HW3MuyangShi

Muyang Shi

2024-02-22

Problem 1

(a)

Define $Q(\theta|\theta^{(t)})$ to be the expectation of the joint log likelihood for the complete data $X = (Y, \Delta)$, conditioned on the observed data $\boldsymbol{Y} = \boldsymbol{y}$,

$$
\begin{aligned}
Q(\theta|\theta^{(t)}) &= \mathbb{E}\left\{\log L(\theta|X)|y_i, \theta^{(t)}\right\} \\
&= \mathbb{E}\left\{\log f_X(x|\theta)|y_i, \theta^{(t)}\right\} \\
&= \int \left[\log f_X(x|\theta)\right] f_{\delta|y_i}(\delta|y_i, \theta^{(t)}) d\delta \\
(\star) &= \sum_{i=1}^{n} \Bigg(\log\left((1-p)\mu\exp(-\mu y_i)\right) \cdot \frac{(1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)} \\
&\quad + \log\left(p\lambda\exp(-\lambda y_i)\right) \cdot \frac{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)} \Bigg)
\end{aligned}
$$

So for the E-step we just compute $(\star)$.

(b)

We maximize $Q(\theta|\theta^{(t)})$ with respect to $\theta = (p, \lambda, \mu)$; so, differentiating with respect to p, λ, μ yields:

$$
\begin{aligned}
\frac{dQ(\theta|\theta^{(t)})}{dp} &= \sum_{i=1}^{n} \Bigg(-\frac{1}{1-p} \cdot \frac{(1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)} \\
&\quad + \frac{1}{p} \cdot \frac{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)} \Bigg) \\
\frac{dQ(\theta|\theta^{(t)})}{d\lambda} &= \sum_{i=1}^{n} \left[\frac{1}{\lambda} - y_i\right] \cdot \frac{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)} \\
\frac{dQ(\theta|\theta^{(t)})}{d\mu} &= \sum_{i=1}^{n} \left[\frac{1}{\mu} - y_i\right] \cdot \frac{(1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)}
\end{aligned}
$$

Setting these derivatives equal to zero and solving for $\hat{p}, \hat{\mu}, \hat{\lambda}$ completes the M step; we set $\theta^{(t+1)}$ to be these $\hat{p}, \hat{\mu}, \hat{\lambda}$, where

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n}\frac{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)}$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{n}\frac{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)}}{\sum_{i=1}^{n} y_i\frac{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)}}$$

$$\hat{\mu} = \frac{n - \sum_{i=1}^{n}\frac{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)}}{\sum_{i=1}^{n} y_i\frac{(1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)}{p^{(t)}\lambda^{(t)}\exp(-\lambda^{(t)}y_i) + (1-p^{(t)})\mu^{(t)}\exp(-\mu^{(t)}y_i)}}$$

(Note: to perform the EM, we'd now be returning to the E-step, unless a stopping criteria, e.g. $||\theta^{(t+1)} - \theta^{(t)}||_2 < \epsilon$, has been met.

Problem 2

Note: the cpp source code to this document can be found on my Github, listed as `EM.cpp`, here.

Problem 3

Dataset is simulated as such:

```r
set.seed(600)
n <- 100
p <- 0.25
lambda <- 1
mu <- 2

data100 <- foreach(i = 1:100) %do% {
  rates <- sample(c(lambda,mu), n, replace = TRUE, prob = c(p, 1-p))
  Y <- sapply(rates, rexp, n = 1)
  Y
}
```

Problem 4

Parameters are estimated in parallel:

```r
theta100 <- foreach(i = 1:100) %dopar% {
  result <- EM(data100[[i]], c(0.25, 1, 2), eps=1e-8)
  list(p = result$theta[1,],
       lambda = result$theta[2,],
       mu = result$theta[3,])
}
stopImplicitCluster()
```

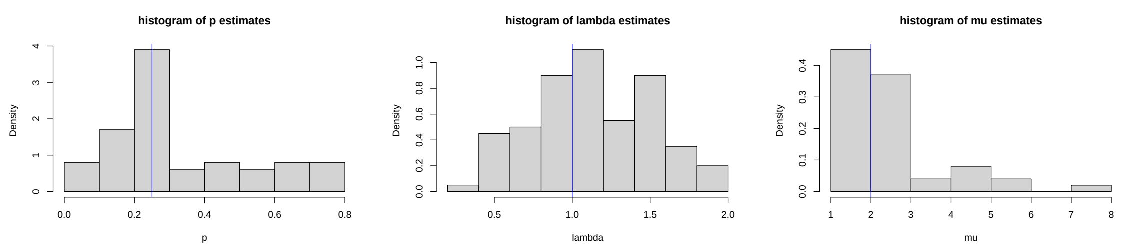
Problem 5

We estimate the standard errors of the parameter estimates using bootstrap method:

```
set.seed(600)
bootstrap <- function(Y){
  n.boot <- 100
  theta.boot <- matrix(NA, nrow = n.boot, ncol = 3)
  theta.boot[1,] <- t(EM(Y, c(0.25, 1, 2), eps=1e-8)$theta)
  for(j in 2:n.boot){
    Y.boot <- sample(Y, 100, replace = TRUE)
    theta.boot[j,] <- t(EM(Y.boot, c(0.25, 1, 2), eps=1e-8)$theta)
  }
  var.p <- var(theta.boot[,1])
  var.lambda <- var(theta.boot[,2])
  var.mu <- var(theta.boot[,3])
  return(c(var.p, var.lambda, var.mu))
}
var.boot <- foreach(i = 1:100) %dopar% {
  bootstrap(data100[[i]])
}
stopImplicitCluster()
```

Problem 6

Below are three histograms showing the estimated parameters p, λ, μ from 100 data sets each of size 100:



The performance for each of the parameters in the model is summarized in the table below:

Table 1: Performance of each of the parameters estimation

	true value	average estimates	bias	standard errors	coverage probability
p	0.25	0.328	0.078	0.150	0.78
lambda	1.00	1.128	0.128	0.321	0.91
mu	2.00	2.486	0.486	5.908	1.00