Bayesian Statistics Large Assignment 1 - Report

Elena Kreis s4841670 Mihaela Neacsu s1002844

April 16, 2019

1 So... does it actually work?

1.1 Recovering the correct model

1.1.1

The solutions for this part can be found in the file `doesitwork_template.R`.

1.1.2

The solutions for this part can be found in the file `doesitwork_template.R`. For $\hat{\mu}$ and $\hat{\sigma}$ we chose 100 and 5 respectively since they are relatively far away from the true values, as was advised.

1.1.3

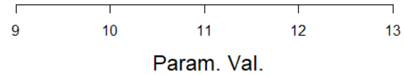


Figure 1: Posterior distribution for N=10

N = 10:
HDI = 9.79-12.3

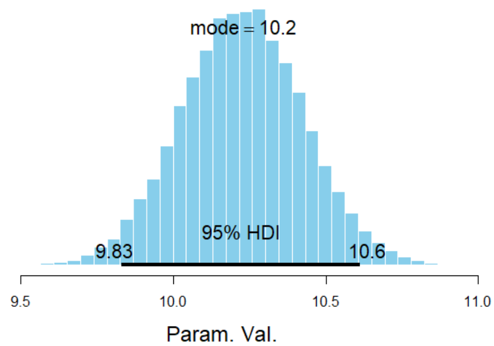


Figure 2: Posterior distribution for N=100

N = 100:
HDI = 9.83-10.6

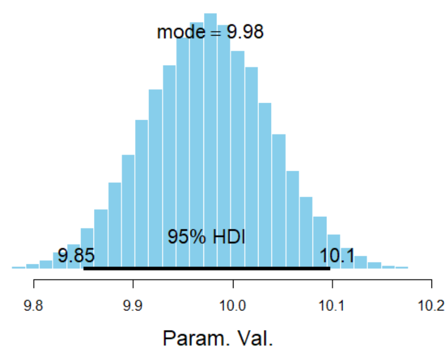


Figure 3: Posterior distribution for N=1000

N = 1000:
HDI = 9.85-10.1

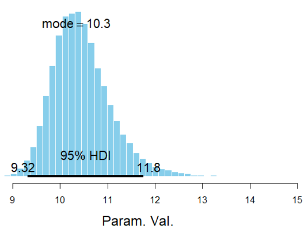
With a higher number of samples N the HDI range gets smaller, which indicates that the confidence in the accuracy of the model increases, which is intuitive since more samples mean more evidence.

1.1.4

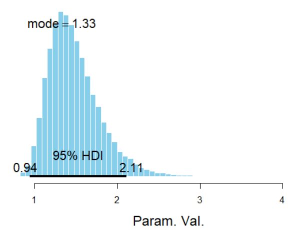
In this exercise the graphs were obtained by alternating the monitored parameter in

```
samples = coda.samples(jagsmodel1,
                       c(''),
                       n.iter=mcmciterations)
```

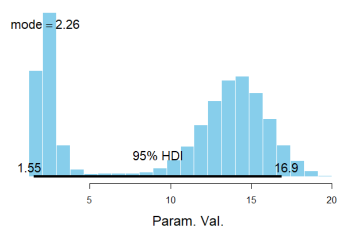
N=10



(a) Posterior distribution for μ



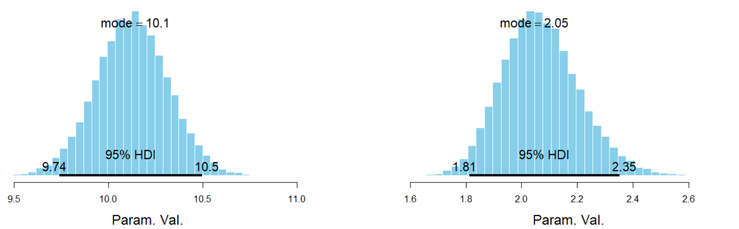
(b) Posterior distribution for σ



(c) Posterior distribution for σ

Figure 4: N=10

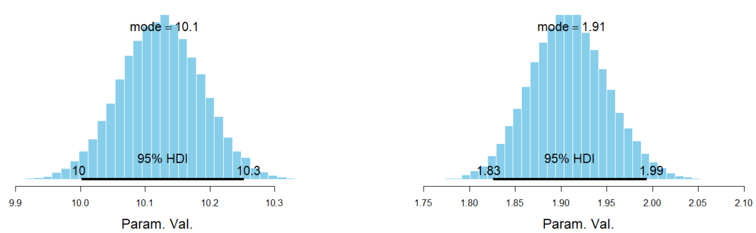
N=100



(a) Posterior distribution for μ (b) Posterior distribution for σ

Figure 5: N=100

N=1000



(a) Posterior distribution for μ (b) Posterior distribution for σ

Figure 6: N=1000

The model is still able to recover μ fairly well. The mode of each posterior distribution for μ is very close to 10, despite the bad estimate of $\hat{\mu}$.

