1 Question 1

(a). Because $x, y \geq 1$, $\lambda, \mu > 0$ and $Z = min\{X, Y\}$. According to $F_X(x;\lambda)$ and $F_Y(y;\mu)$ we can get F_Z that for any $z \geq 1$. Then Z is Pareto-distributed with parameters λ and μ. We can find the density function of Z:

$$\begin{aligned}
F_Z(z) &= P(Z \leq z; \lambda, \mu) \\
&= P(min\{X, Y\} \leq z) \\
&= 1 - P(min\{X, Y\} > z) \\
&= 1 - P(X > z \quad and \quad Y > z) \\
&= 1 - P(X > z; \lambda)\ P(Y > z; \mu) \\
&= 1 - (1 - P(X \leq z))(1 - P(Y \leq z)) \\
&= 1 - (1 - F_X(z))(1 - F_Y(z)) \\
&= 1 - \left(1 - \left(1 - \frac{1}{z^\lambda}\right)\right)\left(1 - \left(1 - \frac{1}{z^\mu}\right)\right) \\
&= 1 - \frac{1}{z^{\lambda+\mu}}
\end{aligned} \tag{1}$$

we then differentiate $F_Z(z)$ to obtain the density function of Z as:

$$\begin{aligned}
f_Z(z; \lambda, \mu) &= \frac{\partial}{\partial z}\left(1 - \frac{1}{z^{\lambda+\mu}}\right) \\
&= (\lambda + \mu) z^{-\lambda-\mu-1}
\end{aligned} \tag{2}$$

From the given cumulative distributions of X and Y, we can find the density functions of X and Y:

$$f_X(x) = \lambda x^{-\lambda-1} \tag{3}$$

$$f_Y(y) = \mu x^{-\mu-1} \tag{4}$$

We then need to find the frequency function of $\delta(f_\delta)$:

$$\begin{aligned}
P(\delta = 1) &= P(X < Y) \\
&= \iint_{X<Y} f_X f_Y \, dy \, dx \\
&= \int_1^\infty \int_x^\infty \lambda x^{-\lambda-1} \mu y^{-\mu-1} \, dy \, dx \\
&= \frac{\lambda}{\lambda + \mu}
\end{aligned} \tag{5}$$

thus, the frequency function is:

$$\begin{aligned} S(1) &= P(\delta = 1) \\ &= \frac{\lambda}{\lambda + \mu} \end{aligned} \tag{6}$$

$$\begin{aligned} S(0) &= P(\delta = 0) \\ &= 1 - P(\delta = 1) \\ &= \frac{\mu}{\lambda + \mu} \end{aligned} \tag{7}$$

After the above analysis we can consider the distribution of Z to be a Pareto distribution with parameters λ and μ, and the distribution of δ to be a Bernoulli distribution with parameter $p = \frac{\lambda}{\lambda+\mu}$.

(b). From (a) we get $f_Z(z; \lambda, \mu) = (\lambda + \mu) z^{-\lambda-\mu-1}$. Now, let $Z_1 \ldots Z_n$ be random sample from $f_Z(z; \theta)$, the log likelihood is thus of the form

$$\begin{aligned} \log L(\theta; z, d) &= \log \prod_{i=1}^{n} f_Z(z; \lambda, \mu) \\ &= \log \prod_{i=1}^{n} f_Z(z; \theta) \\ &= \log \prod_{i=1}^{n} \theta z_i^{-\theta-1} \\ &= \log\, [\theta^n \prod_{i=1}^{n} z_i^{-\theta-1}] \\ &= n \log(\theta) + (-\theta - 1) \sum_{i=1}^{n} \log(z_i) \\ &= n \log(\theta) - \theta \sum_{i=1}^{n} \log(z_i) - \sum_{i=1}^{n} \log(z_i) \end{aligned} \tag{8}$$

