Homework3 Q1

Xiao Cui & Nayun Li

2018/2/26

Question 1

$$
\begin{aligned}
&E-step:\\
&Q(\Psi \mid \Psi^{(k)}) = E_z[l_n^c(\Psi)]\\
&\qquad = \sum_{i=1}^{n}\sum_{i=j}^{m} \underbrace{P(Z_{ij}=1 \mid y_i, x_i, \Psi^{(x)})}_{P_{ij}^{(k)}} lnP(Z_{ij}=1, x_i, y_i \mid \Psi)
\end{aligned}
$$

$$
\begin{aligned}
P_{ij}^{(k+1)} &= \frac{P(Z_{ij}=1 \mid y_i, x_i, \Psi^{(x)})}{P(x_i, y_i \mid \Psi^{(k)})}\\
&= \frac{P(Z_{ij}=1 \mid y_i, x_i, \Psi^{(x)})}{\sum_{s=1}^{m} P(Z_{is}=1, x_i, y_i \mid \Psi^{(k)})}
\end{aligned}
$$

$$
\begin{aligned}
&with\\
&P(Z_{ij}=1 \mid y_i, x_i, \Psi^{(x)}) = P(Z_{ij}=1 \mid \Psi^{(k)})P(x_i, y_i \mid Z_{ij}=1, \Psi^{(k)})\\
&\qquad = \pi_j^{(k)} * \phi(y_i - x_i^T\beta_j; 0, \sigma^2)\\
&\Longrightarrow p_{ij} = \frac{\pi_j^{(k)} * \phi(y_i - x_i^T\beta_j; 0, \sigma^2)}{\sum_{j=1}^{m} \pi_j^{(k)}\phi(y_i - x_i^T\beta_j; 0, \sigma^2)}
\end{aligned}
$$

$$
\begin{aligned}
M-step: Q(\Psi \mid \Psi^{(k)}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} * lnp(Z_{ij}=1, x_i, y_i \mid \Psi)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} * ln\pi_j^{(k)} * \phi(y_i - x_i^T\beta_j; 0, \sigma^2)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} * ln\pi_j^{(k)} * (\frac{1}{\sqrt{2\pi}\sigma)} exp[-\frac{1}{2}(\frac{y_i - x_i^T\beta_j}{\sigma})^2])\\
&= \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} * ln\pi_j^{(k)}\sqrt{2\pi}^{-1}}_{I_1} - \underbrace{\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} ln\sigma^2}_{I_2} - \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} * [\frac{y_i - x_i^T\beta_j}{\sigma}]^2}_{I_3}
\end{aligned}
$$

As for I_3,only I_3 has β_j. So minimize I_3, which means we only need to find $X_i^T\beta_j$ to minimize each I_{3j}. From the property of sample mean, $X_i^T\beta_j$ must be the mean of a weighted sample $y_1, ..., y_n$, each y_i has weight $P_{ij}^{(k+1)}$:

$$
x_i^T\beta_j = \frac{\sum_{i=1}^{n} p_{ij}^{(k+1)y} y_i}{\sum_{i=1}^{n} p_{ij}^{(k+1)}}
$$

$$
\begin{aligned}
x_i x_i^T \beta_j &= \frac{\sum_{i=1}^n x_i p_{ij}^{(k+1)y} y_i}{\sum_{i=1}^n x_i x_i^T p_{ij}^{(k+1)}} \\
&= (\sum_{i=1}^n x_i p_{ij}^{(k+1)y} y_i)
\end{aligned}
$$

Next, only I_2 and I_3 contain σ^2, and I_2 , I_3 is the sum of j terms of the following form, each including a single σ^2,

$$
S_j = \sum_{i=1}^n p_{ij}^{(k+1)} ln\sigma^2 + \sum_{i=1}^n p_{ij}^{(k+1)} \frac{(y_i - x_i^T \beta_j)^2}{\sigma^2}
$$

And we only need to find σ^2 to minimize S_j. Now the $x_i^T \beta_j$ is equal to the weight of y_i, to minimize S_j, σ^2 must be the sample variance of the weighted sample $x_1, ..., x_n$. However, all the variances are equal to σ^2.

