

EM Algorithm and Simulation

Surya Teja Eada and Kristen Sandberg

Abstract

This project has two components. Firstly, we derive the updating rules in the construction of the EM algorithm in application to maximum likelihood estimation in finite mixture regression models. Secondly, we apply the rejection sampling algorithm to simulate from few complicated density functions that are in a way similar to a mixture of Gamma densities and Beta densities

1 Expectation - Maximization

In this question, we will derive the updating rules in the given algorithm based on the construction of an EM algorithm for a finite mixture regression model. Suppose the density of y_i (conditional on \mathbf{x}_i), is given by $f(y_i|\mathbf{x}_i, \Psi) = \sum_{j=1}^m \pi_j \phi(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma^2)$, for $i = 1, \dots, n$, where π_j s are called mixing proportions, $\boldsymbol{\beta}_j$ is the regression coefficient vector for the j th group, $\phi(\cdot; \mu, \sigma^2)$ denotes the density function of $N(\mu, \sigma^2)$, and $\Psi = (\pi_1, \boldsymbol{\beta}_1, \dots, \pi_m, \boldsymbol{\beta}_m, \sigma)^T$ collects all the unknown parameters. Let $\hat{\Psi}_{\text{mle}} = \arg \max_{\Psi} \sum_{i=1}^n \log \{ \sum_{j=1}^m \pi_j \phi(y_i; \mathbf{x}_i^T \boldsymbol{\beta}_j, \sigma^2) \}$ and

$$z_{ij} = \begin{cases} 1 & \text{if } i\text{th observation is from } j\text{th component;} \\ 0 & \text{otherwise.} \end{cases}$$

1.1 E-Step

In the E-Step of the EM algorithm, we compute the conditional expectation of $l_n^c(\Psi)$, where $l_n^c(\Psi)$ is given in equation (1).

$$l_n^c(\Psi) = \sum_{i=1}^n \sum_{j=1}^m z_{ij} \log \{ \pi_j \phi(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j; 0; \sigma^2) \} \quad (1)$$

Observe,

$$\begin{aligned} Q(\Psi|\Psi^{(k)}) &= \mathbb{E}_z \{ l_n^c(\Psi) \} \\ &= \mathbb{E}_z \left[\sum_{i=1}^n \sum_{j=1}^m z_{ij} \log \{ \pi_j \phi(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j; 0; \sigma^2) \} \right] \\ &= \sum_{i=1}^n \sum_{j=1}^m \log \{ \pi_j \phi(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j; 0; \sigma^2) \} \mathbb{E}_z \left[z_{ij} | y_i, \mathbf{x}_i; \Psi^{(k)} \right] \\ &= \sum_{i=1}^n \sum_{j=1}^m \log \{ \pi_j \phi(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j; 0; \sigma^2) \} \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \log \{ \pi_j \phi(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j; 0; \sigma^2) \}. \end{aligned}$$

Here p_{ij} represents the probability that the observation belongs to component j . Since this is a finite mixture of normal distributions, we first consider the probability under model j , which is $\pi_j \phi(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k)}; 0, \sigma^2)$. This is the normal density multiplied by the mixing proportion. We then normalize this to sum to one by dividing this by the summation of probabilities over all components j . Thus, we have

$$p_{ij} = \frac{\pi_j \phi(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k)}; 0, \sigma^2)}{\sum_{j=1}^m \pi_j \phi(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k)}; 0, \sigma^2)}. \quad (2)$$

Therefore, $\sum_{j=1}^m p_{ij} = 1$.

1.2 M-Step

For the M-Step, we want to maximize $Q(\Psi|\Psi^{(k)})$ to obtain $(\boldsymbol{\beta}^{(k+1)}, \sigma^{2(k+1)})$, so we will compute the following partial derivatives, set them equal to zero, and solve for each variable individually.

1.3 M-Step (2.a) Compute $\pi_j^{(k+1)}$

First, we will show that $\pi_j^{(k+1)} = \frac{\sum_{i=1}^n p_{ij}^{(k+1)}}{n}$. Recall that the mixing proportions, π_1, \dots, π_m must satisfy $\pi_1 + \dots + \pi_m = 1$. Therefore, we will find the maximum by finding a solution for $\mathcal{L}'(\pi_1, \dots, \pi_m) = 0$ where $\mathcal{L}(\pi_1, \dots, \pi_m) = \sum_{i=1}^n \sum_{j=1}^m z_{ij} \log \{\pi_j \phi(y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j; 0; \sigma^2)\} - \lambda \left[\sum_{j=1}^m \pi_j - 1 \right]$ with λ a Lagrange multiplier. Observe,

$$\frac{\partial \mathcal{L}}{\partial \pi_j} = \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \frac{1}{\pi_j^{(k+1)}} - \lambda. \quad (3)$$

Now recall that $\sum_{j=1}^m p_{ij} = 1$ and $\sum_{j=1}^m \pi_j = 1$, so now equation (3) becomes

$$\frac{\partial \mathcal{L}}{\partial \pi_j^{(k+1)}} = n - \lambda. \quad (4)$$

We set equation (4) equal to zero, subtract λ from both sides, and find that $\lambda = n$. Thus, from equation (3) we have

$$\frac{\partial \mathcal{L}}{\partial \pi_j^{(k+1)}} = \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \frac{1}{\pi_j^{(k+1)}} - n. \quad (5)$$

We set equation (5) equal to zero, subtract n from both sides, and we have

$$n = \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \frac{1}{\pi_j^{(k+1)}},$$

which implies that

$$\pi_j^{(k+1)} = \frac{\sum_{i=1}^n p_{ij}^{(k+1)}}{n}. \quad (6)$$

1.4 M-Step (2.b) Compute $\beta_j^{(k+1)}$

Second, we will show that $\beta_j^{(k+1)} = \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T p_{ij}^{(k+1)} \right)^{-1} \left(\sum_{i=1}^n \mathbf{x}_i p_{ij}^{(k+1)} y_i \right)$ for $j = 1, \dots, m$.

