IDA Assignment 2

Johnny Lee, s1687781

25th March 2022

Q1.

Suppose $Y_1, \cdots, Y_n$ are independent and identically distributed with cumulative distribution function given by

$$F(y;\theta) = 1 - e^{-y^2/(2\theta)}, \quad y \geq 0, \quad \theta > 0.$$

Further suppose that observations are (right) censored if $Y_i > C$, for some known $C > 0$, and let

$$X_i = \begin{cases} Y_i & \text{if} \quad Y_i \leq C, \\ C & \text{if} \quad Y_i > C, \end{cases} \qquad R_i = \begin{cases} 1 & \text{if} \quad Y_i \leq C \\ 0 & \text{if} \quad Y_i > C \end{cases}$$

a)

Show that the maximum likelihood estimator based on the observed data $\{(x_i, r_i)\}_{i=1}^n$ is given by

$$\hat{\theta} = \frac{\sum_{i=1}^n X_i^2}{2\sum_{i=1}^n R_i}.$$

Answer :

We first define the Survival function (from **Workshop 3**)as

$$S(C;\theta) = \mathbb{P}(Y_i > C;\theta) = 1 - F(y_i;\theta)$$

which also represents the censored observations. For the uncensored observations, we have

$$f(y_i;\theta) = \frac{d}{dy_i}F(y_i;\theta) = \frac{ye^{-y^2/2\theta}}{\theta}$$

Given that $Y_1, \ldots, Y_n$ are independent and identically distributed, we have the likelihood function as,

$$\begin{aligned} L(\theta|\boldsymbol{y},\boldsymbol{r}) &= \prod_{i=1}^n \left([f(y_i;\theta)]^{r_i} [S(C;\theta)]^{1-r_i} \right) \\ &= \prod_{i=1}^n \left([\frac{y_i e^{-y_i^2/2\theta}}{\theta}]^{r_i} [e^{-C^2/\theta}]^{1-r_i} \right) \\ &= \left(\prod_{i=1}^n y_i^{r_i} \right) \left(\frac{1}{\theta} \right)^{\sum_{i=1}^n r_i} \exp\left(\frac{\sum_{i=1}^n (r_i y_i^2 + (1-r_i)C^2)}{2\theta} \right) \end{aligned} \tag{1}$$

Now we can rewrite the expression inside the exponential in terms of X_i. X_i can be expressed as

$$x_i = r_i y_i + C(1 - r_i)$$

Then by taking square on both sides we have,

$$x_i^2 = r_i^2 y_i^2 + (1-r_i)^2 C^2 + 2 r_i y_i C (1-r_i)$$

Noting that R_i is binary, we can then conclude with the expression as

$$x_i^2 = r_i y_i^2 + (1-r_i) C^2 \tag{2}$$

Now we substitute (2) into (1) to have,

$$\begin{aligned} L(\theta|\boldsymbol{y},\boldsymbol{r}) &= \left(\prod_{i=1}^n y_i^{r_i}\right)\left(\frac{1}{\theta}\right)^{\sum_{i=1}^n r_i} \exp\left(\frac{\sum_{i=1}^n (r_i y_i^2 + (1-r_i)C^2)}{2\theta}\right) \\ \implies \log(L(\theta|\boldsymbol{y},\boldsymbol{r})) &= \sum_{i=1}^n r_i \log(y_i) - \log\theta \sum_{i=1}^n r_i - \frac{\sum_{i=1}^n x_i^2}{2\theta} \qquad (3) \\ \implies \frac{d}{d\theta} \log L(\theta|\boldsymbol{y},\boldsymbol{r}) &= \frac{1}{\theta}\sum_{i=1}^n r_i + \frac{1}{2\theta^2}\sum_{i=1}^n x_i^2 \end{aligned}$$

By equating the derivative (3) to 0, we can obtain the maximum likelihood estimate of θ as below.

$$\hat{\theta}_{MLE} = \frac{\sum_{i=1}^n x_i^2}{2\sum_{i=1}^n r_i} \quad \text{(shown)}$$

.

b)

Show that the expected Fisher Information for the observed data likelihood is

$$I(\theta) = \frac{n}{\theta^2}(1 - e^{-C^2/(2\theta)})$$

Note: $\int_0^C y^2 f(y;\theta)dy = -C^2 e^{-C^2/(2\theta)} + 2\theta(1 - e^{-C^2/(2\theta)})$, where $f(y;\theta)$ is the density function corresponding to the cumulative distribution function $F(y;\theta)$ defined above.

