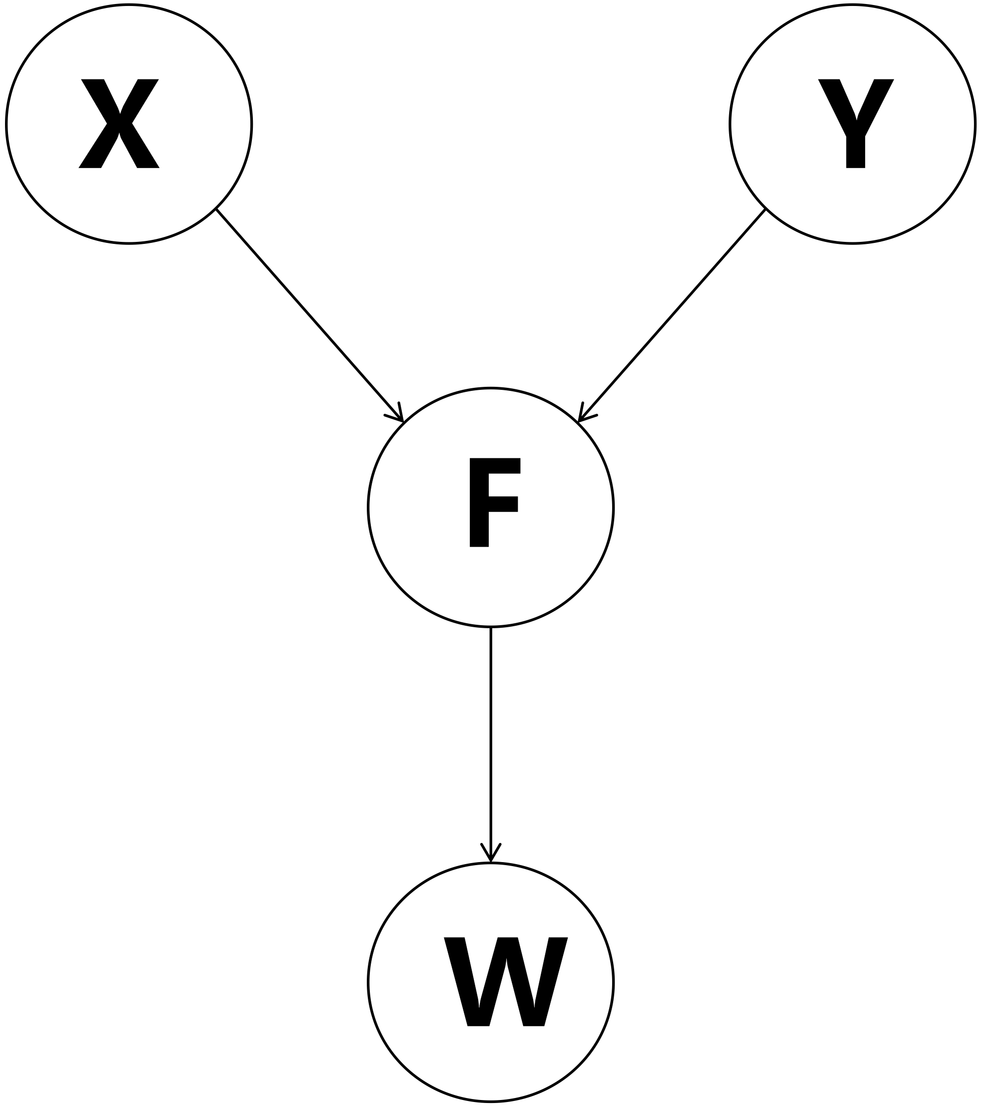
宁波大学
NINGBO UNIVERSITY

Causaul Model and covariate adjustment

汇报人：王天元

日期：2025年12月3日

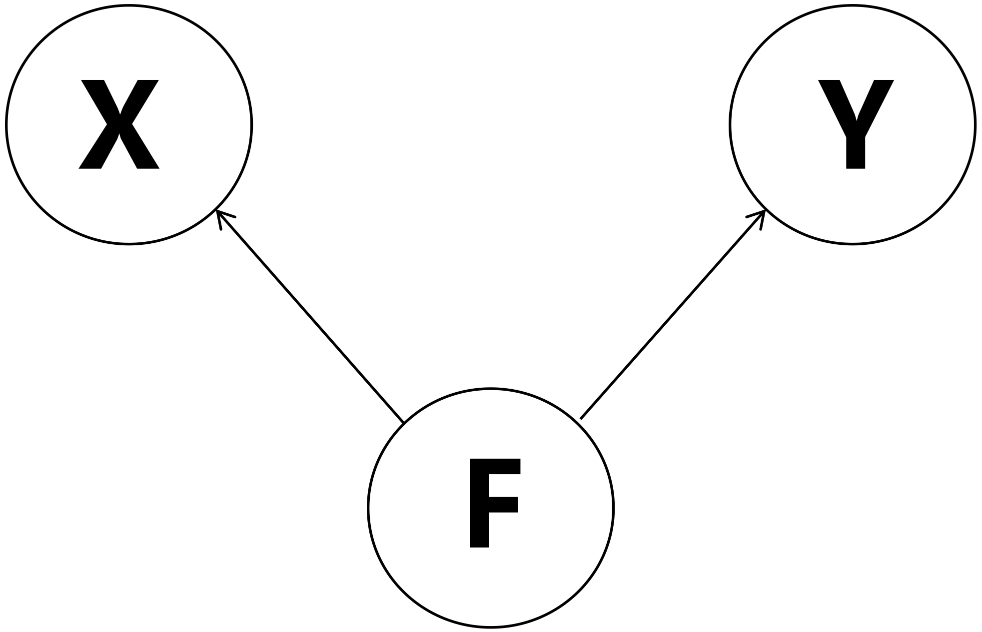
collider



X到Y这个路径被F这个Clollider阻塞，X与Y是独立的

若Condition在F上，或者是F的后代上，则X与Y是相关(dependent)的，即路径被打开（unblock)

