HW_3_Maheshwari

Saurabh Maheshwari - 100%

May 3, 2018

Problem 1

E and M steps are derived in writing. The result for the MLE and EM algorithms are as follows:

```
## [1] "MLE estimate of theta =  0.601549641643992"
```

```
## [1] "EM estimate of theta =  0.60157127643068"
```

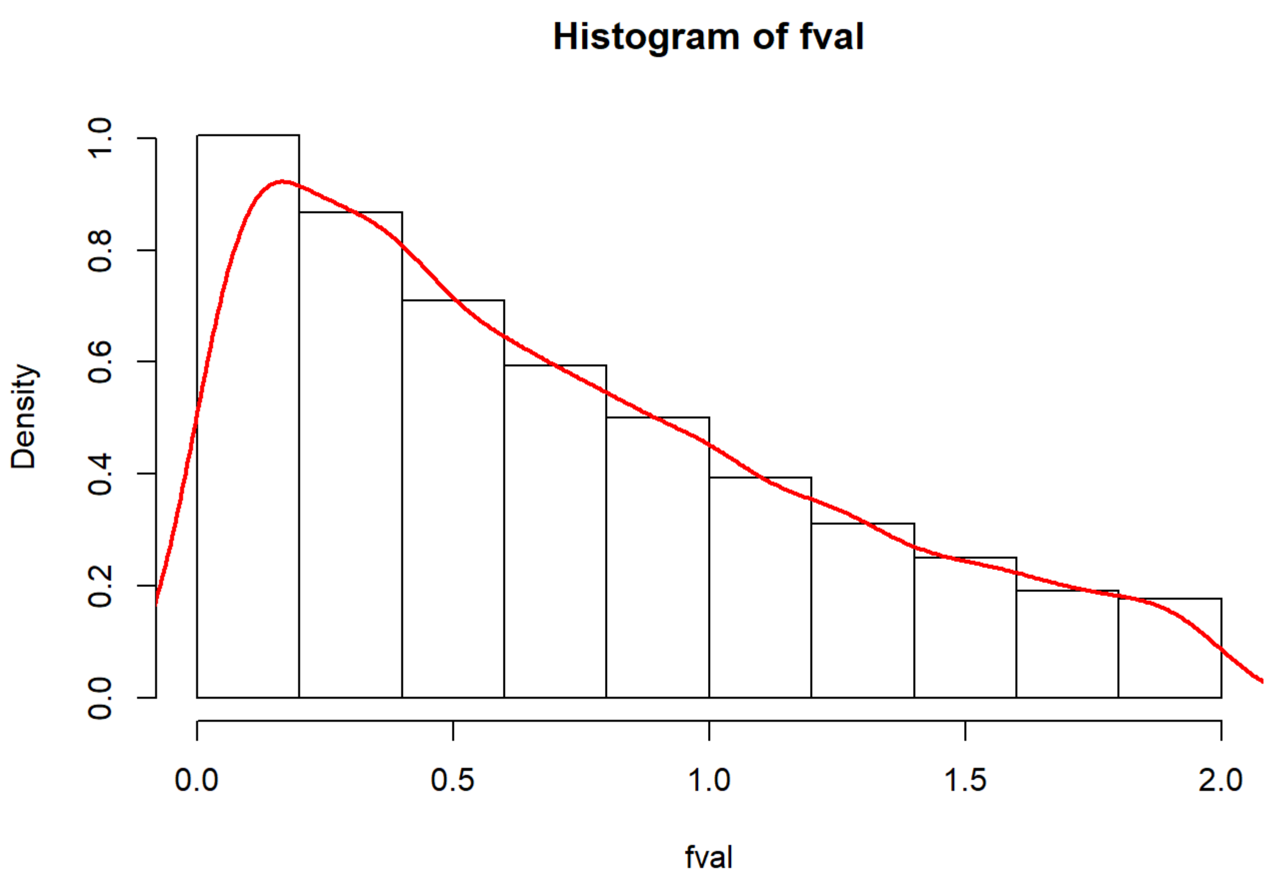
Thus, estimates in part a and b are approximately equal.

