Bios372 Final Project

Minchun Zhou

December 2, 2014

I agree to share this report within the use of course Bios372.

1 Introduction

Functional magnetic resonance imaging (fMRI) is a functional neuroimaging procedure that measures brain activity. fMRI detects the changes in blood flow associated with the functional connectivity. Typical fMRI data contain a set of discrete time series measured on three-dimensional volume elements, called voxels. There are two type of data in fMRI analysis, task induced and resting state. In this project, we will focus on task induced type of fMRI data analysis. Task induced type fMRI data are collected when subjects are asked to perform some tasks, such as tapping finger and pressing button. We want to test if some parts of brain are activated when the subject is performing the task. Resting state type fMRI data are collected when subjects are rested. We want to test if some parts of brain are functionally connected. In this project, we are interested in task induced fMRI data. We want to see whether one specfic region of interest(ROI) in brain are associated with distinguishing perceptual dimensions(i.e., shape or texture).

2 Methods

2.1 Data description and Exploratory analysis

All data in this project was from one subject. Our main interest was one ROI, which containing 20 voxels. The subject was trained to press the keypress to select the one dimension shape(D1) or two dimensions texture(D2). In the experiment, the stimulus(shape or texture) were prejected on the screen and lasted 2 seconds, with a variable inter-trial interval of 08 sec. The whole experiment lasted 128 seconds. The goal was to test if D2 effect was greater than D1 effect.

In the data, we had 20 discrete time series for 20 voxel. Each time series included 128 values corresponding to the signal value at each second. We also had three covariates value $X_1(t)$, $X_2(t)$ and $X_3(t)$ at each second, where $X_i(t), i = 1, 2, 3$ was the convolution between the ith impulse function and the hemodynamic response function (HRF). $X_1(t)$ indicated stimuli 1(D1). $X_2(t)$ indicated stimuli 2(D2). $X_3(t)$ was the instruction period. Figure 1 shows the stimulus and the mean signal at each time point. In the upper plot(D1), the black solid line was the impulse function, which shows when the stiluli was on. The red solid line was the convolution between the impulse function and HRF. The dotted black line was the mean signal. In the bottom plot(D2), the black solid line was the impulse function, which shows when the stiluli was on. The blue solid line was the convolution between the impulse function and HRF. The dotted black line was the mean signal.

2.2 Models

We proposed three models to test the hypothesis that D2 effect was greater than D1 effect. In Model 1, we combined 20 time series into one time series, which means we assumed that all voxels are the same and used its mean time series. In Model 2, we assumed that all voxels are independent of each other. In Model 3, we assumed that all voxels are correlated to each other based on their Euclidean distance.

- Model 1

$$\begin{aligned}
\bar{Y(t)} &\sim N(\mu(t), \sigma_y^2) \\
\mu(t) &= \beta_1 X_1(t) + \beta_2 X_2(t) + \beta_3 X_3(t) + w(t)
\end{aligned}$$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \sim MVN\left(\begin{pmatrix} \beta_{10} \\ \beta_{20} \\ \beta_{30} \end{pmatrix}, \Sigma\right)$$

$$\begin{aligned}
w(t) &\sim N(\alpha + \beta w(t-1), \sigma_w^2) \\
\alpha &\sim N(a, b) \\
\beta &\sim N(a, b) \\
\sigma_y &\sim IG(c, d) \\
\sigma_w &\sim IG(c, d)
\end{aligned}$$

where $a = 0$, $b = 10^{-6}$, $c = d = 0.01$, $\beta_{10} = \beta_{20} = \beta_{30} = 0$, $\Sigma^{-1} = 0$.

In Model 1, we combined 20 time series into one mean time series. We assumed that the signal value $Y(t)$ at time t is autoregressive model with order one (AR(1)). $Y(t)$ is also normally distributed with mean $\mu(t)$ and variance σ_y^2. The AR(1) model is modeled on w(t) in the mean component $\mu(t)$. $w(t)$ is normally distributed with mean $\alpha + \beta w(t-1)$ and variance σ_w^2. $\mu(t)$ is the summation of the stimulus and $w(t)$. Our main hypothesis is $H_0 : \beta_2 > \beta_1$ or $H_0 : \beta_2 - \beta_1 > 0$.

The hyperparameter α, β, σ_y and σ_w are modeled using non-informative prior, where α and β are normally distributed with mean 0 and variance 10^{-6}, σ_y and σ_w have inverse-Gamma distribution. Using non-informative prior on hyperparameter can reduce the influence of choosing prior on the final analysis.

- Model 2

$$
\begin{aligned}
Y_v(t) &\sim N(\mu_v(t), \sigma_y^2) \\
\mu_v(t) &= \beta_1 X_1(t) + \beta_2 X_2(t) + \beta_3 X_3(t) + w_v(t)
\end{aligned}
$$

$$
\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \sim MVN\left(\begin{pmatrix} \beta_{10} \\ \beta_{20} \\ \beta_{30} \end{pmatrix}, \Sigma \right)
$$

$$
\begin{aligned}
w_v(t) &\sim N(\alpha + \beta w_v(t-1), \sigma_w^2) \\
\alpha &\sim N(a, b) \\
\beta &\sim N(a, b) \\
\sigma_y &\sim IG(c, d) \\
\sigma_w &\sim IG(c, d)
\end{aligned}
$$

where $a = 0$, $b = 10^{-6}$, $c = d = 0.01$, $\beta_{10} = \beta_{20} = \beta_{30} = 0$, $\Sigma^{-1} = 0$.

In Model 2, we treated 20 voxel independently. We assumed that the signal value $Y_v(t)$ at voxel v at time t is normally distributed with voxel specified mean $\mu_v(t)$ and common variance σ_y^2, where$\mu_v(t)$ is the sumation of the stimulus and voxel specified $w_v(t)$. $w_v(t)$ is normally distributed with voxel specified mean $\alpha + \beta w_v(t-1)$ and common variance σ_w^2. The hyperparameter α, β, σ_y and σ_w are modeled as in Model 1.

- Model 3

$$
\begin{aligned}
\underline{Y(t)} &\sim MVN(\underline{\mu(t)}, R) \\
\underline{\mu(t)} &= c(\mu_1(t), \mu_2(t), ..., \mu_{20}(t)) \\
\mu_v(t) &= \beta_1 X_1(t) + \beta_2 X_2(t) + \beta_3 X_3(t) + w_v(t)
\end{aligned}
$$

$$
\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \sim MVN\left(\begin{pmatrix} \beta_{10} \\ \beta_{20} \\ \beta_{30} \end{pmatrix}, \Sigma \right)
$$

$$
\begin{aligned}
w_v(t) &\sim N(\alpha + \beta w_v(t-1), \sigma_w^2) \\
\alpha &\sim N(a, b) \\
\beta &\sim N(a, b) \\
R^{-1} &\sim Wishart((\rho\Omega)^{-1}, \rho) \\
\sigma_w &\sim IG(c, d)
\end{aligned}
$$

where $a = 0$, $b = 10^{-6}$, $c = d = 0.01$, $\beta_{10} = \beta_{20} = \beta_{30} = 0$, $\Sigma^{-1} = 0$, $\rho = 20$.