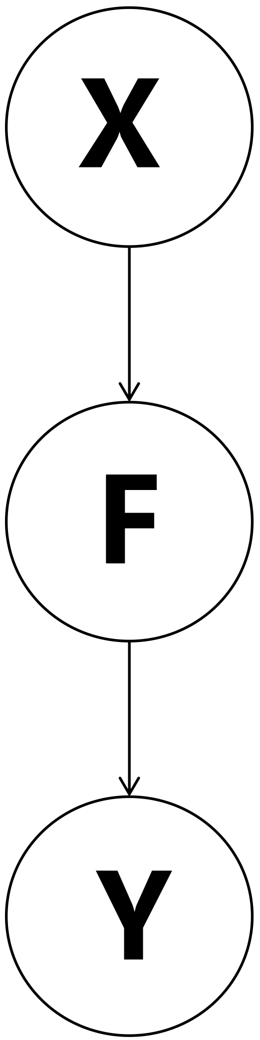
fork



X到Y这个路径是打开的，
X与Y是相关的
（dependent）

若Condition在F上则
X与Y是独立
(independent)的，即路
径被阻塞（block）

chain

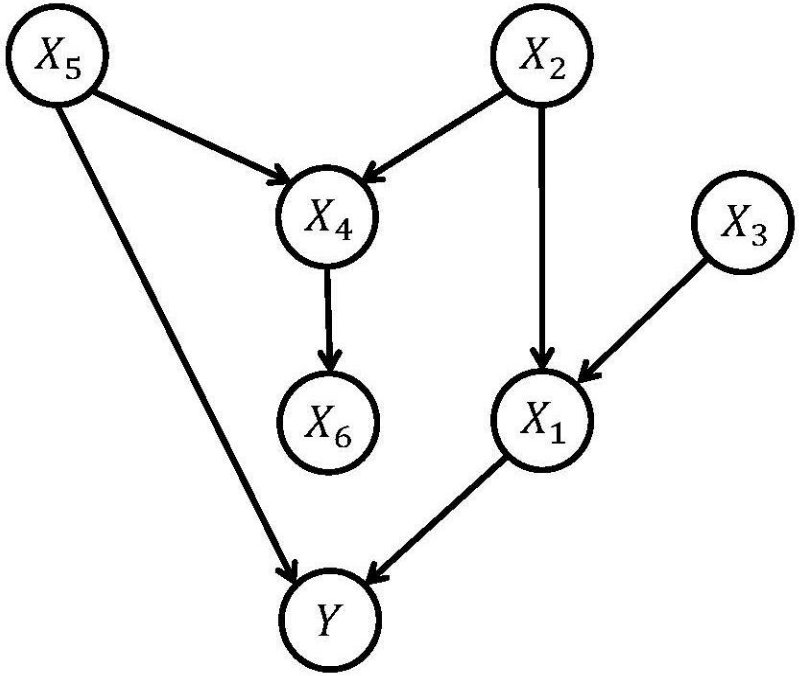


X到Y这个路径是打开的，
X与Y具有因果关系

若Condition在F上，则
X与Y是独立
(independent)的，即路
径被阻塞（block）

Clicker question – d-separation

- Consider the following graph G. Assume $p(x_V)$ factorizes according to G. Which of the following (conditional) independencies are implied by the global Markov property?



$X_2 \perp\!\!\!\perp X_5$? Yes

$X_2 \perp\!\!\!\perp X_5 | X_6$? No

$X_2 \perp\!\!\!\perp X_5 | X_3$? Yes

$X_3 \perp\!\!\!\perp X_5 | X_2$? Yes

$X_2 \perp\!\!\!\perp X_3 | X_5$? Yes

$Y \perp\!\!\!\perp X_3$? No

$Y \perp\!\!\!\perp X_3 | X_1$? Yes

2. do算子和因果DAG模型

Do-operator and causal DAG models

■ **干预的本质就是切断被干预变量与其父节点之间的因果链，并强制其取值**

- Let $G = (\boldsymbol{V}, \boldsymbol{E})$ be a DAG and P be the distribution of X_V with density p
- The pair (G, P) is a causal DAG model or a causal Bayesian network if for any $\boldsymbol{W} \subset \boldsymbol{V}$

$$p(x_V \mid do(x_W = x'_W)) = \boxed{\prod_{i \in V \setminus W} p\left(x_i \middle| x_{\mathrm{pa}(i)}\right)} 1\{x_W = x'_W\}$$

未被干预部分的按照
因果结构生成

Causal DAGs

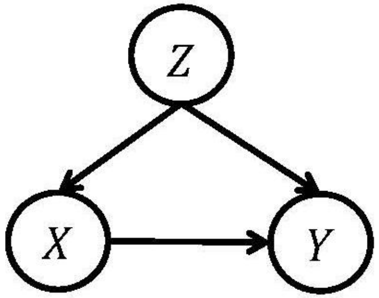
- Causal DAGs imply strong assumptions, allowing us to estimate post-intervention distributions from observational data

