Assignment 3

Analyzing hierarchical normal model of rats' growth over time using Bayesian approach

Table of Contents

Introduction

In this report, I will analyse 30 rats' growth whose weight measured weekly for five weeks, starting from age 8 of weeks. The data is taken from section 6 of Gelfand et al (1990; JASA 85: 972-985) and also it is included among the examples of WinBugs). To analyse the data, I will run hierarchical normal model using Bayesian hierarchical modelling.

The aim of this report is to estimate the effect of time over unit growth of rats; to estimate overall variation between rats over the weeks of measurements; and to estimate overall growth rate of rats.

Methods

The code for this report is based NormalGammaHierarchicalModel.R and sourcing the code from DBDA2Eprograms folder from Doing Bayesian Analysis book by John K. Kruschke. All of the analysis of this report is using R software, with extensive usage of coda, rjags, and runjags packages.

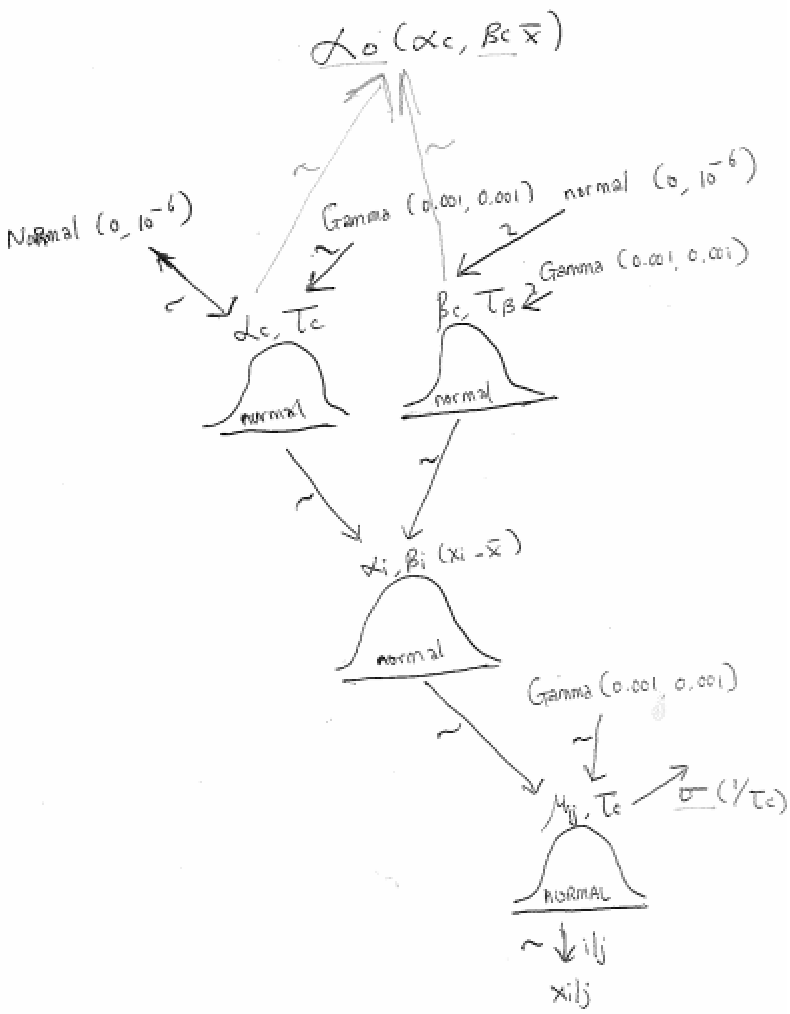
Discussion & Results

All of the rats' growth data type in this report are integer with defined as Yij, as the weight of rat *i* at week *j*. Yij is distributed as Normal distribution with mean μ_{ij} and standard distribution τ_c. And μ_{ij} is specified by this equation $\alpha_i + \beta_i(x_j - \bar{x})$, resulting **regression model on the mean of normal likelihood**.

α_i also distributed as Normal distribution, with mean α_c and standard distribution τ_α, and β_i is distributed as Normal distribution with mean β_c and standard deviation τ_β.

τ_c, τ_α, and τ_β are distributed as Gamma distribution with shape parameter 0.001 and scale parameter 0.001, while α_c and β_c is distributed as Normal distribution with 0 mean and 10^{-6} standard deviation.