In Model 3, we assumed that 20 voxels were correlated to each other based on their Euclidean distance. We also assumed that the time series $\underline{Y(t)}$ across 20 voxel at time t was multivariate normally distributed with mean vector $\underline{\mu(t)}$ and covariance matrix R, where $\underline{\mu(t)}$ was a vector consist of voxel specified mean $\mu_v(t)$ as in Model 2. The covariance matrix R had inverse-Wishart distribution, where the scale matrix parameter Ω was a distance matrix. $\Omega_{i,j}$ was the Euclidean distance between voxel i and voxel j. The hyperparameter α, β and σ_w were modeled as in Model 1.

- Sampling Method

  All three models were analyzed using Gibbs sampling by OpenBUGS. Three MCMC chains were performed for each model.

  The initial valuesfor Model 1 were:
  Chain 1: $\beta_\mu = c(0, 0, 0), \sigma_y^2 = 1, \alpha = 1.0, \beta = 0.95, \sigma_w^2 = 1.0$
  Chain 2: $\beta_\mu = c(10, -10, 10), \sigma_y^2 = 1/3, \alpha = 2.0, \beta = 0, \sigma_w^2 = 1/3$
  Chain 3: $\beta_\mu = c(-10, 10, -10), \sigma_y^2 = 2, \alpha = -1.0, \beta = -0.95, \sigma_w^2 = 1/2$

  The initial valuesfor Model 2 were:
  Chain 1: $\beta_\mu = c(0, 0, 0), \sigma_y^2 = 1, \alpha = 1.0, \beta = 0.95, \sigma_w^2 = 1.0$
  Chain 2: $\beta_\mu = c(-10, -10, -10), \sigma_y^2 = 1/3, \alpha = -1.0, \beta = 0, \sigma_w^2 = 1/2$
  Chain 3: $\beta_\mu = c(10, 10, 10), \sigma_y^2 = 2, \alpha = 0, \beta = 0, \sigma_w^2 = 1/3$

  The initial valuesfor Model 3 were:
  Chain 1: $\beta_\mu = c(0, 0, 0), \sigma_y^2 = 1, \alpha = 1.0, \beta = 0.95, \sigma_w^2 = 1.0, R = diag(rep(1, 20))$
  Chain 2: $\beta_\mu = c(10, 10, 10), \sigma_y^2 = 1/3, \alpha = -1.0, \beta = -0.95, \sigma_w^2 = 1/2, R = diag(rep(0.1, 20))$
  Chain 3: $\beta_\mu = c(-10, -10, -10), \sigma_y^2 = 2, \alpha = 0, \beta = 0, \sigma_w^2 = 1/3, R = diag(rep(10, 20))$

- Model Checking and Sensitivity analysis for the prior distribution

  We chose Unif(-1000,1000) as an alternative non-informative prior for the hyperparameter α, β, σ_y and σ_w to perform the sensitivity analysis. For model checking, there were no good creteria in time series analysis using bayesian approach. We chose the last observation Y(128) as the test statistics in Model 1 and last observation in the first voxel $Y_1(128)$ in Model 2 and Model 3 to perform the model checking.

3 Results

Summary statistics for β_1, β_2 and $\beta_2 - \beta_1$ were listed in Table 1 for Model 1, Table 2 for Model 2 and Table 3 for Model 3. In Model 1, the estimate of β_1 was -0.159(95% CI [-0.544, 0.226]). The estimate of β_2 was 0.383(95% CI [-0.024, 0.771]). The difference between β_2 and β_1 was 0.542(95% CI [0.184, 0.893]). In Model 2, the estimate of β_1 was -0.014(95% CI [-0.243, 0.174]). The estimate of β_2 was 0.346(95% CI [0.133, 0.546]).The difference between β_2 and β_1 was 0.361(95% CI [0.147, 0.567]). In Model 3, the estimate of β_1 was -15.359(95% CI [-2043.525, 1912.625]). The estimate of β_2

was 0.064(95% CI [-0.021, 0.181]). The difference between β_2 and β_1 was 15.423(95% CI [-1912.594, 2043.588]). The DIC for Model 1, 2, 3 were 186.2, 7046 and -2.097×10^4 respectively. If we just looked at DIC, then Model 1 might be the best. We still wanted to choose Model 2 or Model 3 because Model 1 ignored information and correlation between voxels.

The posterior distribution of $\beta_1, \beta_2, \beta_3$ and $\beta_2 - \beta_1$ were plotted in Figure 2 for Model 1, Figure 3 for Model 2 and Figure 4 for Model 3. All posterior distributions looked like normally distributed. Figure 5, Figure 6 and Figure 7 were the trace plots for three models. In Model 1 and 2, β_1, β_2 and β_3 all converged. In Model 3, β_2 and β_3 converged but β_1 did not.

Sensitivity analysis for the prior distributions were only perfromed on Model 1 and Model 2 since β_1 and β_2 were both converged in these two models. Summary statistics for β_1, β_2 and $\beta_2 - \beta_1$ by using new prior for α, β, σ_y and σ_w were listed in Table 4 for Model 1, Table 5 for Model 2. The posterior distribution of $\beta_1, \beta_2, \beta_3$ and $\beta_2 - \beta_1$ were plotted in Figure 8 for Model 1, Figure 9 for Model 2. Figure 10 and Figure 11 were the trace plots for Model 1 and Model 2. With new prior Unif(-1000,1000), Model 1 did not converge but Model 2 did. The result from Model 2 with new prior in Table 5 were similar to the result in original analysis Table 2.

Model checking were only perfromed on Model 1 and Model 2 since β_1 and β_2 were both converged in these two models. The p-value for model 1 was 0.475. The p-value for model 2 was 0.223. Both p-value indicateed that the model was appropriate for the test statistics we chose.

4 Discussion

In this project, we proposed three models to analyze the task induced fMRI data. In the first model, we simplified the data using mean signals across 20 voxels. In the second model, we ignored the spatial correlation between voxels. In the third model, we assumed that voxels were correlated based on their Euclidean distance. We found that the ROI was activated based on the contrast($D2 > D1$) in Model 1 and Model 2. In Model 3, β_1 did not converge and the activation was not found.

Kang et al(2012) proposed a frequentist method to model the spatial-temporal correlation in the fMRI data. They transformed the time series into frequency domain and chose two frequency band, one in high frequency and one in low frequency. Their inference were made on the frequency band. By using thier method, the estimate was $\hat{\beta_1} = -0.8233, \hat{\beta_2} = -2.9478, \hat{\beta_2} - \hat{\beta_1} = -2.1245, sd_{\hat{\beta_2} - \hat{\beta_1}} = 8.3488$ p-value was 0.6. We found the estimates from our models in Table 1, Table 2 and Table 3 were higher than the estimate from their method. One reason could be that we did not consider the correlation between voxels in Model 1 and Model 2, which overstimated the effects of stimulus. In Model 2, when we added more information using all voxels, the estimates of β_1 and β_2 were shrunk towards to zero compwere to Model 1. Although β_1 was not converged in Model 3, mean $\beta_2 = 0.064$ was even closer to zero than β_2 in Model 1 and Model We expected to see similar results as in their method if β_1 in Model 3 converged.

5 References

1. Spatio-Spectral Mixed-Effects Model for Functional Magnetic Resonance Imaging Data, Hakmook Kang, Hernando Ombao, Crystal Linkletter, Nicole Long d, David Badre, JASA, 2012, 568-577.