- How do we know the causal DAG?
  - Now: assume it is given, e.g. from background knowledge
  - Later: consider learning causal DAG (under some assumptions)
  - In any case, causal DAG provides clear framework to state causal assumptions for analysis
    - Allows for an honest debate about such assumptions
    - Can draw several possible causal DAGs, conduct the analysis for each of them and perform a sensitivity analysis

Confounding

- **不严格的来说，我们把X与Y的共同原因称为“混淆因子”**

  严格来说，就是观察到的相关性不等于实际的因果



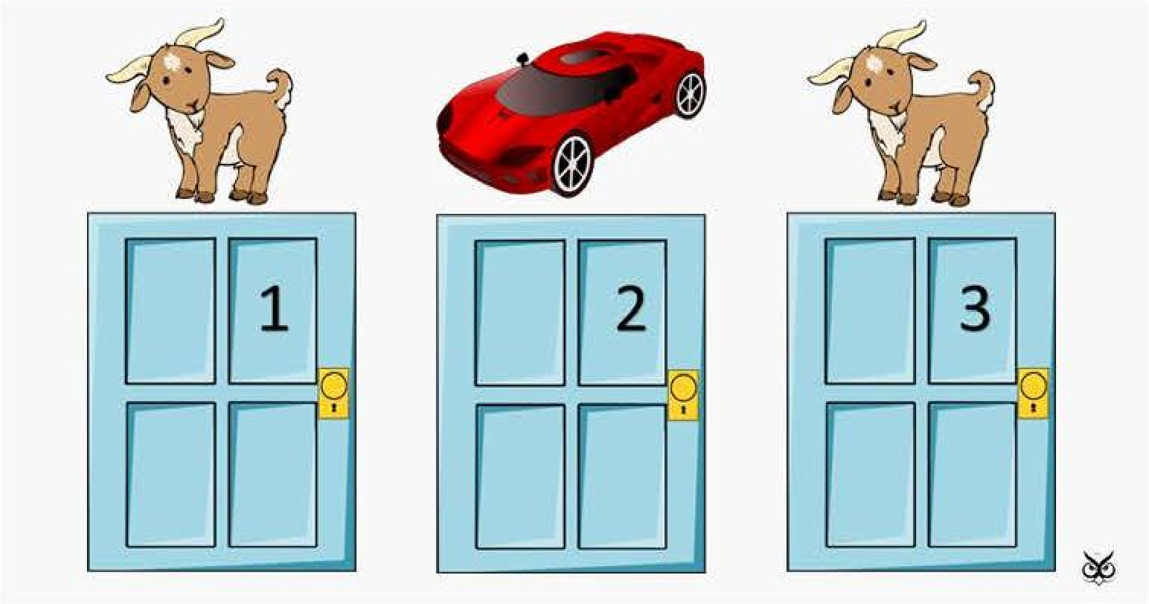
- More formally, let (G, P) with $G = (\boldsymbol{V}, \boldsymbol{E})$ be a causal Bayesian network
- $(i, k) \in \boldsymbol{V}, i \neq k$ and suppose there is a directed path from i to k
- Then the causal effect from i to k is said to be confounded if

$$p(x_k|x_i) \neq p(x_k|do(x_i))$$

- Confounding is a causal concept
- Definitions relying on associations only fall short

