathrm{T}} A, \\
Y_{x^*, M_{x^*}} &= \theta_1 x^* + \theta_2 M_{x^*} + \theta_3 x^* M_{x^*} + \epsilon_1 + \theta_4^{\mathrm{T}} A,
\end{aligned}
$$

在线性模型下的中介分析 II

$$\begin{aligned}
\mathrm{ATE}(x, x^*; x) &= \mathbb{E}(Y_{x, M_x} - Y_{x^*, M_{x^*}}) \\
&= \theta_1(x - x^*) + \mathbb{E}\{\theta_2(M_x - M_{x^*})\} + \mathbb{E}\{\theta_3(xM_x - x^* M_{x^*})\} \\
&= \theta_1(x - x^*) + \theta_2\mathbb{E}\{(M_x - M_{x^*})\} \\
&\quad + \theta_3\mathbb{E}\Big\{x(\beta_0 + \beta_1 x + \beta_2^{\mathrm{T}} A + \epsilon_2) - x^*(\beta_0 + \beta_1 x^* + \beta_2^{\mathrm{T}} A + \epsilon_2)\Big\} \\
&= \theta_1(x - x^*) + \theta_2\beta_1(x - x^*) \\
&\quad + \theta_3\mathbb{E}\Big\{(x - x^*)\beta_0 + \beta_1(x^2 - x^* x^*) + (x - x^*)\beta_2^{\mathrm{T}} A)\Big\} \\
&= (\theta_1 + \theta_2)\beta_1(x - x^*) + \theta_3\{\beta_0 + \beta_1(x + x^*) + \beta_2^{\mathrm{T}}\mathbb{E}(A)\}(x - x^*)
\end{aligned}$$

在线性模型下的中介分析 III

- 考虑控制直接作用 $\mathrm{CDE}(x, x^*; m)$，在 $M = m$ 的条件下，

$$
\begin{aligned}
Y_{x,m} &= \theta_1 x + \theta_2 m + \theta_3 xm + \epsilon_1 + \theta_4^{\mathrm{T}} A, \\
\mathrm{CDE}(x, x^*; m) &= \mathbb{E}(Y_{1,m} - Y_{0,m}) \\
&= (\theta_1 + \theta_3 m)(x - x^*).
\end{aligned}
$$

代入具体的数值，比如 $x = 1$，$x^* = 0$，$m = 0$，我们有：

$$
\mathrm{CDE}(1, 0; 0) = (\theta_1 + \theta_3 \cdot 0)(1 - 0) = \theta_1
$$

在线性模型下的中介分析 IV

- 自然直接作用 (NDE)

$$
\begin{aligned}
M_x &= \beta_0 + \beta_1 x + \beta_2^{\mathrm{T}} A + \epsilon_2, \\
Y_{x,{\color{red}m}} &= \theta_1 x + \theta_2 {\color{red}m} + \theta_3 x {\color{red}m} + \epsilon_1 + \theta_4^{\mathrm{T}} A, \\
Y_{x,{\color{red}M_{x^*}}} &= \theta_1 x + \theta_2 {\color{red}M_{x^*}} + \theta_3 x {\color{red}M_{x^*}} + \epsilon_1 + \theta_4^{\mathrm{T}} A, \\
Y_{x^*,{\color{red}M_{x^*}}} &= \theta_1 x^* + \theta_2 {\color{red}M_{x^*}} + \theta_3 x^* {\color{red}M_{x^*}} + \epsilon_1 + \theta_4^{\mathrm{T}} A,
\end{aligned}
$$

在线性模型下的中介分析 V

$$\begin{aligned}
\mathrm{NDE}(x, x^*; x^*) &= \mathbb{E}(Y_{x, M_{x^*}} - Y_{x^*, M_{x^*}}) \\
&= \mathbb{E}\{(\theta_1 + \theta_3 M_{x^*})(x - x^*)\} \\
&= \{\theta_1 + \theta_3 \mathbb{E}(M_{x^*})\}(x - x^*) \\
&= \Big[\theta_1 + \theta_3\{\beta_0 + \beta_1 x^* + \beta_2^{\mathrm{T}} E(A)\}\Big](x - x^*).
\end{aligned}$$

- 自然间接作用 (NIE):

$$\begin{aligned}
M_x &= \beta_0 + \beta_1 x + \beta_2^{\mathrm{T}} A + \epsilon_2, \\
Y_{x,m} &= \theta_1 x + \theta_2 m + \theta_3 x m + \epsilon_1 + \theta_4^{\mathrm{T}} A, \\
Y_{x,M_x} &= \theta_1 x + \theta_2 M_x + \theta_3 x M_x + \epsilon_1 + \theta_4^{\mathrm{T}} A, \\
Y_{x,M_{x^*}} &= \theta_1 x + \theta_2 M_{x^*} + \theta_3 x M_{x^*} + \epsilon_1 + \theta_4^{\mathrm{T}} A,
\end{aligned}$$

