

第四次作业

1. a. 任一有向无环图必有一点无任何入度, 将此点放在第一个, 去掉所有由该点发出的线, 后再找, 必有新点无任何入度, 重复此操作, 最后连上线可得图1所示。

b. 由描述可知, 可能同时有多个入度为零的点, 因此不唯一。

c. 4个结点可分成: 4个1结点; 1个2结点+2个1结点; 1个3结点+1个1结点; 2个2结点; 4个结点。

分别为 4; $C_4^2 \times (3 \times 1 \times 1)$; $4 \times (2 \times 1)$; $C_4^2 \times (3 \times 3)$; 367.

共543种。

d. $a_n = \sum_{i=1}^n (-1)^{i+1} 2^{n-i-1} C_n^i a_{n-i}$.

e. 是。

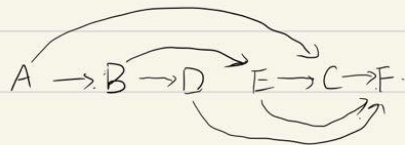
f. 上三角阵。

2. ① 有环路的话, 首先是无法判断先后的因果关系, 其次可能重复计算影响。

② 不能知前再后, 因为不能由果推因, 不能认为果对因有关联影响。

③ 相关系数只能用一个, 因为相关系数表示隐性关系, 多次与其他隐性关系相乘会失去其意义。

3.



4. 相同点: 都是正向结果与负向结果之间的差值。

不同点: 第一个只有概率, 第二个取期望, 第三个在OP基础上加权, 第四个作了60操作。

5, a, 可忽略性指对缺失数据的忽略.

交换性指实验组里个体与互换.

confounding 指由于多个前置条件, 使得对后面的因果预测造成干扰的情况.

b, 因为假设有一个一直在样本中的变量, 则它只有1% 概率被访问, 因为它代表了其他99个人, 这样使数据更均衡.

$\mathcal{P}(z)$ 为在各预测值下某条件的概率.

c, 先计算 score: $\text{Propensity Score} = P(z = \text{treatment assignment} | x \in \mathcal{X})$.

之后基于 score 对样本进行聚合或匹配或加权.

6, x, y 是 0,1 变量时, 有 $1-x = 0$ 或 1 . 又 $(1-x)^2$ 为 0 或 1. 同理 x 为 0 或 1 则 x^2 为 0 或 1.

则显然 $\sum x_t^2$ 与 $\sum x_t$ 相等, 且 $\sum (1-x_t)^2 = \sum (1-x_t)$.

因此一阶矩和二阶矩是一样的.