# 统计中的计算方法
# 第一次作业

于慧倩
14300180118

2017 年 4 月

1. 假设 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 是 $n$ 个来自于三个二元正态分布的混合分布的独立样本，推导出用 EM 方法估计三个二元正态分布参数的迭代步骤。

   二元正态分布的概率密度函数 $f$ 为

   $$f(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}|}} \exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))$$

   令 $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)$ 代表 $\boldsymbol{x}$ 属于哪一个独立分布，即：

   $x_i|(z_i = 1) \sim N(\boldsymbol{x}_1, \boldsymbol{\sigma_1}), x_i|(z_i = 2) \sim N(\boldsymbol{x}_2, \boldsymbol{\sigma}_1), x_i|(z_i = 3) \sim N(\boldsymbol{x}_3, \boldsymbol{\sigma}_3)$

   有

   $$P(z_i = 1) = \tau_1, P(z_i = 2) = \tau_2, P(z_i = 3) = \tau_3$$

   其中

   $$\tau_1 + \tau_2 + \tau_3 = 1$$

   令 $z_{ij} = 1$ 表示第 $i$ 个样本属于第 $j$ 个分布。则有似然函数为

   $$L(\theta|\boldsymbol{x}, \boldsymbol{z}) = P(\boldsymbol{x}, \boldsymbol{z}|\theta) = \prod_{i=1}^{n}\prod_{j=1}^{3}(\tau_j f(\boldsymbol{x}_i; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_i))^{z_{ij}}$$

   那么有 log 似然函数：

   $$\log L(\theta|\boldsymbol{x}, \boldsymbol{z}) = \sum_{i=1}^{n}\sum_{j=1}^{3} z_{ij}[\log\tau_j - \frac{1}{2}\log|\boldsymbol{\Sigma}_j|] - \frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_j) - \log(2\pi)]$$

   (a) E-step:

   $$\begin{aligned} Q(\theta|\theta^{(t)}) &= E(\log L(\theta|\boldsymbol{x}, \boldsymbol{z})) \\ &= E(\sum_{i=1}^{n}\sum_{j=1}^{3} z_{ij}[\log\tau_j - \frac{1}{2}\log|\boldsymbol{\Sigma}_j|] - \frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_j) - \log(2\pi)]) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{3} T_{ij}^{(t)}[\log\tau_j - \frac{1}{2}\log|\boldsymbol{\Sigma}_j|] - \frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_j) - \log(2\pi)] \end{aligned}$$

   其中

   $$T_{ij}^{(t)} = P(z_{ij} = 1|X_i = \boldsymbol{x}_i; \theta^{(t)}) = \frac{\tau_j^{(t)} f(\boldsymbol{x}_i; \boldsymbol{\mu}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)})}{\sum_{j=1}^{3}\tau_j^{(t)} f(\boldsymbol{x}_i; \boldsymbol{\mu}_j^{(t)}, \boldsymbol{\Sigma}_j^{(t)})}$$

(b) M-step:

i. 对于 $\tau_j$ 利用 MLE 方法进行估计：

$$\begin{aligned}\tau^{(t+1)} &= \arg\max_{\tau} Q(\theta|\theta^{(t)}) \\ &= \arg\max_{\tau}\{\sum_{j=1}^{3}[\sum_{i=1}^{n} T_{i,j}^{(t)}]\log\tau_j\}\end{aligned}$$

利用 MLE 方法与 $\tau_1+\tau_2+\tau_3=1$，令 $\tau_3=1-\tau_1-\tau_2$，得到对 $\tau_1$ 求导式子并令其为零：

$$\frac{\sum_{i=1}^{n} T_{1,i}^{(t)}}{\tau_1} - \frac{\sum_{i=1}^{n} T_{3,i}^{(t)}}{1-\tau_1-\tau_2} = 0$$

