Homework 3

Anurag Amin - Q.2 & Q.3(b) &
Michael Noyes - Q.1 & Q.3(a)

3/1/2018

Problem 1: E-M algorithm

The complete log-likelihood can be written as:

$$l_n^c(\Psi) = \sum_{i=1}^{n} \log[p(y_i, \mathbf{x}_i, z_i; \Psi)] = \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij} \log[\pi_j \phi(y_i - \mathbf{x}_i^T \beta_j; 0, \sigma^2] = \sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij} \log[\pi_j * \phi(y_i - \mathbf{x}_i^T \beta_j; 0, \sigma^2)]$$

E-Step:

$$Q(\Psi \mid \Psi^{(k)}) = E_z[l_n^c(\Psi)] = E_z[\sum_{i=1}^{n}\sum_{j=1}^{m} z_{ij} \log\{\pi_j^{(k)} * \phi(y_i - \mathbf{x}_i^T \beta_j{}^{(k)}; 0, {\sigma^2}^{(k)}\}] = \sum_{i=1}^{n}\sum_{j=1}^{m} E[z_{ij} \mid y_i, \mathbf{x}_i; \Psi^{(k)}][\log\{\pi_j^{(k)} * \phi(y_i - \mathbf{x}_i^T \beta_j{}^{(k}$$

So,

$$Q(\Psi \mid \Psi^{(k)}) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} [log\{\pi_j^{(k)} * \phi(y_i - \mathbf{x}_i^T \beta_j{}^{(k)}; 0, {\sigma^2}^{(k)})\}]$$

Note, using Bayes' Rule we can see that:

$$p_{ij}^{(k+1)} = E[z_{ij} \mid y_i \mathbf{x}_i; \Psi^{(k)}] = \frac{p(z_{ij}; \pi) p(\mathbf{x}_i \mid y_i, z_i; \Psi^{(k)})}{p(y_i, \mathbf{x}_i; \Psi^{(k)})} = \frac{\pi_j^{(k)} \phi(y_i - \mathbf{x}_i^T \beta_j^{(k)}; 0, {\sigma^2}^{(k)})}{\sum_{j=1}^{m} \pi_j^{(k)} \phi(y_i - \mathbf{x}_i^T \beta_j^{(k)}; 0, {\sigma^2}^{(k)})}$$

Also note that $zij = 1$ if observation i is from component j; $zij = 0$ otherwise

M-Step: We maximize the conditional expectation with respect to Ψ to obtain $(\beta^{(k+1)}, \sigma^{2(k+1)})$

$$Q(\Psi \mid \Psi^{(k)}) = \sum_{i=1}^{n}\sum_{j=1}^{m} E[z_{ij} \mid y_i, \mathbf{x}_i; \Psi^{(k)}][\log \pi_j^{(k)} * \phi(y_i - \mathbf{x}_i^T \beta_j^{(k)}; 0, {\sigma^2}^{(k)})] = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \log[\pi_j^{(k)}] + \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \log[\phi(y_i - \mathbf{x}_i^T$$

$$Q(\Psi \mid \Psi^{(k)}) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \log[\pi_j^{(k)}] + \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \log[\frac{1}{\sqrt{2\pi * {\sigma^2}^{(k)}}}] - \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}^{(k+1)} \frac{(y_i - \mathbf{x}_i^T \beta_j^{(k)})^2}{2 * {\sigma^2}^{(k)}}$$

From lecture notes we know that maximization can be obtained by finding a solution for L'(q1,...,qk)=0, i.e.

$$\frac{\partial L(q_1, ...q_k)}{\partial q_k} = 0$$

where

$$L(q_1, ..., q_k) = \sum_{k=1}^{K} W_k log[q_k] - \lambda \sum_{k=1}^{K} (q_k - 1)$$

with λ a Lagrange multiplier.

Now, we must apply the Lagrange equation to the solution above with the constraint that $\sum_{j=1}^{m} \pi_j = 1$. We get the following:

$$L[{\pi_1}^{(k)}, ..., {\pi_m}^{(k)}; \lambda] = Q(\Psi \mid \Psi^{(k)}) - \lambda(\sum_{j=1}^{m} {\pi_j}^{(k)} - 1)$$

Next, we must take the derivative of the L with respect to $\pi_j^{(k)}$. Setting the resulting derivative equal to 0 we get the following:

$$\frac{\partial L}{\partial \pi_j^{(k)}} = \sum_{i=1}^{n} {p_{ij}}^{(k+1)} \frac{1}{\pi_j^{(k+1)}} - \lambda = 0$$

We get

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

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

Now, we must refer back to the equation equal to $Q(\Psi \mid \Psi^{(k)})$. First, to get $\beta_j^{(k+1)}$ we must take the derivative of the equation with respect to $\beta_j^{(k)}$. Setting this equal to 0, we get:

$$\frac{\partial Q}{\partial \beta_j^{(k)}} \propto \sum_{i=1}^{n} p_{ij}^{(k+1)} \mathbf{x}_i (y_i - \mathbf{x}_i^T \beta_j^{(k+1)}) = 0$$

Rearranging this equation, we get:

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

Ultimately solving to get:

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

where j ranges from 1 to m.

