STA 243 Assignment 3

Chen Zihao 915490404

Problem 1.

(a)

$$
\begin{aligned}
l(\theta) &= x_1 \log(2+\theta) + (x_2 + x_3)\log(1-\theta) + x_4 \log\theta + c \\
l'(\theta) &= \frac{x_1}{2+\theta} + \frac{x_2 + x_3}{1-\theta} + \frac{x_4}{\theta} = 0
\end{aligned}
$$

plug $(125, 21, 20, 33)$ in , and we can get

$$199\theta^2 - 10\theta - 66 = 0$$

by solving the binary equation above, we can get

$$\theta = \frac{10 \pm \sqrt{100 + 4 \times 199 \times 66}}{2 \times 199} \approx 0.6015713$$

(b)

For the E step, we have,

$$
\begin{aligned}
Q(\theta, \theta^{(t)}) &= E[l_c(\theta)|\curvearrowleft] \\
&= E[(x_{12} + x_4)\log\theta + (x_2 + x_3)\log(1-\theta)|\curvearrowleft] \\
&= (E[x_{12}|\curvearrowleft] + x_4)\log\theta + (x_2 + x_3)\log(1-\theta) \\
&= (x_1 \frac{\theta^{(t)}}{2+\theta^{(t)}} + x_4)\log\theta + (x_2 + x_3)\log(1-\theta)
\end{aligned}
$$

for the M step, we have,

$$
\begin{aligned}
Q(\theta, \theta^{(t)}) &= (x_1 \frac{\theta^{(t)}}{2+\theta^{(t)}} + x_4)\log\theta + (x_2 + x_3)\log(1-\theta) \\
Q'_\theta(\theta, \theta^{(t)}) &= (x_1 \frac{\theta^{(t)}}{2+\theta^{(t)}} + x_4)\frac{1}{\theta} - (x_2 + x_3)\frac{1}{1-\theta} = 0 \\
\Rightarrow \theta^{t+1} &= \frac{x_1 \frac{\theta^{(t)}}{2+\theta^{(t)}} + x_4}{x_1 \frac{\theta^{(t)}}{2+\theta^{(t)}} + x_4 + x_2 + x_3}
\end{aligned}
$$

```
## [1] 0.6015713
```

the number above is the E-M algorithm results.

(c) they are the same.

Problem 2.

step 1. calculate the conditional expectation and replace the missing data with it

step 2. calculate the sample mean and sample variance

$$f_X(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$$\log L = \log f_X(x_1, x_2 | \Sigma, \mu) = -\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)$$

$$E[X_{i1} | X_{i2}] = \mu_1^{(k)} + \sigma_{12}^{(k)} (\sigma_2^{(k)})^{-2} (x_{i2} - \mu_2^{(k)})$$
$$E[X_{i2} | X_{i1}] = \mu_2^{(k)} + \sigma_{12}^{(k)} (\sigma_1^{(k)})^{-2} (x_{i1} - \mu_1^{(k)})$$

for (i in $1:p$) :

(1) $E[(x_i-\mu)^T \Sigma^{-1} (x_i - \mu) \mid x_{\text{observed}}, \Sigma^{(k)}, \mu^{(k)}]$

$= [E[x_{i1}|x_{i2}] - \mu_1,\ x_{i2} - \mu_2]\, \Sigma^{-1} [E[x_{i1}|x_{i2}] - \mu_1,\ x_{i2} - \mu_2]^T$

for (i in $p+1 : p+q$) :

(2) $E[(x_i-\mu)^T \Sigma^{-1} (x_i - \mu) \mid x_{\text{observed}}, \Sigma^{(k)}, \mu^{(k)}]$

$= [x_{i1} - \mu_1,\ E[x_{i2}|x_{i1}] - \mu_2]\, \Sigma^{-1} [x_{i1} - \mu_1,\ E[x_{i2}|x_{i1}] - \mu_2]^T$

for (i in $p+q+1 : n$) :

(3) $E[x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \mid X]$

$= (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$

E part :

