

Monte Carlo Methods HW 4

Kylie Taylor

4/4/2019

We have the weighted normal density of

$$k(x|\theta) = wN(x|\mu_1, \sigma^2) + (1 - w)N(x|\mu_2, \sigma^2)$$

With priors $\lambda = \frac{1}{\sigma^2} \sim Ga(1, 1)$, and $f(\mu_1) \sim N(0, 100)$, also $f(\mu_2) \sim N(0, 100)$, and lastly $f(w) \sim U(0, 1)$

1. Generate data of size $n=100$ from a standard normal distribution and run the Markov chain Gibbs sampler for the above model using the data you obtained. The appropriate latent variable here is the d which determines which group each observation came from.

This means that the d , the latent variable, will show up as a subscript denoting which group (1 or 2) that each observation for x came from.

Let $d_i = 1$ if x_i comes from $N(\mu_1, \sigma^2)$, and $d_i = 2$ if x_i comes from $N(\mu_2, \sigma^2)$. Where $w = w_1$ and $1 - w = w_2$. Keep in mind that $f(d_i)$ does not have a prior distribution, as it is a latent variable.

The goal is to find the joint density

$$k(x_i, d_i, \mu_1, \mu_2, w, \lambda)$$

In order for the latent variable, d to be included, the following must be true

$$\sum_{d_i=1,2} k(x_i, d_i, \mu_1, \mu_2, w, \lambda) = k(x_i, \mu_1, \mu_2, w, \lambda)$$

The full likelihood is :

$$\begin{aligned} k(x_i, d_i, \mu_1, \mu_2, w, \lambda) &= f(\mu_1)f(\mu_2)f(w)f(\lambda) \prod_{i=1}^n k(x_i|\mu_1, \mu_2, w, \lambda) \\ &= N(0, 100) * N(0, 100) * Beta(1, 1) * Ga(1, 1) * \prod_{i=1}^n wN(x|\mu_1, \sigma^2) + (1 - w)N(x|\mu_2, \sigma^2) \end{aligned}$$

I will continue to find the posterior conditional densities of the parameters

I will start with the posterior for the weights

$$f(w|x, d) = f(w) \prod_i w_{d_i} N(x_i|u_{d_i}, \sigma^2) \propto 1 * \frac{w_1^{n_1} w_2^{n_2}}{\sqrt{2\pi\sigma^2}} e^{-\frac{\lambda}{2} \sum_i (x_i - \mu_{d_i})^2}$$

$$\propto w_1^{n_1} w_2^{n_2} = w^{n_1} (1-w)^{N-n_1}$$

This means that

$$f(w|x_i, d) \sim \text{Beta}(n_1 + 1, n_2 + 1)$$

We are also interested in $f(d|x_i, \lambda, \mu_1, \mu_2, w)$ for $d_i = 1, 2$ which we know is

$$Pr(d_i = 1|x_i, \lambda, \mu_1, \mu_2, w) = \frac{w_1 N(x_1|\mu_1, \sigma^2)}{w_1 N(x_1|\mu_1, \sigma^2) + w_2 N(x_2|\mu_2, \sigma^2)}$$

Next I will calculate the posterior for

$$\begin{aligned} f(\mu_1|x_i, d) &= f(\mu_1) \prod_i w_{d_i} N(x_i|u_{d_i}, \sigma^2) \\ &\propto e^{-\frac{\mu_1}{100}} e^{-\frac{\lambda}{2} \sum_{d_i=1} (x_i - \mu_1)^2} \propto e^{-\frac{\mu_1}{100} - \frac{\lambda}{2} \sum_{d_i=1} (x_i - \mu_1)^2} \\ &e^{-\mu_1^2 (\frac{1}{100} + \frac{\sum_{d_i=1} 1}{2}) + \lambda \mu_1 \sum_{d_i=1} x_i - \frac{\lambda}{2} \sum_{d_i=1} x_i^2} \\ &\propto e^{-\frac{1}{100} + \frac{n_1 \lambda}{2}} (\mu_1^2 - \frac{\lambda \mu_1 \sum_{d_i=1} x_i}{\frac{1}{100} + \frac{n_1 \lambda}{2}} + \frac{\frac{\lambda}{2} \sum_{d_i=1} x_i^2}{(\frac{1}{100} + \frac{n_1 \lambda}{2})^2}) \\ &\propto e^{-\frac{1}{100} + \frac{n_1 \lambda}{2}} (\mu_1 - \frac{\frac{\lambda}{2} \sum_{d_i=1} x_i}{(\frac{1}{100} + \frac{n_1 \lambda}{2})})^2 \end{aligned}$$