For σ this looks slightly different. For N=100 and N=1000 it is recovered pretty well, both times having a mode close to 2. For N=10 the recovering differs strongly between runs. Some runs are distributed somewhat around 2, while others show a bimodal distribution with an extremely wide HDI.

1.1.5

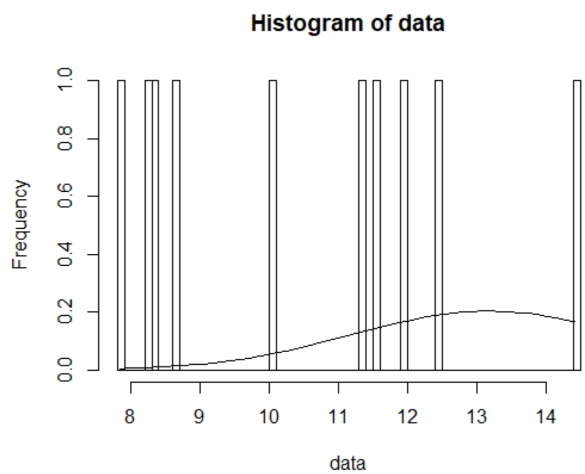


Figure 7: Data vs. Gaussian distribution with $\mu = 13.18419$, $\sigma = 1.959472$, for N = 10

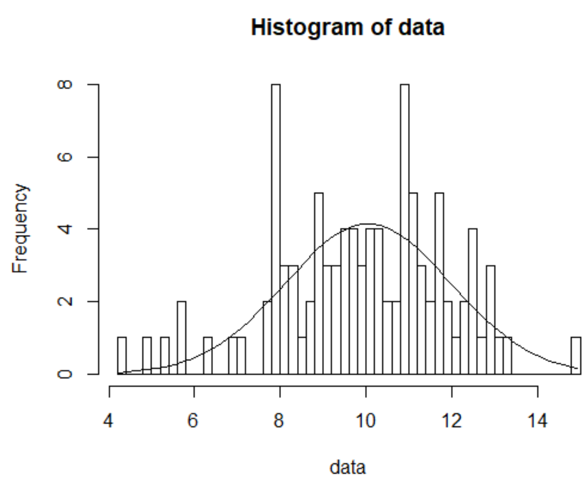


Figure 8: Data vs. Gaussian distribution with $\mu = 10.05071$, $\sigma = 1.924041$, for N = 100

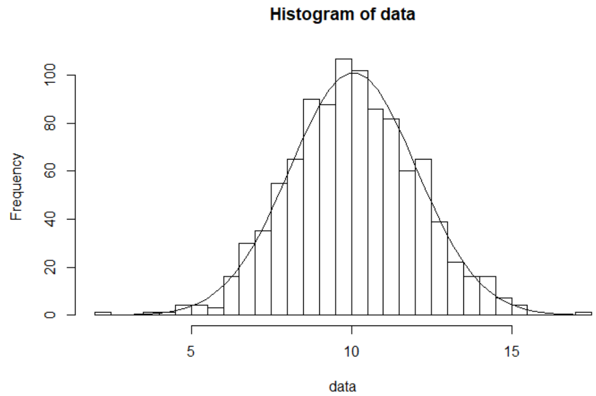


Figure 9: Data vs. Gaussian distribution with $\mu = 10.0556$, $\sigma = 1.971928$, for N = 1000

It is clear that N has a strong influence on how well the recovered distribution matches the data.
For N=10 there does not seem to be a visible fit, but there is also no visible distribution among the data.
For N=100 the data starts showing a distribution much more clearly and the recovered Gaussian distribution seems to match it fairly well, although not ideally.
For N=1000 we can see an almost perfect match between data and distribution.

