Assignment 2

Zekai Li S2040608

Question 1

(a) **My solution**

Rewrite the observations in the form of $X_i^2 = Y_i^2 I(Y_i \leq C) + C^2 I(Y_i > C) = Y_i^2 R_i + C^2(1 - R_i)$. And contribution of a non censored observation to the likelihood is $P(Y_i > C) = 1 - F(C) = e^{-\frac{C^2}{2\theta}}$.

Therefore, the likelihood of the observed data is,

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \{f(y_i;\theta)^{R_i}[P(Y_i > C)]^{1-R_i}\} \\
&= \frac{\prod_{i=1}^{m} y_i}{\theta^m} e^{-\sum_{i=1}^{m} \frac{y_i^2}{2\theta}} e^{-\frac{(n-m)C^2}{2\theta}}
\end{aligned}
$$

where $m = \sum_{i=1}^{n} R_i$, so the formula of the log-likelihood equal to 0 is:

$$
\begin{aligned}
-\frac{m}{\theta} + \sum_{i=1}^{m} \frac{y_i^2}{2\theta^2} + \frac{(n-m)C^2}{2\theta^2} &= 0 \\
\sum_{i=1}^{n} \frac{R_i^2 y_i^2 + (1 - R_i)C^2}{2\theta^2} &= \sum_{i=1}^{n} R_i \\
\sum_{i=1}^{n} \frac{X_i^2}{2\theta^2} &= \sum_{i=1}^{n} R_i
\end{aligned}
$$

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

(b) **My solution**

From (a), we know the first gradient of the liklihood is $-\frac{\sum_{i=1}^{n} R_i}{\theta} + \sum_{i=1}^{n} \frac{X_i^2}{2\theta^2}$. Therefore, the fisher information would be:

$$
\begin{aligned}
\mathcal{I}(\theta) &= -\mathbb{E}\left[\frac{\sum_{i=1}^{n} R_i}{\theta^2} - \sum_{i=1}^{n} \frac{X_i^2}{\theta^3}\right] \\
&= \sum_{i=1}^{n} \frac{\mathbb{E}[X_i^2]}{\theta^3} - \sum_{i=1}^{n} \frac{\mathbb{E}[R_i]}{\theta^2} \\
&= \sum_{i=1}^{n} \frac{\left(\int_0^C y^2 f(y;\theta)\, dy + C^2(1 - F(C))\right)}{\theta^3} - \sum_{i=1}^{n} \frac{F(C)}{\theta^2} \\
&= \frac{2n}{\theta^2}(1 - e^{-\frac{C^2}{2\theta}}) - \frac{n}{\theta^2}(1 - e^{-\frac{C^2}{2\theta}}) \\
&= \frac{n}{\theta^2}(1 - e^{-\frac{C^2}{2\theta}})
\end{aligned}
$$

(c) **My solution**

From the asymptotic normality of the maximum likelihood estimator, we know that $\hat{\theta} \sim \mathcal{N}(\theta, \frac{1}{n\mathcal{J}(\theta)}) = \mathcal{N}(\theta, \frac{\theta^2}{1-e^{-C^2/2\theta}})$. And the 95% confidence interval of normalized normal distribution z is $[-1.96, 1.96]$. Therefore, the interval of $\hat{\theta}$ is $[\theta - \frac{1.96\theta}{\sqrt{1-e^{-C^2/2\theta}}}, \theta + \frac{1.96\theta}{\sqrt{1-e^{-C^2/2\theta}}}]$.

Question 2

(a) **My solution**

the contribution of a non censored observation to the likelihood is $P(Y < D|\mu, \sigma^2) = \Phi(D; \mu, \sigma^2)$.

