Incomplete Data Analysis: Assignment 2

Xiao Heng (s2032451)

1.

(a) By the definition, we have $F(y;\theta) + S(y;\theta) = 1$, where $F(y;\theta) = Pr(Y \leq y)$ is the given cumulative distribution function. In this case, we can easily obtain $S(y;\theta) = 1 - F(y;\theta) = e^{-y^2/(2\theta)}, y \geq 0, \theta > 0$, as well as $Pr(Y > C) = S(C;\theta) = e^{-C^2/(2\theta)}$. As for the density of the observed, it follows $f(y;\theta) = F'(y;\theta) = \frac{y}{\theta}e^{-y^2/(2\theta)}$.

Now we can write:

$$X_i = Y_i \times I(Y_i \leq C) + C \times I(Y_i > C) = Y_i R_i + C(1 - R_i)$$

The likelihood is thus of the form:

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n}\{f(y_i;\theta)^{r_i} S(C;\theta)^{1-r_i}\} \\
&= \prod_{i=1}^{n}\{[\frac{y_i}{\theta}e^{-y_i^2/(2\theta)}]^{r_i}[e^{-C^2/(2\theta)}]^{1-r_i}\} \\
&= (\frac{y_i}{\theta})^{\sum_{i=1}^{n} r_i} e^{-\frac{1}{2\theta}\sum_{i=1}^{n} y_i^2 r_i} e^{-\frac{1}{2\theta}\sum_{i=1}^{n} C^2(1-r_i)} \\
&= (\frac{y_i}{\theta})^{\sum_{i=1}^{n} r_i} e^{-\frac{1}{2\theta}\sum_{i=1}^{n}[y_i^2 r_i + C^2(1-r_i)]}
\end{aligned}$$

Noting that $x_i^2 = y_i^2 r_i + C^2(1 - r_i)$, we have:

$$L(\theta) = (\frac{y_i}{\theta})^{\sum_{i=1}^{n} r_i} e^{-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2}$$

The corresponding loglikelihood is:

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

We thus have:

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

From the definition of $F(y;\theta)$ we know that $\theta > 0 (\neq 0)$, so that by multiplying the θ^2 on both sides and letting the equation to be 0, we obtain:

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

(b) We have:

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

The expected information is:

$$
\begin{aligned}
I(\theta) &= -E(\frac{1}{\theta^2}\sum_{i=1}^{n} R_i - \frac{1}{\theta^3}\sum_{i=1}^{n} X_i^2) \\
&= -\frac{n}{\theta^2}E(R) + \frac{1}{\theta^3}E(\sum_{i=1}^{n} X_i^2) \\
&= -\frac{n}{\theta^2}E(R) + \frac{1}{\theta^3}E(\sum_{i=1}^{n}[Y_i^2 R_i + C^2(1-R_i)]) \\
&= -\frac{n}{\theta^2}E(R) + \frac{1}{\theta^3}E(\sum_{i=1}^{n} Y_i^2 R_i) + \frac{nC^2}{\theta^3}(1-E(R)) \\
&= -\frac{n}{\theta^2}(1+\frac{C^2}{\theta})E(R) + \frac{1}{\theta^3}E(\sum_{i=1}^{n} Y_i^2 R_i) + \frac{nC^2}{\theta^3}
\end{aligned}
$$

For $E(R)$, note that R is a binary random variable and so:

$$
\begin{aligned}
E(R) &= 1 \times Pr(R=1) + 0 \times Pr(R=0) \\
&= Pr(R=1) \\
&= Pr(Y \leq C) \\
&= F(C;\theta) \\
&= 1 - e^{-C^2/(2\theta)}
\end{aligned}
$$

As for $E(\sum_{i=1}^{n} Y_i^2 R_i)$:

$$
\begin{aligned}
E(\sum_{i=1}^{n} Y_i^2 R_i) &= n\int_0^C y^2 f(y;\theta)dy \\
&= n(-C^2 e^{-C^2/(2\theta)} + 2\theta(1-e^{-C^2/(2\theta)}))
\end{aligned}
$$