1.2 Recovery with the wrong model

1.2.1

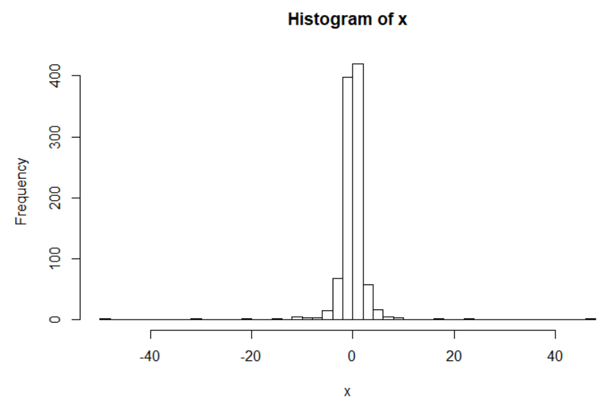


Figure 10: Student-t distribution with df=2

We can see that the distribution is very peaked around 0 but there are still outliers, visible as very small bars, far away from the center of the distribution, up to around -40.

1.2.2

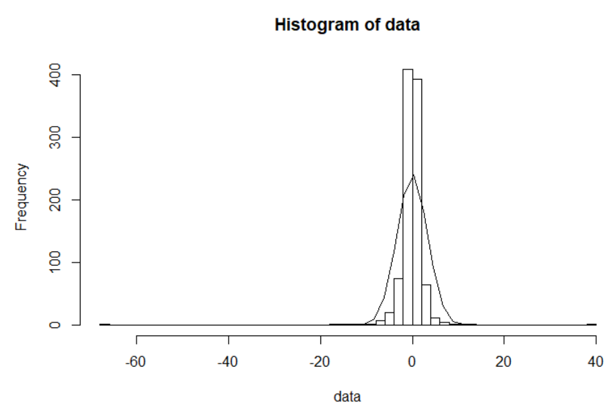


Figure 11: Gaussian distribution with μ = -0.1124487, σ = 3.304665, fitted to student-t distribution with df=2

We can see that the Gaussian distribution does not fit the data so well. It seems that the parameters that are learned are trying to compensate for the outliers and loose some of the peak of the distribution because of that. Also the furthest outliers are actually not accounted for in the Gaussian distribution, since it doesn't 'reach' that far.
We can see that assuming the wrong model to recover a distribution can have fairly negative effects.

2 Modeling and parameter learning for exam results

2.1 To study or not to study

2.1.1

Generative model:

$$
\begin{gathered}
\omega = 0.5 \\
\psi = 0.5 \\
n = 40 \\
z_i|\omega \sim Bernoulli(\omega), \quad i = 1...p \\
\phi|a, b \sim Beta(a, b) \\
\theta_h|z_i, \phi, \psi = (1-h) \cdot \psi + h \cdot \phi, \text{ where } h = \{0, 1\}, \; i = 1...p \\
k_i|n, \theta_h \sim Binom(n, \theta_{z_i}), \quad i = 1...p
\end{gathered}
$$

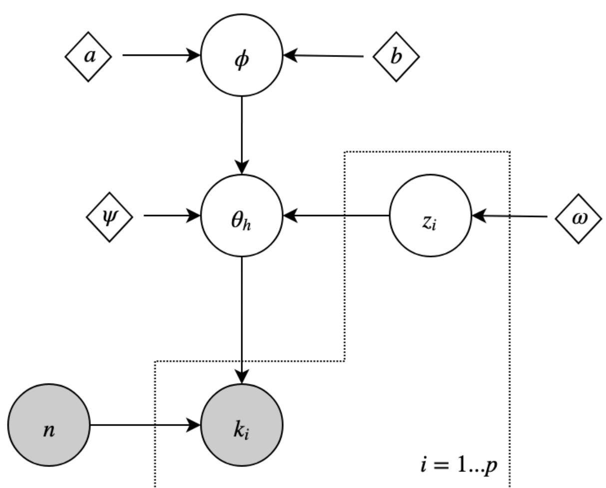


Figure 12: Graphical model

θ here is used as a link function to express that there are two different probabilities used, depending on the label z_i. If $z_i = 0$ then $\theta = \psi$ and if $z_i = 1$ then $\theta = \phi$.