First, we will rewrite $Q(\Psi|\Psi^{(k)})$ as follows:

$$\begin{aligned} Q(\Psi|\Psi^{(k)}) &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left\{ \log \pi_j + \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} (y_i - \mathbf{x}_i^T \beta_j^{(k+1)})^T (y_i - \mathbf{x}_i^T \beta_j^{(k+1)}) \right\} \right] \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left\{ \log \pi_j + \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} (y_i^T - (\mathbf{x}_i^T \beta_j^{(k+1)})^T) (y_i - \mathbf{x}_i^T \beta_j^{(k+1)}) \right\} \right] \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left\{ \log \pi_j + \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} (y_i^T - (\beta_j^{(k+1)})^T \mathbf{x}_i) (y_i - \mathbf{x}_i^T \beta_j^{(k+1)}) \right\} \right] \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left\{ \log \pi_j + \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} (y_i^T y_i - (\beta_j^{(k+1)})^T \mathbf{x}_i y_i - y_i^T \mathbf{x}_i^T \beta_j^{(k+1)} + (\beta_j^{(k+1)})^T \mathbf{x}_i \mathbf{x}_i^T \beta_j^{(k+1)}) \right\} \right] \right\} \end{aligned}$$

Now, we note that y_i is not a vector, so $y_i^T = y_i$. Also, we will need to know

$$\frac{\partial \left(\beta_j^{(k+1)} \mathbf{x}_i y_i \right)}{\partial \beta_j^{(k+1)}} = \frac{\partial \left((\mathbf{x}_i y_i)^T \beta_j^{(k+1)} \right)}{\partial \beta_j^{(k+1)}} = (\mathbf{x}_i y_i)^T = y_i^T \mathbf{x}_i^T = y_i \mathbf{x}_i^T,$$

and

$$\frac{\partial \left(\left(\beta_j^{(k+1)} \right)^T \mathbf{x}_i \mathbf{x}_i^T \beta_j^{(k+1)} \right)}{\partial \beta_j^{(k+1)}} = \left(\beta_j^{(k+1)} \right)^T \left[\mathbf{x}_i \mathbf{x}_i^T + (\mathbf{x}_i \mathbf{x}_i^T)^T \right] = \left(\beta_j^{(k+1)} \right)^T \left[\mathbf{x}_i \mathbf{x}_i^T + \mathbf{x}_i \mathbf{x}_i^T \right] = 2 \left(\beta_j^{(k+1)} \right)^T \mathbf{x}_i \mathbf{x}_i^T$$

Thus, the partial derivative of $Q(\Psi|\Psi^{(k)})$ with respect to $\beta_j^{(k+1)}$ is:

$$\begin{aligned} \frac{\partial Q(\Psi|\Psi^{(k)})}{\partial \beta_j^{(k+1)}} &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left(\frac{\sqrt{2\pi\sigma^2}}{\exp \left\{ \frac{-1}{2\sigma^2} (y_i - \mathbf{x}_i^T \beta_j^{(k+1)})^T (y_i - \mathbf{x}_i^T \beta_j^{(k+1)}) \right\}} \right) \left(\frac{1}{\sqrt{2\pi\sigma^2}} \left(\frac{-1}{2\sigma^2} (0 - y_i \mathbf{x}_i^T - y_i \mathbf{x}_i^T \right. \right. \\ &\quad \left. \left. + 2 \left(\beta_j^{(k+1)} \right)^T \mathbf{x}_i \mathbf{x}_i^T \right) \exp \left\{ \frac{-1}{2\sigma^2} (y_i - \mathbf{x}_i^T \beta_j^{(k+1)})^T (y_i - \mathbf{x}_i^T \beta_j^{(k+1)}) \right\} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left(\frac{-1}{2\sigma^2} \right) \left(-2 y_i \mathbf{x}_i^T + 2 \left(\beta_j^{(k+1)} \right)^T \mathbf{x}_i \mathbf{x}_i^T \right) \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left(\frac{1}{\sigma^2} \right) \left(y_i \mathbf{x}_i^T - \left(\beta_j^{(k+1)} \right)^T \mathbf{x}_i \mathbf{x}_i^T \right). \end{aligned}$$

We set this partial derivative equal to 0, multiply both sides by σ^2 , and add $\left(\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left(\beta_j^{(k+1)} \right)^T \mathbf{x}_i \mathbf{x}_i^T \right)$ to both sides to obtain:

$$\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left(\beta_j^{(k+1)} \right)^T \mathbf{x}_i \mathbf{x}_i^T = \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} y_i \mathbf{x}_i^T$$

