MTH-511A Numerical Assignment Report

Gaurav Kumar Mahajan (160264)

4 November 2019

Notes

This document contains
1. Observations for numerical experiments for all the questions.
2. Mathematical details for questions 1 and 2
In detail proofs and theoretical explanations for Question 1(a) and Question 2(a) are attached in the supplementary document.
All the results and figures will be generated on running the code.

Abstract

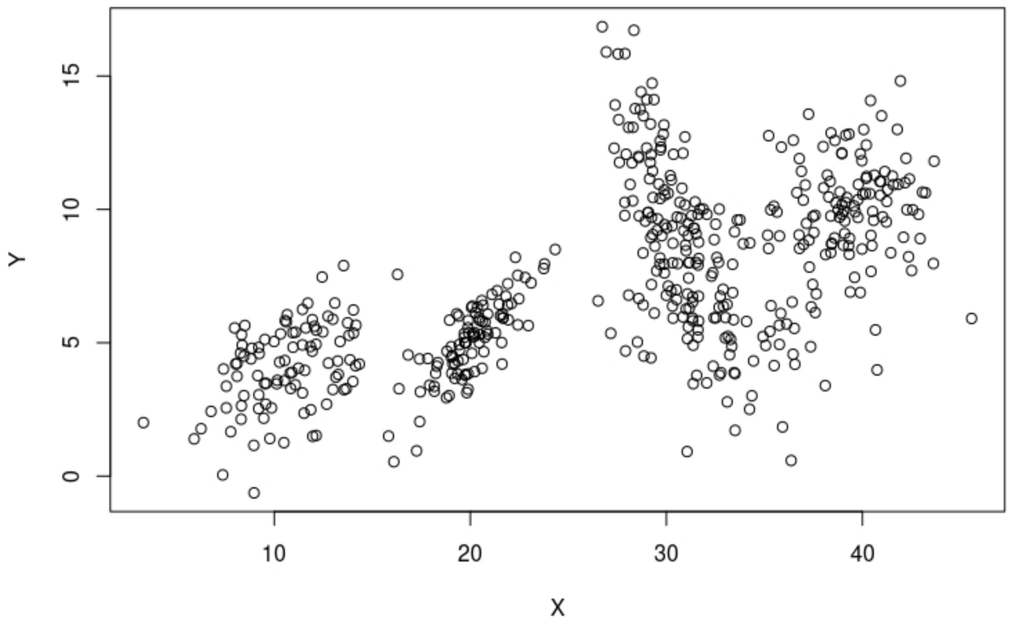
In Numerical Assignment, in Q1 We have (X, Y) from a **bi-variate mixture of Gaussian** with some unknown number of classes C, Our goal are to choosing the appropriate number of classes C and estimate **mean** $\mu = (\mu_1, \mu_2, ..., \mu_C)$, **Co-variance Matrix** $\Sigma = (\Sigma_1, \Sigma_2, ..., \Sigma_C)$, and **Mixture probability** $\pi = (\pi_1, \pi_2, ..., \pi_C)$

In Q2, Problem is a Bayesian logistic regression model, response y_i is a realization of Y_i , and Y_i follow the **Bernoulli distribution** with probability p_i, p_i is a function of β , Our goal is to find the **posterior distribution of** β **and posterior mean of** β **: E(**β **/** $\underline{\mathbf{Y}}$**).**

1 Question 1(a)

From the plot of Z = (X, Y) , we can easily seen that there is 4 Classes, so we choose C = 4 .

Scatter plot of data (X, Y)



$$f(x, y/\mu_1,\ .....,\ \mu_4,\ \Sigma_1,\ .....,\ \Sigma_4) = \sum_{c=1}^{c=4} \pi_c f(x, y/\mu_c,\ \Sigma_c) \quad (1)$$

Where

$$f(Z/\mu_c,\ \Sigma_c) = \frac{1}{(2\pi)\ |\Sigma_c|^{1/2}}\ exp(-\frac{1}{2}(Z-\mu_c)^T\ \Sigma_c^{-1}\ (Z-\mu_c)) \quad (2)$$

$$Z = (X, Y) \qquad \mu_c \in \mathbb{R} \qquad \Sigma_c \in \mathbb{R}^2$$

$$\text{Let's } \theta = (\pi_1,\ .....,\ \pi_4\ \mu_1,\ .....,\ ,\mu_4,\ \Sigma_1,\ .....,\ \Sigma_4)$$

We solved this problem by using of EM Algorithm

1.1 Estimation Step (E - step)

For given parameter vales we can compute the expected values of latent variable

1.2 Maximization Step (M - step)

In this step, we updates the parameters of our model based on the latent variable calculate using ML method

For given **gaussian mixture model**, our goal is to maximize the likelihood function with respect to the parameters comprising the **means** and **co-variances** of the components and **mixing probability**

Log-likelihood function

$$L(\theta/\underline{Z}\ ) = \log(f(\ \underline{Z}\ /\ \theta\ )) \tag{3}$$

$$L(\theta/\underline{Z}\ ) = \log(\ \prod_{i=1}^{n} f(\ Z_i\ /\ \theta)) \tag{4}$$

$$L(\theta/\underline{Z}\ ) = \sum_{i=1}^{n} \log(\ f(\ Z_i\ /\ \theta)) \tag{5}$$

$$L(\theta/\underline{Z}\ ) = \sum_{i=1}^{n} \log(\sum_{c=1}^{4}\ f(\ Z_i\ /\ \mu_c,\ \Sigma_c)) \tag{6}$$

Let's for any fix class c

$$\gamma_{i,c,(k)} = \frac{\pi_{c,(k)}\ f(\ Z_i\ /\ \mu_{c,(k)},\ \Sigma_{c,(k)}))}{\sum_{j=1}^{4}\ \pi_{j,(k)}\ f(\ Z_i\ /\ \mu_{j,(k)},\ \Sigma_{j,(k)})} \tag{7}$$