Take these expectations back, we finally obtain:

$$
\begin{aligned}
I(\theta) &= -\frac{n}{\theta^2}(1+\frac{C^2}{\theta})(1-e^{-C^2/(2\theta)}) + \frac{n}{\theta^3}(-C^2 e^{-C^2/(2\theta)} + 2\theta(1-e^{-C^2/(2\theta)})) + \frac{nC^2}{\theta^3} \\
&= \frac{n}{\theta^2}[-(1-e^{-C^2/(2\theta)}) - \frac{c^2}{\theta} + \frac{c^2}{\theta}e^{-C^2/(2\theta)} - \frac{c^2}{\theta}e^{-C^2/(2\theta)} + 2(1-e^{-C^2/(2\theta)}) + \frac{c^2}{\theta}] \\
&= \frac{n}{\theta^2}(1-e^{-C^2/(2\theta)})
\end{aligned}
$$

(c) Following the asymptotic normality, we have:

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

And it also holds with more convenience that:

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

The corresponding standard error of maximum likelihood estimator $\hat{\theta}_{MLE}$ then follows:

$$(\hat{\theta}_{MLE} - \theta) \sim N_p(0, (J(\hat{\theta}_{MLE}))^{-1})$$

In this case, the 95% confidence interval ($z_{0.025} = 1.96$) for θ is:

$$\hat{\theta}_{MLE} \pm \frac{1.96}{\sqrt{J(\hat{\theta}_{MLE})}}$$

2.

(a) Given the normal density function $\phi(\cdot;\mu,\sigma^2)$ and cumulative distribution function $\Phi(\cdot;\mu,\sigma^2)$, we then have the process:

$$\begin{aligned}
L(\mu,\sigma^2|\mathbf{x},\mathbf{r}) &= f(\mathbf{r}|\theta,\sigma^2)L(\theta,\sigma^2|\mathbf{x}) \\
&= \{\prod_{i=1}^{n} Pr(X_i < x)^{1-r_i}\}\{\prod_{i=1}^{n}\phi(x_i;\mu,\sigma^2)^{r_i}\} \\
&= \{\prod_{i=1}^{n} \Phi(x_i;\mu,\sigma^2)^{1-r_i}\}\{\prod_{i=1}^{n}\phi(x_i;\mu,\sigma^2)^{r_i}\} \\
&= \prod_{i=1}^{n}\phi(x_i;\mu,\sigma^2)^{r_i}\Phi(x_i;\mu,\sigma^2)^{1-r_i}
\end{aligned}$$

In this case, the corresponding log likelihood of the observed data $\{(x_i, r_i)\}_{i=1}^n$ is:

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

(b)

```
log_like_norm <- function(X, R, mu, sigma = 1.5)
{
  sum(R*dnorm(x = X, mean = mu, sd = sigma, log = TRUE)
      + (1-R)*pnorm(X, mean = mu, sd = sigma, log.p = TRUE))
}
mle_2b <- maxLik(logLik = log_like_norm, X = dataex2$X, R = dataex2$R, sigma = 1.5,
                 start = mean(dataex2$X))
summary(mle_2b)
```

```
## --------------------------------------------
## 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)
## [1,]   5.5328     0.1075   51.48  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## --------------------------------------------
```

Note that the initial value of μ is set as the mean of observed data, which is intuitively close to the true value (although not necessary). In this case, with package `maxLik`, the maximum likelihood estimate of μ based on the data available in the file `dataex2.Rdata` is 5.5328.

3.

Following the lecture, a missing data mechanism is **ignorable** for likelihood inference if:

- the missing data are MAR (or MCAR) and
- the parameter ψ (missingness mechanism) and θ (data model) are distinct/disjoint, in the sense that the joint parameter space of (ψ,θ) is the product of the parameter spaces Ψ and Θ (separability condition).

(a) **Ignorable.** 1. The missing mechanism is MAR, since it is only related with fully observed data Y_1 and independent of Y_2. 2. The parameters $\psi = (\psi_0, \psi_1)$ are distinct from θ.

(b) **Not ignorable.** The missing mechanism is MNAR, since it is not independent of Y_2 (which contains missing values). In this case, there is no need to do further check for parameters.

