Incomplete Data Analysis
Assignment 2

Josephine Li s2346729

Question 1

Sub Question 1-a

First, we try to obtain the density function of Z, which is f_Z. First, calculate the cumulative distribution function of Z, which is F_Z. The cdf of Z with $z \geq 1$ and $\lambda, \mu > 0$ is:

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(min\{X, Y\} \leq z) \\ &= 1 - P(min\{X, Y\} > z) \\ &= 1 - P(X > z, Y > z) \\ &= 1 - P(X > z) \times p(Y > z) \\ &= 1 - (1 - P(X \leq z)) \times (1 - P(Yleqz)) \\ &= 1 - (1 - F_X(z)) \times (1 - F_Y(z)) \\ &= 1 - z^{-(\lambda+\mu)} \end{aligned}$$

And the density function of Z with $z \geq 1$ and $\lambda, \mu > 0$ is:

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{\lambda + \mu}{z^{\lambda+\mu+1}}$$

Then, try to obtain the frequency function of δ, which is f_δ. From 2 cumulative density function, we can calculate the density functions of X and Y with $x, y \geq 1$ and $\lambda, \mu > 0$, which is:

$$\begin{aligned} f_X(x; \lambda) &= \frac{d}{dx} F_X(x) = \frac{\lambda}{x^{\lambda+1}} \\ f_Y(y; \mu) &= \frac{d}{dy} F_Y(y) = \frac{\mu}{y^{\mu+1}} \end{aligned}$$

As 2 distributions are independent, he joint density function of 2 distribution X, Y is:

$$f_{XY}(x, y; \lambda, \mu) = f_X(x; \lambda) \times f_Y(y; \mu) = \frac{\lambda \times \mu}{x^{\lambda+1} \times y^{\mu+1}}$$

To calculate the frequency when $\delta = 1$, we need to calculate $P(X < Y)$, which is:

$$P(X < Y) = \int_1^\infty \int_x^\infty \frac{\lambda \times \mu}{x^{\lambda+1} \times y^{\mu+1}} dydx = \frac{\lambda}{\lambda + \mu}$$

And the frequency when $\delta = 0$ is:

$$P(X \geq Y) = 1 - P(X < Y) = \frac{\mu}{\lambda + \mu}$$

Therefore, the frequency function of δ is:

$$f_\delta = \begin{cases} \frac{\lambda}{\lambda+\mu}, & \delta = 1 \\ \frac{\mu}{\lambda+\mu}, & \delta = 0 \end{cases}$$

The distribution of Z is Pareto-distribution with parameter $\lambda + \mu$, and the distribution of δ is Bernoulli distribution with parameter $\frac{\lambda}{\lambda+\mu}$.

Sub Question 1-b

For samples $Z_1, Z_2, \ldots, Z_n$ from $f_Z(z; \theta)$ with $\theta = \lambda + \mu$, the joint density of observation $\mathbb{z} = (z_1, z_2, ..., z_n)$ is:

$$f_Z(\mathbb{z}; \theta) = \prod_{i=1}^{n} f_Z(z_i; \theta) = \theta^n \prod_{i=1}^{n} z_i^{-(\theta+1)} = L(\theta; \mathbb{z})$$

$L(\theta; z)$ is the likelihood function, and the log likelihood is:

$$\begin{aligned} \log L(\theta; z) &= \log \theta^n \prod_{i=1}^{n} z_i^{-(\theta+1)} \\ &= \log \theta^n + \log \prod_{i=1}^{n} z_i^{-(\theta+1)} \\ &= n \log \theta - (\theta + 1) \sum_{i=1}^{n} \log z_i \end{aligned}$$

The maximum likelihood estimator of θ is:

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} L(\theta; z)$$

To calculate the value of $\hat{\theta}_{MLE}$, we need to calculate the first derivatives for log likelihood function and let it equal to zero, which is:

$$\frac{d}{d\theta} \log L(\theta; \mathbb{z}) = \frac{n}{\theta} - \sum_{i=1}^{n} \log z_i = 0$$

We get the result $\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^{n} \log z_i}$. To ensure that we have obtained a maximum, we need to confirm the derivative of the score function is negative. The derivative of the log likelihood function is:

$$\frac{d^2}{d\theta d\theta^T} \log L(\theta; \mathbb{z}) = -\frac{n}{\theta^2}$$

As $z_i \geq 1$, the second derivative is negative. Hence, the value of $\hat{\theta}_{MLE}$ is:

$$\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^{n} \log z_i}$$

We can do the same steps for deriving maximum likelihood estimator for p.