With help of M - step , we re-estimate the parameter using the current parameters values. We find For fix any class c

$$\mu_{c,(k+1)} = \frac{\sum_{i=1}^{n}\ \gamma_{i,c,(k)}\ Z_i}{\sum_{i=1}^{n}\ \gamma_{i,c,(k)}} \tag{8}$$

$$\Sigma_{c,(k+1)} = \frac{\sum_{i=1}^{n}\ \gamma_{i,c,(k)}\ (Z_i - \mu_{c,(k)})\ (Z_i - \mu_{c,(k)})^T}{\sum_{i=1}^{n}\ \gamma_{i,c,(k)}} \tag{9}$$

$$\pi_{c,(k+1)} = \frac{\sum_{i=1}^{n}\ \gamma_{i,c,(k)}}{n} \tag{10}$$

Mean and Co-variance of all data are

$$\mu = \begin{pmatrix} 26.97 \\ 7.27 \end{pmatrix} \tag{11}$$

$$\Sigma = \begin{pmatrix} 108 & 22 \\ 22 & 11 \end{pmatrix} \tag{12}$$

From the plot, we take a guessing a initial value of $\mu,\ \Sigma,\ \pi$

$$\mu_1 = \begin{pmatrix} 11.97 \\ 3.52 \end{pmatrix} \tag{13}$$

$$\mu_2 = \begin{pmatrix} 21.97 \\ 6.02 \end{pmatrix} \tag{14}$$

$$\mu_3 = \begin{pmatrix} 31.97 \\ 8.52 \end{pmatrix} \tag{15}$$

$$\mu_4 = \begin{pmatrix} 41.97 \\ 11.02 \end{pmatrix} \tag{16}$$

$$\Sigma_1 = \begin{pmatrix} 108 & 22 \\ 22 & 11 \end{pmatrix} \tag{17}$$

$$\Sigma_2 = \begin{pmatrix} 108 & 22 \\ 22 & 11 \end{pmatrix} \tag{18}$$

$$\Sigma_3 = \begin{pmatrix} 108 & 22 \\ 22 & 11 \end{pmatrix} \tag{19}$$

$$\Sigma_4 = \begin{pmatrix} 108 & 22 \\ 22 & 11 \end{pmatrix} \tag{20}$$

$$\pi_1 = \pi_2 = \pi_3 = \pi_4 = 0.25 \tag{21}$$

1.3 Parameter values from the code is

Estimated values of mean μ

$$\mu_1 = \begin{pmatrix} 10.735437 \\ 4.084471 \end{pmatrix} \tag{22}$$

$$\mu_2 = \begin{pmatrix} 20.122703 \\ 5.084493 \end{pmatrix} \tag{23}$$

$$\mu_3 = \begin{pmatrix} 30.737572 \\ 8.508568 \end{pmatrix} \tag{24}$$

$$\mu_4 = \begin{pmatrix} 39.041787 \\ 9.597558 \end{pmatrix} \tag{25}$$

Estimated values of Co-variance Σ

$$\Sigma_1 = \begin{pmatrix} 5.436589 & 1.695577 \\ 1.695577 & 2.577215 \end{pmatrix} \tag{26}$$

$$\Sigma_2 = \begin{pmatrix} 2.706711 & 2.021562 \\ 2.021562 & 2.219912 \end{pmatrix} \tag{27}$$

$$\Sigma_3 = \begin{pmatrix} 4.098295 & -4.200178 \\ -4.200178 & 10.668113 \end{pmatrix} \tag{28}$$

$$\Sigma_4 = \begin{pmatrix} 5.484697 & 1.343040 \\ 1.343040 & 5.072358 \end{pmatrix} \tag{29}$$

Estimated values of Mixture Probability π

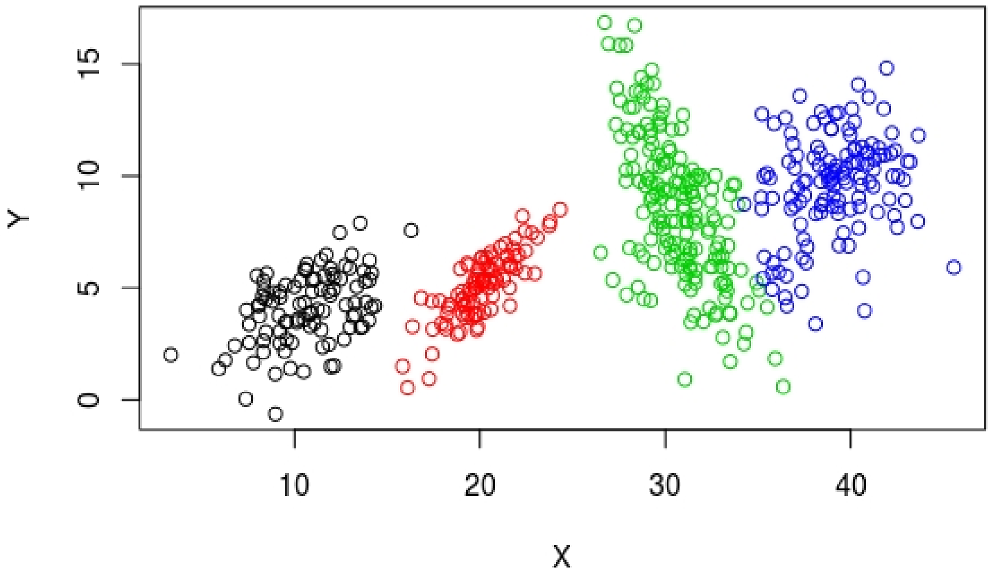
$$\pi_1 = 0.1953434 \tag{30}$$

$$\pi_2 = 0.1906508 \tag{31}$$

$$\pi_3 = 0.3528420 \tag{32}$$

$$\pi_4 = 1 - \pi_1 - \pi_2 - \pi_3 = 0.2611638 \tag{33}$$

Scatterplot of data (X, Y) for C = 4



2 Question 1(b)

To evaluate the performance of any model, We need to test it on some unseen data. Based on the models performance on unseen data we can say weather our model is under - fitting / over - fitting / well generalized.

Cross Validation (CV) is one of the technique used to test the effectiveness of a model, if we have a limited data.