$Q((\mu, \Sigma), (\mu^{(k)}, \Sigma^{(k)})) =$

$-\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}\left(\sum_{i=1}^{p}(1) + \sum_{i=p+1}^{p+q}(2) + \sum_{i=p+q}^{n}(3)\right)$

M-part ~~$= -\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}($~~

$$\frac{\partial Q}{\partial \mu} = \frac{1}{2}\left(-\sum_{i=1}^{p} 2\Sigma^{-1}[E(x_{i1}|x_{i2}] - \mu_1, x_{i2} - \mu_2]^T - \sum_{i=p+1}^{p+q} 2\Sigma^{-1}[x_{i1} - \mu_1, E(x_{i2}|x_{i1}) - \mu_2]^T - \sum_{i=p+q+1}^{p+q+r} 2\Sigma^{-1}[x_i - \mu)\right) = 0$$

$\Rightarrow \Sigma^{-1}(\frac{1}{n}\sum x_i^* - \mu) = 0 \Rightarrow \mu^{(k+1)} = \frac{1}{n}\sum x_i^*$ where x^* is the matrix that using the $E[x_{i1}|x_{i2}]$ and $E[x_{i2}|x_{i1}]$ to replace the missing data in original observed x.

$$\Rightarrow \mu_1^{(k+1)} = \frac{1}{n}\left[\sum_{i=1}^{p}(\mu_1^{(k)} + \sigma_{12}^{(k)} (\sigma_2^{(k)})^{-2} (x_{i2} - \mu_2^{(k)}) + \sum_{i=p+1}^{n} x_{i1}\right]$$
$$\mu_2^{(k+1)} = \frac{1}{n}\left[\sum_{i=1}^{p} x_{i1} + \sum_{i=p+1}^{p+q}(\mu_2^{(k)} + \sigma_{12}^{(k)} (\sigma_1^{(k)})^{-2} (x_{i1} - \mu_1^{(k)}) + \sum_{i=p+q+1}^{n} x_{i1}\right]$$

$\frac{\partial Q}{\partial \Sigma} =$

$(\Sigma^{-1} - \Sigma^{-1} s^* \Sigma^{-1})^T = 0 \Rightarrow \Sigma^{(k+1)} = s^*$ where $s^* = \frac{1}{n}(x^* - \bar{x}^*)^T (x^* - \bar{x}^*)$

so that, the update rule is ① replace the missing part by $E(x_{i1}|x_{i2})$ and $E(x_{i2}|x_{i1})$ as above, to form x^* ② calculate $\bar{x}^*_{(k)}$ and $s^*_{(k)}$ of x^* ③ $\mu^{(k+1)} = \bar{x}^*_{(k)}$ and $\Sigma^{(k+1)} = s^*_{(k)}$

Figure 1: Probelm 2

step 3. update the 5 unknown variables by sample mean and sample variance.

step 4. go to step 1 using the new variable given by step3, unless it converges.

details are in the Figure 1.

Problem 3.

$$
\begin{aligned}
f(x) &\propto e^{-x}, 0 < x < 2 \\
\int_0^2 ce^{-x}dx &= 1 \\
-ce^{-x}|_0^2 &= 1 \\
c(1-e^{-2}) &= 1 \\
c &= \frac{1}{1-e^{-2}}
\end{aligned}
$$

