CS7323 从数据学习因果关系

第三次作业

1. 通径分析（path analysis）。
   a. 请计算图 1 中 A-D，A-E，A-F 和 B-F 的相关系数。
   b. 在通径分析中，图 1 的 f，g，h 为已知观测数据，现列出求解 a，b，c，d，e 的方程组并回答该方程组的求解是线性方程求解，还是二次方程求解或是其他。
   c. 更一般地，可以将 b 小题写成如图 2 所示的下三角矩阵形式，问：
   这种形式下有几个方程（不考虑边是否存在）。
   方程是否有解，有几个解？

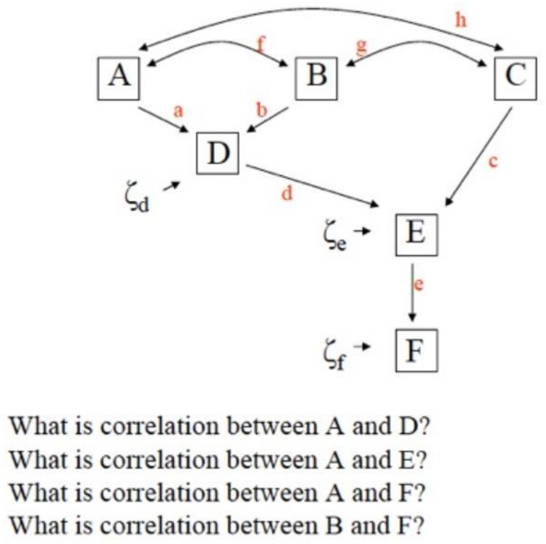


图 1。

- Model hypothesis: $\Sigma = \Sigma(\theta)$

Estimation of Path Models:

- Choose $\hat{\theta}$ so that $\Sigma(\hat{\theta})$ is close to S

sample covariance matrix of the observed variables: S

$$\overbrace{\begin{bmatrix} \mathrm{var}(Y_1) & & \\ \mathrm{cov}(Y_2,Y_1) & \mathrm{var}(Y_2) & \\ \mathrm{cov}(X_1,Y_1) & \mathrm{cov}(X_1,Y_2) & \mathrm{var}(X_1) \end{bmatrix}}^{\Sigma} = \overbrace{\begin{bmatrix} \gamma_{11}^2\mathrm{var}(X_1) + \mathrm{var}(\zeta_1) & & \\ \beta_{21}(\gamma_{11}^2\mathrm{var}(X_1) + \mathrm{var}(\zeta_1)) & \beta_{21}^2(\gamma_{11}^2\mathrm{var}(X_1) + \mathrm{var}(\zeta_1)) + \mathrm{var}(\zeta_2) & \\ \gamma_{11}\mathrm{var}(X_1) & \beta_{21}\gamma_{11}\mathrm{var}(X_1) & \mathrm{var}(X_1) \end{bmatrix}}^{\Sigma(\theta)}$$

$$\mu(\theta) = \begin{bmatrix} \alpha_y + \Lambda_y(\mathbf{I}-\mathbf{B})^{-1}(\alpha_\eta + \Gamma\mu_\xi) \\ \alpha_x + \Lambda_x\mu_\xi \end{bmatrix}$$

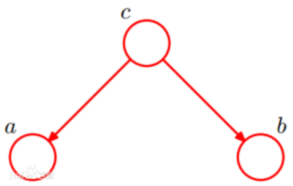
$$\Sigma_{xx}(\theta) = \Lambda_x\Phi\Lambda_x' + \Theta_\delta$$
$$\Sigma_{xy}(\theta) = \Lambda_x\Phi\Gamma'(\mathbf{I}-\mathbf{B})^{-1'}\Lambda_y'$$
$$\Sigma_{yy}(\theta) = \Lambda_y(\mathbf{I}-\mathbf{B})^{-1}(\Gamma\Phi\Gamma' + \Psi)(\mathbf{I}-\mathbf{B})^{-1'}\Lambda_u' + \Theta_\epsilon$$

图 2。

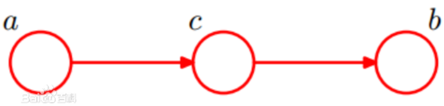
Tips;

https://www.publichealth.columbia.edu/research/population-health-methods/path-analysis

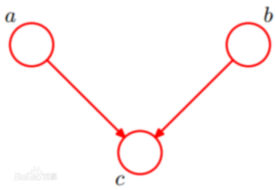
2. D 分离（D-Separation）。计算证明，图 3（a），（b），（c）中，节点 c 在被观测（作为条件）和未被观测两种的情况下，节点 a 和节点 b 是否相互独立。换言之，即证明分别$P(a,b|c)=p(a|c)p(b|c)$和$P(a,b)=p(a)p(b)$是否成立。



（a）



（b）



（c）

图 3。

3. （选做）尝试类比分析纯电阻电路（或图论中的图结构等）和图 4 所示的因果电路图。

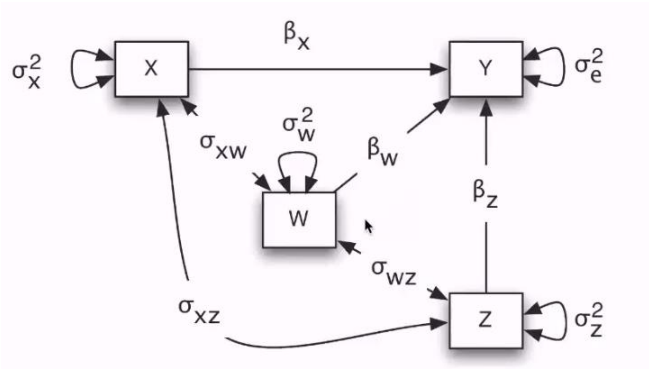


图 4 因果电路。

4. 有两组人得了某种病，一组人吃药，一组人不吃药。假设这种药会影响人体的某项指标的数值，当该药起正向作用时，该指标将正向提高。反之，指标将降低。换言之，吃药的人指标平均值减去不吃药的人指标的平均值，若为 0，则药不起作用；若小于 0 则药起反向作用；若大于 0，则药起正向作用。一个影响实验结果的重要因素是，对吃药和不吃药两组病人的划分，请讨论在划分两组病人时，哪些因素可能破坏实验结果，并说明规避的方法。

