STAT 515 Spring 2015
Take-home Final
Wanjun Liu

1.(a) Let $g(\beta_1; \beta_0, \sigma, \lambda, \boldsymbol{x}, \boldsymbol{y})$ be the posterior for β_1. $h(.)$ is proportional to $g(.)$ up to a normalizing constant.

$$\begin{aligned} h(\beta_1) &\propto g(\beta_1; \beta_0 = 5, \sigma = 1, \lambda = 0.4, \boldsymbol{x}, \boldsymbol{y}) \\ &= \left(\prod_{i=1}^{n} f(y_i; \mu = 5 + \beta_1 x_i, \sigma = 1, \lambda = 0.4) \right) \phi(\beta_1 | 0, 10) \end{aligned}$$

where $f(y|\mu, \sigma, \lambda)$ is the density of EMG(μ, σ, λ), $\phi(\beta_1|\mu, \sigma)$ is the density of $N(\mu, \sigma^2)$ (with mean μ and standard deviation σ). I use normal distributions with variance τ^2 as the proposal. A brief description of the algorithm is as follows,

Initialization: sample size n, initial value $x^{(1)}$, τ^2, flag= 1;
for i *in* $2 : n$ **do**
  draw $y \sim N(x^{(i-1)}, \tau^2)$;
  compute $p = \min\left(1, \exp\{\log h(y) - \log h(x^{(i-1)})\}\right)$;
  accept $x^{(i)} \leftarrow y$ with probability p, otherwise $x^{(i)} \leftarrow x^{(i-1)}$;
  if $x^{(i)} \leftarrow y$ **then**
    flag $\leftarrow$ flag+1
  end
end
Return $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$; acceptance rate = flag/n.

The starting value is chosen as 7 based on several preliminary MCMC runs. The variance of the normal proposal τ^2 should not be too large or too small. I use $\tau = 0.9$ based on MCMC standard error and autocorrelation. It also gives a reasonable acceptance rate, which is around 0.38.

1.(b) & (c)
The point estimates, MCMC standard errors, 95% credible intervals and ESS for β_1 with different sample sizes are summarized in Table 1.

	Sample size	Mean (MCse)	95% CI	ESS
1	1000	7.3638 (0.0258)	(6.7580, 7.9315)	235
2	5000	7.3499 (0.0112)	(6.7223, 7.9302)	1290
3	10000	7.3472 (0.0082)	(6.7148, 7.9317)	2461
4	20000	7.3431 (0.0060)	(6.7223, 7.9403)	4642
5	50000	7.3431 (0.0039)	(6.7206, 7.9265)	11231
6	100000	7.3402 (0.0027)	(6.7183, 7.9246)	22386

Table 1: Summary of samples with different sizes

1.(d) Figure 1 gives the smoothed density plot of the sample (with sample size 100000).

1.(e) Based on Figure.2, an M-H sample with sample size 10^5 should be accurate. The upper left plot shows that MCMC standard error decreases dramatically when sample size is relatively small and approaches to 0 stably when sample size becomes larger. The

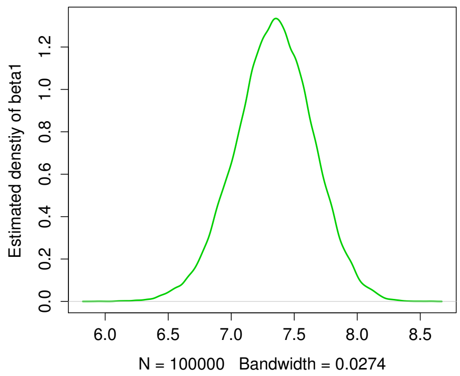

Figure 1: Estimated density of β_1

upper right plot shows the posterior mean for β_1 with three different starting values 0, 7 , 14. Even though the starting values are different, all the three point estimates converge to the same value (around 7.34) as sample size increases. The plot of acf suggests a weak autocorrelation in the sample. For the lower right plot, the green solid line is the estimated density using the entire sample ($n = 10^5$) while the blue dash line the estimated density using half of the sample. These two estimated densities match very well. Also, the ESS with $n = 10^5$ is around 22000 (greater than 5000). As a result, we can say with sample size 10^5, the estimate of β_1 should be accurate.