(c) **Not ignorable.** Although the missing mechanism is MAR, since it is only related with fully observed data Y_1 and independent of Y_2, the parameters of missing mechanism contains μ_1, which is not distinct (though ψ is distinct), so that the second condition is not satisfied.

4.

Given the bernoulli distribution, we know that

$$f(y;\boldsymbol{\beta}) = (\frac{e^{\beta_0+x\beta_1}}{1+e^{\beta_0+x\beta_1}})^y(\frac{1}{1+e^{\beta_0+x\beta_1}})^{1-y}$$

The likelihood of the complete data is:

$$\begin{aligned}
L(\boldsymbol{\beta}|\mathbf{y}_{obs},\mathbf{y}_{mis}) &= f(\mathbf{y}_{obs};\boldsymbol{\beta})f(\mathbf{y}_{mis};\boldsymbol{\beta}) \\
&= \prod_{i=1}^{m} f(y_i;\boldsymbol{\beta}) \prod_{j=m+1}^{n} f(y_j;\boldsymbol{\beta}) \\
&= \prod_{i=1}^{m}[(\frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}})^{y_i}(\frac{1}{1+e^{\beta_0+x_i\beta_1}})^{1-y_i}] \prod_{j=m+1}^{n}[(\frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}})^{y_j}(\frac{1}{1+e^{\beta_0+x_j\beta_1}})^{1-y_j}]
\end{aligned}$$

The complete log likelihood is then:

$$\begin{aligned}
\log L(\boldsymbol{\beta}|\mathbf{y}_{obs},\mathbf{y}_{mis}) &= \sum_{i=1}^{m}\{y_i[\beta_0+x_i\beta_1-\log(1+e^{\beta_0+x_i\beta_1})]+(1-y_i)[-\log(1+e^{\beta_0+x_i\beta_1})]\} \\
&+ \sum_{j=m+1}^{n}\{y_j[\beta_0+x_j\beta_1-\log(1+e^{\beta_0+x_j\beta_1})]+(1-y_j)[-\log(1+e^{\beta_0+x_j\beta_1})]\} \\
&= \sum_{i=1}^{m}[y_i(\beta_0+x_i\beta_1)-\log(1+e^{\beta_0+x_i\beta_1})] + \sum_{j=m+1}^{n}[y_j(\beta_0+x_i\beta_1)-\log(1+e^{\beta_0+x_i\beta_1})] \\
&= \sum_{i=1}^{m} y_i(\beta_0+x_i\beta_1) + \sum_{j=m+1}^{n} y_j(\beta_0+x_j\beta_1) - \sum_{k=1}^{n}\log(1+e^{\beta_0+x_k\beta_1})
\end{aligned}$$

At iteration $(t+1)$ of the iterative procedure, the `E-step` calculates the conditional expectation, with respect to the missing observations, of the complete data log-likelihood given the observed data and the estimate of $\boldsymbol{\beta}$ from iteration t, $\boldsymbol{\beta}^{(t)}$, that is:

$$\begin{aligned}
Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}) &= E_{Y_{mis}}[\log L(\boldsymbol{\beta}|\mathbf{y}_{obs},\mathbf{y}_{mis})|\mathbf{y}_{obs},\boldsymbol{\beta}^{(t)}] \\
&= E_{Y_{mis}}[\sum_{i=1}^{m} y_i(\beta_0+x_i\beta_1) + \sum_{j=m+1}^{n} y_j(\beta_0+x_j\beta_1) - \sum_{k=1}^{n}\log(1+e^{\beta_0+x_k\beta_1})|\mathbf{y}_{obs},\boldsymbol{\beta}^{(t)}] \\
&= \sum_{i=1}^{m} y_i(\beta_0+x_i\beta_1) - \sum_{k=1}^{n}\log(1+e^{\beta_0+x_k\beta_1}) + E_{Y_{mis}}[\sum_{j=m+1}^{n} y_j(\beta_0+x_j\beta_1)|\boldsymbol{\beta}^{(t)}]
\end{aligned}$$

In the Bernoulli distribution, we know that $E[Y_i] = p_i(\boldsymbol{\beta}) = \frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}}$