This equation can be modelled with graphical representation as illustrated in the graph below:



α_c, τ_α, β_c, τ_β,and τ_c are become the prior and set to the non-informative value: 150, 1, 10, 1, and 1 respectively. The distribution of this value is already mentioned above.

After taking 500 tuning steps followed by 500 updates of burn-in period, first, we will analyse **the effect of time over unit growth of rats (posterior β_c)**, with the data presented below:

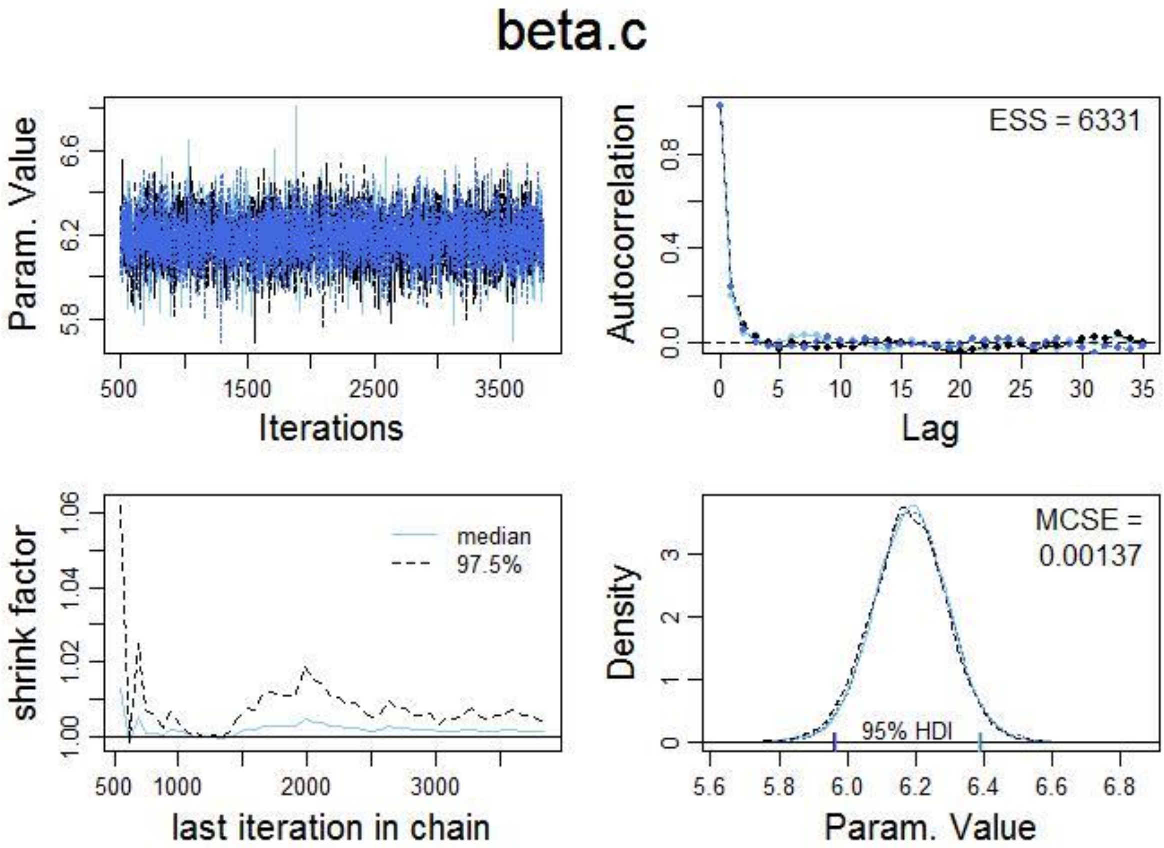
```
            ESS     mean   median     mode hdiMass   hdiLow  hdiHigh compVal pGtCompVal
beta.c 6320.584 6.184551 6.184782 6.169148    0.95 5.968104 6.394493      NA         NA
       ROPElow ROPEhigh pLtROPE pInROPE pGtROPE
beta.c      NA       NA      NA      NA      NA
```