We used K - Cross Validation

Cross Validation function is

$$CV(\hat{f}, \lambda) = \frac{1}{n} \sum_{i=1}^{n} L(y_i,\ \hat{f}^{-i}(Z,\ \lambda)) \tag{34}$$

Where L is a loss function
and $\lambda \in \Lambda \rightarrow$ All model being compared

We used the negative log-likelihood as a loss function, choose the best model among the 5 models with C = (2, 3, 4, 5, 6)

2.1 Negative log-likelihood values that obtained from the code for different Classes

S.No	Model	Negative log-likelihood value
1	**C = 2**	**5.873263**
2	**C = 3**	**5.665248**
3	**C = 4**	**5.505543**
4	**C = 5**	**5.480109**
5	**C = 6**	**5.486371**

Here We can clearly see that negative log-likelihood take minimum value at C = 5, hence we chose C = 5 model.

3 Question 1(c)

The **Akaike information criterion (AIC)** is an estimator of the relative quality of models for a given set of data. Given a collection of models for the data, **AIC** estimates the quality of each model, relative to each of the other models.

Loss function is , Akaike information criterion

$$AIC(\ \hat{\theta}\ /\ Z) = -2\ \log(\ L(\ \hat{\theta}\ /Z))\ \ +\ \ 2K \tag{35}$$

Where K = Number of parameter in θ
Since we have a bivariate data, hence in our case

$$K = \mathbf{5}C\ \textbf{- 1}$$

We applied AIC method into whole data which does not use **cross-validation**

We used Akaike information criterion ,to find the best model among the 5 models with C = (2, 3, 4, 5, 6)

S.No	Model	AIC Value
1	**C = 2**	**5875.007**
2	**C = 3**	**5660.143**
3	**C = 4**	**5498.310**
4	**C = 5**	**5467.764**
5	**C = 6**	**5467.915**

Here We can clearly see that AIC take minimum value at C = 5, hence we chose C = 5 model.

4 Question 1(d)

From the both method, we have found the same model $C = 5$. Hence there are 5 classes in our **Gaussians Mixture** .

5 Question 1(e)

Since we have 5 classes, with the help of EM Algorithm ,we will find the **Maximum Likelihood** values of different parameter

5.1 MLE Values from the code is

Estimated values of mean μ

$$\mu_1 = \begin{pmatrix} 10.73567 \\ 4.08446 \end{pmatrix} \tag{36}$$

$$\mu_2 = \begin{pmatrix} 20.114937 \\ 5.079003 \end{pmatrix} \tag{37}$$

$$\mu_3 = \begin{pmatrix} 33.668097 \\ 6.247307 \end{pmatrix} \tag{38}$$

$$\mu_4 = \begin{pmatrix} 30.264956 \\ 9.556029 \end{pmatrix} \tag{39}$$

$$\mu_5 = \begin{pmatrix} 39.54258 \\ 10.36426 \end{pmatrix} \tag{40}$$

$$\Sigma_1 = \begin{pmatrix} 5.437740 & 1.695457 \\ 1.695457 & 2.577144 \end{pmatrix} \tag{41}$$

$$\Sigma_2 = \begin{pmatrix} 2.684671 & 2.005951 \\ 2.005951 & 2.208859 \end{pmatrix} \tag{42}$$

$$\Sigma_3 = \begin{pmatrix} 12.5891336 & -0.2855273 \\ -0.2855273 & 4.6835966 \end{pmatrix} \tag{43}$$

$$\Sigma_4 = \begin{pmatrix} 2.616236 & -3.715132 \\ -3.715132 & 9.320141 \end{pmatrix} \tag{44}$$

$$\Sigma_5 = \begin{pmatrix} 4.07763550 & 0.05886963 \\ 0.05886963 & 2.59971108 \end{pmatrix} \tag{45}$$

$$\pi_1 = 0.1953510 \tag{46}$$

$$\pi_2 = 0.1901778 \tag{47}$$

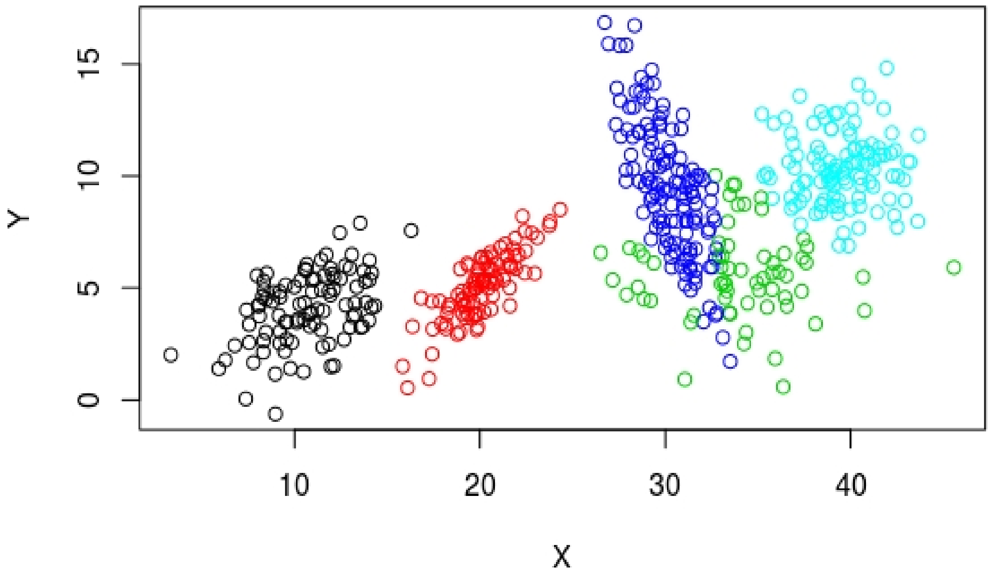
$$\pi_3 = 0.1591327 \tag{48}$$

$$\pi_4 = 0.2490160 \tag{49}$$

$$\pi_5 = 0.2063225 \tag{50}$$

6 Question 1(f)

Scatterplot of data (X, Y) for C = 5



7 Question 2(a)

Given a **Bayesian logistic regression model** . For $i = 1$, 2, ..., n and $x_i = (1, x_{2i}, x_{3i}, x_{4i}, x_{5i})^T$, Where x_i be the vector of covariates for the i^{th} observation and $\beta \in \mathbb{R}^5$ be the corresponding vector of regression coefficients, response y_i is a realization of Y_i with

$$Y_i \sim Bern(p_i) \tag{51}$$

$$Where\ p_i = \frac{exp(x_i^T\beta)}{1 + exp(x_i^T\beta)} \tag{52}$$

Given that β has the prior distribution

$$\beta \sim N_5(0, 100\ I_5) \tag{53}$$

$$Let's\ \underline{Y} = (Y_1, Y_2, ..., Y_n)^T \tag{54}$$

$$f(y_i/\beta)\ =\ P(Y_i = y_i)\ =\ p_i^{y_i}(1 - p_i)^{y_i} \tag{55}$$

$$f(\ \underline{Y}/\beta)\ =\ \prod_{i=1}^{n} f(y_i/\beta) \tag{56}$$

$$\Pi(\beta) = f(\beta) = \frac{1}{(2\pi)^{5/2}\ |\Sigma|^{1/2}}\ exp(-\frac{\beta^T\Sigma^{-1}\beta}{2}) \tag{57}$$

$$\textbf{Where}\ \Sigma = \begin{pmatrix} 100 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 0 & 100 \end{pmatrix}$$