so that we have

$$f(x) = \frac{e^{2-x}}{e^2-1}, 0 < x < 2$$

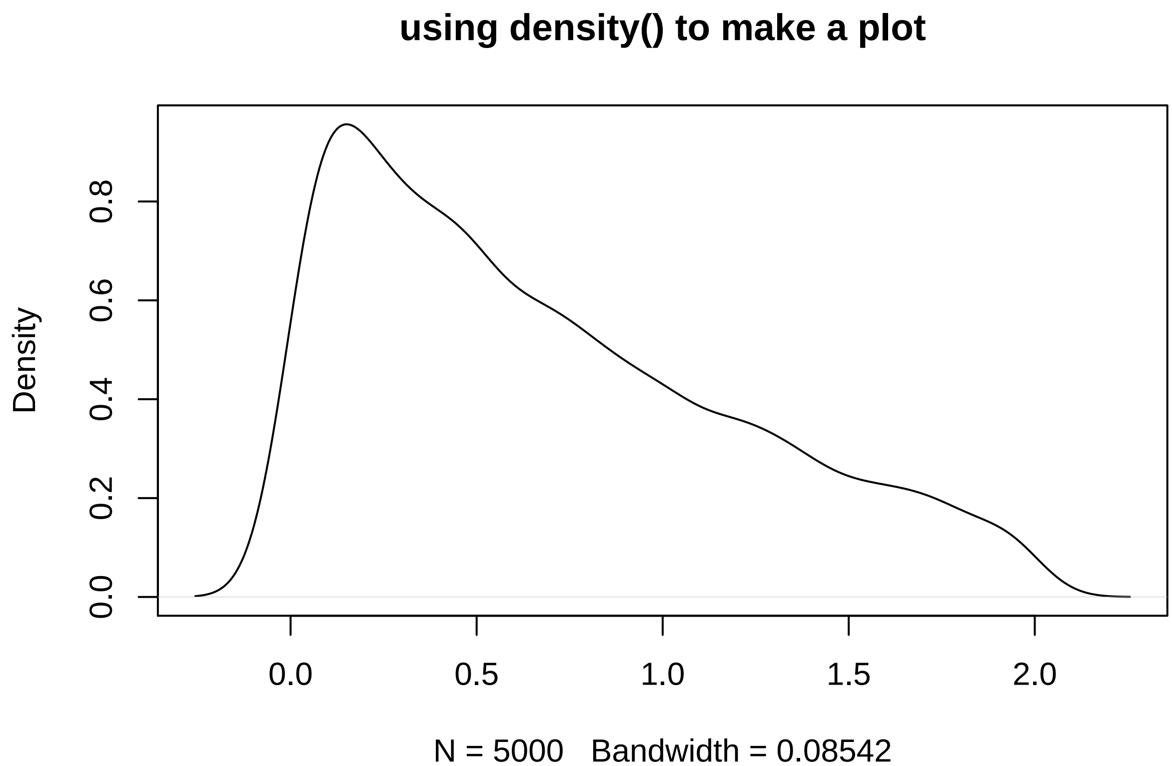
also we can get the cdf (for 0<x<2):

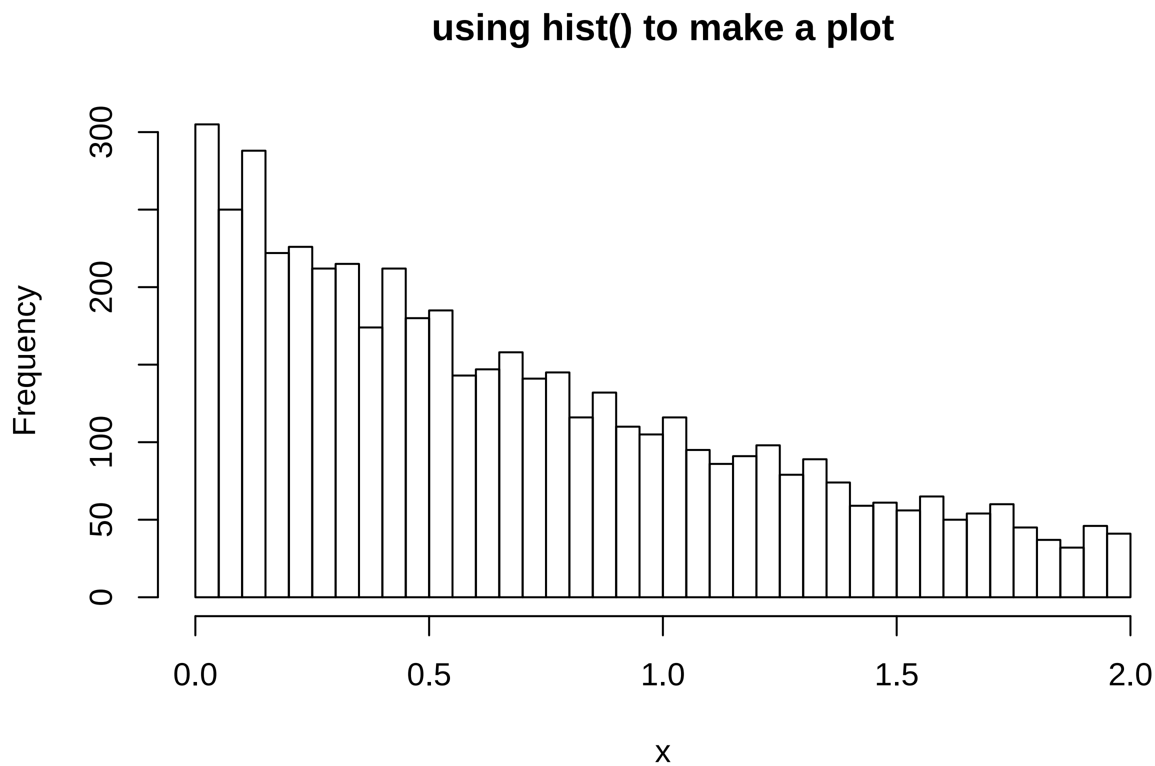
$$F(x) = \int_0^x f(t)dt = \frac{e^2}{e^2-1}\int_0^x e^{-t}dt = \frac{1-e^{-x}}{1-e^{-2}}$$