在线性模型下的中介分析 VI

$$
\begin{aligned}
\mathrm{NIE}(x, x^*; x) &= \mathbb{E}(Y_{x,M_x} - Y_{x,M_{x^*}}) \\
&= \mathbb{E}\{(\theta_2 + \theta_3 x)(M_x - M_{x^*})\} \\
&= (\theta_2 + \theta_3 x)\mathbb{E}\{(M_x - M_{x^*})\} \\
&= (\theta_2\beta_1 + \theta_3\beta_1 x)(x - x^*)
\end{aligned}
$$

- 如果感兴趣的是条件 NDE，则有：

$$
\mathbb{E}(Y_{x,M_{x^*}} - Y_{x^*,M_{x^*}} \mid V = v) = \left[\theta_1 + \theta_3\{\beta_0 + \beta_1 x^* + \beta_2^{\mathrm{T}} v\}\right](x - x^*)
$$

在线性模型下的中介分析 VII

- 注意，如果原因变量和中介变量对结果变量没有交互作用，使得 $\theta_3 = 0$，则这些表达式简化为：

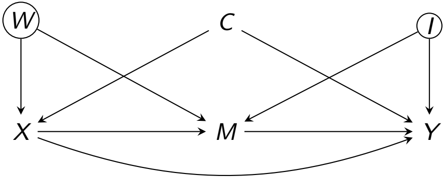
$$\mathrm{CDE}(x, x^*; m) = \theta_1(x - x^*)$$

$$\mathrm{NDE}(x, x^*; x^*) = \theta_1(x - x^*)$$

$$\mathrm{NIE}(x, x^*; x) = \theta_2\beta_1(x - x^*)$$

这些是社会科学文献中通常用于直接和间接效应的表达式（Baron 和 Kenny，1986）- “乘积法”。然而，与 Baron 和 Kenny (1986) 方法不同，即使在存在交互作用的情况下，也可以使用基于反事实定义和估计的这种直接和间接效应的方法。

总作用的识别性 I



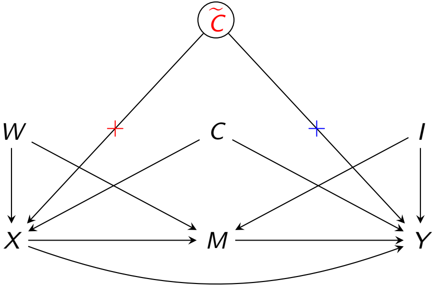
1. W 为 X 和 M 之间的混杂因素（允许未观测）
2. I 为 M 和 Y 之间的混杂因素（允许未观测）
3. 我们需要假定 X 和 Y 之间没有公共未观测混杂因素，即 $X \perp\!\!\!\perp Y_x \mid C$
4. 但注意 $X \not\!\perp\!\!\!\perp Y_x \mid (C, M)$，这是因为 $W \to M \leftarrow I$ 是一个 v 结构，条件在 M 将打开未观测混杂的路径，使得 X 和 Y 受到未观测混杂的影响.

总作用的识别性 II

5. 此时我们可以如下识别总作用

$$
\begin{aligned}
\mathbb{E}(Y_{1M_1} - Y_{0M_0}) &= \mathbb{E}(Y_1 - Y_0) \\
&= \mathbb{E}\{\mathbb{E}(Y_1 - Y_0 \mid C)\} \\
&= \mathbb{E}\{\mathbb{E}(Y_1 \mid C)\} - \mathbb{E}\{\mathbb{E}(Y_0 \mid C)\} \\
&= \mathbb{E}\{\mathbb{E}(Y_1 \mid C, X = 1)\} - \mathbb{E}\{\mathbb{E}(Y_0 \mid C, X = 0)\} \\
&= \mathbb{E}\{\mathbb{E}(Y \mid C, X = 1)\} - \mathbb{E}\{\mathbb{E}(Y \mid C, X = 0)\}
\end{aligned}
$$

