1. EMG – MCMC – Uni-Variate with Data 1

   (a) The prior and posterior distribution is derived in the following:

$$\begin{aligned} f(\beta_1) &= \frac{1}{\sqrt{200\pi}} exp\left\{-\frac{\beta_1^2}{200}\right\} \\ \pi(\beta_1|\sigma_i, \mathbf{Y}, \mathbf{X}) &\propto L(\beta_1|\sigma_i, \beta_0, \lambda, \mathbf{Y}, \mathbf{X}) f(\beta_1) \\ &= f(\beta_1) \prod_{i=1}^{100} f(Y_i, X_i|\sigma_i, \beta_0, \lambda) \end{aligned}$$

   It is noticeable that, the EMG pdf would generate very small numbers such that the product of all 100 observations would lead the value of posterior pdf to be very close to 0 whichever the value of β_1 would be (due to the level of tolerance in R). As a result, the M-H algorithm would "accept" all the time. In this case, the MCMC sample generated would be purely from the proposed distribution such that the mean would not converge no matter how many iterations were run. To solve this problem, I proposed a naive approach – add a scaling factor k to the EMG pdf (with M-H, one can add any constant as a scaling factor to the posterior pdf without altering the sampling properties). That is to say:

$$\begin{aligned} \pi(\beta_1|\sigma_i, \mathbf{Y}, \mathbf{X}) &\propto h(\beta_1|\sigma_i, \mathbf{Y}, \mathbf{X}) \\ &= f(\beta_1) \prod_{i=1}^{100} k f(Y_i, X_i|\sigma_i, \beta_0, \lambda) \\ &= f(\beta_1) k^{100} \prod_{i=1}^{100} f(Y_i, X_i|\sigma_i, \beta_0, \lambda) \end{aligned}$$

   With several trials, the A-M-H algorithm that produces stable outcome is proposed in the following:

   i. Start with $\beta_1^1 = 0$ (this is a value based on several preliminary MCMC runs)

   ii. For t = 1,. . . , n, propose Z such that

$$q(z|\beta_1^t) = \frac{1}{\sqrt{2\pi\tau^2}} exp\left\{\frac{(z-\beta_1^t)^2}{2\tau^2}\right\} \sim N(\beta_1^t, \tau^2)$$

   where $\tau^2 = 1$

   iii. Select $k = 22$. Accept Z with probability

$$p(z|\beta_1^t) = \begin{cases} \min\left\{1, \frac{h(z)}{h(\beta_1^t)} \frac{q(z|\beta_1^t)}{q(\beta_1^t|z)}\right\} & \text{if} \quad h(\beta_1^t) q(\beta_1^t|z) > 0 \\ 1 & \text{if} \quad h(\beta_1^t) q(\beta_1^t|z) = 0 \end{cases}$$

   if $p(z|\beta_1^t) > U \sim unif(0, 1)$. Otherwise reject Z, set $Z = \beta_1^t$

   (b) I would pick the sample size at which the chain shows stable mean and MCSE in case of $n = 10000$. Under such condition, the posterior expectation $\mu_{\beta_1} = 7.345387$ and the $MCSE_{\beta_1} = 0.009446465$.

   (c) The 95% credible interval [2.5%, 97.5%] for β_1 is $[6.736405, 7.926770]$.

   (d) The plot is shown in Figure 1.

   (e) The algorithm selection is based on the plot of mean, mcse and acf against the number of iterations (as is shown in Figure 2 – 4). As is shown in these figures, mean and mcse all converge reasonably around the value calculated in part b and acf drop within reasonable interval in a quick manner. Moreover, even as it is shown in Figure 1, when the sample size is cut by half, the density shape is quite consistent. (run the code for more graphs)

2. EMG – MCMC – Multi-Variate with Data 2

(a) The prior and posterior distribution is derived in the following:

$$\begin{aligned}
f(\beta_0) &= \frac{1}{\sqrt{200\pi}} exp\left\{-\frac{\beta_0^2}{200}\right\} \\
f(\beta_1) &= \frac{1}{\sqrt{200\pi}} exp\left\{-\frac{\beta_1^2}{200}\right\} \\
f(\lambda) &= \frac{1}{\Gamma(0.01)\lambda^{0.01}} \lambda^{0.01-1} exp\left\{-\frac{\lambda}{100}\right\} \\
\pi(\beta_0, \beta_1, \lambda|\sigma_i, \mathbf{Y}, \mathbf{X}) &\propto L(\beta_0, \beta_1, \lambda|\sigma_i, \mathbf{Y}, \mathbf{X}) f(\beta_0) f(\beta_1) f(\lambda) \\
&= f(\beta_0) f(\beta_1) f(\lambda) \prod_{i=1}^{500} f(Y_i, X_i|\sigma_i, \beta_0, \beta_1, \lambda)
\end{aligned}$$