Answer :

From (3), we take another derivative of it and thus obtain as below

$$\frac{d^2}{d\theta^2}\log L(\theta) = \frac{1}{\theta^2}\sum_{i=1}^{n} r_i - \frac{x_i^2}{\theta^3}$$

Then, the Fisher Information for the observed data likelihood is,

$$\begin{aligned}
I(\theta) &= -\mathbb{E}\left(\frac{\sum_{i=1}^n r_i}{\theta^2} - \frac{x_i^2}{\theta^3}\right) \\
&= -\frac{n\mathbb{E}(R)}{\theta^2} + \frac{n\mathbb{E}(X^2)}{\theta^3} \\
&= -\frac{n\mathbb{E}(R)}{\theta^2} + \frac{1}{\theta^3}\left(n\mathbb{E}(RY^2) + nC^2\mathbb{E}(1-R)\right)
\end{aligned} \tag{4}$$

Again, noting that R_i is binary,

$$\begin{aligned}
\mathbb{E}(R) &= 1 \cdot \mathbb{P}(R=1) + 0 \cdot \mathbb{P}(R=0) \\
&= \mathbb{P}(R=1) = \mathbb{P}(Y \le C) \\
&= F(C;\theta) = 1 - e^{-C^2/2\theta} \\
\implies \mathbb{E}(1-R) &= e^{-C^2/2\theta}
\end{aligned} \tag{5}$$

With the given equation, $\mathbb{E}(RY^2) = \int_0^C y^2 f(y;\theta)dy = -C^2 e^{-C^2/(2\theta)} + 2\theta(1 - e^{-C^2/(2\theta)})$, we can combine all the above equations as express the expected Fisher Information again,

$$\begin{aligned}
I(\theta) &= \frac{n\mathbb{E}(R)}{\theta^2} + \frac{1}{\theta^3}\left(n\mathbb{E}(RY^2) + nC^2\mathbb{E}(1-R)\right) \\
&= \frac{-n}{\theta^2}(1 - e^{-C^2/2\theta}) - \frac{n}{\theta^3}(C^2 e^{-C^2/2\theta}) + \frac{n}{\theta^3}(2\theta(1 - e^{-C^2/2\theta})) + \frac{n}{\theta^3}(C^2 e^{-C^2/2\theta}) \\
&= \frac{n}{\theta^2}(1 - e^{-C^2/2\theta}) \quad \text{(shown)}
\end{aligned} \tag{6}$$

c)

Appealing to the asymptotic normality of the maximum likelihood estimator, provide a 95% confidence interval for θ.

Answer :

By the Central Limit Theorem, asymptotic normality of the maximum likelihood estimator is given as,

$$\hat{\theta}_{MLE} \sim N_p(\theta, I(\theta)^{-1})$$

Thus, with 0 and $\frac{1}{I(\theta)}$ as the asymptotic mean and variance respectively, we can obtain the 95% confidence interval as below,

$$\hat{\theta}_{MLE} \pm \frac{1.96}{\sqrt{I(\theta)}} = \hat{\theta}_{MLE} \pm \frac{1.96 \cdot \theta_{MLE}}{\sqrt{n(1 - e^{-C^2/2\theta_{MLE}})}}$$

Q2.

Suppose that a dataset consists of 100 subjects and 10 variables. Each variable contains 10% of missing values. What is the largest possible subsample under a complete case analysis? What is the smallest? Justify.

Suppose that $Y_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ are iid for $i = 1, ..., n$. Further suppose that now observations are (left) censored if $Y_i < D$, for some known D and let

$$X_i = \begin{cases} Y_i & \text{if } Y_i \geq D, \\ D & \text{if } Y_i < D, \end{cases} \qquad R_i = \begin{cases} 1 & \text{if } Y_i \geq D \\ 0 & \text{if } Y_i < D \end{cases}$$

a)

Show that the log-likelihood of the observed data $\{(x_i, r_i)\}_{i=1}^n$ is given by

$$\log L(\mu, \sigma^2 | \boldsymbol{x}, \boldsymbol{r}) = \sum_{i=1}^{n} \left\{ r_i \log \phi(x_i; \mu, \sigma^2) + (1 - r_i) \log \Phi(x_i; \mu, \sigma^2) \right\}$$

where $\phi(\cdot; \mu, \sigma^2)$ and $\Phi(\cdot; \mu, \sigma^2)$ stands, respectively, for the density function and cumulative distribution function of the normal distribution with mean μ and variance σ^2.