As for the expectation term, we have:

$$
\begin{aligned}
E_{Y_{mis}}[\sum_{j=m+1}^{n} y_j(\beta_0 + x_j\beta_1)|\boldsymbol{\beta}^{(t)}] &= \sum_{j=m+1}^{n} E_{Y_{mis}}[y_j(\beta_0 + x_j\beta_1)|\boldsymbol{\beta}^{(t)}] \\
&= \sum_{j=m+1}^{n} \frac{(\beta_0 + x_j\beta_1)e^{\beta_0^{(t)}+x_j\beta_1^{(t)}}}{1 + e^{\beta_0^{(t)}+x_j\beta_1^{(t)}}}
\end{aligned}
$$

Therefore:

$$
Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}) = \sum_{i=1}^{m} y_i(\beta_0 + x_i\beta_1) - \sum_{k=1}^{n} \log(1 + e^{\beta_0+x_k\beta_1}) + \sum_{j=m+1}^{n} \frac{(\beta_0 + x_j\beta_1)e^{\beta_0^{(t)}+x_j\beta_1^{(t)}}}{1 + e^{\beta_0^{(t)}+x_j\beta_1^{(t)}}}
$$

We can now proceed to the `M-step` and we have that:

$$
\frac{\partial}{\partial\beta_0} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}) = \sum_{i=1}^{m} y_i - \sum_{k=1}^{n} \frac{e^{\beta_0+x_k\beta_1}}{1 + e^{\beta_0+x_k\beta_1}} + \sum_{j=m+1}^{n} \frac{e^{\beta_0^{(t)}+x_j\beta_1^{(t)}}}{1 + e^{\beta_0^{(t)}+x_j\beta_1^{(t)}}}
$$

$$
\frac{\partial}{\partial\beta_1} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}) = \sum_{i=1}^{m} x_i y_i - \sum_{k=1}^{n} \frac{x_k e^{\beta_0+x_k\beta_1}}{1 + e^{\beta_0+x_k\beta_1}} + \sum_{j=m+1}^{n} \frac{x_j e^{\beta_0^{(t)}+x_j\beta_1^{(t)}}}{1 + e^{\beta_0^{(t)}+x_j\beta_1^{(t)}}}
$$

For the `M-step`, letting $\frac{\partial}{\partial\beta_0} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}) = 0$ and $\frac{\partial}{\partial\beta_1} Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}) = 0$, with updating iterations, we can maximise $Q(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)})$ respect to $\boldsymbol{\beta}$. However, here the partial derivatives are not easy to calculate, so we adopt the `maxLik` method directly:

```
ind_mis <- which(is.na(dataex4$Y))
X <- c(dataex4$X[-ind_mis],dataex4$X[ind_mis])
Y_obs <- dataex4$Y[-ind_mis]
m_obs <- length(Y_obs)
n <- length(X)
diff <- 1
eps <- 0.00001
beta01t_old <- c(1,1)
Q <- function(X, Y_obs, beta01, beta01t, m_obs, n)
{
  sum(Y_obs*(beta01[1]+X[1:m_obs]*beta01[2]))-sum(log(1+exp(beta01[1]+
  X*beta01[2])))+sum((beta01[1]+X[(m_obs+1):n]*beta01[2])*exp(beta01t[1]+
  X[(m_obs+1):n]*beta01t[2])/(1+exp(beta01t[1]+X[(m_obs+1):n]*beta01t[2])))
}
while(diff > eps)
{
  beta01t_new <- maxLik(logLik = Q, X = X, Y_obs = Y_obs, m_obs = m_obs, n = n,
                        beta01t = beta01t_old, start = beta01t_old)$estimate
  diff <- sum(abs(beta01t_old-beta01t_new))
  beta01t_old <- beta01t_new
}
beta01t_new
```

```
## [1]  0.9755281 -2.4803722
```

In this case, $\beta_0 = 0.9755$ and $\beta_1 = -2.4803$.

5.