Next, we need to solve for ${\sigma^2}^{(k+1)}$. Again, we must take the derivative of the equation equal to $Q(\Psi \mid \Psi^{(k)})$, but this time with respect to ${\sigma^2}^{(k)}$. Setting this equal to 0 we get:

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

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

Problem 2: Gamma mixture models and Rejection Sampling

Our goal here is to simulate observations from a given density function $f(x)$. We will achieve this by using the Rejection Sampling Method with instrumental density function $g(x)$.

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

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

(a) Here, we try to find the value of the normalizing 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 "1". Hence:

$$\int_0^\infty g(x)dx = 1$$
$$\int_0^\infty C * (2x^{\theta-1} + x^{\theta-\frac{1}{2}})e^{-x}dx = 1$$
$$C\Big[\int_0^\infty 2x^{\theta-1}e^{-x}dx + \int_0^\infty x^{\theta-\frac{1}{2}}e^{-x}dx\Big] = 1$$

$$C\Big[2\Gamma(\theta)\int_0^\infty \frac{x^{\theta-1}e^{-x}}{\Gamma(\theta)}dx + \Gamma(\theta+\frac{1}{2})\int_0^\infty \frac{x^{(\theta+\frac{1}{2})-1}e^{-x}}{\Gamma(\theta+\frac{1}{2})}dx\Big] = 1 \tag{3}$$

It can be observed that the integral in the above equation is a mixture of two gamma density integral functions, namely $\Gamma(rate = 1, shape = \theta)$ and $\Gamma(rate = 1, shape = \theta + \frac{1}{2})$.

$$C\Big[2\Gamma(\theta)\int_0^\infty \frac{x^{\theta-1}e^{-x}}{\Gamma(\theta)}dx + \Gamma(\theta+\frac{1}{2})\int_0^\infty \frac{x^{(\theta+\frac{1}{2})-1}e^{-x}}{\Gamma(\theta+\frac{1}{2})}dx\Big] = 1$$
$$C\Big[2\Gamma(\theta) + \Gamma(\theta+\frac{1}{2})\Big] = 1$$
$$C = \Big[\frac{1}{2\Gamma(\theta) + \Gamma(\theta+\frac{1}{2})}\Big]$$

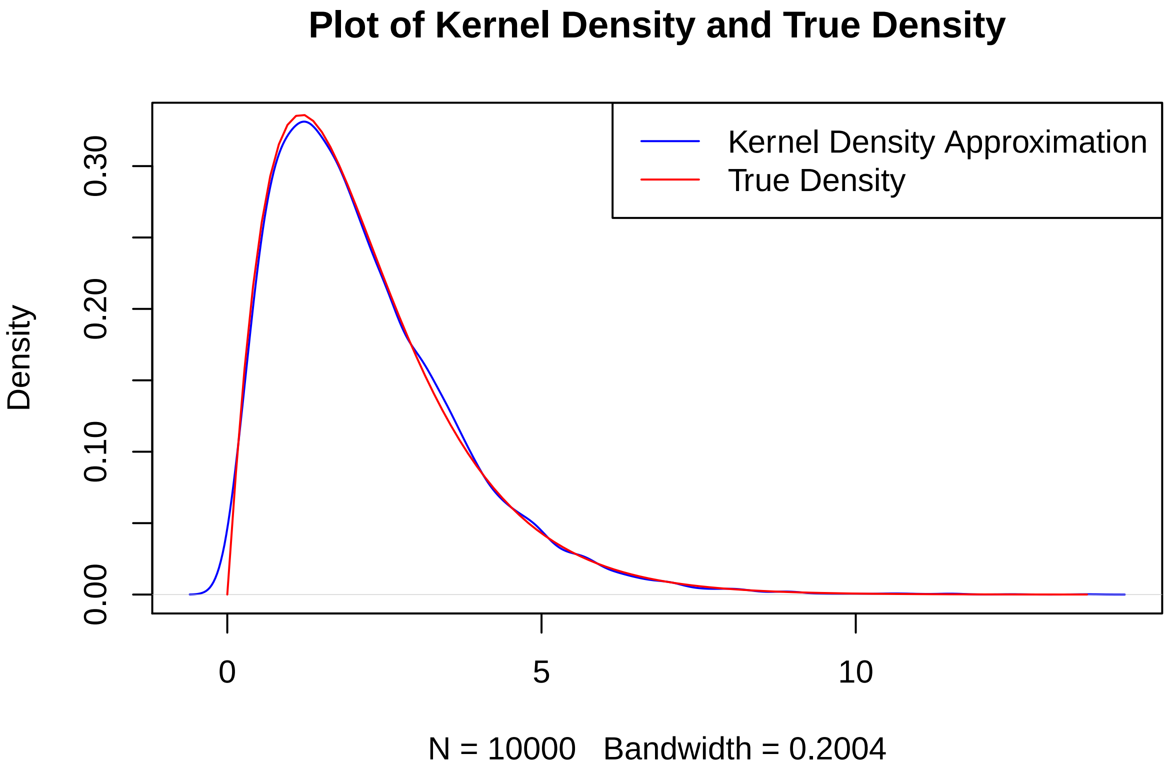
So, the density function $g(x)$ is given as follows:

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

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

$$\alpha_1 + \alpha_2 = 1$$

(b) Now, we will sample from g and plot the kernel density approximation and true density.



\

Hence, from the plot we can see that our kernel density approximation is pretty good.

(c) Now, we will use Rejection Sampling to sample from $f(x)$

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

$$Let\ q(x) = \sqrt{(4+x)}x^{\theta-1}e^{-x} \tag{5}$$

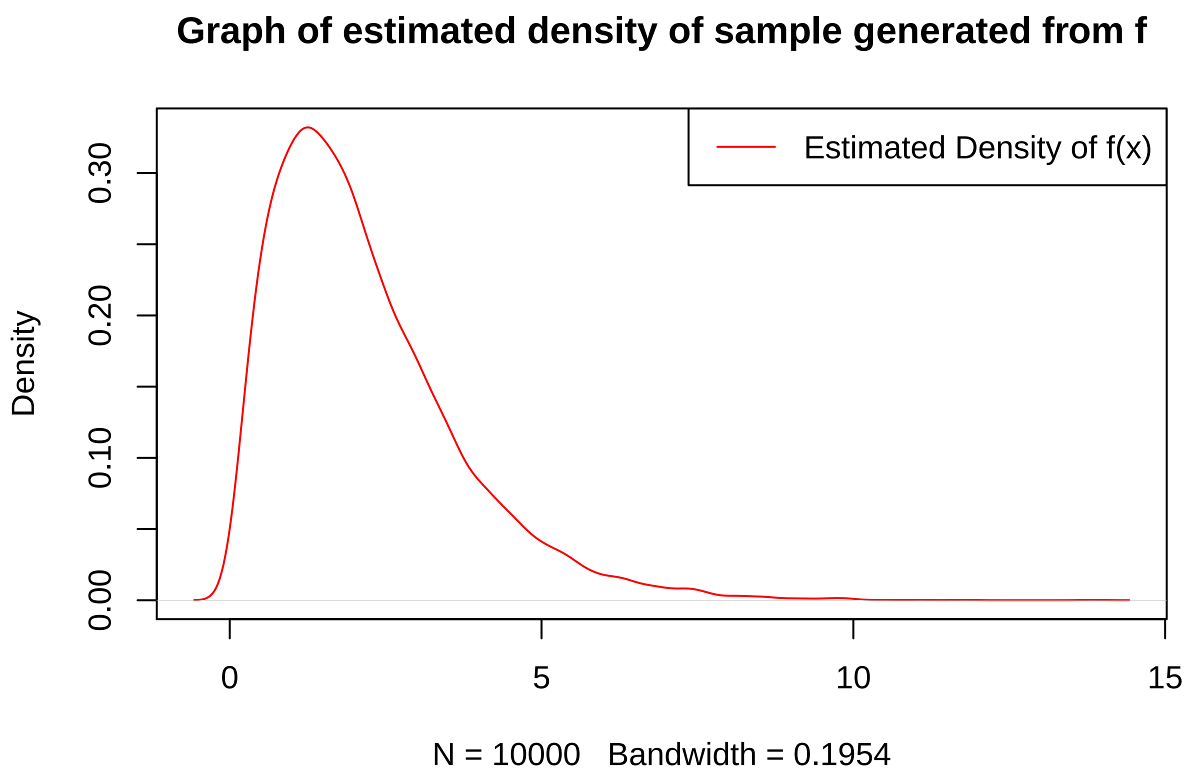
$$q(x) \leq (\sqrt{4} + \sqrt{x})x^{\theta-1}e^{-x} \tag{6}$$

$$q(x) \leq (2x^{\theta-1} + x^{\theta-\frac{1}{2}})e^{-x} \tag{7}$$

$$q(x) \leq g(x) \tag{8}$$

$$\alpha = 1 \tag{9}$$

$$\tag{10}$$


Graph of estimated density of sample generated from f
Estimated Density of f(x)
Density
0.00
0.10
0.20
0.30
0
5
10
15
N = 10000 Bandwidth = 0.1954