This makes

$$f(\mu_1|x_i, d) \sim N(\frac{\frac{\lambda}{2} \sum_{d_i=1} x_i}{(\frac{1}{100} + \frac{n_1 \lambda}{2})}, \frac{1}{(\frac{1}{100} + \frac{n_1 \lambda}{2})})$$

By symmetry, we know that the posterior for μ_2 is

$$f(\mu_2|x_i, d) \sim N(\frac{\frac{\lambda}{2} \sum_{d_i=2} x_i}{(\frac{1}{100} + \frac{n_2 \lambda}{2})}, \frac{1}{(\frac{1}{100} + \frac{n_2 \lambda}{2})})$$

The last posterior I need to calculate is

$$\begin{aligned} f(\lambda|x_i, d) &= f(\lambda) \prod_i w_{d_i} N(x_i|u_{d_i}, \sigma^2) \propto e^{-\lambda} \prod_i \lambda^{\frac{1}{2}} e^{-\frac{\lambda}{2} (x_i - \mu_{d_i})^2} \\ &\propto \lambda^{\frac{n}{2}} e^{-\lambda} e^{-\frac{\lambda}{2} \sum_i (x_i - \mu_{d_i})^2} \end{aligned}$$

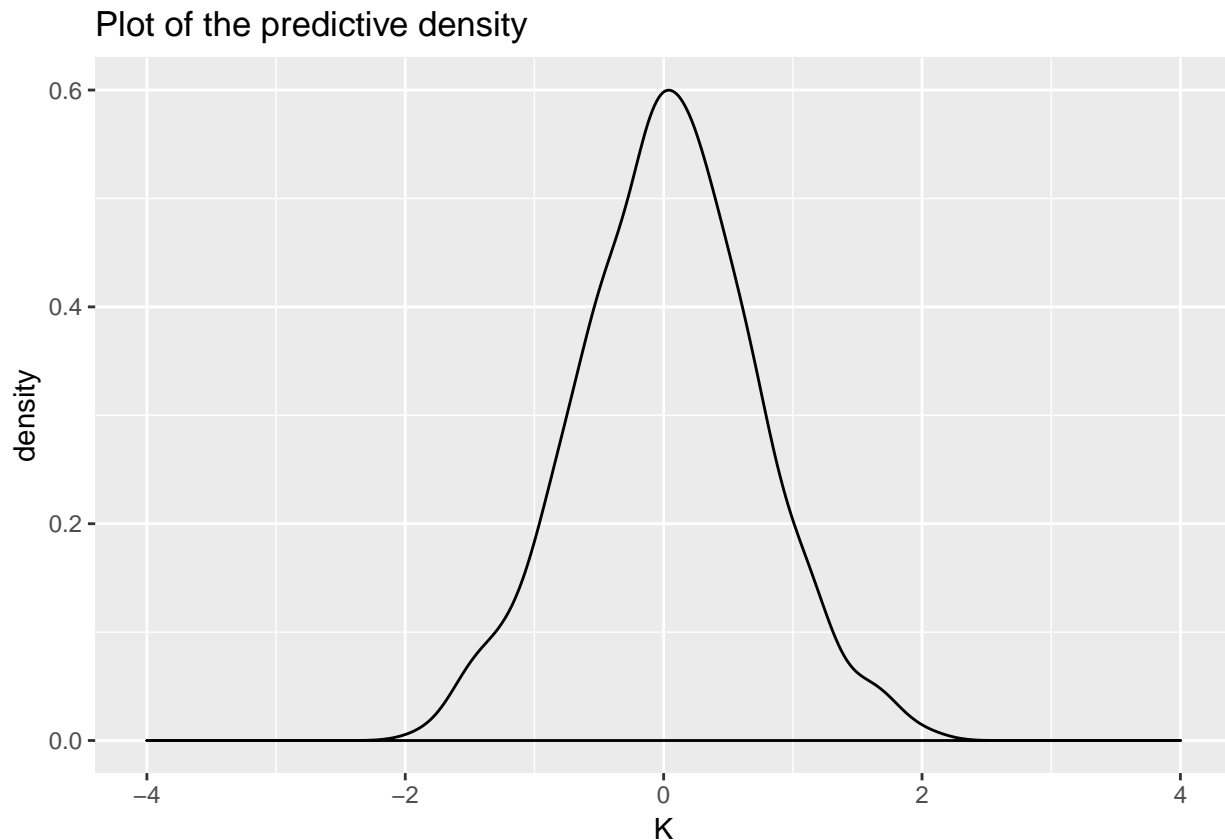
This makes

$$f(\lambda|x_i, d) \sim Ga(1 + \frac{n}{2}, 1 + \frac{1}{2} \sum_i (x_i - \mu_{d_i})^2)$$

I generated the data x_i as 100 randomly sampled observations from a standard normal distribution. I will now use a Gibbs Framework to sample from the posterior distributions. I incorporated the latent variable, d , which determines which group each observation came from, for each iteration in the Gibbs framework. The d_i 's determine how many of the 100 observations come from group 1, n_1 or group 2, n_2 . This