Problem 2

Method derived in writing, attached.

Problem 3

Details of the algorithm is provided in writing, attached. The estimated density curve has the following shape.



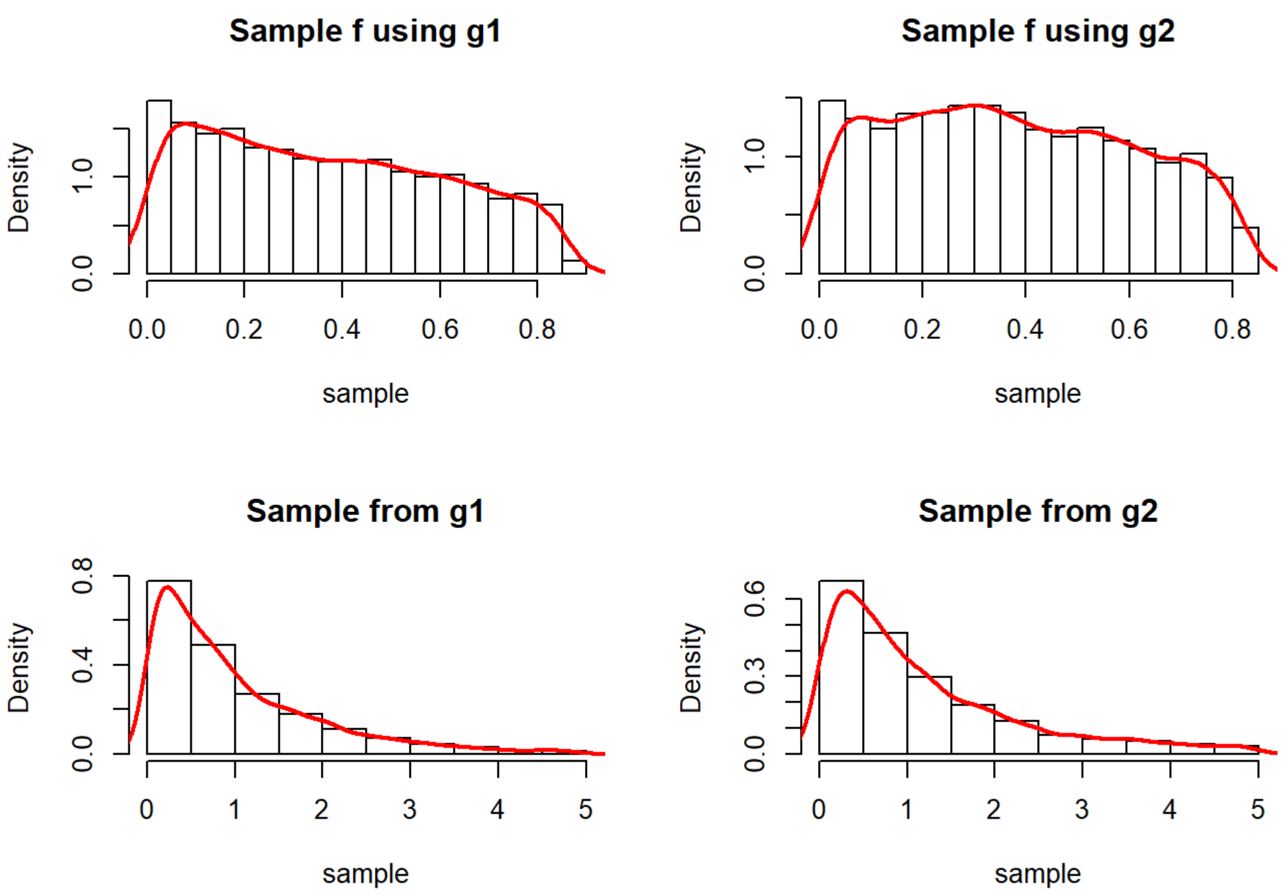
Problem 4

Using rejection sampling, f was sampled using g1 and g2. Also, g1 and g2 were sampled using inverse transform methods. The implemented code is provided. The derivations regarding alphas is in writing and attached. The results are as follows:

```
## [1] "time taken for sampling f using g1 = 1.176139 secs"
```