For samples $\delta_1, \delta_2, \ldots, \delta_n$ from $f_\delta(d; p)$ with $p = \frac{\lambda}{\lambda+\mu}$, the probability mass function for $\delta = (d_1, d_2, ..., d_n)$ is:

$$f_\delta(\mathbb{d}; p) = p^d (1-p)^{1-d}, d \in \{0, 1\}$$

The likelihood function for p is:

$$L(p; \mathbb{d}) = \prod_{i=1}^{n} p^{d_i}(1-p)^{1-d_i} = p^{\sum_{i=1}^{n} d_i} \times (1-p)^{n-\sum_{i=1}^{n} d_i}$$

$$\log L(p; \mathbb{d}) = \log(p^{\sum_{i=1}^{n} d_i} \times (1-p)^{n-\sum_{i=1}^{n} d_i}) = \sum_{i=1}^{n} d_i \log p + (n - \sum_{i=1}^{n} d_i) \log(1-p)$$

Calculating the first derivatives for log likelihood function and let it equal to zero, which is:

$$\frac{d}{d\theta} \log L(p; \mathbb{d}) = \frac{1}{p} \sum_{i=1}^{n} d_i - \frac{1}{1-p}(n - \sum_{i=1}^{n} d_i) = 0$$

We get the result $\hat{p}_{MLE} = \frac{\sum_{i=1}^{n} d_i}{n} = \overline{\delta}$. To ensure that we have obtained a maximum, we need to confirm the derivative of the score function is negative. The derivative of the log likelihood function is:

$$\frac{d^2}{dpdp^T} \log L(p; \mathbb{d}) = -\frac{1}{p^2} \sum_{i=1}^{n} d_i - \frac{1}{(1-p)^2}(n - \sum_{i=1}^{n} d_i)$$

As $d_i \in \{0, 1\}$, n is counting number of d_i, therefore, $n \geq \sum_{i=1}^{n}$. And $p^2, (1-p)^2 > 0$, we can conclude that the second derivative is negative. Hence, the value of $\hat{p}_{MLE}$ is:

$$\hat{p}_{MLE} = \frac{\sum_{i=1}^{n} d_i}{n} = \overline{\delta}$$

Sub Question 1-c

By the asymptotic normality of the MLE, we know that:

- for large sample sizes, the MLE follows an approximate normal distribution;
- the mean of the distribution is equal to the true value of the parameter being estimated;
- the standard deviation of the distribution is approximately equal to the inverse of the Fisher information, which can be estimated using the observed information.

For θ, the expected Fisher information matrix is defined as:

$$I(\theta) = -E[\frac{d^2}{d\theta d\theta^2} \log L(\theta; \mathbb{z})] = \frac{n}{\theta^2}$$

The observed second derivatives evaluated at the MLE is:

$$J(\hat{\theta}_{MLE}) = \frac{n}{\hat{\theta}^2} = \frac{(\sum_{i=1}^{n} \log z_i)^2}{n}$$

As we known from the asymptotic normality of the MLE, we know that:

$$\hat{\theta} \sim N(\theta, J(\hat{\theta}_{MLE})^{-1})$$
$$\hat{\theta} \sim N(\hat{\theta}, \frac{\theta^2}{n})$$

Therefore, a 95% confidence $(\alpha = 0.95)$ interval for θ is given by:

$$[\hat{\theta} - Z_{0+(1-0.95)/2} \times \sqrt{\frac{\theta^2}{n}}, \hat{\theta} + Z_{1-(1-0.95)/2} \times \sqrt{\frac{\theta^2}{n}}]$$