Therefore, the likelihood of the observed data is,

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

Therefore, the log likelihood of the observed data is given by:

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

(b) **My solution**

```r
load(file = "dataex2.Rdata")

# log likelihood of dataex2
log_like_dataex2 <- function(mean){
  X <- dataex2[[1]]; R <- dataex2[[2]]
  sum(R*dnorm(X,mean = mean,sd = 1.5, log = TRUE) + (1-R)*pnorm(X,mean=mean,sd=1.5,log=TRUE))
}

mle <- maxLik(logLik = log_like_dataex2, start = c(15))
summary(mle)
```

```
## --------------------------------------------
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 3 iterations
## Return code 1: gradient close to zero
## 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
## --------------------------------------------
```

Therefore, the maximum likelihood estimate of μ is 5.5328.

Question 3

(a) **My solution**

Since $\text{logit}\{Pr(R = 0|y1, y2, \theta, \psi)\} = \psi_0 + \psi_1 y_1$, the missing mechanism is MAR and $\psi = (\psi_0, \psi_1)$ distinct from θ. Therefore, the missing indicator can be ignored for likelihood estimation.

(b) **My solution**

Since $\text{logit}\{Pr(R = 0|y1, y2, \theta, \psi)\} = \psi_0 + \psi_1 y_2$, the missing mechanism is MNAR. Therefore, the missing indicator cannot be ignored for likelihood estimation.

(c) **My solution** Since $\text{logit}\{Pr(R = 0|y1, y2, \theta, \psi)\} = 0.5(\mu_1 + \psi y_1)$, the missing mechanism is MAR. However, (μ_1, ψ) are not distinct from θ. Therefore, the missing indicator cannot be ignored for likelihood estimation.

Question 4

The log likelihood of complete data is:

$$\log L(\beta|\mathbf{y_{obs}}, \mathbf{y_{mis}}) = \sum_{i=1}^{m} [y_i \log(P_i(\beta)) + (1-y_i)\log(1-P_i(\beta))] + \sum_{i=m+1}^{n} [y_i \log(P_i(\beta)) + (1-y_i)\log(1-P_i(\beta))]$$

At iteration $t+1$, the E step is given by:

$$Q(\beta|\beta^{(t)}) = \sum_{i=1}^{m} \left[E_{\mathbf{Y_{mis}}}(y_i|\beta^{(t)})\log(P_i(\beta)) + (1 - E_{\mathbf{Y_{mis}}}(y_i|\beta^{(t)}))\log(1-P_i(\beta))\right] + \sum_{i=m+1}^{n} [y_i \log(P_i(\beta)) + (1-y_i)\log(1-P_i(\beta))]$$

And we have:

$$E_{\mathbf{Y_{mis}}}(y_i|\beta^{(t)}) = P_i(\beta^{(t)})$$

Therefore,

$$Q(\beta|\beta^{(t)}) = \sum_{i=1}^{m} \left[P_i(\beta^{(t)})\log(P_i(\beta)) + (1 - P_i(\beta^{(t)}))\log(1-P_i(\beta))\right] + \sum_{i=m+1}^{n} [y_i \log(P_i(\beta)) + (1-y_i)\log(1-P_i(\beta))]$$

```r
load(file = "dataex4.Rdata")

# probability of ß
P_beta <- function(x,beta0,beta1){
  exp(beta0+x*beta1) / (1+exp(beta0+x*beta1))
}

# log likelihood
log_like_dataex4 <- function(param){
  beta0<-param[1]; beta1<-param[2]
  x <- dataex4[[1]]; y_modified <- purrr::map2_dbl(dataex4$X,
```

```r
                                                         as.double(dataex4$Y),
                                                         ~ if_else(is.na(.y),P_beta(.x,beta0,beta1),.y))
  sum(y_modified*log(P_beta(x,beta0,beta1))+(1-y_modified)*log(1-P_beta(x,beta0,beta1)))
}

# EM
multi <- function(beta,eps=1e-5){
  diff <- 1

  while(diff>eps){
    beta.old <- beta

    # M-step
    mle <- maxLik(logLik = log_like_dataex4, start = beta)
    beta <- mle[[2]]

    diff <- sum(abs(beta-beta.old))
  }

  return(beta)
}

multi(c(1,-5))
```

```
## [1]  0.7635572 -4.1509698
```

Therefore, the maximum likelihood of β is $\hat{\beta}_0 = 0.7635572, \hat{\beta}_1 = -4.1509698$.