Inverse transformation

$$F^{-1}(x) = -\log(1-(1-e^{-2})x)$$

so that sample u from unif(0,1) 5000 times and put it into the equation above to get x.


using density() to make a plot
Density
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
2.0
N = 5000   Bandwidth = 0.08542



Problem 4.

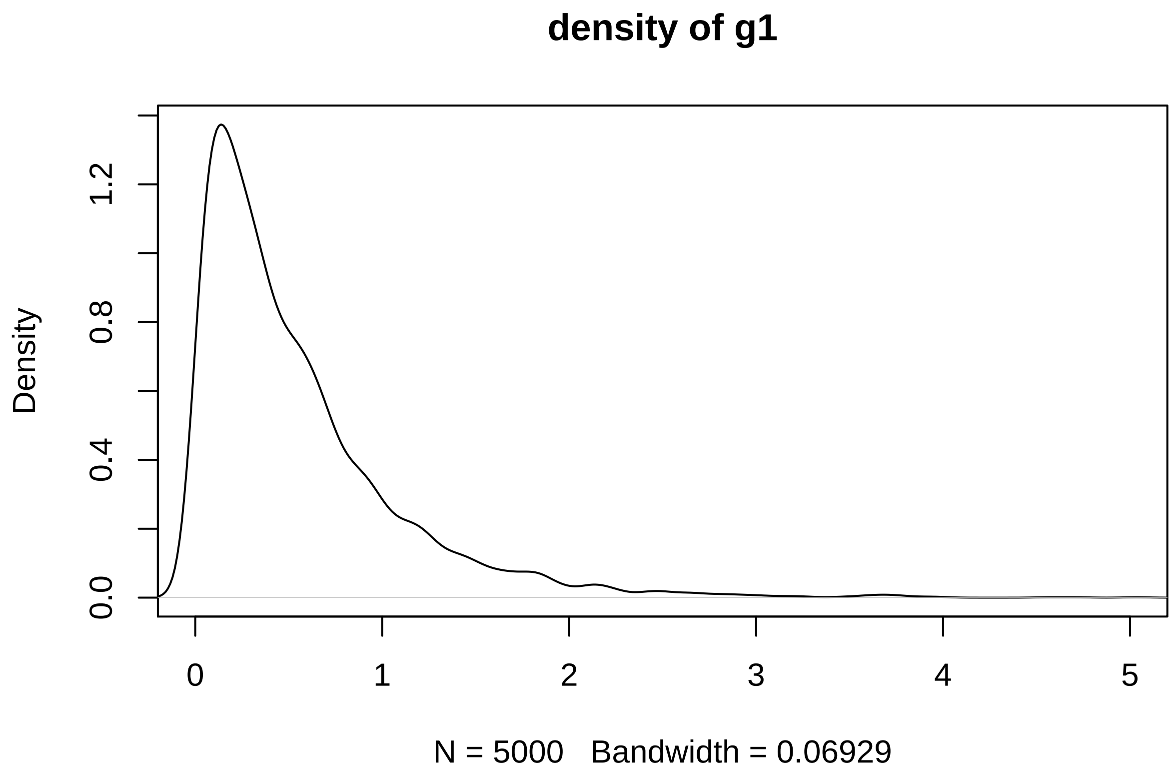
(1)

