Homework 3

Yue Gu & Chenyi Yu

March 1, 2018

Question 1

(a)

The log likelihood function is:

$$
\begin{aligned}
l_n^c(\Psi) &= \sum_{i=1}^{n} \log p(y_i, \mathbf{x}_i, z_i; \Psi) &(1)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij}\{\log \pi_j \phi(y_i - \mathbf{x}_i^T \beta_j; 0, \sigma^2)\} &(2)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij}\{\log \pi_j + \log \phi(y_i - \mathbf{x}_i^T \beta_j; 0, \sigma^2)\} &(3)
\end{aligned}
$$

E-Step: if we treat Z as a random variable and take the conditional expression of $l_n^c(\Psi)$, we get:

$$
\begin{aligned}
Q(\Psi \mid \Psi^{(k)}) &= E_z[l_n^c(\Psi)] &(4)\\
&= E_z[\sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij}\{\log \pi_j^{(k)} + \log \phi(y_i - \mathbf{x}_i^T {\beta_j}^{(k)}; 0, {\sigma^2}^{(k)}\}] &(5)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} E[z_{ij} \mid y_i, \mathbf{x}_i; \Psi^{(k)}]\{\log \pi_j^{(k)} + \log \phi(y_i - \mathbf{x}_i^T {\beta_j}^{(k)}; 0, {\sigma^2}^{(k)}\} &(6)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)}\{\log \pi_j^{(k)} + \log \phi(y_i - \mathbf{x}_i^T {\beta_j}^{(k)}; 0, {\sigma^2}^{(k)}\} &(7)\\
& &(8)
\end{aligned}
$$

where

$$
\begin{aligned}
p_{ij}^{(k+1)} &= E[z_{ij} \mid y_i, \mathbf{x}_i; \Psi^{(k)}] &(9)\\
&= \frac{p(z_{ij}; \pi) p(\mathbf{x}_i \mid y_i, z_i; \Psi^{(k)})}{p(y_i, \mathbf{x}_i; \Psi^{(k)})} \mathit{By\ Bayes'\ Rule} &(10)\\
&= \frac{\pi_j^{(k)} \phi(y_i - \mathbf{x}_i^T \beta_j^{(k)}; 0, {\sigma^2}^{(k)})}{\sum_{j=1}^{m} \pi_j^{(k)} \phi(y_i - \mathbf{x}_i^T \beta_j^{(k)}; 0, {\sigma^2}^{(k)})} &(11)\\
& &(12)
\end{aligned}
$$

(b)

M-Step: Maximize $Q(\Psi|\Psi^{(k)})$ to obtain $(\beta^{(k+1)}, {\sigma^2}^{(k+1)})$

$$
\begin{aligned}
Q(\Psi \mid \Psi^{(k)}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} E[z_{ij} \mid y_i, \mathbf{x}_i; \Psi^{(k)}]\{\log \pi_j^{(k)} + \log \phi(y_i - \mathbf{x}_i^T \beta_j^{(k)}; 0, {\sigma^2}^{(k)})\} && (13)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)}\{\log \pi_j^{(k)} + \log \phi(y_i - \mathbf{x}_i^T \beta_j^{(k)}; 0, {\sigma^2}^{(k)})\} && (14)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \log \pi_j^{(k)} + \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \log \phi(y_i - \mathbf{x}_i^T \beta_j^{(k)}; 0, {\sigma^2}^{(k)}) && (15)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \log \pi_j^{(k)} + \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \log(\frac{1}{\sqrt{2\pi}\sigma^{(k)}}) - \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \frac{(y_i - \mathbf{x}_i^T \beta_j^{(k)})^2}{2{\sigma^2}^{(k)}} && (16)
\end{aligned}
$$

Let

$$A_1 = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \log \pi_j^{(k)} \tag{17}$$

$$A_2 = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \log(\frac{1}{\sqrt{2\pi}\sigma^{(k)}}) \tag{18}$$

$$A_3 = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \frac{(y_i - \mathbf{x}_i^T \beta_j^{(k)})^2}{2{\sigma^2}^{(k)}} \tag{19}$$