另对 $\tau_2$ 求导，可以得到 $\tau_j$ 之间的关系式：

$$\tau_1 = \frac{\sum_{i=1}^{n} T_{1,i}^{(t)}}{\sum_{i=1}^{n} T_{3,i}^{(t)}}\tau_3$$

$$\tau_2 = \frac{\sum_{i=1}^{n} T_{2,i}^{(t)}}{\sum_{i=1}^{n} T_{3,i}^{(t)}}\tau_3$$

并且有 $\tau_1+\tau_2+\tau_3=1=\dfrac{\sum_{i=1}^{n}(T_{1,i}^{(t)}+T_{2,i}^{(t)}+T_{3,i}^{(t)})}{\sum_{i=1}^{n}(T_{1,i}^{(t)}+T_{2,i}^{(t)}+T_{3,i}^{(t)})}$ 所以得到：

$$\tau_j^{(t+1)} = \frac{\sum_{i=1}^{n} T_{j,i}^{(t)}}{\sum_{i=1}^{n}(T_{1,i}^{(t)}+T_{2,i}^{(t)}+T_{3,i}^{(t)})}$$

ii. 对于 $(\mu_j, \mathbf{\Sigma}_j)$ 同样利用 MLE 方法进行估计：

$$\begin{aligned}(\boldsymbol{\mu}_1^{(t+1)}, \mathbf{\Sigma}_1(t+1)) &= \underset{\boldsymbol{\mu}_1, \mathbf{\Sigma}_1}{\arg\max}\, Q(\theta|\theta^{(t)}) \\ &= \underset{\boldsymbol{\mu}_1, \mathbf{\Sigma}_1}{\arg\max} \sum_{i=1}^{n} T_{1,i}^{(t)} \{-\frac{1}{2}\log|\mathbf{\Sigma}_1| - \frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)^T \mathbf{\Sigma}_1^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)\}\end{aligned}$$

对于 $\mu_1$ 求导令其为 0，得到：

$$\sum_{i=1}^{n} T_{1,i}^{(t)} \{\mathbf{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1) = 0$$

进一步得到：

$$\boldsymbol{\mu}_1^{(t+1)} = \frac{\sum\limits_{i=1}^{n} T_{1,i}^{(t)} \boldsymbol{x}_i}{\sum\limits_{i=1}^{n} T_{1,i}^{(t)}}$$

另对 $\mathbf{\Sigma}_1$ 求导，得到：

$$\sum_{i=1}^{n} T_{1,i}^{(t)} [\mathbf{\Sigma} - (\boldsymbol{x}_i - \boldsymbol{\mu}_1)(\boldsymbol{x}_i - \boldsymbol{\mu}_1)^T] = 0$$

$$\mathbf{\Sigma}_1^{(t+1)} = \frac{\sum\limits_{i=1}^{n} T_{1,i}^{(t)} (\boldsymbol{x}_i - \boldsymbol{\mu}_1^{(t+1)})(\boldsymbol{x}_i - \boldsymbol{\mu}_1^{(t+1)})^T}{\sum\limits_{i=1}^{n} T_{1,i}^{(t)}}$$