2.1.2

First, we need to draw a label z_i for each $i \in \{1...p\}$ from a Bernoulli distribution, which has ω as a given probability. Second, we need to draw ϕ from a Beta distribution. Next, using ϕ and ψ we can calculate the two different θ, one for the studying students and one for the non-studying students, which we can pass as a parameter for the Binomial distribution, in order to obtain a k_i for each $i \in \{1...p\}$.

2.1.3

The implementation can be found in the file `examresults_template.R`.

2.1.4

We plot the convergence plots using a different function than the one in the slides, namely:

```
plot(model1samples)
dev.off()
```

This yields the following result:

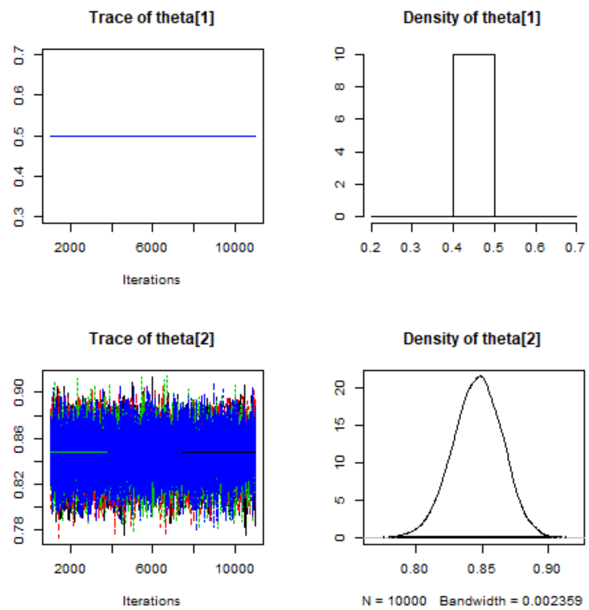


Figure 13

The results show convergence since the lines are not distinguishable. We have let the sampler run for 10000 iterations, since this number has proven suitable in previous examples and this seems to be sufficient, as the sampler has converged. We might be able to achieve convergence with a lower number of iterations, since the plots seem to show that the sampler already converged around 2000 iterations.
You can note that for `theta[1]`, which is the success probability of the guessing group, there is no burn-in, since it is a set value.

2.1.5

By calling the `summary` function, we can see that the mean of the distribution for the probability of a correct answer for the students that studies is 0.8467, while the SD is 0.01858.

2.1.6

Again using `summary` we see that the students 1 to 5 get group label 0 (guessing) while students 6 to 15 get label 1 (studying). This seems to be a sensible choice since, looking at `k`, the first 5 students have a low number of correct answers, all mostly around 20, while the latter 10 students have much higher numbers, starting at 29.

2.1.7

To learn the probability of a student having studied we need to modify the definition of omega in the model. Instead of being a set value it would then follow a beta distribution.

$$\omega|a_2, b_2 \sim Beta(a_2, b_2)$$

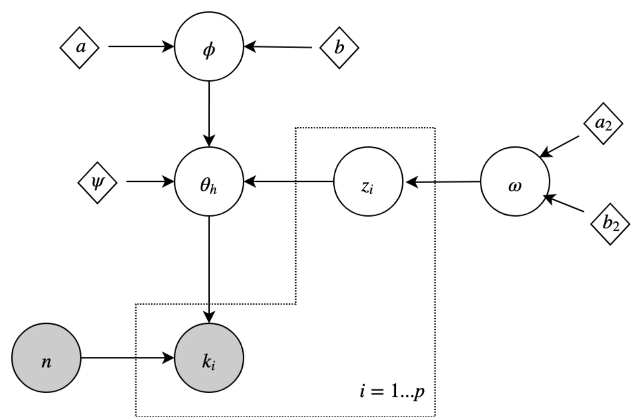


Figure 14: Graphical model

2.2 Individual variation using a hierarchical model

2.2.1

Both the parameters ϕ and θ_1 (theta for group 1: studying) capture the assumption that all studying students share the same success probability since they are both set values.