Then we apply Lagrange equation with the constraint $\sum\limits_{j=1}^{m} \pi_j = 1$, we have

$$L({\pi_1}^{(k)}, ..., {\pi_m}^{(k)}; \lambda) = Q(\Psi \mid \Psi^{(k)}) - \lambda(\sum_{j=1}^{m} {\pi_j}^{(k)} - 1) \tag{20}$$

By taking the derivative of L with repective to $\pi_j^{(k)}$ and set to zero, we obtain

$$\frac{\partial L}{\partial \pi_j^{(k)}} = \sum_{i=1}^{n} {p_{ij}}^{(k+1)} \frac{1}{\pi_j^{(k+1)}} - \lambda = 0 \tag{21}$$

$$\Rightarrow \pi_j^{(k+1)} = \frac{\sum\limits_{i=1}^{n} p_{ij}^{(k+1)}}{n} \tag{22}$$

And

$$\frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} p_{ij}^{(k+1)}}{\lambda} = \frac{n}{\lambda} = 1 \tag{23}$$

$$\Rightarrow \lambda = n \tag{24}$$

Therefore, we get

$$\pi_j^{(k+1)} = \frac{\sum\limits_{i=1}^{n} p_{ij}^{(k+1)}}{n} \tag{25}$$

As required.

By taking the derivative of A_3 with repective to $\beta_j^{(k)}$ and set to zero, we get

$$\frac{\partial A_3}{\partial \beta_j^{(k)}} \propto \sum_{i=1}^{n} p_{ij}^{(k+1)} \mathbf{x}_i (y_i - \mathbf{x}_i^T \beta_j^{(k+1)}) = 0 \tag{26}$$

$$\Rightarrow \sum_{i=1}^{n} p_{ij}^{(k+1)} \mathbf{x}_i y_i = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \beta_j^{(k+1)} p_{ij}^{(k+1)} \tag{27}$$

$$\Rightarrow \beta_j^{(k+1)} = (\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T p_{ij}^{(k+1)})^{-1} (\sum_{i=1}^{n} \mathbf{x}_i p_{ij}^{(k+1)} y_i), j = 1, ..., m \tag{28}$$

As required.

By taking the derivative of $A_2 + A_3$ with repective to $\sigma^{2(k)}$ and set to zero, we get

$$\sigma^{2(k+1)} = \frac{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} p_{ij}^{(k+1)} (y_i - \mathbf{x}_i^T \beta_j^{(k+1)})^2}{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} p_{ij}^{(k+1)}} \tag{29}$$

$$\Rightarrow \sigma^{2(k+1)} = \frac{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} p_{ij}^{(k+1)} (y_i - \mathbf{x}_i^T \beta_j^{(k+1)})^2}{n} \tag{30}$$

As required.

Question 2

(a)

When calculating the normalizing constant C, we can seperate the formula $g(x) \propto (2x^{\theta-1} + x^{\theta-1/2})e^{-x}$ into two Gamma distributions. Then we can get $2C\Gamma(\theta) + C\Gamma(\theta + \frac{1}{2} = 1)$. Thus $C = \frac{1}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})}$

$$g(x) = \frac{1}{2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2})} 2x^{\theta-1} e^{-x} + \frac{1}{2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2})} 2x^{\theta-1/2} e^{-x}$$