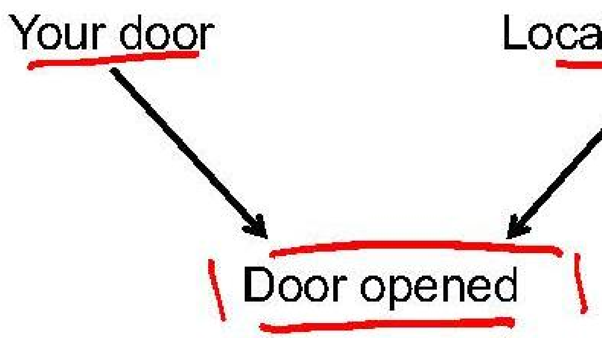
Selection bias – Example: Monty Hall

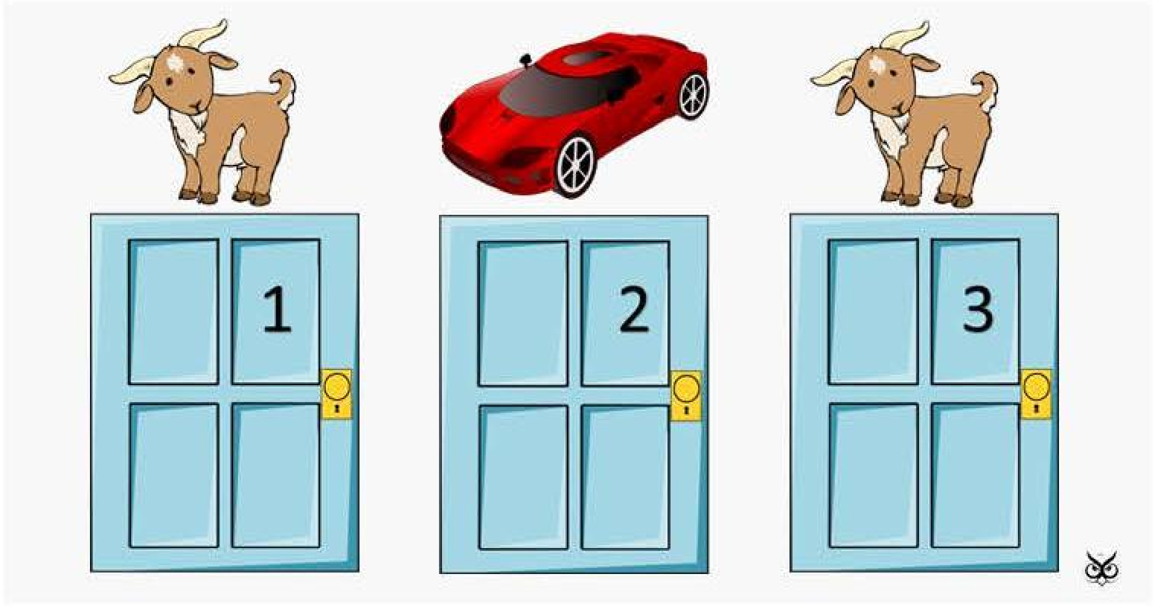
- Quiz show hosted by Monty Hall
- Setting:
  - 3 closed doors with 1 car and 2 goats
  - You pick one door
  - Monty Hall opens one of the remaining doors with a goat
  - Then he asks you whether you want to switch
- How do you decide?



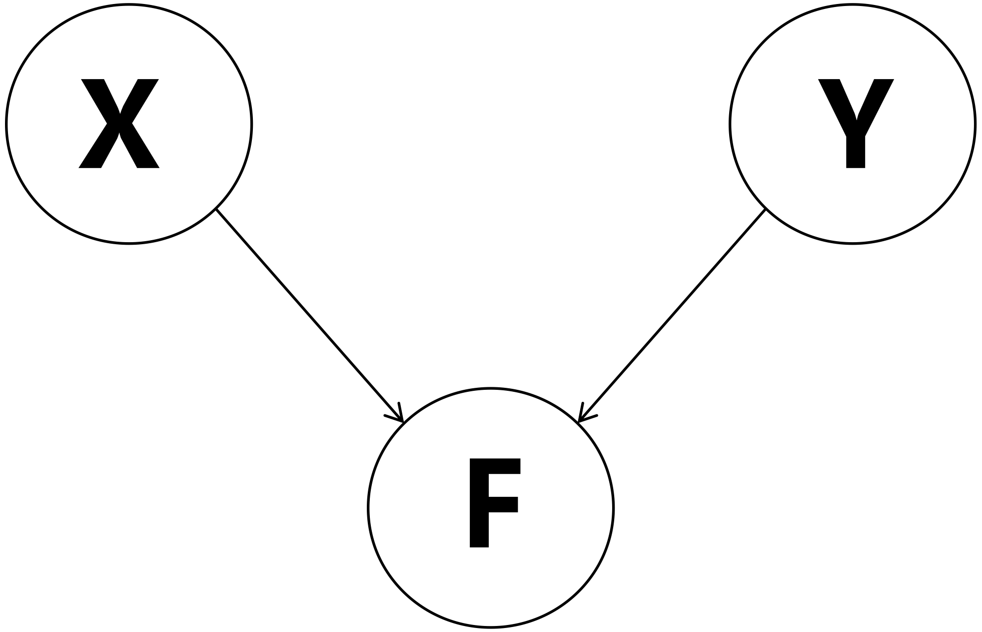
Selection bias – Example: Monty Hall

- Can show with Bayes rule:
  - If switch: success probability is 2/3
  - If stay: success probability is 1/3
- Instance of “selection bias”
  - “collider bias”, “Berkson’s paradox”