(a) Folloowing the process in lecture, we define an augmented complete dataset where $\mathbf{y}_{obs} = (y_1 \ldots, y_n)$ and $\mathbf{y}_{mis} = (y_1 \ldots, y_n)$ is a vector of unobserved/latent group data indicator, such that:

$$z_i = \begin{cases} 1, & \text{if } y_i \text{ if belongs to the first component} \\ 0 & \text{if } y_i \text{ if belongs to the second component} \end{cases}$$

Note that $Z_i \overset{\text{iid}}{\sim}$ Bernoulli(p)\$ (so that $Pr(Z_i = 1) = p$). Thus, the complete data likelihood follows:

$$L(\theta|y,z) = \prod_{i=1}^{n} \{[p f_{\text{LogNormal}}(y;\mu,\sigma^2)]^{Z_i}[(1-p)f_{\text{Exp}}(y;\lambda)]^{1-Z_i}\}$$

Corresponding log likelihood is given by:

$$\log L(\theta|y,z) = \sum_{i=1}^{n} Z_i\{\log p + \log f_{\text{LogNormal}}(y;\mu,\sigma^2)\} + \sum_{i=1}^{n}(1-Z_i)\{\log(1-p) + \log f_{\text{Exp}}(y;\lambda)\}$$

For the `E-step` we would need to compute:

$$\begin{aligned} Q(\theta|\theta^{(t)}) &= E_Z[\log L(\theta|y,z)|y,\theta^{(t)}] \\ &= \sum_{i=1}^{n} E[Z_i|y,\theta^{(t)}]\{\log p + \log f_{\text{LogNormal}}(y;\mu,\sigma^2)\} + \sum_{i=1}^{n}(1 - E[Z_i|y,\theta^{(t)}])\{\log(1-p) + \log f_{\text{Exp}}(y;\lambda)\} \end{aligned}$$

Now:

$$\begin{aligned} E[Z_i|y,\theta^{(t)}] &= E[Z_i|y_i,\theta^{(t)}] \\ &= 1 \times Pr(Z_i = 1|y_i,\theta^{(t)}) + 0 \times Pr(Z_i = 0|y_i,\theta^{(t)}) \\ &= \frac{p^{(t)} f_{\text{LogNormal}}(y;\mu^{(t)},(\sigma^{(t)})^2)}{p^{(t)} f_{\text{LogNormal}}(y;\mu^{(t)},(\sigma^{(t)})^2) + (1-p^{(t)})f_{\text{Exp}}(y;\lambda^{(t)})} \\ &= \widetilde{p}_i^{(t)} \end{aligned}$$

In this case:

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n} \widetilde{p}_i^{(t)}\{\log p + \log f_{\text{LogNormal}}(y;\mu,\sigma^2)\} + \sum_{i=1}^{n}(1-\widetilde{p}_i^{(t)})\{\log(1-p) + \log f_{\text{Exp}}(y;\lambda)\}$$

Now, for the `M-step`:

$$
\begin{aligned}
\frac{\partial}{\partial p} Q(\theta|\theta^{(t)}) &= \frac{1}{p}\sum_{i=1}^{n}\widetilde{p}_i^{(t)} - p\sum_{i=1}^{n}(1-\widetilde{p}_i^{(t)}) = 0 \\
\Rightarrow p^{(t+1)} &= \frac{\sum_{i=1}^{n}\widetilde{p}_i^{(t)}}{n} \\
\frac{\partial}{\partial \mu} Q(\theta|\theta^{(t)}) &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\widetilde{p}_i^{(t)}(\log y_i - \mu) = 0 \\
\Rightarrow \mu^{(t+1)} &= \frac{\sum_{i=1}^{n}\widetilde{p}_i^{(t)}\log y_i}{\sum_{i=1}^{n}\widetilde{p}_i^{(t)}} \\
\frac{\partial}{\partial \sigma^2} Q(\theta|\theta^{(t)}) &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\widetilde{p}_i^{(t)} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\widetilde{p}_i^{(t)}(\log y_i - \mu)^2 = 0 \\
\Rightarrow (\sigma^{(t+1)})^2 &= \frac{\sum_{i=1}^{n}\widetilde{p}_i^{(t)}(\log y_i - \mu^{(t+1)})^2}{\sum_{i=1}^{n}\widetilde{p}_i^{(t)}} \ (\sigma > 0) \\
\frac{\partial}{\partial \lambda} Q(\theta|\theta^{(t)}) &= \frac{1}{\lambda}\sum_{i=1}^{n}(1-\widetilde{p}_i^{(t)}) - \sum_{i=1}^{n}(1-\widetilde{p}_i^{(t)})y_i = 0 \\
\Rightarrow \lambda^{(t+1)} &= \frac{\sum_{i=1}^{n}(1-\widetilde{p}_i^{(t)})}{\sum_{i=1}^{n}(1-\widetilde{p}_i^{(t)})y_i}
\end{aligned}
$$

