Take Home Final Report

Gang Yang (gzy105)

Q1

(A) Posterior distribution $\pi(\beta_1|\mathbf{Y},\mathbf{X}) \sim \mathrm{p}(\mathbf{Y},\mathbf{X}|\beta_1,\beta_0,\sigma_i,\lambda,\mu,\sigma)\mathrm{p}(\beta_1|\beta_0,\sigma_i,\lambda,\mu,\sigma)$

Thus $\pi(\beta_1|\mathbf{Y},\mathbf{X}) \sim \mathrm{h}(\beta_1|\mathbf{Y},\mathbf{X}) = N(\beta_1;\mu,\sigma)\times\prod_i EMG(Y_i,\beta_1 X_i+\beta_0,\sigma_i,\lambda)$,

where $\mathrm{h}(\beta_1|\mathbf{Y},\mathbf{X})$ is the unnormalized distribution.

Metropolis-Hasting Algorithm for approximating posterior distribution:

0)Initialize fixed parameters, such as sample size, $\beta_0,\sigma_i,\lambda,\mu,\sigma$, load data, load external function such as density of EMG variable, bm function to calculate Monte Carlo standard error.

1) Select a initial value for $\beta_{1,1}$=7 as the first state of the Markov Chain, which is chosen based on preliminary trials.

2) Generate the next state through the following way:

2a) Locate the Nth state of the Markov Chain of $\beta_{1,N}$. Propose the (N+1)th state$\beta^*_{1,N+1}$ for β_1 drawn from a normal distribution $\mathrm{N}(\beta_{1,N},\tau=1)$, where τ is the standard deviation.

2b) Calculate the Metropolis-Hasting acceptance probability $\alpha\left(\beta_{1,N},\beta^*_{1,N+1}\right)$ =min(1,$\frac{h(\beta^*_{1,N+1})}{h(\beta_{1,N})}$)

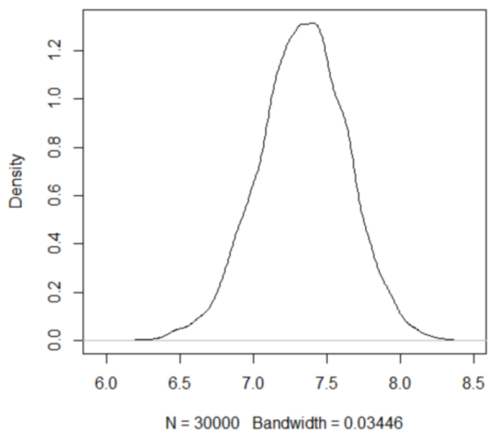
2c) Accept $\beta^*_{1,N+1}$ with probability $\alpha\left(\beta_{1,N},\beta^*_{1,N+1}\right)$, else reject y and stay at x. That is, simulate U~uniform distribution(0,1) and if U<$\alpha\left(\beta_{1,N},\beta^*_{1,N+1}\right)$ set $\beta_{1,N+1}=\beta^*_{1,N+1}$, else set $\beta_{1,N+1}=\beta_{1,N}$. Use log scale to avoid numerical issues, that is calculate

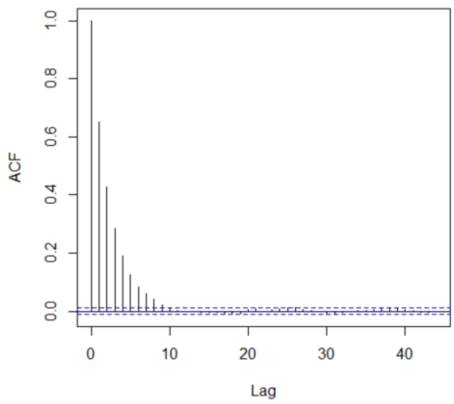
log$\alpha\left(\beta_{1,N},\beta^*_{1,N+1}\right)$=min(0,log($h(\beta^*_{1,N+1})$)-log($h(\beta_{1,N})$)) and compare logU with log$\alpha\left(\beta_{1,N},\beta^*_{1,N+1}\right)$. Repeat procedure 2) until draw enough samples from the posterior distribution.

(B) MCMC Sample size N =30000, estimate of posterior expectation of β_1 is 7.3385. MCMC SE is 0.00389.