Which can be simplified as:

$$[\hat{\theta} - 1.96\frac{\theta}{\sqrt{n}}, \hat{\theta} + 1.96\frac{\theta}{\sqrt{n}}]$$

For p, the expected Fisher information matrix is defined as:

$$I(p) = -E[\frac{d^2}{dpdp^2}\log L(p; \mathbb{d})] = \frac{1}{p^2}\sum_{i=1}^{n} d_i + \frac{1}{(1-p)^2}(n - \sum_{i=1}^{n} d_i)$$

The observed second derivatives evaluated at the MLE is:

$$\begin{aligned} J(\hat{p}_{MLE}) &= \frac{1}{\hat{p}^2}\sum_{i=1}^{n} d_i + \frac{1}{(1-\hat{p})^2}(n - \sum_{i=1}^{n} d_i) \\ &= \frac{n}{\bar{\delta}} + \frac{n}{1-\bar{\delta}} \\ &= \frac{n}{\bar{\delta}(1-\bar{\delta})} \end{aligned}$$

As we known from the asymptotic normality of the MLE, we know that:

$$\hat{p} \sim N(p, J(\hat{p}_{MLE})^{-1})$$

$$\hat{p} \sim N(\hat{p}, \frac{\bar{\delta}(1-\bar{\delta})}{n})$$

Therefore, a 95% confidence $(\alpha = 0.95)$ interval for θ is given by:

$$[\hat{p} - Z_{0+(1-0.95)/2} \times \sqrt{\frac{\bar{\delta}(1-\bar{\delta})}{n}}, \hat{p} + Z_{1-(1-0.95)/2} \times \sqrt{\frac{\bar{\delta}(1-\bar{\delta})}{n}}]$$

Which can be simplified as:

$$[\bar{\delta} - 1.96\sqrt{\frac{\bar{\delta}(1-\bar{\delta})}{n}}, \bar{\delta} + 1.96\sqrt{\frac{\bar{\delta}(1-\bar{\delta})}{n}}]$$

Question 2

Sub Question 2-a

The likelihood function of the uncensored data is given by:

$$L(\mu, \sigma^2 | y) = \prod_{i=1}^{n} f_Y(y_i; \mu, \sigma^2)$$

where $f_Y(y_i; \mu, \sigma^2)$ is the probability density function of Y.

For the censored observations, we have that $x_i = D$ if $y_i < D$, and $x_i = y_i$ if $y_i \geq D$. The probability of observing $x_i = D$ is equal to the probability of $y_i < D$, which is given by $F_Y(D; \mu, \sigma^2)$. Therefore, the likelihood function for the censored data is:

$$L(\mu, \sigma^2|x, r) = \prod_{i=1}^{n}[f_X(x_i; \mu, \sigma^2)]^{ri} \times [F_Y(D; \mu, \sigma^2)]^{1-ri}$$

where $f_X(x_i; \mu, \sigma^2)$ is the probability density function of X, which is equal to:

$$f_X(x_i; \mu, \sigma^2) = \begin{cases} f_Y(x_i; \mu, \sigma^2) & \text{if } x_i \geq D \\ F_Y(D; \mu, \sigma^2) & \text{if } x_i < D \end{cases}$$

Taking the logarithm of this expression yields the log-likelihood function:

$$\begin{aligned} \log L(\mu, \sigma^2; x, r) &= \sum_{i=1}^{n}\{r_i \log f_X(x_i; \mu, \sigma^2) + (1 - r_i) \log F_Y(D; \mu, \sigma^2)\} \\ &= \sum_{i=1}^{n}\{r_i \log \phi(x_i; \mu, \sigma^2) + (1 - r_i) \log \Phi(x_i; \mu, \sigma^2)\} \end{aligned}$$

Sub Question 2-b

The log likelihood for μ and σ^2 is given by:

$$\begin{aligned} \log L(\mu, \sigma^2; x, r) &= \sum_{i=1}^{n}\{r_i \log \phi(x_i; \mu, \sigma^2) + (1 - r_i) \log \Phi(x_i; \mu, \sigma^2)\} \\ &= \sum_{i=1}^{n}\{r_i \log(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}) + (1 - r_i) \log(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_i} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}})\} \end{aligned}$$