2. Bayesian Biostatistics, Emmanuel Lesaffre, Andrew B. Lawson, 2012.

3. Bayesian Data analysis, Second Edtion, Andrew Gelman, John Carlin, et.all 2003.

4. Bayesian Methon for Data Analysis, Third Edition, Brandley Carlin, Thomas Louis, 2009.

6 Appendix

6.1 Tables

	mean	sd	2.5%	25%	50%	75%	97.5%
mu.beta[1]	-0.159	0.197	-0.544	-0.289	-0.156	-0.025	0.226
mu.beta[2]	0.383	0.203	-0.024	0.250	0.383	0.521	0.771
diff.beta	0.542	0.182	0.184	0.419	0.541	0.663	0.893

Table 1: Summary statistics for Model 1

	mean	sd	2.5%	25%	50%	75%	97.5%
mu.beta[1]	-0.014	0.107	-0.243	-0.081	-0.008	0.060	0.174
mu.beta[2]	0.346	0.109	0.133	0.270	0.351	0.426	0.546
diff.beta	0.361	0.107	0.147	0.289	0.360	0.434	0.567

Table 2: Summary statistics for Model 2

	mean	sd	2.5%	25%	50%	75%	97.5%
mu.beta[1]	-15.359	1003.868	-2043.525	-678.525	0.056	672.925	1912.625
mu.beta[2]	0.064	0.050	-0.021	0.031	0.060	0.092	0.181
diff.beta	15.423	1003.868	-1912.594	-672.878	-0.002	678.595	2043.588

Table 3: Summary statistics for Model 3

	mean	sd	2.5%	25%	50%	75%	97.5%
mu.beta[1]	0.923	3.380	-3.629	-1.774	-0.157	5.238	5.759
mu.beta[2]	1.529	3.459	-3.063	-1.303	0.458	5.773	6.549
diff.beta	0.606	0.384	-0.009	0.344	0.542	0.820	1.401

Table 4: Summary statistics for Model 1 with New prior

	mean	sd	2.5%	25%	50%	75%	97.5%
mu.beta[1]	-0.038	0.092	-0.219	-0.100	-0.036	0.027	0.139
mu.beta[2]	0.326	0.089	0.152	0.266	0.327	0.386	0.504
diff.beta	0.364	0.104	0.153	0.295	0.366	0.434	0.562

