



中国科学院大学
University of Chinese Academy of Sciences

基于 Python 的因果效应识别

CCF 因果科学暑期学校

计算技术研究所 李奉治

lifengzhi20z@ict.ac.cn

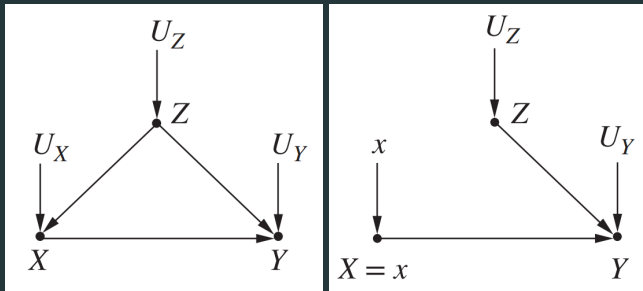
2021 年 8 月 16 日

大纲

1. 校正公式
2. 后门准则
3. 前门准则
4. Do-演算

校正公式

使用药物 (X)、痊愈 (Y) 和性别 (Z) 之间的关系



平均因果效应 Average Causal Effect, ACE

$$P(Y = 1 \mid do(X = 1)) - P(Y = 1 \mid do(X = 0))$$

校正公式

- 边缘概率 $P(Z = z)$ 在干预后不变，因为移除 Z 到 X 的箭头不会影响确定 Z 值的过程。

$$P_m(Z = z) = P(Z = z)$$

- 不管 X 是自发变化还是被故意操控发生变化， Y 对 X 和 Z 的响应函数 $Y = f(x, z, u_Y)$ 不变。

$$P_m(Y = y \mid Z = z, X = x) = P(Y = y \mid Z = z, X = x)$$

- Z 和 X 在修改后的模型中是 **d**-分离的。

$$P_m(Z = z \mid X = x) = P_m(Z = z) = P(Z = z)$$

$$P_m(Z = z) = P(Z = z)$$

$$P_m(Y = y \mid Z = z, X = x) = P(Y = y \mid Z = z, X = x)$$

$$P_m(Z = z \mid X = x) = P_m(Z = z) = P(Z = z)$$

$$P(Y = y \mid do(X = x))$$

$$= P_m(Y = y \mid X = x)$$

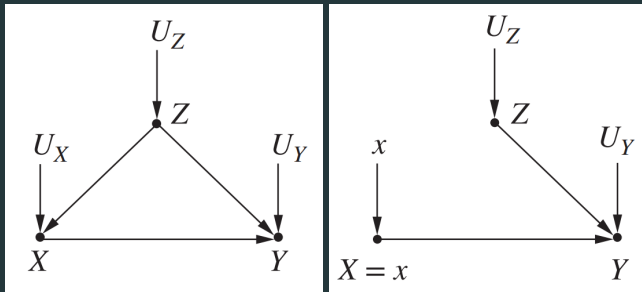
$$= \sum_z P_m(Y = y \mid X = x, Z = z) P_m(Z = z \mid X = x)$$

$$= \sum_z P_m(Y = y \mid X = x, Z = z) P_m(Z = z)$$

$$= \sum_z P(Y = y \mid X = x, Z = z) P(Z = z)$$

校正公式

使用药物 (X)、痊愈 (Y) 和性别 (Z) 之间的关系

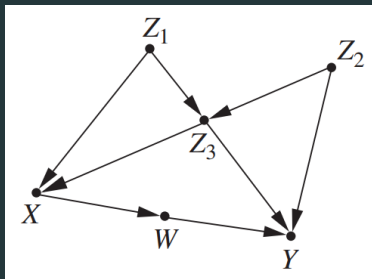


$$P(Y = y \mid do(X = x)) = \sum_z P(Y = y \mid X = x, Z = z)P(Z = z)$$

校正公式

- 边缘概率 $P(Z = z)$ 在干预后不变，因为移除 Z 到 X 的箭头不会影响确定 Z 值的过程。
- 不管 X 是自发变化还是被故意操控发生变化， Y 对 X 和 Z 的响应函数 $Y = f(x, z, u_Y)$ 不变。
- Z 和 X 在修改后的模型中是 **d**-分离的。