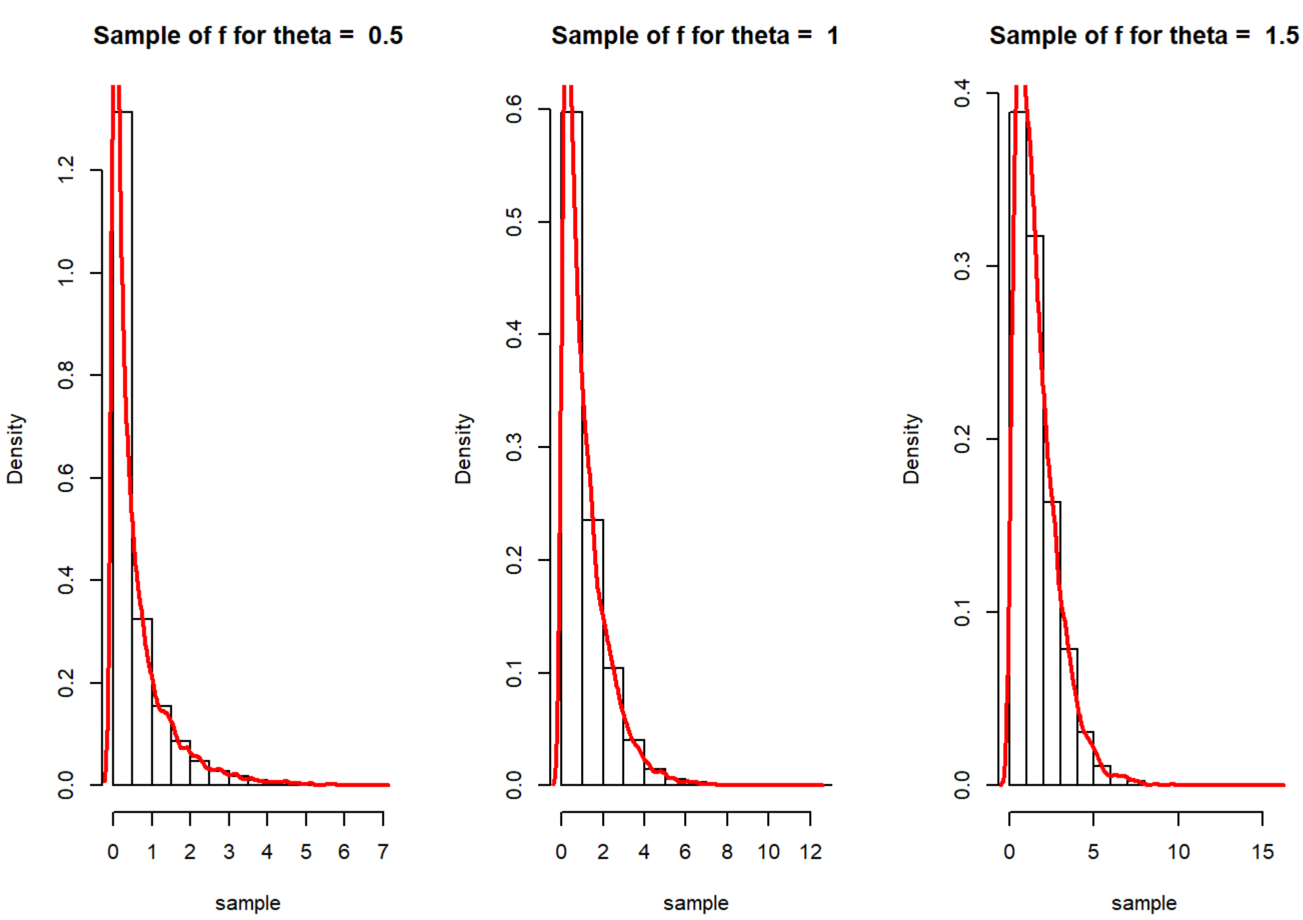
```
## [1] "time taken for sampling f using g2 = 1.861408 secs"
```

As the alpha used for g1 (1) is smaller than alpha used for g2 (pi/2), the probability of acceptance is greater for the algorithm using g1, hence, it is faster.