(C) A 95% credible interval for β_1 is (6.724,7.921) based on my samples.

(D) Estimate of posterior distribution of β_1 is plot as the left one below, where the horizontal axis is β_1:



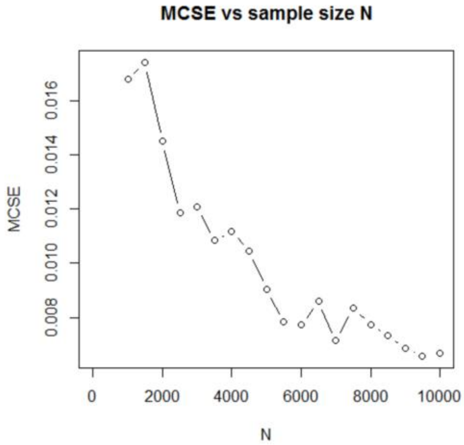


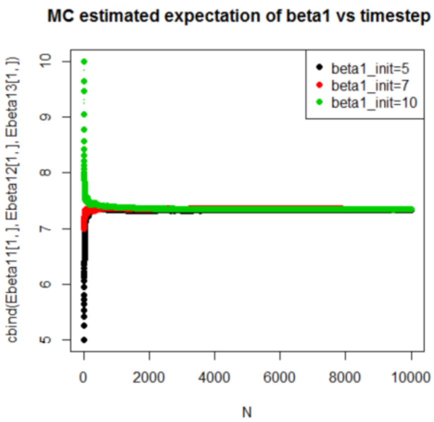
(E) We show the validity of result through the following evidence:

(1) Auto Correlation of the states in Markov Chain is appropriate as shown in the right figure above. The correlation length is not so large that effective sampled state is limited. The acceptance rate is around 35%, which is also appropriate. ESS is around 2000 when N=10000, ESS is 6244 when N=30000 in the MCMC run we used to answer (b)-(d).

(2) I increase the sample size N from 1000 to 10000 by a step of 500, draw the MCMC SE

versus Sample size N as shown in the left graph below. One can see the MCMC SE tend to decrease when N increase, which is a good indication of the validity of the result. Also the MCMC SE is much smaller in magnitude than the expectation when N=30000.





(3) Also I try to compare Markov Chain states sampled from different initial states, MC estimated expectation of β_1 all converge to the same interval as shown in the right figure above, which is also another good indication. In the figure, y axis is the MC estimated expectation of β_1 by averaging samples up to N time steps..

Based on all these evidence, we show our MCMC estimate is accurate.

Q2.

(A) Posterior distribution $\pi(\beta_0,\beta_1,\lambda|\mathbf{Y},\mathbf{X},\boldsymbol{\alpha}) \sim \mathrm{p}(\mathbf{Y},\mathbf{X}|\beta_0,\beta_1,\lambda,\boldsymbol{\alpha})\mathrm{p}(\beta_0,\beta_1,\lambda|\boldsymbol{\alpha})$,

where $\alpha = (\mu_{\beta_0},\sigma_{\beta_0},\mu_{\beta_1},\sigma_{\beta_1},\alpha_\lambda,\beta_\lambda,\sigma_i)$ and $\beta_0 \sim N(\mu_{\beta_0},\sigma_{\beta_0}), \beta_1 \sim N(\mu_{\beta_1},\sigma_{\beta_1}), \lambda \sim \Gamma(\alpha_\lambda,\beta_\lambda)$

Thus $\pi(\beta_0,\beta_1,\lambda|\mathbf{Y},\mathbf{X},\boldsymbol{\alpha}) \sim \mathrm{h}(\beta_0,\beta_1,\lambda|\mathbf{Y},\mathbf{X},\boldsymbol{\alpha})$

$$= N(\beta_0;\mu_{\beta_0},\sigma_{\beta_0}) \times N(\beta_1;\mu_{\beta_1},\sigma_{\beta_1}) \times \Gamma(\lambda;\alpha_\lambda,\beta_\lambda) \times \prod_i EMG(Y_i;\beta_1 X_i + \beta_0,\sigma_i,\lambda),$$

where $\mathrm{h}(\beta_0,\beta_1,\lambda|\mathbf{Y},\mathbf{X},\boldsymbol{\alpha})$ is the unnormalized distribution.

