第六次作业

PB18000029 舒文炫

目录

6.1 题

解： 设第二次试验中第 i 个灯泡在 t 只会还能亮的事件为 $L_i = 1$,$L = (L_1, L_2, ..., L_m)$ 可以容易得到 $P(L_i = 1) = \int_t^{\infty} \theta e^{-\theta x} dx = e^{-\theta t}$ 这里缺失的数据为第二次试验 m 个灯泡的寿命，我记为 $(x_1, x_2, ..., x_m)$, 从而完全数据就是 $(y_1, ..., y_n, x_1, ...x_m)$ 在完全数据下的后验分布 $\pi(\theta|Y, X)$ 可计算为

$$
\begin{aligned}
\pi(\theta|Y, X) &\propto \pi(\theta) p(X, Y|\theta) \\
&\propto \pi(\theta) p(X|\theta) p(Y|\theta) \\
&\propto \theta^{-1} \theta^n e^{-\theta \sum_{i=1}^n x_i} \theta^m e^{-\theta \sum_{j=1}^m y_j} \\
&\propto \theta^{m+n-1} e^{-\theta(\sum_{i=1}^n x_i + \sum_{j=1}^m y_j)}
\end{aligned}
$$

这是一个 Gamma 分布 $\Gamma(m + n, \sum_{i=1}^n x_i + \sum_{j=1}^m y_j))$, 我们知道该 Gamma 分布的众数是 $\frac{m+n-1}{\sum_{i=1}^n x_i + \sum_{j=1}^m y_j}$ 但在这里 Y 的值我们是缺失的，只能通过 EM 算法来迭代逼近

E-M 算法中的 E 步为

$$\begin{aligned} Q(\theta,\hat{\theta}^{(i)}) &= E\{[(m+n-1)ln\theta-\theta(\sum_{i=1}^{n}x_i+\sum_{j=1}^{m}y_j)]|L,\hat{\theta}^{(i)}\} \\ &= (m+n-1)ln\theta-\theta\sum_{i=1}^{n}x_i-\theta E[\sum_{j=1}^{m}y_j|L,\hat{\theta}^{(i)}] \end{aligned}$$

M 步为最大化 Q，为此令

$$\frac{\partial Q(\theta,\hat{\theta}^{(i)})}{\partial\theta}=0$$

得到

$$\hat{\theta}^{(i+1)}=\frac{m+n-1}{\sum_{i=1}^{n}x_i+E[\sum_{j=1}^{m}y_j|L,\hat{\theta}^{(i)}]}$$

实际上由期望的线性性，可以对每一个 y_j 求期望，这时 y_j 只取决于条件 L_j

$$E[y_j|L_j,\hat{\theta}^{(i)}]=\begin{cases}\int_0^{\infty}yp(y|L_j=0,\hat{\theta}^{(i)})dy=\frac{1/\hat{\theta}^{(i)}-(t+1/\hat{\theta}^{(i)})e^{-\hat{\theta}^{(i)}t}}{p(L_j=0|\hat{\theta}^{(i)})}=\frac{1/\hat{\theta}^{(i)}-(t+1/\hat{\theta}^{(i)})e^{-\hat{\theta}^{(i)}t}}{1-e^{-\hat{\theta}^{(i)}t}},L_j=0\\ \int_0^{\infty}yp(y|L_j=1,\hat{\theta}^{(i)})dy=\frac{(t+1/\hat{\theta}^{(i)})e^{-\hat{\theta}^{(i)}t}}{p(L_j=1|\hat{\theta}^{(i)})}=t+1/\hat{\theta}^{(i)},L_j=1\end{cases}$$

从而可以总记为

$$E[y_j|L,\hat{\theta}^{(i)}]=(t+1/\hat{\theta}^{(i)})^{L_j}[\frac{1/\hat{\theta}^{(i)}-(t+1/\hat{\theta}^{(i)})e^{-\hat{\theta}^{(i)}t}}{1-e^{-\hat{\theta}^{(i)}t}}]^{1-L_j}$$

将该式代入上面 $\theta^{(i)}$ 的迭代公式，如果有具体数据进行数值计算，得到收敛后的值即为后验众数估计

6.2 题

证明：与 6.1 题相同的假设，设第二次试验中第 i 个灯泡在 t 只会还能亮的事件为 $L_i=1$,$L=(L_1,L_2,...,L_m)$，这里缺失的数据为第二次试验 m 个灯泡的寿命，我记为 $(x_1,x_2,...,x_m)$, 从而完全数据就是 $(y_1,...,y_n,x_1,...x_m)$ 在完全数据下的后验分布 $\pi(\theta|Y,X)$ 可计算为

$$\begin{aligned} \pi(\theta|Y,X) &\propto \pi(\theta)p(X,Y|\theta) \\ &\propto \pi(\theta)p(X|\theta)p(Y|\theta) \\ &\propto \theta^{-n}\theta^{-m}I_{x_{(n)}<\theta}I_{y_{(m)}<\theta} \\ &\propto \theta^{-m-n}I_{x_{(n)}<\theta}I_{y_{(m)}<\theta} \end{aligned}$$

这个 $x_{(n)}$ 是顺序统计量，表示 $x_1, x_2, ..., x_n$ 从小到大排列第 n 个数, 也就是其中最大的数，$y_{(m)}$ 同理

在 E 步，要求

$$\begin{aligned} Q(\theta, \hat{\theta}^{(i)}) &= E\{[-(m+n)ln\hat{\theta}^{(i)} + lnI_{x_{(n)}<\hat{\theta}^{(i)}} + lnI_{y_{(m)}<\hat{\theta}^{(i)}}]|L, \hat{\theta}^{(i)}\} \\ &= -(m+n)ln\hat{\theta}^{(i)} + lnI_{x_{(n)}<\hat{\theta}^{(i)}} + E[lnI_{y_{(m)}<\hat{\theta}^{(i)}}|L, \hat{\theta}^{(i)}] \end{aligned}$$

但这这里就可以注意到问题所在,

$$I_{y(m)<\hat{\theta}^{(i)}} = \prod_{j=1}^{m} I_{y_j<\hat{\theta}^{(i)}}$$

从而

$$lnI_{y_{(m)}<\hat{\theta}^{(i)}}|L, \hat{\theta}^{(i)} = \sum_{j=1}^{m} lnI_{y_j<\hat{\theta}^{(i)}}|L_j, \hat{\theta}^{(i)}$$

这里就要求对每一项 $I_{y_j<\hat{\theta}^{(i)}}|L_j, \hat{\theta}^{(i)}$ 都要大于 0，不然没有意义。但是如果存在 $L_j = 1$，注意到 $y_j|L_j = 1, \hat{\theta}^{(i)}$ 的分布实际上为 $U(t, \hat{\theta}^{(i)})$, 也就是说在 $L_j = 1$ 条件下，y_i 以正概率取 0，那么期望 $E[lnI_{y_{(m)}<\hat{\theta}^{(i)}}|L, \hat{\theta}^{(i)}]$ 将不存在，从而在 E 步便无法进行下去，这样使得 EM 算法失效了