STAT-5361
Statistical Computing
HW #3

Instructor

Prof. Jun Yan

Students

Hee Seung Kim (2539662)

Department of ECE, University of Connecticut

e-mail: hee.s.kim@uconn.edu

Due March/1/2018

***All figures and table are the results of R code**

Problem 1.

(a)

< From a lecture note, >

$$Q(\theta|\theta_t) = \sum_z p(z|x,\theta_t)\ln p(x,z|\theta) = \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}\ln p(z_i = k, x_i|\theta)$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{K} p(z_i = k|x_i,\theta_t)\ln p(z_i = k, x_i|\theta) = \sum_{i=1}^{n}\sum_{k=1}^{K} q_{tk}\phi(x_i|\mu_{tk},\Sigma_{tk})\ln p(z_i = k, x_i|\theta)$$

$$= I_1 - \frac{I_2}{2} - \frac{I_3}{2}$$

$I_1 = \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}\ln[(2\pi)^{-d/2}q_k]$,

$I_2 = \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}\ln|\Sigma_k|$,

$I_3 = \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ik}(x_i - \mu_k)'\Sigma_k^{-1}(x_i - \mu_k)$,

$I_{3k} = \sum_{i=1}^{n} w_{ik}(x_i - \mu_k)'\Sigma_k^{-1}(x_i - \mu_k), \leftarrow because\ only\ I_3\ contains\ \mu_k\ which\ is\ a\ quadratic\ form.$

$\mu_k = \frac{\sum_{i=1}^{n} w_{ik}x_i}{\sum_{i=1}^{n} w_{ik}}$

cf 1) $p(z_i = k|x_i,\theta_t) = q_{tk}\phi(x_i|\mu_{tk},\Sigma_{tk}) = q_k n(x_i|\mu_k,\Sigma_k)$

cf 2) $n(x_i|\mu_k,\Sigma_k) = \frac{(2\pi)^{-d/2}}{\sqrt{|\Sigma_k|}} exp\{-\frac{1}{2}(x_i - \mu_k)'\Sigma_k^{-1}(x_i - \mu_k)\}$

cf 3) $w_{ik} = \frac{p(z_i=k|x_i,\theta_t)}{\sum_{s=1}^{K} p(z_i=k|x_i,\theta_t)}$

For the given problem, $w_{ik} = p_{ij}$

$$Q(\psi|\psi^k) = E_2\{l_n^c(\psi)\} = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)}\{\ln \pi_j + \ln\frac{1}{\sqrt{2\pi}\sigma} + \frac{1}{2\sigma^2}(y_i - X_i^T\beta_j)^2\} = I_1 - \frac{I_2}{2} - \frac{I_3}{2}$$

$I_1 = \sum_{i=1}^{n}\sum_{j=1}^{m} p^{(k+1)}\ln[\pi_j(2\pi)^{-1/2}]$,

$I_2 = \sum_{i=1}^{n}\sum_{j=1}^{m} p^{(k+1)}\ln\sigma^2$,

$I_3 = \sum_{i=1}^{n}\sum_{j=1}^{m} p^{(k+1)}(y_i - X_i^T\beta_j)^2\,\sigma^2$,

because only I_3 contains β_j,

$$I_{3j} = \sum_{i=1}^{n} p^{(k+1)}(y_i - X_i^T\beta_j)^2\,\sigma^2$$

$$\Rightarrow \frac{d}{d\beta_j} I_{3j} = \sum_{i=1}^{n} p^{(k+1)}(-2x_i y_i + 2x_i X_i^T\beta_j)^2\,\sigma^2 = 0$$

$$\Rightarrow \beta^{(k+1)} = \sum_{i=1}^{n}(p^{(k+1)}x_i X_i^T)^{-1}\sum_{i=1}^{n} p^{(k+1)}x_i y_i$$

also because only I_2, I_3 contain σ^2,

$$\Rightarrow \frac{d}{d\sigma^2}(I_2 + I_3) = \sum_{i=1}^{n}\sum_{j=1}^{m} p^{(k+1)}\sigma^{-2} - \sum_{i=1}^{n}\sum_{j=1}^{m} p^{(k+1)}(y_i - X_i^T\beta_j)^2(\sigma^{-2})^{-2} = 0$$

$$\Rightarrow \sigma^{2(k+1)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} p^{(k+1)}(y_i - X_i^T\beta_j)^2}{n}$$

the minimized result of π_j is,

$$\pi_j = \frac{1}{n}\sum_{i=1}^{n} p^{(k+1)}$$

Problem 2.