$q(x) \leq \alpha g_1(x)$ take $\alpha = sup\frac{q(x)}{g_1(x)} = 1$

step 1. $g_1(x) = e^{-x}$ and we can get the inverse transformation that $G_1^{-1}(x) = -\log(1-x)$ and using the inverse transformation method, we generate a samples from unif(0,1) and plug in the function above (then we get a sample x from $g_1(x)$)and get another unif(0,1) sample u.

step 2. if $u > \frac{q(x)}{g_1(x)}$ then return to step 1, otherwise keep x and go the step 3.

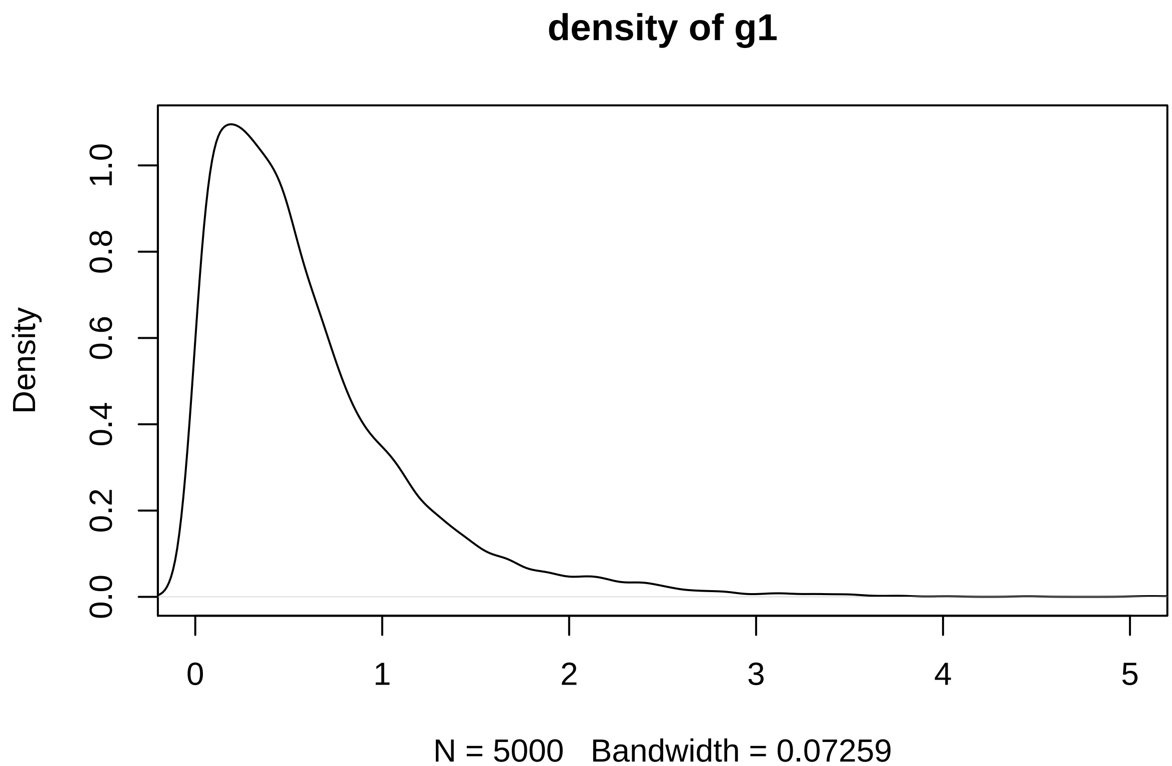
step 3. if we have already generate 5000 random observations plot them, otherwise go to step 1.