For similar idea as in question 1, propose scaling factor k such that

$$\begin{aligned}
\pi(\beta_0, \beta_1, \lambda|\sigma_i, \mathbf{Y}, \mathbf{X}) &\propto h(\beta_0, \beta_1, \lambda|\sigma_i, \mathbf{Y}, \mathbf{X}) \\
&= f(\beta_0) f(\beta_1) f(\lambda) \prod_{i=1}^{500} k f(Y_i, X_i|\sigma_i, \beta_0, \beta_1, \lambda) \\
&= f(\beta_0) f(\beta_1) f(\lambda) k^{500} \prod_{i=1}^{500} f(Y_i, X_i|\sigma_i, \beta_0, \beta_1, \lambda)
\end{aligned}$$

Since the three variables $\beta_0, \beta_1, \lambda$ were distributed independently in prior, V-M-H should have similar efficiency with that of A-M-H. Hence here I would still apply A-M-H algorithm. With several trials, the A-M-H algorithm that produces stable outcome is proposed in the following:

i. Start with $\beta_0^1 = 0, \beta_1^1 = 0, \lambda^1 = 0.01$ (this is a value based on several preliminary MCMC runs)

ii. For t = 1,. . . , n, since the prior distributions are independent with each other, propose Z_1, Z_2, Z_3 such that

$$\begin{aligned}
q(z_1|\beta_0^t) &= \frac{1}{\sqrt{2\pi\tau_1^2}} exp\left\{\frac{(z_1 - \beta_0^t)^2}{2\tau_1^2}\right\} \sim N(\beta_0^t, \tau_1^2) \\
q(z_2|\beta_1^t) &= \frac{1}{\sqrt{2\pi\tau_2^2}} exp\left\{\frac{(z_2 - \beta_1^t)^2}{2\tau_2^2}\right\} \sim N(\beta_1^t, \tau_2^2) \\
q(z_3|\lambda^t) &= \frac{1}{\sqrt{2\pi\tau_3^2}} exp\left\{\frac{(z_3 - \lambda^t)^2}{2\tau_3^2}\right\} \sim N(\lambda^t, \tau_3^2)
\end{aligned}$$

where $\tau_1^2 = 0.1, \tau_2^2 = 0.1, \tau_3^2 = 0.1$, note that for z_3, only positive number is allowed to be produced

iii. Select $k = 16$. Accept Z_1, Z_2, Z_3 with probability

$$p(z_1, z_2, z_3|\beta_0^t, \beta_1^t, \lambda^t) = \begin{cases} \min\left\{1, \frac{h(z_1,z_2,z_3)}{h(\beta_0^t,\beta_1^t,\lambda^t)} \frac{q(z_1|\beta_0^t))}{q(\beta_0^t|z_1)} \frac{q(z_2|\beta_1^t))}{q(\beta_1^t|z_2)} \frac{q(z_3|\lambda^t)}{q(\lambda^t|z_3)}\right\} \\ \text{if} \\ \text{h}(\beta_0^t, \beta_1^t, \lambda^t) q(\beta_0^t|z_1) q(\beta_1^t|z_2) q(\lambda^t|z_3) \ > \ 0 \\ 1 \\ \text{if} \\ \text{h}(\beta_0^t, \beta_1^t, \lambda^t) q(\beta_0^t|z_1) q(\beta_1^t|z_2) q(\lambda^t|z_3) \ = \ 0 \end{cases}$$

if $p(z_1, z_2, z_3|\beta_0^t, \beta_1^t, \lambda^t) > U \sim unif(0, 1)$. Otherwise reject Z_1, Z_2, Z_3, set $\{Z_1, Z_2, Z_3\} = \{\beta_0^t, \beta_1^t, \lambda^t\}$

(b) I would pick the sample size at which the chain shows stable mean and MCSE in case of $n = 10000$. Under such condition, the posterior expectation, MCSE and 95% confidence interval are reported in the following:

	β_0	β_1	λ
Mean	2.0676660	3.4620131	0.5872572
MCSE	0.01585171	0.03378696	0.00378453
95%CI	[1.789142, 2.353423]	[2.968436, 3.900523]	[0.5150406, 0.6642538]

(c) The correlation between β_0 and β_1 is calculated as 0.2125158.

