Compare response time of non-schizophrenics and schizophrenics

With Bayesian Hierarchical Model

&

Estimating MCMC Convergence

YI DIAO

Masters Paper

**Faculty Sponsor Approval**:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Faculty Sponsor Signature Date

**Faculty Reader Approval**:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Faculty Reader Signature Date

**Department Chair Approval**:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Department Chair Signature Date

**ABSTRACT**

Recently, there has been dramatic growth in the development and application of Bayesian inference in Statistics. In Bayesian field, Monte Carlo Markov Chain (MCMC) are often used to simulated direct draws from complicated distributions of interest. R is a well-developed simple programming language with a wide range of functions for data manipulation and graphs display. The purpose of this master paper is using Gibbs sampling simulation in R program to generate three posterior distribution of parameters from Bayesian hierarchical modeling based on response time from non-schizophrenics and schizophrenics and predict the probability of non-schizophrenic reacting faster than schizophrenic. Due to lack of information of prior, MCMC samples we simulated maybe or maybe not approximate the true distribution. So there were several statistical diagnostic methods were adopted to estimate the convergence of the Markov chain simulated.

**ACKNOWEDGEMENTS**

First, I cannot express enough thanks to my advisor Dr. Jun Ye’s careful guidance and continued support for this master paper. Then I would like to express sincere gratitude to my reader Dr. Sujay Datta for reviewing this paper and giving suggestions. The knowledge and skills that I have gained from this paper with their help are priceless and very important for my future development. Finally, to my family, especially my husband, I am so grateful to your support and encouragement.

TABLE OF CONTENTS

[1. INTRODUCTION 5](#_Toc418766113)

[2. DATA AND BAYESIAN METHODS 7](#_Toc418766114)

[2.1 Data 7](#_Toc418766115)

[2.2 Bayesian Statistical Method](#_Toc418766116) 10

[2.2.1 Gibbs Sampling 1](#_Toc418766117)1

[2.2.2 Hierarchical Normal Model 1](#_Toc418766117)2

[2.2.3 Analysis and Results 13](#_Toc418766119)

[2.3 Analysis with different Prior Parameters 1](#_Toc418766116)7

[2.4 Summary 21](#_Toc418766115)

[3. ESTIMATING CONVERGENCE OF MCMC 22](#_Toc418766114)

[3.1 Convergence of MCMC](#_Toc418766115) 22

[3.2 Analysis and Results of different Digonosis 2](#_Toc418766116)2

[3.2.1 Trace Plots 23](#_Toc418766119)

[3.2.2 Stationarity Boxplots 24](#_Toc418766118)

3.2.3 Autocorrelation 25

[3.2.4 Effective sample size 26](#_Toc418766119)

[3.2.5 Raftery and Lewis 27](#_Toc418766119)

[3.2.6 Geweke 28](#_Toc418766115)

[3.2.7 Heidelberger and Welch 30](#_Toc418766121)

[4. DISCUSSION](#_Toc418766114) 32

[REFERENCES 33](#_Toc418766129)

[APPENDIX 34](#_Toc418766130)

[R-script: 34](#_Toc418766131)

# INTRODUCTION

Bayesian inference is a method of [statistical inference](https://en.wikipedia.org/wiki/Statistical_inference) based on Bayes’ theorem which shows the relation between two conditional probabilities that are the reverse of each other. The main difference between frequentist (classical) inference and Bayesian inference is, for the former, the unknown parameters are fixed and the probabilities are objective, for the latter, parameters are treated as random variables and probabilities” are subjective which are considered as “degree of belief” (Rathnayake, R.C., 2010). In Bayesian statistics, based on a prior distribution over the unknown parameters formulated according to some beliefs and observing data , a posterior distribution can be obtained which takes account of both the prior and data. The posterior probability of a model is proportional to the prior probability times the likelihood which can be expressed as p(θ|y) ∝ p(θ) x p(y|θ). We don’t always do Bayesian since the calculations needed for Bayesian statistics can be overwhelming and prior distribution is subjective and maybe result in different results (Parker, M., 2005).

The secret behind the increasing popularity of Bayesian analysis lies in the application of

Markov Chain Monte Carlo (MCMC) to compute the posterior probability density. And reasonably efficient MCMC algorithm exist to sample from the posterior distribution for most classes of models (Martin, A.D., Quinn, K.M. & Park, J.H., 2011). R program is used for computation of Bayesian modeling, such as hierarchical modeling, since it contains a huge collection of statistical functions and good summary of probability distributions (Rathnayake, R.C., 2010).

A critical issue for users of MCMC methods in applications is how to determine when it is safe to stop sampling and use the samples to estimate characteristics of the distribution of interest (Cowels, M.K. and Carlin, B.P., 1996). In this paper, several diagnostic methods will be applied to estimate whether the simulated sample size of posterior distribution is large enough to make sure it is a good representation of the true distribution.

# DATA AND BAYESIAN METHODS

## 2.1 Data

The dataset of this paper presented response times (in milliseconds) for 11 non-schizophrenics and 6 schizophrenics (30 measurements for each person) (Belin and Rubin, 1990). The dataset had been used in Figure 18.1 from the book “Bayesian Data Analysis”, third edition by Dr. Gelman. Psychologists at Harvard University performed an experiment measuring thirty reaction times for each of seventeen male subjects: eleven non-schizophrenics and six schizophrenics. Manual reaction times to visual cues, where subjects watch a screen and move their fingers from one button to another when a signal appears, were measured. There are 30 measurements per individual. Fig. 1 and Fig. 2 are histograms from these two groups with normal curves.

Fig.1, Histogram of response time (in milliseconds) for 11 Non-schizophrenic individuals

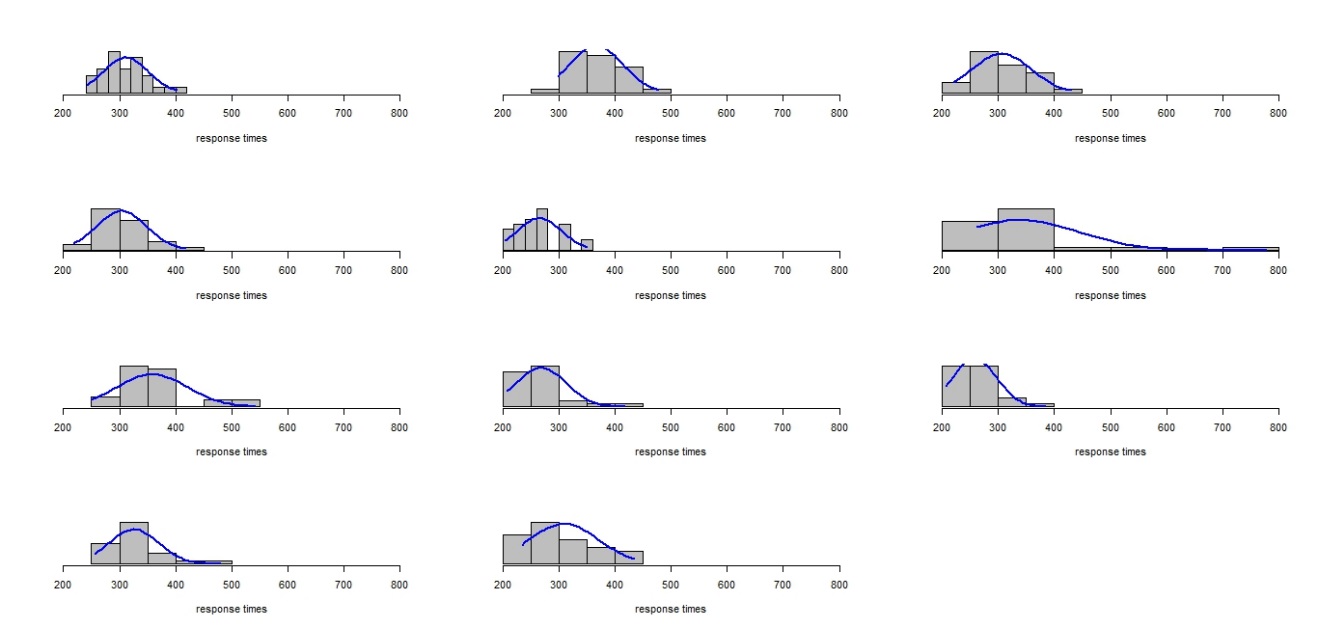
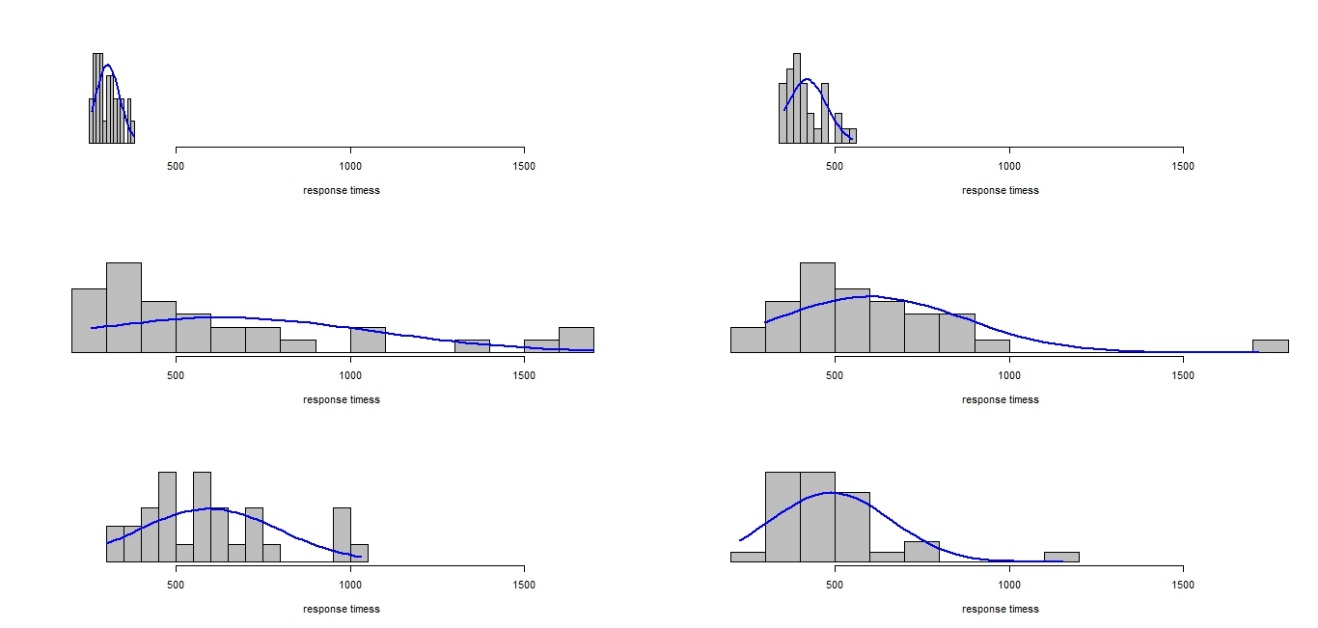