As we already known $\sigma 2 = 1.5^2$, the MLE value for μ is given by:

$$\frac{d}{d\mu}\log L(\mu; x, r) = 0$$

Read data from `dataex2.Rdata` file.

```
load('dataex2.Rdata')
# library(maxLik)
require(maxLik)
```

```
## Loading required package: maxLik
```

```
## Loading required package: miscTools
```

```
##
## Please cite the 'maxLik' package as:
## Henningsen, Arne and Toomet, Ott (2011). maxLik: A package for maximum likelihood
estimation in R. Computational Statistics 26(3), 443-458. DOI 10.1007/s00180-010-0217
-1.
##
## If you have questions, suggestions, or comments regarding the 'maxLik' package, pl
ease use a forum or 'tracker' at maxLik's R-Forge site:
## https://r-forge.r-project.org/projects/maxlik/
```

```
log_like <- function(mu,data,sd=1.5){
  x <- data[1]
  r <- data[2]
  sum(r*log(dnorm(as.matrix(x),mean = mu,sd))
      + (1-r)*log(pnorm(as.matrix(x),mean = mu,sd )))
}
mle <- maxLik(logLik = log_like,data = dataex2, start = c(mu=5))
summary(mle)
```

```
## --------------------------------------------
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 3 iterations
## Return code 1: gradient close to zero (gradtol)
## Log-Likelihood: -336.3821
## 1  free parameters
## Estimates:
##    Estimate Std. error t value Pr(> t)
## mu   5.5328     0.1075   51.46  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## --------------------------------------------
```

The maximum likelihood estimator for μ is 5.5328.

Question 3

Sub Question 3-a

The missing data mechanism can be disregarded for likelihood inference, given that it is MAR (Missing At Random) and only associated with the observed value of Y_1 , while ψ is different from θ. The missing mechanism can be represented as:

$$L(\theta, \psi | \mathbf{y}_{obs}, \mathbf{r}) = f(\mathbf{r} | \mathbf{y}_{obs}, \psi) L(\theta | \mathbf{y}_{obs})$$

Therefore, it is ignoreable for likelihood estimation.

Sub Question 3-b

The missing data mechanism cannot be ignored for likelihood inference, as it is MNAR (Missing Not At Random) and associated with the missing value of Y_2 , while ψ is different from θ. The missing mechanism can be represented as:

$$L(\theta, \psi | \mathbf{y}_{mis}, \mathbf{r}) = f(\mathbf{r} | \mathbf{y}_{obs}, \psi) L(\theta | \mathbf{y}_{obs})$$

Therefore, it is non-ignoreable for likelihood estimation.

Sub Question 3-c

The missing data mechanism cannot be disregarded for likelihood inference, as it is MAR (Missing At Random) and associated only with the observed value of Y_1 , but not distinct from θ. The mechanism can be represented as:

$$L(\theta, \psi | \mathbf{y}_{obs}, \mathbf{r}) = f(\mathbf{r} | \mathbf{y}_{obs}, \psi, \mu_1) L(\theta | \mathbf{y}_{obs})$$

Therefore, it is ignoreable for likelihood estimation.

Question 4

We calculate the log likelihood for β of the complete data first.

As $Y_i \sim Bernoulli\{p_i(\beta)\}$, the density function of Y is:

$$f(Y_i;\beta) = p_i(\beta)^{Y_i}(1-p_i(\beta))^{1-Y_i}$$

with $p_i(\beta) = (exp(\beta_0+\beta_1)) \div (1+exp(\beta_0+\beta_1))$, the density function is:

$$f(Y_i;\beta) = [\frac{exp(\beta_0+x_i\beta_1)}{1+exp(\beta_0+x_i\beta_1)}]^{Y_i} \times [\frac{1}{1+exp(\beta_0+x_i\beta_1)}]^{1-Y_i}$$

the complete likelihood function is:

$$\begin{aligned} L(\beta;Y_i) &= \prod_{i=1}^{n} f(Y_i;\beta) \\ &= \prod_{i=1}^{n} [\frac{exp(\beta_0+x_i\beta_1)}{1+exp(\beta_0+x_i\beta_1)}]^{Y_i} \times [\frac{1}{1+exp(\beta_0+x_i\beta_1)}]^{1-Y_i} \end{aligned}$$