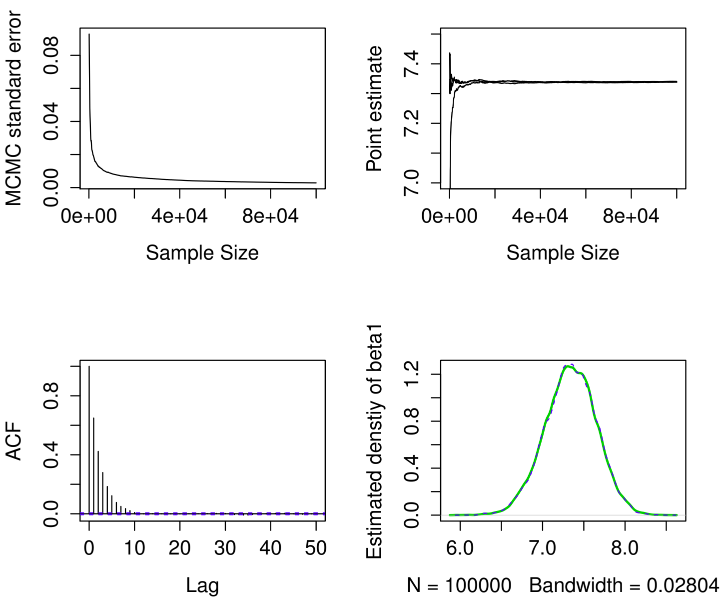

Figure 2: Estimated density of β_1

2.(a) Let $g(\beta_0, \beta_1, \lambda; \sigma, \boldsymbol{x}, \boldsymbol{y})$ be the posterior for $(\beta_0, \beta_1, \lambda)$ given $\boldsymbol{x}$ and $\boldsymbol{y}$. Let $h_0(\beta_0)$, $h_1(\beta_1)$ and $h_2(\lambda)$ be the full conditional densities for β_0, β_1 and λ (up to normalizing constants), all of which are all proportional to $g(.)$. Thus a lazy way is treating $g(.)$ (with other parameters fixed) as h_0, h_1 and h_2.

$$g(\beta_0, \beta_1, \lambda; \sigma = 1, \boldsymbol{x}, \boldsymbol{y})$$
$$= \left(\prod_{i=1}^{n} f(y_i; \mu = \beta_0 + \beta_1 x_i, \sigma = 1, \lambda) \right) \phi(\beta_0|0, 10)\phi(\beta_1|0, 10)\gamma(\lambda|0.01, 100)$$

where $f(y|\mu, \sigma, \lambda)$ and $\phi(\beta|\mu, \sigma)$ are the same as above. $\gamma(\lambda|k, \theta)$ is the density for Gamma distribution with shape $= k$ and scale $= \theta$. Let starting values be $\beta_0^{(1)} = 2.35$, $\beta_1^{(1)} = 3.45$ and $\lambda^{(1)} = 0.8$ (based on several preliminary MCMC runs). I use normal proposals for β_0 and β_1 and use a truncated normal proposal (only takes positive values) for λ . Let $\tau_0 = 0.2$, $\tau_1 = 0.3$, $\tau_2 = 0.1$ be the tuning parameters for β_0, β_1 and λ, respectively.

I use a variable-at-a-time M-H algorithm for this problem. A brief description of the algorithm is as follows:

Initialization: n, $\beta_0^{(1)}$, $\beta_1^{(1)}$, $\lambda^{(1)}$, τ_0^2, τ_1^2, τ_2^2, $\text{flag}_0 = \text{flag}_1 = \text{flag}_2 = 1$;
for i *in* $2:n$ **do**
draw $y_0 \sim N(\beta_0^{(i-1)}, \tau_0^2)$;
let $p_0 = \min\left(1, \exp\{\log h_0(y_0|\beta_1^{(i-1)}, \lambda^{(i-1)}) - \log h_0(\beta_0^{(i-1)}|\beta_1^{(i-1)}, \lambda^{(i-1)})\}\right)$;
accept $\beta_0^{(i)} \leftarrow y_0$ with probability p_0, $\text{flag}_0 \leftarrow \text{flag}_0 + 1$;
otherwise $\beta_0^{(i)} \leftarrow \beta_0^{(i-1)}$;
draw $y_1 \sim N(\beta_1^{(i-1)}, \tau_1^2)$;
let $p_1 = \min\left(1, \exp\{\log h_1(y_1|\beta_0^{(i)}, \lambda^{(i-1)}) - \log h_1(\beta_1^{(i-1)}|\beta_0^{(i)}, \lambda^{(i-1)})\}\right)$;
accept $\beta_1^{(i)} \leftarrow y_1$ with probability p_1, $\text{flag}_1 \leftarrow \text{flag}_1 + 1$;
otherwise $\beta_1^{(i)} \leftarrow \beta_1^{(i-1)}$;
draw $y_2 \sim N(\lambda^{(i-1)}, \tau_2^2)I(x > 0)$;
let $p_2 = \min\left(1, \exp\{\log h_2(y_2|\beta_0^{(i)}, \beta_1^{(i)}) - \log h_1(\lambda^{(i-1)}|\beta_0^{(i)}, \beta_1^{(i)})\}\right)$;
accept $\lambda^{(i)} \leftarrow y_2$ with probability p_2, $\text{flag}_2 \leftarrow \text{flag}_2 + 1$;
otherwise $\lambda^{(i)} \leftarrow \lambda^{(i-1)}$;
end
Return $\boldsymbol{\beta}_0 = (\beta_0^{(1)}, \beta_0^{(2)}, \ldots, \beta_0^{(n)})$, $\boldsymbol{\beta}_1 = (\beta_1^{(1)}, \beta_1^{(2)}, \ldots, \beta_1^{(n)})$,
$\boldsymbol{\lambda} = (\lambda^{(1)}, \lambda^{(2)}, \ldots, \lambda^{(n)})$; flag_0/n, flag_1/n, flag_2/n.