Fig.2, Histogram of response time (in milliseconds) for 6 Non-schizophrenic individuals



In order to obtain the approximate normal distribution for 30 measurements per each person, especially for 6 schizophrenics, logarithms of these data are calculated and made Fig.3 and Fig.4 with normal curves shown below:

Fig.3, Histogram of Y = log (response time in milliseconds)

for 11 Non-schizophrenic individuals

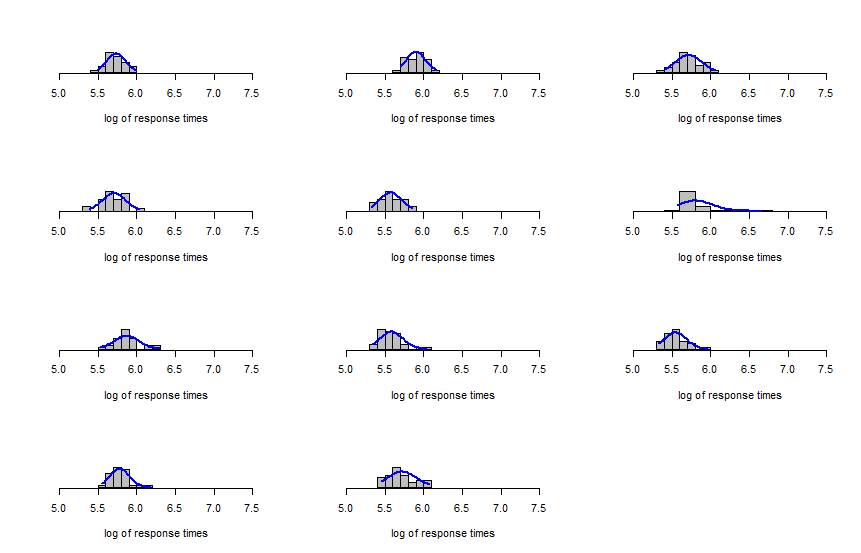
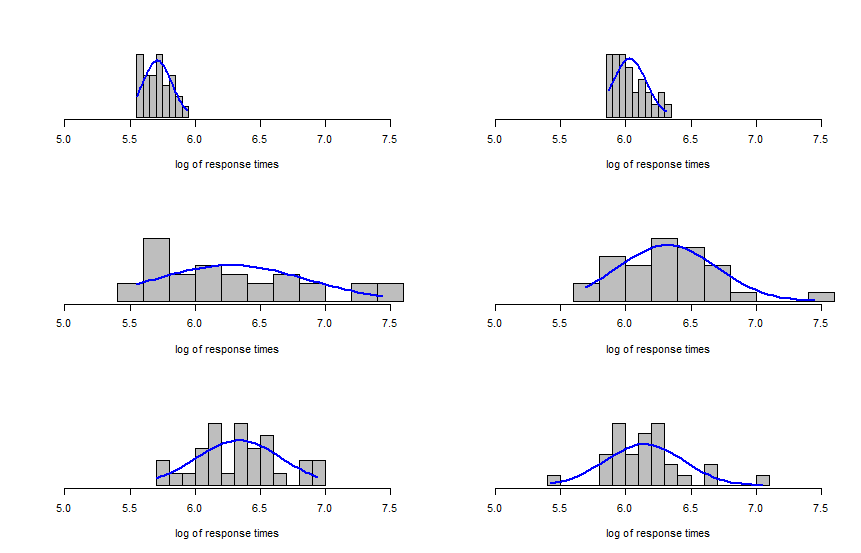


Fig.4. Histogram of Y = log (response time in milliseconds) 6 Schizophrenic individuals



So for future analysis in this paper, Y= log(response time in milliseconds) for these two groups will be adopted.

And boxplots for each group and each individual are produced in R as Fig. 5 and Fig.6.

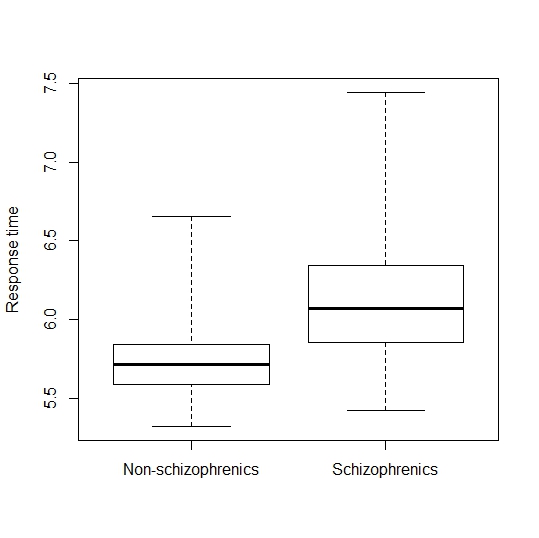
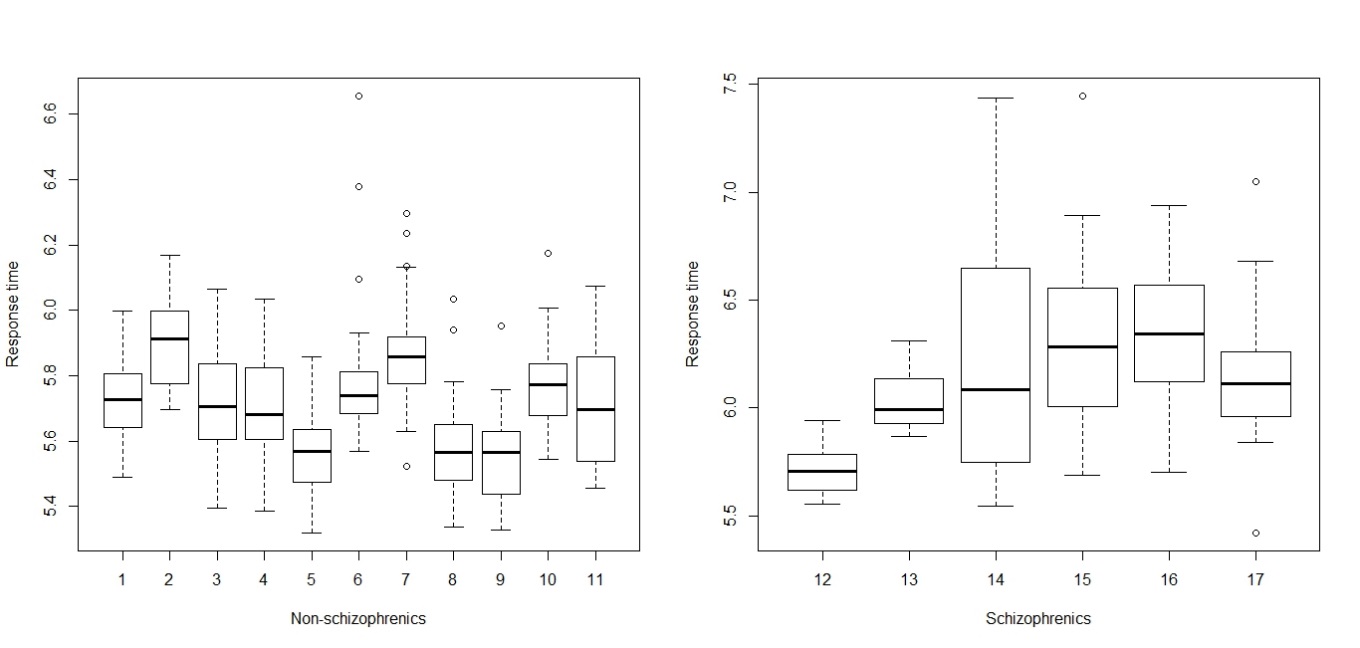
Fig.5, Boxplot of response time for non-schizophrenics and schizophrenics 

Fig.6, Boxplot of response time for 11 non-schizophrenics and 6 schizophrenics



According to these two boxplots, we conclude that

* Response times appear longer for schizophrenics than for non-schizophrenics.
* Response times for most of the schizophrenics appear more variability than those for non-schizophrenics.

And these two results can also be obtained from Fig. 3 and Fig. 4.

### 2.2 Bayesian Statistical Methods

In frequentist inference, parameters are fixed quantities. But in Bayesian inference, the true value of a parameter can be thought of as being a random variable to which we assign a probability distribution (Stevens, J.W., 2009). Bayesian inference has got its name from the use of Bayes theorem. If we are interested in estimating from data Bayes’ rule can be expressed as

Here, is a probability distribution for which is known as the prior distribution. It can express your beliefs about the parameter before you examine the data. is the likelihood function combining the data and prior information. is posterior distribution that you can update your beliefs about . Since is the probability of data that is a constant, posterior probability is always proportional to prior probability timing likelihood. This means posterior distribution is under the same distribution family of the prior which called conjugate prior.

### 2.2.1 Gibbs Sampling

Markov Chain Monte Carlo (MCMC) is a class of methods in which we can simulate draws that are slightly dependent and are approximately from a (posterior) distribution. It is so-named because one uses the previous sample values to randomly generate the next sample value to generating a Markov chain. Markov chain isa sequence x1, x2,… of random elements of some set if the conditional distribution of xn+1 given x1, . . . ,xn depends on xn only (Geyer, C.J., 2011). Gibbs sampling is one particular MCMC methods which is very widely applicable in Bayesian analysis. The first step of Gibbs sampling is to derive the posterior conditional for each of the random variables. Then we simulate posterior samples from the target joint posterior by iteratively sampling a value for a random variable from its corresponding posterior conditional while other variables are fixed to their current values (Yildirim, I., 2012). For example, if we are interested in sampling from the posterior , where is observed data and contains three parameter Gibbs sampling steps are:

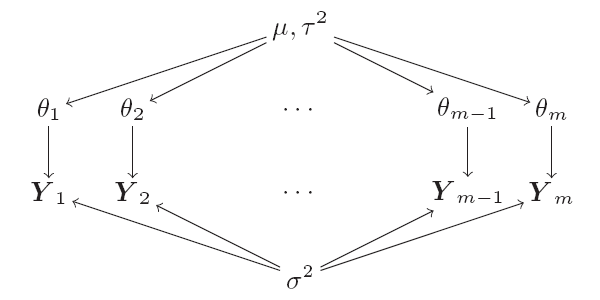
1. Define the set of full conditional distributions of , and
2. Choose a vector for starting values
3. Start with any (can also start with or , order does not matter), then draw a value from the full conditional distribution
4. Using value , draw a value (also can draw ) from the full conditional distribution
5. Using both updated value and , draw a value from conditional distribution
6. Using to draw with most updated values
7. Repeat until we get draws for vector with number we expected.

