hw05

曾舸舵 2020302121048

2022-05-16

1

(1)

E 步

$$E(Z_{ij}|x,\theta^{(t)}) = z_{ij}^{(k)}$$

$$Q = \sum\sum z_{ij}^{(t)} \log p_i - \frac{1}{2}\sum\sum z_{ij}^{(t)}[2\log\sigma + (x_j-\mu_i)^2/\sigma^2] + const$$

M 步

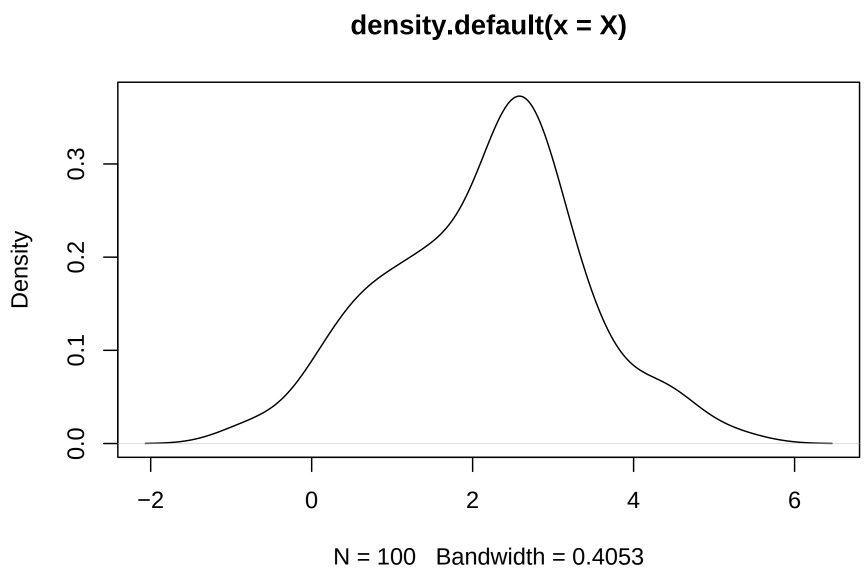
$$p_i^{(t+1)} = \frac{1}{n}\sum_j z_{ij}^{(t)} \mu_i^{(t+1)} = \frac{\sum_j z_{ij}x_j}{\sum_j z_{ij}}(\sigma^2)^{(t+1)} = \sum\sum z_{ij}^{(t)}(x_j-\mu_i^{(t+1)})^2/n$$

```
set.seed(1919)
n <- 100
p1 <- .2
p2 <- .3
p3 <- .5
p <- c(p1,p2,p3)
```

```r
mu_1 <- 1
mu_2 <- 2
mu_3 <- 3
mu <- c(mu_1, mu_2, mu_3)

sigma <- 1

X <- rnormmix(n, p, mu, sigma)
plot(density(X))
```




```r
# initial value
p10 <- .11
p20 <- .38
p30 <- .51
mu10 <- 1.1
mu20 <- 2.1
mu30 <- 3.1
```

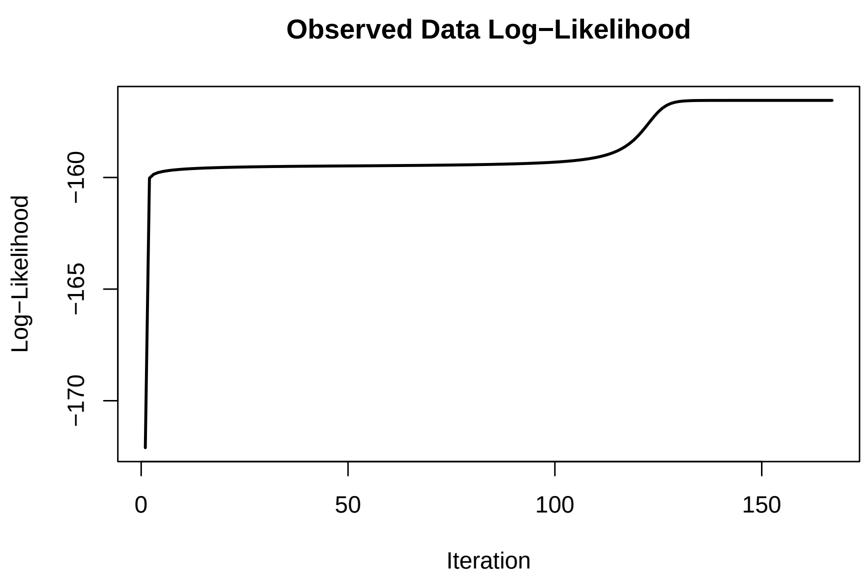
```
gm <- normalmixEM(X, lambda=c(p10,p20,p30),
                  mu=c(mu10, mu20, mu30),
                  arbvar = FALSE,eps=1e-10,
                  maxit=1e10)
```