Problem 5

All the derivations are provided in writing, attached. The code is provided. The density plots are as follows:



As seen in the density graph above, the graph with theta = 0.5 is skewed to left the most, as probability density at small x in high.

Also, the density shown on y axis should not be confused with the actual probability densities, rather used as an indicative of the curve shape. The y axis is such that the total area under the histogram bins equals 1, due to small bins, sometimes the density scales above 1.

Problem 6

The methadology is provided in writing, attached.

Ans. 1)

a) $l(\theta) = x_1 \log(2+\theta) + (x_2+x_3)\log(1-\theta) + x_4 \log\theta + c$

$$\frac{dl(\theta)}{d\theta} = \frac{x_1}{2+\theta} + \frac{x_2+x_3}{1-\theta}(-1) + \frac{x_4}{\theta} = 0 \quad \text{(For MLE)}$$

$\because \; x = (125, 21, 20, 33)$

$\Rightarrow \; x_1(\theta)(1-\theta) - (x_2+x_3)(2+\theta)\theta + x_4(2+\theta)(1-\theta) = 0$

$\Rightarrow \; 125(\theta)(1-\theta) - 41(2+\theta)\theta + 33(2+\theta)(1-\theta) = 0$

Solving for θ: ~~$\theta = -0.443, 0.895$~~ $\theta = -0.5512, 0.6015$

$\because \theta = 4p_4$, hence can't be negative

thus, ~~$\theta = 0.895$~~ $\theta = 0.6015$

b) Now, $l_c(\theta) = (x_{12}+x_4)\log\theta + (x_2+x_3)\log(1-\theta)$

Expectation Step:

$$Q(\theta,\theta^k) = E_{\theta^k}[\,l_c(\theta) \mid X]$$

$$= E_{\theta^k}\Big[\big[(x_{12}+x_4)\log\theta + (x_2+x_3)\log(1-\theta)\big] \,/\, (X_1, X_2, X_3, X_4)\Big]$$

$$= E_{\theta^k}[x_{12}/x_1]\log\theta + x_4\log\theta + (x_2+x_3)\log(1-\theta)$$

$$\Rightarrow E_{\theta^k}[x_{12} \mid x_1] = x_1 \times Pr(x_{12}/x_1)$$

$$= x_1 \times \frac{\theta^k/4}{\theta^k/4 + 1/2} = \frac{\theta^k x_1}{2+\theta^k}$$

Thus, $$Q(\theta,\theta^k) = \frac{\theta^k x_1}{2+\theta^k}\log\theta + x_4\log\theta + (x_2+x_3)\log(1-\theta)$$

M step: $Q'(\theta, \theta^k) = 0$

$$\Rightarrow \left(\frac{\theta^k x_1}{2+\theta^k}\right)\frac{1}{\theta} + \frac{x_4}{\theta} + \frac{(x_2+x_3)}{1-\theta}(-1) = 0$$

$$\Rightarrow \left(\frac{\theta^k x_1}{2+\theta^k} + x_4\right)(1-\theta) - (x_2+x_3)\,\theta = 0$$

$$\Rightarrow \theta^{k+1} = \frac{\frac{\theta^k x_1}{2+\theta^k} + x_4}{\frac{\theta^k x_1}{2+\theta^k} + x_4 + x_2 + x_3}$$

Solving iteratively using R

$\theta = 0.6015$

Hence, θ in part a $= 0.601546$

θ in part b using EM algorithm $= 0.6015713$

Thus, θ are almost the same.

Ans.2) $\mu = (\mu_1, \mu_2)$, $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$

$$l_c(\mu, \Sigma) = -n \log 2\pi - \frac{1}{2} n \log|\Sigma| - \frac{1}{2}\sum_{j=1}^{n} (w_j - \mu)^T \Sigma^{-1} (w_j - \mu)$$

where, $w_j = (w_{j1}, w_{j2})$ are the observations.
for the first p observations, w_{j1} is missing
for the Second q observations, w_{j2} is missing
last r observations, complete data available.

