Exam Report

Bing Yao

Problem 1

(a) Metropolis-Hasting algorthm

(1) Read data. Set X=the first column, Y=the second colum.
(2) Calculate log of the joint distribution

$$log.h = log(\pi(\beta_1|Y,X)) = \sum_i log(EMF(5+\beta_1 X, 1, 0.4)) - \frac{\beta_1^2}{200} \tag{1}$$

(3) Start off the initial value $\beta_1^{(0)} = 0$ (suggested, doesn't matter which value is picked).
(4) Generate a candidate $y* \sim N(\beta_1^{(t)}, \tau)$, ($\tau$ is the tuning parameter).
(5) Let $\beta_1^{(t+1)} = y*$ with $probability = min(1, exp(log.h(y*) - log.h(\beta_1^{(t)})))$.
(6) Loop back to step (4).

(b) Estimation and M.C.se with sample size $N = 40000$

$$E(\beta_1|Y,X) = 7.3414 \tag{2}$$
$$MC.se = 0.00390 \tag{3}$$

with sample size $N = 10000$, tuning parameter $\tau = 1$

(c) 95% credible interval, with N=40000

$$\beta_1 = (6.7259, 7.9241) \tag{4}$$

(d) Density plot of β_1 in figure1

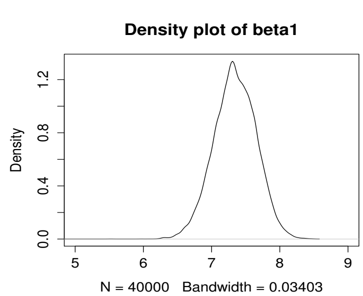


Figure 1: Density plot of β_1 with MC sample size $N = 40000$.

(e) supporting plots

According to figure2, MC.se is quite small when sample size $N > 20000$, and auto-correlation plot guarantees the quality of my sample. $ESS = 7580$ for my sample, which is another evidence.

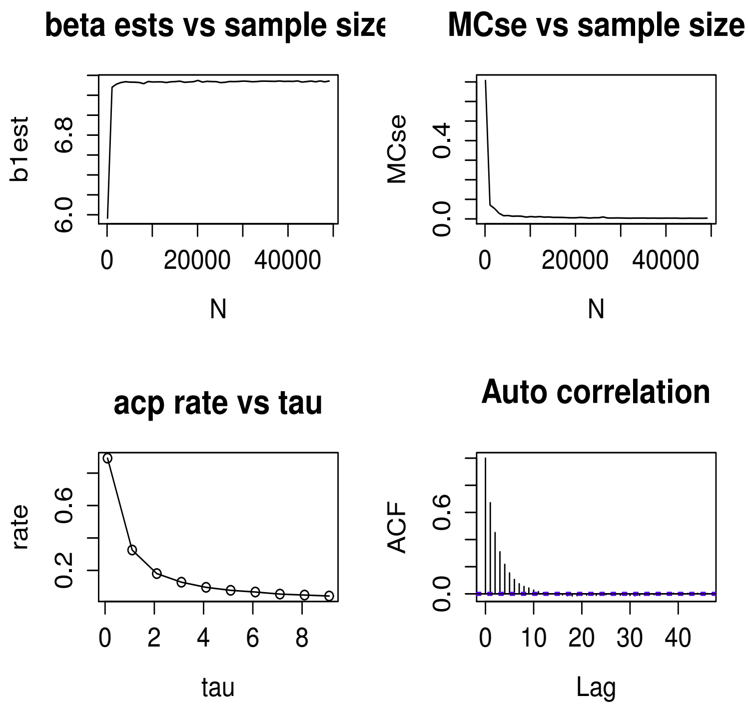


Figure 2: Estimations and MC.ses vs sample size .

Problem 2

(a) Variable at a time Metropolis Hasting algorithm

(1) Read data. Set X=the first column, Y=the second colum.
(2) Calculate log of the joint distribution

$$\begin{aligned} log.h &= log(\pi(\beta_0, \beta_1, \lambda|Y, X)) \\ &= \sum_i log(EMF(\beta_0 + \beta_1 X, 1, \lambda)) - \frac{\beta_0^2}{200} - \frac{\beta_1^2}{200} - \frac{\lambda}{100} + (0.01 - 1) * log(\lambda) \end{aligned} \tag{5}$$

(3) Start off the initial vector $(\beta_0, \beta_1, \lambda) = (1, 1, 1)$ (suggested, doesn't matter which initial vector is picked.)