(d) The plot is shown in Figure 5 – 7 for β_0, β_1, λ respectively.

(e) The algorithm selection is based on the plot of mean, mcse and acf against the number of iterations (as is shown in Figure 8 – 16). As is shown in these figures, mean and mcse all converge reasonably around the value calculated in part b and acf drop within reasonable interval in a quick manner. Moreover, even as it is shown in Figure 5 – 7, when the sample size is cut by half, the density shape is quite consistent. (run the code for more graphs)

3. EMG – MCMC – Multi-Variate with Data 3

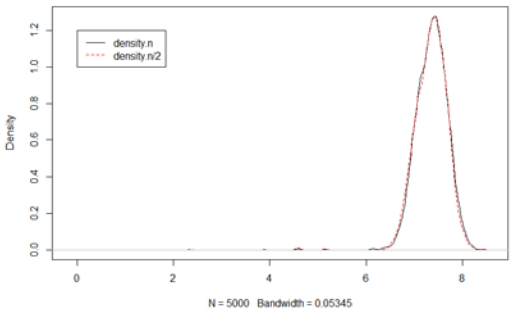
(a) I would pick the sample size at which the chain shows stable mean and MCSE in case of $n = 10000$. Under such condition, the posterior expectation, MCSE and 95% confidence interval are reported in the following:

	β_0	β_1	λ
Mean	0.1252536	2.4621539	0.1576834
MCSE	0.020081811	0.045420711	0.000359027
95%CI	[-0.2536091, 0.6833649]	[0.9162041, 3.1079836]	[0.1460686, 0.1693677]

(b) The plot is shown in Figure 17 – 19 for β_0, β_1, λ respectively.

(c) Similar as the previous two questions, since the sample size of the observations were increased to 1000 while the magnitude of the value of each sample point does not vary much, the scaling factor k needs to be redesigned as well as that the variance for the proposed normal distributions should be re-calibrated to make reasonable sample set. In general, following the previous experience, the calibration should not differ much from what has been done in the previous two questions. However, Figure 19 shows that the posterior density for λ is highly inconsistent with sample size (however many initial values or calibrations were tried). My thinking of this would be that the observations may contain redundant information which may need to be truncated. Or else, the sample generated for λ is not efficient after certain number of iterations such that all iterations after 5000 need to be truncated. (run the code for more graphs)

Figure 1



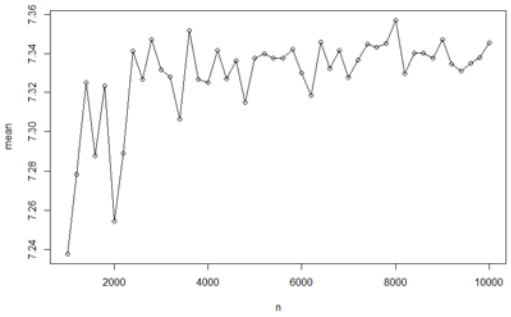
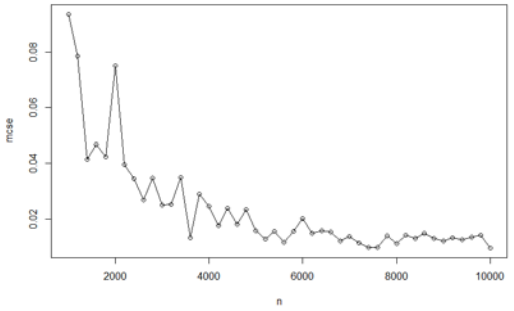


Figure 2

Figure 3



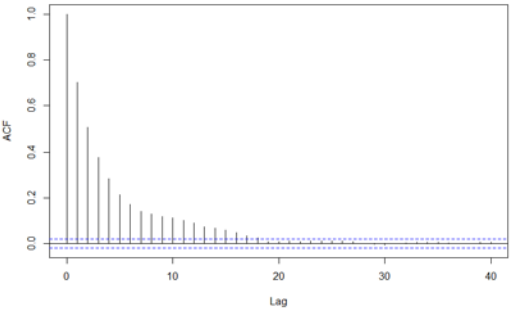
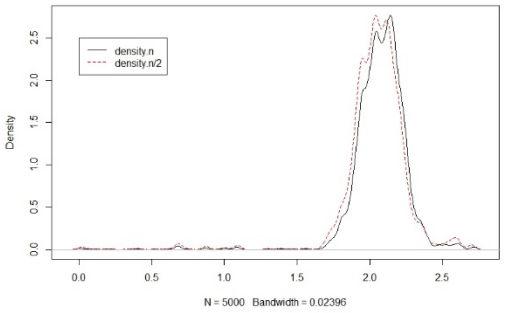