number of observations from each group were then used to take a sample from the posteriors of the parameters. I ran 11,000 Gibbs iterations, with a 1,000 observation burn in (my computer couldn't reasonably handle more iterations).

2. Plot the predictive density; this is done by taking the average of the densities you get at each iteration of the Markov chain. Does it look like it is a standard normal density?

I averaged the posterior densities over each iteration of my Markov chain and used those converged estimates to generate the predictive density, which is plotted below. The density looks very much like a standard normal density to me, centered around 0, with a variance of approx 1. Obviously with many more Monte Carlo iterations, the predictive density will look increasingly more like a standard normal density.

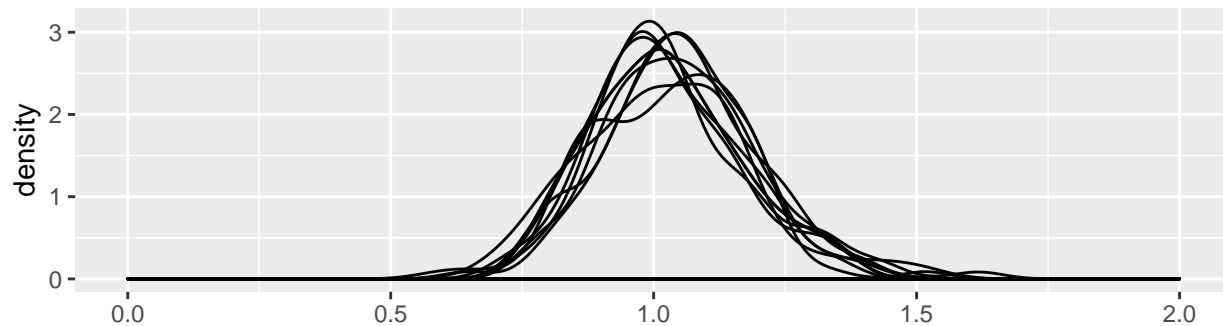


3. Plot some of the posterior distributions for μ_1 and μ_2 and σ and w . Is there any indication from these that the true density is standard normal?

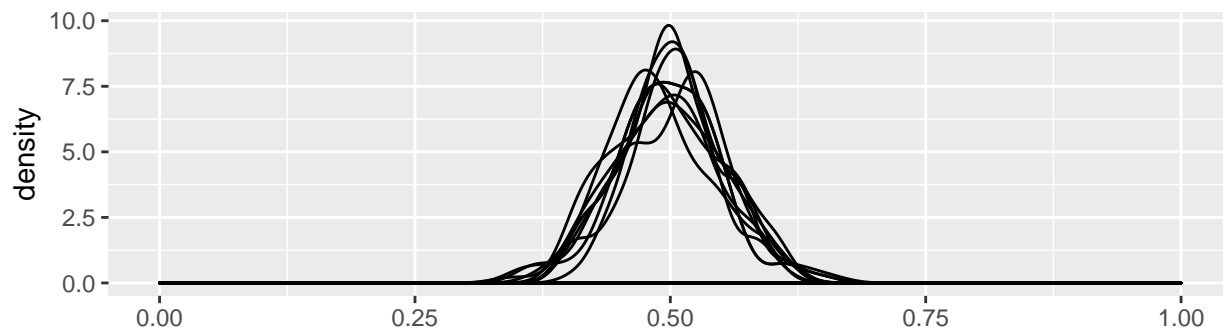
The next plots are the posterior distributions for the four parameters. Each parameter appears to be distributed normally, but not necessarily standard normal, as they are not centered over 0 with a variance of approx 1.