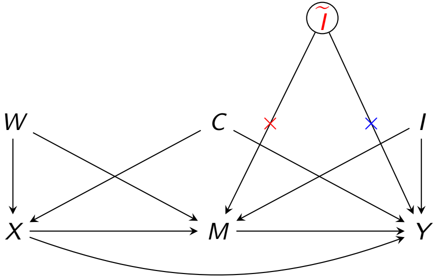
自然直接作用和自然间接作用的识别性 I



1. 在给定 C 的情况下，不存在未观测的原因-结果混杂因素 $\tilde{C}$,

$$Y_{x,m} \perp\!\!\!\perp X \mid C$$

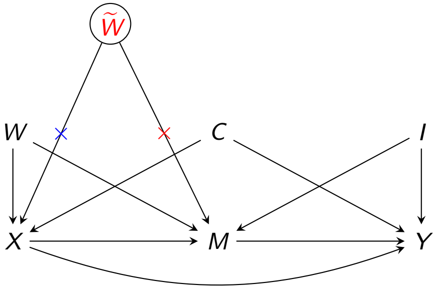
自然直接作用和自然间接作用的识别性 II



❷ 在给定 (X, I) 的情况下，不存在未观测的中介-结果混杂因素 $\widetilde{I}$,

$$Y_{x,m} \perp\!\!\!\perp M \mid (I, X)$$

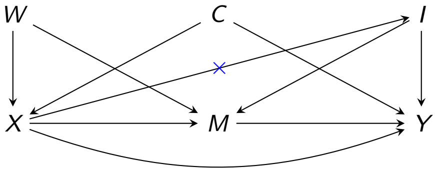
自然直接作用和自然间接作用的识别性 III



❸ 在给定 W 的情况下，不存在未观测的原因-中介混杂因素 $\widetilde{W}$,

$$M_a \perp\!\!\!\perp X \mid W$$

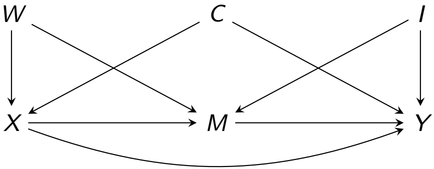
自然直接作用和自然间接作用的识别性 IV



4. 不存在受原因影响的中介-结果混杂因素（即不存在从 X 到 I 的箭头）

$$Y_{x,m} \perp\!\!\!\perp M_x^* \mid I$$

自然直接作用和自然间接作用的识别性 I



更正式地，以反事实符号表示，这些假设是：

1. $Y_{x,m} \perp\!\!\!\perp X \mid C$
2. $Y_{x,m} \perp\!\!\!\perp M \mid (I, X)$
3. $M_x \perp\!\!\!\perp X \mid W$
4. $Y_{x,m} \perp\!\!\!\perp M_{x^*} \mid I$

- 对于控制直接效应而言，只需要假设（1）和（2）。
- 当处理变量是随机分配时，假设（1）和（3）自然成立。

自然直接作用和自然间接作用的识别性 II

- 令 $V=(W,C,I)$, 我们接下来介绍 $E(Y_{1,M_0})$ 的识别性，$E(Y_{0,M_1})$ 可类似识别。

$$
\begin{aligned}
E(Y_{1,M_0}) &= \sum_{v=1}\sum_{m=1} E(Y_{1,M_0} \mid V=v, M_0=m)\mathrm{pr}(M_0=m, V=v) \\
&= \sum_{v=1}\sum_{m=1} E(Y_{1,m} \mid V=v, M_0=m)\mathrm{pr}(M_0=m, V=v) \\
&= \sum_{v=1}\sum_{m=1} E(Y_{1,m} \mid V=v)\mathrm{pr}(M_0=m, V=v) \\
&= \sum_{v=1}\sum_{m=1} E(Y_{1,m} \mid V=v, X)\mathrm{pr}(M_0=m, V=v) \\
&= \sum_{v=1}\sum_{m=1} E(Y_{1,m} \mid V=v, X, M)\mathrm{pr}(M_0=m, V=v) \\
&= \sum_{v=1}\sum_{m=1} E(Y \mid V=v, X=1, M=m)\mathrm{pr}(M_0=m, V=v) \\
&= \sum_{v=1}\sum_{m=1} E(Y \mid V=v, X=1, M=m)\mathrm{pr}(M_0=m \mid V=v)\mathrm{pr}(V=v) \\
&= \sum_{v=1}\sum_{m=1} E(Y \mid V=v, X=1, M=m)\mathrm{pr}(M=m \mid V=v, X=0)\mathrm{pr}(V=v),
\end{aligned}
$$

非参识别性

- 控制直接作用 (CDE):

$$\mathrm{CDE} = E\left(Y_{x,m} - Y_{x^*,m}\right) = \sum_{v} \left\{E(Y \mid x, m, v) - E\left(Y \mid x^*, m, v\right)\right\} P(V = v)$$