Then we can get a Markov chain with bunch of draws of that are approximately from the posterior distribution.

### 2.2.2 Hierarchical Normal Model

Bayesian hierarchical model is a [statistical model](https://en.wikipedia.org/wiki/Statistical_model) written in multiple levels that estimates the [parameters](https://en.wikipedia.org/wiki/Parameters) of the [posterior distribution](https://en.wikipedia.org/wiki/Posterior_probability) using the [Bayesian method](https://en.wikipedia.org/wiki/Bayesian_inference). It is a popular model for describe the heterogeneity of means across several populations, in which the within- and between- group sampling models are both normal (Hoff, P.D., 2009). And the structure for model is as Fig.7.

Fig.7, A graphical representation of the basic hierarchical normal model



In this graphical representation,

* The observations {,…} are independent and identical where follows , where{, …, } is within-group sampling distribution.
* {, ..., } is a random variable with parameter which is between-group heterogeneity in population means.

### 2.2.3 Analysis and Results

Here we will separately analyze non-schizophrenics and schizophrenics with Bayesian hierarchical modeling. To compare the responsible time for each group, we will represent the distribution posterior of mean for them. From the data, response times for each individuals are distributed normally. And in each group, we assume the population variance is same across individuals. In this model, we have three parameters . Because the gamma family does turn out to be a conjugate class of densities for which is referred to as precision (Hoff, P.D., 2009). For the variance parameter, inverse gamma commonly considered as a prior distribution which can lead to proper posterior distribution (Gelman, A., 2006). The Gibbs sampling procedure is (Hoff, P.D., 2009):

1. Specify three prior distributions:

~

~

~

1. We already knew principle of Gibbs sampling in section 2.2.2.1. Given a current state of the unknown , a new state is generated as:
2. Sample ;
3. Sample ;
4. Sample
5. for each , sample
6. After complicated calculation, we get each posterior as full conditional distribution:

~

~

~

~

1. Specify the starting values and prior parameter values:
2. For starting values of parameters in model, we set them based on the observed data. R-script (see appendix) of this paper shows how exactly starting values are set.

* for calculation of full conditional distributions of is overall mean of variance for each individuals.
* for calculation of full conditional distributions of is overall mean of all observed data.
* for calculation of full conditional distributions of and is variance for the mean value of each individuals.

1. For the parameters of three priors, values are set as:

* For prior of of non-schizophrenics, is 5.7 is from the overall mean of this group is 5.7. is 3 since for the normal distribution, most of the probability is within two standard deviation. Response time is higher than zero, which is > 0, so we choose is 2.85. The same rule is for schizophrenics, is 6.1 and is 3.
* For prior of (between-group) and (within-group), and are considered as prior sample size meanwhile and are considered as the sample variances, which implies when → 0, the posterior distribution is more objective (Hoff, P.D., 2009). The range of sample data for non-schizophrenics is around 1 and range for schizophrenics is around 2. So for group of non-schizophrenics and for group of schizophrenics are set. Since different sample will generate different variance, we specify as weakly informative prior parameter which is set purposely to include less information than whatever actual prior knowledge is available (Gelman, A., 2006) so that prior distribution can weakly centered around this value from other populations (Hoff, P.D., 2009).

1. Results from Gibbs sampling in R:

After Gibbs sampling of 10000 iterations based on hierarchical normal modeling with these prior information, the marginal posterior distributions for are shown as Fig.8 and Fig.9.

Fig.8, posterior distributions for of non-schizophrenics

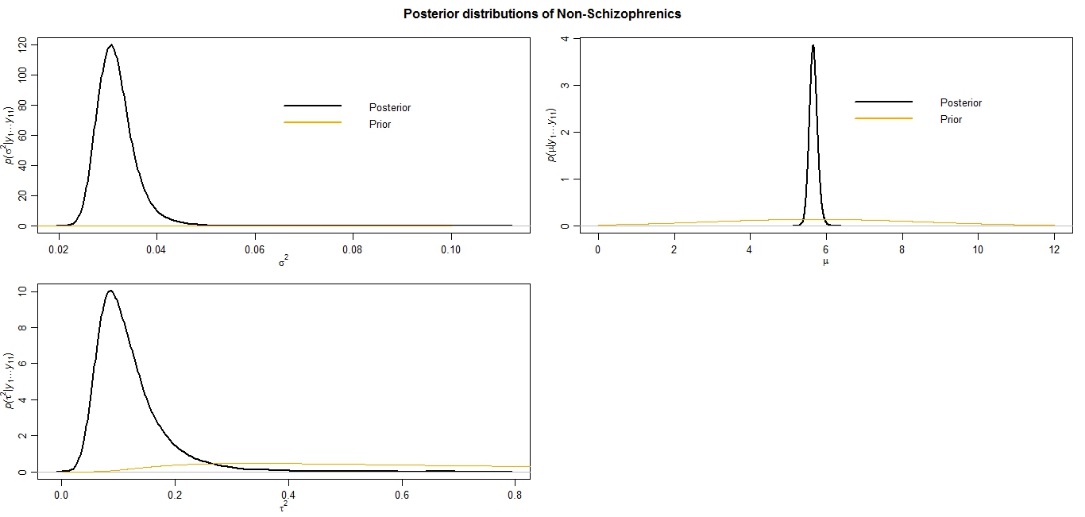
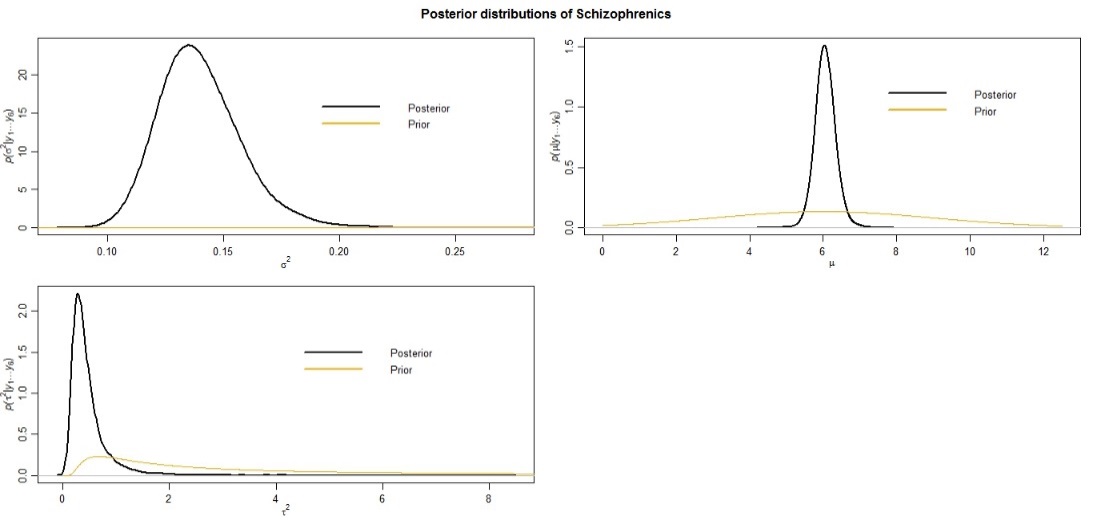
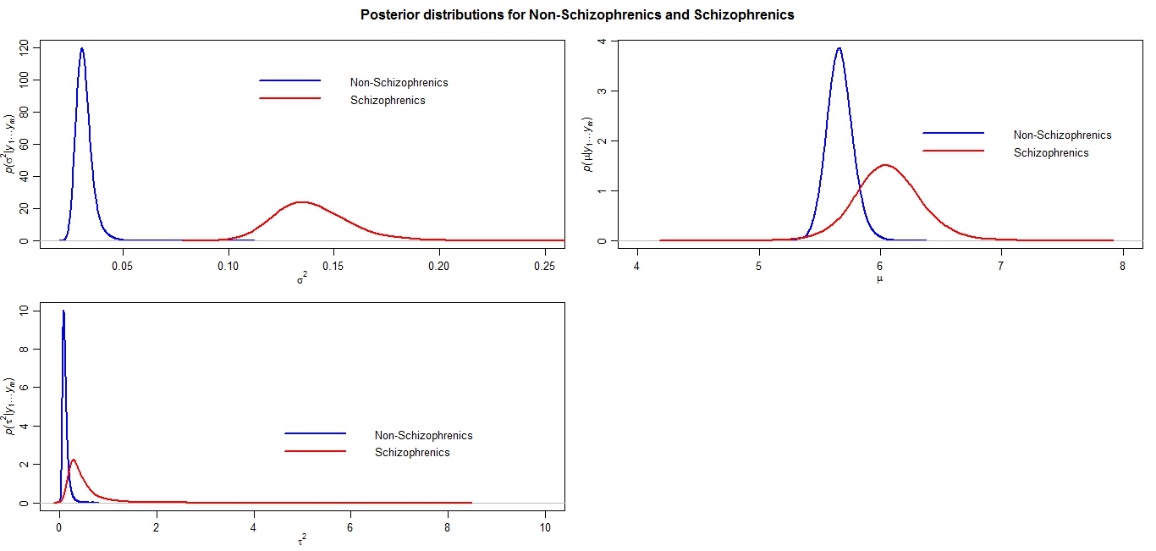


Fig.9, posterior distributions for of schizophrenics



As shown in Fig.8 and Fig.9, all posterior distributions are more concentrated than their corresponding prior distributions. The posterior means of are 0.03174972, 5.66786683 and 0.11774262 respectively to non-schizophrenics. In schizophrenics group, the posterior means of are 0.1396830, 6.0653227 and 0.4646806. Compare two left panels for each group, we can see the range for most of posterior of is lower than the range for most posterior of which implicates between-group variability () is higher than within-group variability (). This makes sense since generally variabilities from different individuals is higher than the variabilities from the same individuals on repeated measures.