Table 5: Summary statistics for Model 2

6.2 Figures

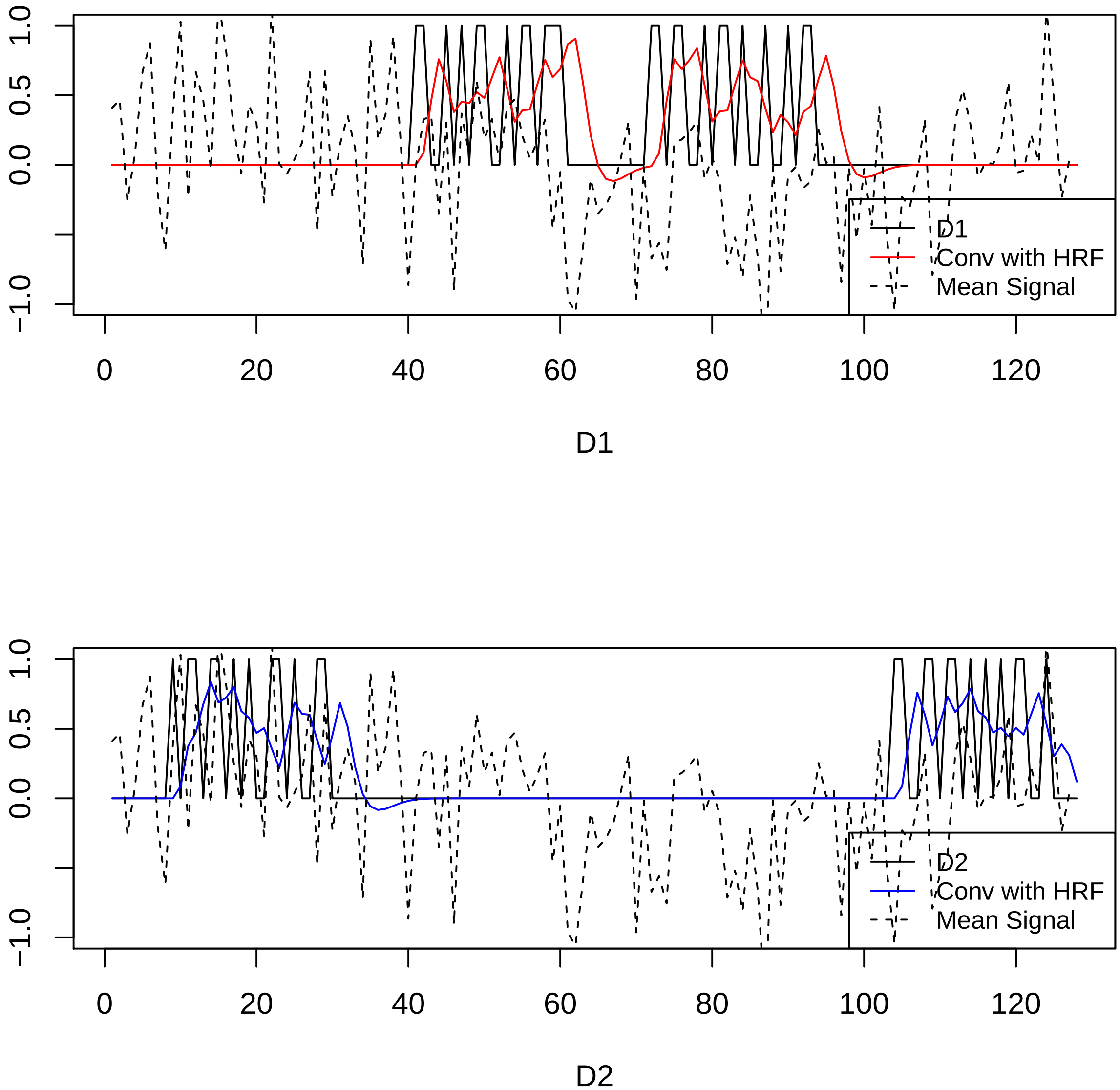

Figure 1: Stimulus D1 and D2

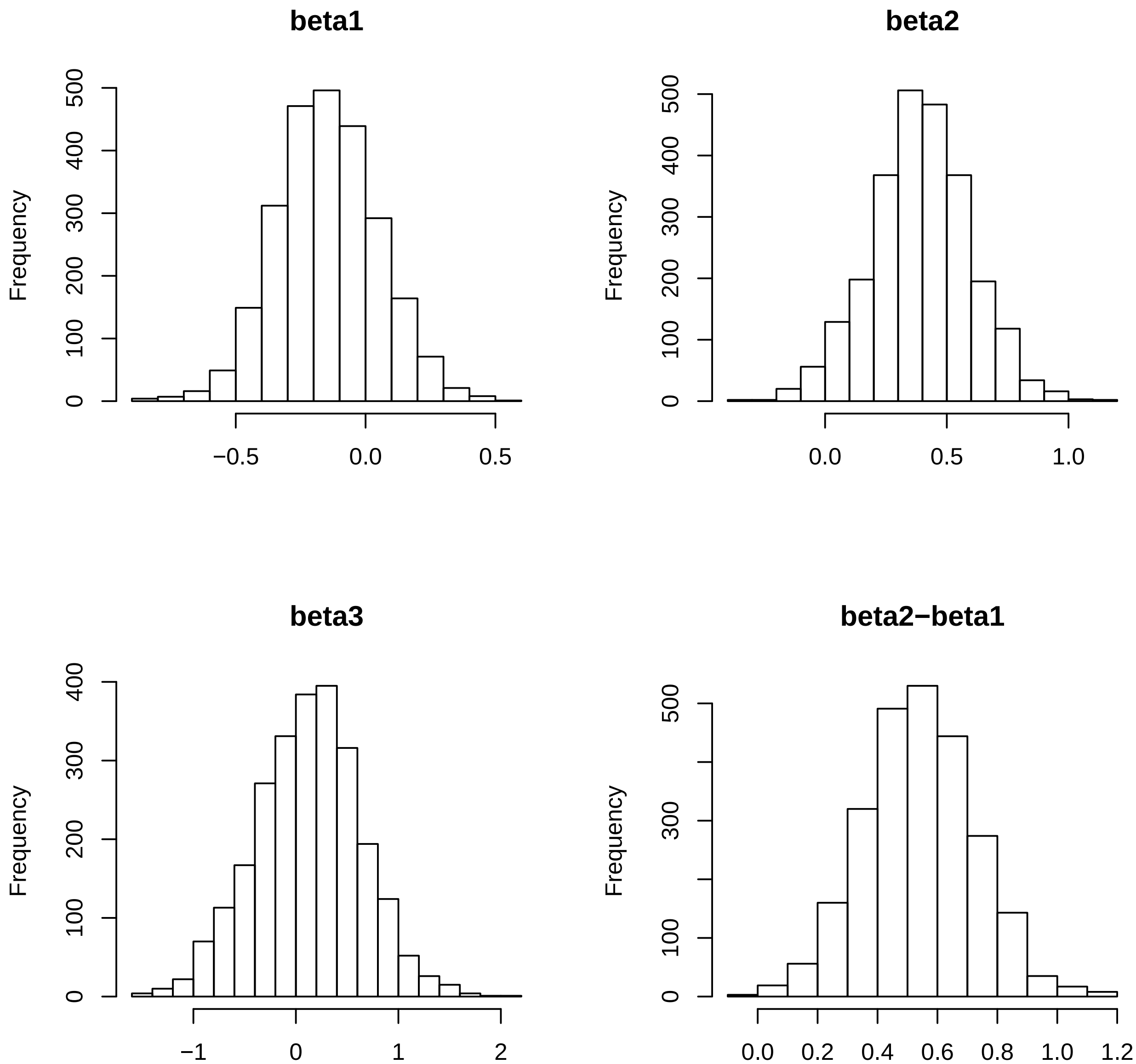

Figure 2: Model 1 Posterior distribution