Since $\beta_j^{(k+1)}$ does not depend on i and $p_{ij}^{(k+1)}$ is a scalar, we can move $\left(\beta_j^{(k+1)} \right)^T$ outside of the summation, and we have:

$$\left(\beta_j^{(k+1)} \right)^T \sum_{i=1}^n p_{ij}^{(k+1)} \mathbf{x}_i \mathbf{x}_i^T = \sum_{i=1}^n p_{ij}^{(k+1)} y_i \mathbf{x}_i^T,$$

for $j = 1, \dots, m$. Now we right multiply by $\left(\sum_{i=1}^n p_{ij}^{(k+1)} \mathbf{x}_i \mathbf{x}_i^T\right)^{-1}$ to obtain:

$$\left(\beta_j^{(k+1)}\right)^T = \left(\sum_{i=1}^n p_{ij}^{(k+1)} y_i \mathbf{x}_i^T\right) \left(\sum_{i=1}^n p_{ij}^{(k+1)} \mathbf{x}_i \mathbf{x}_i^T\right)^{-1}.$$

Now, we would like the formula for $\beta_j^{(k+1)}$, so we will take the transpose of both sides as we have:

$$\begin{aligned} \beta_j^{(k+1)} &= \left[\left(\sum_{i=1}^n p_{ij}^{(k+1)} y_i \mathbf{x}_i^T\right) \left(\sum_{i=1}^n p_{ij}^{(k+1)} \mathbf{x}_i \mathbf{x}_i^T\right)^{-1} \right]^T \\ &= \left[\left(\sum_{i=1}^n p_{ij}^{(k+1)} \mathbf{x}_i \mathbf{x}_i^T\right)^{-1} \right]^T \left[\left(\sum_{i=1}^n p_{ij}^{(k+1)} y_i \mathbf{x}_i^T\right) \right]^T \\ &= \left[\left(\sum_{i=1}^n p_{ij}^{(k+1)} \mathbf{x}_i \mathbf{x}_i^T\right)^T \right]^{-1} \left[\sum_{i=1}^n \left(p_{ij}^{(k+1)} y_i \mathbf{x}_i^T\right)^T \right] \\ &= \left[\sum_{i=1}^n \left(p_{ij}^{(k+1)} \mathbf{x}_i \mathbf{x}_i^T\right)^T \right]^{-1} \left[\sum_{i=1}^n \mathbf{x}_i \left(p_{ij}^{(k+1)}\right)^T y_i^T \right] \\ &= \left[\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T p_{ij}^{(k+1)} \right]^{-1} \left[\sum_{i=1}^n \mathbf{x}_i p_{ij}^{(k+1)} y_i \right] \end{aligned}$$

for $j = 1, \dots, m$.

1.5 M-Step (2.c) Compute $\sigma^{2(k+1)}$

Third, we will show that $\sigma^{2(k+1)} = \frac{\sum_{j=1}^m \sum_{i=1}^n p_{ij}^{(k+1)} (y_i - \mathbf{x}_i^T \beta_j^{(k+1)})^2}{n}$.

Let $\phi(\cdot; \mu, \sigma^2)$ denote the density function of $N(\mu, \sigma^2)$. Thus, $\phi(y_i - \mathbf{x}_i^T \beta_j^k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-1}{2\sigma^2} (y_i - \mathbf{x}_i^T \beta_j^k - 0)^2\right\}$. Therefore, we have

$$Q(\Psi|\Psi^{(k)}) = \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \log \left\{ \pi_j \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} (y_i - \mathbf{x}_i^T \beta_j^k - 0)^2 \right\} \right\}.$$

When we take the partial derivative of $Q(\Psi|\Psi^{(k)})$ with respect to $\sigma^{2(k+1)}$, we have

$$\begin{aligned} \frac{\partial Q}{\partial \sigma^{2(k+1)}} &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left(\frac{\sqrt{2\pi\sigma^{2(k+1)}}}{\exp \left[\frac{-1}{2\sigma^{2(k+1)}} (y_i - \mathbf{x}_i^T \beta_j^{(k)})^2 \right]} \right) \left(\frac{-\pi}{(2\pi\sigma^{2(k+1)})^{(3/2)}} \exp \left[\frac{-1}{2\sigma^{2(k+1)}} (y_i - \mathbf{x}_i^T \beta_j^{(k)})^2 \right] \right) \\ &\quad + \frac{1}{\sqrt{2\pi\sigma^{2(k+1)}}} \left(\frac{1}{2(\sigma^{2(k+1)})^2} (y_i - \mathbf{x}_i^T \beta_j^{(k)})^2 \right) \exp \left[\frac{-1}{2\sigma^{2(k+1)}} (y_i - \mathbf{x}_i^T \beta_j^{(k)})^2 \right] \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left(\frac{-\pi}{2\pi\sigma^{2(k+1)}} + \frac{1}{2(\sigma^{2(k+1)})^2} (y_i - \mathbf{x}_i^T \beta_j^{(k)})^2 \right) \\ &= \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left(\frac{-1}{2\pi\sigma^{2(k+1)}} + \frac{1}{2(\sigma^{2(k+1)})^2} (y_i - \mathbf{x}_i^T \beta_j^{(k)})^2 \right). \end{aligned}$$

Now, we can set this partial derivative equal to zero and solve for $\sigma^{2(k+1)}$. Thus, we have

$$0 = \frac{1}{2\sigma^{2(k+1)}} \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} \left(-1 + \frac{1}{\sigma^{2(k+1)}} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k)})^2 \right)$$