Answer :

We first define the Survival function (from **Workshop 3**)as

$$S(D; \mu, \sigma^2) = \mathbb{P}(Y_i < D; \mu, \sigma^2) = \Phi(x_i; \mu, \sigma^2)$$

which also represents the censored observations. For the uncensored observation, we have

$$\phi(x_i; \mu, \sigma^2)$$

Given that $X_1, \ldots, X_n$ are independent and identically distributed, we have the likelihood function as,

$$\begin{aligned} L(\mu, \sigma^2 | \boldsymbol{x}, \boldsymbol{r}) &= \prod_{i=1}^{n} \left([\phi(x_i; \mu, \sigma^2)]^{r_i} [\Phi(x_i; \mu, \sigma^2)]^{1-r_i} \right) \\ \implies l(\mu, \sigma^2 | \boldsymbol{x}, \boldsymbol{r}) &= \log \prod_{i=1}^{n} \left(\phi(x_i; \mu, \sigma^2)]^{r_i} [\Phi(x_i; \mu, \sigma^2)]^{1-r_i} \right) \qquad (7) \\ &= \sum_{i=1}^{n} \left(r_i \log \phi(x_i; \mu, \sigma^2) + (1 - r_i) \log \Phi(x_i; \mu, \sigma^2) \right) \quad \text{(shown)} \end{aligned}$$

b)

Determine the maximum likelihood estimate of μ based on the data available in the file `dataex2.Rdata`. Consider σ^2 known and equal to 1.5^2. **Note**: You can use a built in function such as `optim` or the `maxLik` package in your implementation.

Answer :

```
#defining a function to simulate the log likelihood
log.lik <- function(mu, data){
  x <- data[, 1]
  r <- data[, 2]
  sum((r*dnorm(x, mu, 1.5, log = TRUE) +
         (1-r)*pnorm(x, mu, 1.5, log = TRUE)))
}

#computing the maximum likelihood estimate of mu
mle <- maxLik(logLik = log.lik, data = dataex2, start = c(mu = mean(dataex2$X)))
summary(mle)
```

```
## --------------------------------------------
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 2 iterations
## Return code 8: successive function values within relative tolerance limit (reltol)
## Log-Likelihood: -336.3821
## 1  free parameters
## Estimates:
##    Estimate Std. error t value Pr(> t)
## mu   5.5328     0.1075   51.48  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## --------------------------------------------
```

We built a function `log.lik()` that produces the log likelihood and then used `maxLik()` to simulate μ based on the data. With Newton-Raphson method, we estimated $\hat{\mu} = 5.5328$ and standard error of 0.1075

Q3.

Consider a bivariate normal sample (Y_1, Y_2) with parameters $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_{12}, \sigma_2^2)$. The variable Y_1 is fully observed, while some values of Y_2 are missing. Let R be the missingness indicator, taking the value 1 for observed values and 0 for missing values. For the following missing data mechanisms state, justifying, whether they are ignorable for likelihood-based estimation.

a)

$\text{logit}\{\mathbb{P}(R = 0|y_1, y_2, \theta, \psi)\} = \psi_0 + \psi_1 y_1, \quad \psi = (\psi_0, \psi_1)$ distinct from θ.

Answer :

Referring to the ignorability assumption (from **Lecture 6.1**), the missing in Y_2 is either **MAR** or **MCAR** and its model parameters, $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_{12}, \sigma_2^2)$ and missing mechanism parameter, ψ.

First, the missing mechanism is **MAR**. This is because the missingness is only dependent on Y_1 which is a fully observed variable. The parameters, $\{\theta, \psi\}$ are also distinct. Therefore, the ignorability assumption holds here and (a) is ignorable for likelihood-based estimation.

b)

$\text{logit}\{\mathbb{P}(R = 0|y_1, y_2, \theta, \psi)\} = \psi_0 + \psi_1 y_2, \quad \psi = (\psi_0, \psi_1)$ distinct from θ.

Answer :

The missing mechanism is **MNAR** as the mechanism is only dependent on Y_2. Therefore, the missing value is depending on itself and possibly other factors. Hence, by referring to the ignorability assumption (from **Lecture 6.1**), we conclude that (b) is not ignorable for likelihood-based estimation.

c)

$\text{logit}\{\mathbb{P}(R = 0|y_1, y_2, \theta, \psi)\} = 0.5(\mu_1 + \psi y_1)$, scalar ψ distinct from θ.

Answer :

The missing mechanism here is dependent on both μ_1 and Y_1 thus **MAR**. We can observe similarity to (a). Distinctness of the parameters means that the parameter space of $\{\theta, \psi\}$ is equal to the Cartesian product of their individual product spaces. However, the μ_1 also exists in the parameter space. This violates the ignorability assumption. Hence, (c) is not ignorable for likelihood-based estimation.