2.2.2

To capture individual variability, each student should be assigned a 'personal' ϕ_i, drawn from a beta distribution, i.e.:

$$\phi_i|\mu, \kappa \sim Beta(\mu \cdot \kappa, (1-\mu) \cdot \kappa), \quad i = 1...p$$

And then θ would change to:

$$\theta_i|z_i, \phi_i, \psi = (1-z_i) \cdot \psi + z_i \cdot \phi_i, \quad i = 1...p$$

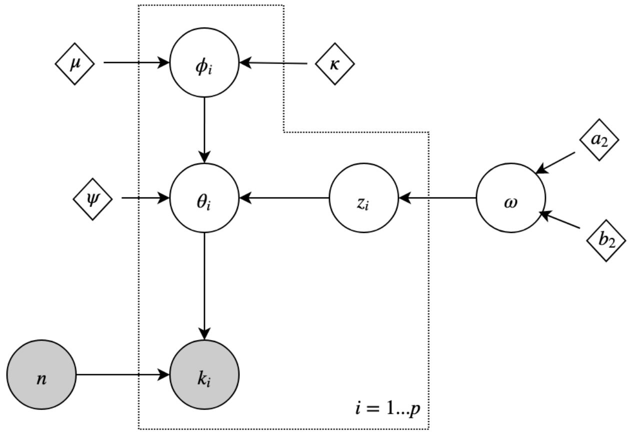


Figure 15: Graphical model

Note that instead of a and b as before, we now use μ and κ for the beta distribution of ϕ, since it is easier to give a good prior when thinking in terms of mean, as was advised in the assignment.

2.2.3

For students 6-15, who have studied, their respective probabilities are:

6. 0.745
7. 0.945
8. 0.705
9. 0.851
10. 0.874
11. 0.874
12. 0.806
13. 0.921
14. 0.898
15. 0.827

2.3 Easy and tough exam questions

2.3.1

For a beta distribution, a flat prior means a = b = 1.
Generative model:

$$p_i|a_1, b_1 \sim Beta(1,1), \quad i = 1...n$$

$$q_j|a_2, b_2 \sim Beta(1,1), \quad j = 1...m$$

$$\theta_{ij}|p_i, q_j = p_i \cdot q_j, \quad i = 1...n, j = 1...m$$

$$k_{ij}|\theta_{ij} \sim Bern(\theta_{ij}), \quad i = 1...n, j = 1...m$$

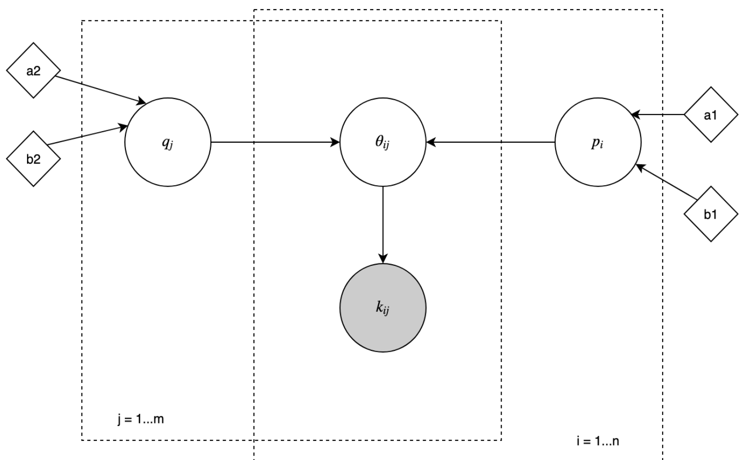


Figure 16: Graphical Model

2.3.2

Using the `summary` function on the samples and looking at p we see the different probabilities with which each student answers questions correctly, which represents how well they studied.
The three best students are Student 1 (0.89592), Student 5 (0.84933) and Student 6 (0.82227), while the three worst-performing students are Student 8 (0.08536), Student 7 (0.17549) and Student 2 (0.28208).

2.3.3

Now we monitor k instead of p, and get the individual predictions of student i answering question j.
For $k_{1,13}$ the mean is 0.7252 which makes the predicted outcome 1.
For $k_{8,5}$ the mean is 0.0140, making the expected outcome 0.
For $k_{10,18}$ the mean is 0.4062, which notably is a less 'clear' decision than the other two and makes the expected outcome 0.

To express that the questions get harder, we draw q from a beta distribution with $b > a$. We chose to change the values to a=1, b=5.
For $k_{1,13}$ the mean is now 0.45200, making the expected outcome 0.
For $k_{8,5}$ the mean is 0.01425, making the expected outcome 0.
For $k_{10,18}$ the mean is 0.30550, making the expected outcome 0.