The result show that the posterior mean of β_c is 6.12, change slightly different from the prior β_c with the value of 10. This posterior value is supported with 95% HDI data (5.97:6.40). Because the prior data is non-informative, we can guess that posterior β_c value will be much affected by likelihood. Now we can take a look at value of σ as a part of likelihood described below:

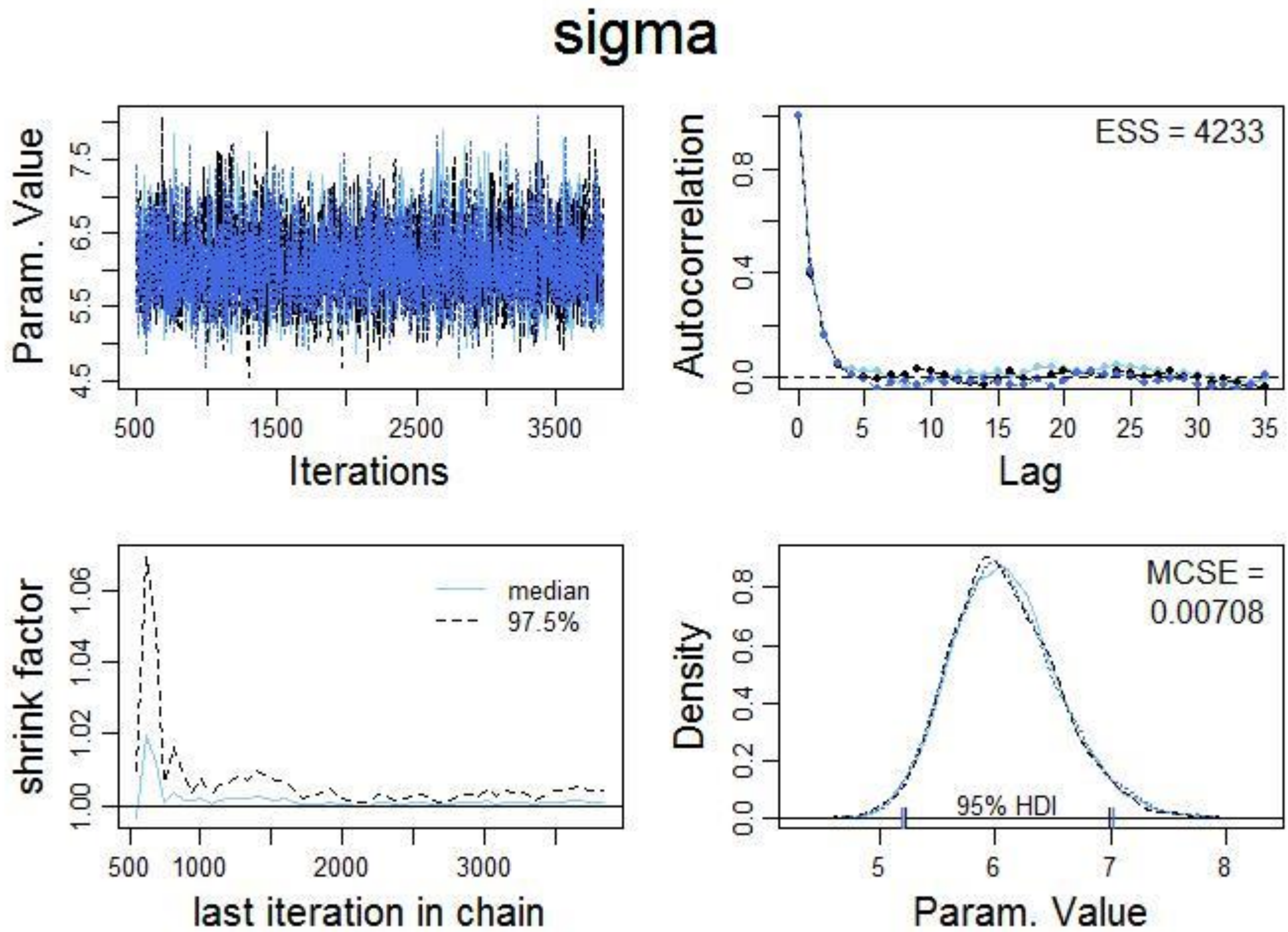
```
           ESS     mean   median     mode hdiMass   hdiLow  hdiHigh compVal pGtCompVal
sigma 4229.526 6.082678 6.049024 5.990339    0.95 5.204465 6.997794      NA         NA
      ROPElow ROPEhigh pLtROPE pInROPE pGtROPE
sigma      NA       NA      NA      NA      NA
>
```

The mean of overall variation between rats over the weeks of measurement (σ) is 6.1, validated by 95% HDI (5.2:6.9). Because the prior cannot affect posterior β_c much, the value of posterior β_c will be much closer to σ than to prior β_c .