Q4.

$$Y_i \overset{\text{ind.}}{\sim} \text{Bernoulli}(p_i(\boldsymbol{\beta}))$$

$$p_i(\boldsymbol{\beta}) = \frac{exp(\beta_0 + x_i\beta_1)}{1 + exp(\beta_0 + x_i\beta_1)},$$

for $i = 1, \cdots, n$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. Although the covariate x is fully observed, the response variable Y has missing values. Assuming ignorability, derive and implement the EM algorithm to compute the MLE of $\boldsymbol{\beta}$ based on the data available in `dataex4.Rdata`. **Note**: 1) For simplicity, and without loss of generality because we have a univariate pattern of missingness, when writing down your expressions, you can assume that the first m values of Y are observed and the remaining $n - m$ are missing. 2) You can use a built in function such as `optim` or the `maxLik` package for the M-step.

Answer :

```r
head(dataex4)
```

```
##            X  Y
## 1 -0.4689827  1
## 2 -0.2557522  1
## 3  0.1457067  1
## 4  0.8164156 NA
## 5 -0.5966361  1
## 6  0.7967794 NA
```

```r
cat("Number of missing values in Y:", sum(is.na(dataex4)))
```

```
## Number of missing values in Y: 95
```

Scrutinising on the dataset, we can observe that the missing value only occurs in Y and there are 95 missing values occurring in a univariate pattern

We first derive the likelihood function to implement the EM algorithm given that $\boldsymbol{y_{obs}} = y_1, \ldots, y_m$ and $\boldsymbol{y_{mis}} = y_{m+1}, \ldots, y_n$.

$$\begin{aligned}
L(\beta_0, \beta_1 | \boldsymbol{x, y_{obs}, y_{mis}}) &= \prod_{i=1}^{n} \left([p_i(\beta_0, \beta_1)]^{y_i} [1 - p(\beta_0, \beta_1)]^{1-y_i} \right) \\
&= \prod_{i=1}^{n} \left(\frac{e^{\beta_0 + x_i\beta_1}}{1 + e^{\beta_0 + x_i\beta_1}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta_0 + x_i\beta_1}} \right)^{1-y_i} \\
\implies \log L(\beta_0, \beta_1 | \boldsymbol{x, y_{obs}, y_{mis}}) &= \sum_{i=1}^{n} \left(y_i \log \left(\frac{e^{\beta_0 + x_i\beta_1}}{1 + e^{\beta_0 + x_i\beta_1}} \right) + (1 - y_i) \log \left(\frac{1}{1 + e^{\beta_0 + x_i\beta_1}} \right) \right) \\
&= \sum_{i=1}^{n} \left(y_i \log(e^{\beta_0 + x_i\beta_1}) - \log(1 + e^{\beta_0 + x_i\beta_1}) - y_i \log(1 + e^{\beta + x_i\beta_1}) + y_i \log(1 + e^{\beta + x_i\beta_1}) \right) \\
&= \sum_{i=1}^{n} \left(y_i(\beta_0 + x_i\beta_1) - \log(1 + e^{\beta_0 + x_i\beta_1}) \right) \\
&= l(\boldsymbol{\beta} | \boldsymbol{x, y_{obs}, y_{mis}})
\end{aligned} \tag{8}$$

Now we proceed to implement the EM algorithm by calculating $Q(\boldsymbol{\beta}|\boldsymbol{\beta^{(t)}})$

$$
\begin{aligned}
Q(\boldsymbol{\beta}|\boldsymbol{\beta^{(t)}}) &= \mathbb{E}_{\boldsymbol{y_{mis}}}[l(\boldsymbol{\beta}|\boldsymbol{x}, \boldsymbol{y_{obs}}, \boldsymbol{y_{mis}})|\boldsymbol{y_{obs}}, \boldsymbol{x}, \boldsymbol{\beta^{(t)}}] \\
&= \sum_{i=1}^{m}\Big(y_i(\beta_0 + x_i\beta_i)\Big) - \sum_{i=1}^{n}\Big(\log(1+e^{\beta_0+x_i\beta_1})\Big) + \sum_{i=m+1}^{n}\Big((\beta_0+x_i\beta_1)\mathbb{E}_{\boldsymbol{y_{mis}}}[y_i|\boldsymbol{x}, \boldsymbol{y_{obs}}, \boldsymbol{\beta^{(t)}}]\Big) \\
&= \sum_{i=1}^{m}\Big(y_i(\beta_0 + x_i\beta_i)\Big) - \sum_{i=1}^{n}\Big(\log(1+e^{\beta_0+x_i\beta_1})\Big) + \sum_{i=m+1}^{n}\Big((\beta_0+x_i\beta_1)p_i(\boldsymbol{\beta}^{(t)})\Big) \\
&(\mathbb{E}(Y_i) = p_i(\boldsymbol{\beta}) \text{ as } Y_i \sim \text{Bernoulli}(p_i(\boldsymbol{\beta})))
\end{aligned} \tag{9}
$$

