Bayesian Analysis on Gravity Data

Author – Sandip Dutta

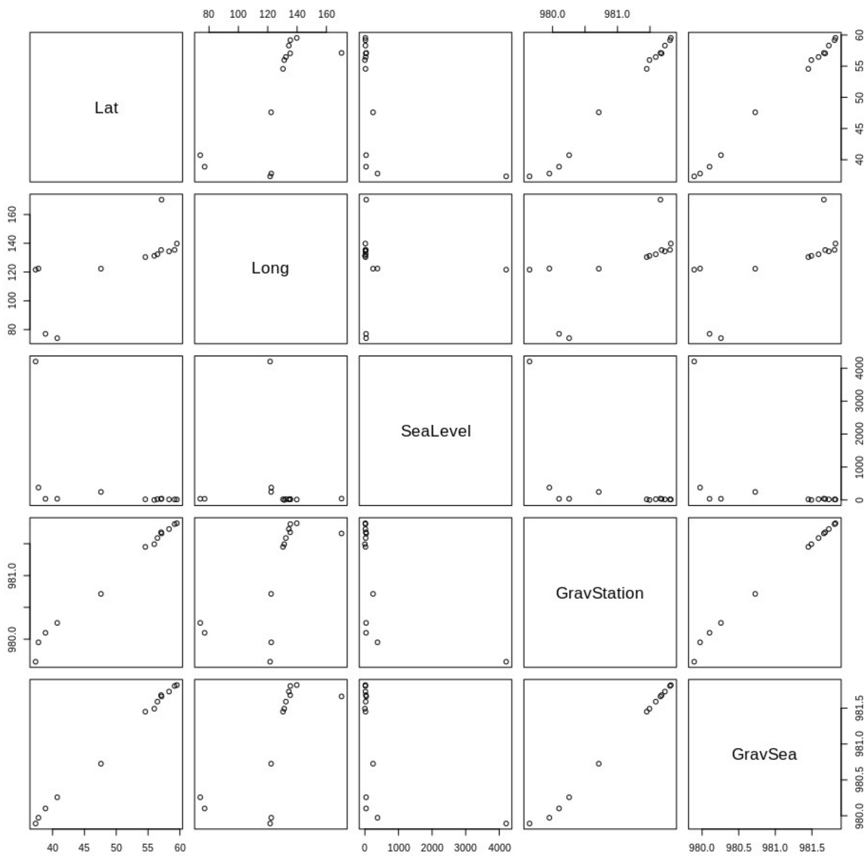
SUMMARY

In this report, we used gravity data, which lists data from multiple spots in the United States. We tried to fit a model to the data so that gravity at some other spot can be measured given latitude, longitude and elevation above sea level. This is useful if we wish to deduce ‘g’ at some spot without direct experiments. We compared two models and then used DIC to assess the better model.

The value of ‘g’ is very useful for certain physical experiments. The value of ‘g’ or the gravity constant varies from place to place on earth. In this report, we wish to make a model that takes in various parameters and gives a good estimate of ‘g’.

ABOUT THE DATA

The data is taken from http://users.stat.ufl.edu/~winner/datasets.html. All credits are to the original authors.The data consists of the values of gravitational constnat ‘g’ at 13 places in the US, along with their latitude, longitude and elevation above sea level values.



GravSea and GravStation are the 2 gravity values.

Since GravSea is the value of ‘g’ at the sea level, predicting it will not be aligned with out interests.

So we will make a model that predicts GravStation given values of latitude, longitude, and elevation above sea level (Lat, Long and SeaLevel columns respectively).

The data has no N/A values ie no data points are missing. The data looks good to use for our purpose.

It can be seen that GravStation and Lat have a good correlation. Longitude and SeaLevel are somewhat related to GravStation. One outlier related to SeaLevel might be in the data. We will however keep it and see how well our models handle it.

MODELLING

We will fit 2 models to the data

1. GravSea[i] = b_0 + b[1] * SeaLevel[i] + b[2] * Lat[i] + b[3] * Long[i] with Normal priors for each of the ‘b’ coefficients. An inverse Gamma prior was chosen for the standard deviation.
[Model - 1]

```
mod_string = "model {
     for(i in 1:length(GravStation)){
          GravStation[i] ~ dnorm(theta[i], prec)
          theta[i] = b_0 + b[1]*SeaLevel[i] + b[2]*Lat[i] + b[3]*Long[i]
     }

     b_0 ~ dnorm(0.0, 1.0/5e6)

     for(j in 1:4) {
          b[j] ~ dnorm(0.0, 1.0 / 5e6)
     }

     prec ~ dgamma(1.0, 1.0)
     sig = sqrt(1.0/prec)
}"
```

2. GravSea[i] = b_0 + b[1] * SeaLevel[i] + b[2] * Lat[i] * Long[i] with Normal priors for each of the ‘b’ coefficients. An inverse Gamma prior was chosen for the standard deviation.
[Model - 2]

```
mod_string_2 = "model {
     for(i in 1:length(GravStation)){
          GravStation[i] ~ dnorm(theta[i], prec)
          theta[i] = b_0 + b[1]*SeaLevel[i] + b[2]*Lat[i] * Long[i]
     }

     b_0 ~ dnorm(0.0, 1.0/5e6)

     for(j in 1:3) {
          b[j] ~ dnorm(0.0, 1.0 / 5e6)
     }
     prec ~ dgamma(1.0, 1.0)
     sig = sqrt(1.0/prec)
}"
```