now change to $g_2(x) = \frac{2}{\pi(1+x^2)}$, it is easy to get the inverse of the CDF of g_2, $G_2^{-1}(x) = \tan(\frac{\pi}{2}x)$

$\alpha = sup\frac{q(x)}{g_2(x)} = sup\frac{\pi e^{-x}}{2} = \frac{\pi}{2}$

the steps are the same as the previous one.



both graph have some values greater than 5 which is not included.

(b) Instead of using the system.time function (which is not useful since it is not stable as i tried), I use the acceptance ratio to roughly judge which one is better.

The acceptance ratio of g_1

```
5000/t1
```

```
## [1] 0.6248438
```

The acceptance ratio of g_2

```
5000/t2
```

```
## [1] 0.5537099
```

g_1 is faster than g_2 as the acceptance ratio of g_1 is greater than g_2

Problem 5.

(a)

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

(b)

As shown in (a), I use the gamma distribution to let the integral parts to become 1. it can be also shown as below.

$$g(x) = c_1\frac{1}{\Gamma(\theta)}x^{\theta-1}e^{-x} + c_2\frac{1}{\Gamma(\theta+\frac{1}{2})}x^{\theta+1/2-1}e^{-x}$$

It is a mixture of Gamma(θ,1) and Gamma($\theta+\frac{1}{2}$,1)

where the weight $c_1 = \frac{2\Gamma(\theta)}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})}$,$c_2 = \frac{\Gamma(\theta+\frac{1}{2})}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})}$

(c)

we can make a two steps sampling.

step 1. Get U from Uniform(0,1) step 2. If $U > c_1 = \frac{2\Gamma(\theta)}{2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2})}$ get X from Gamma($\theta+\frac{1}{2}$,1), otherwise get X from Gamma(θ,1).

proof:

let $h(x)$ be the pdf of x from the above procedure

$$
\begin{aligned}
h(x) &= P(U > c_1)f(x|U > c_1) + P(U \le c_1)f(x|U \le c_1) \\
&= c_2\frac{1}{\Gamma(\theta+\frac{1}{2})}x^{\theta+1/2-1}e^{-x} + c_1\frac{1}{\Gamma(\theta)}x^{\theta-1}e^{-x} \\
&= g(x)
\end{aligned}
$$

(d)

$$
\begin{aligned}
\frac{q(x)}{g(x)} &= \frac{\sqrt{4+x}x^{\theta-1}e^{-x}}{c(2x^{\theta-1}+x^{\theta-1/2})e^{-x}} \\
&= \frac{\sqrt{4+x}}{c(2+x^{1/2})} \\
&= (2\Gamma(\theta)+\Gamma(\theta+\frac{1}{2}))\frac{\sqrt{4+x}}{(2+x^{1/2})}
\end{aligned}
$$