同理由对称性得到：

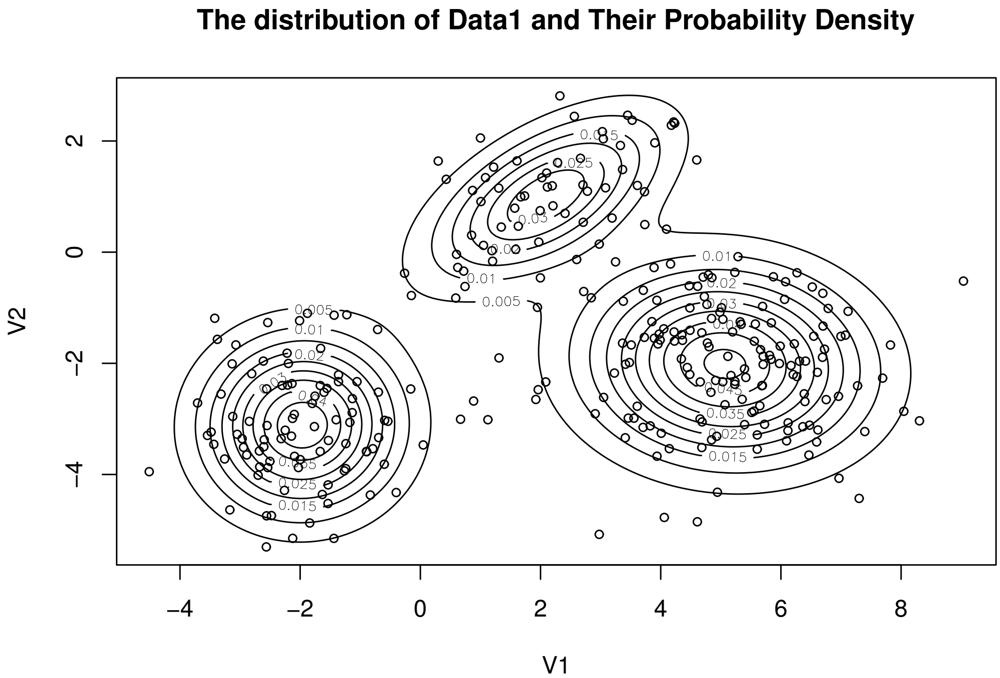
$$\boldsymbol{\mu}_2^{(t+1)} = \frac{\sum\limits_{i=1}^{n} T_{2,i}^{(t)} \boldsymbol{x}_i}{\sum\limits_{i=1}^{n} T_{2,i}^{(t)}}$$

$$\mathbf{\Sigma}_2^{(t+1)} = \frac{\sum\limits_{i=1}^{n} T_{2,i}^{(t)} (\boldsymbol{x}_i - \boldsymbol{\mu}_2^{(t+1)})(\boldsymbol{x}_i - \boldsymbol{\mu}_2^{(t+1)})^T}{\sum\limits_{i=1}^{n} T_{2,i}^{(t)}}$$

$$\boldsymbol{\mu}_3^{(t+1)} = \frac{\sum\limits_{i=1}^{n} T_{3,i}^{(t)} \boldsymbol{x}_i}{\sum\limits_{i=1}^{n} T_{3,i}^{(t)}}$$

$$\boldsymbol{\Sigma}_3^{(t+1)} = \frac{\sum\limits_{i=1}^{n} T_{3,i}^{(t)} (\boldsymbol{x}_i - \boldsymbol{\mu}_3^{(t+1)})(\boldsymbol{x}_i - \boldsymbol{\mu}_3^{(t+1)})^T}{\sum\limits_{i=1}^{n} T_{3,i}^{(t)}}$$

2. 对数据 Data1.csv 用 1. 中方法进行估计参数。




按照 1. 的方法进行估计迭代，在选取迭代初值的时候，将全部数据平均分为三份，分别计算它们的期望和协方差矩阵。迭代获得最终分布如图，三个二元正态分布的参数如下表：

```
[[1]]
[[1]]$u
          [,1]
[1,] -1.964060
[2,] -3.113237

[[1]]$sigmaf
           [,1]       [,2]
[1,] 0.99377298 0.01606908
[2,] 0.01606908 0.96803571

[[1]]$tau
[1] 0.3025644
```

```
[[2]]
[[2]]$u
         [,1]
[1,] 2.087015
[2,] 0.986269

[[2]]$sigmaf
          [,1]      [,2]
[1,] 1.4512558 0.6590272
[2,] 0.6590272 0.8731801

[[2]]$tau
[1] 0.1970871
```

```
[[3]]
[[3]]$u
          [,1]
[1,]  5.068487
[2,] -2.013263

[[3]]$sigmaf
          [,1]      [,2]
[1,]  2.049798 -0.111201
[2,] -0.111201  1.172887

[[3]]$tau
[1] 0.5003484
```