We can multiply both sides by $2\sigma^{2(k+1)}$ and add $\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)}$ to both sides. Therefore, $\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} = \frac{1}{\sigma^{2(k+1)}} \sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k)})^2$. Lastly, we multiply both sides by σ^2 and divide by $\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)}$. This gives us

$$\sigma^{2(k+1)} = \frac{\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k)})^2}{\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)}}.$$

Again, since $\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} = n$, we can express this as

$$\sigma^{2(k+1)} = \frac{\sum_{i=1}^n \sum_{j=1}^m p_{ij}^{(k+1)} (y_i - \mathbf{x}_i^T \boldsymbol{\beta}_j^{(k)})^2}{n}.$$

2 Using Gamma Mixture Model in Rejection Sampling

In this question, we want to create a procedure that simulates observations from a complex density function $f(x)$, as given in equation (7). For this task, we utilize the rejection sampling method with the instrumental density function $g(x)$, as given in (8). The idea is to simulate from $g(x)$, which will be observed to be a mixture of gamma density functions and then utilize the rejection sampling method.

$$f(x) \propto \sqrt{(4+x)}x^{\theta-1}e^{-x} \quad (7)$$

$$g(x) \propto (2x^{\theta-1} + x^{\theta-\frac{1}{2}})e^{-x} \quad (8)$$

First, we try to find the value of constant C in the density function $g(x)$.

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

Since $g(x)$ is a density function, its integral over all values of x will result to a constant value “1”. This implies:

$$\begin{aligned} \int_0^{\infty} g(x)dx &= 1 \\ \int_0^{\infty} C * (2x^{\theta-1} + x^{\theta-\frac{1}{2}})e^{-x}dx &= 1 \\ C \left[\int_0^{\infty} 2x^{\theta-1}e^{-x}dx + \int_0^{\infty} x^{\theta-\frac{1}{2}}e^{-x}dx \right] &= 1 \\ C \left[2\Gamma(\theta) \int_0^{\infty} \frac{x^{\theta-1}e^{-x}}{\Gamma(\theta)}dx + \Gamma(\theta + \frac{1}{2}) \int_0^{\infty} \frac{x^{(\theta+\frac{1}{2})-1}e^{-x}}{\Gamma(\theta + \frac{1}{2})}dx \right] &= 1 \end{aligned} \quad (9)$$

It can be observed that the above integral in equation (9) is a mixture of two gamma density integral functions, namely $\Gamma(\text{rate} = 1, \text{shape} = \theta)$ and $\Gamma(\text{rate} = 1, \text{shape} = \theta + \frac{1}{2})$. Integrals of gamma density functions is equal to “1”, results in the following value for C.

$$\begin{aligned} C \left[2\Gamma(\theta) \int_0^{\infty} \frac{x^{\theta-1}e^{-x}}{\Gamma(\theta)}dx + \Gamma(\theta + \frac{1}{2}) \int_0^{\infty} \frac{x^{(\theta+\frac{1}{2})-1}e^{-x}}{\Gamma(\theta + \frac{1}{2})}dx \right] &= 1 \\ C \left[2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2}) \right] &= 1 \\ C &= \left[\frac{1}{2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2})} \right] \end{aligned}$$

So, the density function $g(x)$ is given as follows:

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

We observe that $g(x)$ is a mixture density function of two gamma densities, $\Gamma(\text{rate} = 1, \text{shape} = \theta)$ and $\Gamma(\text{rate} = 1, \text{shape} = \theta + \frac{1}{2})$ with weights $\alpha_1 = \frac{2\Gamma(\theta)}{2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2})}$, $\alpha_2 = \frac{\Gamma(\theta + \frac{1}{2})}{2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2})}$ s.t.

$$\alpha_1 + \alpha_2 = 1$$

2.1 Simulation from ‘g(x)’ density

In order to simulate samples from $g(x)$ mixture density, we need to sample from $\Gamma(\text{rate} = 1, \text{shape} = \theta)$ and $\Gamma(\text{rate} = 1, \text{shape} = \theta + \frac{1}{2})$ with different probabilities. We also simulate 10000 from this mixture $g(x)$ density when $\theta = 2$.

```
## simulating using theta = 2 values
## alpha in this code refers to probability that describes which distribution
## among the mixture does an observation belong to and not prob of acceptance

simulate_g <- function(sample_size, theta) {
  alpha_0 = 2*gamma(theta)/(2*gamma(theta) + gamma(theta + (1/2)))
  indicator = rbinom(sample_size, 1, alpha_0)
  simulate_g = numeric(sample_size)
  for (i in 1:sample_size) {
    if (indicator[i] == 1) {
      simulate_g[i] <- rgamma(1, shape = theta, rate = 1)
    }
    else { simulate_g[i] <- rgamma(1, shape = theta + (1/2), rate = 1)}
  }
  return(simulate_g)
}

sample_size_sol <- 10000
theta_sol <- 2
g_sample <- simulate_g(sample_size_sol, theta_sol)
alpha = 2*gamma(theta_sol)/(2*gamma(theta_sol) + gamma(theta_sol + (1/2)))

plot(density(g_sample), col = "red")
curve(alpha*dgamma(x, shape = 2, rate = 1) + (1-alpha)*dgamma(x, shape = 5/2, rate = 1),
      xlim = c(0, max(g_sample)), add = TRUE, col = "green")
legend("topright", legend = c("Kernel Density", "True Density"), col = c("red", "green"),
      lty = c(1, 1))

## source("Simulate_Gamma_Mixture.R")
## source("g_Kernel_Density_Plot.R")
```

The kernel density is a good approximation of the true density when 10000 sample points have been simulated from $g(x)$, when $\theta = 2$ is utilized and this can be observed from Figure (1).