$$
\begin{aligned}
\sigma^{2(k+1)} &= \frac{\sum_{i=1}^n p_{ij}(y_i - x_i^T \beta_j)^2}{\sum_{i=1}^n p_{ij}^{(k+1)}} \\
\Longrightarrow \sum_{i=1}^n p_{ij}^{(k+1)} \sigma^{2(k+1)} &= \sum_{i=1}^n p_{ij}(y_i - x_i^T \beta_j)^2 \\
\Longrightarrow \sigma^{2(k+1)} \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} (y_i - x_i^T \beta_j)^2 \\
\Longrightarrow \sigma^{2(k+1)} &= \frac{\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} (y_i - x_i^T \beta_j)^2}{n}
\end{aligned}
$$

Finally, only I_1 contains π_j^k,since $\pi_1 + \cdots + \pi_m = 1$. Therefore, the maximization can be obtained by finding a solution $L'(\pi_1 \cdots \pi_m) = 0$.

$$
\begin{aligned}
&i.e: \\
&\quad \frac{\partial L(\pi_1 \cdots \pi_m)}{\partial \pi_j} = 0 \\
&where \\
&\quad L(\pi_1 \cdots \pi_m) = \sum_{j=1}^m (\sum_{i=1}^n p_{ij}^{(k+1)}) * ln(\pi_j) - \lambda(\sum_{j=1}^m \pi_j - 1) \\
&\Longrightarrow \pi_j = \frac{\sum_{i=1}^n p_{ij}^{(k+1)}}{\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)}} \\
&\Longrightarrow \pi_j = \frac{\sum_{i=1}^n p_{ij}^{(k+1)}}{n}
\end{aligned}
$$

Question 2

(a)

We find the value of C, and proof g(x) is mixtrue of Gamma distribution:

$$
\begin{aligned}
& C\int_0^\infty (2x^{\theta-1}+x^{\theta-1/2})e^{-x}dx = 1 \\
\Longrightarrow & \int_0^\infty (2x^{\theta-1}+x^{\theta-1/2})e^{-x}dx = C^{-1} \\
\Longrightarrow & \, 2\int_0^\infty x^{\theta-1}e^{-x}dx + \int_0^\infty x^{\theta-1/2}e^{-x}dx = C^{-1} \\
\Longrightarrow & \, 2\Gamma(\theta) + \int_0^\infty x^{\theta-1/2}e^{-x}dx = C^{-1} \\
\Longrightarrow & \, 2\Gamma(\theta) + \Gamma(\theta+\frac{1}{2}) = C^{-1} \\
\Longrightarrow & \, C = \frac{1}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})}
\end{aligned}
$$

$$
\begin{aligned}
& g(x) \propto (2x^{\theta-1}+x^{\theta-1/2})e^{-x} \\
\Longrightarrow & \, g(x) \propto (2x^{\theta-1}e^{-x}+x^{\theta-1/2}e^{-x}) \\
\Longrightarrow & \, g(x) \propto 2\Gamma(\theta) * \frac{x^{\theta-1}e^{-x}}{\Gamma(\theta)} + \Gamma(\theta+\frac{1}{2}) * \frac{x^{(\theta+1/2)-1}e^{-x}}{\Gamma(\theta+\frac{1}{2})} \\
\Longrightarrow & \, g(x) = \frac{2\Gamma(\theta)}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})} * \frac{x^{\theta-1}e^{-x}}{\Gamma(\theta)} + \frac{\Gamma(\theta+\frac{1}{2})}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})} * \frac{x^{(\theta+1/2)-1}e^{-x}}{\Gamma(\theta+\frac{1}{2})}
\end{aligned}
$$

Therefore, g(x) is a mixture of Gamma distribution, one of which is $g_1(x) = \frac{x^{\theta-1}e^{-x}}{\Gamma(\theta)}$,the other one is $g_2(x) = \frac{x^{(\theta+1/2)-1}e^{-x}}{\Gamma(\theta+\frac{1}{2})}$,with corresponding weight $w_1 = 2\Gamma(\theta)$ and $w_2 = \Gamma(\theta+\frac{1}{2})$, and probability $\pi_1 = \frac{2\Gamma(\theta)}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})}$ and $\pi_2 = \frac{\Gamma(\theta+\frac{1}{2})}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})}$,respectively.