Posterior Distribution of β

$$\Pi(\beta/\ \underline{Y})\ \propto\ \Pi(\beta)\ \ f(\underline{Y}/\beta) \tag{58}$$

$$\Pi(\beta/\ \underline{Y})\ \propto\ \ exp(-\frac{\beta^T\Sigma^{-1}\beta}{2})\ \prod_{i=1}^{n} p_i^{y_i}\ (1 - p_i)^{1-y_i} \tag{59}$$

$$\Pi(\beta/\ \underline{Y})\ \propto\ \ exp(-\frac{\beta^T\Sigma^{-1}\beta}{2})\ \prod_{i=1}^{n} \{(\frac{exp(x_i^T\beta)}{1 + exp(x_i^T\beta)})^{y_i}\ \ (\frac{1}{1 + exp(x_i^T\beta)})^{1-y_i}\} \tag{60}$$

$$\Pi(\beta/\underline{Y}) \propto exp(-\frac{\beta^T\Sigma^{-1}\beta}{2}) \prod_{i=1}^{n}\{\frac{exp(x_i^T\beta y_i)}{1+exp(x_i^T\beta)}\} \tag{61}$$

$$\Pi(\beta/\underline{Y}) \propto \frac{exp(-\frac{\beta^T\Sigma^{-1}\beta}{2} + \sum_{i=1}^{n} x_i^T\beta)}{\prod_{i=1}^{n}(1+exp(x_i^T\beta))} \tag{62}$$

8 Question 2(b)

Starting values that are closer to the mode of the posterior distribution will ensure faster converging to the target density. It can be difficult in practice to find starting points near the posterior mode .Prior value is a good starting value for the MCMC that samples from the posterior distribution of β. because MCMC sampling always updates its value from previous iteration, by moving closer to estimated value. IF we take the expected prior value as the starting value, the model will itself update it to correct estimate.

$$\beta_{Start} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{63}$$

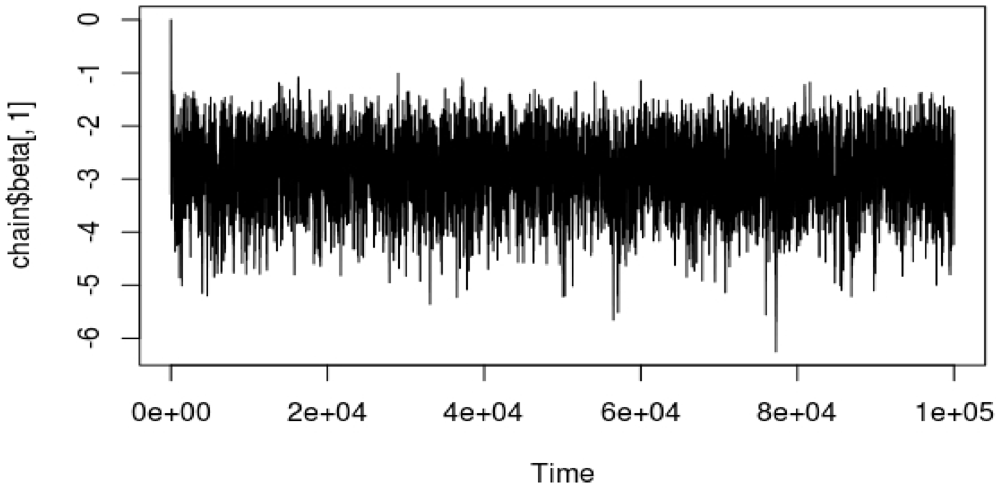
9 Question 2(c)

9.1 Values that obtained from the code is

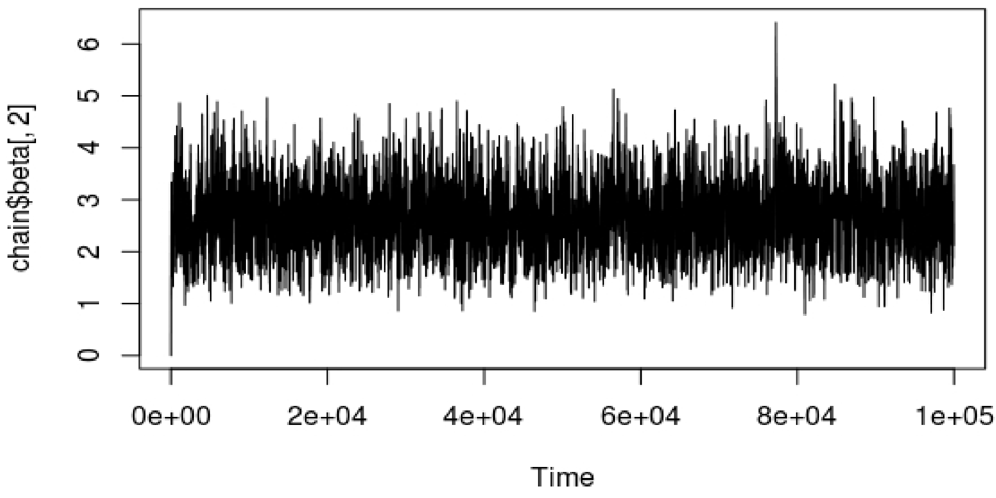
$$Acceptance\ probability = 0.28556 \tag{64}$$

