Lista 3

Técnicas Computacionais em Estatística

Tailine J. S. Nonato

April 22, 2025

Lista 3 - Algoritmo EM

Exercício 1

Seja T uma v.a. exponencial com FDP e FDA, respectivamente:

$$f(t;\theta) = \theta \exp(-\theta t), \quad F(t;\theta) = 1 - \exp(-\theta t), \quad t > 0, \theta > 0.$$

Em análise de sobrevivência, confiabilidade ou mesmo economia, um estudo para observar uma amostra aleatória proveniente de uma população pode terminar na prática antes de ser possível observar toda a amostra, ou seja, podemos ter observações censuradas.

Considere uma amostra referente à duração do desemprego, t, em dias, de 30 indivíduos. A duração do desemprego é o evento de interesse, ou seja, o tempo até o indivíduo deixar a situação de desemprego. No entanto, alguns indivíduos podem não experimentar o evento de interesse (não encontrarem emprego ao final do estudo ou podem ter por algum motivo saído do estudo), resultando em censura à direita.

Considere que temos acesso às observações:

$$\{x_i = (t_i, \delta_i), i = 1, \dots, n\},$$

sendo:

- $\delta_i = 1$ se a observação t_i não é censurada;
- $\delta_i = 0$ se a observação t_i é censurada.

Considere o modelo exponencial com censura e obtenha a estimativa de θ pelo algoritmo EM baseada nos dados abaixo:

```
time <- c(8,  5,  2,  4,  2,  3,  6,  1,  5,  5, 10,  8,  5,  2,  1, 12,
4,  2,  4,  2,  7,  1,  6,  3,  9,  8,  3,  2, 14,  4)

status <- c(1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0,
            1, 0, 1, 0, 1, 1, 1, 1, 1)
```

Dicas:

Note que

$$f(z|y_i;\theta_0) = \frac{f(z;\theta_0)}{1-F(y_i;\theta_0)} = \frac{\theta_0 e^{-\theta_0 z}}{e^{-\theta_0 y_i}} = \theta_0 e^{-\theta_0(z-y_i)}, \quad z > y_i;$$

Note que $E_{\theta_0}[Z_i] = y_i + \frac{1}{\theta_0}$.

A verossimilhança da amostra amplicada (completa) (y, z) é dada por

$$L^c(\theta; y, z) = \prod_{i=1}^{m} f(y_i;\theta) \prod_{i=m+1}^{n} f(z_i;\theta)$$

Note que a estimativa de MV baseada na log-verossimilhança com dados censurados é

$$\hat{\theta} = \left(\frac{n}{m}\bar{y}\right)^{-1}.$$

Resolução

Sabe-se que

$L^c(\theta; y, z) = \prod_{i=1}^{m} f(y_i;\theta) \prod_{i=m+1}^{n} f(z_i;\theta)$

A log-verossimilhança é dada por

$l(\theta; y, z) = \sum_{i=1}^{m} \log f(y_i;\theta) + \sum_{i=m+1}^{n} \log f(z_i;\theta)$

$l(\theta; y, z) = m\log\theta - \theta\sum_{i=1}^{m} y_i + (n-m)\log\theta - \theta\sum_{i=m+1}^{n} z_i$

$l(\theta; y, z) = (n)\log\theta - \theta\left(\sum_{i=1}^{m} y_i - \sum_{i=m+1}^{n} z_i\right)$

Derivando em relação a θ e igualando a zero, temos que:

$\frac{\partial l(\theta;y,z)}{\partial\theta} = \frac{n}{\theta} - \left(\sum_{i=1}^{m} y_i - \sum_{i=m+1}^{n} z_i\right) = 0$

$\theta = \frac{n}{\sum_{i=1}^{m} y_i - \sum_{i=m+1}^{n} z_i}$

```r
# computar y e z
y <- time[status==1]
z <- time[status==0]

# computar n e m
n <- length(time)
m <- length(y)
ybar <- mean(y)

# algoritmo EM
i=1
nonstop <- TRUE
theta <- rep(0,100)
theta[i] <- rnorm(1,mean=ybar,sd=sd(y))
cat("theta[",i,"]=",theta[i],"\n")
```

```
theta[ 1 ]= 7.029548
```

```r
# Algoritmo EM
while (nonstop) {
  # E-step: calcular valor esperado dos dados censurados
  z_esperado <- z + (1/theta[i])

  # M-step: atualizar a estimativa de theta
  #arg max L(theta;Y,z)
  # prod f(y_i;theta) * prod f(z_i;theta)
  theta[i+1] <- n/(sum(y) + sum(z_esperado))

  i=i+1
  cat("theta[", i, "] =", theta[i], "\n")

  # critério de parada
  nonstop=(abs(theta[i]-theta[i-1])>10^(-5))
}
```

```
theta[ 2 ] = 0.2009642
theta[ 3 ] = 0.1556145
theta[ 4 ] = 0.1457476
theta[ 5 ] = 0.143027
theta[ 6 ] = 0.1422305
theta[ 7 ] = 0.1419933
```

```
theta[ 8 ] = 0.1419223
theta[ 9 ] = 0.141901
theta[ 10 ] = 0.1418946
```

```r
cat("# de iterações:", i, "\n")
```

```
# de iterações: 10
```

```r
cat("theta^ =", theta[i], "\n")
```

```
theta^ = 0.1418946
```