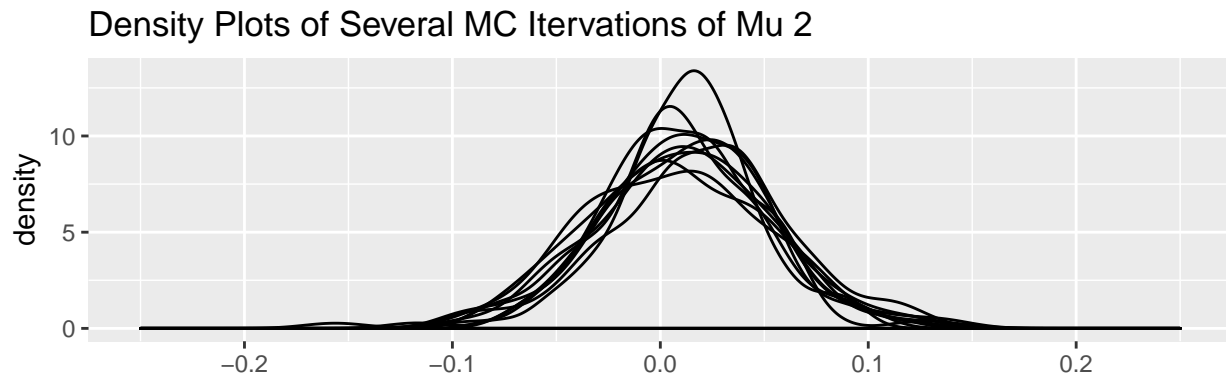
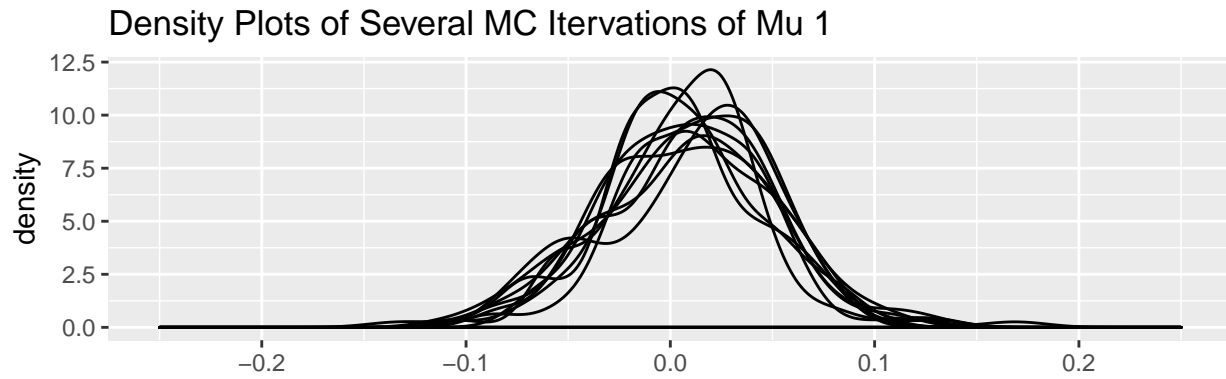
The plots below are density plots of 10 individual MC estimates for λ, w, μ_1 , and μ_2 , respectively. As we can see, they all seem to follow a similar shape, just centered around different means.