(b)(c)

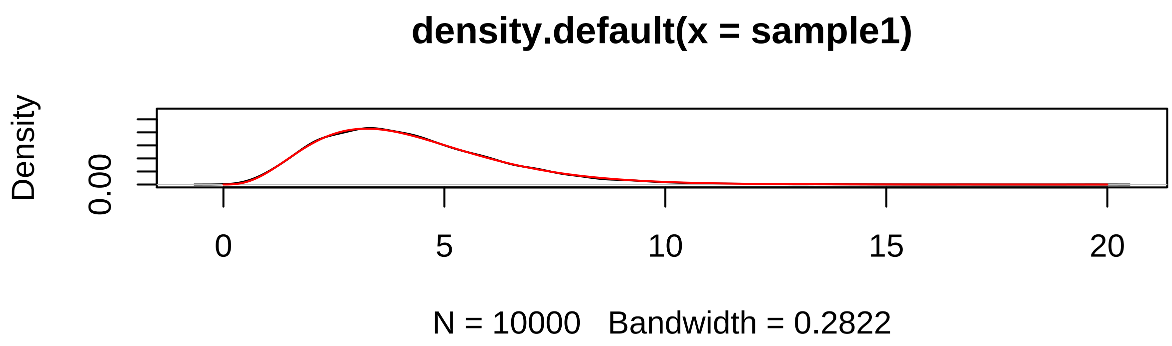
We first define our parameters: $\theta = 4, \alpha = 2$ and the size of sample is 10,000. Then we deal with (b): According to (a), we know g(x) is a mixture of Gamma distribution, so we define g_1, g_2, then we can get the result of estimation. Finally, we use the original function to deserve the true value, and plot two densitys in one figure.

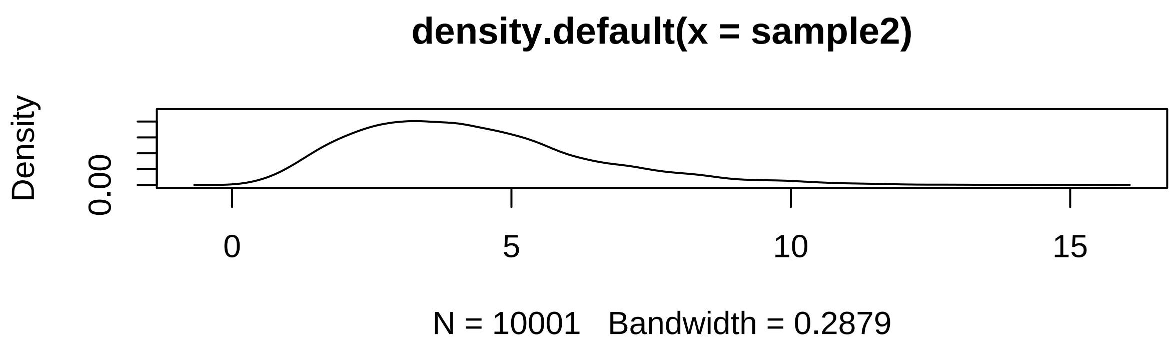
In (c), we need to use rejection sampling to sample. We get $\alpha = 2$, we have the rule X ~ g and U ~ Unif(0,1), if $U > \frac{q(X)}{\alpha g(X)}$, then go to next step, otherwise return X. So the same as the method learned in class, we can derive the result of the estimated density.