To see how representative, effective and efficient of posterior β_c distribution, we can take a look at this MCMC examination graph below. From the iteration graph in the upper-left, we can see that after burn-in period, most of the chain is overlap each other and mix well, means the posterior representativeness from this graph is good. The density plot in the lower-right panel also overlap well. The shrink factor in the lower-left showed us that after burn-in period, the measure quickly went to 1.0. However, we can see that the chain value of 97.5% is more unstable than the chain value of median. It went up to 1.02, before it goes down again to the value near 1.0. However, the value after burn-in period never go greater than 1.1, means overall chains is converged adequately. Last, from autocorrelation in upper-right, we see the autocorrelation remain near 0 for large lags, means the chain has low autocorrelation. The parameter value at successive steps in the chain are providing independent information about posterior β_c, and each successive steps is not redundant with the previous steps.



From the previous explanation, we know that posterior β_c is highly affected by . So now we will also take a look at posterior σ MCMC examination graph provided below.



Not much different with posterior β_c, we can see in the iteration graph in upper-left that after burn-in period most of the chain is also overlap each other and mix well, means the posterior representativeness from this graph is good. The density plot in the lower-right panel also overlap well. The shrink factor in the lower-left showed us that after burn-in period, the measure quickly went to 1.0. However, the 97.5% value in posterior σ look more stable than in posterior β_c, and also it never go greater than 1.1, means overall chains is converged adequately. Last, from autocorrelation in upper-right, we see the autocorrelation remain near 0 for large lags, means the chain has low autocorrelation. The parameter value at successive steps in the chain are providing independent information about posterior σ, and each successive steps is not redundant with the previous steps.

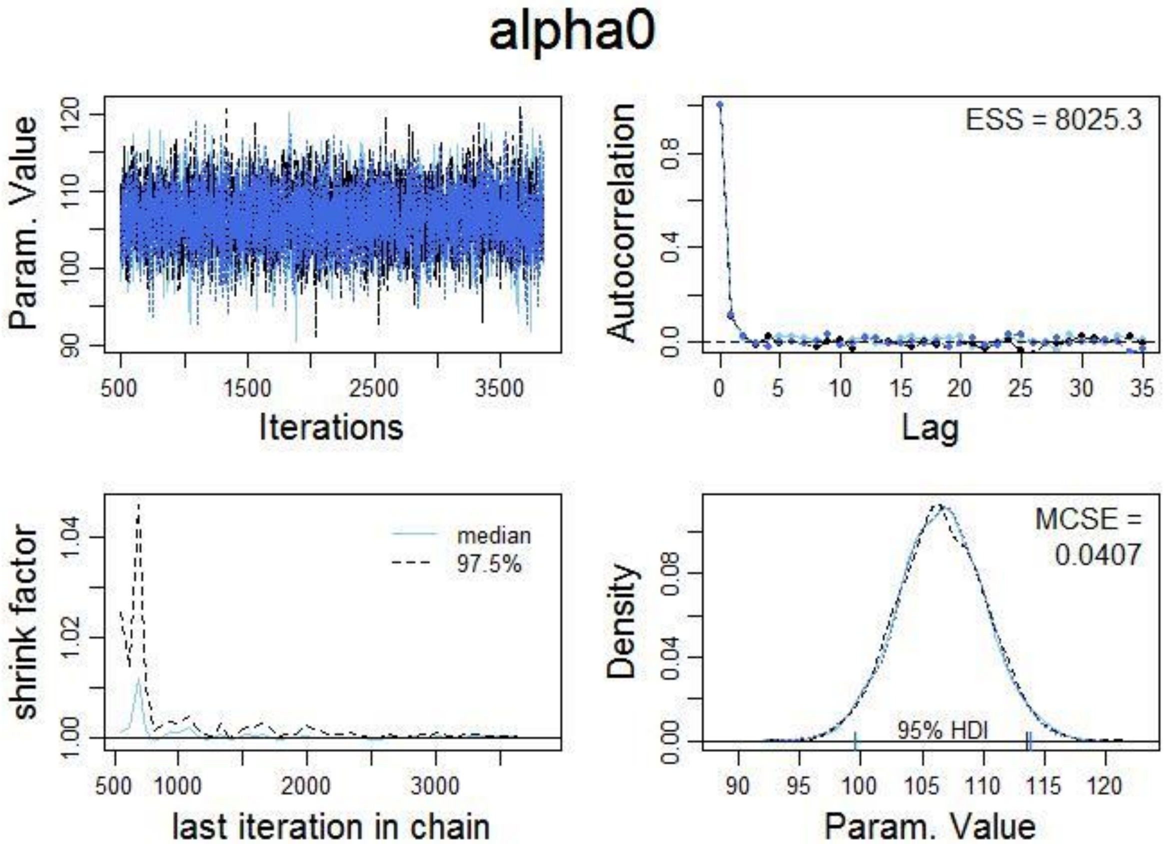
The third part to examine is the overall growth rate of rats, with the result provided below:

```
            ESS     mean    median     mode hdiMass   hdiLow  hdiHigh compVal pGtCompVal
alpha0 8029.709 106.6049 106.5903 106.6163    0.95 99.61337 113.8886      NA         NA
       ROPElow ROPEhigh pLtROPE pInROPE pGtROPE
alpha0      NA       NA      NA      NA      NA
```