5. 说明为什么一个有向无环图对应的邻接矩阵在经过行置换（permutation）后，一定能得到一个下三角矩阵。

6. 假设随机变量 A，B，C 的联合概率分布为$P(A,B,C)$，请举例$P(A,B,C)$的一个条件概率分解，该分解对应的概率图并不唯一（画出概率图，提示：马尔可夫等价类）。

7. 写出图 5 中三个等式计算的具体矩阵，向量的表示形式，如$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$。

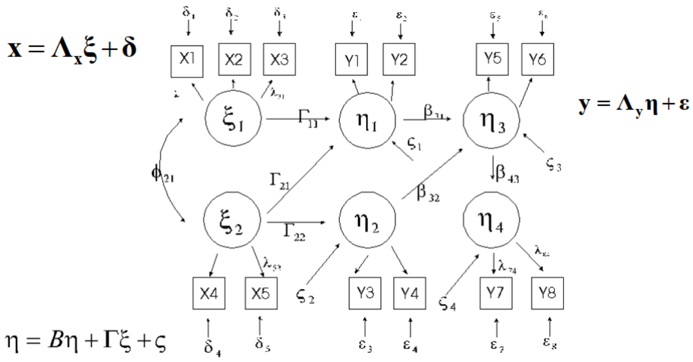


图 5。

8. 根据图 6 的概率图模型，可以得到以下分解形式：

$$P(A,B,C,D,E,F) = P(F|C,D,E)P(C|A,E)P(E|B)P(D|B)P(B|A)P(A),$$

任选以上条件概率分解中的三个条件概率为例，讨论当图中的各个节点都服从高斯分布时，Judea Pearl 的因果图模型和通径分析中的线性方程之间的关系。

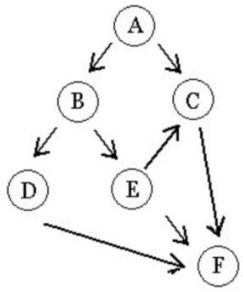


图 6。

9. 图 7a，b 给出了一个贝叶斯网及其相应的条件概率分布，请补全图 8a 的推理过程。

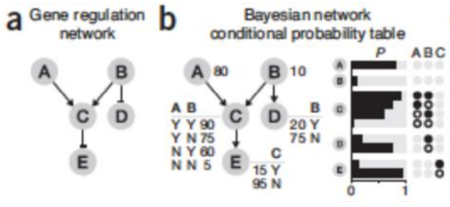


图 7。

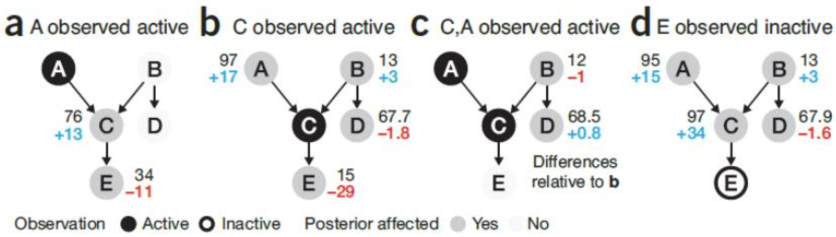


图 8。

10. 因果图如图 9 所示，证明有如下关系，$P(\cdot\,|do\ X) = P(\cdot\,|X)$。

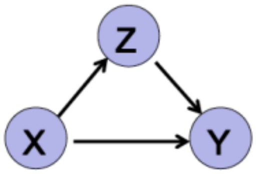
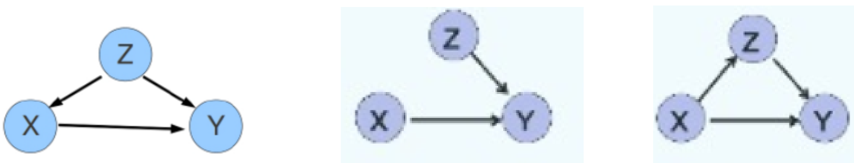


图 9。

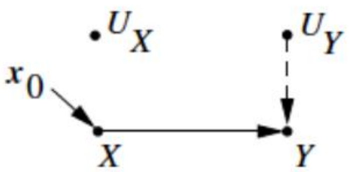
11. 证明：对图 10(a)中的随机变量X进行do操作后，将等价于图 10(b)。若不进行do操作，而是直接进行概率推断，则等价于图 10(c)。



(a) (b) (c)

图 10。

12. 将图 11 中的期望表达$E(Y|X = x_0)$写成对应的积分表达形式。



$$
\begin{aligned}
z &= f_Z(u_Z) \\
x &= x_0 \\
y &= f_Y(x, u_Y)
\end{aligned}
$$

$$E(Y|do(x_0)) = E(f_Y(x_0, u_Y))$$

$$
\begin{aligned}
E(Y|X = x_0)) &= E(f_Y(X, u_Y)|X = x_0) \\
&= E(f_Y(x_0, u_Y)|X = x_0) \\
&= E(f_Y(x_0, u_Y))
\end{aligned}
$$

图 11。