2.(b) The posterior mean, MCMC standard error, 95% CI, acceptance rate and ESS for β_0, β_1 and λ are summarized in Table.2 (with sample size 2×10^5).

	expectation (MCse)	95% CI	Acpt.rate	ESS
β_0	2.3522 (0.001667)	(2.0791, 2.6078)	0.36	6455
β_1	3.4600 (0.002432)	(3.0593, 3.8725)	0.40	7196
λ	0.8078 (0.0004530)	(0.6979, 0.9326)	0.45	17507

Table 2: Mean, MCse and 95% CI for β_0, β_1 and λ

2.(c) The approximate correlation between β_0 and β_1 is -0.7841073.
2.(d) Figure 3 gives the estimated marginal densities for for β_0, β_1 and λ.

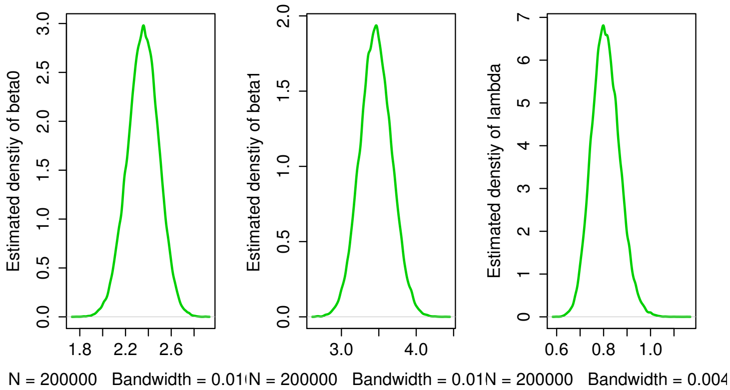

Figure 3: Estimated densities of marginal distributions

2.(e) Based on Figure 4, a sample with sample size 2×10^5 should reliable. The three plots in the first column shows that MCMC standard errors approach to 0 stably as sample sizes increase for all three parameters. The plots in the middle column show that even though the starting values are different (there are 3 lines in each of the plots in the middle column, I use different starting values for each line), the posterior mean converges to the same value as sample size increases. The plots in the last column shows the autocorrelations in samples, which is not very strong. Also, the ESS for each parameter is greater than 5000 (see Table 2). Thus an M-H sample with sample size 2×10^5 should reliable.

3.(a) The posterior mean, MCMC standard error, 95% CI, acceptance rate and ESS for β_0, β_1 and λ are summarized in Table.3 (with sample size 2×10^5).

	Mean (MCse)	95% CI	Acpt.rate	ESS
β_0	0.1489 (2.0092×10^{-3})	(-0.1770, 0.4670)	0.32	6700
β_1	2.4757 (3.3831×10^{-3})	(1.9258, 3.0180)	0.33	6791
λ	0.1613 (3.1053×10^{-5})	(0.1508, 0.1724)	0.30	31879

Table 3: Mean, MCse and 95% CI for β_0, β_1 and λ

3.(b) Figure 5 gives the estimated marginal densities for for β_0, β_1 and λ.