(a)

As you can see following result, The $g(x)$ can be expressed as a mixture of two Gamma distributions with different weights.

$$g(x) = \mathrm{c} \times (2x^{\theta-1} + x^{\theta-\frac{1}{2}})e^{-x}$$

$$c\int_0^\infty \left(2x^{\theta-1} + x^{\theta-\frac{1}{2}}\right)e^{-x}dx = 1 \leftarrow because\ integral\ result\ should\ be\ 1\ for\ whole\ range$$

$$\Rightarrow \int_0^\infty \left(2x^{\theta-1} + x^{\theta-\frac{1}{2}}\right)e^{-x}dx = 2\int_0^\infty x^{\theta-1}\,e^{-x}dx + \int_0^\infty x^{\theta-\frac{1}{2}}\,e^{-x}dx \ = \frac{1}{c}$$

$$Since\ \Gamma(z) = \int_0^\infty x^{Z-1}\,e^{-x}dx \quad (\leftarrow Definition\ of\ Gamma\ function),$$

$$i)\ 2\int_0^\infty x^{\theta-1}\,e^{-x}dx = 2\Gamma(\theta)$$

$$ii)\ \int_0^\infty x^{\theta-\frac{1}{2}}\,e^{-x}dx = \int_0^\infty x^{(\theta+\frac{1}{2})-1}\,e^{-x}dx \ = \Gamma\left(\theta+\frac{1}{2}\right)$$

$$thus, 2\int_0^\infty x^{\theta-1}\,e^{-x}dx + \int_0^\infty x^{\theta-\frac{1}{2}}\,e^{-x}dx \ = \ 2\Gamma(\theta) + \Gamma\left(\theta+\frac{1}{2}\right) = \frac{1}{c}$$

$$\Rightarrow c = \frac{1}{2\Gamma(\theta) + \Gamma\left(\theta+\frac{1}{2}\right)}$$

in conclusion,

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

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

(b)

This function requires two parameters. (N: desired number of sample, θ: Gamma scale)

$\boldsymbol{STEP\ 0}.\ Prepare\ slots\ for\ the\ sample$

$X \leftarrow rep(0, N);$

$\boldsymbol{STEP\ 1}.\ Calculate\ the\ weight\ of\ each\ Gamma\ distribution\ (\pi 1, \pi 2)$

$\pi 1 \leftarrow 2 * gamma(\theta)/(2 * gamma(\theta) + gamma(\theta + 0.5))$

$\boldsymbol{STEP\ 2}.\ For\ given\ N\ times, draw\ a\ Gamma\ observation\ according\ to\ the\ chosen\ index$

$for\ (i\ in\ 1{:}\,N)\ \{$

$if\ (runif(1)\ >\ pi1)\ \{\ X[i]\ \leftarrow\ rgamma(1, shape = \theta, scale = 1)\}$

$else\ \{\ X[i]\ \leftarrow\ rgamma(1, shape = \theta + 0.5, scale = 1)\ \}\ \}$

$\boldsymbol{STEP\ 3}.\ Return\ generated\ values$

$return(X);$

(c)

This function requires two parameters. (N: desired number of sample, θ: Gamma scale)

$\boldsymbol{STEP\ 0}.\ Prepare\ slots\ for\ the\ sample$

$X \leftarrow rep(0, N);$

$\boldsymbol{STEP\ 1}.\ Draw\ a\ sample\ from\ g, using\ the\ rs1b\ function\ defined\ above.$

