# Take Home Final

Jacob Parsons

April 29, 2015

1. (a) The posterior distribution in the problem is such that:

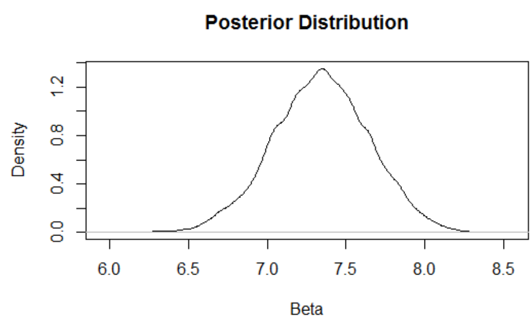
$$\pi(\beta_1 \mid X, Y) \propto \pi(\beta_1) \prod_{i=1}^{n} f(Y_i; \mu_i, 1, .4) \tag{1}$$

where $f(x, \mu, \sigma, \lambda)$ is the density given in the problem, $\pi(x)$ is a normal density with mean 0 and standard deviation 10, and $\mu_i = 5 + \beta_1 X_i$. The right hand side of equation 1 shall be taken to be the unnormalized posterior, $h(\beta_1)$, and is our target density. We shall form a Markov chain, *chain*. After running the algorithm multiple times, the choice of starting point was not too important. So, I chose the starting point $chain_1 = 7.34$. The starting point was the end result of a previous run. I update the chain until the effective sample size is greater than 5000 (where the effective sample size is updated every 1000 updates to the chain). To update the chain from $chain_i$ to $chain_{i+1}$, a proposal value, $Z$, is selected randomly according to a $N(chain_i, \tau^2)$ distribution, where I settled on $\tau = 1$. Next, a random variable $U$ is sampled according to $Uniform(0, 1)$. If $U < \frac{h(\beta_1)}{h(Z)}$ then we set $chain_{i+1} = Z$, otherwise we set $chain_{i+1} = chain_i$. This condition is checked on the log scale. It is at this point that the effective sample size is computed using the provided function is $i + 1$ is a multiple of 1000.

   (b) Taking the mean of the sample posterior gives us the estimate $\widehat{\beta_1} = 7.342098$. The associated MCMC standard error is 0.004393592.

   (c) I obtain a 95% credible by taking the $2.5th$ and $97.5th$ sample percentiles. This gives the credible interval$(6.723071, 7.935310)$

   (d) An estimate of the posterior pdf of $\beta_1$ is shown below:




   (e) The first thing that I considered was how the MCMC standard error changed with sample size. Figure 2 is a plot of the standard errors showing that as the sample size increased the standard error rapidly approached zero.Next, I looked at how different choices of starting values changed the estimates given by the chain. Figure 3 shows how the estimate changed with sample size for three different chains each with unique starting values. It can be seen that in the long run the estimate does not depend on the choice of starting values. Finally, a look at a plot of the autocorrelation function (figure 4) shows that the autocorrelations quickly vanish as the lag increases.

Figure 2

MCMC Standard Error vs Sample Size

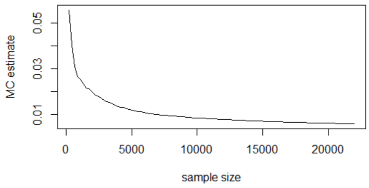



Figure 3

Estimates versus sample size

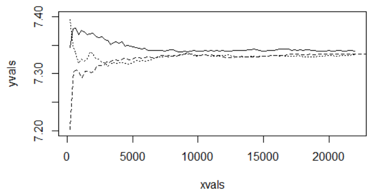
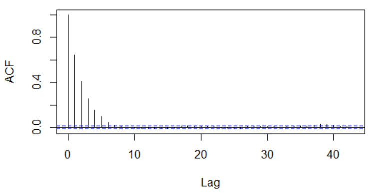



Figure 4

Autocorrelation Function




2. (a) The posterior distribution in the problem is such that:

$$\pi(\beta_0, \beta_1, \lambda \mid X, Y) \propto \pi_0(\beta_0)\pi_0(\beta_1)\pi_1(\lambda) \prod_{i=1}^{n} f(Y_i; \mu_i, 1, \lambda) \tag{2}$$