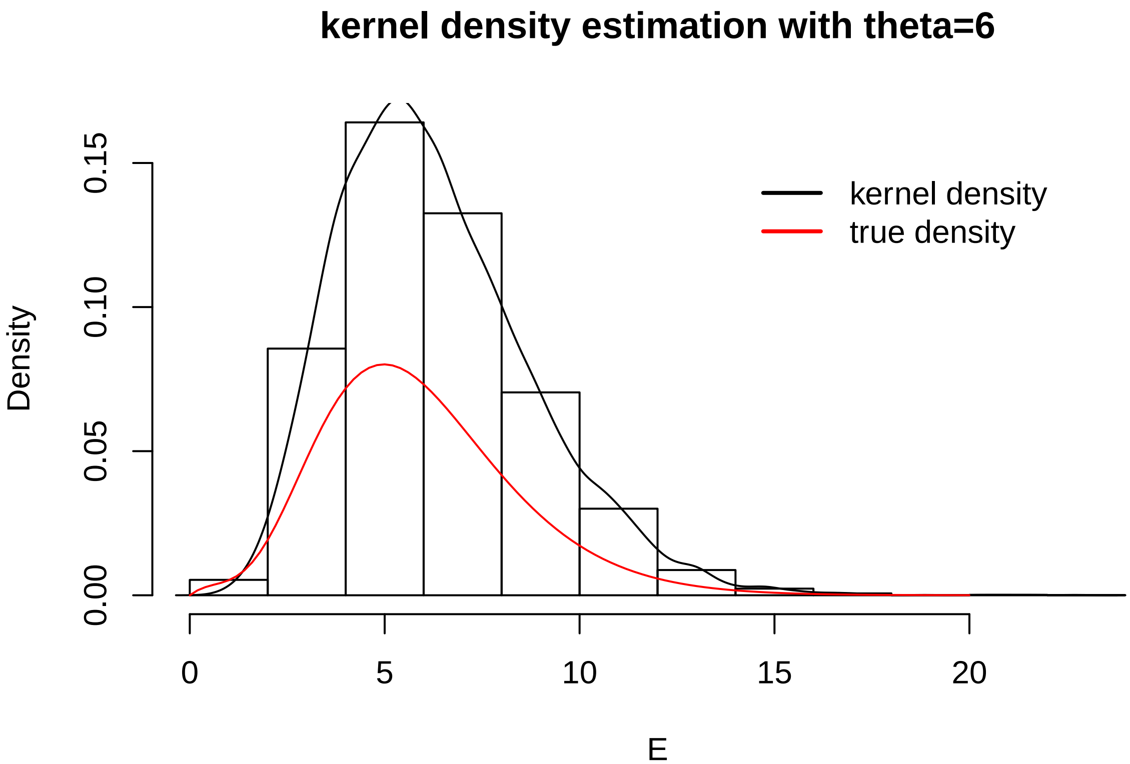
$$= \frac{2\Gamma(\theta)}{2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2})} \cdot \frac{1}{\Gamma(\theta)} x^{\theta-1} e^{-x} + \frac{\Gamma(\theta + \frac{1}{2})}{2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2})} \cdot \frac{1}{\Gamma(\theta + \frac{1}{2})} 2x^{\theta-1/2} e^{-x}$$

g is a mixture of Gamma(θ,1) and Gamma($\theta+\frac{1}{2}$,1) with the weights $\frac{2\Gamma(\theta)}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})}$ and $\frac{\Gamma(\theta+\frac{1}{2})}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})}$ respectively.

(b)

```r
sample_g <- function(theta){
  w <- 2*gamma(theta)/(2*gamma(theta)+gamma(theta+1/2))
  m <- 10000
  counts <- 0
  draws <- c()
  u <- runif(m,0,1)
  for(i in 1:m){
    if(u[i]<w){
      x <- rgamma(1,theta,1)
      counts <- counts+1
      draws <- c(draws,x)
    }
    else{
      x <- rgamma(1,theta+0.5,1)
    counts <- counts+1
    draws <- c(draws,x)
    }
  }
  return(draws)
}
E <- sample_g(6)
hist(E,prob=TRUE,main = "kernel density estimation with theta=6")
lines(density(E))

g <-function(x,theta=6){
  (1/(2*gamma(theta)+gamma(theta+0.5)))*(2*x^(theta-1)+x*(theta-0.5))*exp(-x)
}
curve(g,from = 0,to=20,add=T,col="red")
legend("topright", inset=.1,
         legend = c("kernel density","true density"),
         bty = "n", lty = 1, lwd = 2, col = c("black", "red"))
```