Question 5

(a) **My solution**

Create a vector of observed/latent group data indicator:

$$Z_i = \begin{cases} 1, & y_i \sim LogNormal \\ 0, & y_i \sim Exp \end{cases}$$

Therefore, the log likelihood of complete data would be:

$$\log L(\theta|y,z) = \sum_{i=1}^{n} z_i \left[\log p - \log(y_i\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2}(\log y_i - \mu)^2\right] + \sum_{i=1}^{n}(1-z_i)\left[\log(1-p) + \log\lambda - \lambda y_i\right]$$

For the E-step, we need to compute:

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n} E_Z[z_i|y,\theta^t]\left[\log p - \log(y_i\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2}(\log y_i - \mu)^2\right] + \sum_{i=1}^{n}(1-E_Z[z_i|y,\theta^t])\left[\log(1-p) + \log\lambda - \lambda y_i\right]$$

where $E_Z[z_i|y,\theta^t] = \dfrac{p^{(t)}\frac{1}{y_i\sqrt{2\pi(\sigma^{(t)})^2}}e^{-\frac{1}{2(\sigma^{(t)})^2}(\log y_i - \mu^{(t)})^2}}{p^{(t)}\frac{1}{y_i\sqrt{2\pi(\sigma^{(t)})^2}}e^{-\frac{1}{2(\sigma^{(t)})^2}(\log y_i - \mu^{(t)})^2} + (1-p^{(t)})\lambda^{(t)}e^{-\lambda^{(t)}y_i}} = \tilde{p}_i^{(t)}$.

Thus, for the M-step,

$$\frac{\partial}{\partial p}Q(\theta|\theta^{(t)}) = 0 \Rightarrow p^{(t+1)} = \frac{\sum_{i=1}^{n}\tilde{p}_i^{(t)}}{n}$$

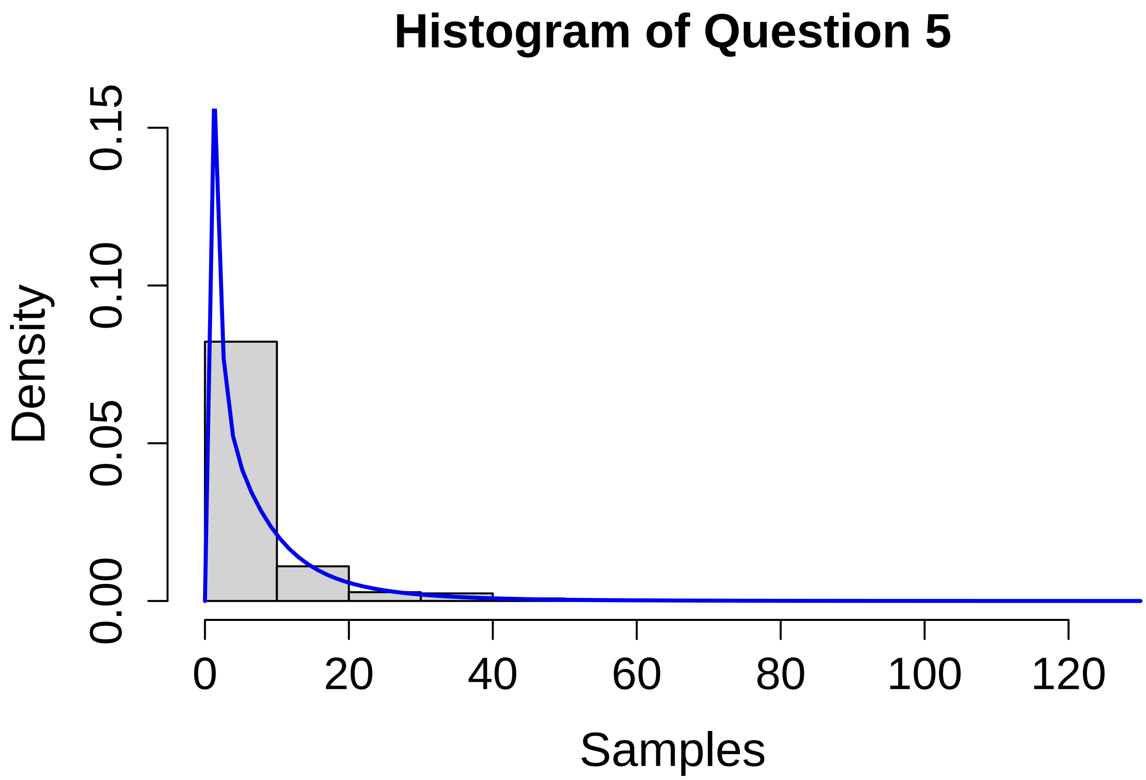
$$\frac{\partial}{\partial \mu} Q(\theta|\theta^{(t)}) = 0 \Rightarrow \mu^{(t+1)} = \frac{\sum_{i=1}^{n} \tilde{p}_i^{(t)} \log y_i}{\sum_{i=1}^{n} \tilde{p}_i^{(t)}}$$