collider

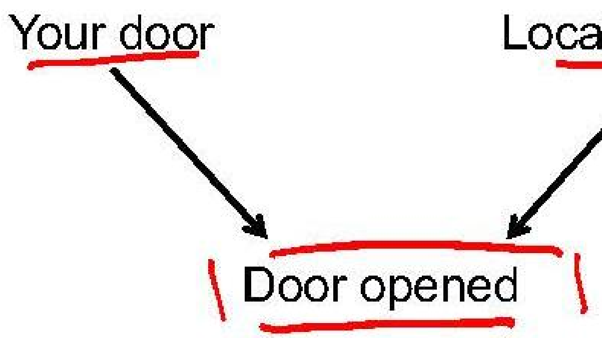


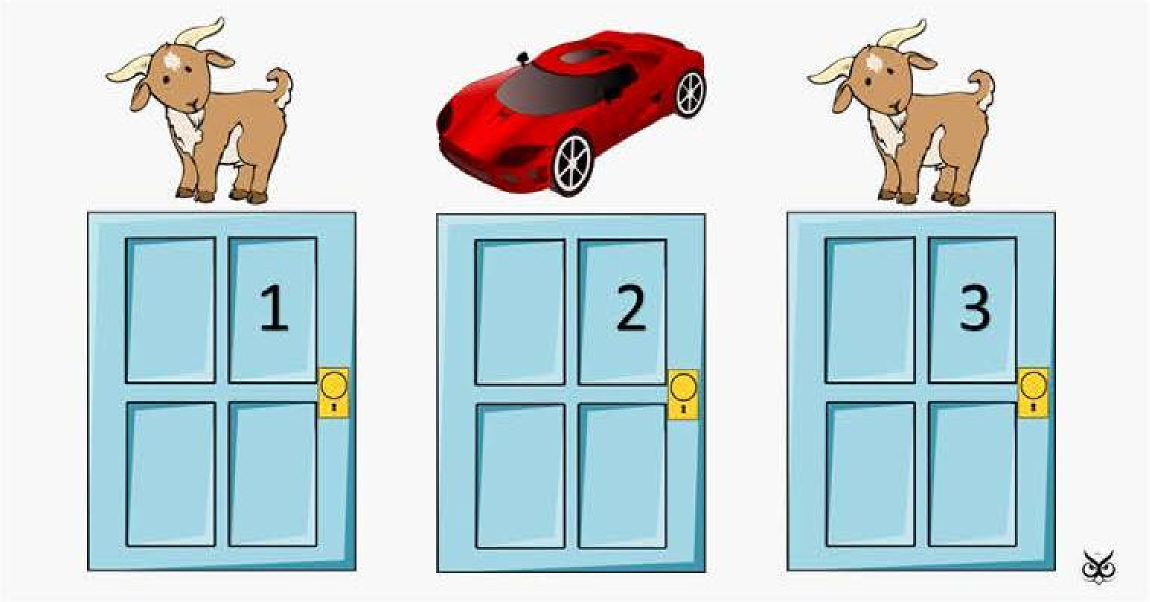
X到Y这个路径被F这个Clollider阻塞，X与Y是独立的

若Condition在F上，则X与Y是相关(dependent)的，即路径被打开（unblock)

Selection bias – Example: Monty Hall

- Can show with Bayes rule:
  - If switch: success probability is 2/3
  - If stay: success probability is 1/3
- Instance of "selection bias"
  - "collider bias", "Berkson's paradox"





Structural equation models

- Let X_V be a collection of variables, and $G = (\boldsymbol{V}, \boldsymbol{E})$ be a DAG
- Each X_i is generated as a function of its graphical parents in G and noise ϵ_i:

$$X_i \leftarrow h_i(X_{\mathrm{pa}(i)}, \epsilon_i), \qquad i \in \boldsymbol{V}$$

where $\epsilon_i, i \in \boldsymbol{V}$, are jointly independent

- The structural equations and the distribution of ϵ_V yield a distribution P for X_V

Structural equation models

- 对变量Xj的干预，通过替换其结构函数hj来建模
  其他变量的结构方程保持不变
  每个变量Xi被这个 $X_i = h_i(\text{parents of } X_i, \varepsilon_i)$ 结构方程式描述

$$X_j \leftarrow h_j(X_{\mathrm{pa}(j)}, \epsilon_j) \quad \text{by} \quad X_j \leftarrow x_j'.$$

Intervention on X_2

$$\begin{cases} X_1 \leftarrow h_1(X_{pa(1)}, \varepsilon_1) \\ X_2 \leftarrow h_2(X_{pa(2)}, \varepsilon_2) \\ \vdots \\ X_p \leftarrow h_p(X_{pa(p)}, \varepsilon_p) \end{cases} \qquad \begin{cases} X_1 \leftarrow h_1(X_{pa(1)}, \varepsilon_1) \\ X_2 \leftarrow x_2' \\ \vdots \\ X_p \leftarrow h_p(X_{pa(p)}, \varepsilon_p) \end{cases}$$

5.1 举例

假设x,y都是被SEM描述的且独立的

$$\begin{cases} x \leftarrow N(0,1) \\ y \leftarrow 6x + N_y(0,1) \end{cases}$$

DAG: $x \to y$

$\therefore var(y) = 6^2+1 = 37$

$E(y) = 6\times 0 + 0 = 0$

$\therefore y \sim N(0,37)$

(1) 若$do(x=3)$: $E(y)=18$

$var(y)=1$

$\therefore y|do(x=3) \sim \tilde{N}(18,1)$

对x做干预会改变y的分布

(3) 若$x=3$: $E(y)=18$

$var(y)=1$

$\therefore y|x=3 \sim N'(18,1)$

(3) 若干预y: $do(y=3)$

$x|do(y=3) \sim N(0,1)$

对y做干预不会改变x的分布

(4) 若令$y=3$:

$x|y=3 \sim N(\frac{18}{37}, \frac{1}{37})$

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{bmatrix}\right)$$

$$\Sigma = \begin{bmatrix} 1 & 6 \\ 6 & 37 \end{bmatrix}$$

$$X \mid Y = y \sim \mathcal{N}\left(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y-\mu_Y),\ \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right)$$

5.1 举例(4)的推导

$$\begin{cases} x \sim N_x(0,1) \\ \varepsilon \sim N_y(0,1) \end{cases} \rightarrow \text{定义 } Y = 6X + \varepsilon$$

令 $Z = \begin{bmatrix} x \\ \varepsilon \end{bmatrix} \sim \tilde{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$

考虑 $Y = 6X + \varepsilon$ 这个变换

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X \\ 6X + \varepsilon \end{bmatrix} = \hat{O} \begin{bmatrix} X \\ \varepsilon \end{bmatrix}$$