$while\ (1)\ \{$

$Y\ =\ rs1b(1, \theta);$

$\boldsymbol{STEP\ 2}.\ If\ randomly\ generated\ uniform\ number\ exceed\ given\ thresh$
$-\ hold, then\ sample\ Y\ again, o.w, retain\ Y.$

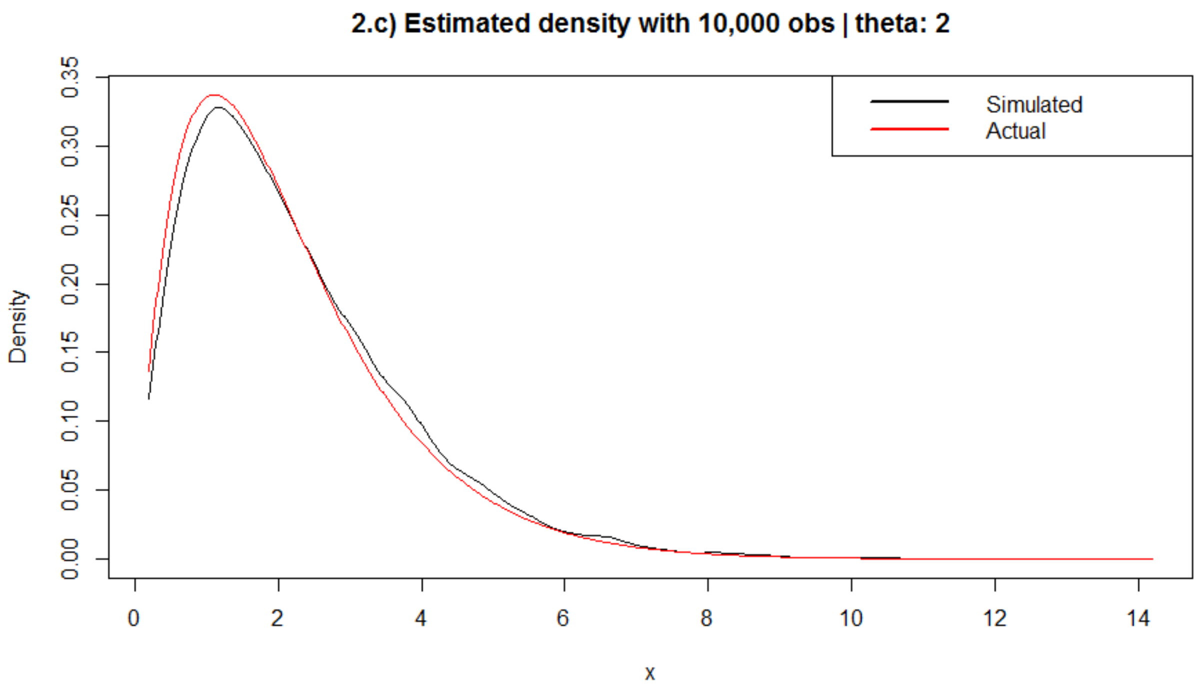
$if\ (runif(1)\ <\ sqrt(4 + Y)/(sqrt(Y) + 2))\ \{$

$X[i]\ =\ Y;\ break;$

$\}$

$\boldsymbol{STEP\ 3}.\ Return\ generated\ values$

$return(X);$



To see whether generated sample values are fit into given distribution, we have to draw a plot for 'Simulated' and 'Actual' values.

(1) Specify the number of points in the interval on which density is to be estimated. The points are equally spaced in the interval, with the first one and last one being the end points of the interval. Note that the maximum value of interval is set by maximum value of generated sample.

```
intvl=c(.2, floor(max(sample1c)+1)+.2); n.p=1+(intvl[2]-intvl[1])/0.02;
```

(2) Estimate the density. The returned is

```
df.obj = density(sample1c, from=intvl[1], to=intvl[2], n=n.p), x = df.obj$x; df = df.obj$y;
```

(3) Calculate the value of q(x) and make the two on the same scale

```
q.val = (4+x)^0.5 * x^(theta-1)*exp(-x); q.val = q.val*mean(df)/mean(q.val);
```

(4) Draw a plot for both 'Simulated' and 'Actual' values with legend

```
plot(ts(cbind(df, q.val), start=intvl[1], deltat=diff(intvl)/(n.p-1)),
```

plot.type="single", col=c("black", "red"),

ylab="Density", xlab="x", main = paste("2.c) Estimated density with 10,000 obs | theta:", theta)); legend('topright', legend=c("Simulated","Actual"),lty=c(1,1), lwd=c(2.5,2.5), col=c('black','red'))

Problem 3.