The score function of θ is given by

$$\begin{aligned} U(\theta) &= \frac{\partial}{\partial \theta} \left(n \log(\theta) - \theta \sum_{i=1}^{n} \log(z_i) - \sum_{i=1}^{n} \log(z_i) \right) \\ &= \frac{n}{\theta} - \sum_{i=1}^{n} \log(z_i) \end{aligned} \tag{9}$$

we then have

$$
\begin{aligned}
U(\theta) &= \frac{\partial}{\partial\theta}\left(n\log(\theta) - \theta\sum_{i=1}^{n}\log(z_i) - \sum_{i=1}^{n}\log(z_i)\right) \\
&= \frac{n}{\theta} - \sum_{i=1}^{n}\log(z_i)
\end{aligned} \tag{10}
$$

$$
U(\hat{\theta}) = 0 \tag{11}
$$

$$
\hat{\theta} = \frac{n}{\sum_{i=1}^{n}\log z_i} \tag{12}
$$

Let $\delta_1 \ldots \delta_n$ be random sample from $f_\delta(d;p)$, with $p = \frac{\lambda}{\lambda+\mu}$, the log likelihood is thus of the form:

$$
\begin{aligned}
\log L(p;z,d) &= \log\prod_{i=1}^{n} f_\delta(d;\lambda,\mu) \\
&= \log\prod_{i=1}^{n}\left(\frac{\lambda}{\lambda+\mu}\right)^{d_i}\left(\frac{\mu}{\lambda+\mu}\right)^{1-d_i} \\
&= \log\prod_{i=1}^{n} p^{d_i}(1-p)^{1-d_i} \\
&= \log\prod_{i=1}^{n} f_\delta(d;p) \\
&= \sum_{i=1}^{n} d_i\log p + \sum_{i=1}^{n}(1-d_i)\log(1-p)
\end{aligned} \tag{13}
$$

The score function of p is given by

$$
\begin{aligned}
U(p) &= \frac{\partial}{\partial\theta}\left(\sum_{i=1}^{n} d_i\log p + \sum_{i=1}^{n}(1-d_i)\log(1-p)\right) \\
&= \frac{\sum_{i=1}^{n} d_i}{p} - \frac{\sum_{i=1}^{n}(1-d_i)}{1-p}
\end{aligned} \tag{14}
$$

$$
U(\hat{p}) = 0 \tag{15}
$$

$$
\begin{aligned}
\hat{p} &= \frac{\sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} d_i + \sum_{i=1}^{n}(1-d_i)} \\
&= \frac{\sum_{i=1}^{n} d_i}{n}
\end{aligned} \tag{16}
$$

(c). Under certain conditions, the maximum likelihood estimate approximates a normal distribution with a mean equal to the true value of the parameter and a variance equal to the inverse of the Fisher information.

$$\begin{aligned} I(\theta) &= -E\left[\left(\frac{\partial^2}{\partial\theta^2}\log L(\theta, p; z, d)\right)\right] \\ &= \frac{n}{\theta^2} \end{aligned} \tag{17}$$

Thus, the asymptotic distribution of $\hat{\theta}$ is $N\left(\theta,\ \frac{\theta^2}{n}\right)$. The Fisher information $I(p)$ is:

$$\begin{aligned} I(p) &= -E\left[\left(\frac{\partial^2}{\partial p^2}\log L(\theta, p; z, d)\right)\right] \\ &= \frac{n}{p(1-p)} \end{aligned} \tag{18}$$

Thus, the asymptotic distribution of $\hat{p}$ is $N\left(p,\ p\frac{1-p}{n}\right)$.

For the 95% confidence interval, we used the standard normal distribution and the asymptotic normality of the maximum likelihood estimator, for θ obtained:

$$\left[\hat{\theta} - 1.96 \times \sqrt{\frac{\hat{\theta}^2}{n}},\ \hat{\theta} + 1.96 \times \sqrt{\frac{\hat{\theta}^2}{n}}\right] \tag{19}$$

For the 95% confidence interval, we used the standard normal distribution and the asymptotic normality of the maximum likelihood estimator, for p obtained:

$$\left[\hat{p} - 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + 1.96 \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \tag{20}$$

These confidence intervals are only approximations and so may be inaccurate for small sample sizes or when regularity conditions are not met.

2 Question 2

(a). According to the conditions given in the question, we have $\phi(x_i; \mu, \sigma^2)$ from noncensored observation and $P(X < D; \mu, \sigma^2) = \Phi(x_i; \mu, \sigma^2)$ from censored observation. Thus the log likelihood of the observed data is:

$$\begin{aligned}
\log L(\mu, \sigma^2; x, r) &= \log \prod_{i=1}^{n} \{(\phi(x_i; \mu, \sigma^2))^{r_i} (\Phi(x_i; \mu, \sigma^2))^{1-r_i}\} \\
&= \log\{(\phi(x_i; \mu, \sigma^2))^{\sum_{i=1}^{n} r_i} (\Phi(x_i; \mu, \sigma^2))^{\sum_{i=1}^{n}(1-r_i)}\} \qquad (21) \\
&= \sum_{i=1}^{n} \{r_i \log(\phi(x_i; \mu, \sigma^2)) + (1 - r_i)(\Phi(x_i; \mu, \sigma^2))\}
\end{aligned}$$

(b). See the code in GitHub. The maximum likelihood estimate of μ based on the data available in the file dataex2.Rdata is 5.5328 under $\sigma^2 = 1.5^2$.