3. 一组随机抽样中随机变量 $\xi$ 取值为 0，1，2，3，4，5，6，观察到 $\xi$ 取 0，1，2，3，4，5，6 的次数为 $n_0, n_1, n_2, n_3, n_4, n_5, n_6$。假设随机变量实际服从两个总体的混合分布：总体 A：以概率 $p$，随机变量取值为 0；总体 B：以概率 $1-p$，随机变量服从均值为 $\lambda$ 的泊松分布；设计 EM 算法估计 $p$ 和 $\lambda$。

   我们有 $\xi$ 取值 $(0,1,2,3,4,5,6)$ 的概率为 $p+(1-p)e^{-\lambda}, (1-p)\lambda e^{-\lambda}, (1-p)\frac{\lambda^2 e^{-\lambda}}{2}, (1-p)\frac{\lambda^3 e^{-\lambda}}{3!}, (1-p)\frac{\lambda^4 e^{-\lambda}}{4!}, (1-p)\frac{\lambda^5 e^{-\lambda}}{5!}, (1-p)\frac{\lambda^6 e^{-\lambda}}{6!}$。我们观察到的数据为 $\boldsymbol{\xi} = (n_0, n_1, n_2, n_3, n_4, n_5, n_6)^T$。未知参数向量 $\theta = (\lambda, p)^T$。

   有似然函数：

$$L(\theta) = (p + (1-p)e^{-\lambda})^{n_0} \prod_{i=1}^{6} [(1-p^{(t)}) \frac{{\lambda^{(t)}}^i e^{-\lambda^{(t)}}}{i!}]^{n_i} \prod_{i \geqslant 7}^{\infty} 1$$

   除去常数后的 log 似然函数为

$$\log L(\theta) = n_0 \log(p + (1-p)e^{-\lambda}) + \sum_{i=1}^{6} n_i \log((1-p)\lambda^i e^{-\lambda})$$

   令 $n_0 = n_{0A} + n_{0B}$，其中 $n_{0A}, n_{0B}$ 分别表示 $\xi$ 取值为 0 且属于 A 分布的个数与 $\xi$ 取值为 0 且属于 B 分布的个数。

   则有似然函数

$$L(\theta|\theta^{(t)}) = {p^{(t)}}^{n_{0A}^{(t)}} [(1-p^{(t)})e^{-\lambda^{(t)}}]^{n_{0B}^{(t)}} \prod_{i=1}^{6} [(1-p^{(t)}) \frac{{\lambda^{(t)}}^i e^{-\lambda^{(t)}}}{i!}]^{n_i}$$

   则去除常数的 log 似然函数为

$$\log L(\theta) = n_{0A} \log p + n_{0B}[\log(1-p) - \lambda] + \sum_{i=1}^{6} n_i [\log(1-p) + i \log(\lambda) - \lambda]$$

(a) E-step:

$$\begin{aligned}Q(\theta|\theta^{(t)}) &= E(\log L(\theta;\boldsymbol{\xi}))\\ &= E\{n_{0A}\log p + n_{0B}[\log(1-p)-\lambda] + \sum_{i=1}^{6} n_i[\log(1-p)+i\log(\lambda)-\lambda]\}\\ &= n_{0A}^{(t)}\log p + n_{0B}^{(t)}[\log(1-p)-\lambda] + \sum_{i=1}^{6} n_i[\log(1-p)+i\log(\lambda)-\lambda]\end{aligned}$$

有概率如下:

$$P(\xi\in A|\xi=0) = \frac{p}{p+(1-p)e^{-\lambda}}, P(\xi\in B|\xi=0) = \frac{(1-p)e^{-\lambda}}{p+(1-p)e^{-\lambda}}$$