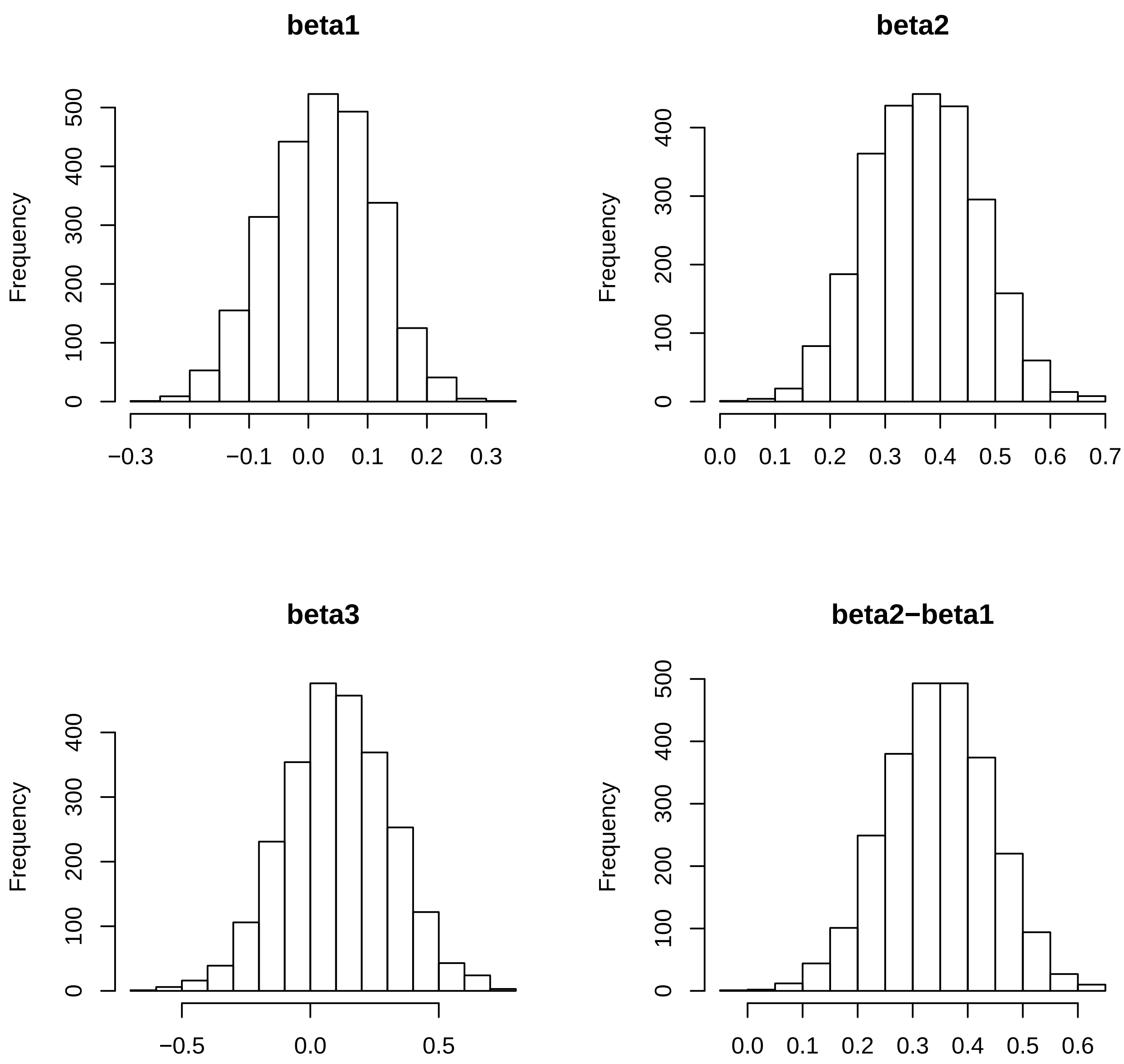

Figure 3: Model 2 Posterior distribution

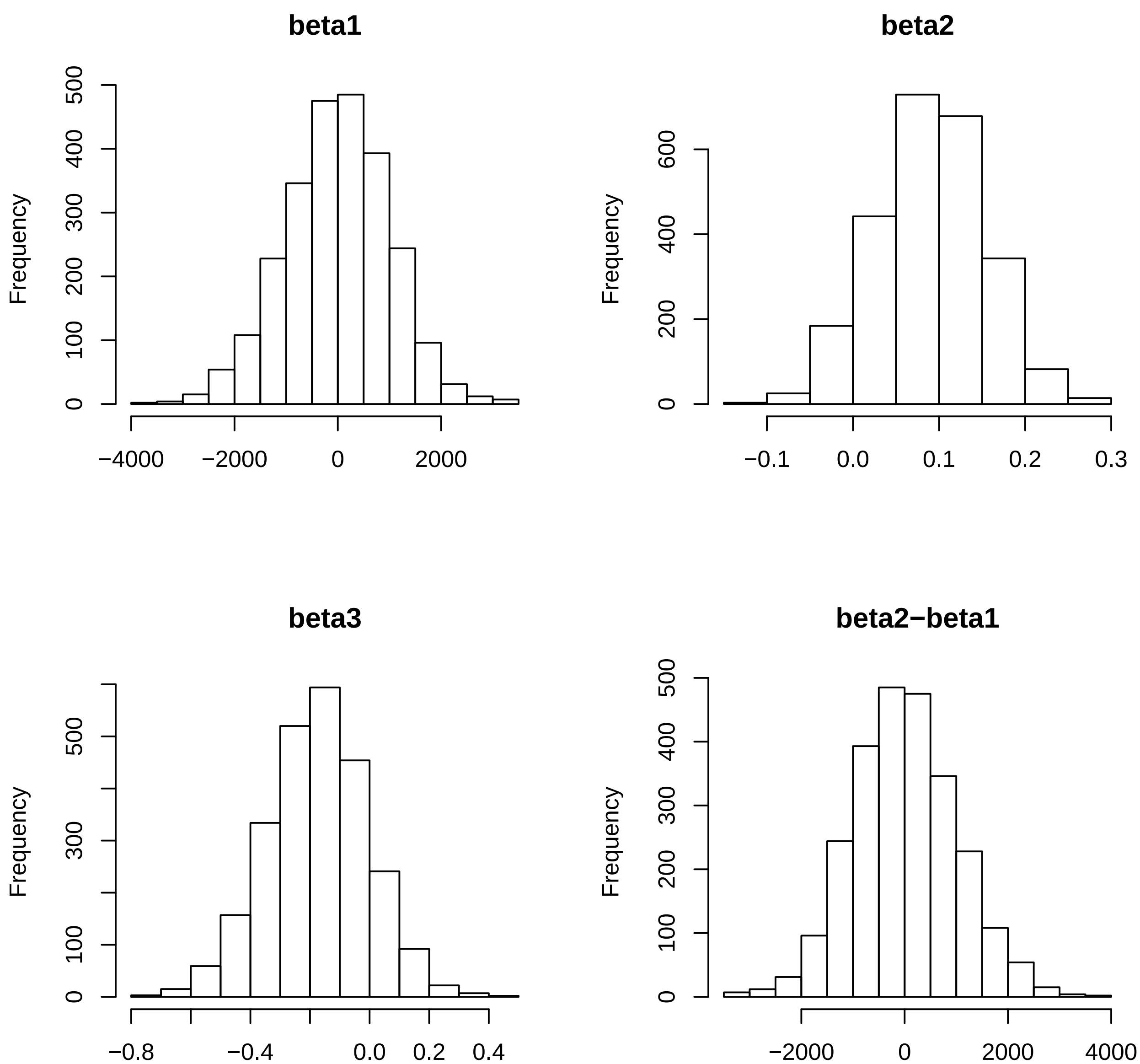

Figure 4: Model 3 Posterior distribution