where $f(x, \mu, \sigma, \lambda)$ is the density given in the problem, $\pi_0(x)$ is a normal density with mean 0 and standard deviation 10, $\pi_1(x)$ is the density of a $Gamma(.01, 100)$ distribution , and $\mu_i = \beta_0 + \beta_1 X_i$. In order to form the Markov chain, I first took $\beta_0^1 = 2.289$, $\beta_1^1 = 3.4631$, and $\lambda^1 = .7841$. These values were values for the last result in a previous run. Next, while the minimum effective sample size among the chains of coordinates of the chain is less than 5000 (this is updated every 10,000 samples) we update the chain as follows: first update $\beta_0$. We sample a proposal,$P$ , from a $N(\beta_0^i, .8)$ distribution. The tuning constant was chosen after several tries. Since the proposal distribution is symmetric we generate $U$ according to $uniform(0, 1)$ and accept the proposal if $U$ is less than

$$\frac{\pi(P \mid \beta_1^i, \lambda^i, X, Y)}{\pi(\beta_0^i \mid \beta_1^i, \lambda^i, X, Y)} = \frac{\pi(P, \beta_1^i, \lambda^i \mid X, Y)}{\pi(\beta_0^i, \beta_1^i, \lambda^i \mid X, Y)}. \tag{3}$$

As before, if we accept then we set $\beta_0^{i+1} = P$ otherwise we set $\beta_0^{i+1} = \beta_0^i$. Next, we update $\beta_1$ in the same matter except that we compare $U$ to

$$\frac{\pi(P \mid \beta_0^{i+1}, \lambda^i, X, Y)}{\pi(\beta_1^i \mid \beta_0^{i+1}, \lambda^i, X, Y)} = \frac{\pi(\beta_0^{i+1}, P, \lambda^i \mid X, Y)}{\pi(\beta_0^{i+1}, \beta_1^i, \lambda^i \mid X, Y)} \tag{4}$$

and $P$ is chosen according to $N(\beta_1^i, .8)$. Next we update $\lambda$ using the same method with $P$ chosen according to $exp(1/\lambda)$ and where $U$ is compared to

$$\frac{\pi(P \mid \beta_0^{i+1}, \beta_1^{i+1}, X, Y)}{\pi(\lambda^i \mid \beta_0^{i+1}, \beta_1^{i+1}, X, Y)} = \frac{\pi(\beta_0^{i+1}, \beta_1^{i+1}, P \mid X, Y)}{\pi(\beta_0^{i+1}, \beta_1^{i+1}, \lambda^i \mid X, Y)}. \tag{5}$$

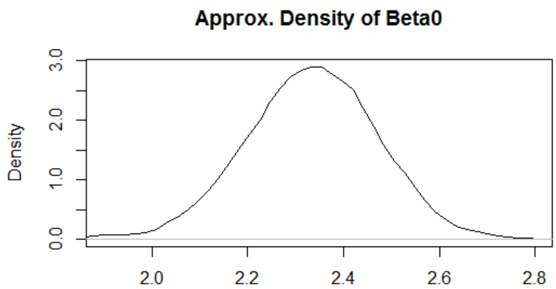
Note that each of the three previous equalities hold since each of the full conditionals is proportional to the posterior (and so the constants of proportionality cancel out in the ratios). Similarly we may use the ratios of the right hand side of (2) in the ratio instead of the posterior, which is what is done in the code. Also, each of the comparisons is done on the log scale. Indeed, the posterior is actually calculated on this scale. It is at this point that the effective sample size is updated if the chain's length is a multiple of 10,000.

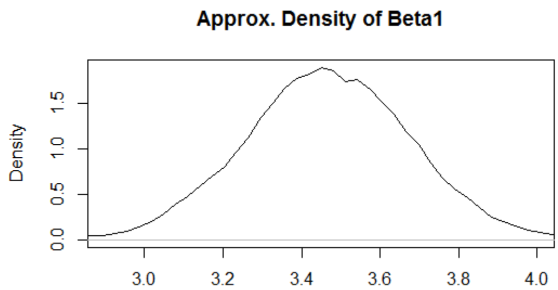
(b)

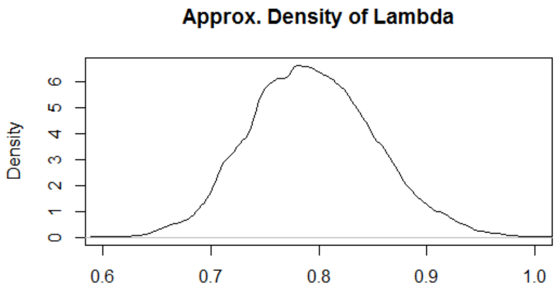
| | | $\hat{\mu}$ (MCMC Standard Error) | 95% Credible interval |
|---|---|---|---|
| | $\beta_0$ | 2.289 (.00597) | (1.906, 2.598) |
| | $\beta_1$ | 3.448 (.00520) | (3.005, 3.924) |
| | $\lambda$ | .7819 (.00169) | (.653, .917) |

(c) An estimate of the correlation between $\beta_0$ and $\beta_1$ is .2358.