(a)

Writing a function to sample for f using g(mixture of Beta) with rejection sampling. This function requires following parameters. (N: desired number of sample, θ: 1st beta scale, Beta: 2nd beta scale)

$\boldsymbol{STEP\ 0}.\ Prepare\ slots\ for\ the\ sample$

$X \leftarrow rep(0, N);$

$\boldsymbol{STEP\ 1}.\ Calculate\ the\ weight\ of\ each\ Beta\ distribution\ (\lambda, 1-\lambda)$

$\lambda \leftarrow 3\text{^}0.5 * \theta/(beta + 3\text{^}0.5 * \theta)$

$\boldsymbol{STEP\ 2}.\ Draw\ a\ sample\ from\ g.$

$while\ (1)\ \{$

$Z \leftarrow runif(1)$

$if\ (runif(1) > \lambda)\ \{\ Y \leftarrow Z\text{^}(1/\theta)\}$

$else\ \{\ Y \leftarrow 1 - Z\text{^}(1/beta)\ \}$

$\boldsymbol{STEP\ 3}.\ If\ randomly\ generated\ uniform\ number\ exceed\ given\ thresh - hold, then\ sample\ Y\ again, o.w, retain\ Y.$

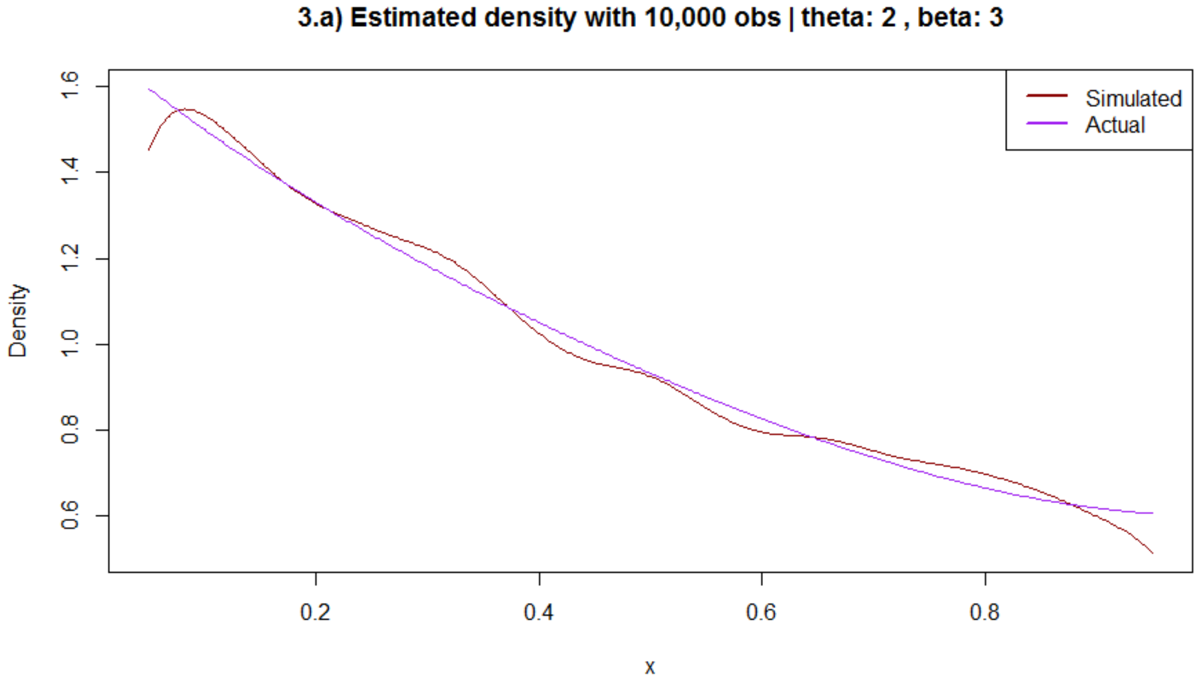
$$if\ (runif(1) < (Y\text{^}(\theta-1)/(1+Y\text{^}2) + sqrt(Y\text{^}2+2)*(1-Y)\text{^}(beta-1))/(Y\text{^}(\theta-1) + sqrt(3)*(1-Y)\text{^}(beta-1)))$$