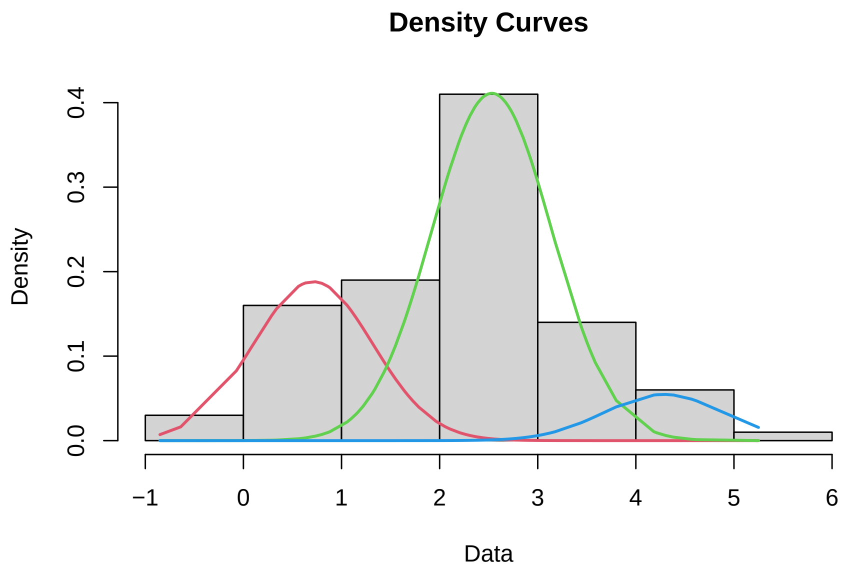
```
## number of iterations= 166
```

```
summary(gm)
```

```
## summary of normalmixEM object:
##          comp 1   comp 2    comp 3
## lambda 0.287727 0.628640 0.0836333
## mu     0.710877 2.532419 4.2841311
## sigma  0.610065 0.610065 0.6100651
## loglik at estimate:  -156.5466
```

```
plot(gm,density=TRUE)
```





运用 bootstrap 分析方差

```
se_bar <- boot.se(gm, arbvar=FALSE, B=50)
```

```
## number of iterations= 44
## number of iterations= 139
## number of iterations= 81
## number of iterations= 430
## number of iterations= 148
## WARNING! NOT CONVERGENT!
## number of iterations= 1000
## number of iterations= 70
## number of iterations= 43
## number of iterations= 747
## number of iterations= 218
## number of iterations= 96
## number of iterations= 70
## number of iterations= 77
## WARNING! NOT CONVERGENT!
```

```
## number of iterations= 1000
## One of the variances is going to zero;  trying new starting values.
## WARNING! NOT CONVERGENT!
## number of iterations= 1000
## number of iterations= 213
## number of iterations= 173
## number of iterations= 225
## number of iterations= 152
## number of iterations= 581
## number of iterations= 135
## number of iterations= 117
## number of iterations= 89
## number of iterations= 236
## number of iterations= 86
## number of iterations= 432
## number of iterations= 768
## number of iterations= 858
## number of iterations= 52
## number of iterations= 196
## number of iterations= 156
## number of iterations= 190
## number of iterations= 127
## number of iterations= 62
## number of iterations= 51
## number of iterations= 90
## One of the variances is going to zero;  trying new starting values.
## number of iterations= 466
## number of iterations= 124
## number of iterations= 163
## WARNING! NOT CONVERGENT!
## number of iterations= 1000
## number of iterations= 266
## number of iterations= 110
```

```
## number of iterations= 310
## number of iterations= 103
## number of iterations= 104
## number of iterations= 276
## number of iterations= 134
## number of iterations= 57
## number of iterations= 122
## number of iterations= 657
## number of iterations= 72
## number of iterations= 180
```

```
se_bar$lambda.se
```

```
## [1] 0.1293219 0.1811428 0.1658843
```

```
se_bar$mu.se
```

```
##           [,1]      [,2]      [,3]
## [1,] 0.3622007 0.2917186 0.6805389
```

```
se_bar$sigma.se
```

```
## [1] 0.1845597 0.1790061 0.2937424
```