对于 X 的父节点集合是否满足呢？



- 边缘概率 $P(Z = z)$ 在干预后不变，因为移除 Z 到 X 的箭头不会影响确定 Z 值的过程。
- 不管 X 是自发变化还是被故意操控发生变化， Y 对 X 和 Z 的响应函数 $Y = f(x, z, u_Y)$ 不变。
- Z 和 X 在修改后的模型中是 **d**-分离的。

因果效应规则 - 给定一个图 G ，设变量 X 的父节点集合为 PA ，则 X 对 Y 的因果效应为

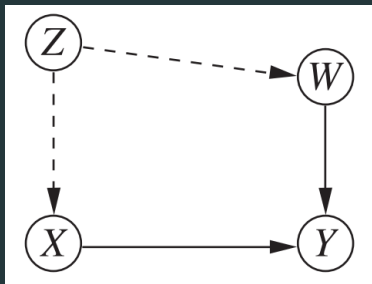
$$\begin{aligned} P(Y = y \mid do(X = x)) &= \sum_z P(Y = y \mid X = x, PA = z) P(PA = z) \\ &= \sum_z \frac{P(X = x, Y = y, PA = z)}{P(X = x \mid PA = z)} \end{aligned}$$

后门准则

后门准则

变量通常有不可观察的父节点，即虽然该父节点在图中表现出来了，但是节点的值却无法得到。在这些情况下，需要找到一个替代的变量集合用于校正。

新药物 (X)、痊愈率 (Y)、体重 (W) 和经济社会地位 (Z) 之间的关系



后门准则

后门准则 - 给定有向无环图中的一对有序变量 (X, Y) ，如果变量集合 Z 满足：

1. Z 中没有 X 的后代结点，
2. 且 Z 阻断了 X 与 Y 之间的每条含有指向 X 的路径，

则称 Z 满足关于 (X, Y) 的后门准则。此时 X 对 Y 的因果效应可以由以下**后门校正公式**计算：

$$P(Y = y \mid do(X = x)) = \sum_z P(Y = y \mid X = x, Z = z)P(Z = z)$$

(中文版第 87 页公式多了一个右括号)

【证明】 Judea Pearl. *Comment: Graphical Models, Causality and Intervention*. Statistical Science, 8(3):266–269, 1993.

为了阻断那些使 X 和 Y 相关但不传递 X 产生的因果效应的路径，我们希望找到满足如下要求的变量集合用于校正：

1. 阻断 X 和 Y 之间的所有伪路径
2. 保持所有从 X 到 Y 的有向路径不变
3. 不会产生新的伪路径

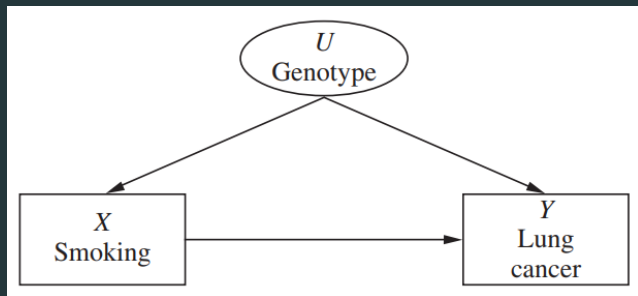
伪路径/后门路径 - X 和 Y 之间的、以指向 X 的箭头为开始的、未被阻断的路径。

在图中进行干预 X 操作之后 X 和 Y 之间连通的路径是可传递 X 产生的因果效应的路径。

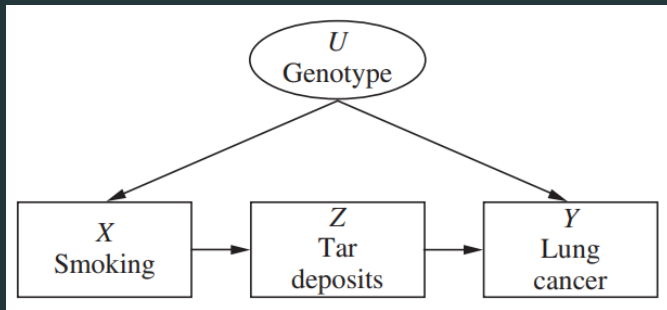
由于 X 的某些父节点无法测量，因此只能尝试切断这一路径。

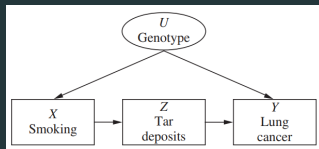
前门准则

还有许多因果效应估计问题无法使用后门校正进行解决。



再引入一个中介变量之后也许可以解决？





1. X 对 Z 的因果效应可识别，没有从 X 到 Z 的后门路径，控制空集就可以使用后门校正公式计算：

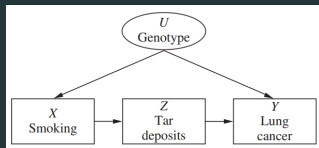
$$P(Z = z \mid do(X = x)) = P(Z = z \mid X = x)$$