Thus the full conditional distribution for each posterior variable can be written as below:

$\mathrm{h}(\beta_0| - \boldsymbol{\beta_0}) \sim N(\beta_0;\mu_{\beta_0},\sigma_{\beta_0}) \times \prod_i EMG(Y_i;\beta_1 X_i + \beta_0,\sigma_i,\lambda)$

$\mathrm{h}(\beta_1| - \boldsymbol{\beta_1}) \sim N(\beta_1;\mu_{\beta_1},\sigma_{\beta_1}) \times \prod_i EMG(Y_i;\beta_1 X_i + \beta_0,\sigma_i,\lambda)$

$\mathrm{h}(\lambda| - \boldsymbol{\lambda}) \sim \Gamma(\lambda;\alpha_\lambda,\beta_\lambda) \times \prod_i EMG(Y_i;\beta_1 X_i + \beta_0,\sigma_i,\lambda)$

Metropolis-Hasting Algorithm for approximating posterior distribution:

0)Initialize fixed parameters, such as sample size, $\mathbf{Y},\mathbf{X},\boldsymbol{\alpha}$, load data, load external function such as density of EMG variable, bm function to calculate Monte Carlo standard error.

1) Select a initial value for $(\beta_{01},\beta_{11},\lambda_1)$ = (2,3.5,0.6) as the first state of the Markov Chain, which is chosen based on preliminary trials.

2) Generate the next state through the following way:

2a) Locate the Nth state of the Markov Chain of $(\beta_{0N},\beta_{1N},\lambda_N)$. Propose the (N+1)th state $\beta^*_{0,N+1}$ for β_0 drawn from a normal distribution $\mathrm{N}(\beta_{0,N},\tau_{\beta_0})$, where τ_{β_0} is the standard deviation.

2b) Calculate the Metropolis-Hasting acceptance probability

$$\alpha(\beta_{0,N},\beta^*_{0,N+1}) = \min(1,\frac{h(\beta^*_{0,N+1}|-\boldsymbol{\beta^*_{0,N+1}})}{h(\beta_{1,N}|-\boldsymbol{\beta_{1,N}})})$$

2c) Accept $\beta^*_{0,N+1}$ with probability $\alpha(\beta_{0,N},\beta^*_{0,N+1})$, else reject y and stay at x. That is, simulate

U~uniform distribution(0,1) and if U<$\alpha(\beta_{0,N}, \beta^*_{0,N+1})$ set $\beta_{0,N+1} = \beta^*_{0,N+1}$, else set $\beta_{0,N+1} = \beta_{0,N}$. Use log scale to avoid numerical issues, that is calculate log$\alpha(\beta_{0,N}, \beta^*_{0,N+1})$=min(0,log($h(\beta^*_{0,N+1} | -\beta^*_{0,N+1})$)-log($h(\beta_{1,N} | -\beta_{1,N})$)) and compare logU with log$\alpha(\beta_{0,N}, \beta^*_{0,N+1})$.

2d)Repeat (2a)-(2c) for β_1 and λ. Notice the difference between the following $-\boldsymbol{\beta}^*_{\mathbf{0},\boldsymbol{N+1}}$=($\beta_{1N}, \lambda_N, \boldsymbol{\alpha}$), $-\boldsymbol{\beta}^*_{\mathbf{1},\boldsymbol{N+1}}$=($\beta^*_{0,N+1}, \lambda_N, \boldsymbol{\alpha}$) and $-\boldsymbol{\lambda}^*_{\mathbf{1},\boldsymbol{N+1}}$=($\beta^*_{0,N+1}, \beta^*_{1,N+1}, \boldsymbol{\alpha}$)