设 $\hat{O} = \begin{bmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{bmatrix}$

$$\therefore \begin{bmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{bmatrix} \begin{bmatrix} X \\ \varepsilon \end{bmatrix} = \begin{bmatrix} X \\ 6x + \varepsilon \end{bmatrix}$$

$$\begin{bmatrix} O_{11} X + O_{12} \varepsilon \\ O_{21} X + O_{22} \varepsilon \end{bmatrix} = \begin{bmatrix} X \\ 6x + \varepsilon \end{bmatrix}$$

$$\Rightarrow \begin{cases} O_{11} = 1, \ O_{12} = 0 \\ O_{21} = 6, \ O_{22} = 1 \end{cases}$$

$$\therefore \hat{O} = \begin{bmatrix} 1 & 0 \\ 6 & 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 6 & 1 \end{bmatrix} \begin{bmatrix} X \\ \varepsilon \end{bmatrix}$$

$\therefore$ 这个变换是线性的

5.1 举例(4)的推导

☆ 线性变换保持多元正态性

$$\mu_{x,y} = \hat{O}\mu_z = \begin{bmatrix} 1 & 0 \\ 6 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Sigma_{x,y} = \hat{O}\Sigma_z\hat{O}^T = \begin{bmatrix} 1 & 0 \\ 6 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 6 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 6 \\ 6 & 37 \end{bmatrix}$$

$$\therefore \begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}'\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 6 \\ 6 & 37 \end{bmatrix}\right)$$

$$\therefore x \sim \mathcal{N}_x(0, 1)$$

$$y \sim \mathcal{N}_y(0, 37)$$

$$f_X(\underline{x}) = \left[(2\pi)^{k/2}|\Sigma|^{1/2}\right]^{-1} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu})\right\}$$

要求X、Y的联合密度函数 $f_{X,Y}(x,y)$

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]$$

$$\therefore f_{X,Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$\Rightarrow X|Y=y \sim N\left(\mu_x + \frac{Cov(X,Y)}{Var(Y)}(y-\mu_y),\ Var(X) - \frac{Cov(X,Y)^2}{Var(Y)}\right)$$

$$X|Y=3 \sim N\left(\frac{18}{37}, \frac{1}{37}\right)$$

两个边缘分布都为正态分布的随机变量，其联合分布是否一定为二元正态分布？

不一定

多元正态分布的每个分量对应的随机变量满足一元正态分布

两个边缘分布都为正态分布的随机变量，联合分布一定为二元正态分布的条件：

（1）(X,Y) 是多元正态分布的线性变换——example5.1(4)

（2）所有线性组合 aX+bY 都满足正态分布

（3）每个随机变量之间是相互独立的

两个边缘分布都为正态分布的随机变量，联合分布不为二元正态分布情况：Next

$$\begin{cases} x \sim N_x(0,1) \\ \varepsilon \sim \tilde{N}(0,1) \end{cases}$$

，但是定义 $Y = X^2 + \varepsilon$

$$\begin{bmatrix} x \\ \varepsilon \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

$$\begin{bmatrix} x \\ x^2+\varepsilon \end{bmatrix} = \hat{O}\begin{bmatrix} x \\ \varepsilon \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ \varepsilon \end{bmatrix}$$

$$\Rightarrow \hat{O} = \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix}$$

$$\begin{cases} \mu_{X,Y} = \hat{O}\,\mu_{X,\varepsilon} = \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \Sigma_{X,Y} = \hat{O}\,\Sigma_{X,\varepsilon}\,\hat{O}^T = \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix}\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & x \\ x & x^2+1 \end{bmatrix} \end{cases}$$

$$\Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & x \\ x & x^2+1 \end{bmatrix}\right)$$

明显不是正态分布

Clicker question – Observational and interventional distributions

- Suppose the distribution of (X, Y) is entailed by a SEM:

$$X \leftarrow N_X$$
$$Y \leftarrow 1 + 2X + N_Y$$

with $N_X, N_Y \sim \mathcal{N}(0,1)$ and DAG $X \rightarrow Y$

- TRUE or FALSE?

● Y 的边缘分布是 N(1,5)	T
● 在干预 do(X=2) 下，Y 的干预分布等于 Y 的边缘分布	F
● 在干预 do(Y=1) 下，X 的干预分布等于 X 的边缘分布	T
● 在干预 do(X=2) 下，Y 的干预分布等于条件分布 Y|X=2	T

Interventions

1.外科式干预（surgical interventions）

人为的切断所有指向被干预变量的因果路径，然后把被干预变量固定为常数

2.不完美干预（imperfect interventions）

不是固定为一个值，而是将其设为一个随机变量N，且满足一个分布 $\tilde{N}_i \sim \tilde{F}$

3.移位干预（shift interventions）

修改被干预变量的生成机制

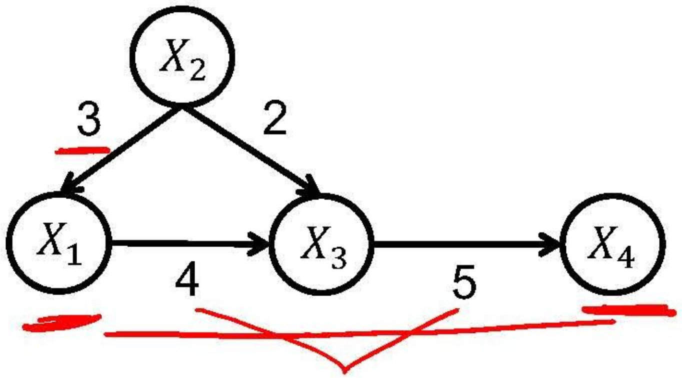
$$X_i \leftarrow h_i(X_{\mathrm{pa}(i)}, \varepsilon_i) + \tilde{N}_i$$