Fig.10, posterior distributions for of non-schizophrenics and schizophrenics



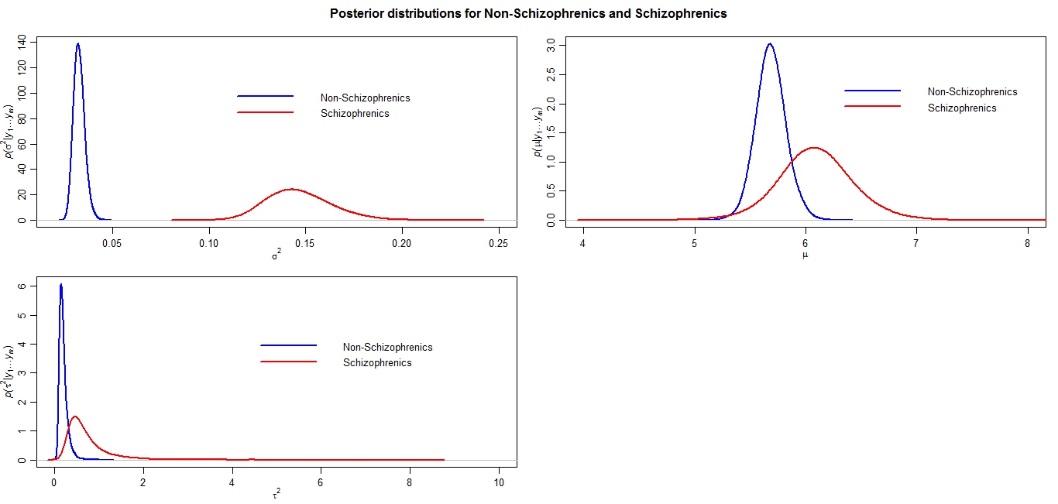
Comparing non-schizophrenics and schizophrenics with all posterior distributions of is shown as Fig.10. According to the posterior distribution of which is in the right plot, we can get there’s big chance that schizophrenics has longer response time than non-schizophrenics. And this probability is obtained in R program as 0.9206 (see appendix). From the right two plots, we conclude that within-group () variabilities and between-group variabilities () of schizophrenics are all higher than them from non-schizophrenics which is consistent with the result from the observed data in section 2.1 of this paper.

### 2.3 Analysis with different Prior Parameters

Since in Bayesian inference, posterior distributions is under the same family of prior distributions. So posterior is sensitive with prior distributions with proper parameter values. And several sets of parameter values for prior distributions of and will be tried here based on the rules mentioned in section 2.2.3. Different parameter values for prior won’t be re-considered here because and are from fixed observed data with normal distribution.

* Non-schizophrenics,

Schizophrenics,

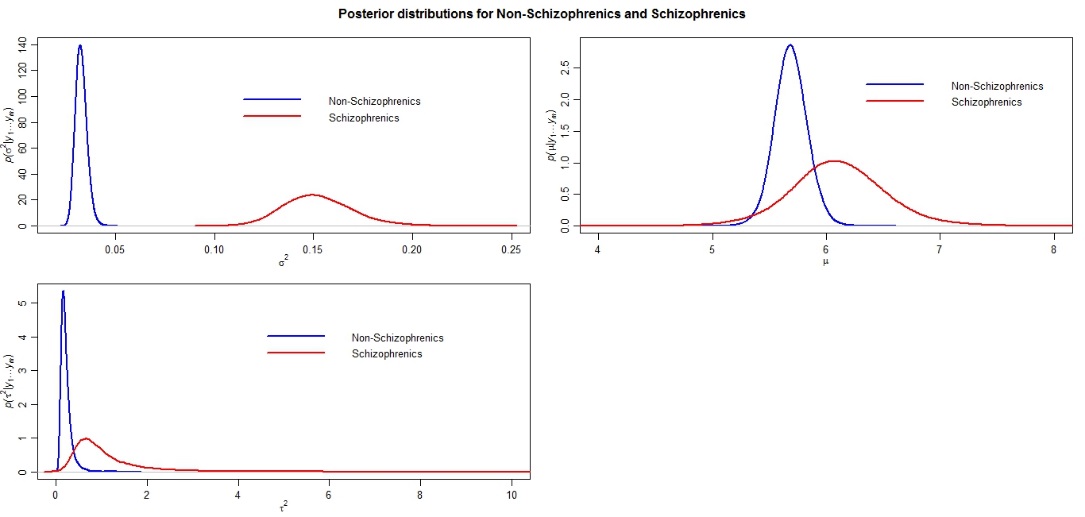
Fig.11, comparison of the posterior distribution for two groups

The posteriors of non-schizophrenics and schizophrenics comparison is as Fig.11. From R program, the probability that schizophrenics has longer response time than non-schizophrenics is 0.8804.

* Non-schizophrenics,

Schizophrenics,

Fig.12, comparison of the posterior distribution for two groups

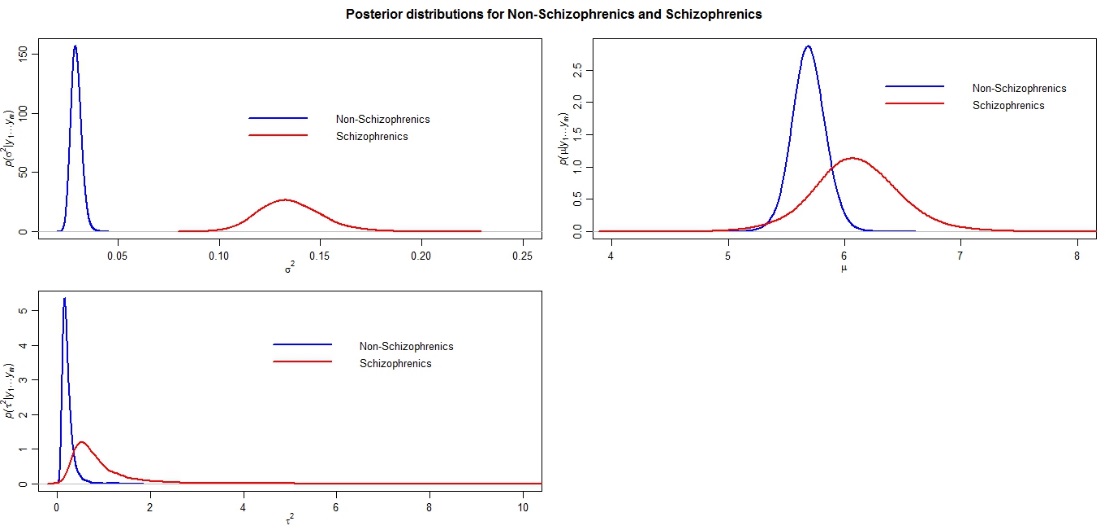


If we considered variabilities from population of schizophrenics and non-schizophrenics are higher than it form the observed data, we get the comparison as Fig.12. And in this situation, the probability that schizophrenics has longer response time than non-schizophrenics is 0.8466.

* Non-schizophrenics,

Schizophrenics,

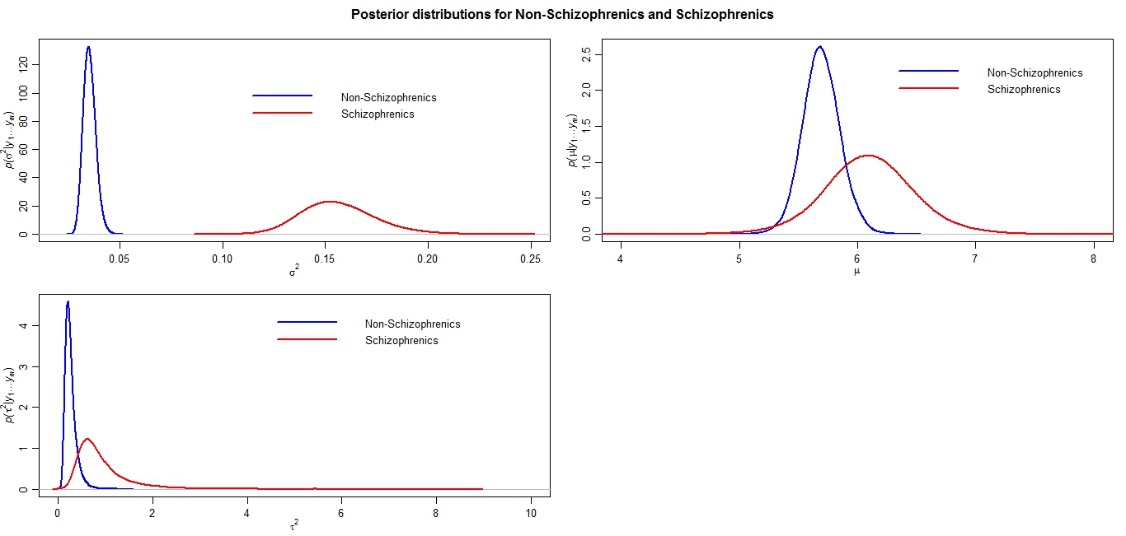
Fig.13, comparison of the posterior distribution for two groups



If we considered between-group variabilities are higher than within-group variabilities for non-schizophrenics and schizophrenics, we get the comparison as Fig.13. Here, the probability that schizophrenics has longer response time than non-schizophrenics is 0.8620.

* Non-schizophrenics,

Schizophrenics,

Fig.14, comparison of the posterior distribution for two groups 

The chance that schizophrenics has longer response time than non-schizophrenics here is 0.8552. Based on analysis with different parameter value for prior and , we can see there’s no big differences among these posterior distributions from these situations. But with higher parameter values, the posterior distributions is more spread out and the posterior of mean for schizophrenics has relative lower probability than posterior of mean for schizophrenics.

### 2.4 Summary

We used a published dataset to present a Gibbs sampling example in R program. In hierarchical normal model, MCMC samples of were simulated. Based on the plot programmed, we can see the posteriors are obviously more concentrated than the priors and non-schizophrenics have shorter response time than schizophrenics. Also, several simulation with reasonable priors were compared and we found different prior didn’t impact our model simulation.