the complete log likelihood function is:

$$\begin{aligned} \log L(\beta;Y_i) &= \prod_{i=1}^{n} \log f(Y_i;\beta) \\ &= \prod_{i=1}^{n} \log\{[\frac{exp(\beta_0+x_i\beta_1)}{1+exp(\beta_0+x_i\beta_1)}]^{Y_i} \times [\frac{1}{1+exp(\beta_0+x_i\beta_1)}]^{1-Y_i}\} \\ &= \sum_{i=1}^{n} Y_i(\beta_0+x_i\beta_1) - \sum_{i=1}^{n} \log(1+exp(\beta_0+x_i\beta_1)) \end{aligned}$$

Then, applying the EM algorithm to estimate the values of β.

For **E-step** at iteration $t+1$, we need to calculate the expectation, with respect to what is missing, Y_i, of the above complete data likelihood, given what is observed x_i and the current estimate of β, that is:

$$\begin{aligned} Q(\beta|\beta^t) &= E_Y[\log L(\beta|Y,x)|Y_{obs},x,\beta^t] \\ &= E_{Y_{m+1..n}}[\log L(\beta|Y,x)|Y_{1..m},x_{1..m},\beta^t] \\ &= E[\sum_{i=1}^{m} Y_i(\beta_0+x_i\beta_1) - \sum_{i=1}^{m} \log(1+exp(\beta_0+x_i\beta_1)) \\ &\quad + \sum_{i=m+1}^{n} Y_i(\beta_0+x_i\beta_1) - \sum_{i=m+1}^{n} \log(1+exp(\beta_0+x_i\beta_1))] \\ \\ &= \sum_{i=1}^{m} Y_i(\beta_0+x_i\beta_1) - \sum_{i=1}^{n} \log(1+exp(\beta_0+x_i\beta_1)) \\ &\quad + \sum_{i=m+1}^{n} \frac{exp(\beta_0^t+x_i\beta_1^t)}{1+exp(\beta_0^t+x_i\beta_1^t)}(\beta_0+x_i\beta_1) \end{aligned}$$

where the first m value of Y is observed and $m+1..n$ value of Y is missing.

For **M-step**, we obtain β^{t+1}, the value of β that maximice $Q(\beta|\beta^t)$.

To calculate β^{t+1}, we need to do first partial derivatives for β_0, β_1 and let them equal to 0, which is:

$$\frac{\partial}{\partial\beta_0}Q = \frac{\partial}{\partial\beta_0}[\sum_{i=1}^{m} Y_i(\beta_0 + x_i\beta_1) - \sum_{i=1}^{n} \log(1 + exp(\beta_0 + x_i\beta_1))$$
$$+ \sum_{i=m+1}^{n} \frac{exp(\beta_0^t + x_i\beta_1^t)}{1 + exp(\beta_0^t + x_i\beta_1^t)}(\beta_0 + x_i\beta_1)]$$
$$= \sum_{i=1}^{m} Y_i - \sum_{i=1}^{n} \frac{exp(\beta_0 + x_i\beta_1)}{1 + exp(\beta_0 + x_i\beta_1)} + \sum_{i=m+1}^{n} \frac{exp(\beta_0^t + x_i\beta_1^t)}{1 + exp(\beta_0^t + x_i\beta_1^t)} = 0$$

$$\frac{\partial}{\partial\beta_1}Q = \frac{\partial}{\partial\beta_1}[\sum_{i=1}^{m} Y_i(\beta_0 + x_i\beta_1) - \sum_{i=1}^{n} \log(1 + exp(\beta_0 + x_i\beta_1))$$
$$+ \sum_{i=m+1}^{n} \frac{exp(\beta_0^t + x_i\beta_1^t)}{1 + exp(\beta_0^t + x_i\beta_1^t)}(\beta_0 + x_i\beta_1)]$$
$$= \sum_{i=1}^{m} Y_i x_i - \sum_{i=1}^{n} \frac{exp(\beta_0 + x_i\beta_1)}{1 + exp(\beta_0 + x_i\beta_1)}x_i + \sum_{i=m+1}^{n} \frac{exp(\beta_0^t + x_i\beta_1^t)}{1 + exp(\beta_0^t + x_i\beta_1^t)}x_i = 0$$