We can see that the means got much lower, resulting in all 3 questions being predicted as a wrong answer. This seems sensible since whether a question is answered correctly is dependant on both the student studying and how hard the question is. If questions become harder then it also is less likely that a students answers them incorrectly.
$k_{1,13}$ shows the biggest decrease, since it contradicts the prior belief that the questions are very difficult and thus hard to answer. $k_{8,5}$ supports this belief and therefore does not suffer a lot of change.

2.4 Inferring a difference

2.4.1

$$\theta_i | a_i, b_i \sim Beta(a_i, b_i), \quad i \in \{1, 2\}$$
$$k_i | \theta_i, n_i \sim Binom(\theta_i, n_i), \quad i \in \{1, 2\}$$
$$\delta = |\theta_1 - \theta_2|$$

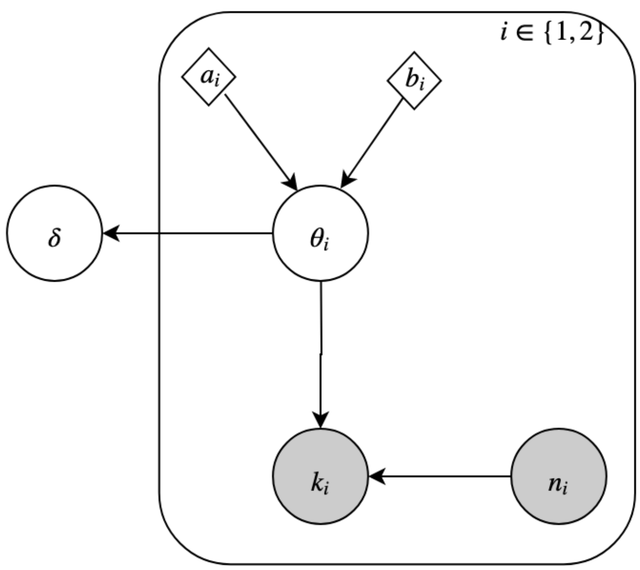


Figure 17: Graphical model

2.4.2

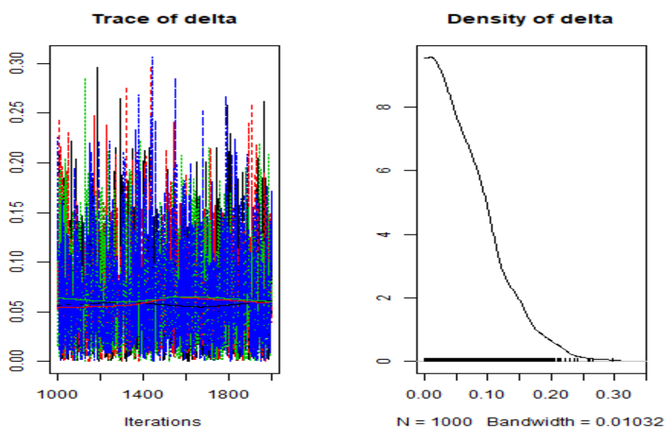


Figure 18: Distribution of δ when $n_2 = 63$

From the posterior distribution in Figure 18 we see that delta is distributed very close to 0, and the mean is only slightly above, at 0.0666622 (obtained with `summary` function). This leads to the conclusion that there is not really a difference between the groups.

If we change n_2 to 49 we get the following plots:

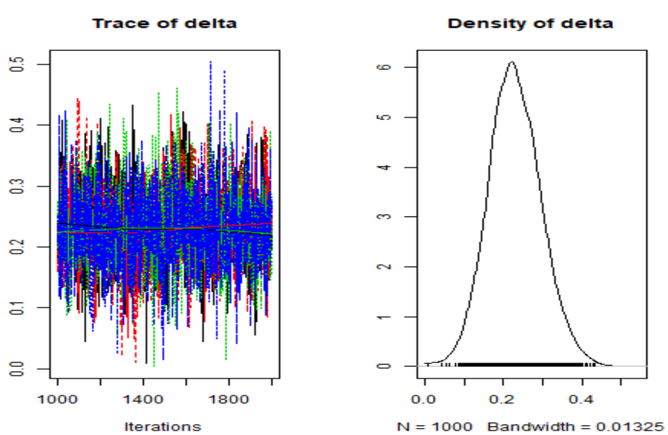


Figure 19: Distribution of δ when $n_2 = 49$

From the posterior distribution in figure 19 we now see a distribution that is centered around a much higher value. The mean here is 0.231938, which means that there now does seem to be a notable difference between the two groups.