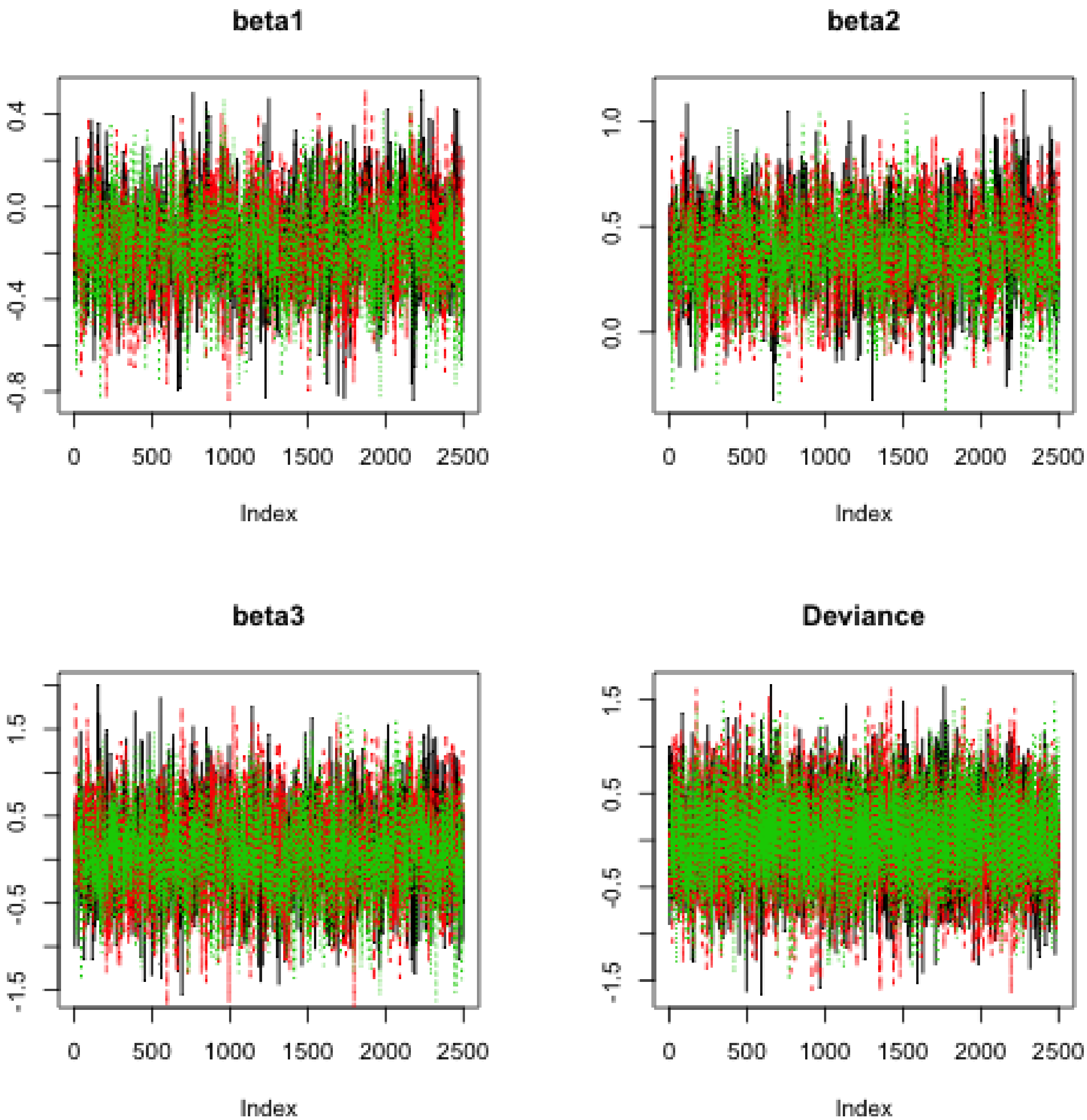

Figure 5: Model 1 Convergence

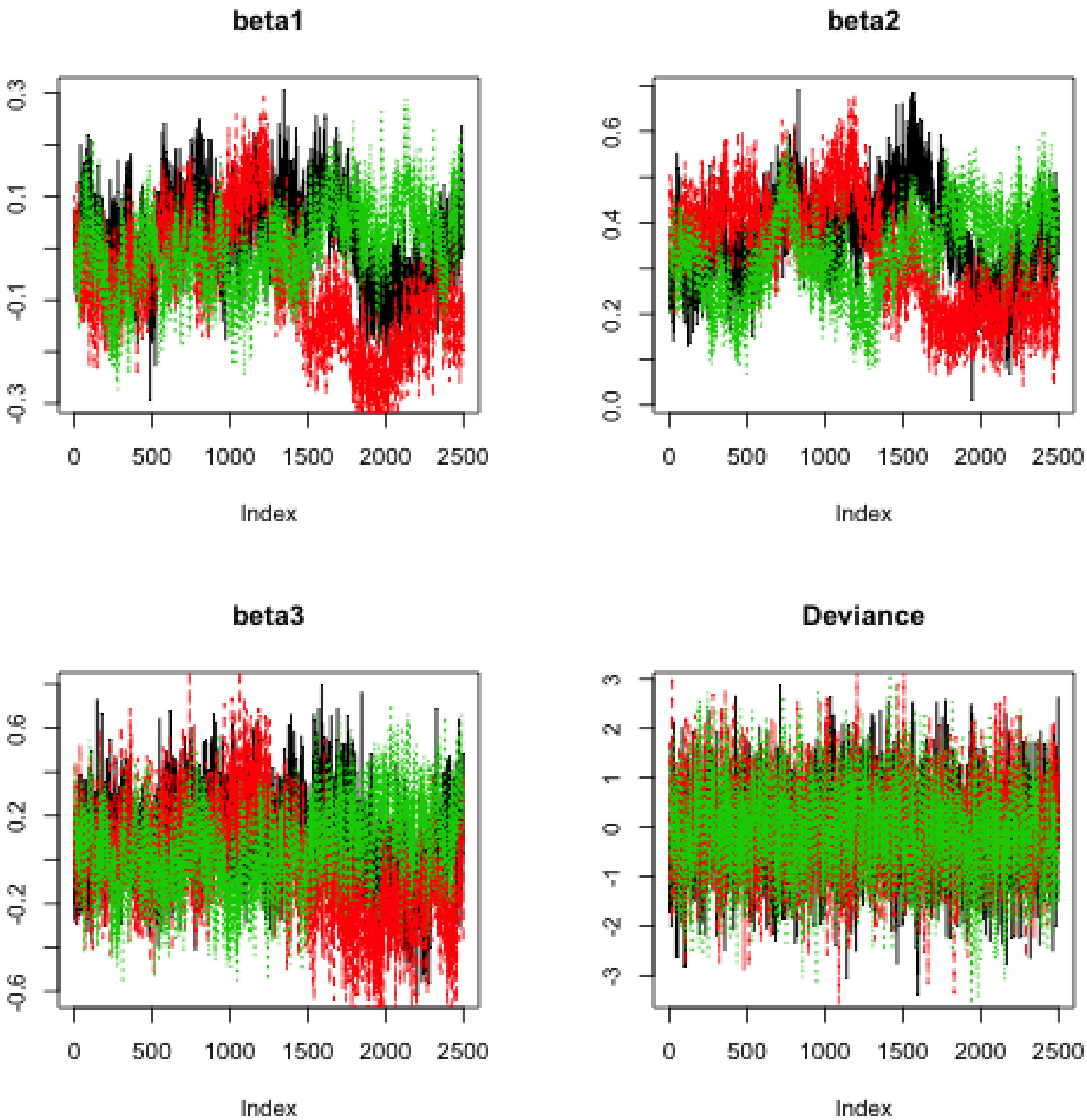

Figure 6: Model 2 Convergence

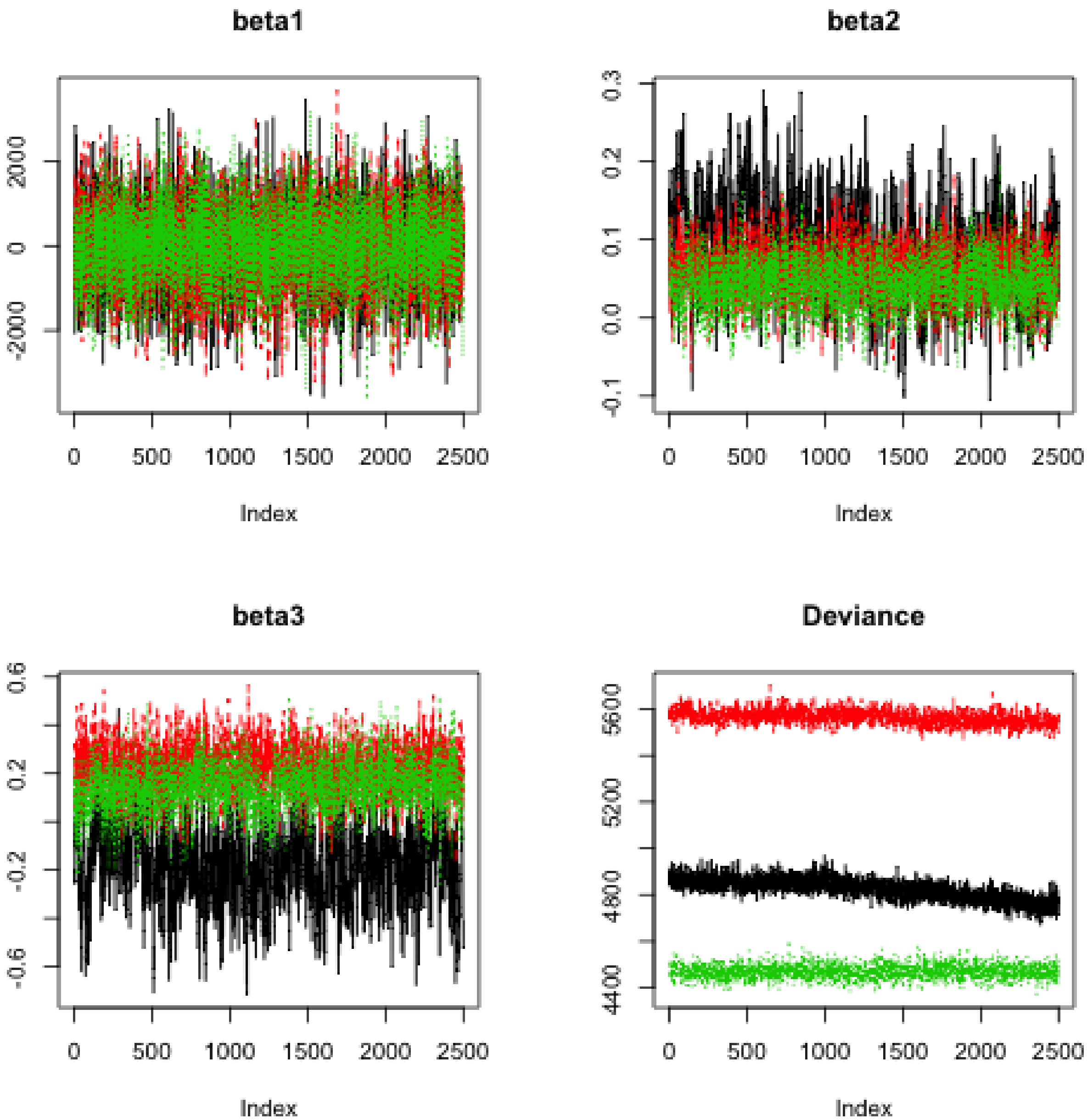