Code and Graphs are as follows:

```
##define parameter
problem1<-function(){
  rm(list=ls())
  theta = 4
  alpha = 2
```

```r
  nsamples = 10000


  c = 1/(2*gamma(theta) + gamma(theta + 0.5))
  w1 = 2*c*gamma(theta)
  w2 = c*gamma(theta + 0.5)
  f<- function(x,theta) sqrt(4+x)*x^(theta - 1)*exp(-x)
  g<- function(x,theta) (2*x^(theta -1) + x^(theta - 0.5))*exp(-x)

  ##Question b
  sample_g <- function(n, theta){
    uni<- runif(n,0,1)
    len_g1 <- length(uni[uni < w1])
    len_g2 <- n - len_g1
    g1 <- rgamma(len_g1, theta, 1)
    g2 <- rgamma(len_g2, theta+0.5, 1)
    return(append(g1,g2))
  }
  ##Question C
  sample_f <- function(n, theta){
    fsample <- c()
    while(length(fsample) <= n){
      x<- sample_g(1, theta = theta)
      u<- runif(1, 0, 1)
      if (u < (f(x=x, theta = theta)/(alpha*g(x=x, theta = theta)))){
        fsample<- append(fsample,x)}
    }
    return(fsample)
  }

  par(mfrow=c(2,1))
  x = seq(0, 20, .1)
  truth = w1*dgamma(x,theta,1) + (1-w1)*dgamma(x,theta+0.5,1)
  sample1 <- sample_g(n = nsamples, theta = theta)
  plot(density(sample1), ylim = c(0, 0.28))
  lines(x,truth, col="red", lwd=1)
  #legend('topright', lw = 1,  c('true density','sample density'), col = c('red','black'))

  sample2 <- sample_f(n = nsamples, theta = theta)
  plot(density(sample2), ylim = c(0, 0.23))
}
problem1()
```





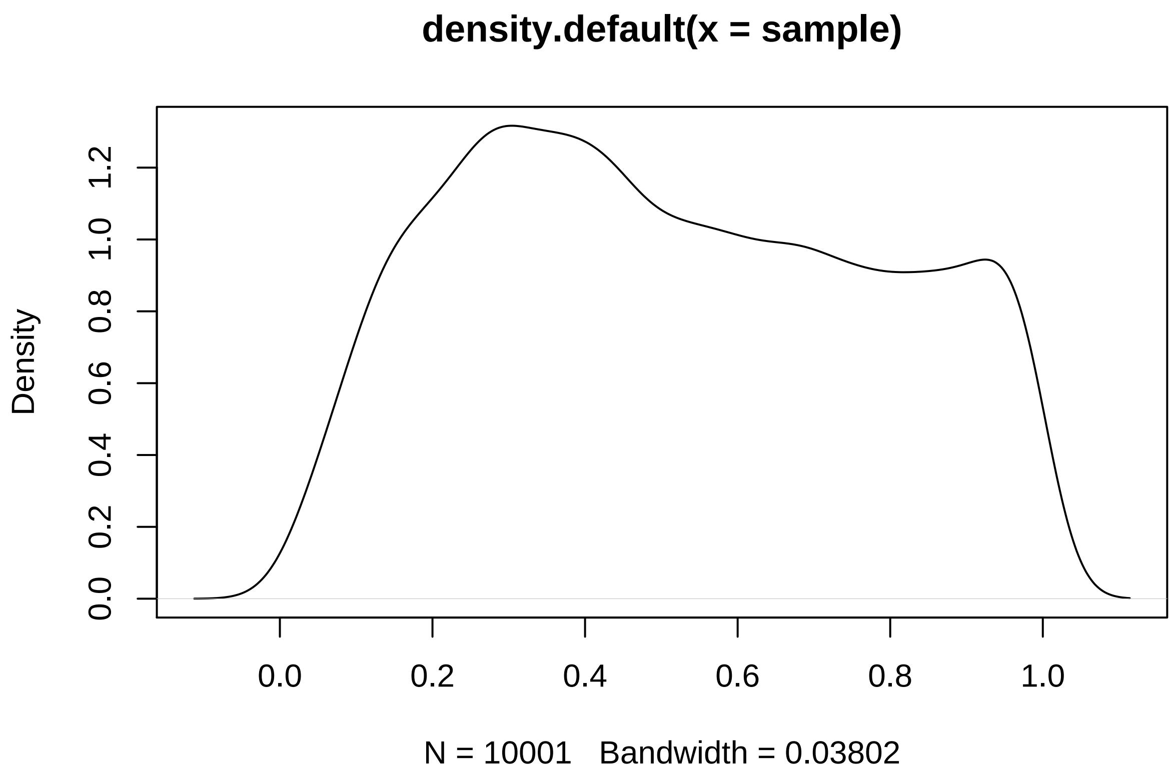
Question 3

(a)