Solving $l_c(\mu, \Sigma)$ more:

$$l_c(\mu, \Sigma) = -n \log 2\pi - \frac{1}{2} n \log|\Sigma| - \frac{1}{2|\Sigma|}\sum_{j=1}^{n} (w_{j1}-\mu_1)^2 \sigma_2^2 - (w_{j2}-\mu_2)(w_{j1}-\mu_1) \times \sigma_{12}$$
$$- (w_{j2}-\mu_2)(w_{j1}-\mu_1)\sigma_{12}$$
$$+ (w_{j2}-\mu_2)^2 \sigma_{11}$$

$$\Rightarrow l_c(\mu, \Sigma) = -n \log 2\pi - \frac{1}{2} n \log|\Sigma| - \frac{1}{2|\Sigma|}\sum_{j=1}^{n} \sigma_2^2 w_{j1}^2 + 2w_{j1}(\mu_2\sigma_{12} - \mu_1\sigma_2^2)$$
$$+ 2w_{j2}(\mu_1\sigma_{12} - \mu_2\sigma_1^2)$$
$$- 2w_{j1}w_{j2}\sigma_{12} + \sigma_1^2 w_{j2}^2$$

Expectation step:

$$\Rightarrow \cancel{E_{\theta^k}((\mu, \Sigma), (\mu^{(k)}, \Sigma^{(k)}))} \equiv Q(\theta, \theta^{(k)}) = E_{\theta^{(k)}}(l_c(\mu, \Sigma))$$
where $\theta^{(k)} = \{\mu_1^{(k)}, \mu_2^{(k)}, \sigma_1^{2(k)}, \sigma_{12}^{(k)}, \sigma_2^{2(k)}\}$

For first p observations:

$$E_{\theta^{(k)}}(l_c(\mu, \Sigma))_p = -n \log 2\pi - \frac{1}{2} n \log|\Sigma^{(k)}| - \frac{1}{2|\Sigma|}\sum_{j=1}^{p} \Big[E(w_{j1}^2 | w_{j2})\sigma_2^2 + 2E(w_{j1}|w_{j2}) \times (\mu_2\sigma_{12} - \mu_1\sigma_2^2)$$
$$+ 2w_{j2}(\mu_1\sigma_{12} - \mu_2\sigma_1^2)$$
$$- 2w_{j2}\sigma_{12} E(w_{j1}|w_{j2}) - \sigma_1^2 w_{j2}^2 \Big]$$

$$E(w_{j2}|w_{j2}) = w_{j2} + \frac{\sigma_{12}^{(k)}}{\sigma_2^{(k)2}}(w_{j2} - \mu_2^{(k)}) = w_{j1}^{(k)} \quad - ①$$

and $$E(w_{j1}^2|w_{j2}) = w_{j1}^{(k)2} + \sigma_2^{(k)2} \quad - ②$$

<u>for second q observations:</u>

$$E_{\theta^{(k)}}(l_c(\mu,\Sigma))_q = -n\log 2\pi - \frac{1}{2}n\log|\Sigma^{(k)}| - \frac{1}{2|\Sigma|}\sum_{j=p+1}^{p+q+1}\Big[\sigma_{22}^{(k)}w_{j1}^2 + 2w_{j1}(\mu_2^k\sigma_{12}^k - \mu_1^k\sigma_{22}^k)$$
$$+ 2E(w_{j2}|w_{j1})(\mu_1^k\sigma_{12}^k - \mu_2^k\sigma_{11}^k)$$
$$- 2E(w_{j2}|w_{j1})\,w_{j1}\sigma_{12}^k$$
$$+ \sigma_{11}^k E(w_{j2}^2|w_{j1})\Big]$$

similarily, $$E(w_{j2}|w_{j1}) = w_{j1} + \frac{\sigma_{12}^{(k)}}{\sigma_1^{(k)2}}(w_{j1} - \mu_1^{(k)}) = w_{j2}^{(k)} \quad - ③$$

$$E(w_{j2}^2|w_{j1}) = w_{j2}^{2(k)} + \sigma_1^{2(k)} \quad - ④$$

<u>for last r observations:</u>

$$E_{\theta^{(k)}}(l_c(\mu,\Sigma))_r = l_c(\mu^k,\Sigma^k)$$

Thus, $$Q(\theta,\theta^{(k)}) = E_{\theta^{(k)}}(l_c(\mu,\Sigma))_p + E_{\theta^{(k)}}(l_c(\mu,\Sigma))_r + l_c(\mu^k,\Sigma^k)$$