(b) Based on the updating rules above, we can construct the EM iterative function as below:

```
EM_mixture <- function(y, theta0, eps){
  n <- length(y)
  theta <- theta0
  p <- theta[1]
  mu <- theta[2]; sigma <- theta[3]; lambda <- theta[4]
  diff <- 1
  while(diff > eps){
    theta.old <- theta
    #E-step
    ptilde1 <- p * dlnorm(y, meanlog = mu, sdlog = sigma)
    ptilde2 <- (1 - p) * dexp(y, rate = lambda)
    ptilde <- ptilde1 / (ptilde1 + ptilde2)
    #M-step
    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((1 - ptilde) * y)
    theta <- c(p, mu, sigma, lambda)
    diff <- sum(abs(theta - theta.old))
  }
  return(theta)
}
```

Now we adopt the starting values and apply the **EM_mixture**() function, showing the result of estimated parameters as follows:

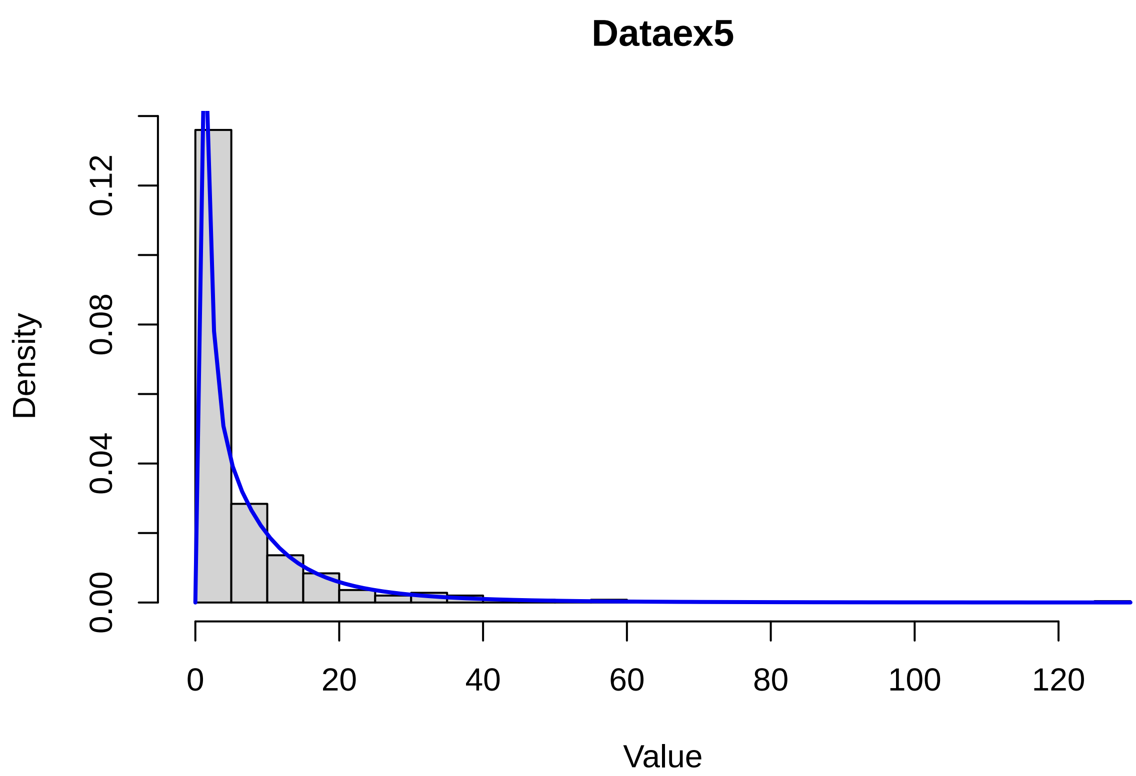
```
res <- EM_mixture(y = dataex5, c(0.1, 1, 0.5, 2), 0.00001)
p <- res[1]
mu <- res[2]
sigma <- res[3]
```

```
lambda <- res[4]
p; mu; sigma; lambda
```

```
## [1] 0.4795916

## [1] 2.013147

## [1] 0.9294364

## [1] 1.033139
```