10 Question 2(d)

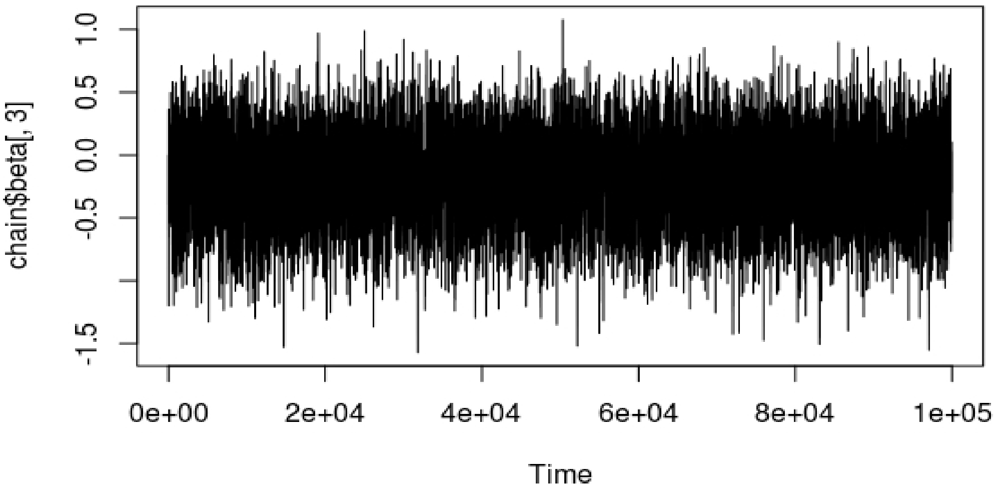
10.1 Time Plot of β_1



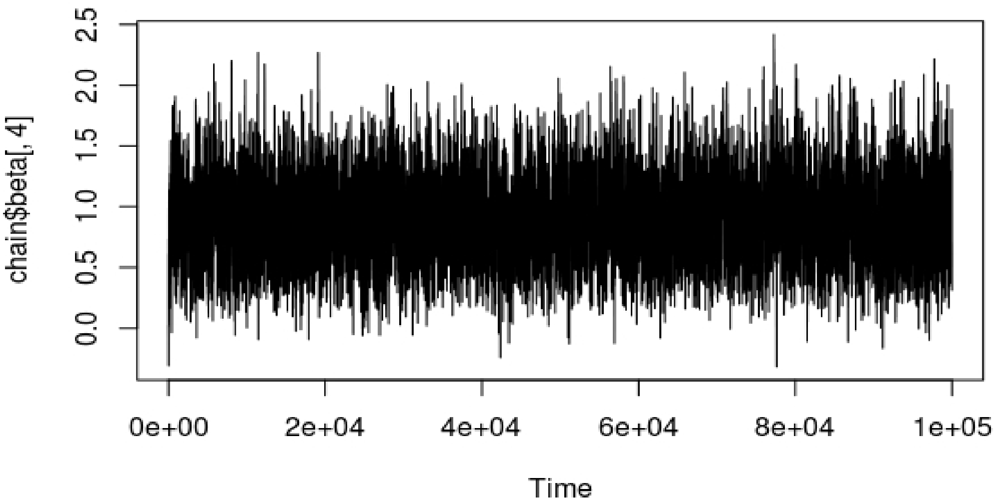
10.2 Time Plot of β_2



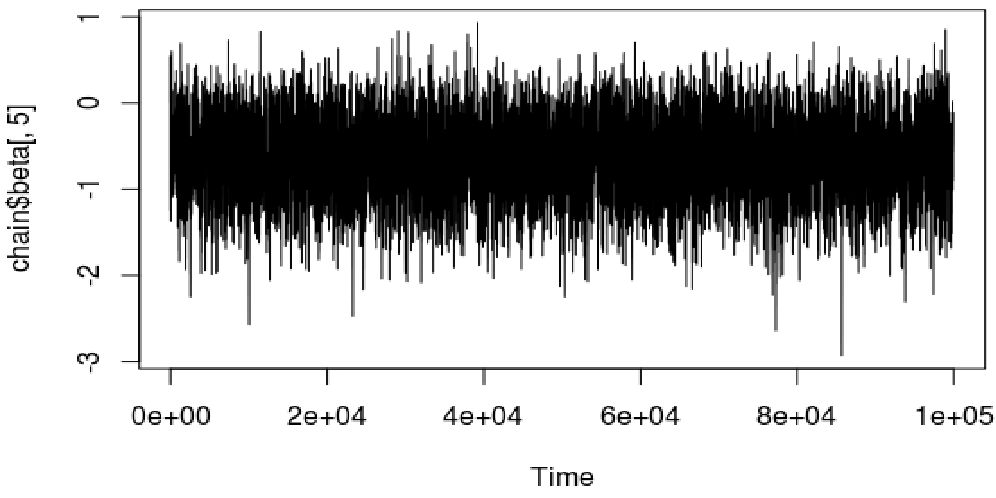
10.3 Time Plot of β_3



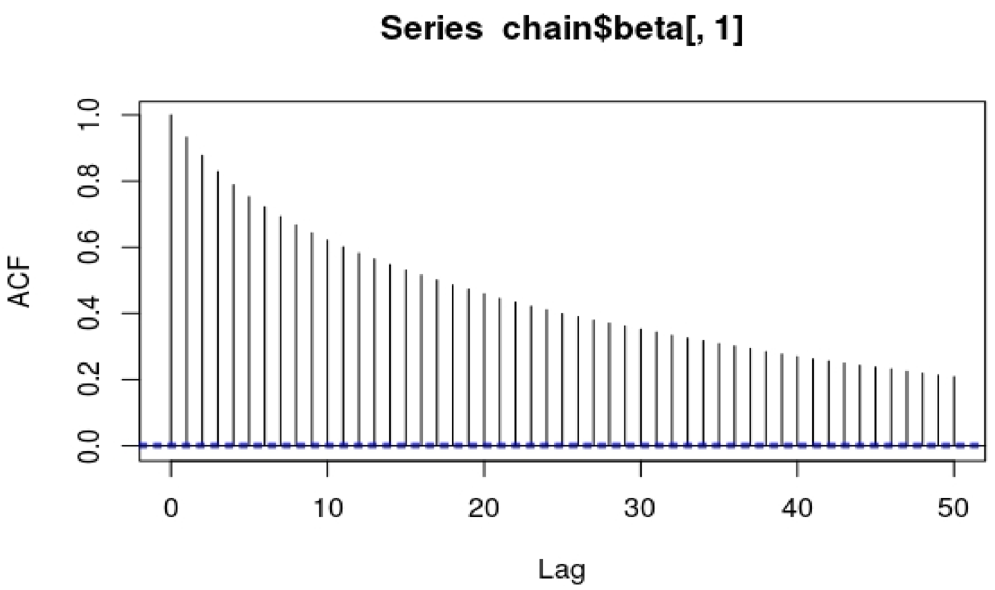
10.4 Time Plot of β_4



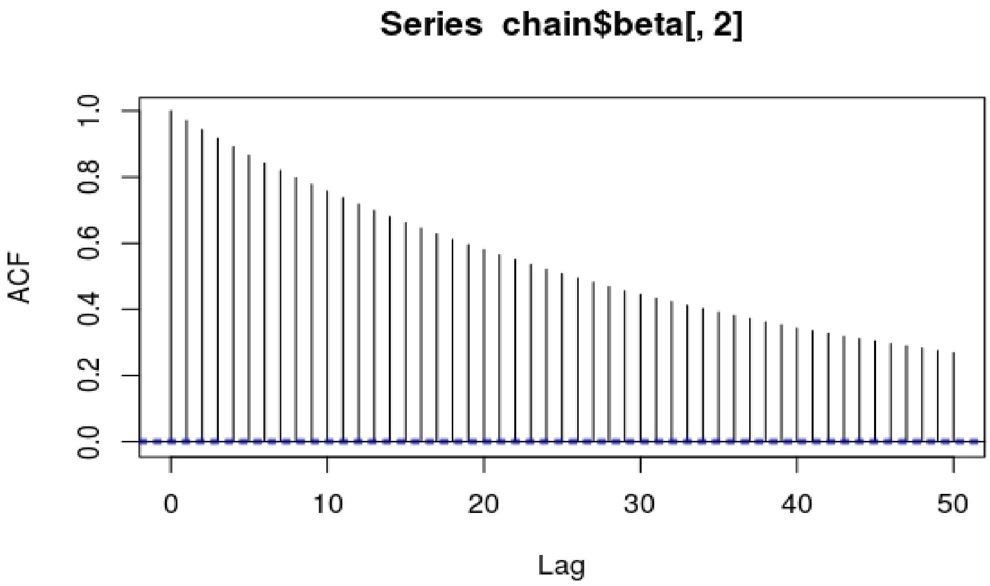
10.5 Time Plot of β_5