<u>M step:-</u>

differentiating $Q(\theta,\theta^{(k)})$ partially wrt $\theta^{(k)} = (\mu_1^k, \mu_2^k, \sigma_{12}^k, \sigma_1^{2(k)}, \sigma_2^{2(k)})$

$$\Rightarrow \mu_1^{(k+1)} = \sum_{j=1}^{n} w_{j1}^{(k)}/n \ ; \quad \mu_2^{(k+1)} = \sum_{j=1}^{n} w_{j2}/n$$

$$\sigma_1^{2(k+1)} = \sum_{j=1}^{n} w_{j1}^{(k)2} - \frac{1}{n}\Big[\sum_{j=1}^{n} w_{j1}^{(k)}\Big]^2$$

$$\sigma_2^{2(k+1)} = \sum_{j=1}^{n} w_{j2}^{(k)2} - \frac{1}{n}\Big[\sum_{j=1}^{n} w_{j2}^{(k)}\Big]^2$$

$$\sigma_1\sigma_2^{(k+1)} = \sum_{j=1}^{n} w_{j1}^k w_{j2}^k - \frac{1}{n}\sum_{j=1}^{n} w_{j1}^{(k)} \sum_{j=1}^{n} w_{j2}^{(k)}$$

, where, $w_{j1}^{(k)}$ and $w_{j2}^{(k)}$ are defined in equations ① and ③ respectively.

Ans. 3) $f(x) \propto e^{-x}$

$F(x) \propto 1 - e^{-x}$ (exponential distribution)

$F^{-1}(x) \propto -\log(1-x)$

$\because$ x is conditioned, such that $x \in (0,2)$

$x \to A + (B-A)u$, where $A = F(0)$, $B = F(1)$, $u \sim unif(0,1)$

$\Rightarrow F^{-1}(A + (B-A)u) \propto -\log(1 - (1-e^{-2})u)$

Thus, generating 5000 samples from the above function, estimated density plot is provided.

Ans. 4) To sample from g_1 and g_2, inverse transformation method is used.

To sample from $f(x)$ using g_1, $\alpha_1 = \sup \frac{q(x)}{g_1(x)} = \frac{e^{-x}}{(1+x^2) \cdot e^{-x}} = \frac{1}{1+x^2}$

$\Rightarrow \alpha_1 = 1$

To sample from $f(x)$ using g_2, $\alpha_2 = \sup \frac{q(x)}{g_2(x)} = \sup\left(\frac{\pi}{2} e^{-x}\right) = \pi/2$

$\Rightarrow \alpha_2 = \pi/2$

Algorithm and results are depicted on the page attached.

Ans. 5) a) Let $g(x) = c\,(2x^{\theta-1} + x^{\theta-1/2})\,e^{-x}$, $x>0$

where, c is the normalising constant

$\Rightarrow \int g(x)dx = 1$ (as $g(x)$ is a density function)

$\Rightarrow c\int (2x^{\theta-1} + x^{\theta-1/2})\, e^{-x}\, dx = 1$

$\Rightarrow 2c\int x^{\theta-1} e^{-x} dx + c\int x^{(\theta+1/2)-1} e^{-x} dx = 1$

$\Rightarrow 2c\,\Gamma(\theta)\underbrace{\int \text{Gamma}(\theta,1)dx}_{1} + c\underbrace{\int \text{Gamma}(\theta+1/2,1)\,dx}_{1} \cdot \Gamma(\theta+1/2) = 1$

$\Rightarrow c = \dfrac{1}{2\Gamma(\theta) + \Gamma(\theta+1/2)}$, where $\Gamma(\theta)$ is the gamma function.