# Estimating Convergence of MCMC

### 3.1 Convergence of MCMC

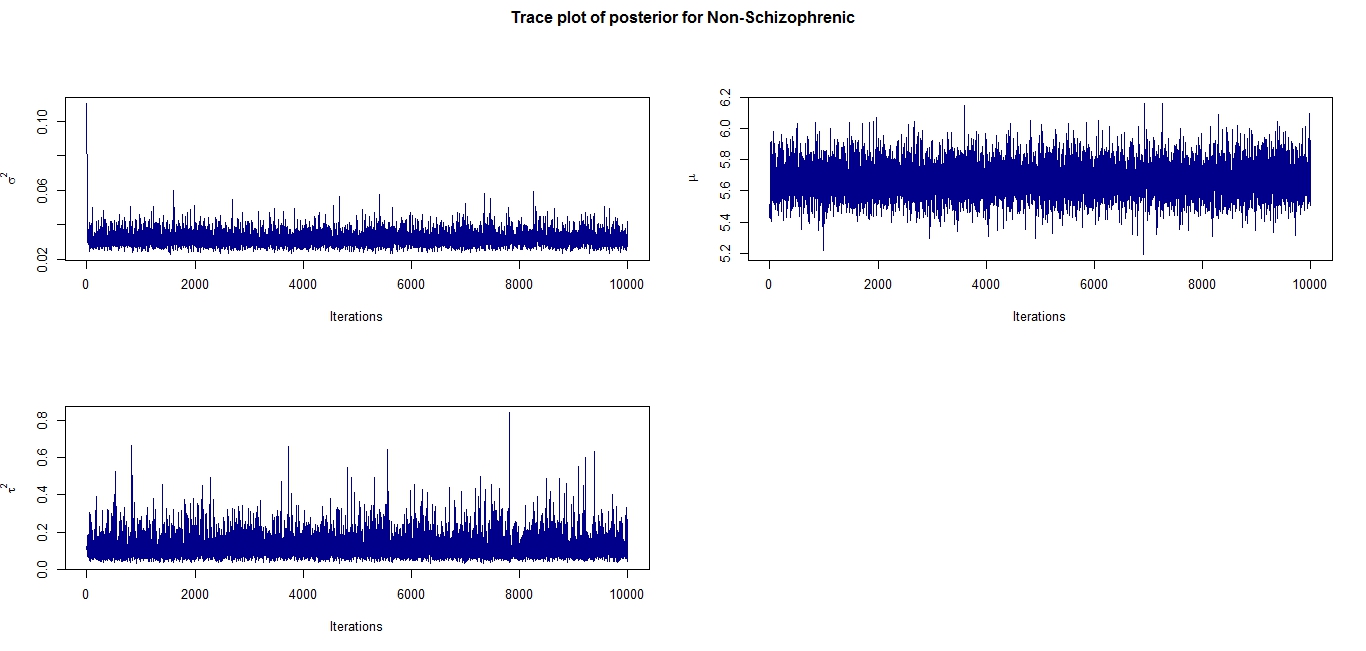
Like all statistical methods, MCMC method has its own disadvantages. The key one is the difficulty to determine the convergence of this algorithm. After the sample chain from MCMC algorithm being constructed, we should assess whether the chain is under a stationary distribution, how many steps are needed to converge to the stationary distribution and whether the stationary distribution is reached quickly and so on. This stationary distribution is the true posterior distribution that we are interested. The convergence estimation is important since the MCMC simulation is often started at a random point in parameter space and is often far from the true high density regions of the posterior distribution (Sahli, K., 2011) and we should make sure the efficiency of our simulated samples before making any inference.

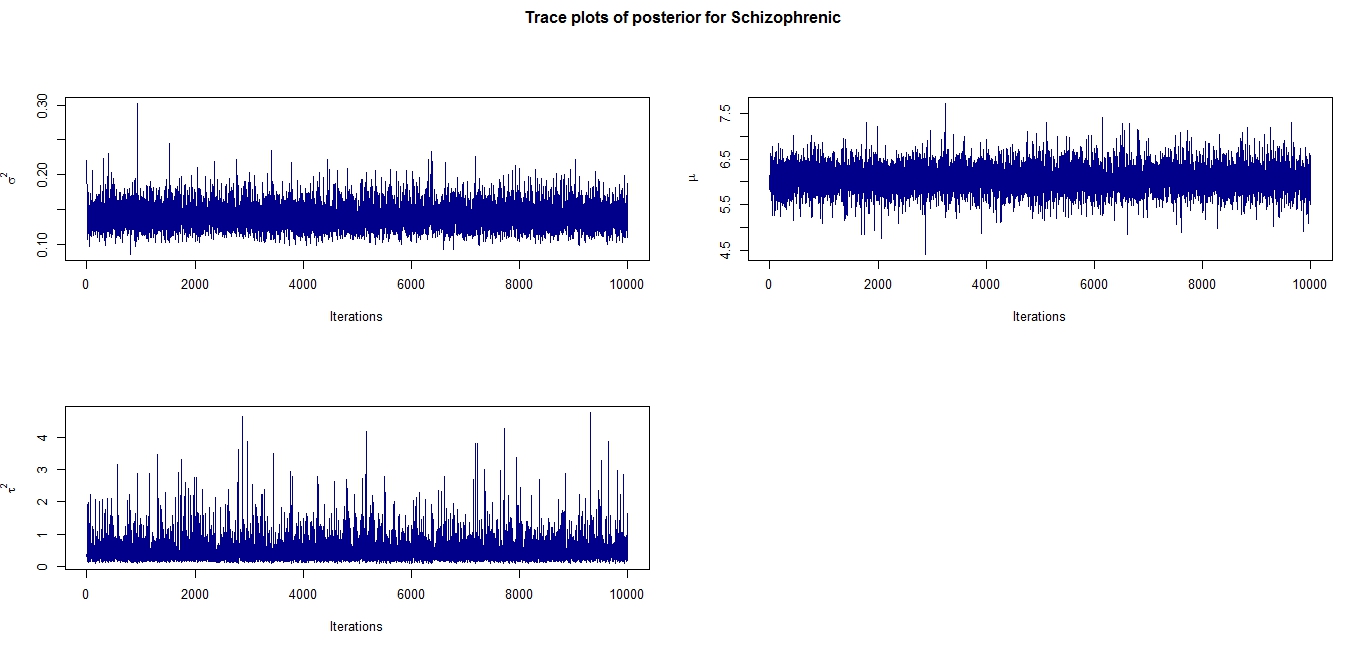
### 3.2 Analysis and Results with different methods

CODA is a menu-driven set of S-Plus functions which serves as an output processor of BUGS (Bayesian inference Using Gibbs Sampling) software (Best, N., Cowles, M.K. and Vines, K., 1995). It can be used to perform convergence diagnostics. So for this paper, we mainly use functions in this R package to estimate convergence of MCMC. We use visual inspection at first, such as trace plots and stationary boxplots, to see how the chain mixing and how the samples distribute. Then we use “geweke. diag”, “heidel.diag” and so on (in coda) which are statistical diagnostics to generate output for assessing MCMC convergence.

### 3.2.1 Trace Plots

A trace plot is a plot of the iteration number against the value of each sample of MCMC at each iteration. This kind of plots is very useful for assessing convergence. A stationary distribution can be inferred from a trace plot with relatively constant mean and variance. A chain that mixes well traverses its posterior space rapidly, and it can jump from one remote region of the posterior to another in relatively few steps ([1]). Fig.15 and Fig.16 are trace plots for each Gibbs sample of posterior for non-schizophrenics and schizophrenics. These plots all show good mixing which implies our MCMC converge.

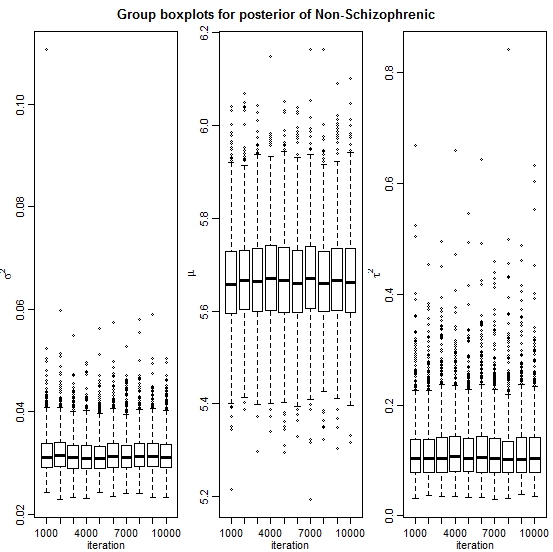
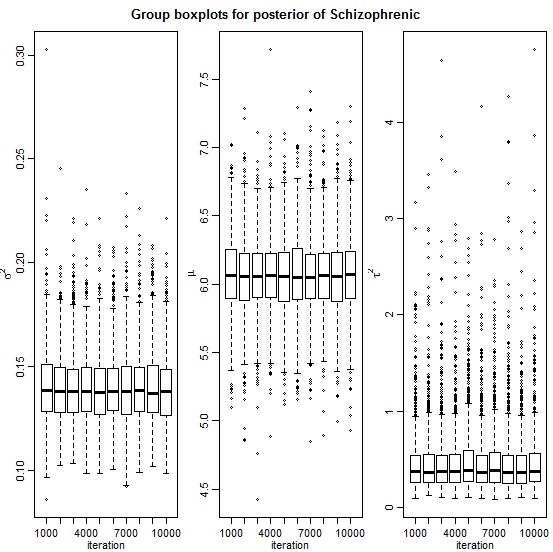
Fig.15, Trace plots for MCMC samples of non-schizophrenics Fig.16, Trace plots for MCMC samples of schizophrenics



### 3.2.2 Stationary Boxplots

The direct stationary boxplot shows us the relationship between the simulated parameter values and iteration number. This approach is to produce boxplots of sequential groups of samples. If the distribution of samples in any one boxplot is the same as that in any other, and the medians or interquartile ranges of the boxplots were not moving in a consistent direction with iteration number, we can conclude that stationarity has been achieved (Hoff, P.D., 2009). Otherwise, we would suspect that stationary has not been reached and we need run long chain. Fig.17 is the boxplots of sequential groups of MCMC samples for non-schizophrenics and schizophrenics (see R-script in appendix).

Fig.17, sequential boxplots for the sample of posterior distribution from MCMC

The stationarity plots in Fig.17 above gave us a clear presentation about the simulated MCMC sequential sample boxplots. Each boxplot contains 1000 posterior samples. Based on the rules of this method mentioned above, we conclude that stationarity has been achieved and we don’t need to run longer chain.

### 3.2.3 Autocorrelation

One nature of Markov chain is that members of a sample will generally be correlated with each other which will slow the algorithm in its attempt to sample from the entire stationary distribution (Cowels, M.K. and Carlin, B.P., 1996). Autocorrelation is a measure of how independent different samples from your posterior distribution are. A Markov chain with high autocorrelation moves around the parameter space very slow, taking a long time to achieve the correct balance among the different regions of the parameter space. The higher the autocorrelation, the more MCMC samples we need to attain a given level of precision for our approximation (Hoff, P.D., 2009). So lower autocorrelation means higher efficiency in your chain. In this paper, we use R function “acf” to represent autocorrelation plot for each MCMC output (see R-script in appendix).