Now we differentiate Q with respect to β_0 and β_1 for the M-Step

$$
\begin{aligned}
\frac{d}{d\beta_0}Q(\boldsymbol{\beta}|\boldsymbol{\beta^{(t)}}) &= \sum_{i=1}^{m} y_i - \sum_{i=1}^{n}\left(\frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}}\right) \\
&+ \sum_{i=m+1}^{n}\left(\frac{e^{\beta_0^{(t)}+x_i\beta_1^{(t)}}}{1+e^{\beta_0^{(t)}+x_i\beta_1^{(t)}}} + x_i\beta_1\frac{e^{\beta_0^{(t)}+x_i\beta_1^{(t)}}}{(1+e^{\beta_0^{(t)}+x_i\beta_1^{(t)}})^2} + \beta_0\frac{e^{\beta_0^{(t)}+x_i\beta_1^{(t)}}}{(1+e^{\beta_0^{(t)}+x_i\beta_1^{(t)}})^2}\right) \\
\frac{d}{d\beta_1}Q(\boldsymbol{\beta}|\boldsymbol{\beta^{(t)}}) &= \sum_{i=1}^{m} y_i x_i - \sum_{i=1}^{n}\left(x_i\frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}}\right) \\
&+ \sum_{i=m+1}^{n}\left(\beta_0\frac{x_i e^{\beta_0^{(t)}+x_i\beta_1^{(t)}}}{(1+e^{\beta_0^{(t)}+x_i\beta_1^{(t)}})^2} + x_i\frac{e^{\beta_0^{(t)}+x_i\beta_1^{(t)}}}{1+e^{\beta_0^{(t)}+x_i\beta_1^{(t)}}} + x_i\beta_1\frac{x_i e^{\beta_0^{(t)}+x_i\beta_1^{(t)}}}{(1+e^{\beta_0^{(t)}+x_i\beta_1^{(t)}})^2}\right)
\end{aligned} \tag{10}
$$

However, the solutions of the derivatives have no closed form expressions and thus we need to resort to numerical methods. Before proceeding to the implementation, we first need to preprocess `dataex4` due to the `NA` values by rearranging it.

```r
#rearranging the order where the NA values go the last
dataex4 <- dataex4[order(dataex4$Y),]
row.names(dataex4) <- NULL
head(dataex4,5)
```

```
##            X Y
## 1  0.3215956 0
## 2  0.2582281 0
## 3  0.4352370 0
## 4 -0.2277718 0
## 5 -0.3193020 0
```

```r
tail(dataex4,5)
```

```
##             X  Y
## 496 0.6989142 NA
## 497 0.8936356 NA
## 498 0.7551561 NA
## 499 0.8782734 NA
## 500 0.6924915 NA
```

We can confirm that the order of `dataex4` is changed where the `NA` values are at the last. Now we proceed to the implementation of the EM algorithm. In the code, we have used for the stopping criterion as below

$$|\beta_0^{(t+1)} - \beta_0^{(t)}| + |\beta_1^{(t+1)} - \beta_1^{(t)}| < \varepsilon$$

```r
#function to compute the Q function
qfn <- function(param, data){
  beta0 <- param[1]; beta1 <- param[2]
  x <- data[, 1]; y <- data[, 2]
  express <- beta0 + x[1:405]*beta1
  express.na <- beta[1] + x[406:500]*beta[2]
  express.all <- beta0 + x*beta1
  express.fix <- beta0 + x[406:500]*beta1
  #expression of the Q function
  sum(y[1:405]*express) - sum(log(1 + exp(express.all))) +
    sum(express.fix*exp(express.na)/(1 + exp(express.na)))
}

#using maxlik
beta <- c(0,0)
diff <- 1
while(diff > 0.000001){
    mle <- maxLik(logLik = qfn, data = dataex4, start = beta)
    diff <- sum(abs(mle$estimate - beta))
    beta <- mle$estimate
}
cat("The value for beta0:", beta[1], "\n",
    "The value for beta1:", beta[2])
```

```
## The value for beta0: 0.9755263
##  The value for beta1: -2.480383
```

```r
#using optim
beta <- c(0,0)
diff <- 1
while(diff > 0.000001){
    mle <- optim(beta, qfn, data = dataex4,
                 control = list(fnscale = -1), hessian = TRUE)
    diff <- sum(abs(mle$par - beta))
    beta <- mle$par
}
cat("The value for beta0:", beta[1], "\n",
    "The value for beta1:", beta[2])
```

```
## The value for beta0: 0.9757079
##  The value for beta1: -2.479987
```

Using two different methods, `maxLik()` and `optim()` with $\varepsilon = 1e-6$, we can see that we both obtained similar results by a small difference (0.001). Thus, we conclude that the MLE of $\hat{\beta}_{\text{MLE}} = (\hat{\beta}_{0\text{MLE}}, \hat{\beta}_{1\text{MLE}})$ is $(0.976, -2.48)$ respectively.