Besides, we need to make sure that the second derivatives are negative.

$$\frac{\partial^2}{\partial\beta_0^2}Q = -\sum_{i=1}^{n} \frac{exp(\beta_0 + x_i\beta_1)}{(1 + exp(\beta_0 + x_i\beta_1))^2} - \sum_{i=m+1}^{n} \frac{exp(\beta_0^t + x_i\beta_1^t)x_i^2}{(1 + exp(\beta_0^t + x_i\beta_1^t))^2}$$

$$\frac{\partial^2}{\partial\beta_1^2}Q = -\sum_{i=1}^{n} \frac{x_i^2 exp(\beta_0 + x_i\beta_1)}{(1 + exp(\beta_0 + x_i\beta_1))^2} - \sum_{i=m+1}^{n} \frac{exp(\beta_0^t + x_i\beta_1^t)x_i^2}{(1 + exp(\beta_0^t + x_i\beta_1^t))^2}$$

Therefore, we can obtain the values of $\beta_0^{t+1}, \beta_1^{t+1}$ from the first derivatives when they equal to 0.

Then, repeating M-step and E step until some convergence criterion is met. Here we set criterion like:

$$\left|\beta_0^{t+1} - \beta_0^t\right| + \left|\beta_1^{t+1} - \beta_1^t\right| \leq 0.0001$$

Then, we try to encode EM algorithm and calculate the values of β .

```
require(maxLik)
log_like_ex4 <- function(beta,beta_t,data){
  complete <- data[complete.cases(data), ]
  missing <- data[!complete.cases(data), ]

  beta0 <- beta[1];beta1 <- beta[2]
  beta0_t <- beta_t[1];beta1_t <- beta_t[2]

  x <- data[,1]; y <- data[,2]
  x_1m <- complete[,1]; y_1m <- complete[,2]
  x_mn <- missing[,1]; y_mn <- missing[,2]

  sum(y_1m*(beta0+x_1m*beta1)) - sum(log(1+exp(beta0+beta1*x)))+
    sum(exp(beta0_t+beta1_t*x_mn)/(1+exp(beta0_t+beta1_t*x_mn))*
          (beta0+x_mn*beta1))
}
```

```
# workshop4
load('dataex4.Rdata')
EMex4 <- function(beta_ini, data_EM, eps){
diff <- 1
beta <- beta_ini
i <- 0
while(diff > eps){
beta.old <- beta
mle_beta <- maxLik(logLik = log_like_ex4, data =
                     data_EM, beta_t = beta.old,start = beta_ini)
beta <- mle_beta$estimate
diff <- sum(abs(beta-beta.old))
i <- i+1
}
return(beta)
}
beta= EMex4(beta_ini = c(0, 0), data_EM = dataex4,0.0001)
beta
```

```
## [1]  0.9755477 -2.4802557
```

The result for β_0 is 0.9755, for β_1 is -2.4803.

Question 5

Sub Question 5-a

We calculate the log likelihood for θ of the complete data first. According to the question, CDF of F_Y is:

$$
\begin{aligned}
F(y;\lambda,\mu,p) &= p \times F_X(y;\lambda) + (1-p) \times F_Y(y;\mu) \\
&= p \times (1-y^{-\lambda}) + (1-p) \times (1-y^{-\mu}) \\
&= 1 - p \times y^{-\lambda} - y^{-u} + p \times y^{-\mu}
\end{aligned}
$$

with $y \geq 1$ and $\lambda, \mu > 0$.