Figure 4

Figure 5



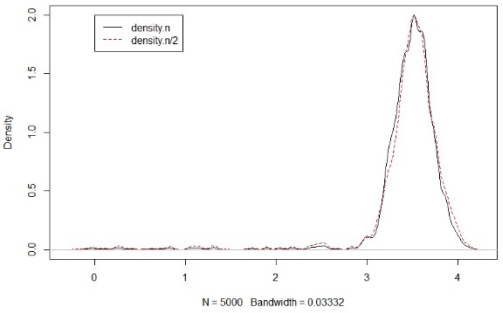
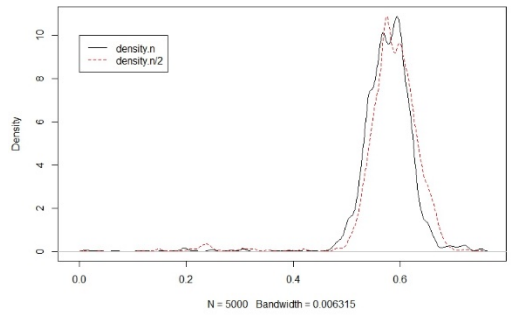


Figure 6

Figure 7



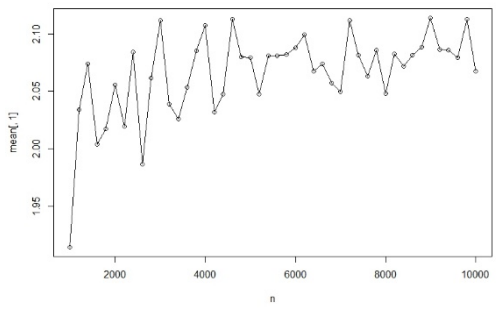
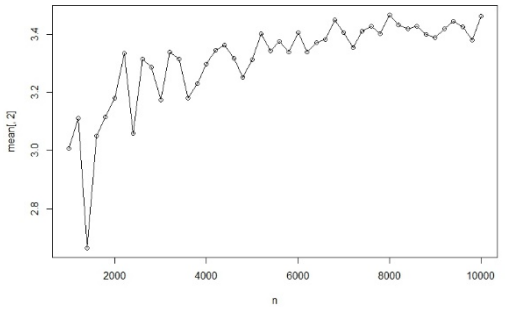


Figure 8

Figure 9



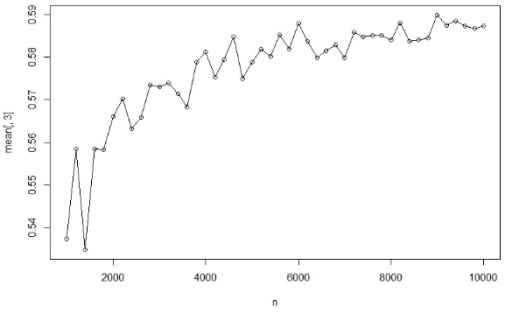
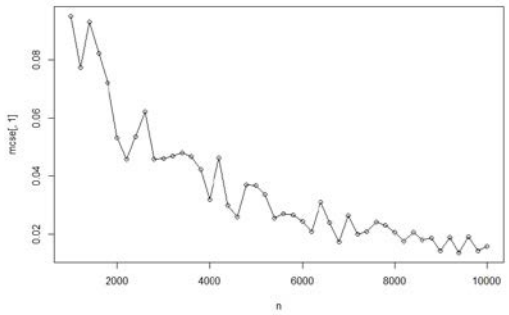