- 自然直接作用 (NDE):

$$\begin{aligned}\mathrm{NDE}(x^*) &= E\left(Y_{x,M_{x^*}} - Y_{x^*}, M_{x^*}\right) \\ &= \sum_{v=1}\sum_{m=1}\{E(Y \mid x, m, v) - E(Y \mid x^*, m, v)\} P(M = m \mid x^*, v) P(V = v)\end{aligned}$$

- 自然间接作用 (NIE):

$$\begin{aligned}\mathrm{NIE}(x) &= E\left(Y_{x,M_x} - Y_{x,M_{x^*}}\right) \\ &= \sum_{v=1}\sum_{m=1} E(Y \mid x, m, v)\left\{P(M = m \mid x, v) - P\left(M = m \mid x^*, v\right)\right\} P(V = v)\end{aligned}$$

三稳健估计量 I

Tchetgen Tchetgen & Shpitser (2012) 提出了一个关于 NDE, NIE 的三稳健估计量，我们以 NDE 为例介绍，它的估计量为：

$$
\begin{aligned}
\widehat{\mathrm{NDE}}^{\mathrm{tp}} =& \widehat{E}(Y_{1,M_0}) - \widehat{E}(Y_{0,M_0}) \\
=& \mathbb{P}_n\left[\frac{X\widehat{\mathrm{pr}}(M \mid X=0, V)}{\widehat{\mathrm{pr}}(X=1 \mid V)\widehat{\mathrm{pr}}(M \mid X=1, V)}\left\{Y - \widehat{E}(Y \mid X=1, M, V)\right\}\right. \\
&\left. + \frac{1-X}{\widehat{\mathrm{pr}}(X=0 \mid V)}\left\{\widehat{E}(Y \mid X=1, M, V) - \widehat{h}(1,0,V)\right\} + \widehat{h}(1,0,V)\right] \\
&- \underbrace{\mathbb{P}_n\left[\frac{1-X}{\widehat{\mathrm{pr}}(X=0 \mid V)}\left\{Y - \widehat{h}(0,0,V)\right\} + \widehat{h}(0,0,V)\right]}_{\widehat{E}(Y_{0,M_0})},
\end{aligned}
$$

其中

$$
\widehat{h}(x, x^*, V) = \int \widehat{E}(Y \mid X=x, M=m, V)\widehat{\mathrm{pr}}(m \mid X=x^*, V)dm
$$

三稳健估计量 II

当下述三个条件之一成立时，$\widehat{\mathrm{NDE}}^{\mathrm{tp}}$ 便是 NDE 的相合估计：

1. 结果变量模型 $\widehat{E}(Y \mid x, m, v)$ 以及中间变量条件密度 $P(M = m \mid x, v)$ 被正确设定;
2. 结果变量模型 $\widehat{E}(Y \mid x, m, v)$ 以及倾向评分 $\mathrm{pr}(X = 1 \mid v)$ 被正确设定;
3. 中间变量条件密度 $\mathrm{pr}(M = m \mid x, v)$ 以及倾向评分 $\mathrm{pr}(X = 1 \mid v)$ 被正确设定。

- 由于 $E\left(Y_{0,M_0}\right) = E\left(Y^0\right)$, 以及前一页估计量 $\widehat{E}\left(Y_{0,M_0}\right)$ 正是其双稳健估计量, 所以当上述三个条件之一成立时, $\widehat{E}\left(Y_{0,M_0}\right) \to E\left(Y_{0,M_0}\right)$ 。
- 我们下面只考虑 $\widehat{E}\left(Y_{1,M_0}\right)$ 。

三稳健估计量 III

- 假如 $\widehat{E}(Y \mid x, m, v)$ 以及 $\text{pr}(M = m \mid x, v)$ 被正确设定, 那么

$$\widehat{E}(Y \mid x, m, v) \to E(Y \mid x, m, v),$$
$$\widehat{\text{pr}}(M = m \mid x, v) \to \text{pr}(M = m \mid x, v),$$
$$\widehat{\text{pr}}(X = 1 \mid z) \to \text{pr}^*(X = 1 \mid z).$$

我们有:

三稳健估计量 IV

$$
\begin{aligned}
\widehat{E}(Y_{1,M_0}) = \mathbb{P}_n \Bigg[& \frac{X\widehat{\text{pr}}(M \mid X=0, V)}{\widehat{\text{pr}}(X=1 \mid V)\widehat{\text{pr}}(M \mid X=1, V)} \left\{ Y - \widehat{E}(Y \mid X=1, M, V) \right\} \\
& + \frac{1-X}{\widehat{\text{pr}}(X=0 \mid V)} \left\{ \widehat{E}(Y \mid X=1, M, V) - \widehat{h}(1,0,V) \right\} + \widehat{h}(1,0,V) \Bigg] \\
\to E \Bigg[& \frac{X\text{pr}(M \mid X=0, V)}{\text{pr}^*(X=1 \mid V)\text{pr}(M \mid X=1, V)} \{ Y - E(Y \mid X=1, M, V) \} \\
& + \frac{1-X}{\text{pr}^*(X=0 \mid V)} \{ E(Y \mid X=1, M, V) - h(1,0,V) \} + h(1,0,V) \Bigg] \\
= E\{ & 0 + 0 + h(1,0,V) \} = E(Y_{M_0}) \,.
\end{aligned}
$$

三稳健估计量 V

- 假如 $E(Y \mid x, m, v)$ 以及 $\text{pr}(X = 1 \mid z)$ 被正确设定, 那么

$$\widehat{E}(Y \mid x, m, v) \to E(Y \mid x, m, v)$$

$$\widehat{\text{pr}}(M = m \mid x, v) \to \text{pr}^*(M = m \mid x, v)$$

$$\widehat{\text{pr}}(X = 1 \mid z) \to \text{pr}(X = 1 \mid z)$$

三稳健估计量 VI

我们有:

$$
\begin{aligned}
\widehat{E}(Y_{1,M_0}) = \mathbb{P}_n \Bigg[& \frac{X\widehat{\text{pr}}(M \mid X=0,V)}{\widehat{\text{pr}}(X=1 \mid V)\widehat{\text{pr}}(M \mid X=1,V)} \left\{ Y - \widehat{E}(Y \mid X=1,M,V) \right\} \\
& + \frac{1-X}{\widehat{\text{pr}}(X=0 \mid V)} \left\{ \widehat{E}(Y \mid X=1,M,V) - \widehat{h}(1,0,V) \right\} + \widehat{h}(1,0,V) \Bigg] \\
\to E \Bigg[& \frac{X\text{pr}^*(M \mid X=0,V)}{P(X=1 \mid V)\text{pr}^*(M \mid X=1,V)} \{ Y - E(Y \mid X=1,M,V) \} \\
& + \frac{1-X}{P(X=0 \mid V)} \{ E(Y \mid X=1,M,V) - h^*(1,0,V) \} + h^*(1,0,V) \Bigg] \\
= E & \{ 0 + h(1,0,V) - h^*(1,0,V) + h^*(1,0,V) \} \\
= E & (Y_{1,M_0})
\end{aligned}
$$

三稳健估计量 VII

- 假如 $\text{pr}(X = 1 \mid z)$ 以及 $\text{pr}(X = 1 \mid z)$ 被正确设定, 那么

$$\widehat{E}(Y \mid x, m, v) \to E^{*}(Y \mid x, m, v),$$

$$\widehat{\text{pr}}(M = m \mid x, v) \to \text{pr}(M = m \mid x, v),$$

$$\widehat{\text{pr}}(X = 1 \mid z) \to \text{pr}(X = 1 \mid z)$$

三稳健估计量 VIII

我们有:

$$
\begin{aligned}
\widehat{E}(Y_{1,M_0}) &= \mathbb{P}_n\left[\frac{X\widehat{\mathrm{pr}}(M \mid X=0,V)}{\widehat{\mathrm{pr}}(X=1 \mid V)\widehat{\mathrm{pr}}(M \mid X=1,V)}\left\{Y-\widehat{E}(Y \mid X=1,M,V)\right\}\right.\\
&\qquad\left.+\frac{1-X}{\widehat{\mathrm{pr}}(X=0 \mid V)}\left\{\widehat{E}(Y \mid X=1,M,V)-\widehat{h}(1,0,V)\right\}+\widehat{h}(1,0,V)\right]\\
&\to E\left[\frac{X\mathrm{pr}(M \mid X=0,V)}{\mathrm{pr}(X=1 \mid V)\mathrm{pr}(M \mid X=1,V)}\{Y-E^*(Y \mid X=1,M,V)\}\right.\\
&\qquad\left.+\frac{1-X}{\mathrm{pr}(X=0 \mid V)}\{E^*(Y \mid X=1,M,V)-h^*(1,0,V)\}+h^*(1,0,V)\right]\\
&= E\{h(1,0,V)-h^*(1,0,V)+0+h^*(1,0,V)\}\\
&= E(Y_{1,M_0})
\end{aligned}
$$