Q5

Consider a random sample $Y_1, ..., Y_n$ from the mixture distribution with density

$$f(y) = pf_{\text{LogNormal}}(y; \mu, \sigma^2) + (1-p)f_{\text{Exp}}(y; \lambda),$$

with

$$\begin{aligned} f_{\text{LogNormal}}(y; \mu, \sigma^2) &= \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left\{\frac{1}{2\sigma^2}(\log y - \mu)^2\right\}, \quad y > 0, \quad \mu \in \mathbb{R},\ \sigma > 0 \\ f_{\text{Exp}}(y; \lambda) &= \lambda e^{-\lambda y}, \quad y \geq 0, \quad \lambda > 0 \end{aligned}$$

and $\boldsymbol{\theta} = (p, \mu, \sigma^2, \lambda)$

a)

Derive the EM algorithm to find the updating equations for $\boldsymbol{\theta^{(t+1)}} = (p^{(t+1)}, \mu^{(t+1)}, (\sigma^{(t+1)})^2, \lambda^{(t+1)})$.

Answer :

Let us consider a mixture model of Log-Normal and Exponential distributions.

$$\mathbb{P}(Y \leq y) = p \cdot \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left\{\frac{1}{2\sigma^2}(\log y - \mu)^2\right\} + (1-p) \cdot \lambda e^{-\lambda y}$$

Let z_i be the binary latent variables indicating component membership, i.e.

$$z_i = \begin{cases} 1 & \text{if } y_i \text{ belong to } f_{\text{LogNormal}}(y; \mu, \sigma^2) \\ 0 & \text{if } y_i \text{ belong to } f_{\text{Exp}}(y; \lambda) \end{cases}$$

The observed data in this context is $\boldsymbol{y} = (y_1 \ldots y_n)$ and the missing data is $\boldsymbol{z} = (z_1 \ldots z_n)$. The likelihood of the complete data $(\boldsymbol{y}, \boldsymbol{z})$ is

$$\begin{aligned} L(\theta; \boldsymbol{y}, \boldsymbol{z}) &= \prod_{i=1}^{n} \left(p \cdot \frac{1}{y_i\sqrt{2\pi\sigma^2}} \exp\left\{\frac{1}{2\sigma^2}(\log y_i - \mu)^2\right\}\right)^{z_i} \left((1-p) \cdot \lambda e^{-\lambda y_i}\right)^{1-z_i} \\ \implies \log L(\theta; \boldsymbol{y}, \boldsymbol{z}) &= \sum_{i=1}^{n} z_i \log\left(p \cdot \frac{1}{y_i\sqrt{2\pi\sigma^2}} \exp\left\{\frac{1}{2\sigma^2}(\log y_i - \mu)^2\right\}\right) + \sum_{i=1}^{n}(1-z_i)\log\left((1-p) \cdot \lambda e^{-\lambda y_i}\right) \end{aligned} \tag{11}$$

with the corresponding log likelihood (11), we proceed to E-Step,

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta^{(t)}}) &= \mathbb{E}_Z(\log L(\theta; \boldsymbol{y}, \boldsymbol{z})|\boldsymbol{y}, \boldsymbol{\theta^{(t)}}) \\ &= \sum_{i=1}^{n} \mathbb{E}(Z_i|y_i, \theta^{(t)}) \log\left(p \cdot \frac{1}{y_i\sqrt{2\pi\sigma^2}} \exp\left\{\frac{1}{2\sigma^2}(\log y_i - \mu)^2\right\}\right) \\ &+ \sum_{i=1}^{n}(1 - \mathbb{E}(Z_i|y_i, \theta^{(t)}))\log\left((1-p) \cdot \lambda e^{-\lambda y_i}\right) \end{aligned} \tag{12}$$