Question 3

(a)

Because $x \in (0,1)$

$$f \propto \frac{x^{\theta-1}}{1+x^2} + \sqrt{2+x^2}(1-x)^{\beta-1} \le x^{\theta-1} + \sqrt{3}(1-x)^{\beta-1}$$

$$C\int x^{\theta-1} + \sqrt{3}(1-x)^{\beta-1}dx = 1$$

$$C = \frac{1}{q/\theta + \sqrt{3}/\beta}$$

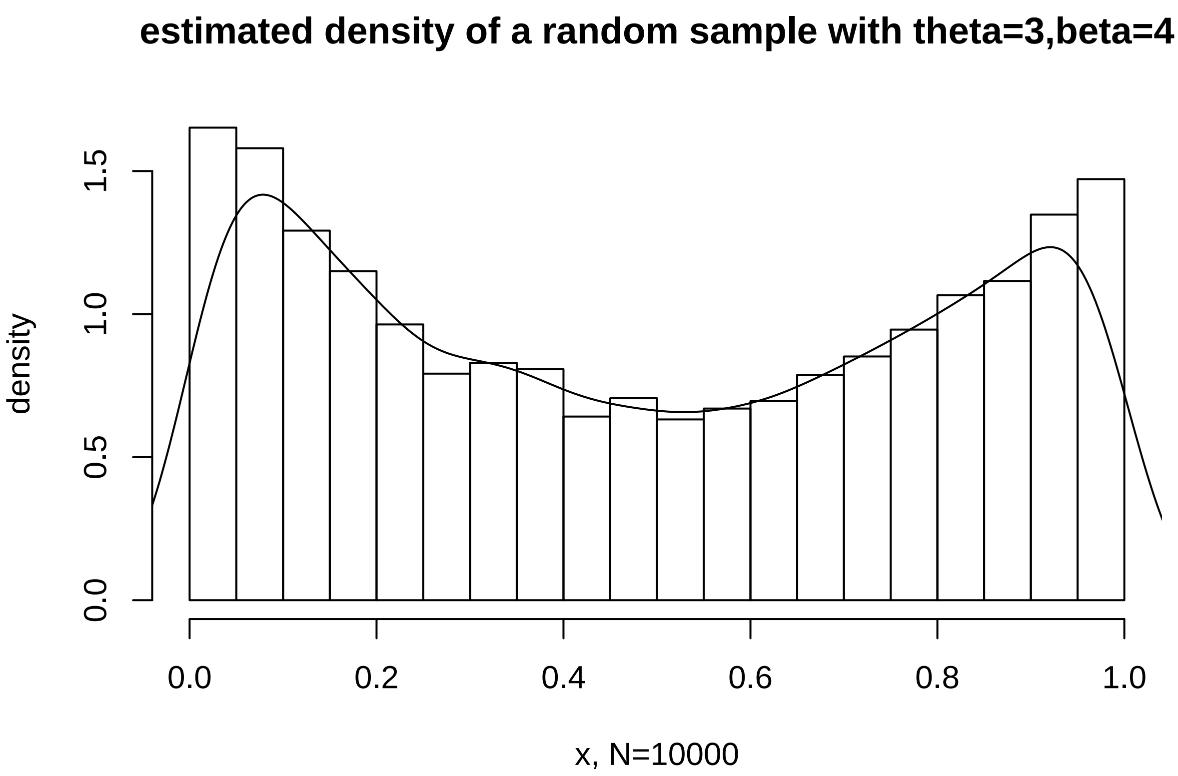
$$g(x) = \frac{1/\theta}{1/\theta+\sqrt{3}/\beta}\theta x^{\theta-1}dx + \frac{\sqrt{3}/\beta}{1/\theta+\sqrt{3}/\beta}\beta(1-x)^{\beta-1}$$

$$= \frac{1/\theta}{1/\theta+\sqrt{3}/\beta}\frac{x^{\theta-1}}{B(\theta-1)} + \frac{\sqrt{3}/\beta}{1/\theta+\sqrt{3}/\beta}\frac{(1-x)^{\beta-1}}{B(1,\beta)}$$