3.(c) Basically, I use the same algorithm in problem# 2 but with different starting values and tuning parameters. The starting values are chosen to be $\beta_0^{(1)} = 0.15$, $\beta_1^{(1)} = 2.45$ and $\lambda^{(1)} = 0.16$ based on several preliminary MCMC runs. The tuning parameters are chosen to be $\tau_0 = 0.3$, $\tau_1 = 0.5$ and $\tau_2 = 0.02$ based on the MCMC standard error, autocorrelation and acceptance rates. Compared with problem #2, I use smaller variances for proposals for this problem.

Alternatively, if we take a look at the scatterplot of `EMG3.dat`, we notice that the data is roughly split into two groups. We may consider that the data is actually from two models. To be more specific, we can divide the data into two groups with some criterion and then we can fit the data from the same group separately.

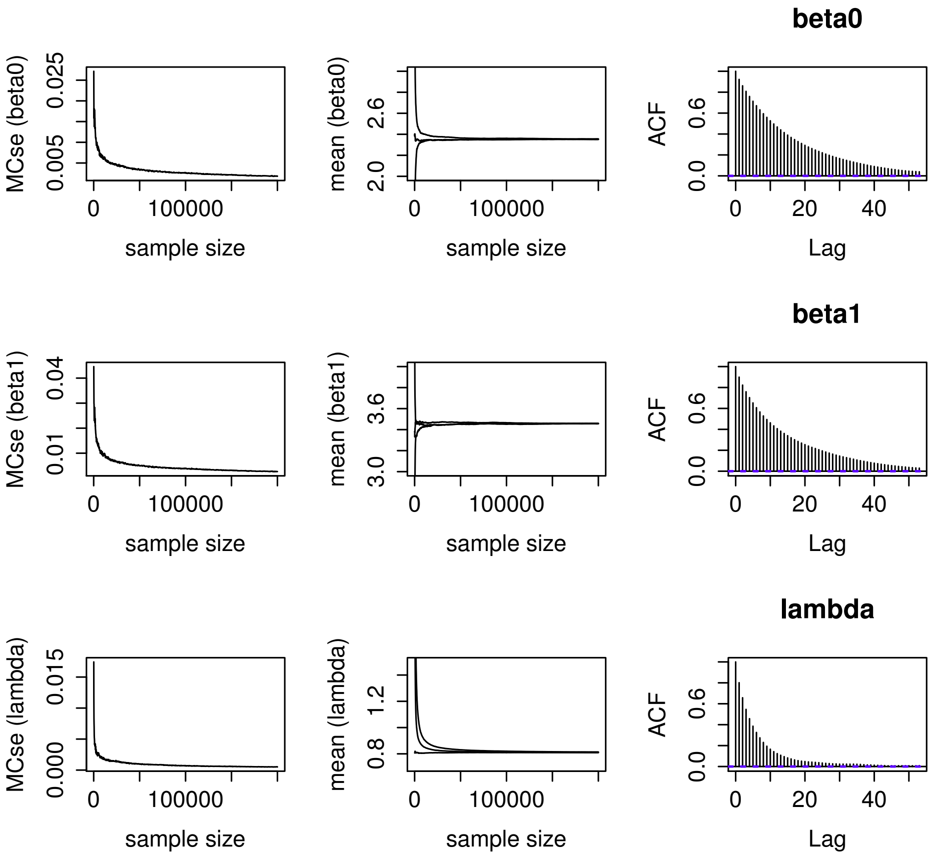

Figure 4: MCse, posterior mean and acf for β_0, β_1 and λ

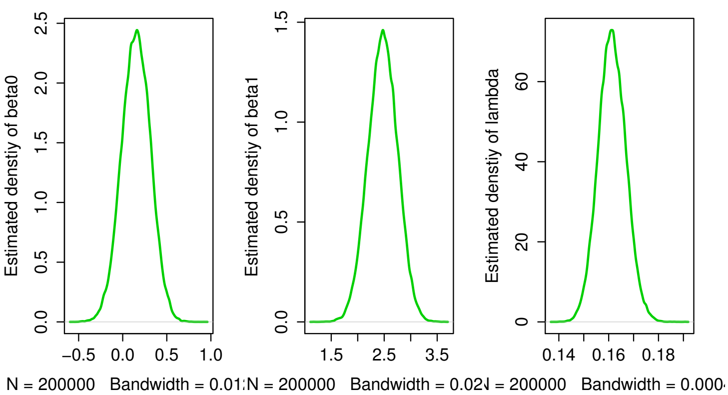

Figure 5: Estimated densities for β_0, β_1 and λ