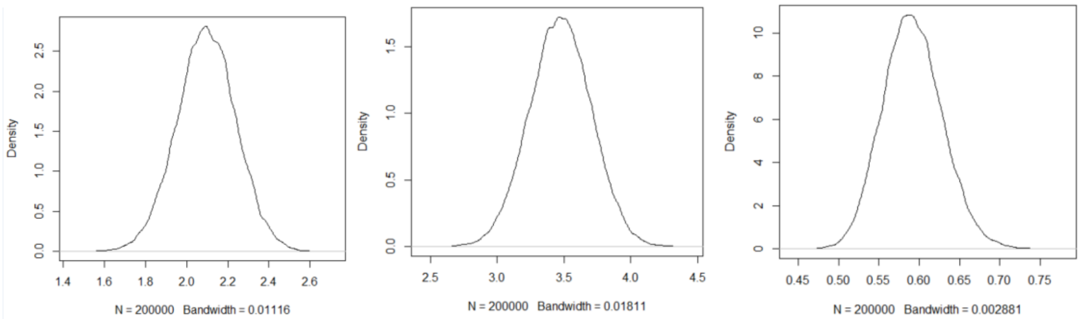
Repeat procedure 2) until draw enough samples from the posterior distribution.

(b) MCMC Sample size N =200000,

	Posterior mean	Estimated MCMC SE	95%credible intervals
β_0	2.0956	0.0019	(1.8098,2.3787)
β_1	3.4742	0.0030	(3.0178,3.9234)
λ	0.5919	0.00027	(0.5237,0.6671)

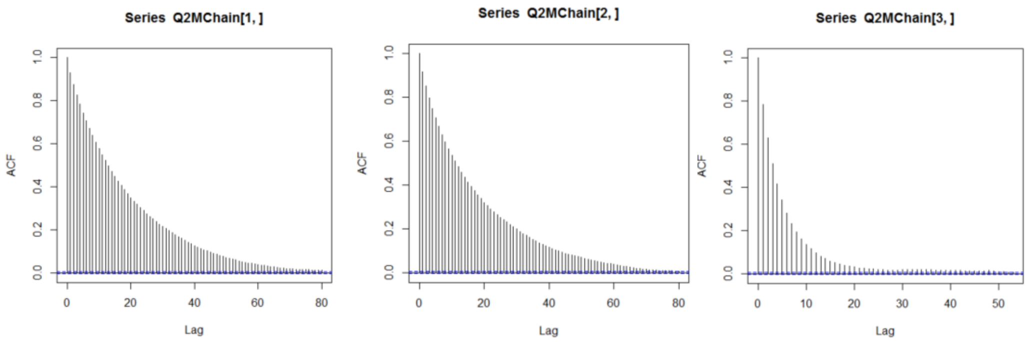
(c) Pearson Correlation between β_0 *and* β_1 is -0.8059.

(d)Marginal density plot for $\beta_0, \beta_1, \lambda$ respectively.



(e) We show the validity of result through the following evidence:

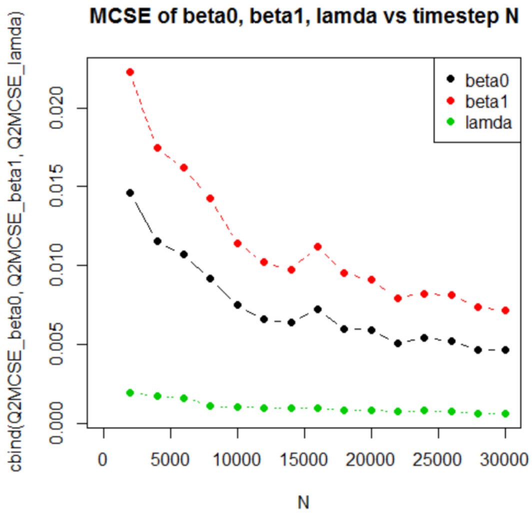
(1) Auto Correlation of the states ($\beta_0, \beta_1, \lambda$) in Markov Chain is shown in the figure below respectively. The correlation decrease slowly with lag, however, it did decay to 0. Thus one could increase the sample size to generate more effective draws. The acceptance rate for $\beta_0, \beta_1, \lambda$ is around 27%, 28%, 31% when N=200000, which is also appropriate. ESS is 5655, 6042 and 20242 respectively for $\beta_0, \beta_1, \lambda$.