Density Plots of Several MC Iterations of Lambda



Density Plots of Several MC Iterations of Lambda



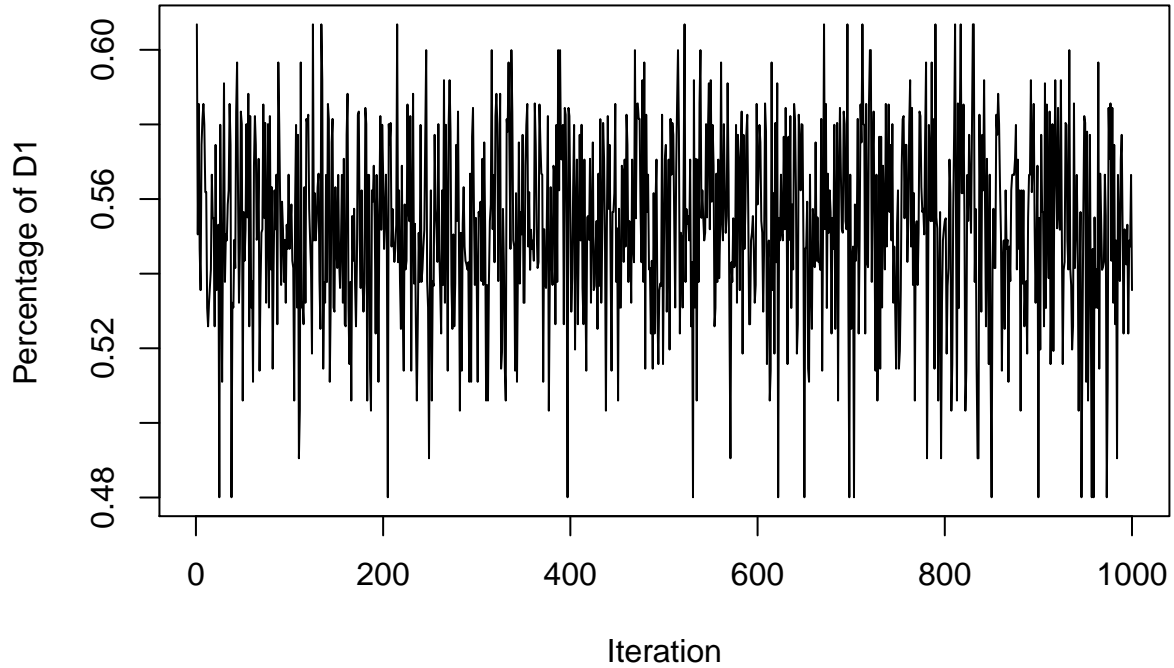


4. Now consider the (d_i) from the output of your Markov chain. Is there any indication that these show there is a single group?

The plot below is a trace plot for $d_i = 1$. I only included the plot for d_1 since, if the observation does not come from $d_i = 1$, it clearly will come from $d_i = 2$. The plot below reveals that there is variation in the probability of a particular observation coming from a certain group, thus number of observations coming from a particular group also varies.

```
## Using parallel package.
## * Set seed with set.seed().
## * Disable this message with options(`mosaic:parallelMessage` = FALSE)
```

Probability that an observation comes from d=1



5. Evaluate the integral

$$I = \int x g_p(x) dx$$

#where g_p is the predictive density you got from part 2.

Here we know that the integral cannot be solved analytically, or using standard calculus, this is why I implement a Monte Carlo Markov Chain. By the definition of a Monte Carlo integral, the estimated I will equal