Fig.18, autocorrelation for each MCMC samples of non-schizophrenics

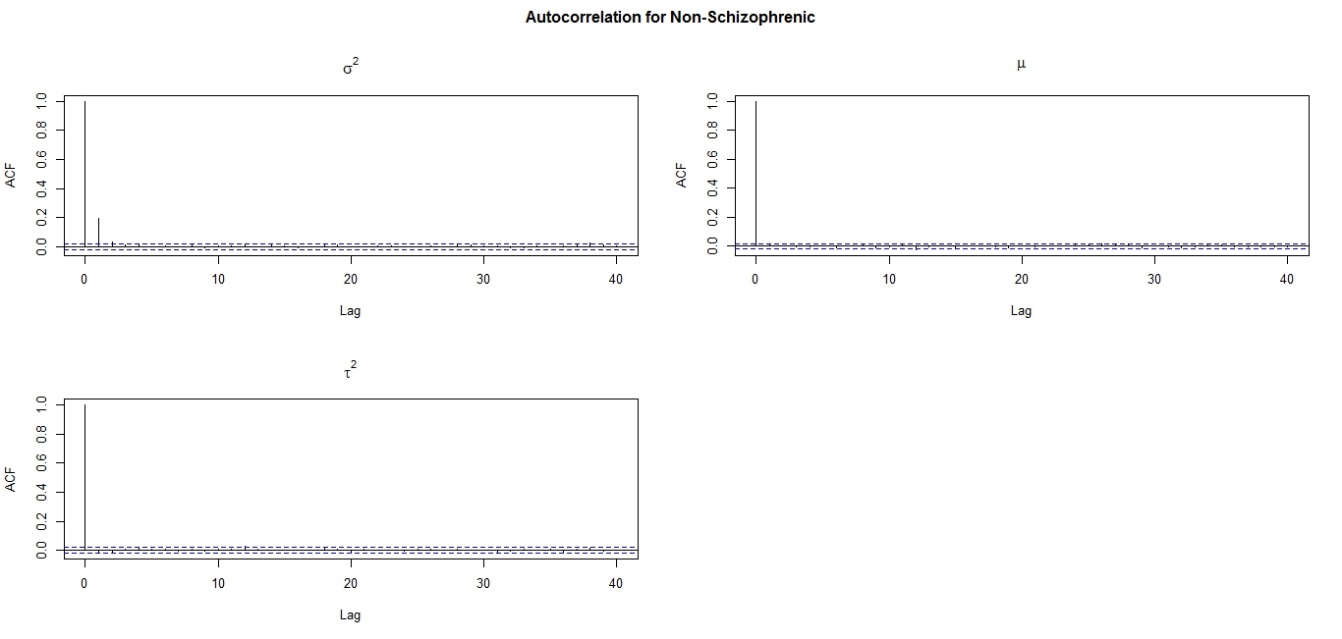


Fig.19, autocorrelation for each MCMC samples of non-schizophrenics

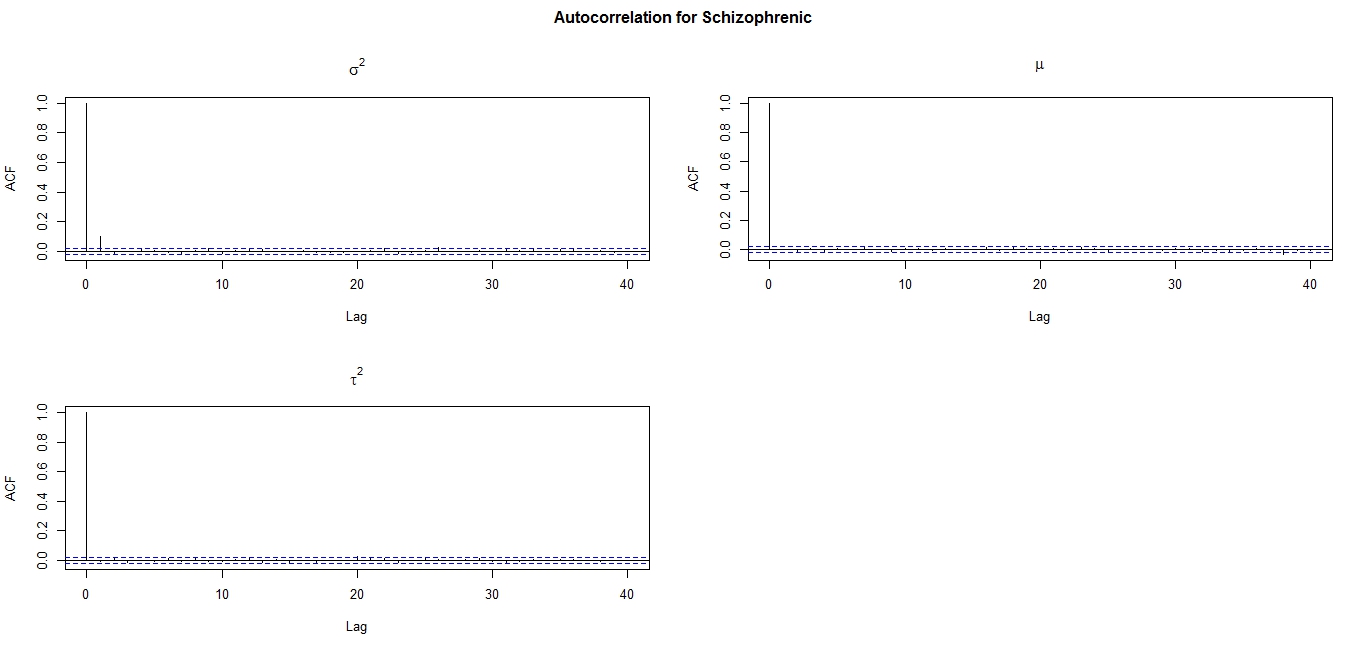


Fig.18 and Fig.19 are plots of showing the autocorrelation of MCMC sample for each parameter posterior. Y-axis represents the autocorrelation coefficients. The blue dotted line is the 95% confidence interval. We can see all autocorrelations are small and closer to 0 from lag-1 or lag-2. This indicates our six Markov chain converge to stationary distribution very quickly and take a short time to achieve the correct balance among the different regions of the parameter space.

### 3.2.4 Effective Sample Size

Effective sample size is a relative concept of autocorrelation of MCMC. It is he number of effectively independent samples from the posterior distribution which can measure how much information you are really get. Large difference between effective sample size and simulation sample size from one chain means high autocorrelation. Here we use R-function “effectiveSize” (in package “coda”) to compute the effective sample size for each Markov chain. R result is as below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Paremeters** |  | **µ** |  | **Group** |
| **Effective sample size** | 6734.682 | 10000 | 10000 | Non-schizophrenic |
| **Effective sample size** | 8639.629 | 10000 | 9822.771 | Schizophrenic |

For the of non-schizophrenics, we can explain the sample size as the precision of the MCMC estimate of the posterior of based on 10000 samples is good as taking 6734.682 independent samples from the distribution. We can see effective sample size from all these MCMC are relative high after 10000 iterations. For Markov chain of µ and in non-schizophrenics and µ in schizophrenics, effective sample sizes even are 10000. High effective sample size means our MCMC variance is small and the approximations are good enough.

### 3.2.5 Raftery and Lewis

This diagnosis is to evaluate the accuracy of the posterior percentile you are interested and detect the convergence to the stationary distribution. We need specify a posterior quantile you want to estimate, the desired degree of accuracy for this estimation and the required probability of getting this degree of accuracy (Best, N., Cowles, M.K. and Vines, K., 1995). We use function “raftery.diag” in R coda package to assess convergence. The results from R are shown as below:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MCMC | Quantile (q)  0.025 | | | Accuracy (r)  +/- 0.005 | | | | Probability (s)  0.95 | | |
| Non-schizophrenics | | | | | Schizophrenics | | | | |
| M | N | Nmin | | I | M | N | | Nmin | I |
|  | 2 | 3962 | 3746 | | 1.06 | 2 | 3771 | | 3746 | 1.010 |
|  | 2 | 3819 | 3746 | | 1.02 | 2 | 3771 | | 3746 | 1.010 |
|  | 2 | 3802 | 3746 | | 1.01 | 2 | 3650 | | 3746 | 0.974 |

* M – number of initial iterations to discard as burn-in
* N – total number should be run for Markov chain
* Nmin – the minimum number of iterations that would be necessary to estimate the quantile with certain accuracy and probability; it will increase as the probability and degree of accuracy increase
* I – dependency factor (), measures the increase of the iteration number needed to reach convergence due to dependence between samples; I > 5 indicates influential starting point, high correlations between samples or poor mixing

From the output, we can see in our R procedure, q, r and s is 0.025, 0.005 and 0.95. For all MCMC, we estimate 0.025 quantile with an accuracy ±0.005 with 95% probability, we need run minimum of 3746 iterations which are all lower than our 10000 iterations. The small burn-in values 2 suggest all of the Markov chains converged immediately.

### 3.2.6 Geweke

The Geweke diagnostic is a time-series approach that compares the mean and variance of segments from the beginning and end of a single chain (usually the first 0.1 and last 0.5 proportions), through a difference of means test to see if the two parts of the chain are from the same distribution which is null hypothesis. The test statistic is a standard Z-score with the standard errors adjusted for autocorrelation. As the chain length → ∞, the sampling distribution of Z → N(0,1) if the chain has converged. In R coda package, function “geweke.diag” is used to calculate this Z-score for MCMC samples.

|  |  |  |  |
| --- | --- | --- | --- |
| MCMC | Fraction in 1st window = 0.1  Fraction in 2nd window = 0.5 | | |
|  |  |  |
| Non-schizophrenics | 0.3583 | -0.9002 | -0.5294 |
| Schizophrenics | 1.1219 | 0.5408 | -0.2863 |

The above is results from R for the result of MCMC for non-schizophrenics and schizophrenics. Since for these six chains, we have 10000 iterations, first 1000 samples and last 5000 samples for each chain are compared. The absolute value of these Z-score are all less than 1.96 (p-value>0.05) indicates failing to reject null hypothesis. So the two parts of each chain are under same distribution.

We can also use function “geweke.plot” to plot Geweke Z-scores for different segments of the whole MCMC sample to estimating the convergence. The first half of the Markov chain is divided into nbins - 1 segments (default value for nbins is 20), then Geweke's Z-score is repeatedly calculated. The first Z-score is calculated with all iterations in the chain, the second after discarding the first segment, the third after discarding the first two segments, and so on. The last Z-score is calculated using only the samples in the second half of the chain.

Fig.20, Geweke diagnostic plot for non-schizophrenics

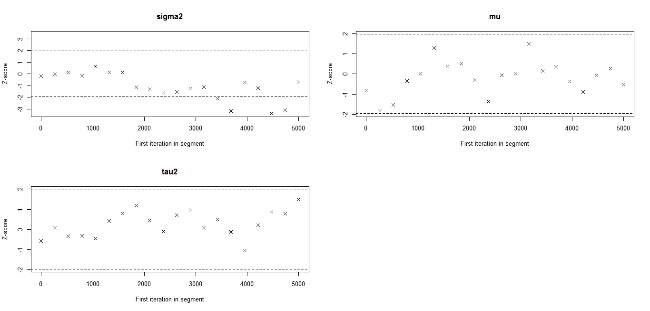
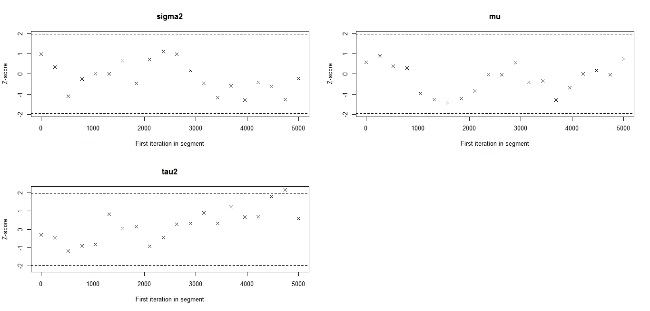


Fig.21, Geweke diagnostic plot for schizophrenics

In Fig.20 and Fig.21, horizontal dotted lines at Z=-1.96, 1.96 represent the 95% confidence interval for an N(0, 1) distribution. We can see all Z-scores of each MCMC are distributed from 0 to 2 except 4 outliers of the chain for within-group variance ( of non-schizophrenics and 1 outlier from the chain of between-group variance () of schizophrenics. Since a large number of Z-scores falling outside the interval (-1.96, 1.96) suggests possible convergence failure (Best, N., Cowles, M.K. and Vines, K., 1995). Our six simulated MCMC samples are converged. If there’s lots of Z-scores is outside the interval, we should run longer chain.