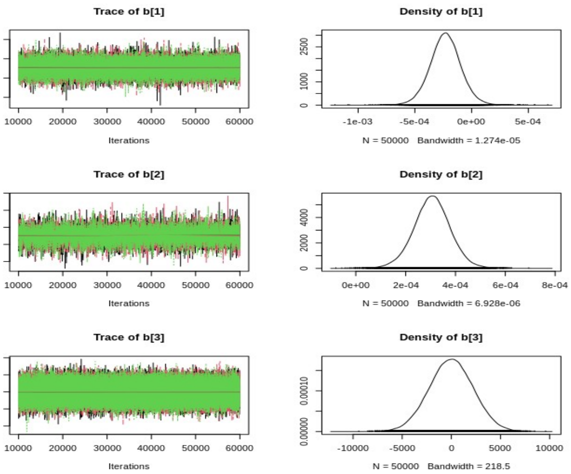
For both the models, Gelman Rubin diagnostics was used to assess convergence, along with trace plots. DIC of both models were calculated and compared to assess the better model

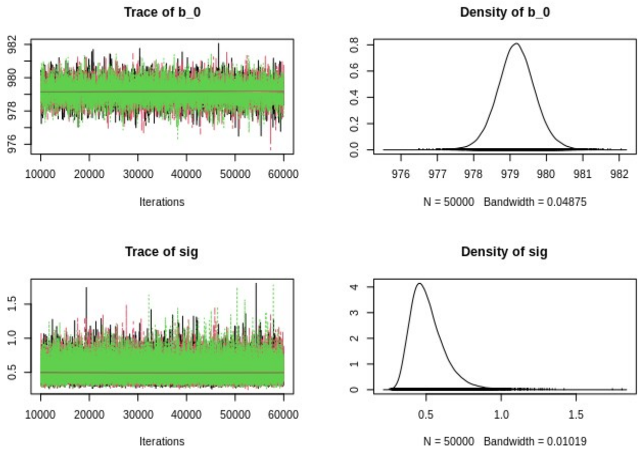
RESULTS

Seed value was set at 0 { set.seed(0) in R}

1. Model – 1

Trace Plots:



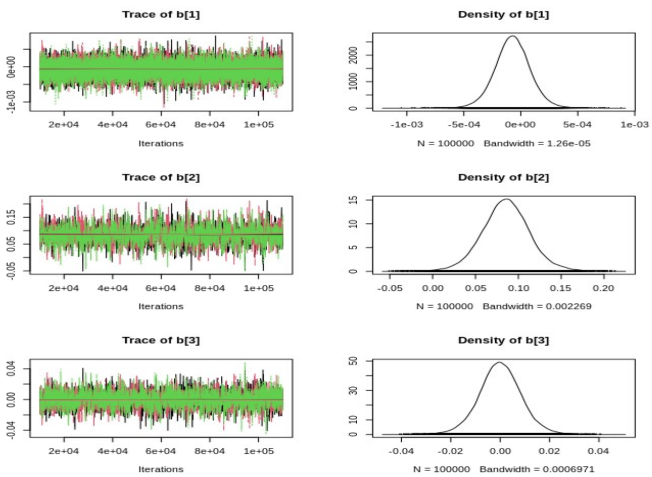


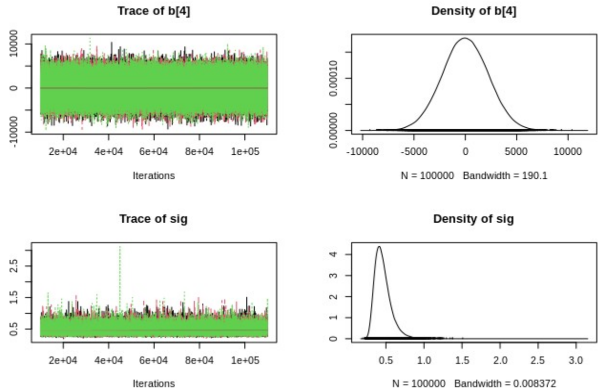
Other important parameter values:

- Burn-in = 1,000 iteraitions
- Number of Chains = 3
- Samples kept = 50,000 for each chain
- Gelman Rubin Diagnostics : 1.01

2. Model – 2:

Trace Plots:





Other important parameter values:

- Burn-in = 1,000 iteraitions
- Number of Chains = 3
- Samples kept = 50,000 for each chain
- Gelman Rubin Diagnostics : 1.00

CONCLUSION

1. DIC for Model 1 – 12.73
2. DIC for Model 2 – 16.73

The DIC values clearly indicate that model 1 is a better performing model than model 2. So our preferred model would be model 1.

Parameters for Model 1

b[1]	b[2]	b[3]	b[4]	b_0	sig
-7.170134e-05	8.574718e-02	-1.469177e-04	9.293978e-04	9.767588e+02	4.624263e-01

The parameter values indicate the following:

1. The intercept or b_0 predicts the mean value for gravity at each point.
2. The other intercepts do small changes to the mean value.

The findings are consistent with our data. If we see, the value of gravity changes by very small amounts between places. So the coefficients do small changes to the baseline (intercept term).

Thus our assumptions are justified and our model is good for use.

This report is part of Coursera Course on Bayesian Statistics: Techniques and Models, organised by University of California, San Diego.

Author:
Sandip Dutta
Email : duttasandip11100@gmail.com