3 Question 3

The missing data mechanism is negligible for likelihood inference in the following cases:

- the missing data is MAR or MCAR
- the parameters ϕ and θ are disjoint

(a). $\text{logit}\{P(R = 0 \mid y_1, y_2, \theta, \psi)\} = \psi_0 + \psi_1 y_1, \ \psi = (\psi_0, \psi_1)$, the missingness only depends on Y_1 which is fully observed. The missing data mechanism is MAR. $\psi = (\psi_0, \psi_1)$ is distinct from θ. So, it is ignorable for likelihood-based estimation.

(b). $\text{logit}\{P(R = 0 \mid y_1, y_2, \theta, \psi)\} = \psi_0 + \psi_1 y_2, \ \psi = (\psi_0, \psi_1)$, the missingness only depends on Y_2 which has some missing values. The missing data mechanism is MNAR. For likelihood-based estimation, it is not ignorable, as missing data can be ignorable, while it does not depend on itself.

(c). $\text{logit}\{P(R = 0 \mid y_1, y_2, \theta, \psi)\} = 0.5(\mu_1 + \psi y_1)$, the missingness only depends on Y_1 which is fully observed and a constant. The missing data mechanism is MAR. The missingness also depends on μ_1. Although the scalar ϕ is different from θ, μ_1 is included in the parameter space of the models ψ and θ. Therefore, it is ignorable for likelihood-based estimation.

4 Question 4

The log likelihood of β is

$$
\begin{aligned}
\log L(\beta) &= \log \prod_{i=1}^{n} p_i\beta^{y_i}(1-p_i\beta)^{1-y_i} \\
&= \log \prod_{i=1}^{n} \left[\left(\frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}} \right)^{y_i} + \left(\frac{1}{1+e^{\beta_0+x_i\beta_1}} \right)^{1-y_i} \right] \\
&= \sum_{i=1}^{n} \left[y_i \log \left(\frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}} \right) + (1-y_i) \log \left(\frac{1}{1+e^{\beta_0+x_i\beta_1}} \right) \right] \\
&= \sum_{i=1}^{n} \left[y_i(\beta_0+x_i\beta_1) - \log(1+e^{\beta_0+x_i\beta_1}) \right]
\end{aligned} \tag{22}
$$

Now we proceed to E-step. We can assuming first m values of Y are observed and remaining $n-m$ are missing. Since $Y \overset{\text{i.i.d}}{\sim} \text{Bern}[p_i(\beta)]$ and $E_{Y_i} = p_i(\beta) = \frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}}$.

$$
\begin{aligned}
Q[\beta|\ \beta^{(t)}] &= E_{Y_{mis}}[\log L(\beta|y_{obs}, y_{mis})|\ y_{obs},\ x,\ \beta^{(t)}] \\
&= \sum_{i=1}^{m} [y_i(\beta_0+x_i\beta_1)] + \sum_{i=1}^{n} \log(1+e^{\beta_0+x_i\beta_1}) \\
&+ \sum_{i=m+1}^{n} [y_i(\beta_0+x_i\beta_1)]\ E_{Y_{mis}}[y_i|\ y_{obs},\ x,\ \beta^{(t)}] \\
&= \sum_{i=1}^{m} y_i(\beta_0+x_i\beta_1) - \sum_{i=1}^{n} \log(1+e^{\beta_0+x_i\beta_1}) + \sum_{i=1}^{m} (\beta_0+x_i\beta_1)p_i(\beta)
\end{aligned} \tag{23}
$$

Code in GitHub and we get result(β): Estimate(s): 0.9755261 -2.480384.

```
Maximum Likelihood estimation
Newton-Raphson maximisation, 5 iterations
Return code 1: gradient close to zero (gradtol)
Log-Likelihood: -185.7986 (2 free parameter(s))
Estimate(s): 0.9755261 -2.480384
```