Figure 7: Model 3 Convergence

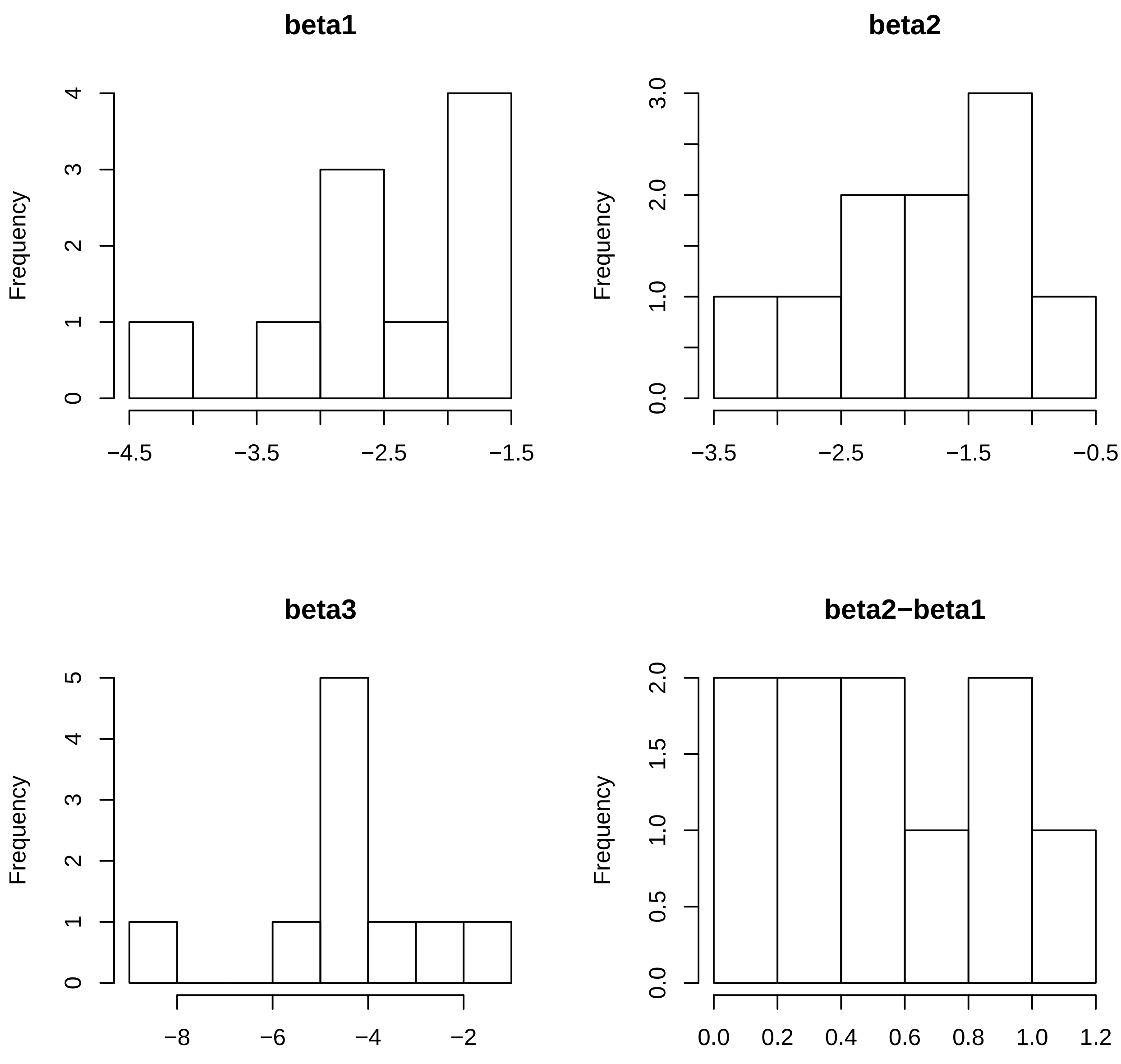

Figure 8: Model 1 Posterior distribution New prior

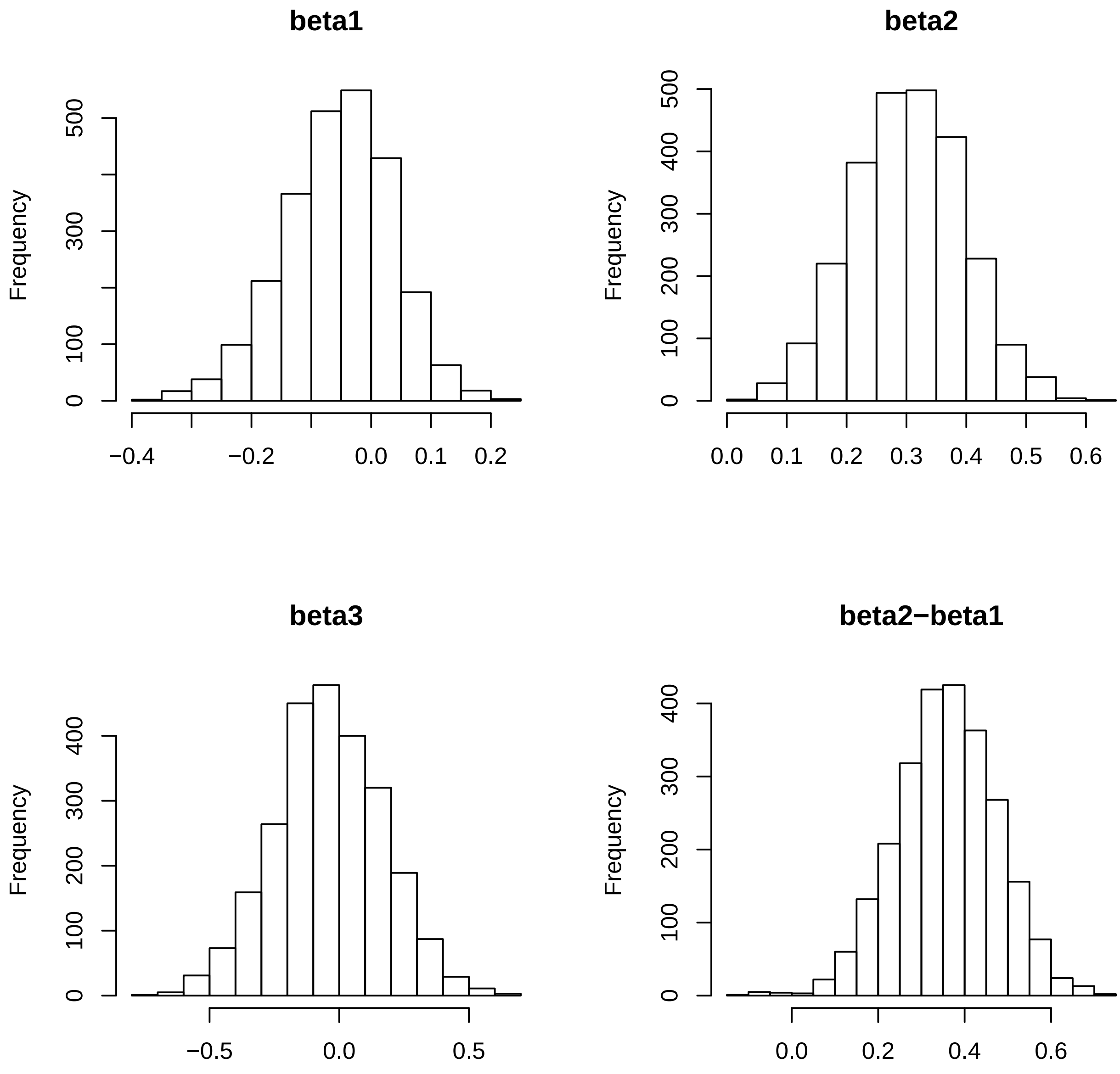

Figure 9: Model 2 Posterior distribution New prior