2. Z 对 Y 的因果效应可识别，后门路径被 X 阻断，使用后门校正公式计算：

$$P(Y = y \mid do(Z = z)) = \sum_x P(Y = y \mid Z = z, X = x)P(X = x)$$

提示：后门校正公式为

$$P(Y = y \mid do(X = x)) = \sum_z P(Y = y \mid X = x, Z = z)P(Z = z)$$



要把两部分效应连接起来获得 X 对 Y 的整体因果效应。

如果固定 Z 为 z ，则 Y 的概率为 $P(Y = y \mid do(Z = z))$

考虑将 X 设置为 x ，那么选择固定 Z 值为 z 的概率为 $P(Z = z \mid do(X = x))$ 。

对 Z 的所有可能取值 z 求和，可以得到

$$\begin{aligned} & P(Y = y \mid do(X = x)) \\ &= \sum_z P(Y = y \mid do(Z = z)) P(Z = z \mid do(X = x)) \\ &= \sum_z \left[\sum_{x'} P(Y = y \mid Z = z, X = x') P(X = x') \right] P(Z = z \mid X = x) \end{aligned}$$

前门准则

前门准则 - 给定有向无环图中的一对有序变量 (X, Y) ，如果变量集合 Z 满足

1. Z 切断了所有 X 到 Y 的有向路径，
2. X 到 Z 没有后门路径，
3. 所有 Z 到 Y 的后门路径都被 X 阻断，

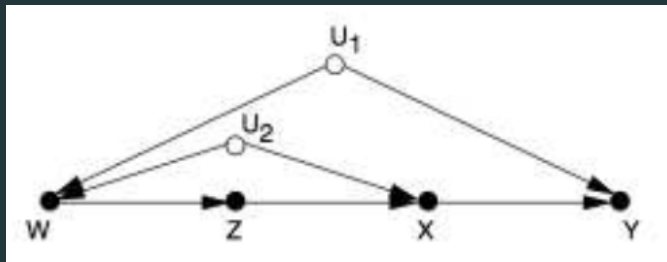
则称 Z 满足关于 (X, Y) 的前门准则。此时 X 对 Y 的因果效应可以由以下**前门校正公式**计算：

$$P(Y = y \mid do(x)) = \sum_z P(z \mid x) \sum_{x'} P(y \mid x', z) P(x')$$

Do-演算

Do-演算

也许前门准则和后门准则都无法解决？



对于因果图 G ，令 $P(\cdot)$ 为所有变量的概率分布函数，对于任意四个互不相交的变量子集 X 、 Y 、 Z 和 W ，有：

增添或删除观察

$$P(y|\hat{x}, z, w) = P(y|\hat{x}, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}$$

干预与观察交换

$$P(y|\hat{x}, \hat{z}, w) = P(y|\hat{x}, z, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{XZ}}}$$

增添或删除干预

$$P(y|\hat{x}, \hat{z}, w) = P(y|\hat{x}, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X, Z(W)}}}$$

其中 $Z(W) = Z \setminus An(W)_{G_{\overline{X}}}$

在以上规则的表述中， $G_{\overline{X}}$ 表示一个删除了 G 中所有指向 X 结点的边后得到的子图， $G_{\overline{XZ}}$ 表示了一个删除了 G 中所有指向 X 结点的边和指出 Z 结点的边后得到的子图。 $An(W)_G$ 表示在图 G 中所有 W 及其祖先结点构成的点集。

增添或删除观察

$$P(y|\hat{x}, z, w) = P(y|\hat{x}, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}$$

此规则说明，如果观察到变量 Z 与 Y 相互独立（可能以其他变量 W 为条件），那么 Y 的概率分布就不会随着变量 Z 而改变，对 Z 观察或不观察的效果都是相同的。