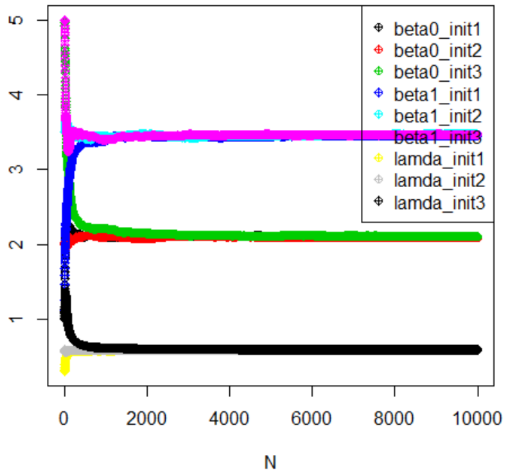


(2) I increase the sample size N from 2000 to 30000 by a step of 2000, draw the MCMC SE versus Sample size N as shown in the left graph below. One can see the MCMC SE tend to decrease when N increase, which is a good indication of the validity of the result. Also the MCMC SE is much smaller in magnitude than the expectation when N=200000.(See table in part (b)).

(3) Also I try to compare Markov Chain states sampled from different initial states, MC estimated expectation of $\beta_0, \beta_1, \lambda$ all converge to the same interval as shown in the right figure below, which is also another good indication. In the figure, y axis is the MC estimated expectation of $\beta_0/\beta_1/\lambda$ by averaging samples up to N time steps. The three initial conditions are (1,1,0.4), (2,3.5,0.6), (5,5,1) in the order of $(\beta_0, \beta_1, \lambda)$

Based on all these evidence, we show our MCMC estimate is accurate.





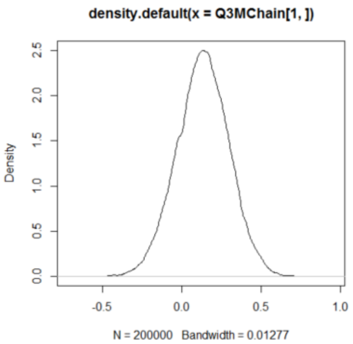
Q3.

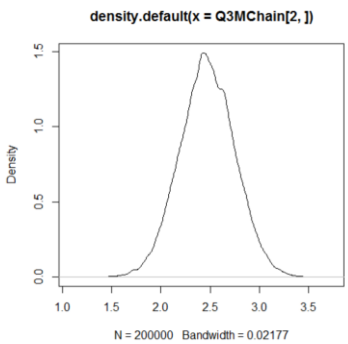
(A)

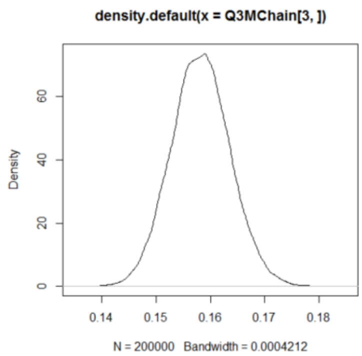
Based on MCMC with sample size N=200000. Initial condition (0.1,2.5,0.15)

	Posterior mean	Estimated MCMC SE	95%credible intervals
β_0	0.1329	0.0022	(-0.1952,0.4510)
β_1	2.4715	0.0037	(1.9289,3.0164)
λ	0.1582	0.00003	(0.1478,0.1690)

(B) Marginal density plot for $\beta_0, \beta_1, \lambda$ respectively.







(C)I changed the standard deviation of the proposed distribution for $(\beta_0, \beta_1, \lambda)$ in the Metropolis-Hasting Algorithm $(\tau_{\beta_0}, \tau_{\beta_1}, \tau_\lambda)$ from (0.3,0.5,0.11) in Q2 to (0.5,0.75,0.025) Q3. This makes the acceptance rate around 25% for every variable and auto-correlation not too large. One can see the MCMC SE is still much smaller than the estimated expectation in magnitude.

ESS is 5048, 5231 and 26566 for $\beta_0, \beta_1, \lambda$ respectively.

I repeat the process to draw how MCMC SE changes when sample size N increases. One can see SE is decreasing as N increase.

Also I try to compare Markov Chain states sampled from different initial states, MC estimated expectation of $\beta_0, \beta_1, \lambda$ all converge to the same interval as shown in the right figure below, which is also another good indication. In the figure, y axis is the MC estimated expectation of $\beta_0/\beta_1/\lambda$ by averaging samples up to N time steps. The three initial conditions are (-1,1,0.05), (0.1,2.5,0.15), (4,4,1) in the order of $(\beta_0, \beta_1, \lambda)$.