We know that $\mathbb{E}(Z_i|\boldsymbol{y}, \boldsymbol{\theta^{(t)}}) = \mathbb{P}(Z_i = 1|y_i, \theta^{(t)})$, and applying Bayes Theorem and the Law of Total Probability, we obtain,

$$
\begin{aligned}
\mathbb{E}(Z_i|\boldsymbol{y},\boldsymbol{\theta^{(t)}}) &= \mathbb{P}(Z_i = 1|y_i, \theta^{(t)}) \\
&= \frac{\left(p^{(t)} \cdot \frac{1}{y_i\sqrt{2\pi(\sigma^2)^{(t)}}}\exp\left\{\frac{1}{2(\sigma^2)^{(t)}}(\log y_i - \mu^{(t)})^2\right\}\right)}{\left(p \cdot \frac{1}{y_i\sqrt{2\pi(\sigma^2)^{(t)}}}\exp\left\{\frac{1}{2(\sigma^2)^{(t)}}(\log y_i - \mu^{(t)})^2\right\}\right)\left((1-p^{(t)})\cdot\lambda^{(t)}e^{-\lambda^{(t)}y_i}\right)} \\
&= \tilde{p}_i^{(t)}, \quad i = 1,\ldots,n
\end{aligned} \tag{13}
$$

Therefore, we substitute 13 into 12

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta^{(t)}}) = &\sum_{i=1}^{n}\tilde{p}_i^{(t)}\log\left(p\cdot\frac{1}{y_i\sqrt{2\pi\sigma^2}}\exp\left\{\frac{1}{2\sigma^2}(\log y_i-\mu)^2\right\}\right) \\
&+\sum_{i=1}^{n}(1-\tilde{p}_i^{(t)})\log\left((1-p)\cdot\lambda e^{-\lambda y_i}\right)
\end{aligned} \tag{14}
$$

For the M-step, we only need to compute the partial derivatives

$$
\begin{aligned}
\frac{\partial}{\partial p}Q(\boldsymbol{\theta}|\boldsymbol{\theta^{(t)}}) = 0 &\implies p^{(t+1)} = \frac{\sum_{i=1}^{n}\tilde{p}_i^{(t)}}{n} \\
\frac{\partial}{\partial \mu}Q(\boldsymbol{\theta}|\boldsymbol{\theta^{(t)}}) = 0 &\implies \mu^{(t+1)} = \frac{\sum_{i=1}^{n}\tilde{p}_i^{(t)}\log(y_i)}{\sum_{i=1}^{n}\tilde{p}_i^{(t)}} \\
\frac{\partial}{\partial \sigma^2}Q(\boldsymbol{\theta}|\boldsymbol{\theta^{(t)}}) = 0 &\implies (\sigma^{(t+1)})^2 = \frac{\sum_{i=1}^{n}\tilde{p}_i^{(t)}(\log(y_i)-\mu^{(t+1)})^2}{\sum_{i=1}^{n}\tilde{p}_i^{(t)}y_i} \\
\frac{\partial}{\partial \lambda}Q(\boldsymbol{\theta}|\boldsymbol{\theta^{(t)}}) = 0 &\implies \lambda^{(t+1)} = \frac{\sum_{i=1}^{n}(1-\tilde{p}_i^{(t)})}{\sum_{i=1}^{n}y_i(1-\tilde{p}_i^{(t)})}
\end{aligned} \tag{15}
$$

b)

Using the dataset `datasetex5.Rdata` implement the EM algorithm and find the MLEs for each component of θ. As starting values, you might want to consider $\theta^{(0)} = (p^{(0)}, \mu^{0)}, (\sigma^{(0)})^2, \lambda^{(0)}) = (0.1, 1, 0.5^2, 2)$. Draw the histogram of the data with the estimated density superimposed.

Answer :

In the code, we have used for the stopping criterion

$$|p^{(t+1)} - p^{(t)}| + |\mu^{(t+1) - \mu(t)}| + |(\sigma^{(t+1)})^2 - (\sigma^{(t)})^2| + |\lambda^{(t+1)} - \lambda^{(t)}| < \varepsilon$$

with $\varepsilon = 0.00001$. For the starting values we use $\theta^{(0)} = (p^{(0)}, \mu^{0)}, (\sigma^{(0)})^2, \lambda^{(0)}) = (0.1, 1, 0.5^2, 2)$ as given.

```
mixture.model <- function(y, theta, eps){
  n <- length(y)
  #initialising the parameters
  p <- theta[1]; mu <- theta[2]; sigma <- theta[3]; lambda <- theta[4]
  diff <- 1
  while(diff > eps){
  theta.old <- theta
  #E-step: computing ptilde
  ptilde1 <- p*dlnorm(y, mean = mu, sd = sigma)
  ptilde2 <- (1 - p)*dexp(y, lambda)
  ptilde <- ptilde1/(ptilde1 + ptilde2)

  #M-step: computing each parameter
  p <- mean(ptilde)
  mu <- sum(log(y)*ptilde)/sum(ptilde)
  sigma <- sqrt(sum(((log(y) - mu)^2)*ptilde)/sum(ptilde))
  lambda <- sum(1 - ptilde)/sum(y*(1 - ptilde))

  #Checking with stopping criterion
  theta <- c(p, mu, sigma, lambda)
  diff <- sum(abs(theta - theta.old))
  }
  return(theta)
}
#Performing the EM algorithm on Mixture Model
res <- mixture.model(y = dataex5, c(0.1, 1, 0.5, 2), 0.000001)
p <- res[1]; mu <- res[2]; sigma <- res[3]; lambda <- res[4]
cat("The value for p is", p)
```

```
## The value for p is 0.47955
```

```
cat("The value for mu is", mu)
```

```
## The value for mu is 2.013262
```

```
cat("The value for sigma^2 is", sigma)
```