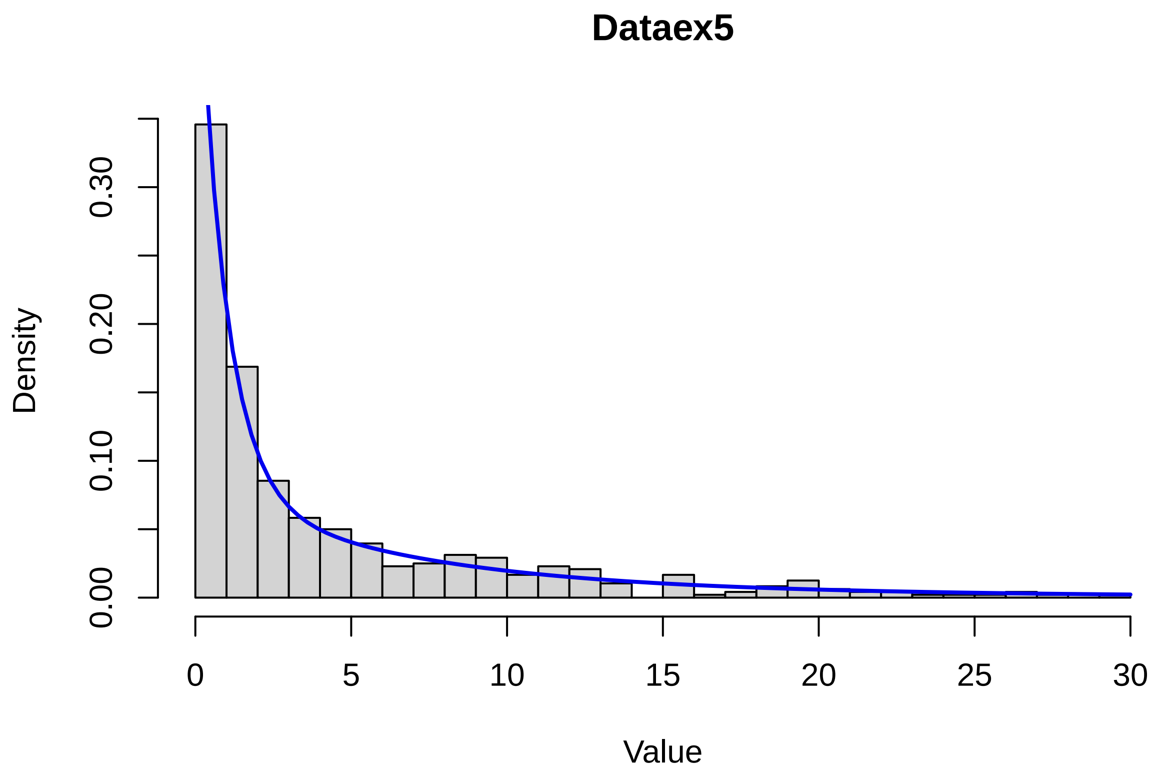
With the estimated parameters, we can draw the histogram of the data with the estimated density superimposed:

```
hist(dataex5, main = "Dataex5", xlab = "Value", ylab = "Density", freq = F, breaks = 40)
curve(p*dlnorm(x, mu, sigma) + (1 - p)*dexp(x, lambda),
      add = TRUE, lwd = 2, col = "blue2")
```



A plot with `x-axis` within $(0, 30)$ is also given, for a clearer check:

```
hist(dataex5[which(dataex5<30)], main = "Dataex5", xlab = "Value", ylab = "Density",
     xlim=c(0, 30), freq = F, breaks = 40)
curve(p*dlnorm(x, mu, sigma) + (1 - p)*dexp(x, lambda),
      add = TRUE, lwd = 2, col = "blue2")
```


Dataex5
Density
0.00
0.10
0.20
0.30
0
5
10
15
20
25
30
Value