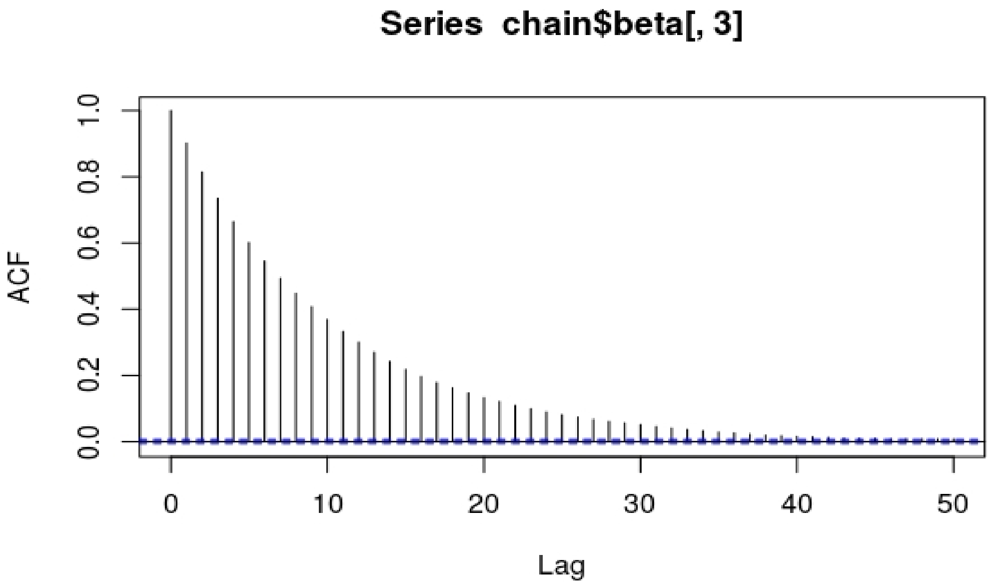
10.6 Auto-correlation Plot of β_1



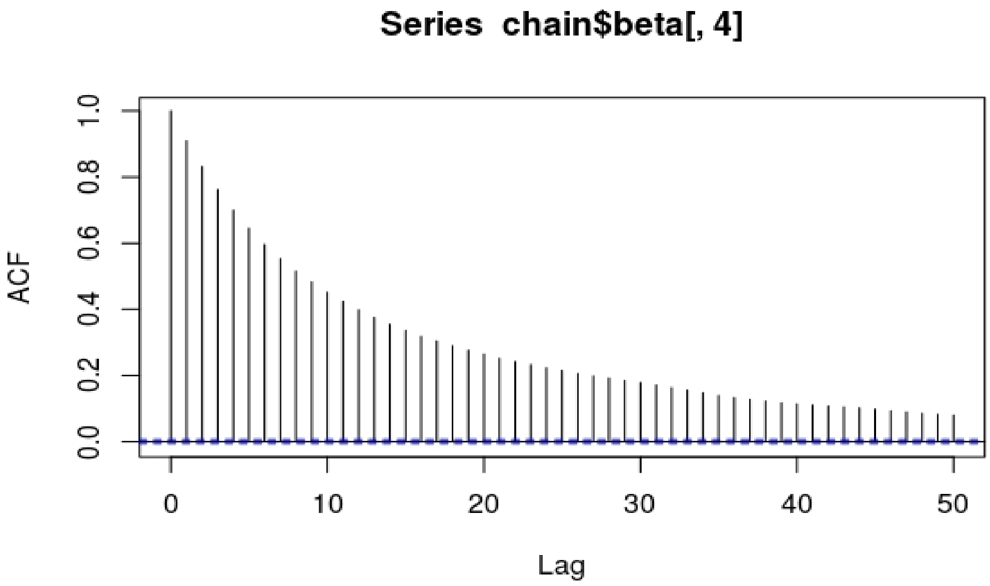
10.7 Auto-correlation Plot of β_2



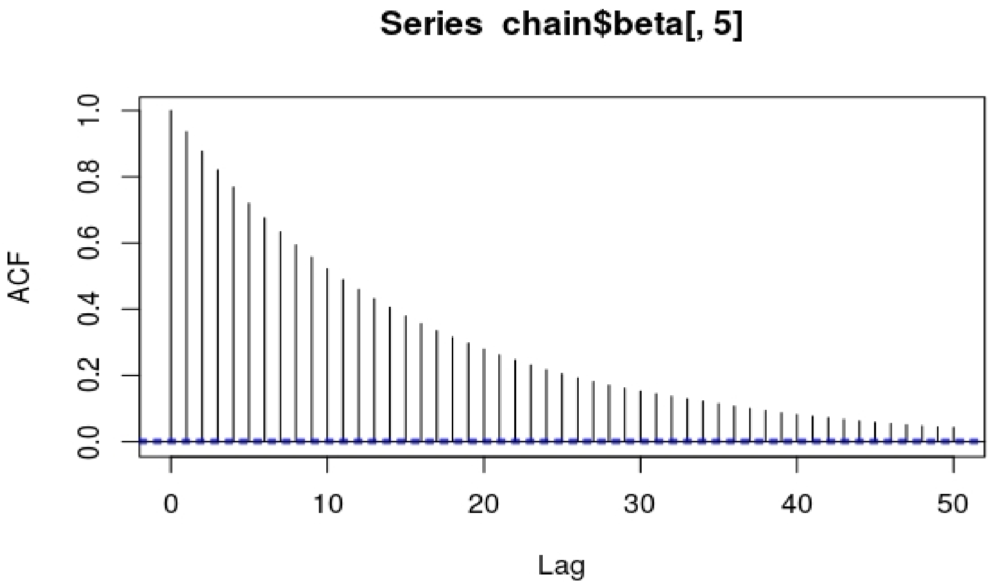
10.8 Auto-correlation Plot of β_3



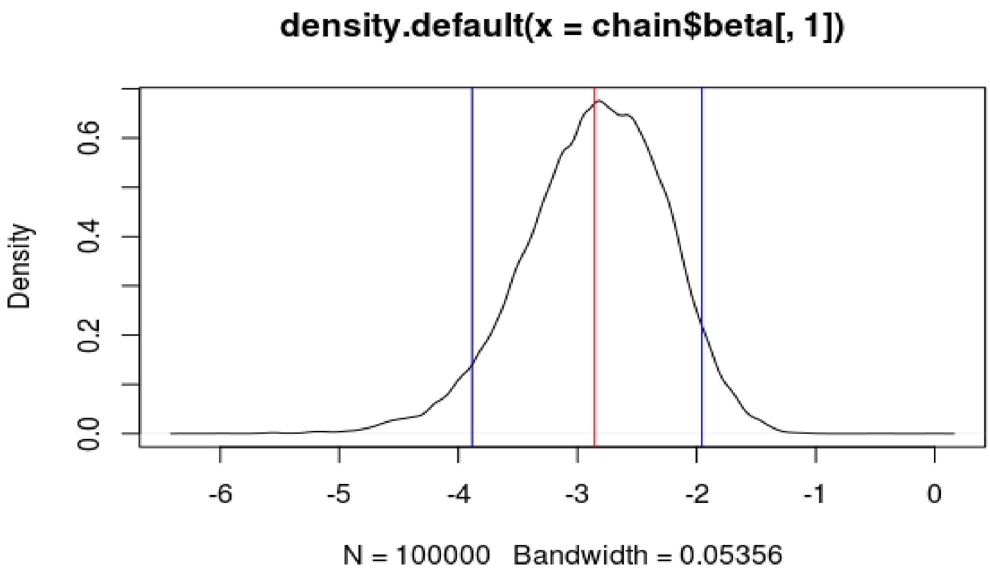
10.9 Auto-correlation Plot of β_4



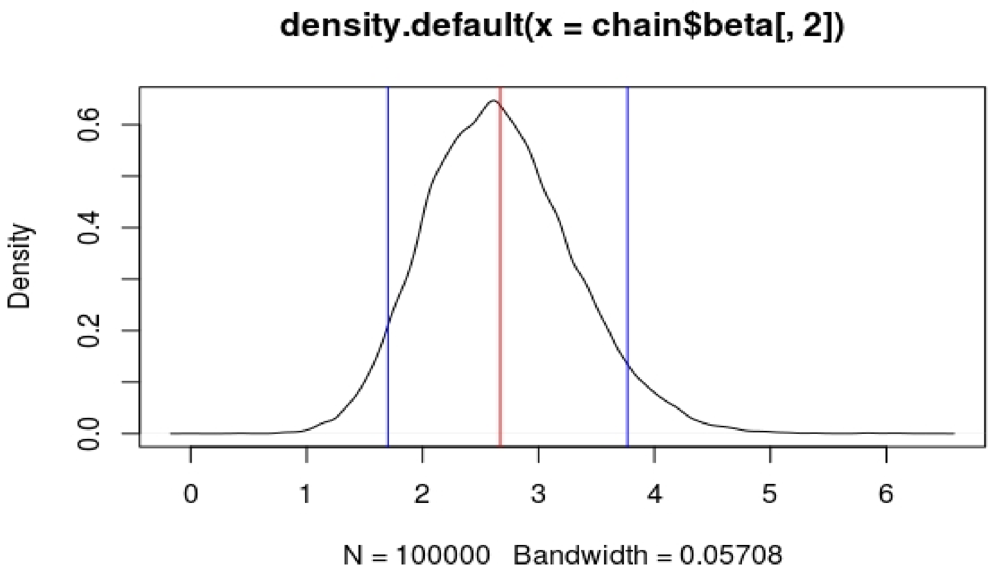