(4) Generate $y_1^* \sim N(\beta_0^{(t)}, \tau)$, and let $\beta_0^{(t+1)} = y_1^*$ with

$$prob = min(1, exp(log.h(y_1^*, \beta_1^{(t)}, \lambda^{(t)}) - log.h(\beta_0^*, \beta_1^{(t)}, \lambda^{(t)}))) \tag{6}$$

Generate $y_2^* \sim N(\beta_1^{(t)}, \tau)$, and let $\beta_1^{(t+1)} = y_2^*$ with

$$prob = min(1, exp(log.h(\beta_0^{(t+1)}, y_2^*, \lambda^{(t)}) - log.h(\beta_0^{(t+1)}, \beta_1^{(t)}, \lambda^{(t)}))) \tag{7}$$

Generate $y_3^* \sim exp(\frac{1}{\lambda^{(t)}})$, and let $\lambda^{(t+1)} = y_3^*$ with

$$prob = min(1, exp(log.h(\beta_0^{(t+1)}, \beta_0^{(t+1)}, y_3^*) - log.h(\beta_0^{(t+1)}, \beta_1^{(t+1)}, \lambda^{(t)}) - \frac{y_3^*}{\lambda^{(t)}} + \frac{\lambda^{(t)}}{y_3^*} + log\frac{y_3^*}{\lambda^{(t)}})) \quad (8)$$

(5) Look back to step (4).

(b) Estimation with sample size $N = 270000$

	Estimation	95% interval	MC.se	Ess
β_0	2.3053	(1.9740, 2.6018)	0.00402	5145
β_1	3.4479	(2.9854,3.9003)	0.00317	7991
λ	0.7819	(0.6504 0.9153)	0.00157	5105

Table 1: Estimations

(c) Auto-correlation pf β_0 and β_1

$$cor(\beta_0, \beta_1) = 0.0478 \quad (9)$$

(d) Density plots

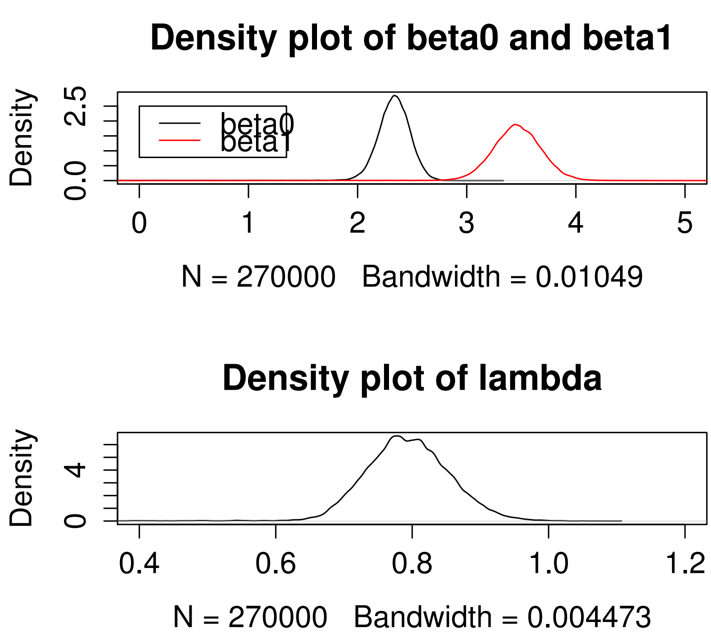


Figure 3: Density plots of β_0, β_1 and λ.

(e) Supporting plots

From figure4 and the ESS values in table1, we can conclude that the estimations are accurate.

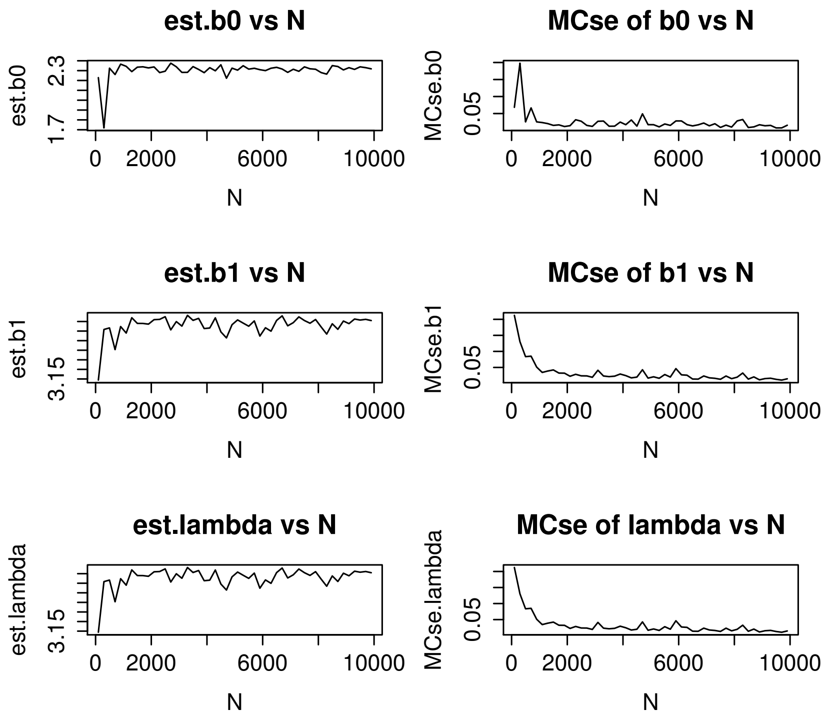


Figure 4: Estimations and MC.se vs sample size.