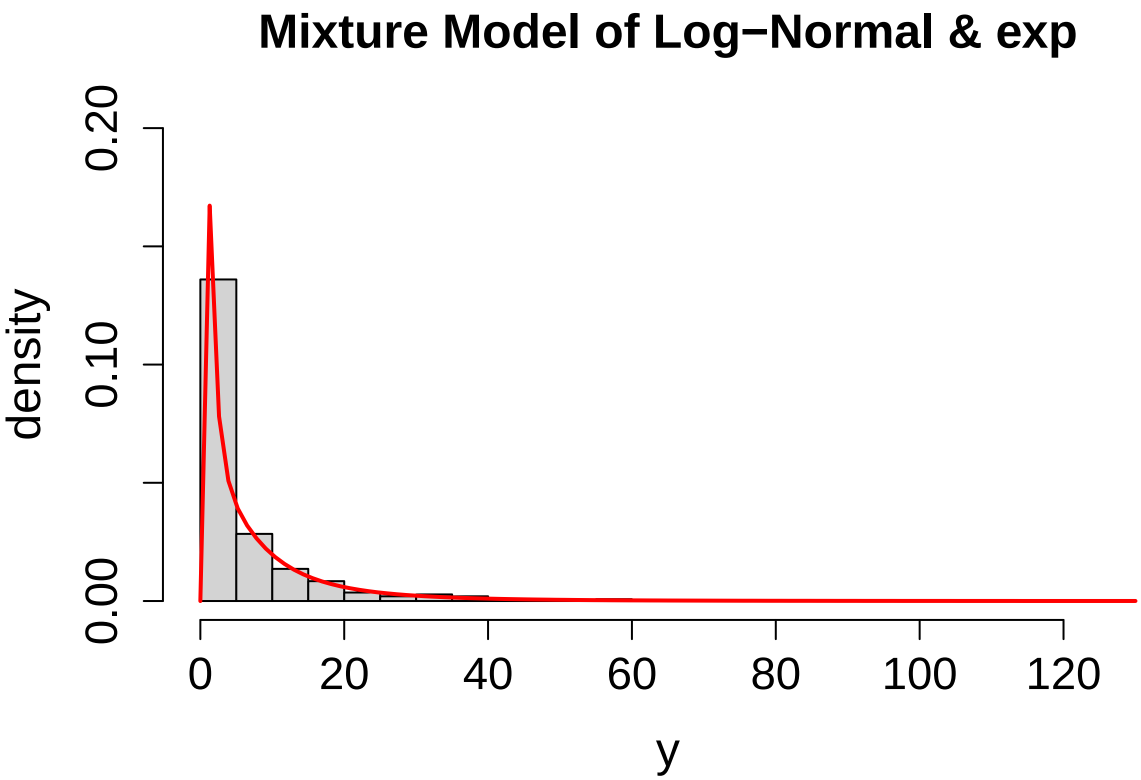
```
## The value for sigma^2 is 0.9293769
```

```
cat("The value for lambda is", lambda)
```

```
## The value for lambda is 1.033019
```

```
#plotting the histogram of the mixture model
hist(dataex5, breaks=45, main = "Mixture Model of Log-Normal & exp",
```

```r
     xlab = "y", ylab = "density", ylim = c(0,0.2),
     cex.main = 1.5, cex.lab = 1.5, cex.axis = 1.4, freq = F)
curve(p*dlnorm(x, mu, sigma)+(1-p)*dexp(x, lambda), add = TRUE, lwd = 2, col = "red")
```



Using the **R** code provided, we obtained $\hat{p} = 0.480, \quad \hat{\mu} = 2.01 \quad \hat{\sigma}^2 = 0.929 \quad \hat{\lambda} = 1.03$. The plot of the observed counts against the expected counts under this mixture model is shown above.