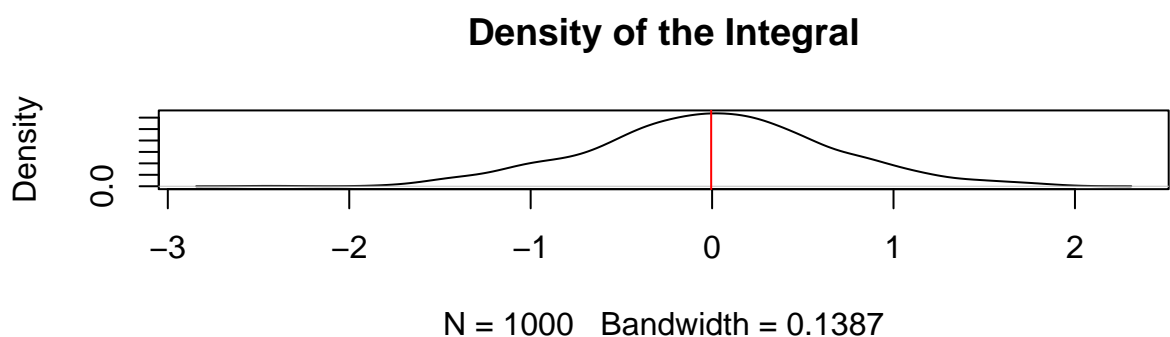
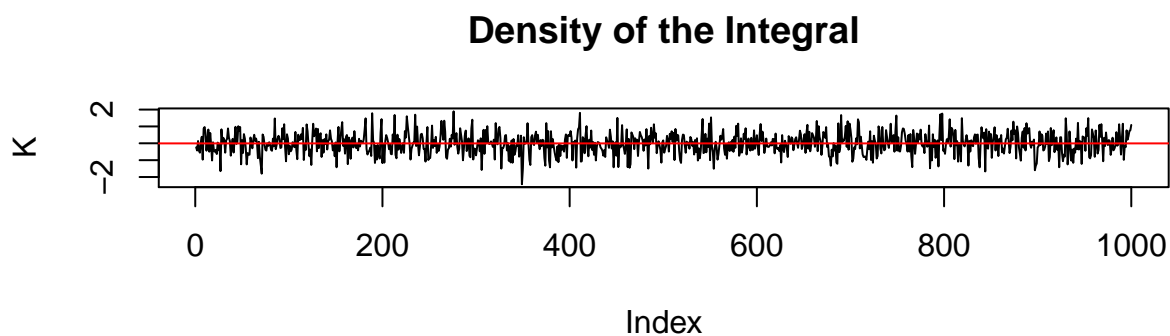
$$\hat{I} = \frac{1}{N} \sum_{i=1}^N x_i$$

$g_p(x)$ is the predictive density from part 2, which is estimated using

$$k(\hat{x}|\theta) = \hat{w}N(x|\hat{\mu}_1, \hat{\sigma}^2) + (1 - \hat{w})N(x|\hat{\mu}_2, \hat{\sigma}^2)$$

The Monte Carlo estimates of the parameters are used to make the Monte Carlo estimates of the predictive density for x.

I implemented the code and got and evaluated the integral to be 0.01275.



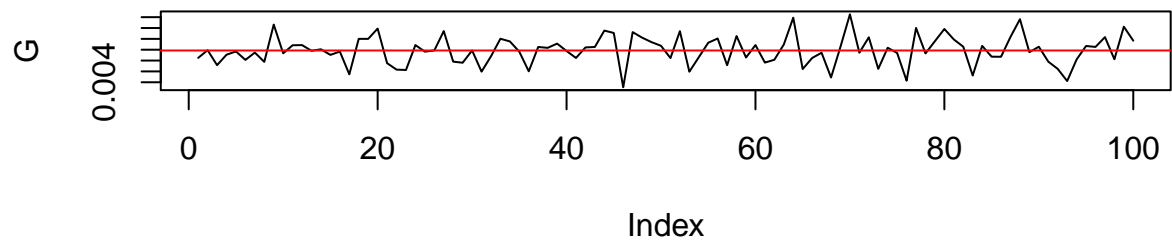
6. Estimate

$$w\mu_1 + (1 - w)\mu_2$$

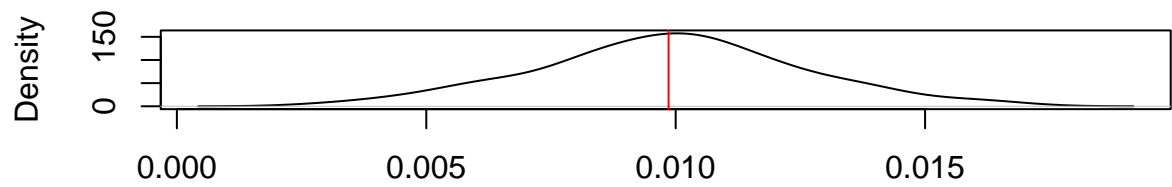
#from the output of the Markov chain. How does this compare with your answer for 5?

I estimated the equation by using the Monte Carlo estimates for w, μ_1 and μ_2 and got a value of 0.009913. This is incredibly close to the estimated value of the integral above, which both round to 0.01.

Plot of the Estimated Equation



Density of the Estimated Equation



N = 100 Bandwidth = 0.0008845