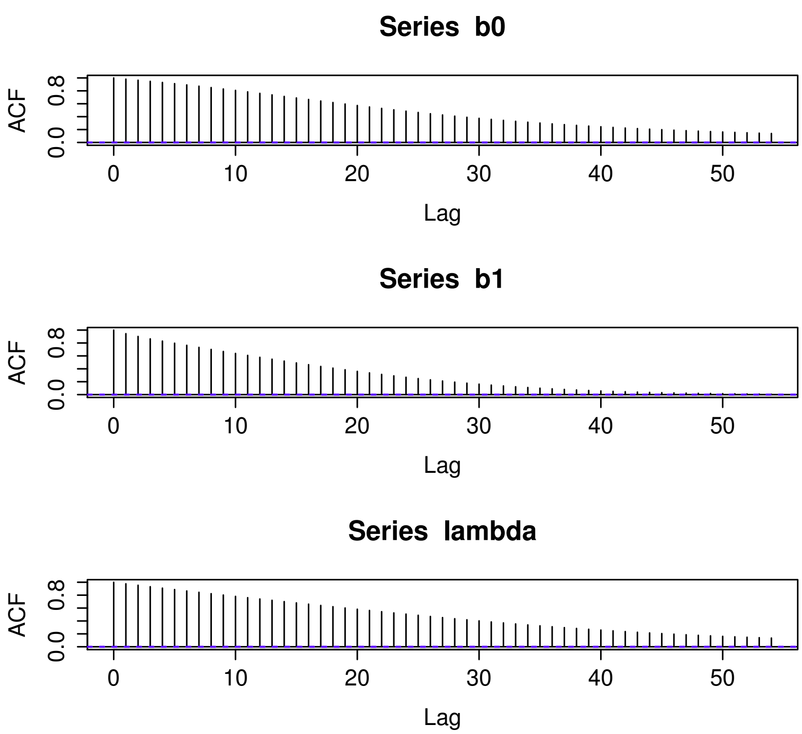


Figure 5: Auto-correlation of samples.

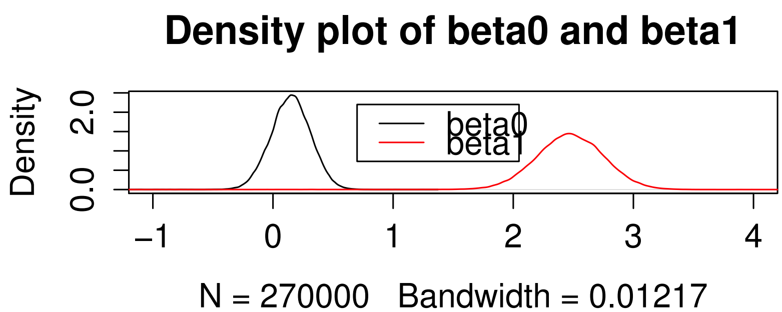
Problem 3

(a) Estimation with sample size $N = 270000$

	Estimation	95% interval	MC.se	Ess
β_0	0.1392	(-0.1890, 0.4645)	0.002600	7204
β_1	2.4639	(1.9007, 3.0161)	0.00379	7718
λ	0.1602	(0.1497, 0.1721)	0.00016	5477

Table 2: Estimations

(b) Density plots



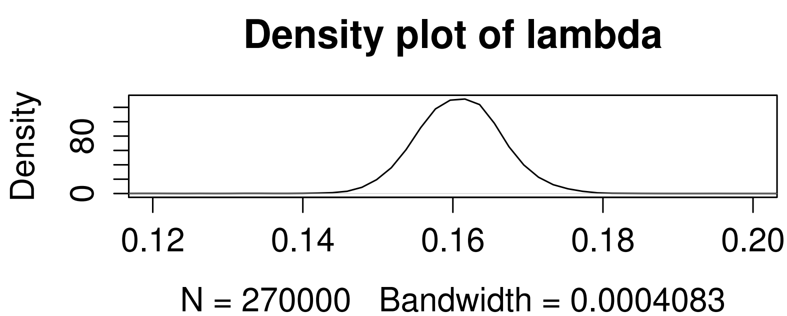


Figure 6: Density plots.

(c) Methods tried for improvement

(1) Adjust tuning parameters in β_0 and β_1 updates to reduce auto-correlation and improve ESS value.

(2) Try different initial vectors to reduce MC.se.

(3) Run long enough to reduce MC.se.