$\{\ X[i] = Y;\ break;\ \}$

$\boldsymbol{STEP\ 4}.\ Repeat\ STEP\ 1\ \&\ 2\ N\ times\ and\ Return\ generated\ values$

$for\ (i\ in\ 1{:}N)\ \{\ STEP\ 1\ \&\ 2\ \}$

$return(X);$



To see whether generated sample values are fit into given distribution, we have to draw a plot for 'Simulated' and 'Actual' values.

(1) Specify the number of points in the interval on which density is to be estimated. The points are equally spaced in the interval, with the first one and last one being the end points of the interval. Note that the maximum value of interval should be less than 1(Propertiy of beta distribution)

```
intvl=c(.05, 0.95); n.p=451;
```

(2).Estimate the density. The returned is

```
df.obj = density(sample2a, from=intvl[1], to=intvl[2], n=n.p), x = df.obj$x; df = df.obj$y;
```

(3) Calculate the value of q(x) and make the two on the same scale

```
q.val = (x^(theta-1))/(1+x^2) + sqrt(2+x^2)*(1-x)^(beta-1);
```

(4) Draw a plot for both 'Simulated' and 'Actual' values with legend

```
plot(ts(cbind(df, q.val), start=intvl[1], deltat=diff(intvl)/(n.p-1)),
plot.type="single", col=c("dark red", "purple"), ylab="Density", xlab="x", main = paste("3.a) Estimated density with 10,000 obs | theta:", theta,", beta:",beta)); legend('topright', legend=c("Simulated","Actual"),lty=c(1,1), lwd=c(2.5,2.5),col=c('dark red','purple'))
```

(b)

$\boldsymbol{STEP\ 0}.\ Prepare\ slots\ for\ the\ sample$

$X \leftarrow rep(0, N);$

$\boldsymbol{STEP\ 1}.\ Calculate\ the\ weight\ of\ each\ Beta\ distribution\ (\lambda, 1-\lambda)$

$lambda \leftarrow 3\text{^}0.5 * \theta/(beta + 3\text{^}0.5 * \theta)$

$\boldsymbol{STEP\ 2}.\ For\ given\ N\ times, draw\ a\ Beta\ observation\ according\ to\ the\ chosen\ index\ and\ if\ randomly$

$generated.\ Uniform\ number\ exceed\ given\ thresh$
$-\ hold\ for\ each\ Beta\ distribution, then\ sample\ Y\ again, o.w, retain\ Y.$