The mean of α_0 is 106.6, validated by 95% HDI (99.61:113.88). From the graph, we can see that the result of α_0 is a mixture of α_c, β_c, and $\bar{x}$, explained with this equation: $\alpha_0 = \alpha_c + \beta_c \bar{x}$. In the prior, the value of α_c is 150. However, posterior value of overall growth rate of rats (α_0) is changed to 106.6.

The big change in the value can be explained by the likelihood equation: $\alpha_i + \beta_i(x_j - \bar{x})$. Here, to reduce dependence between α_i and β_i, we standardise the $x_j$around $\bar{x}$. From the next hierarchy in the graph, we know that α_i will affect α_c, and β_i will affect β_c. Therefore, we know that the dependence between α_c and β_c also reduced, and it affect the value of α_0 that change a little bit far from α_c value.

Again, to see how representative, effective, and efficient of posterior α_0 distribution, we can take a look at this MCMC examination graph below.



We can see that overall, the MCMC representativeness graph in posterior α_0 is giving the good posterior representation. However, the iteration graph in posterior $\alpha_0$showed the best overlap and mix really well than in posterior β_c and σ. The shrink factor in lower-left also fully converge in 1.0 point, better than in posterior β_c and σ. The same thing also can be concluded for the autocorrelation. Although the autocorrelation in posterior β_c and σ remain near 0 for large lags, it still went up and down, and we can see more stable graph in posterior α_0. The density plot in the lower-right panel also overlap well.

Conclusion

To conclude, this report is setting the α_c, τ_α, β_c, τ_β,and τ_c as the prior with non-informative value: 150, 1, 10, 1, and 1 respectively. Because this non-informative priors, the posterior values is following the likelihood. Because the likelihood is affected by τ_c, and τ_c will determine the σ, then the posterior value of β_c (6.18) is following (near) the values of posterior σ (6.1) and not following the prior β_c values (10).

The same things also happen in posterior α_0. The prior α_c (with value 150), should be dependent with β_c. However, with the likelihood equation ($\alpha_i + \beta_i(x_j - \bar{x})$), the dependence between α_c and β_c is reduced. Therefore, value of α_0 (with the equation $\alpha_0 = \alpha_c + \beta_c\bar{x}$), will be affected by likelihood (the value now is 106.6), and without strong prior value, its value will be different than α_c value.

To prove how valid the posterior value, I conduct iterations, shrink factor, density plot, and autocorrelation analysis, with result that all of the posterior result from this MCMC method is well representative, effective, and efficient. However, if we can rank, posterior α_0 has the best representativeness, followed by σ and β_c.