$$\frac{\partial}{\partial \sigma^2} Q(\theta|\theta^{(t)}) = 0 \Rightarrow (\sigma^{(t+1)})^2 = \frac{\sum_{i=1}^{n} \tilde{p}_i^{(t)} (\log y_i - \mu^{(t+1)})^2}{\sum_{i=1}^{n} \tilde{p}_i^{(t)}}$$

$$\frac{\partial}{\partial \lambda} Q(\theta|\theta^{(t)}) = 0 \Rightarrow \lambda^{(t+1)} = \frac{\sum_{i=1}^{n} (1 - \tilde{p}_i^{(t)})}{\sum_{i=1}^{n} (1 - \tilde{p}_i^{(t)}) y_i}$$

(b) **My solution**

```r
load(file = "dataex5.Rdata")

em.mixture.lognorm.exp <-
  function(y,theta0=c(0.1, 1, 0.5**2, 2),eps=1e-5){
  n <- length(y)

  theta <- theta0

  p<-theta[1];mu<-theta[2];sigma<-theta[3];lam<-theta[4]

  diff <- 1
  while(diff>eps){
    theta.old <- theta

    #E-step
    ptilde1 <- p*dlnorm(y, meanlog = mu,sdlog = sqrt(sigma))
    ptilde2 <- (1-p)*dexp(y, rate = lam)
    ptilde <- ptilde1/(ptilde1 + ptilde2)

    #M-step
    p <- mean(ptilde)

    mu <- sum(log(y)*ptilde)/sum(ptilde)
    sigma <- sum(ptilde*(log(y)-mu)**2)/sum(ptilde)

    lam <- sum(1-ptilde)/sum((1-ptilde)*y)

    theta <- c(p,mu,sigma,lam)
    diff <- sum(abs(theta - theta.old))
  }
  return(theta)
}

theta <- em.mixture.lognorm.exp(y = dataex5)

p<-theta[1];mu<-theta[2];sigma<-theta[3];lam<-theta[4]
hist(dataex5, main = "Histogram of Question 5", xlab = "Samples",
ylab = "Density",
cex.main = 1.5, cex.lab = 1.5, cex.axis = 1.4, freq = FALSE, ylim = c(0,0.15))
curve(p*dlnorm(x, mu, sigma)+(1 - p)*dexp(x, lam), add = TRUE, lwd = 2, col = "blue2")
```



Histogram of Question 5
Density
0.00
0.05
0.10
0.15
0
20
40
60
80
100
120
Samples