10.10 Auto-correlation Plot of β_5



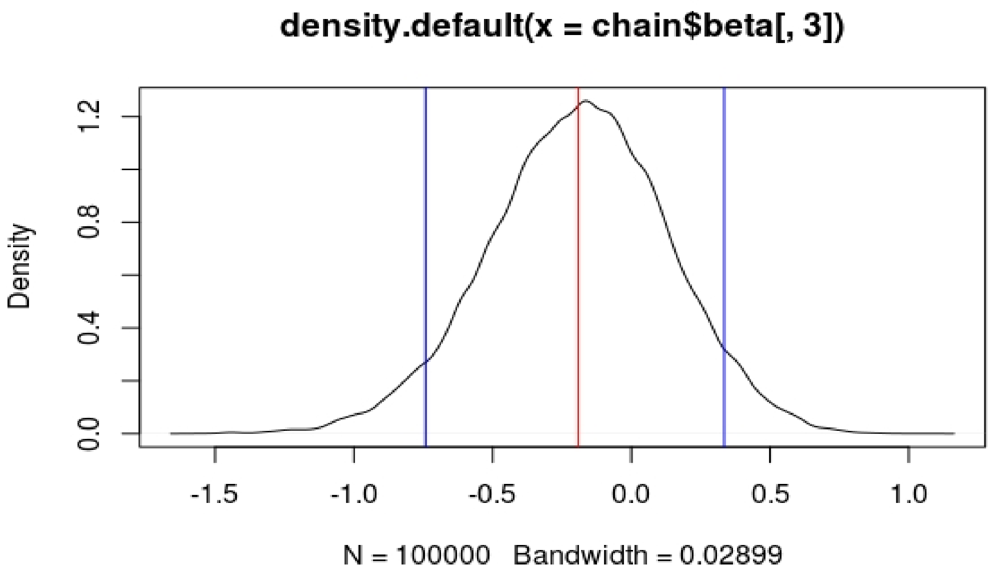
10.11 Density Plot of β_1



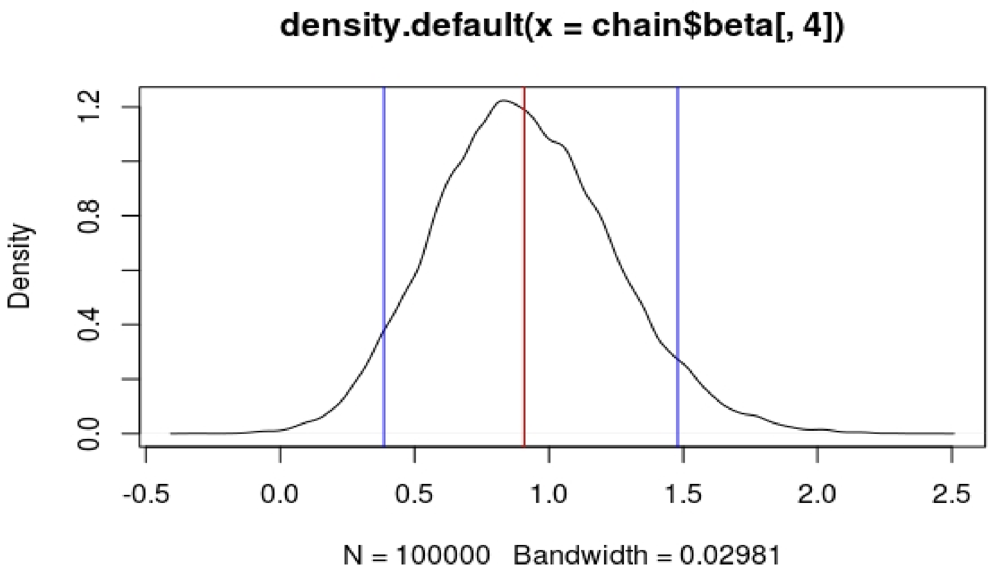
10.12 Density Plot of β_2



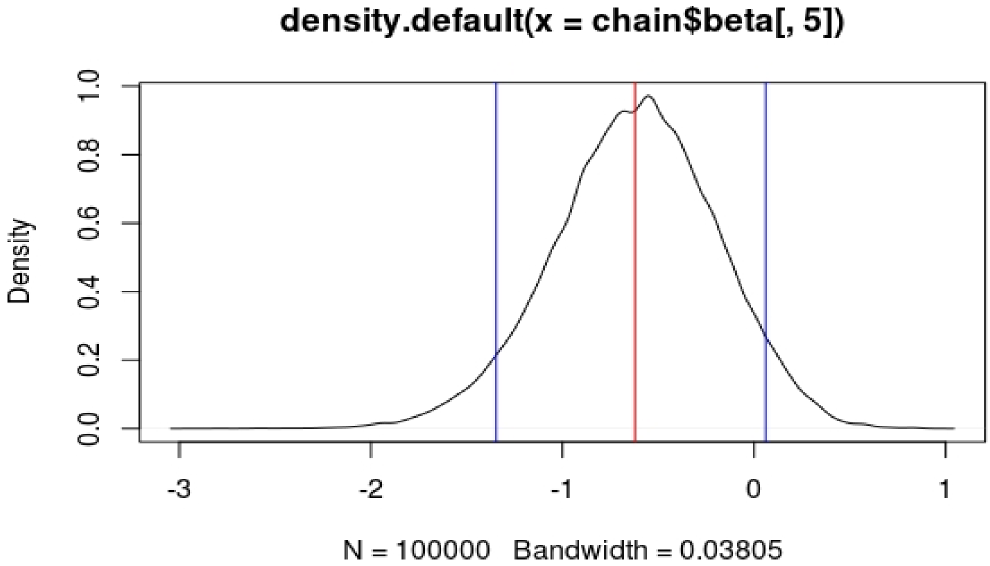
10.13 Density Plot of β_3



10.14 Density Plot of β_4



10.15 Density Plot of β_5



11 Question 2(e)

Values of β_{MLE} and estimated Posterior mean of β from the code is

$$\beta_{MLE} = \begin{pmatrix} -2.5441507 \\ 2.3666499 \\ -0.1724703 \\ 0.8077819 \\ -0.5259096 \end{pmatrix} \tag{65}$$

$$\beta_{Estimated} = \begin{pmatrix} -2.8562572 \\ 2.6698231 \\ -0.1908527 \\ 0.9085875 \\ -0.6217147 \end{pmatrix} \tag{66}$$

Hence We can clearly see that $\beta_{Estimated}$ is very close to the β_{MLE}