

上海交通大学  
(2022 至 2023 学年 第 1 学期 期终考试)

版权所有，不得翻印

课程名称 CS7323 从数据学习因果结构

考试时间: 2 小时 0 分钟

班级号 学号

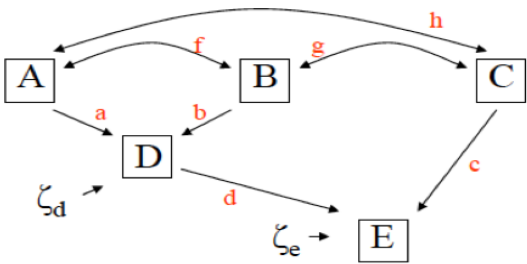
姓名 成绩

1. 回答下面基础问题:

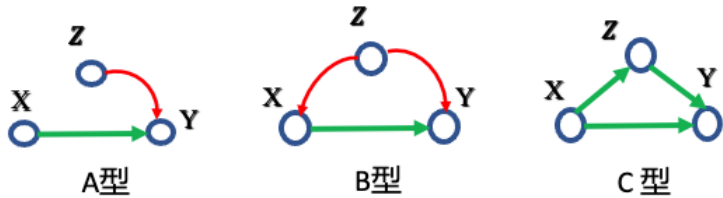
(1) (5%) 举例说明

If  $X \perp\!\!\!\perp Y$ , then  $E[XY] = E[X]E[Y]$ , and  $\text{cov}[X, Y] = E[XY] - E[X]E[Y] = 0$ .

(2) (8%) 用右图解释路径分析中正确路径应满足的三规则，并以 A 与 E 方框为例，计算对应的 direct effect、indirect effect、与 total effect。



(3) (7%) 分别就下面三种 DAG，给出  $P(Y|do\ x)$  的计算公式，说明共同点和不同点，并比较  $P(Y|X)$  和  $P(Y|do\ X)$  之异同。



2. 回答下述问题:

(1) (8%) 用你认为合适的表列解释什么是 Yule-Simpson 悖论? 什么是 Berkson 悖论?

表 A

	康复	未康复	康复率
吃药	20	20	50%
安慰剂	16	24	40%

性别

男性	康复	未康复	康复率
吃药	18	12	60%
安慰剂	7	3	70%

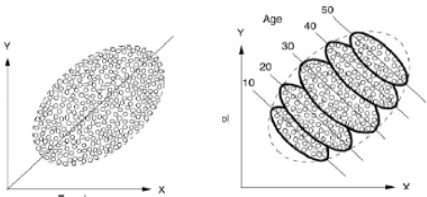
性别

女性	康复	未康复	康复率
吃药	2	8	20%
安慰剂	9	21	30%

表 B

	General Population			Hospitalized in Last Six Months		
Respiratory disease? ↓	Bone disease? ↓			Bone disease? ↓		
	Yes	No	% Yes	Yes	No	% Yes
Yes	17	207	7.6	5	15	25.0
No (control)	184	2,376	7.2	18	219	7.6

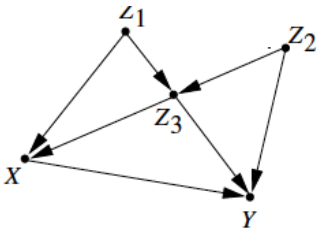
- (2) (8%) 计算  $P(\text{康复}|\text{吃药})$ 和  $P(\text{康复}|\text{do 吃药})$ 。  
(3) (4%) 可以用右图解释 Yule–Simpson 悖论吗？为什么？



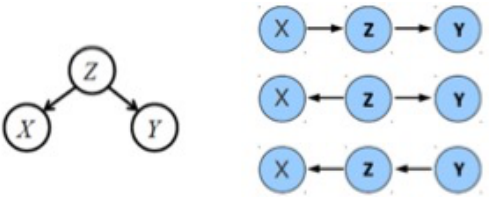
3. 考虑概率图模型的下述问题：

- (1) (8%) 对于右图，推导出计算公式：

$$P(y|do(x_0)) = \sum_{z_1, z_2, z_3} P(z_1)P(z_2)P(z_3|z_1, z_2)P(y|z_2, z_3, x_0)$$