干预与观察交换

$$P(y|\hat{x}, \hat{z}, w) = P(y|\hat{x}, z, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{XZ}}}$$

此规则说明，如果变量集 X 阻断了从 Z 到 Y 的所有路径，那么对于 Y 来说， $do(Z)$ 与 $see(Z)$ （对 Z 进行观测的结果）的效果是相同的。

增添或删除干预

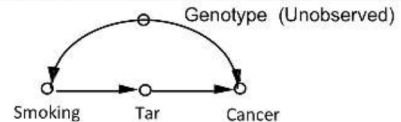
$$P(y|\hat{x}, \hat{z}, w) = P(y|\hat{x}, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}, \overline{Z(w)}}$$

其中 $Z(W) = Z \setminus An(W)_{G_{\overline{X}}}$

此规则说明，如果从 Z 到 Y 之间没有因果路径，那么对于 Y 来说，增添或删除 $do(Z)$ 的效果就是相同的。

DO-CALCULUS AT WORK

Query



Genotype (Unobserved)

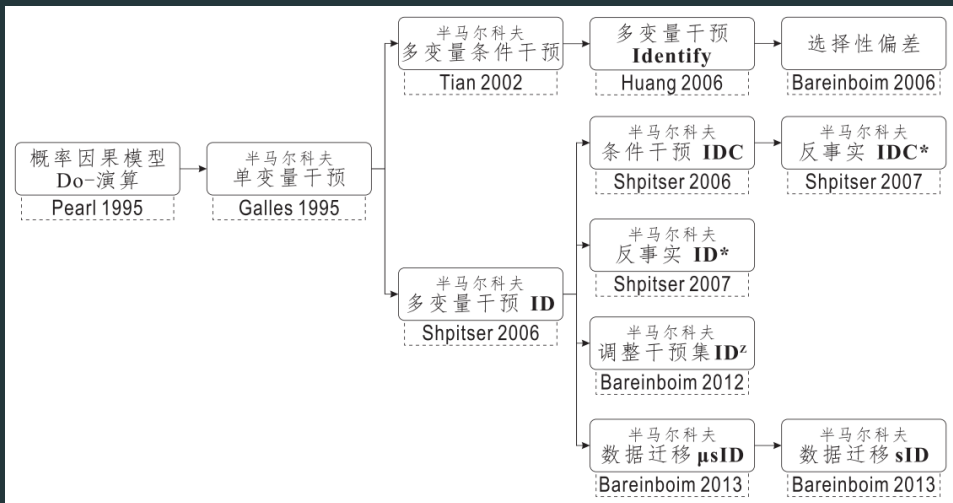
Smoking Tar Cancer

$$\begin{aligned}
 P(c \mid do(s)) &= \sum_t P(c \mid do(s), t) P(t \mid do(s)) && \text{Probability Axioms} \\
 &= \sum_t P(c \mid do(s), do(t)) P(t \mid do(s)) && \text{Rule 2} \\
 &= \sum_t P(c \mid do(s), do(t)) P(t \mid s) && \text{Rule 2} \\
 &= \sum_t P(c \mid do(t)) P(t \mid s) && \text{Rule 3} \\
 &= \sum_{s'} \sum_t P(c \mid do(t), s') P(s' \mid do(t)) P(t \mid s) && \text{Probability Axioms} \\
 &= \sum_{s'} \sum_t P(c \mid t, s') P(s' \mid do(t)) P(t \mid s) && \text{Rule 2} \\
 &= \sum_{s'} \sum_t P(c \mid t, s') P(s') P(t \mid s) && \text{Rule 3}
 \end{aligned}$$

Estimand

$$P(y \mid \hat{x}, \hat{z}, w) = P(y \mid \hat{x}, z, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{XZ}}}$$

Do-演算



1. **校正公式**: 使用被干预结点的父结点集进行校正
2. **后门准则**: 被干预结点有父节点无法被观测, 切断后门路径
3. **前门准则**: 有些后门路径无法被切断, 使用两次后门准则
4. **中介**: 某个中介变量和结果变量之间存在混杂因子? 多变量干预计算 (较难)
5. **Do-演算**: 还解决不了? 终极武器! (具有完备性)

感谢倾听！

Q&A