$if\ (runif(1)\ >\ \lambda)\ \{$

$while\ (1)\ \{$

$Y\ =\ rbeta(1, \theta, 1);$

$if\ (runif(1)\ <\ 1/(1+Y\text{^}2))\ \{$

$X[i]\ =\ Y;\ break;\}$

$\}\}$

$else\ \{$

$while\ (1)\ \{$

$Y\ =\ rbeta(1,1, beta);$

$if\ (runif(1)\ <\ sqrt(2/3 + Y\text{^}2/3))\ \{$

$X[i]\ =\ Y;\ break;$

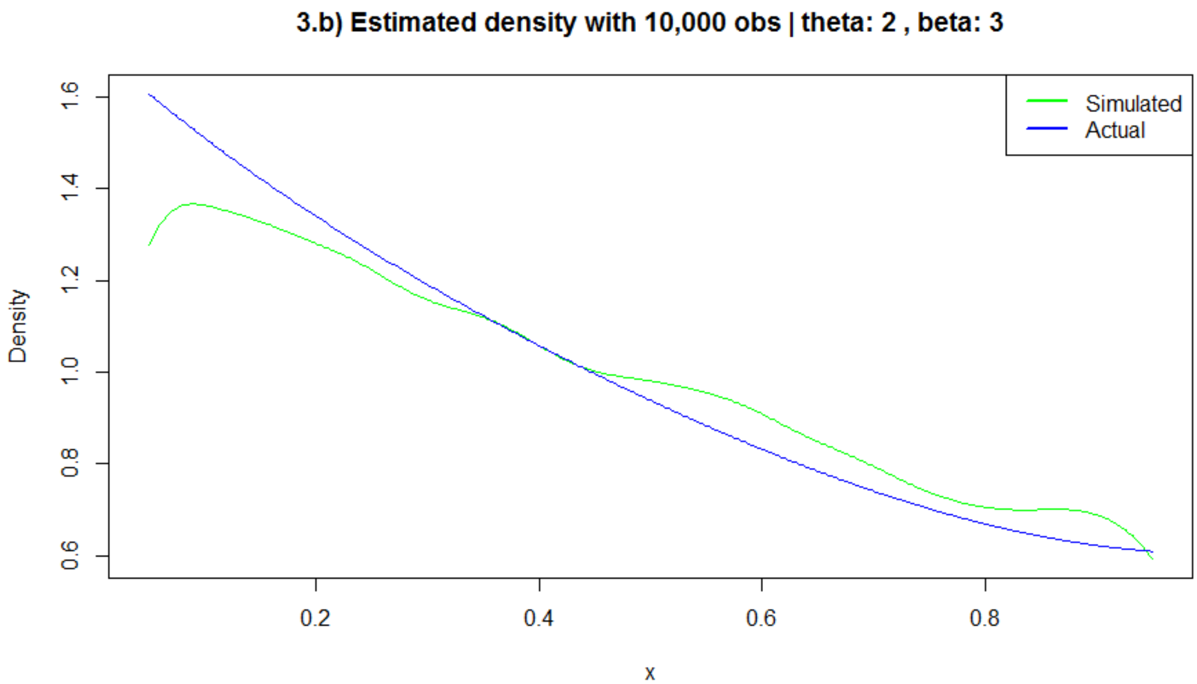
$\}$

$\}\}$

$\boldsymbol{STEP\ 3}.\ Repeat\ STEP\ 2\ N\ times\ and\ Return\ generated\ values$

$for\ (i\ in\ 1{:}N)\ \{\ STEP\ 2\ \}$

$return(X);$



To see whether generated sample values are fit into given distribution, we have to draw a plot for 'Simulated' and 'Actual' values.

(1) Specify the number of points in the interval on which density is to be estimated. The points are equally spaced in the interval, with the first one and last one being the end points of the interval. Note that the maximum value of interval should be less than 1(Property of beta distribution)

intvl=c(.05, 0.95); n.p=451;

(2) Estimate the density. The returned is

df.obj = density(sample2b, from=intvl[1], to=intvl[2], n=n.p), x = df.obj$x; df = df.obj$y;

(3) Calculate the value of q(x) and make the two on the same scale

q.val = (x^(theta-1))/(1+x^2) + sqrt(2+x^2)*(1-x)^(beta-1);

(4) Draw a plot for both 'Simulated' and 'Actual' values with legend

plot(ts(cbind(df, q.val), start=intvl[1], deltat=diff(intvl)/(n.p-1)), plot.type="single", col=c("green", "blue"), ylab="Density", xlab="x", main = paste("3.b) Estimated density with 10,000 obs | theta:", theta,", beta:",beta)); legend('topright', legend=c("Simulated","Actual"),lty=c(1,1), lwd=c(2.5,2.5),col=c('green','blue'))