### 3.2.7 Heidelberg and Welch

This diagnosis is composed of two parts: stationarity test and half-width test. Stationarity test calculates a test statistics to reject or accept the null hypothesis that the Markov chain comes from a stationary process (Best, N., Cowles, M.K. and Vines, K., 1995): Generate a chain of N iterations and define α level; calculate the test statistic on the whole chain and accept or reject null hypothesis that the chain is from a stationary distribution; if null hypothesis is rejected, discard the first 10% of the chain and calculate the test statistic to decide accept or reject null hypothesis; if null hypothesis is rejected, discard the next 10% and calculate the test statistic; repeat until null hypothesis is accepted or 50% of the chain is discarded. If test still rejects null hypothesis, then the chain fails the test and needs to be run longer.

And half-width test checks whether the Markov chain sample size is adequate to estimate the mean values accurately. It calculates half the width of the (1 − ɑ)% credible interval around the mean. If the ratio of the halfwidth and the mean is lower than some ϵ, then the chain passes the test. Otherwise, the chain must be run out longer.

We use function “heidel.diag” (default values for α and ϵ is 0.05 and 0.1) in R coda package, the results are shown as below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| MCMC | Non-schizophrenics | | | Schizophrenics | | |
| Stationary  test | Start  iteration | p-value | Stationary  test | Start  iteration | p-value |
|  | passed | 1 | 0.210 | passed | 1 | 0.746 |
|  | passed | 1 | 0.761 | passed | 1 | 0.949 |
|  | passed | 1 | 0.857 | passed | 1 | 0.664 |
| MCMC | Half-width  test | Mean | Halfwidth | Halfwidth  test | Mean | Half-width |
|  | passed | 0.0317 | 9.28e-05 | passed | 0.140 | 0.000365 |
|  | passed | 5.6679 | 2.05e-03 | passed | 6.065 | 0.005422 |
|  | passed | 0.1177 | 1.15e-03 | passed | 0.465 | 0.006949 |

From the results above, we can see non-schizophrenics and schizophrenics passed stationary test since all p-value > 0.5 that fail to reject the chain comes from a stationary distribution. We also can see all half-width passed since all half-width is less than ϵ times the sample mean So We don’t need run a longer chain.

From results from all seven different MCMC convergence diagnostics above, our six Markov chains all reached convergence and don’t need to run chain longer.

# DISCUSSION

For the response time of schizophrenics, we may need to consider the probability of having response delays. Since current psychological theory suggest that schizophrenics on some measure trials they would delay in starting a reaction or they may have slower reflexes once a reaction has commenced. But it won’t always happen. In future, more and more accuracy measurements for schizophrenic will be necessary. If we can know how the response delays distribute in schizophrenics, more complex and accurate Bayesian model or methods will be more suitable for our analysis.

Many diagnostics for estimating convergence are designed for detecting some kind of convergence. It is possible they would be failure in other statistical practice, even simpler one. So when we assessing MCMC convergence, we need use a variety diagnostic tools rather than any single one. Besides parameter we are interested in, we must check convergence of all parameters. If some of them have bad mixing, we can’t get accurate inference from the one that have good mixing. At last, before applying MCMC algorithm, we should learn as much as possible about the dataset and target density. Deeper understand of data and complexity increasing of model will help us detecting convergence failures.

# REFERENCES

Rathnayake, R.C. (2010) Bayesian Inference and Computation. *Research Papers. Paper 29.*

Parker, M. (2005) Foundations of Statistics – Frequentist and BayesianMartin, A.D., Quinn, K.M. & Park, J.H. (2011) MCMCpack: Markov Chain Monte Carlo in R: *Journal of Statistical Software, volume 42, issue 9*

Cowels, M.K. and Carlin, B.P. (1996) Markov Chain Monte Carlo Convergence Diagnostics: A comparative Review. *Journal of the American Statistical Association*

Belin, T.R. and Rubin, D.B. (1990). Analysis of a finite mixture model with variance components. *Proceeding of the Social Social Statistics Section 211-215. ASA, Alexandria, Va.*

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari A. and Rubin, D.B. (2013) Bayesian Data Analysis: *Chapman & Hall/CRC Texts in Statistical Science, 3rd Edition*

Stevens, J.W. (2009) What is Bayesian Statistics?

Sahli, K. (2011) Estimating convergence of Markov chain Monte Carlo simulations: *Master thesis for Mathematical Statistics, Stockholm University*

Geyer, C.J. (2011) Introduction to Markov Chain Monte Carlo: *Handbook of Markov Chain Monte Carlo, by Chapman & Hall/CRC*

Yildirim, I. (2012) Bayesian Inference: Metropolis-Hastings Sampling

Hoff. P.D. (2009) A First Course in Bayesian Statistical Methods: *Springer*

Gelman, A. (2006) Prior distribution for variance parameters in hierarchical models: *1, Number 3, pp. 515-533*

Best, N., Cowles, M.K. and Vines, K. (1995) CODA\* Convergence Diagnosis and Output Analysis Software for Gibbs sampling output: *Version 0.30*

Other Reference:

[1][http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug\_introbayes\_sect008.htm](http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm" \l "statug_introbayes_sect008.htm)

# APPENDIX

## R-script:

##Read crazy.txt into R

#Name Non-schizophrenics group as y1 and Schizophrenics group as y2

RES=read.table(file.choose())

y1=t(RES[1:11,])

y2=t(RES[12:17,])

#Generate 11 histograms of response time for Non-schizoprhenic individuals

par(mfrow=c(4,3))

q=11

for(q in 1:q){

x <- y1[,q]

h<-hist(x, breaks=6, col="grey", xlab="response times", xlim = c(200, 800), yaxt='n',ylab=NULL, main=NULL)

xfit<-seq(min(x),max(x),length=40)

yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))

yfit <- yfit\*diff(h$mids[1:2])\*length(x)

lines(xfit, yfit, col="blue", lwd=2)}

#Generate 6 histograms of response tim for Schizophrenic individuals

dev.off()

par(mfrow=c(3,2))

w=6

for(w in 1:w){

x <- y2[,w]

h<-hist(x, breaks=12, col="grey", xlab="response timess", xlim = c(200, 1750), yaxt='n',ylab=NULL, main=NULL)

xfit<-seq(min(x),max(x),length=40)

yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))

yfit <- yfit\*diff(h$mids[1:2])\*length(x)

lines(xfit, yfit, col="blue", lwd=2)}

#Make logarithms of y1 and y2 and rename them as y11 and y22

y1=log(y1)

y2=log(y2)

#Generate 11 histograms of log(response time) Non-schizophrenic individuals from y11

par(mfrow=c(4,3))

q=11

for(q in 1:q){

x <- y1[,q]

h<-hist(x, breaks=6, col="grey", xlab="log of response times", xlim = c(5.0, 7.5), yaxt='n',ylab=NULL, main=NULL)

xfit<-seq(min(x),max(x),length=40)

yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))

yfit <- yfit\*diff(h$mids[1:2])\*length(x)

lines(xfit, yfit, col="blue", lwd=2)}

#Generate 6 histograms of log(response time) Schizophrenic individuals from y22

dev.off()

par(mfrow=c(3,2))

w=6

for(w in 1:w){

x <- y2[,w]

h<-hist(x, breaks=12, col="grey", xlab="log of response times", xlim = c(5.0, 7.5), yaxt='n',ylab=NULL, main=NULL)

xfit<-seq(min(x),max(x),length=40)

yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))

yfit <- yfit\*diff(h$mids[1:2])\*length(x)

lines(xfit, yfit, col="blue", lwd=2)}

#Make boxplot about y1 and y2

dev.off()

boxplot(list(y1,y2),range=0,ylab="Response time",

names=c("Non-schizophrenics","Schizophrenics"))

#Overall histogram of y1 and y2 containing 17 individuals

par(mfrow=c(1,2))

boxplot(y1,ylab="Response time",xlab="Non-schizophrenics")

boxplot(y2,ylab="Response time",xlab="Schizophrenics")

##Gibbs sampling

#Set starting values for Non-schizophrenic individuals from y1

m1=11;m2=6;n=30

theta1=y1bar=apply(y1,2,mean)

sv1=apply(y1,2,var)

sigma2\_1=mean(sv1)

mu1=mean(y1)

tau2\_1=var(y1bar)

#Set starting values for Schizophrenic individuals from y2

theta2=y2bar=apply(y2,2,mean)

sv2=apply(y2,2,var)

sigma2\_2=mean(sv2)

mu2=mean(y2)

tau2\_2=var(y2bar)

#Given weakly informative prior parameters for log(response time millseconds)

nu0\_1=1; s20\_1=1 #within gropus (for prior of sigma2)

eta0\_1=1; t20\_1=1 #between groups (for prior of tau2)

mu0\_1=5.7; g20\_1=2.85 #for prior of mu

nu0\_2=1; s20\_2=2

eta0\_2=1; t20\_2=2

mu0\_2=6.1; g20\_2=3

#Setup MCMC

set.seed(234)

S=10000

THETA1=matrix(nrow=S,ncol=m1)

THETA2=matrix(nrow=S,ncol=m2)

MST1=MST2=matrix(nrow=S,ncol=3)

colnames(MST1)=colnames(MST2)=paste(c("sigma2","mu","tau2"))

#MCMC algorithm

for(s in 1:S){

for(j in 1:m1){

## sample new values of the thetas

vtheta=1/(n/sigma2\_1+1/tau2\_1)

etheta=vtheta\*(y1bar[j]\*n/sigma2\_1+1/tau2\_1)

theta1[j]=rnorm(1,etheta,sqrt(vtheta))

}

for(j in 1:m2){

vtheta=1/(n/sigma2\_2+1/tau2\_2)

etheta=vtheta\*(y2bar[j]\*n/sigma2\_2+1/tau2\_2)

theta2[j]=rnorm(1,etheta,sqrt(vtheta))