5 Question 5

(a). First we convert the CDF to its PDF based on the given $F_X(x;\lambda)$ and $F_Y(y;\mu)$: $f_X(x;\lambda) = \lambda x^{-\lambda-1}$ and $f_Y(y;\mu) = \mu y^{-\mu-1}$. We define an augmented complete dataset where $\mathbf{y}_{\text{obs}} = (y_1, \ldots, y_n)$ and $\mathbf{y}_{\text{mis}} = (z_1, \ldots, z_n)$ is a vector of unobserved/latent group data indicator, such that

$$z_i = \begin{cases} 1, & \text{if } y_i \text{ belongs to the first component (short waiting times)} \\ 0 & \text{if } y_i \text{ belongs to the second component (long waiting times).} \end{cases} \tag{24}$$

Note that $Z \overset{\text{i.i.d}}{\sim} \text{Bern}(p)$ and then the complete data likelihood is

$$L(\theta \mid y, z) = \prod_{i=1}^{n} \left\{ [p f_X(x;\lambda)]^{z_i} \left[(1-p) f_Y(y;\mu)\right]^{1-z_i} \right\} \tag{25}$$

Therefore,

$$\begin{aligned} \log L(\theta \mid y, z) &= \sum_{i=1}^{n} z_i \left\{ \log p + \log f_X(x;\lambda) \right\} + \sum_{i=1}^{n} (1-z_i) \left\{ \log(1-p) + \log f_Y(y;\mu) \right\} \\ &= \sum_{i=1}^{n} z_i \left\{ \log p + \log(\lambda x^{-\lambda-1}) \right\} + \sum_{i=1}^{n} (1-z_i) \left\{ \log(1-p) + \log(\mu y^{-\mu-1}) \right\} \end{aligned} \tag{26}$$

For the E-step we should need to compute:

$$\begin{aligned} Q\left(\theta \mid \theta^{(t)}\right) &= E_Z \left[\log L(\theta \mid y, z) \mid y, \theta^{(t)} \right] \\ &= \sum_{i=1}^{n} E\left[Z_i \mid y, \theta^{(t)}\right] \left\{ \log p + \log(f_X(x;\lambda)) \right\} \\ &+ \sum_{i=1}^{n} \left(1 - E\left[Z_i \mid y, \theta^{(t)}\right]\right) \left\{ \log(1-p) + \log(f_Y(y;\mu)) \right\} \\ &= \sum_{i=1}^{n} E\left[Z_i \mid y, \theta^{(t)}\right] \left\{ \log p + \log(\lambda x^{-\lambda-1}) \right\} \\ &+ \sum_{i=1}^{n} \left(1 - E\left[Z_i \mid y, \theta^{(t)}\right]\right) \left\{ \log(1-p) + \log(\mu y^{-\mu-1}) \right\} \end{aligned} \tag{27}$$

Now,

$$
\begin{aligned}
E\left[Z_i \mid y, \theta^{(t)}\right] &= E\left[Z_i \mid y_i, \theta^{(t)}\right] \\
&= 1 \times \operatorname{Pr}\left(Z_i = 1 \mid y_i, \theta^{(t)}\right) + 0 \times \operatorname{Pr}\left(Z_i = 0 \mid y_i, \theta^{(t)}\right) \\
&= \frac{p^{(t)} f_X(x; \lambda)}{p^{(t)} f_X(x; \lambda) + \left(1 - p^{(t)}\right) f_Y(y; \mu)} \\
&= \frac{p^{(t)}(\lambda x^{-\lambda-1})}{p^{(t)}(\lambda x^{-\lambda-1}) + \left(1 - p^{(t)}\right)\left(\mu y^{-\mu-1}\right)} \\
&= \widetilde{p}_i^{(t)}
\end{aligned}
\tag{28}
$$

Thus,

$$
\begin{aligned}
Q\left(\theta \mid \theta^{(t)}\right) &= \sum_{i=1}^{n} \tilde{p}_i^{(t)} \left\{\log p + \log f_X(x; \lambda)\right\} + \sum_{i=1}^{n} \left(1 - \tilde{p}_i^{(t)}\right) \left\{\log(1-p) + \log f_Y(y; \mu)\right\} \\
&= \sum_{i=1}^{n} \tilde{p}_i^{(t)} \left\{\log p + \log(\lambda x^{-\lambda-1})\right\} + \sum_{i=1}^{n} \left(1 - \tilde{p}_i^{(t)}\right) \left\{\log(1-p) + \log(\mu y^{-\mu-1})\right\}
\end{aligned}
\tag{29}
$$