First, we define our parameters: $\theta = 2$, $\beta = 4$, and the size of samples is 10,000. In (a), we use a mixture of Beta functions as our instrumental density, so we define our function g. Then use the rule $u < \frac{f}{g}$ to get the estimated density.

```
##Problem 3
##Question a
problem2a<-function(){
  rm(list = ls())
  theta = 2
  beta = 4
  nsamples = 10000


  f<-function(x,theta,beta) (x^(theta-1))/(1+x^2) + sqrt(2+x^2)*(1-x)^(beta -1)
  g1 <- function(x, theta) x^(theta -1)
  g2 <- function(x, beta) sqrt(3)*(1-x)^(beta-1)
  c <- 1/(beta(theta, 1)+sqrt(3)*beta(1, beta))
  w1<- c*beta(theta,1)
  w2<- sqrt(3)*c*beta(1,beta)
  g <- function(x,theta,beta) (g1(x=x,theta=theta)*w1 + g2(x=x,beta=beta)*w2)
```

```
  ##sample g
  sample_g2<-function(n,theta,beta){
    uni<-runif(n,0,1)
    len_g1<-length(uni[uni<w1])
    len_g2<-n -len_g1
    g1 <- rbeta(len_g1, theta, 1)
    g2 <- rbeta(len_g2, 2, beta)
    return(append(g1,g2))
  }
  ##sample f
  sample_f2<-function(n,theta){
    fsample<-c()
    while (length(fsample)<=n) {
      x<-sample_g2(1,theta = theta, beta = beta)
      u<-runif(1,0,1)
      if(u<(f(x=x, theta=theta,beta=beta)/g(x=x,theta=theta,beta=beta))) fsample <- append(fsam
    }
    return(fsample)
  }
  sample<-sample_f2(n=nsamples,theta = theta)
  plot(density(sample))
}

problem2a()
```



(b)

In (b), we use the rejection sampling to estimate the density. First we define our $\alpha_1 = 1$, which should greater than or equal to 1, $\alpha_2 = 2$, which should be greater than or equal to $\sqrt{3}$, then use the rejection sampling rule($U > \frac{q(X)}{\alpha g(X)}$) to deserve the density. Codes and results are as follows:

```
##question b
problem2b<-function(){
  rm(list = ls())
  theta = 2
  beta = 4
  nsamples = 10000

  alpha_1 = 1 #alpha_1 should be greater than or equal to 1
  alpha_2 = 2 #alpha_2 should be greater than or equal to sqrt(3)
  w1 = alpha_1/(alpha_1 + alpha_2)
  w2 = alpha_2/(alpha_1 + alpha_2)

  #define functions
  f1<- function(x,theta) (x^(theta - 1))/(1+x^2)
  f2<- function(x,beta) sqrt(2+x^2)*(1-x)^(beta -1)
  g1<-function(x,theta) (x^(theta-1))
  g2<-function(x, beta) (1-x)^(beta -1)
  ##part a
  ##sample g first, set n=1
  sample_g1<-function(theta, beta){
    uni<-runif(1,0,1)
    if(uni<w1) g_sample<-rbeta(1,theta,1)
    else       g_sample<-rbeta(1,1,beta)
    return(append(g_sample,uni))
  }
  ##sample f
  sample_f1<-function(n,theta,beta){
    f_sample<-c()
    for(i in 1:n){
      x<-sample_g1(theta,beta)
      u<-runif(1,0,1)
      if(x[2]<w1){
        if (u<f1(x[1],theta)/(alpha_1*g1(x[1],theta))) f_sample<- append(f_sample,x[1])
      }
      else{
        if(u<f2(x[2],beta)/(alpha_2*g2(x[1],beta))) f_sample<- append(f_sample,x[1])
      }
    }
    return(f_sample)
  }
  ##plot rejection sample histogram of f with density of f on same graph for 2a
  sample <-sample_f1(n = nsamples,theta = theta, beta = beta)
```

```r
  plot(density(sample))
}
problem2b()
```