So the instrumental distribution can be seperated into Beta(θ,1) and Beta(1,β) and the weights are $\frac{1/\theta}{1/\theta+\sqrt{3}/\beta}$ and $\frac{\sqrt{3}/\beta}{1/\theta+\sqrt{3}/\beta}$

```
sample_f <- function(theta,beta){
  f <- function(x){
   x^(theta-1)/(1+x^2)+sqrt(2+x^2)*(1-x)^(beta-1)
  }
  g <- function(x){
    x^(theta-1)+sqrt(3)*(1-x)^(beta-1)
  }
```

```r
  w <- (1/theta)/(1/theta+sqrt(3)/beta)
  m <- 10000
  counts <- 0
  draws <- c()
while(counts<=m){
    repeat{
      u <- runif(1,0,1)
      if (u<=w){
        x <- rbeta(1,theta,1)}
      else{
        x <- rbeta(1,1,beta) }
      t <- u*g(x)
      if(t<=f(x)){break}
    }
    counts <- counts+1
    draws <- c(draws,x)
}
  return(draws)
}
E <- sample_f(3,4)
hist(E,freq=F, xlab = "x, N=10000", ylab="density",main="estimated density of a random sample with theta
lines(density(E))
```



(b)

For $0 < x < 1$

$$f_1(x) \propto \frac{x^{\theta-1}}{1+x^2} \leq x^{\theta-1} = g_1(x)$$

$$f_2(x) = \sqrt{2+x^2}(1-x)^{\beta-1} \leq \sqrt{3}(1-x)^{\beta-1} = g_2(x)$$

Set C_1 and C_2 as the normalizing constant for $g_1(x)$ and $g_2(x)$ respectively. $C_1 \int_0^1 g_1(x) = 1$ and $C_2 \int_0^1 g_2(x) = 1$

$$C_1 \int_0^1 x^{\theta-1}dx = \frac{C_1}{\theta}$$

$$C_1 = \theta$$

$$C_2 \int_0^1 \sqrt{3}(1-x)^{\beta-1}dx = \frac{C_2\sqrt{3}}{\beta} = 1$$

$$C_2 = \frac{\beta}{\sqrt{3}}$$

The weight $w_1 = \frac{1/C_1}{1/C_1+1/C2} = \frac{C_2}{C_1+C_2} = \frac{1/\theta}{1/\theta+\sqrt{3}/\beta}$ and $w_2 = \frac{1/C_2}{1/C_1+1/C_2} = \frac{C_1}{C_1+C_2} = \frac{\sqrt{3}/\beta}{1/\theta+\sqrt{3}/\theta}$

```
sample2_f <- function(theta,beta){
  f1 <- function(x){
   x^(theta-1)/(1+x^2)
  }
  f2 <- function(x){
    sqrt(2+x^2)*(1-x)^(beta-1)
  }
  g1<- function(x){
    x^(theta-1)
  }
  g2 <- function(x){
    sqrt(3)*(1-x)^(beta-1)
  }
  w <- (1/theta)/(1/theta+sqrt(3)/beta)
  m <- 10000
  counts <- 0
  draws <- c()
  x <- 0
while(counts<=m){
    flag <- TRUE
    while(flag){
      u <- runif(1,0,1)
      if (u<=w){
        x <- rbeta(1,theta,1)}
        t <- u* g1(x)
        if(t<=f1(x)){
          flag <- FALSE
        }
      else{
        x <- rbeta(1,1,beta)
        t <- u* g2(x)
        if (t<=f2(x)){
          flag <- FALSE
        }
        }
    }
    counts <- counts+1
    draws <- c(draws,x)
}
  return(draws)
```

```
}
E <- sample2_f(3,4)
hist(E,probability = TRUE,main = "estimated density of a random sample with theta=3,beta=4")
lines(density(E))
```