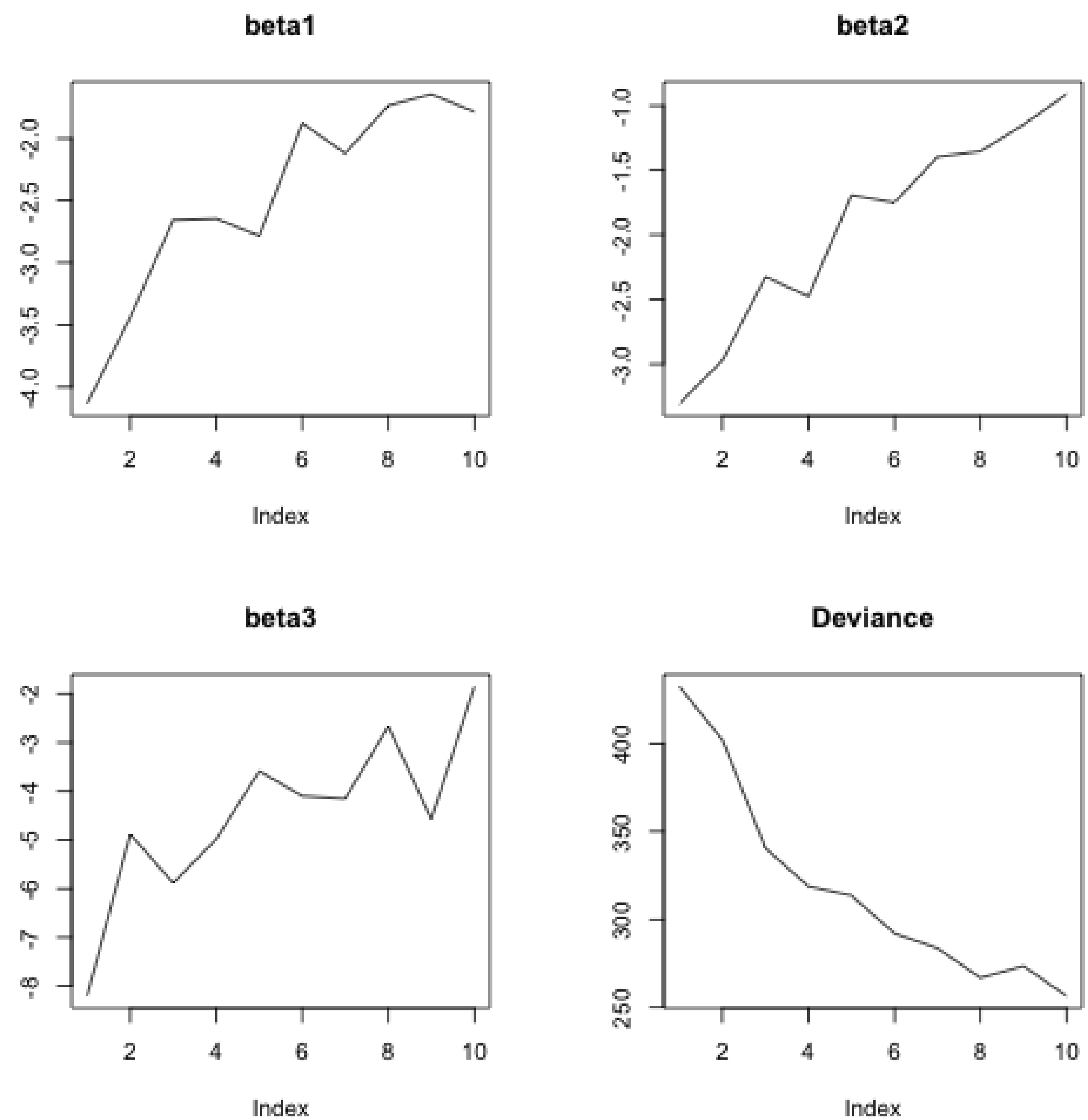

Figure 10: Model 1 Convergence New prior

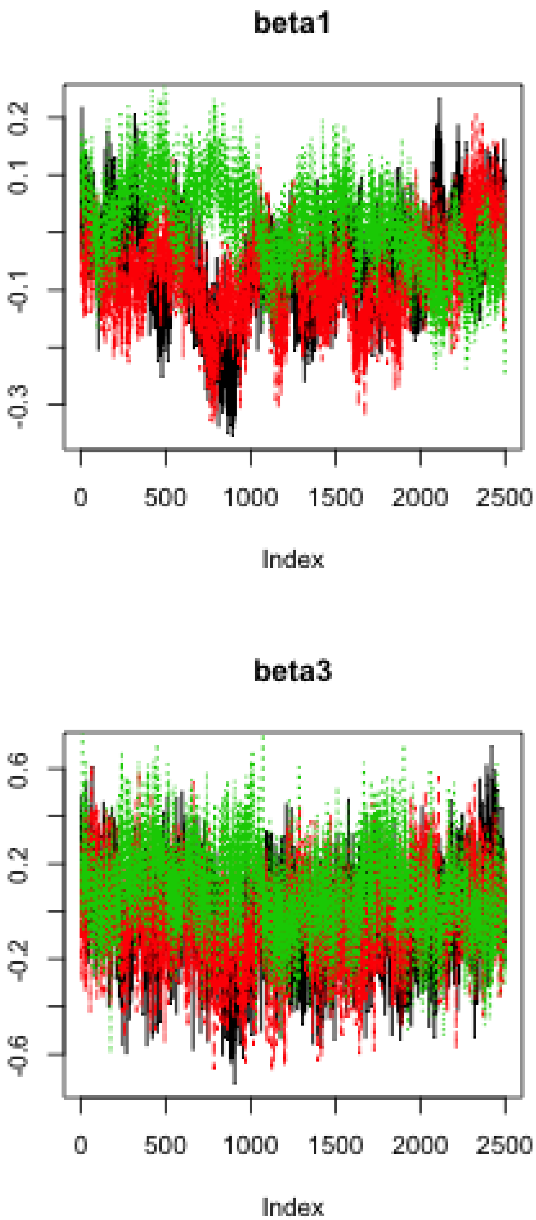

Figure 11: Model 2 Convergence New prior