Do the first derivatives to calculate PDF of Y, which is:

$$
\begin{aligned}
f_Y(y;\lambda,\mu,p) &= \frac{\partial}{\partial y} F(y;\lambda,\mu,p) \\
&= p \times y^{-(\lambda+1)} + y^{-(\mu+1)} - p \times y^{-(\mu+1)} \\
&= p \times \lambda \times y^{-(\lambda+1)} + (1-p) \times \mu \times y^{-(\mu+1)}
\end{aligned}
$$

Define Z to be the indicator of Y, the rule is:

$$
Z_i = \begin{cases} 1, & \text{if } Y_i \text{ is from } F_X \\ 0, & \text{if } Y_i \text{ is from } F_Y \end{cases} \quad f_Z = \begin{cases} p, & Z_i = 1 \\ 1-p, & Z_i = 0 \end{cases}
$$

The PDF of Y and Z can be written as:

$$
f(y,z;\lambda,\mu,p) = [p \times y^{-(\lambda+1)}]^{z_i} + [y-(\mu+1) - p \times y-(\mu+1)]^{1-z_i}
$$

The likelihood function can be written as:

$$
\begin{aligned}
L(\theta; \mathbf{y}, \mathbf{z}) &= \prod_{i=1}^{n} f(y, z; \lambda, \mu, p) \\
&= \prod_{i=1}^{n} [p \times \lambda \times y^{-(\lambda+1)}]^{z_i} \times [(1-p) \times \mu \times y^{-(\mu+1)}]^{(1-z_i)}
\end{aligned}
$$

The log likelihood function is:

$$
\log L = \sum_{i=1}^{n} z_i \log[p \times \lambda \times y^{-(\lambda+1)}] + \sum_{i=1}^{n} (1 - z_i) \log[(1-p) \times \mu \times y^{-(\mu+1)}]
$$

Then, applying the EM algorithm to estimate the values of θ.

For **E-step** at iteration $t+1$, we need to calculate the expectation, with respect to what is from F_X or F_Y of the above complete data likelihood, given what is observed y_i and the current estimate of θ, that is:

$$
\begin{aligned}
Q(\theta|\theta^t) &= E_z[\log L(\theta|\mathbf{y}, z)|\mathbf{y}, \theta^t] \\
&= \sum_{i=1}^{n} E[z_i|\mathbf{y}, \theta^t] \times \log(p \times \lambda \times y^{-(\lambda+1)}) \\
&\quad + \sum_{i=1}^{n} (1 - E[z_i|\mathbf{y}, \theta^t]) \times \log((1-p) \times \mu \times y^{-(\mu+1)})
\end{aligned}
$$

with expectation of Z, which is $E[z_i|\mathbf{y}, \theta^t]$, can be calculated as:

$$
\begin{aligned}
E[z_i|\mathbf{y}, \theta^t] &= 1 \times p_Z(Z=1) + 0 \times (1 - p_Z(Z=1)) \\
&= \frac{p^t \times \lambda^t \times y^{-(\lambda^t+1)}}{p^t \times \lambda^t \times y^{-(\lambda^t+1)} + (1-p^t) \times \mu^t \times y^{-(\mu^t+1)}}
\end{aligned}
$$

Using $Q(\theta|\theta^t)$ and $E[z_i|\mathrm{y}, \theta^t]$, we can complete our calculation in E step.

For **M-step**, we obtain θ^{t+1}, the value of θ that maximise $Q(\theta|\theta^t)$.

To calculate θ^{t+1}, we need to do first partial derivatives for λ, μ, p and let them equal to 0, which is:

$$
\begin{aligned}
\frac{\partial}{\partial p} Q(\theta|\theta^t) &= \frac{\sum_{i=1}^{n} E[z_i|\mathbf{y}, \theta^t]}{p} && = 0 \\
\frac{\partial}{\partial \lambda} Q(\theta|\theta^t) &= \frac{\sum_{i=1}^{n} E[z_i|\mathbf{y}, \theta^t]}{\lambda} - \sum_{i=1}^{n} E[z_i|\mathbf{y}, \theta^t] \times log(y_i) && = 0 \\
\frac{\partial}{\partial \mu} Q(\theta|\theta^t) &= \frac{\sum_{i=1}^{n} (1 - E[z_i|\mathbf{y}, \theta^t])}{\mu} - \sum_{i=1}^{n} (1 - E[z_i|\mathbf{y}, \theta^t]) \times \log(y_i) && = 0
\end{aligned}
$$

