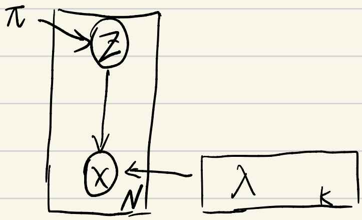
第3次作业

自硕21 崔晏菲 2021210976

1.解：

(1)



$$P(Z=k)=\pi_k \quad k=1,2,\cdots,K$$

$$P_k(x|\theta)=\frac{(\lambda_k)^x}{x!}e^{-\lambda_k}$$

(2)

$$P(Z=k|x)=\frac{P(Z=k,x)}{P(x)}=\frac{P(Z=k)P_k(x|\theta)}{\sum_{i=1}^{K}P(Z=i)P_i(x|\theta)}=\frac{\pi_k\frac{(\lambda_k)^x}{x!}e^{-\lambda_k}}{\sum_{i=1}^{K}\pi_i\frac{(\lambda_i)^x}{x!}e^{-\lambda_i}}$$

$$=\frac{\pi_k(\lambda_k)^x e^{-\lambda_k}}{\sum_{i=1}^{K}\pi_i(\lambda_i)^x e^{-\lambda_i}}$$

(3) 我选择使用泊松分布

红色的基分布为 Poisson $(\lambda_1=0.08)$

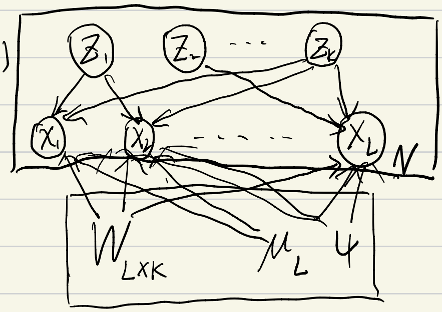
绿色的基分布为 Poisson $(\lambda_2=0.42)$

(4) 方法一：查看频率分布直方图的峰值数。一般来说，单个分布模型的峰值只有1个，因此峰值的数量可以决定K的大小。

方法二：使用K-means方法聚类，计算聚类的SSE，找到SSE减幅最小的K值（即拐点）。

2.解：

(1) (a)



(b) FA模型的独立参数的数量为 $(N\times L+L+L)=L(N+2)$ 个

多元高斯分布：参数不一定是高斯分布；认为自变量就是原因

FA：FA的均值参数 $WZ+\mu$ 服从高斯分布；认为自变量是隐含因子造成的结果

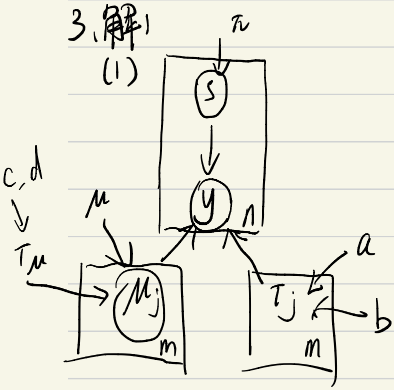
(c) 应用：例如判断骰子是真的还是假的；可以用来数据降维

(2) 不妨取 $Z_1\sim N(0,1)$　$Z_2\sim N(0,1)$

$$x_i\sim N\left((w_{i1}Z_1,\ W_{i2}Z_2),\ \begin{pmatrix}\psi_{i1} & 0\\ 0 & \psi_{i2}\end{pmatrix}\right)\qquad i=1,2,\cdots,5$$

3.解：

(1)



(2)