so that

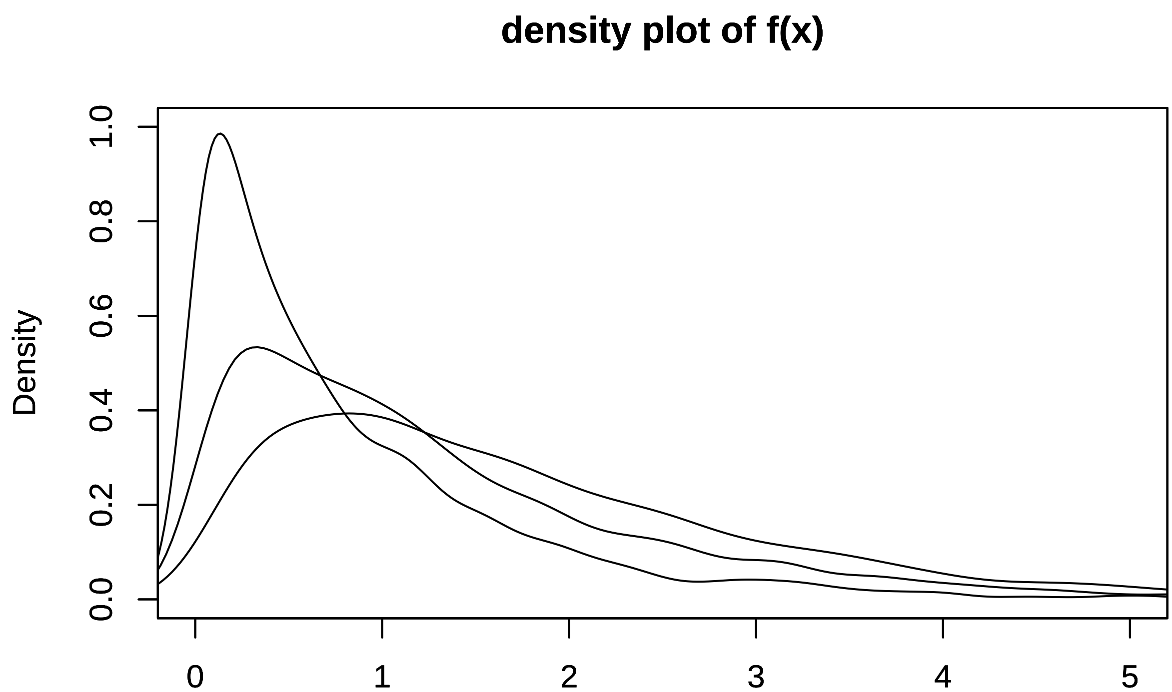
$\alpha = \sup \frac{q(x)}{g(x)} = 2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2})$

for each θ

step 1. sample $X \sim g(x)$, $U \sim Unif(0, 1)$

step 2. if $u > \frac{q(x)}{\alpha g(x)}$ then go to step 1. Otherwise return X.

step 3. repeat step 1 and 2 until got enough samples.



problem 6.

$$f(x, y) \propto x^\alpha y, x > 0, y > 0, x^2 + y^2 \leq 1$$

as we can see $\int x^\alpha dx = \frac{1}{1+\alpha} x^{\alpha+1} < 0$ if $\alpha \leq -1$, so that $\alpha > -1$.

$$1 = \int_0^1 \int_0^{\sqrt{1-x^2}} c x^\alpha y dy dx = \frac{c}{2} \int_0^1 x^\alpha (1 - x^2) dx = \frac{c}{(\alpha + 1)(\alpha + 3)}$$

so that $c = (\alpha + 1)(\alpha + 3)$

$$f(x, y) = (\alpha + 1)(\alpha + 3) x^\alpha y, x > 0, y > 0, x^2 + y^2 \leq 1$$

as the density function can be factorized into two part x^α and y. x and y are independent. we can sample them separately and then we use the rejection algorithm to meet the constrains.

let take an sample $x \sim beta(\alpha+1,1)$ and $y \sim beta(2,1)$, then the joint distribution will be

$$\begin{aligned} f'(x,y) &= \frac{x^{\alpha+1-1}(1-x)^{1-1}}{B(\alpha+1,1)} \frac{y^{2-1}(1-y)^{1-1}}{B(2,1)} \\ &= \frac{1}{B(\alpha+1,1)B(2,1)} x^{\alpha} y \\ &= 2(\alpha+1)x^{\alpha} y \propto x^{\alpha} y \end{aligned}$$

It shows that the sampling is valid. We can use the conditional distribution version of rejection algorithm to modify the constant part of the density.

In addition, $\{(x,y) : x > 0, y > 0, x^2 + y^2 \leq 1\}$ is a subset of $\{(x,y) : 0 < x \leq 1, 0 < y \leq 1\}$

We can do basic rejection algorithm below.

step 1. sample $x \sim beta(\alpha+1,1)$ and $y \sim beta(2,1)$

step 2. if $x^2 + y^2 > 1$, go to step 1. Otherwise, return (x,y).