Check the second derivatives:

$$
\begin{aligned}
\frac{\partial^2}{\partial p \partial p^T} Q(\theta|\theta^t) &= -\frac{\sum_{i=1}^{n} E[z_i|\mathbf{y}, \theta^t]}{p^2} \\
\frac{\partial^2}{\partial \lambda \partial \lambda^T} Q(\theta|\theta^t) &= -\frac{\sum_{i=1}^{n} E[z_i|\mathbf{y}, \theta^t]}{\lambda^2} \\
\frac{\partial^2}{\partial \mu \partial \mu^T} Q(\theta|\theta^t) &= -\frac{\sum_{i=1}^{n} (1 - E[z_i|\mathbf{y}, \theta^t])}{\mu^2}
\end{aligned}
$$

The second derivatives are all negative, therefore, we can obtain $p^{t+1}, \lambda^{t+1}, \mu^{t+1}$, which are:

$$
\begin{aligned}
p^{t+1} &= \frac{\sum_{i=1}^{n} E[z_i|\mathbf{y},\theta^t]}{n} = E[z_i|\mathbf{y},\theta^t] \\
\lambda^{t+1} &= \frac{\sum_{i=1}^{n} E[z_i|\mathbf{y},\theta^t]}{\sum_{i=1}^{n} E[z_i|\mathbf{y},\theta^t] \times \log(y_i)} \\
\mu^{t+1} &= \frac{\sum_{i=1}^{n} (1 - E[z_i|\mathbf{y},\theta^t])}{\sum_{i=1}^{n} (1 - E[z_i|\mathbf{y},\theta^t]) \times \log(y_i)}
\end{aligned}
$$

Then, repeating M-step and E step until some convergence criterion is met. Here we set criterion like:

$$
\left|\lambda^{t+1} - \lambda^t\right| + \left|p^{t+1} - p^t\right| + \left|\mu^{t+1} - \mu^t\right| \le 0.0001
$$

Sub Question 5-b

In this part, try to encode EM algorithm and calculate the values of θ .

```r
rm(list = ls())

EMex5 <- function(ini_theta,data,eps){
  diff <- 1
  y <- data
  theta <- ini_theta
  p <- theta[1]; lambda <- theta[2]; mu <- theta[3]
  n <- length(y)

  while(diff > eps){
    theta.old <- theta
    #E-step
    p1 <- p*lambda*y**(-lambda-1)
    p2 <- (1-p)*mu*y**(-mu-1)
    p_hat <- p1/(p1+p2)
    #M-step
    p <- sum(p_hat)/n
    lambda <- sum(p_hat)/sum(p_hat*log(y))
    mu <- sum(1-p_hat)/sum((1-p_hat)*log(y))
    #theta(t+1)
    theta <- c(p, lambda, mu)
    diff <- sum(abs(theta-theta.old))
  }
  return(theta)
}
load("dataex5.Rdata")

theta <- EMex5(ini_theta = c(0.3,0.3,0.4),data = dataex5, 0.0001)

p <- theta[1]
lambda <- theta[2]
mu<- theta[3]

#estimated p; estimated lambda; estimated mu
theta
```

```
## [1] 0.7939337 0.9762783 6.6705994
```

The estimate values from EM algorithm is $p = 0.7939, \lambda = 0.9763, \mu = 0.6706.$

Then, drawing the histogram of the data with the estimated density superimposed.

```
#function to calculate density function of pareto distribution
dpareto <- function (x, k){
  f <- k*x**(-k-1)
  return(f)
}
load("dataex5.Rdata")
y <- dataex5

hist(y, breaks = "FD", main = "Histogram with Estimated Density",
     xlab = "dataex5",
     ylab = "Density",
     cex.main = 1.5, cex.lab = 1.5, cex.axis = 1.4,
     freq = F, ylim = c(0,1.5), xlim = c(0,15))
#add the estimated density curve
x <- y
curve(p*dpareto(x,lambda)+(1-p)*dpareto(x,mu),
      add = TRUE, lwd = 2, col = "red")
```