}

#sample new value of sigmas

nun\_1=nu0\_1+n\*m1

ss\_1=nu0\_1\*s20\_1

for(j in 1:m1){

ss\_1=ss\_1+sum((y1[,j]-theta1[j])^2)

}

sigma2\_1=1/rgamma(1,nun\_1/2,ss\_1/2)

nun\_2=nu0\_2+n\*m2

ss\_2=nu0\_2\*s20\_2

for(j in 1:m2){

ss\_2=ss\_2+sum((y2[,j]-theta2[j])^2)

}

sigma2\_2=1/rgamma(1,nun\_2/2,ss\_2/2)

#sample new values of mu

vmu\_1=1/(m1/tau2\_1+1/g20\_1)

emu\_1=vmu\_1\*(m1\*mean(theta1)/tau2\_1+mu0\_1/g20\_1)

mu\_1=rnorm(1,emu\_1,sqrt(vmu\_1))

vmu\_2=1/(m2/tau2\_2+1/g20\_2)

emu\_2=vmu\_2\*(m2\*mean(theta2)/tau2\_2+mu0\_2/g20\_2)

mu\_2=rnorm(1,emu\_2,sqrt(vmu\_2))

#sample new values of tau

etam\_1=eta0\_1+m1

ss\_1=eta0\_1\*t20\_1 + sum((theta1-mu1)^2)

tau2\_1=1/rgamma(1,etam\_1/2,ss\_1/2)

etam\_2=eta0\_2+m2

ss\_2=eta0\_2\*t20\_2 + sum((theta2-mu2)^2)

tau2\_2=1/rgamma(1,etam\_2/2,ss\_2/2)

#store results

THETA1[s,]=theta1

THETA2[s,]=theta2

MST1[s,]<-c(sigma2\_1,mu\_1,tau2\_1)

MST2[s,]<-c(sigma2\_2,mu\_2,tau2\_2)

}

#Non-schizophrenic posterior

par(mfrow=c(2,2),mar=c(3,3,1,1),mgp=c(1.75,.75,0),oma=c(0,0,2,0))

plot(density(MST1[,1],adj=2),lwd=2,main="",xlab=expression(sigma^2),

ylab=expression(paste(italic("p("),sigma^2,"|",italic(y[1]),"...",

italic(y[11]),")")))

x11=seq(0,0.1,length=1000)

lines(x11,dinvgamma(x11,nu0\_1/2,nu0\_1\*s20\_1/2),type="l",col="orange")

legend(0.06,90,legend=c("Posterior","Prior"),lwd=c(2,2),

col=c("black","orange"),bty="n")

plot(density(MST1[,2],adj=2),xlab=expression(mu),main="",lwd=2,xlim=c(0,12),

ylab=expression(paste(italic("p("),mu,"|",italic(y[1]),"...",

italic(y[11]),")")))

x21=seq(0,12,length=1000)

lines(x21,dnorm(x21,5.7,2.85),type="l",col="orange")

legend(6,3,legend=c("Posterior","Prior"),lwd=c(2,2),

col=c("black","orange"),bty="n")

plot(density(MST1[,3],adj=2),xlab=expression(tau^2),main="",lwd=2,

ylab=expression(paste(italic("p("),tau^2,"|",italic(y[1]),"...",

italic(y[11]),")")))

x31=seq(0,1.2,length=1000)

lines(x31,dinvgamma(x31,eta0\_1/2,eta0\_1\*t20\_1/2),type="l",col="orange")

legend(0.3,8,legend=c(""Posterior","Prior"),lwd=c(2,2),

col=c("black","orange"),bty="n")

title(main="Posterior distributions of Non-Schizophrenics",outer=TRUE)

#Schizophrenic posterior

par(mfrow=c(2,2),mar=c(3,3,1,1),mgp=c(1.75,.75,0),oma=c(0,0,2,0))

plot(density(MST2[,1],adj=2),lwd=2,main="",xlab=expression(sigma^2),

ylab=expression(paste(italic("p("),sigma^2,"|",italic(y[1]),"...",

italic(y[6]),")")))

x21=seq(0,1,length=1000)

lines(x21,dinvgamma(x21,nu0\_2/2,nu0\_2\*s20\_2/2),type="l",col="orange")

legend(0.18,18,legend=c("Posterior","Prior"),lwd=c(2,2),

col=c("black","orange"),bty="n")

plot(density(MST2[,2],adj=2),xlab=expression(mu),main="",lwd=2,xlim=c(0,12.5),

ylab=expression(paste(italic("p("),mu,"|",italic(y[1]),"...",

italic(y[6]),")")))

x22=seq(0,12.5,length=1000)

lines(x22,dnorm(x22,6,3),type="l",col="orange")

legend(7,1.25,legend=c("Posterior","Prior"),lwd=c(2,2),

col=c("black","orange"),bty="n")

plot(density(MST2[,3],adj=2),xlab=expression(tau^2),main="",lwd=2,

ylab=expression(paste(italic("p("),tau^2,"|",italic(y[1]),"...",

italic(y[6]),")")))

x23=seq(0,10,length=1000)

lines(x23,dinvgamma(x23,eta0\_2/2,eta0\_2\*t20\_2/2),type="l",col="orange")

legend(4,1.7,legend=c("Posterior","Prior"),lwd=c(2,2),

col=c("black","orange"),bty="n")

title(main="Posterior distributions of Schizophrenics",outer=TRUE)

#Compare non-schizophrenics and schizophrenic posterior

par(mfrow=c(2,2),mar=c(3,3,1,1),mgp=c(1.75,.75,0),oma=c(0,0,2,0))

plot(density(MST1[,1],adj=2),lwd=2,col="blue",main="",xlab=expression(sigma^2),

xlim=c(0.02,0.25),ylab=expression(paste(italic("p("),sigma^2,"|",italic(y[1]),"...",

italic(y[m]),")")))

lines(density(MST2[,1],adj=2),lwd=2,col="red")

legend(0.1,110,legend=c("Non-Schizophrenics","Schizophrenics"),lwd=c(2,2),

col=c("blue","red"),bty="n")

plot(density(MST1[,2],adj=2),lwd=2,col="blue",main="",xlab=expression(mu),

xlim=c(4,8),ylab=expression(paste(italic("p("),mu,"|",italic(y[1]),"...",

italic(y[m]),")")))

lines(density(MST2[,2],adj=2),lwd=2,col="red")

legend(6.1,2.5,legend=c("Non-Schizophrenics","Schizophrenics"),lwd=c(2,2),

col=c("blue","red"),bty="n")

plot(density(MST1[,3],adj=2),lwd=2,col="blue",main="",xlab=expression(tau^2),

xlim=c(0,10),ylab=expression(paste(italic("p("),tau^2,"|",italic(y[1]),"...",

italic(y[m]),")")))

lines(density(MST2[,3],adj=2),lwd=2,col="red")

legend(4,4.5,legend=c("Non-Schizophrenics","Schizophrenics"),lwd=c(2,2),

col=c("blue","red"),bty="n")

title(main="Posterior distributions for Non-Schizophrenics and Schizophrenics",

outer=TRUE)

# mean for all posterior

apply(MST1,2,mean)

apply(MST2,2,mean)

#probability for posterior mean of schizophrenics higher than non-schizophrenics

sum(MST1[,2]<MST2[,2])/10000

#trace plots for MCMC

par(mfrow=c(2,2),oma=c(0,0,2,0))

plot(1:S,MST1[,1],type="l",col="dark blue",xlab="Iterations",ylab=expression(sigma^2))

plot(1:S,MST1[,2],type="l",col="dark blue",xlab="Iterations",ylab=expression(mu))

plot(1:S,MST1[,3],type="l",col="dark blue",xlab="Iterations",ylab=expression(tau^2))

title(main="Trace plot of posterior for Non-Schizophrenic",outer=TRUE)

par(mfrow=c(2,2),oma=c(0,0,2,0))

plot(1:S,MST2[,1],type="l",col="dark blue",xlab="Iterations",ylab=expression(sigma^2))

plot(1:S,MST2[,2],type="l",col="dark blue",xlab="Iterations",ylab=expression(mu))

plot(1:S,MST2[,3],type="l",col="dark blue",xlab="Iterations",ylab=expression(tau^2))

title(main="Trace plots of posterior for Schizophrenic",outer=TRUE)

# function of stationarity

stationarity.plot<-function(x,...){

S<-length(x)

scan<-1:S

ng<-min( round(S/100),10)

group<-S\*ceiling( ng\*scan/S) /ng

boxplot(x~group,...)

}

# produce boxplots of sequential groups for s2, µ and t2

par(mfrow=c(1,3),mar=c(2.75,2.75,.5,.5),mgp=c(1.7,.7,0),oma=c(0,0,2,0))

stationarity.plot(MST1[,1],xlab="iteration",ylab=expression(sigma^2))

stationarity.plot(MST1[,2],xlab="iteration",ylab=expression(mu))

stationarity.plot(MST1[,3],xlab="iteration",ylab=expression(tau^2))

title(main="Group boxplots for posterior of Non-Schizophrenic",outer=TRUE)

par(mfrow=c(1,3),mar=c(2.75,2.75,.5,.5),mgp=c(1.7,.7,0),oma=c(0,0,2,0))

stationarity.plot(MST2[,1],xlab="iteration",ylab=expression(sigma^2))

stationarity.plot(MST2[,2],xlab="iteration",ylab=expression(mu))

stationarity.plot(MST2[,3],xlab="iteration",ylab=expression(tau^2))

title(main="Group boxplots for posterior of Schizophrenic",outer=TRUE)

# autocorrelation checking for each posterior of each group

par(mfrow=c(2,2),oma=c(0,0,2,0))

acf(MST1[,1],main=expression(sigma^2))

acf(MST1[,2],main=expression(mu))

acf(MST1[,3],main=expression(tau^2))

title(main="Autocorrelation for Non-Schizophrenic",outer=TRUE)

par(mfrow=c(2,2),oma=c(0,0,2,0))

acf(MST2[,1],main=expression(sigma^2))

acf(MST2[,2],main=expression(mu))

acf(MST2[,3],main=expression(tau^2))

title(main="Autocorrelation for Schizophrenic",outer=TRUE)

#acf(MST1[,1],40)$acf

#Check effetiveSize for sigma, mu and tau

install.packages("lattice")

library(coda)

effectiveSize(MST1)

effectiveSize(MST2)

#Raftery-Lewis diagnosis for convegence

raftery.diag(mcmc(MST1))

raftery.diag(mcmc(MST2))

#Geweke diagnosis for convergency

geweke.diag(MST1)

geweke.diag(MST2)

geweke.plot(mcmc(MST1))

geweke.plot(mcmc(MST2))

#Heidelberger-Welch diagnosis for convergency

heidel.diag(mcmc(MST1))

heidel.diag(mcmc(MST2))