b) $g(x) = c\,(2x^{\theta-1} + x^{\theta-1/2})\,e^{-x}$

$= \dfrac{c.2\,\Gamma(\theta)}{\Gamma(\theta)}\, x^{\theta-1} e^{-x} + \dfrac{c.\,\Gamma(\theta+1/2)}{\Gamma(\theta+1/2)}\, x^{(\theta+1/2)-1} e^{-x}$

$= 2c\,\Gamma(\theta)\,\text{Gamma}(\theta,1) + c\,\Gamma(\theta+1/2).\text{Gamma}(\theta+1/2,1)$

Thus, $g(x)$ is a mixture of two gamma distributions, with weight

$\Rightarrow w_1 = 2c\,\Gamma(\theta) = \dfrac{2\Gamma(\theta)}{2\Gamma(\theta)+\Gamma(\theta+1/2)}$

and

$w_2 = c\,\Gamma(\theta+1/2) = \dfrac{\Gamma(\theta+1/2)}{2\Gamma(\theta)+\Gamma(\theta+1/2)}$

(c) $g(x) = w_1 \text{ gamma}(\theta, 1) + w_2 \text{ gamma}(\theta + 1/2, 1)$

To sample from $g(x)$, the following simple method can be followed

1) Generate a bernoulli trial with probability w_1
2) If success, generate from gamma$(\theta, 1)$, otherwise, generate from gamma$(\theta + 1/2, 1)$
3) Repeat the above steps n times, to generate n random samples.

In order to generate samples from a gamma distribution, statistical packages can be used straight away.

(d) $g(x) = w_1 \text{ gamma}(\theta, 1) + w_2 \text{ gamma}(\theta + 1/2, 1)$

$q(x) = \sqrt{4+x}\, x^{\theta - 1} e^{-x}$

$$\alpha = \sup \frac{q(x)}{g(x)} = \sup \frac{\sqrt{4+x} \cdot x^{\theta-1} e^{-x}}{w_1 \frac{x^{\theta-1} e^{-x}}{\Gamma(\theta)} + w_2 \frac{x^{\theta - 1/2} e^{-x}}{\Gamma(\theta + 1/2)}}$$

$$= \sup \frac{\sqrt{4+x}}{\frac{w_1}{\Gamma(\theta)} + \frac{w_2 \sqrt{x}}{\Gamma(\theta + 1/2)}}$$

Thus, for $\theta = 0.5$, $\alpha = 4.55$

$\theta = 1$, $\alpha = 3.68$

$\theta = 1.5$, $\alpha = 2.78$

The algorithm is designed in R and plots are attached

Ans.6) $f(x,y) \propto x^{\alpha} y$; $x>0,\ y>0$ and $x^2+y^2 \leq 1$

$\Rightarrow q(x,y) = x^{\alpha} y$

let the auxiliary function be $g(x,y) = x^{\alpha}\cdot(\alpha+1)$; $x \in (0,1]$

Thus, $\alpha' = \sup \dfrac{q(x,y)}{g(x,y)}$

$= \sup \dfrac{x^{\alpha} y}{(\alpha+1)x^{\alpha}} = \sup \dfrac{y}{\alpha+1}$

$\because y \leq 1 \Rightarrow \alpha' = \dfrac{1}{\alpha+1}$

Hence, $\alpha' g(x,y) = x^{\alpha} \geq x^{\alpha} y$ (as $y \leq 1$)

$\Rightarrow \alpha' g(x,y) \geq q(x,y)$

Thus, rejection sampling can be used considering $q(x,y) = x^{\alpha}$

and $g(x,y) = x^{\alpha}\cdot(\alpha+1)$

Step 1) Sample $g(x,y)$ or $g(x)$ using inverse transform method

$\because g(x) = (\alpha+1)x^{\alpha}$; $G(x) = x^{\alpha+2}$ [C.D.F]

$X = G^{-1}(u) = \dfrac{\log u}{\alpha+2}$; $u \sim$ unif $[0,1]$

and sample $u \sim$ unif $[0,1]$

Step 2) let val $= q(x,y) / \alpha' g(x,y)$

If $u >$ val $\rightarrow$ go to step 1

else add X to the sample.

Step 3) Repeat 1 and 2 as many times required.