Figure 10

Figure 11



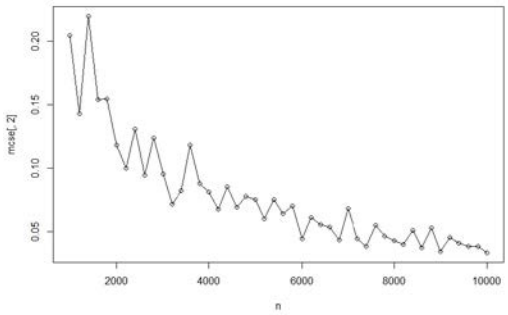
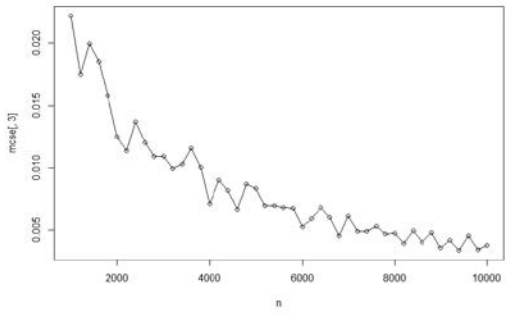


Figure 12

Figure 13



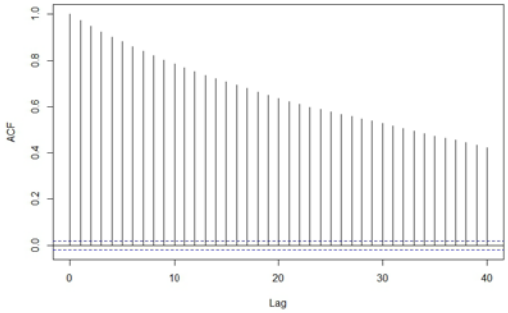
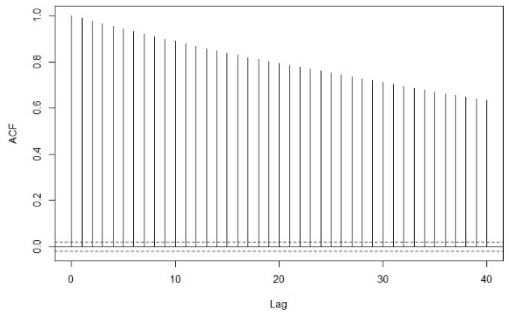


Figure 14

Figure 15



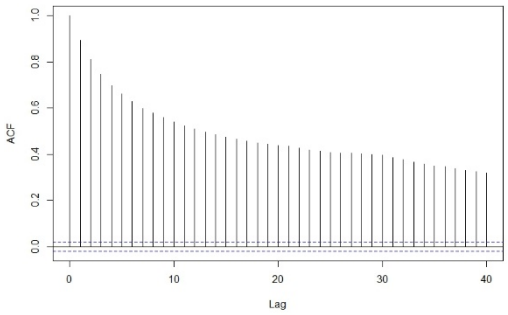
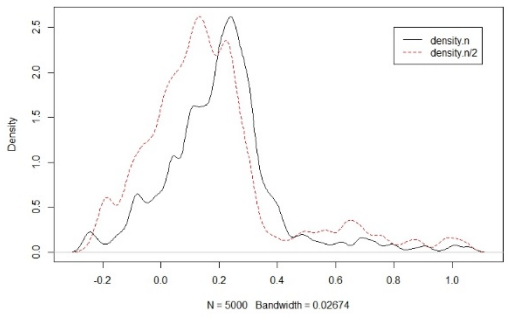


Figure 16

Figure 17



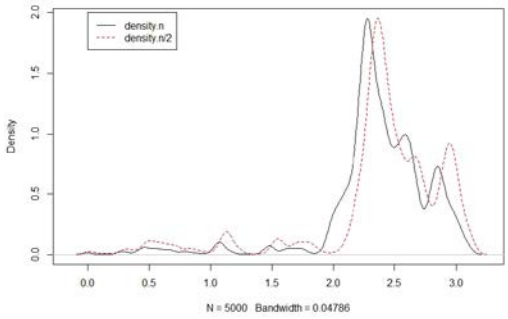


Figure 18

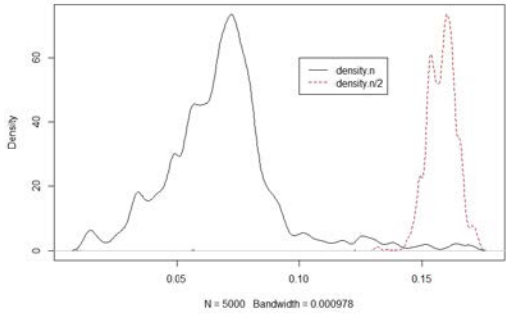


Figure 19