2

(1)

已知数据 $x = (y_1, \cdots, y_n, r)$

设这 m 个灯泡中，失效时间分别为 $t_1, \cdots, t_m$。完全数据对数似然函数为

$$\ln L(\theta) = -(n+m)\ln\theta - \frac{1}{\theta}(\sum_{i=1}^{n} y_i + \sum_{j=1}^{m} t_j)$$

由指数分布的无后效性，$E(t_j|t_j > t) = t + \theta^{(s)}$

下面计算 $E(t_j|t_j<t)$。考虑

$$P(t_j>T|t_j<t)=\frac{P(T<t_j<t)}{P(t_j<t)}=\frac{e^{-T/\theta^{(s)}}-e^{-t/\theta^{(t)}}}{1-e^{-t/\theta^{(s)}}}$$

所以

$$E(t_j|t_j<t)=\int P(t_j>T|t_j<t)dT=\theta^{(s)}+s-\frac{t}{1-e^{-t/\theta^{(s)}}}$$

计算 $E(t_j|\theta^{(t)},x)$。

$$E(t_j|\theta^{(t)},x)=E(t_j|t_j<t)P(t_j<t)+E(t_j|t_j>t)P(t_j>t)=\theta^{(s)}+t-\frac{t}{1-e^{-t/\theta^{(s)}}})\cdot\frac{r}{m}+(t+\theta^{(s)})\cdot\frac{m-r}{m}$$

计算 $Q_t(\theta)$。

$$Q_t(\theta)=E(\ln L(\theta)|y_i,\theta^{(s)})=-(n+m)\ln\theta-\frac{1}{\theta}(\sum_{i=1}^{n}y_i+(\theta^{(s)}+t-\frac{t}{1-e^{-t/\theta^{(s)}}})\cdot r+(t+\theta^{(s)})\cdot(m-r))$$

计算 $\frac{d}{dt}Q_s(\theta)=0$。

$$\frac{d}{dt}Q_s(\theta)=\frac{-(n+m)}{\theta}+\frac{1}{\theta^2}(\sum_{i=1}^{n}y_i+(\theta^{(s)}+t-\frac{t}{1-e^{-t/\theta^{(s)}}})\cdot r+(t+\theta^{(s)})\cdot(m-r))=0$$

得到 $\theta^{(s+1)}=\frac{1}{n+m}(\sum_{i=1}^{n}y_i-\frac{rt}{1-e^{-t/\theta^{(s)}}}+(t+\theta^{(s)})\cdot m)$

(2)

```
set.seed(114514)
n_all <- 1000
X <- rexp(n_all)
t <- 0.5
m <- 50
n <- n_all - m
y_obs <- X[0:n] # 观测数据
y_obs_sum <- sum(y_obs)
t_mis <- X[n+1:n] # 未知数据
```

```r
r <- length(t_mis[t_mis<t])

theta <- mean(y_obs)
k <- 0
k_largest <- 1000
repeat{
  theta0 <- theta # 上一个 $\theta$ 值
  theta <- 1/n_all*(y_obs_sum + m*(theta0 + t) - r*t/(1 - exp(-t/theta0)))
  k <- k+1
  if(abs(theta - theta0) < 1e-10 || k >= k_largest) break
}

theta
```

```
## [1] 0.9241022
```