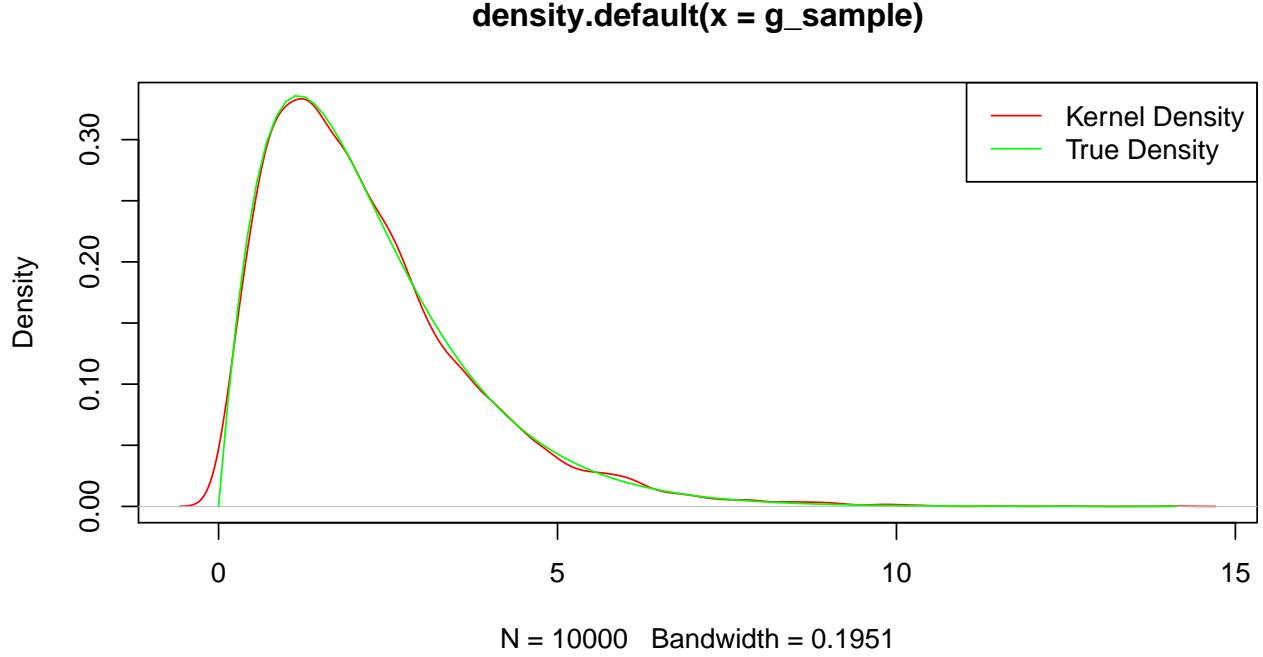


Figure 1: Kernel Density Estimation and True Density of $g(x)$ density function

2.2 Utilizing Rejection Sampling to sample from ‘f(x)’ density

Our ultimate goal is to sample from $f(x)$, a slightly complicated density function. We use rejection sampling for this purpose with instrumental density as $g(x)$. Firstly, we show that $f(x)$ is proportional to $q(x)$ which is less than or equal to $g(x)$

$$\begin{aligned}
 f(x) &\propto \sqrt{(4+x)}x^{\theta-1}e^{-x} \\
 \text{Let } q(x) &= \sqrt{(4+x)}x^{\theta-1}e^{-x} \\
 q(x) &\leq (\sqrt{4} + \sqrt{x})x^{\theta-1}e^{-x} \\
 q(x) &\leq (2x^{\theta-1} + x^{\theta-\frac{1}{2}})e^{-x} \\
 q(x) &\leq (1/C) * g(x) \\
 \text{where } C &= \left[\frac{1}{2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2})} \right] \\
 \alpha &= 1/C \\
 \alpha &= 2\Gamma(\theta) + \Gamma(\theta + \frac{1}{2}) \\
 \alpha g(x) &= (2x^{\theta-1} + x^{\theta-\frac{1}{2}})e^{-x}
 \end{aligned}$$

Now, we design the procedure to sample from $f(x)$ by following the rejection sampling procedure. Then, we plot the estimated density of $f(x)$. It is difficult to plot the actual density of $f(x)$ without evaluating the normalizing constant value that makes it a density.


```

source("Simulate_Gamma_Mixture.R")

simulate_f <- function(sample_size, theta) {
  count <- 0
  count_rejection <- 0
  simulate_f <- numeric(0)
  while(count < sample_size) {
    g_sample <- simulate_g(1, theta)
    uniform <- runif(1,0,1)
    q_x <- sqrt(4 + g_sample)*((g_sample)^(theta-1))*exp(-1*g_sample)
    alpha_g_x <- 2*((g_sample)^(theta-1))*exp(-1*g_sample) +
      ((g_sample)^(theta-(1/2)))*exp(-1*g_sample)
    ratio <- (q_x)/(alpha_g_x)
    if (uniform <= ratio) {
      simulate_f <- append(simulate_f, g_sample)
      count <- count + 1
    }
    else {
      count_rejection <- count_rejection + 1
    }
  }
  acceptance_rate = count/(count + count_rejection)
  return(simulate_f)
}

sample_size_sol <- 10000
theta_sol <- 2
f_sample <- simulate_f(sample_size_sol, theta_sol)

plot(density(f_sample), col = "red")
legend("topright", legend = c("Estimated Density of f(x)"), col = c("red"),
      lty = c(1))