8. 线性结构方程式模型（linear SMEs）

NINGBO UNIVERSITY

Linear structural equation models

- Linear SEMs: all structural equations are linear and the noise is additive
- Example:
  - $X_1 \leftarrow 3X_2 + \varepsilon_1$
  - $X_2 \leftarrow \varepsilon_2$
  - $X_3 \leftarrow 4X_1 + 2X_2 + \varepsilon_3$
  - $X_4 \leftarrow 5X_3 + \varepsilon_4$



- 在线性SEM中，直接平均因果效应由边的权重给出
- What is the total average causal effect of X_1 on X_4?
  - Increasing X_1 by 1 will on average increase X_3 by 4*1=4
  - Increasing X_3 by 4 will on average increase X_4 by 4*5=20

Causal effects in linear SEMs via the path method

- Path method to compute the total causal effect of X_i on X_j in a linear SEM:
  - For each directed path from X_i to X_j, multiply the edge weights along the path
  - Sum up the results over all paths

Determining adjustment sets

观测分布$P(X_i|X_k)$反映相关性
干预分布$P(X_i|do(X_k))$是真正的因果效应

- Can we find a graphical criterion for sets $\boldsymbol{Z} \subset \boldsymbol{V}$ that satisfy

$$p(x_k|do(x_i)) = \int_{x_Z} p(x_k|x_i, x_Z)p(x_Z)dx_Z \qquad (1)$$

for all $p(\cdot)$ such that (G, P) is a causal Bayesian network?

这个式子称为调整公式

满足调整公式的集合Z称为有效调整集合

9.1 如何确定有效调整集

我们如何确定有效的调整集合?
是否有必要进行调整?
要如何调整?

- 我们是否应该尽可能多地调整变量?
- 对“处理前变量（pre-treatment variables）”进行调整是否总是安全的?

No direct causes

- Let (G, P) be a causal Bayesian network
- Assume there are no edges into X_i, i.e., $\mathrm{pa}(i) = \emptyset$
- Then it follows from the truncated factorization formula that:

$$p(x_{V\setminus\{i\}}|do(x_i')) = \prod_{j\in V\setminus\{i\}} p(x_j|x_{\mathrm{pa}(j)})\Big|_{x_i'} = \frac{p(x_V)}{p(x_i)}\Big|_{x_i'} = p(x_{V\setminus\{i\}}|x_i')$$

  - See `Notes week 4 - II.pdf`

- In this special case do-operator is the same as regular conditioning
- For any $k \neq i$, integrating out the other variables yields

$$p(x_k|do(x_i')) = p(x_k|x_i')$$

$$p(x_V \mid do(x_i')) = \left[\prod_{j\in V\setminus\{i\}} p(x_j \mid x_{\mathrm{pa}(j)})\right] \cdot \delta(x_i - x_i')$$

$pa(i) = \phi$

整体的联合分布 $p(x_v) = \prod_{j\in v} p(x_j \mid x_{pa(j)}) = \left\{\prod_{j\in V\setminus\{i\}} p(x_j \mid x_{pa(j)})\right\} \cdot \underbrace{p(x_i \mid x_{pa(i)})}_{p(x_i) \text{ bc } pa(i)=\phi}$

干预后得到联合分布 $p(x_{v\setminus\{i\}} \mid do(x_i')) = \prod_{j\in V\setminus\{i\}} p(x_j \mid x_{pa(j)}) \Big|_{x_i'}$

$= \frac{p(x_v)}{p(x_i)} \Big|_{x_i'}$

$= p(x_{v\setminus\{i\}} \mid x_i) \Big|_{x_i'}$

$= p(x_{v\setminus\{i\}} \mid x_i')$

$$\frac{P(x_v)}{P(x_i)} = \frac{P(x_{v\setminus\{i\}}, x_i)}{P(x_i)} = P(x_{v\setminus\{i\}} \mid x_i)$$

$\Rightarrow$ if $pa(i) = \phi$, then "do" is the same as conditioning

No direct causes

- Note: The situation with $\mathrm{pa}(i) = \emptyset$ arises in a randomized controlled trial where treatment $T = X_i$ is randomized. This shows why causal inference is straightforward in this setting.

Observational

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ T \arrow[rr] & & y \end{tikzcd}$$

Randomized

$$\begin{tikzcd} & Z \arrow[dr] & \\ T \arrow[rr] & & y \end{tikzcd}$$

↳ in graph do(T) is the same as conditioning on T

In a randomized trial, treatment assignment T is determined by a coin toss – effectively erasing the arrow from Z to T.

Reweighting

- Let (G, P) be a causal Bayesian network
- Then it follows from the truncated factorization formula that:

$$p(x_{V\setminus\{i\}} | do(x_i')) = \prod_{j \in V\setminus\{i\}} p(x_j | x_{\mathrm{pa}(j)}) \Big|_{x_i'} = \frac{p(x_V)}{p(x_i | x_{\mathrm{pa}(i)})} \Big|_{x_i'}$$

  - See `Notes week 4 - II.pdf`

- Thus, the interventional distribution is a re-weighted version of the observational distribution, using weights $1/p(x_i | x_{\mathrm{pa}(i)})$
- This is used in inverse probability weighting (IPW) in marginal structural models (Robins, Hernan, Brumback)