- (2) (7%) 写出下面每个 DAG 图对应的  $p(x, y, z)$ 。



4. 回答下面的问题：

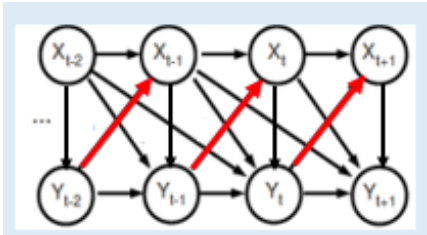
- (1) (8%) 对于给定样本集  $\{X_i, Z_i\}$ ， 根据什么理由认为有概率  $P(X_i, Z_i)=1$ ， 并由  $Y_i = X_i Y_i(1) + (1 - X_i) Y_i(0)$ ， 得到

$$\frac{1}{m} \sum_{i=1}^n x_i Y_i w_i - \frac{1}{n - m} \sum_{i=1}^n (1 - x_i) Y_i w_i$$

- (2) (4%) 说明  $w_i = 1/e(Z_i)$ ，  $e(Z_i) = p(X_i=1 | Z_i)$ ， 并基于给定样本集  $\{X_i, Z_i\}$ ， 解释其含义。  
(3) (3%) 说明  $Z \perp X | e(Z)$  成立的理由。

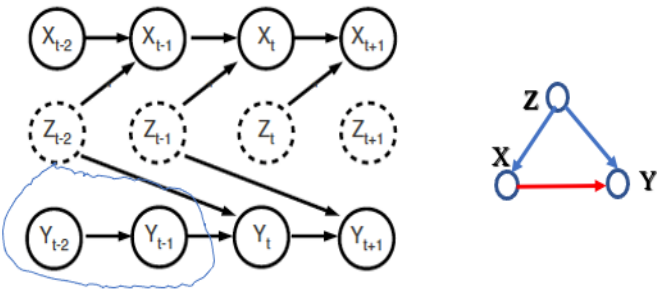
5. 回答下述问题：

- (1) (5%) 借助下图和公式, 说明 Granger 因果概念。



$$X_1(t) = \sum_{j=1}^p A_{11,j} X_1(t-j) + \sum_{j=1}^p A_{12,j} X_2(t-j) + E_1(t)$$
$$X_2(t) = \sum_{j=1}^p A_{21,j} X_1(t-j) + \sum_{j=1}^p A_{22,j} X_2(t-j) + E_2(t)$$

(2) (5%) 借助下图说明关于 Granger 因果的争议



(3) (5%) 借助下面的线性方程组，解释什么是西蒙因果序？

(4.13)  $\alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3$

(4.14)  $\alpha_{24}x_4 + \alpha_{25}x_5$

(4.15)  $\alpha_{32}x_2$

(4.16)  $\alpha_{43}x_3$

(4.17)  $\alpha_{61}x_1 + \alpha_{62}x_2 + \alpha_{63}x_3 + \alpha_{64}x_4$

(4.18)

(4.19)  $\alpha_{71}x_1$

$+ \alpha_{16}x_6$

$\alpha_{66}x_6 + \alpha_{67}x_7$

$= \alpha_{10}$

$= \alpha_{20}$

$= \alpha_{30}$

$= \alpha_{40}$

$= \alpha_{50}$

$= \alpha_{60}$

$= \alpha_{70}$

(4.15)

(4.16)

(4.19)

(4.13)

(4.17)

(4.18)

(4.14)

$x_2$

$x_3$

$x_1$

$x_6$

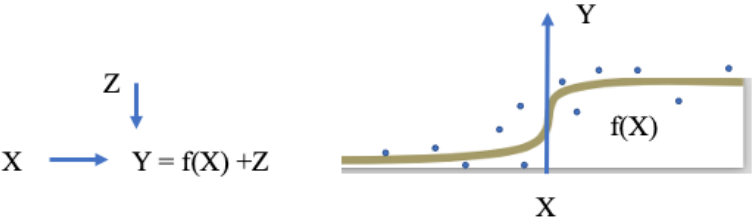
$x_4$

$x_7$

$x_5$

6. 关于叉式 fork 因果结构  $X \rightarrow Y \leftarrow Z$

- (1) (5%) 说明辨识  $Y = f(X, Z)$  所表示 X 至 Y 因果方向，等同辨识叉式结构  $X \rightarrow Y \leftarrow Z$ 。
- (2) (5%) 扼要指出辨识下图中 X 至 Y 因果方向之三个要点，



(3) (5%) 标明下图中所有的叉式节点，并说明理由可由  $Z \rightarrow X$  的因果效应和由  $X \rightarrow Y$  的因果效应估计出 b。