且

$$n_{0A}^{(t)} = P(\xi\in A|\xi=0)n_0 = \frac{p^{(t)}}{p^{(t)}+(1-p^{(t)})e^{-\lambda^{(t)}}}n_0$$

$$n_{0b}^{(t)} = P(\xi\in B|\xi=0)n_0 = \frac{(1-p^{(t)})e^{-\lambda}}{p^{(t)}+(1-p^{(t)})e^{-\lambda^{(t)}}}n_0$$

(b) M-step:

i. 对 $p$ 进行估计:

$$\begin{aligned}p^{(t+1)} &= \arg\max_p Q(\theta|\theta^{(t)})\\ &= \arg\max_p\{n_{0A}^{(t)}\log p + n_{0B}^{(t)}\log(1-p) + \sum_{i=1}^{6} n_i\log(1-p)\}\end{aligned}$$

等号右端对 $p$ 求导，令其为零得到:

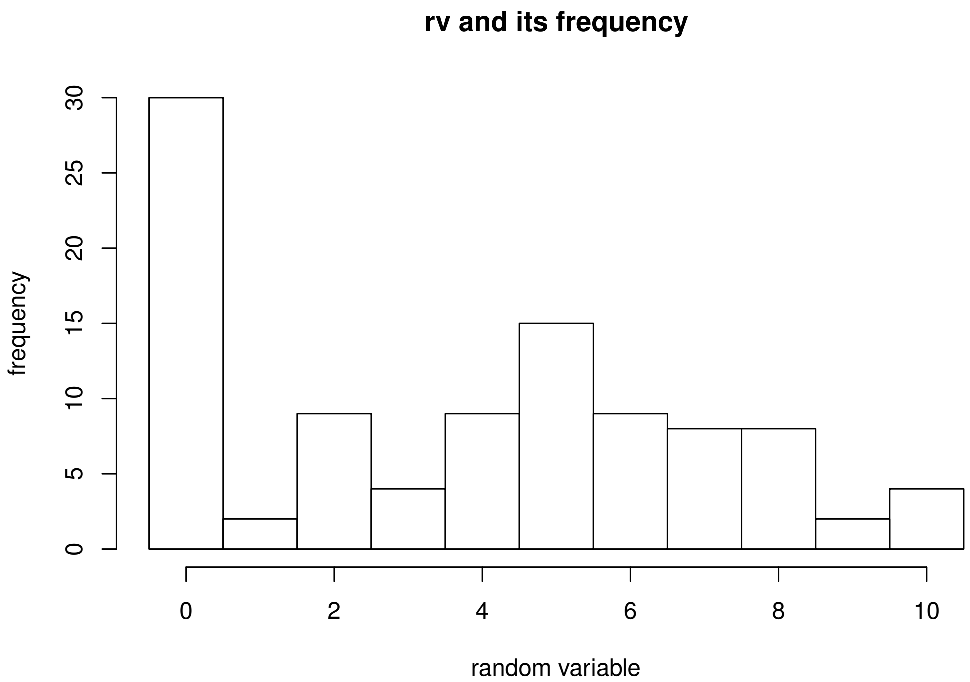
$$p^{(t+1)} = \frac{n_{0A}^{(t)}}{n}$$

ii. 对 $\lambda$ 进行估计:

$$\begin{aligned}\lambda^{(t+1)} &= \arg\max_\lambda Q(\theta|\theta^{(t)})\\ &= n_{0B}^{(t)}(-\lambda) + \sum_{i=1}^{6} n_i[i\log\lambda-\lambda]\end{aligned}$$

$$\lambda^{(t+1)} = \frac{\sum_{i=1}^{6} i*n_i}{n-n_{0A}^{(t)}}$$

4. 对数据 Data2.csv 用 3. 中方法进行估计参数。




将所有数据的均值作为 $\lambda$ 初始值，将 $\xi = 0$ 的频率作为 $p$ 初始值，迭代得到最终结果：

$$p = 0.2965968$$

$$\lambda = 5.331224$$