```

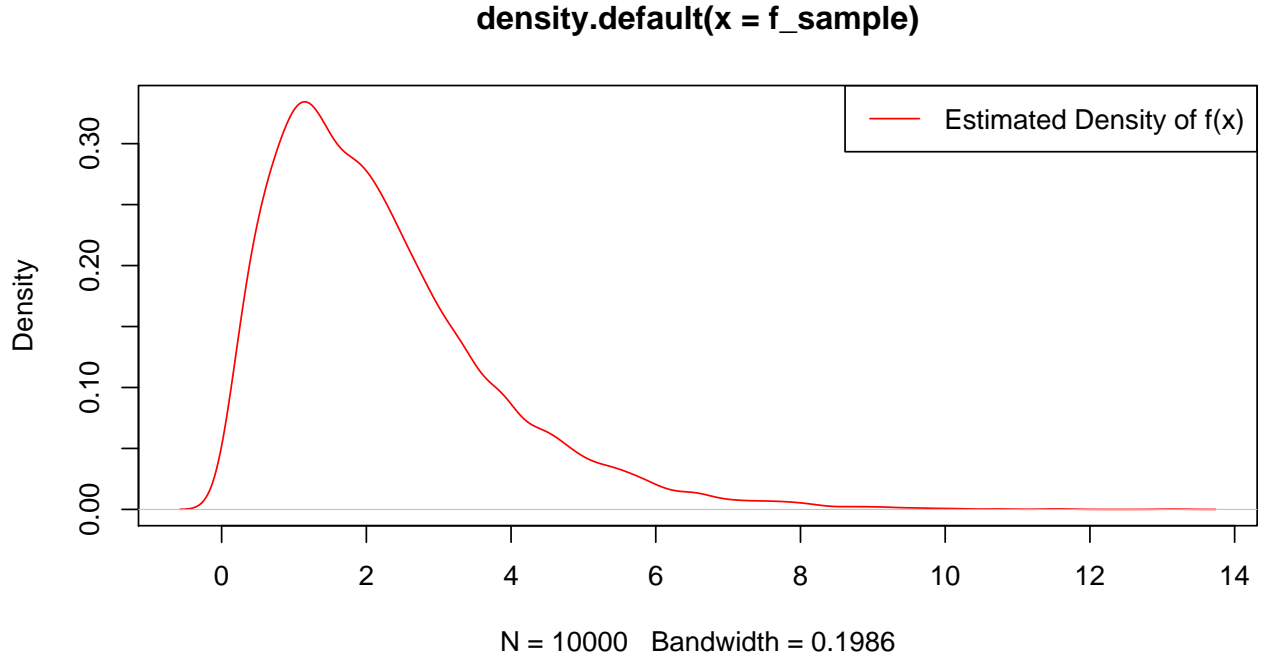


Figure 2: Estimated density of $f(x)$ density function

10000 sample points makes a good estimation of density curve. Based on observing the estimated density curve, we can infer that $f(x)$ has a similar curve to $g(x)$. This also implies a great acceptance probability and an efficient sampling. We evaluated the acceptance rate of the rejection sampling technique when $\theta = 2$ and using 10000 sample points, multiple times and observed that there is approximately 73% acceptance rate. We also evaluated $\frac{1}{\alpha} \int_{\mathcal{X}} q(x)$ and we obtained approximately 73% acceptance rate using integration between 0 and 100, implicating that the sampling procedure was reasonable.

3 Using Beta Mixture Model in Rejection Sampling

In this question, we want to simulate and generate observations from an $f(x)$ density as given in Equation (10), a complicated density that looks similar to a sum of beta densities.

$$f(x) \propto \frac{x^{\theta-1}}{1+x^2} + \sqrt{2+x^2}(1-x)^{\beta-1} \quad (10)$$

3.1 Procedure 1 - Utilizing mixture of Beta distributions

Because, $(1+x^2) \geq 1$ is in the denominator of the first term, and $\sqrt{2+x^2} \leq \sqrt{2} + x$, we can show that $q(x)$ is less than a mixture of beta distributions. Thus we have the following:

$$\begin{aligned} f(x) &\propto \frac{x^{\theta-1}}{1+x^2} + \sqrt{2+x^2}(1-x)^{\beta-1} \\ q(x) &= \frac{x^{\theta-1}}{1+x^2} + \sqrt{2+x^2}(1-x)^{\beta-1} \\ q(x) &\leq x^{\theta-1} + \sqrt{2}(1-x)^{\beta-1} + \sqrt{x^2}(1-x)^{\beta-1} \\ q(x) &\leq \alpha g(x) \end{aligned}$$

$$\begin{aligned} \text{Let } g(x) &\propto x^{\theta-1} + \sqrt{2}(1-x)^{\beta-1} + x(1-x)^{\beta-1} \\ g(x) &= C \left[B(\theta, 1) \frac{x^{\theta-1}}{B(\theta, 1)} + \sqrt{2} B(1, \beta) \frac{(1-x)^{\beta-1}}{B(1, \beta)} + B(2, \beta) \frac{x(1-x)^{\beta-1}}{B(2, \beta)} \right] \\ \text{Since } \int_0^1 g(x) &= 1 \\ C [B(\theta, 1) + B(1, \beta) + B(2, \beta)] &= 1 \\ C &= \frac{1}{B(\theta, 1) + B(1, \beta) + B(2, \beta)} \\ \text{Therefore, } \alpha &= \frac{1}{C} \\ \alpha &= B(\theta, 1) + B(1, \beta) + B(2, \beta) \end{aligned}$$