Problem 3: Two methods to sample from f

(a) Our goal is to simulate observations from a given density. Note, the density looks similar to the sum of Beta distributions.

$$f(x) \propto \frac{x^{\theta-1}}{1+x^2} + \sqrt{2+x^2}(1-x)^{\beta-1} where 0 < x < 1$$

We can show that

$$f(x) \propto \frac{x^{\theta-1}}{1+x^2}+\sqrt{2+x^2}(1-x)^{\beta-1}q(x) = \frac{x^{\theta-1}}{1+x^2}+\sqrt{2+x^2}(1-x)^{\beta-1}q(x) \leq x^{\theta-1}+\sqrt{2}(1-x)^{\beta-1}+\sqrt{x^2}(1-x)^{\beta-1}q(x) \leq alphag$$

because $\sqrt{2+x^2} \leq \sqrt{2}+x$ and $1 \leq (1+x^2)$

We must show that g(x) is less than αg(x) in order to simulate using rejection sampling and the mixture of beta distributions. Solving for alpha we get:

$$g(x) \propto x^{\theta-1}+\sqrt{2}(1-x)^{\beta-1}+x(1-x)^{\beta-1}g(x) = C[B(\theta,1)\frac{x^{\theta-1}}{B(\theta,1)}+\sqrt{2}B(1,\beta)\frac{(1-x)^{\beta-1}}{B(1,\beta)}+B(2,\beta)\frac{x(1-x)^{\beta-1}}{B(2,\beta)}$$

$$\int_0^\infty g(x) = 1/C[B(\theta,1)+B(1,\beta)+B(2,\beta)] = 1/C = \frac{1}{B(\theta,1)+B(1,\beta)+B(2,\beta)}$$

$$\alpha = \frac{1}{C} = B(\theta,1)+B)(1,\beta)+B(2,\beta)$$

Note: I am having a lot of trouble getting this code to run. I'm out of time, however, so I will submit the solution as is.

(b) Here, we use the two parts of the function separately and then sample from f.

$$
\begin{aligned}
f(x) &\propto \left(\frac{x^{\theta-1}}{1+x^2}\right) + \left(\sqrt{2+x^2}(1-x)^{\beta-1}\right) \\
f(x) &\propto \sum_{i=1}^{2} q_i(x) \\
where \;\; q_1(x) &= \frac{x^{\theta-1}}{1+x^2} \\
q_2(x) &= \sqrt{2+x^2}(1-x)^{\beta-1} \\
q_1(x) &\leq B(\theta,1)\frac{x^{\theta-1}}{B(\theta,1)} \\
g_1(x) &= \frac{x^{\theta-1}}{B(\theta,1)} \\
\alpha_1 &= B(\theta,1) \\
q_2(x) &\leq \sqrt{2}(1-x)^{\beta-1} + x(1-x)^{\beta-1} \\
q_2(x) &\leq \sqrt{2}B(1,\beta)\frac{(1-x)^{\beta-1}}{B(1,\beta)} + B(2,\beta)\frac{x(1-x)^{\beta-1}}{B(2,\beta)} \\
q_2(x) &\leq \left(\sqrt{2}B(1,\beta)+B(2,\beta)\right)\left(\frac{\sqrt{2}B(1,\beta)}{\sqrt{2}B(1,\beta)+B(2,\beta)}\frac{(1-x)^{\beta-1}}{B(1,\beta)} + \frac{B(2,\beta)}{\sqrt{2}B(1,\beta)+B(2,\beta)}\frac{x(1-x)^{\beta-1}}{B(2,\beta)}\right) \\
g_2(x) &= \frac{\sqrt{2}B(1,\beta)}{\sqrt{2}B(1,\beta)+B(2,\beta)}\frac{(1-x)^{\beta-1}}{B(1,\beta)} + \frac{B(2,\beta)}{\sqrt{2}B(1,\beta)+B(2,\beta)}\frac{x(1-x)^{\beta-1}}{B(2,\beta)} \\
\alpha_2 &= \sqrt{2}B(1,\beta)+B(2,\beta)
\end{aligned}
$$

Now, we plot the estimated density -