11. 干预后的权重调整（推导）

$$P(x_V) = \left\{ \prod_{j \in V \setminus \{i\}} P(x_j \mid x_{pa(j)}) \right\} \cdot P(x_i \mid x_{pa(i)})$$

$$P(x_{V \setminus \{i\}} \mid do(x_i')) = \prod_{j \in V \setminus \{i\}} P(x_j \mid x_{pa(j)}) \Big|_{x_i'}$$

$$= \frac{P(x_V)}{P(x_i \mid x_{pa(i)})} \Bigg|_{x_i'}$$

$$= \boxed{\frac{P(x_V)\, P(x_{pa(i)})}{P(x_i,\, x_{pa(i)})}} \Bigg|_{x_i'}$$

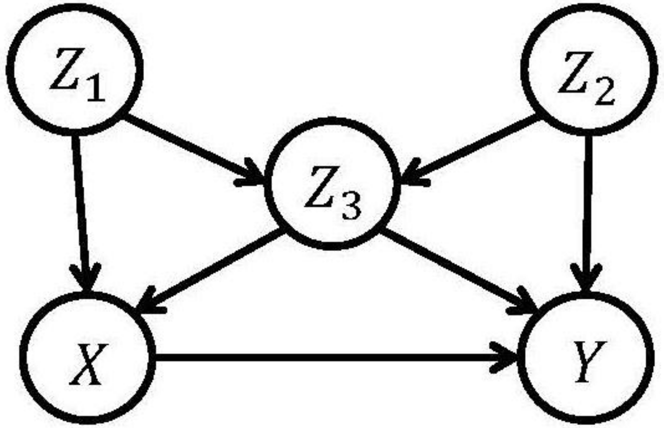
$$= P(x_{V \setminus \{i, pa(i)\}} \mid x_{pa(i)}, x_i)\, P(x_{pa(i)})$$

后门准则

✯必须满足下列两个条件：

1.有效调整集不能包括X的后代

2.有效调整集合阻塞了所有X的后门路径



后门路径define

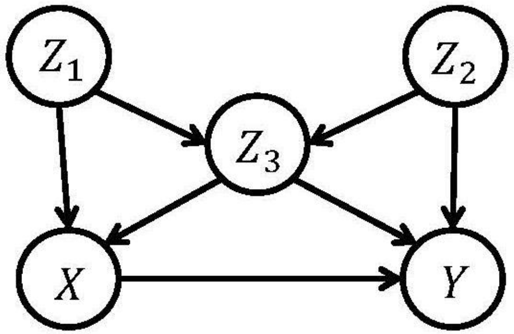
✯**是一条从 X 到 Y 的无向路径(不用考虑方向)**

✯**第一个箭头指向X，最后一个箭头指向Y**

✯**后门路径必须经过X的祖先**

Example

- Interested in the causal effect of X on Y



For $p(y \mid do(x))$, could use

* truncated factorization

* parent adjustment – $Z = \{Z_1, Z_3\}$

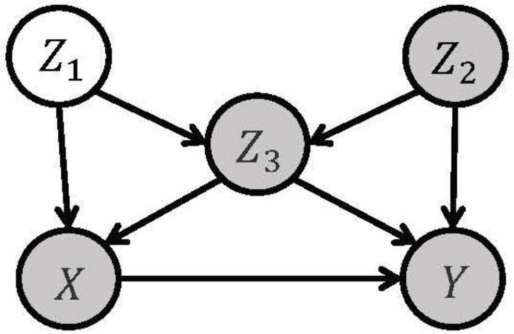
1.截断因子展开

2.父变量调整

Example

- Interested in the causal effect of X on Y
- Can we compute $p(y|do(x))$ if (only) Z_1 is not measured?
  - I.e., is $p(y|do(x))$ identifiable if (only) Z_1 is not measured?

Shaded nodes are observed



通过后门准则进行计算时，我们要对X的父变量进行调整，但是Z_3的其中一个父结点Z_1不可观测，所以我们无法得到Z_3的完整分布

13. 可识别性

Identifiability

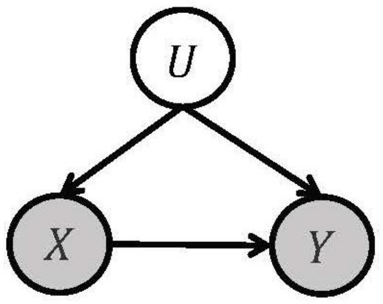
一个统计模型的某个部分是可识别的，当且仅当：

如果改变这个部分，必然会导致观测变量的联合分布发生变化

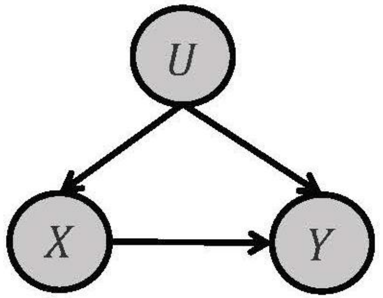
Identifiability

- X and Y observable, U unobservable
  - $p(y|x)$ is identifiable
  - $p(y|do(x))$ not identifiable: can have different $p(y|do(x))$ with same distribution of observables $p(x,y)$ (compensating changes to other parts of the model)
  - Cannot estimate $p(y|do(x))$ from observational data

Shaded nodes are observed



- X, Y and U observable
  - Can write $p(y|do(x))$ in terms of distribution of observables
  - Confounding can be removed by an identification strategy
  - $p(y|do(x))$ identifiable
  - Can estimate $p(y|do(x))$ from observational data



Thank you for listening!
Questions?

汇报人：王天元

日期：2025年12月3日