We've shown that $q(x)$ is less than $\alpha g(x)$, where $\alpha = B(\theta, 1) + B(1, \beta) + B(2, \beta)$. Now, we can simulate using the mixture of beta distributions and the rejection sampling.

```
source("simulate_beta_mix.R")

## We apply the function simulate_f_q3 that simulates samples from f and has
## the following variables: sample_size, c_theta, c_beta
## We chose c_theta = 2, c_beta = 4 for our question

q <- function(x) {
  return((x/(1+x^2)) + sqrt(2 + x^2)*((1-x)^3)) }
## integrate(q,0,1) = 0.7058734
```

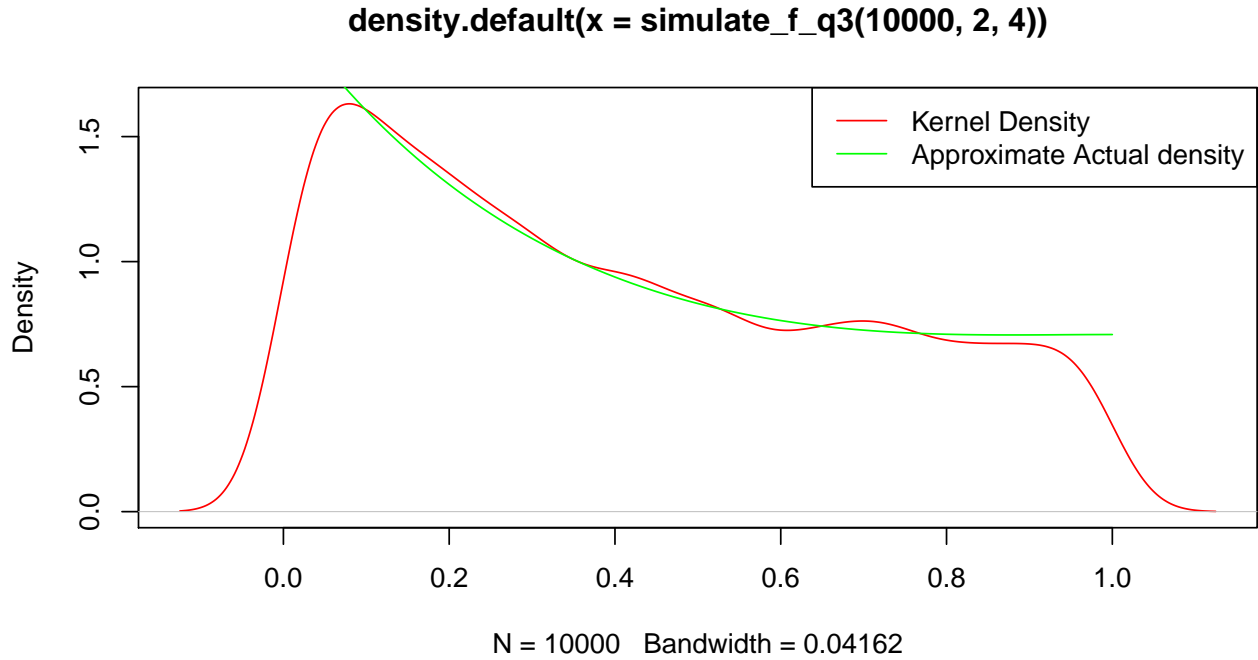


Figure 3: Estimated density of $f(x)$ density function using mixture of beta densities

```
plot(density(simulate_f_q3(10000, 2, 4)), col = "red")
curve(q(x)/0.7058734, xlim = c(0,1), add = TRUE, col = "green")
legend("topright", legend = c("Kernel Density", "Approximate Actual density"),
      col = c("red", "green"), lty = c(1,1))
```

This simulates $f(x)$, however, it can be observed from the plot that contained both the derived density from the simulation and the curve proportional to actual density, that the estimate is slightly perturbed at the end points in terms of shape. While $f(x)$ is proportional to a constantly decreasing function, estimated density has a slight increase at 0 before it decreases and a slight decreasing movement at 1. We use a different procedure as well to simulate and is provided in the next section.