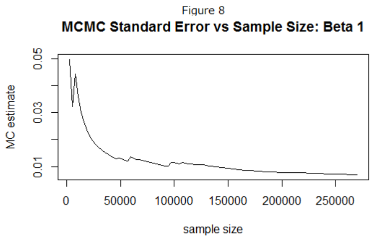
(d) The approximate densities are shown below:

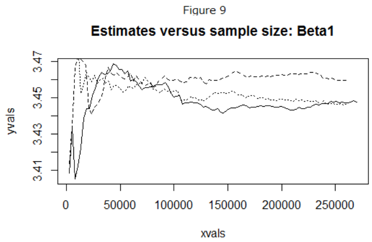


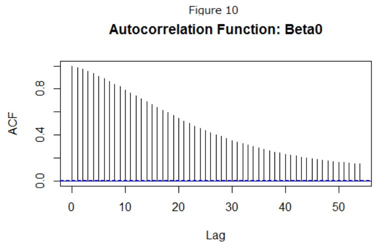




(e) As before, the first thing that I looked at was how the MCMC standard error changed as N increased for each parameter. The least impressive of the plots is shown in figure 8. As can be seen, the only real defect here is a spike near the start but the plot looks more promising later. This defect was not present in the other two charts. So, overall, there is no cause for alarm here. Next, I looked at how different starting values effected the result Figure 9 shows the least impressive plot of three chains for a parameter. Despite having visibly different results in the end (this was the only of the charts to show such a big difference), the total distance between the results is less than a percent of the estimate and is also within two standard error of the other estimates. This suggests that the difference is not cause for worry. The autocorrelation functions were much worse than they were in problem 1. The worst of them is shown in figure 10. This suggests that we might have a bit of bias in our estimates, however, it does decrease substantially with time, and the parameters with the largest correlation issues also seem to be fairly robust to choosing different starting values. So, the estimates are probably fairly accurate.



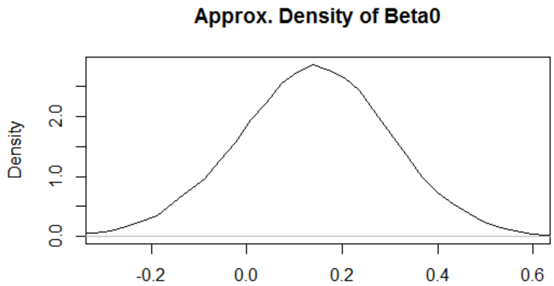



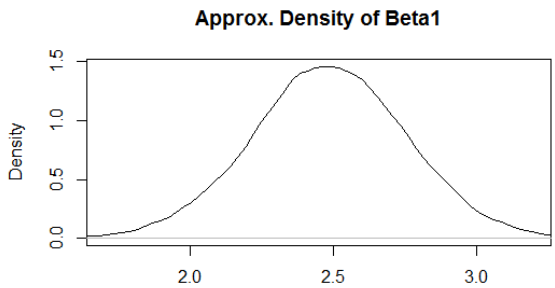



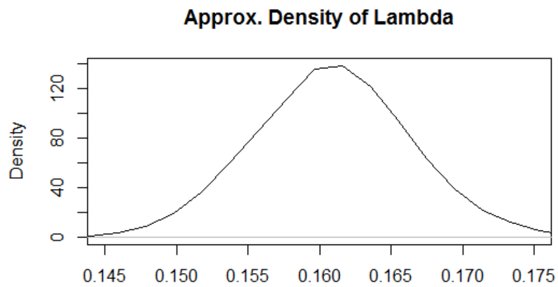

3. (a) .

| | | $\hat{\mu}$ (MCMC Standard Error) | 95% Credible interval |
|---|---|---|---|
| | $\beta_0$ | .1267 (.00390) | (-.2032, .4522) |
| | $\beta_1$ | 2.476 (.00508) | (1.928, 3.033) |
| | $\lambda$ | .1601 (.00018) | (.1496, .1719) |

(b) The approximate posterior densities for the parameters are shown below:




(c) I made very minimal changes to the algorithm. First I noticed that it was taking longer to run, so I lowered the frequency at which it checked the effective sample size. Also, different tuning parameters than before were settled on after testing a few.