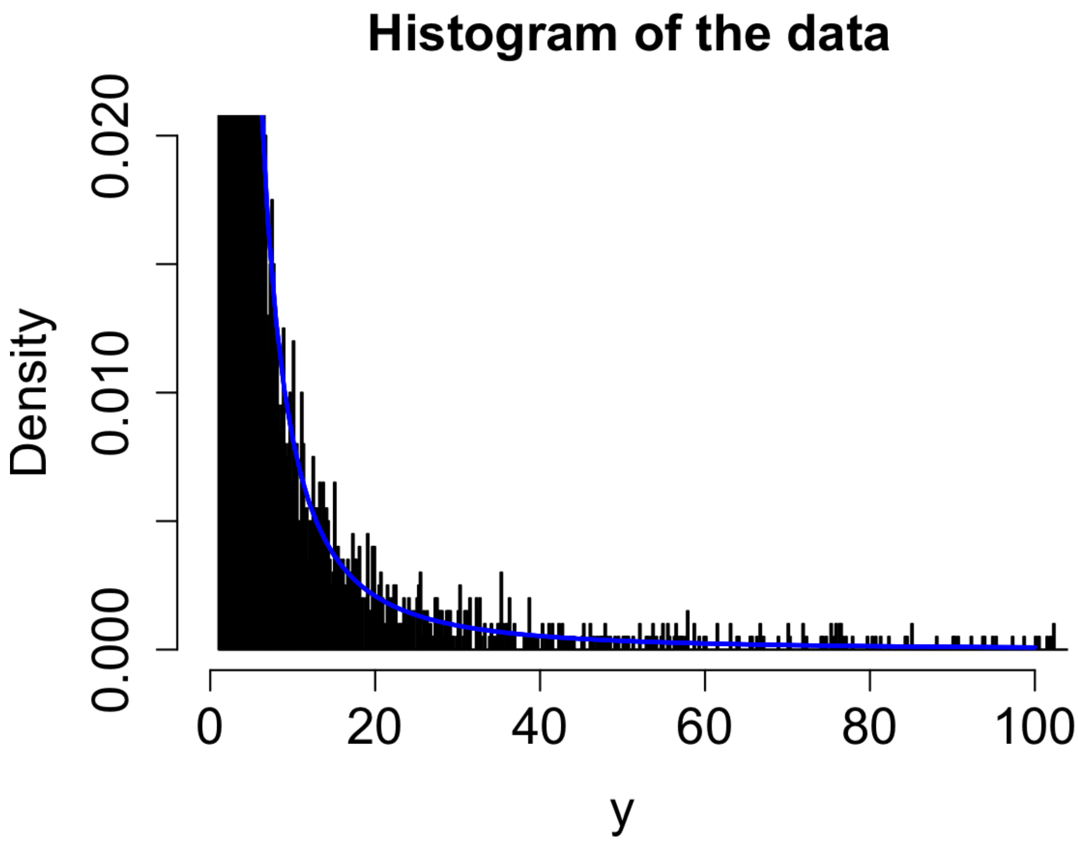
For the M-step,

$$
\frac{\partial}{\partial p} Q\left(\theta \mid \theta^{(t)}\right) = 0 \Rightarrow p^{(t+1)} = \frac{\sum_{i=1}^{n} \tilde{p}_i^{(t)}}{n} \tag{30}
$$

$$
\frac{\partial}{\partial \lambda} Q\left(\theta \mid \theta^{(t)}\right) = 0 \Rightarrow \lambda^{(t+1)} = \frac{\sum_{i=1}^{n} \tilde{p}_i^{(t)}}{\sum_{i=1}^{n} \tilde{p}_i^{(t)} \log(y)} \tag{31}
$$

$$
\frac{\partial}{\partial \mu} Q\left(\theta \mid \theta^{(t)}\right) = 0 \Rightarrow \mu^{(t+1)} = \frac{\sum_{i=1}^{n} \left(1 - \tilde{p}_i^{(t)}\right)}{\sum_{i=1}^{n} \left(1 - \tilde{p}_i^{(t)}\right) \log(y)} \tag{32}
$$

(b). See the code and get result in GitHub.
As can be seen, $\hat{p} = 0.7939337$, $\hat{\lambda} = 0.9762783$ and $\hat{\mu} = 6.670599$, we also draw the histogram of the data with the estimated density superimposed:



Histogram of the data
Density
0.000
0.010
0.020
0
20
40
60
80
100
y