3.2 Procedure 2 - Breaking $f(x)$ into pieces that can be simulated using Beta distributions

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

$$f(x) \propto \sum_{i=1}^2 q_i(x)$$

$$\text{where } q_1(x) = \frac{x^{\theta-1}}{1+x^2} \quad q_2(x) = \sqrt{2+x^2}(1-x)^{\beta-1}$$

$$q_1(x) \leq B(\theta, 1) \frac{x^{\theta-1}}{B(\theta, 1)}$$

$$g_1(x) = \frac{x^{\theta-1}}{B(\theta, 1)}$$

$$\alpha_1 = B(\theta, 1)$$

$$q_2(x) \leq \sqrt{2}(1-x)^{\beta-1} + x(1-x)^{\beta-1}$$

$$q_2(x) \leq \sqrt{2}B(1, \beta) \frac{(1-x)^{\beta-1}}{B(1, \beta)} + B(2, \beta) \frac{x(1-x)^{\beta-1}}{B(2, \beta)}$$

$$q_2(x) \leq \left(\sqrt{2}B(1, \beta) + B(2, \beta) \right) \left(\frac{\sqrt{2}B(1, \beta)}{\sqrt{2}B(1, \beta) + B(2, \beta)} \frac{(1-x)^{\beta-1}}{B(1, \beta)} + \frac{B(2, \beta)}{\sqrt{2}B(1, \beta) + B(2, \beta)} \frac{x(1-x)^{\beta-1}}{B(2, \beta)} \right)$$

$$g_2(x) = \frac{\sqrt{2}B(1, \beta)}{\sqrt{2}B(1, \beta) + B(2, \beta)} \frac{(1-x)^{\beta-1}}{B(1, \beta)} + \frac{B(2, \beta)}{\sqrt{2}B(1, \beta) + B(2, \beta)} \frac{x(1-x)^{\beta-1}}{B(2, \beta)}$$

$$\alpha_2 = \sqrt{2}B(1, \beta) + B(2, \beta)$$

In this procedure, we use simulations from a beta density that influences shape close to “0” and another mixture of beta densities that influences shape close to “1”.

```
source("3_2_Simulate_f.R")

## We apply the function simulate_f_q3 that simulates samples from f and has
## the following variables: sample_size, c_theta, c_beta
## We chose c_theta = 2, c_beta = 4 for our question

q <- function(x) {
  return((x/(1+x^2)) + sqrt(2 + x^2)*((1-x)^3)) }
## integrate(q,0,1) = 0.7058734

plot(density(simulate_f_q3_2(10000, 2, 4)), col = "red")
curve(q(x)/0.7058734, xlim = c(0,1), add = TRUE, col = "green")
legend("topright", legend = c("Kernel Density Procedure-2", "Approximate Actual Density"),
      col = c("red", "green"), lty = c(1,1))
```

Using procedure 2, we obtained the estimated density function again and we have also calculated the estimated acceptance rate which is 0.7845, which is higher than 0.77 obtained by Procedure 1. We observe the distortions at 0 and 1 in utilizing this procedure of simulation as well.

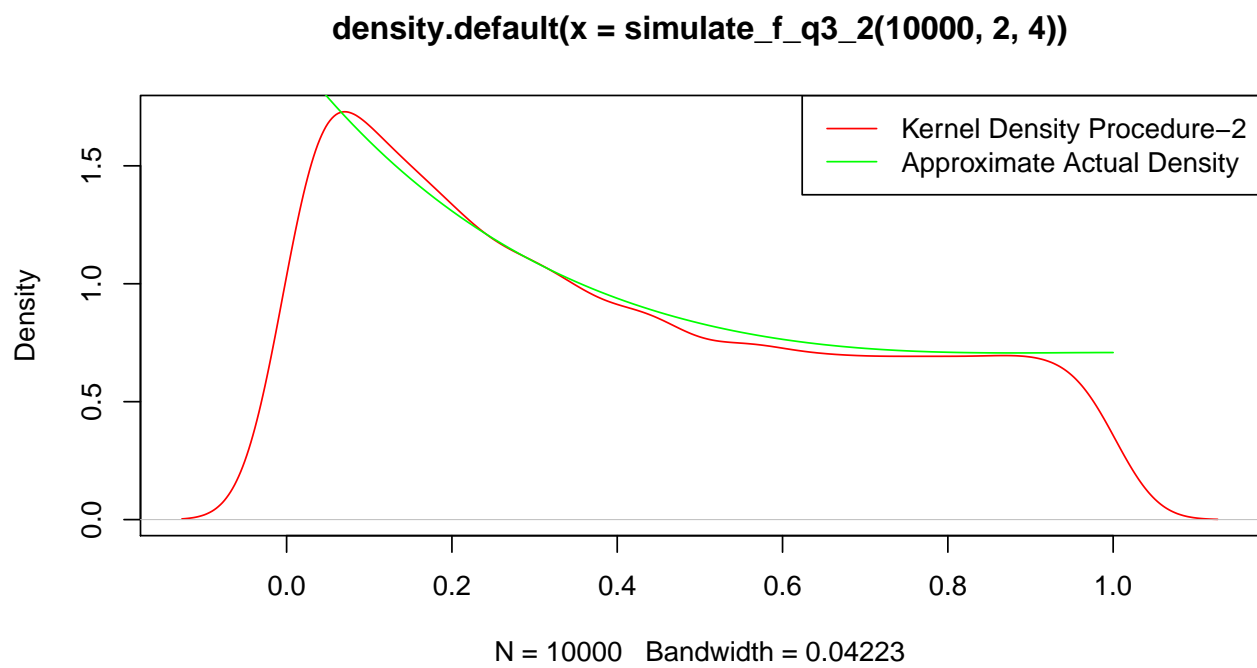


Figure 4: Estimated density of $f(x)$ density function using Procedure 2