Hierarchical Modeling Using Gibbs Sampling

Vadim von Brzeski, Vesselin Diev
AMS 207

1. Introduction

In this project we reproduce the results from Gelfand et. al [1], in which they show the ease and effectiveness of Gibbs sampling applied to a variety of complex inference problems. Specifically, we focus on the hierarchical model example given in [1], and we achieve nearly identical results to those in [1]. We also perform some more predictive model checking and monitoring for convergence using BOA.

2. Hierarchical Model of Rat Growth

We illustrate a full implementation of Bayesian hierarchical modeling using Gibbs sampling, and demonstrate the ease of implementation and accuracy of results. We focus on a population growth problem where the weights of 30 young rats were measured at the end of each week for 5 weeks, first in a control group, and then in a treatment group. Initially, we deal with the control group to illustrate the model and the Gibbs sampling methodology used. The model is quite complex, with 66 total number of parameters.

2.1 Model

A short inspection of the data shows that it is reasonable to assume individual straight-line growth curves. We also assume homoscedastic (i.e. having the same variance) normal measurement error. The full measurement model for the group is then:

$$Y_{ij} \sim N(\alpha_i + \beta_i X_{ij}, \sigma_c^2), \quad i = 1,...,k; \quad j = 1,...,n_i,$$

where: Y_{ij} denotes the weight of i^{th} rat at the end of week j, X_{ij} is age in days of the i^{th} rat at the end of week j, $k = 30$, and $n_i = 5$. Each rat has its own individual growth parameters (α_i, β_i), where: α_i is the initial weight of rat i, and β_i is its growth rate. The population structure is then modeled as:

$$\begin{pmatrix} \alpha_c \\ \beta_c \end{pmatrix} \sim N\left\{ \begin{pmatrix} \alpha_c \\ \beta_c \end{pmatrix}, \sum_c \right\}, \quad i = 1,...,30$$

A full Bayesian analysis requires specification of priors for

$$\sigma_c^2, \mu_c = (\alpha_c, \beta_c^{\ T}), \Sigma_c$$

Assuming independence of the priors, a standard specification is the following Normal-Wishart-InverseGamma form:

$$\mu_g \sim N(\eta, C),$$
$$\Sigma_g^{-1} \sim W((\rho R)^{-1}, \rho),$$
$$\sigma_g^2 \sim \Gamma^{-1}\left(\frac{\nu_o}{2}, \frac{\nu_o \tau_o^2}{2}\right)$$

The measurement model for the i^{th} individual can be rewritten as $Y_i \sim N(X_i \theta_i, \sigma_c^2 I_{n_i})$, where $\theta_i = (\alpha_i, \beta_i)^T$ and X_i denotes the corresponding design matrix.

In order to implement the Gibbs sampler, we need the full conditionals for $\theta_i, \mu_c, \Sigma_c, \sigma_c$. Due to space limitations, we will only show how the full conditionals for θ_i are derived, and will just give the final forms of the rest. The full joint posterior is:

$$p(\theta_i, Y_i, \mu_c, \Sigma_c^{-1}, \sigma_c^2) \propto p(Y_i \mid \theta_i, \mu_c, \Sigma_c^{-1}, \sigma_c^2).p(\theta_i \mid \mu_c, \Sigma_c^{-1}, \sigma_c^2).p(\mu_c, \Sigma_c^{-1}, \sigma_c^2)$$
$$= N(Y_i \mid X_i \theta_i, \sigma_c^2 I_{n_i}).N(\theta_i \mid \mu_c, \Sigma_c).N(\mu_c \mid \eta, C).W(\Sigma_c^{-1} \mid (\rho R)^{-1}, \rho).\Gamma^{-1}(\sigma_c^2 \mid \frac{\nu_0^2}{2}, \frac{\nu_0^2 \tau_0^2}{2})$$

The full conditional for θ_i is then:

$$p(\theta_i \mid Y_i, \mu_c, \Sigma_c^{-1}, \sigma_c^2) \propto N(Y_i \mid X_i\theta_i, \sigma_c^2 I_{n_i}).N(\theta_i \mid \mu_c, \Sigma_c)$$

$$= \exp(-\frac{1}{2}(Y_i - X_i\theta_i)^T(\sigma_c^{-2} I_{n_i})(Y_i - X_i\theta_i)).\exp(-\frac{1}{2}(\theta_i - \mu_c)^T \Sigma_c^{-1}(\theta_i - \mu_c))$$

$$= \exp\left[-\frac{1}{2}\left((Y_i - X_i\theta_i)^T(\sigma_c^{-2} I_{n_i})(Y_i - X_i\theta_i) + (\theta_i - \mu_c)^T \Sigma_c^{-1}(\theta_i - \mu_c)\right)\right]$$

To get a normal kernel, we need something of the form: $\exp\left[-\frac{1}{2}(\theta_i - A)^T B^{-1}(\theta_i - A)\right]$

Expanding the above expression for the full conditional of θ_i, grouping terms together, we get for *A* and *B*:

$$A = (\sigma_c^{-2} X_i^T X_i + \Sigma_c^{-1})^{-1}(\sigma_c^{-2} X_i^T Y_i + \Sigma_c^{-1}\mu_c) \qquad B = (\sigma_c^{-2} X_i^T X_i + \Sigma_c^{-1})^{-1}$$

The full conditional for θ_i is then

$\theta_i \mid Y_i, \mu_c, \Sigma_c^{-1}, \sigma_c^2 = N(A, B)$, for *i = 1,…,k.*

In the same manner the rest of the full conditionals are derived.

$$\mu_c \mid Y, \{\theta\}, \Sigma_c^{-1}, \sigma_c^2 = N\{V(k\Sigma_c^{-1}\bar{\theta} + C^{-1}\eta), V\}, \text{ where } V = (k\Sigma_c^{-1} + C^{-1})^{-1} \text{ and } \bar{\theta} = k^{-1}\sum_{i=1}^{k}\theta_i$$

$$\Sigma_c^{-1} \mid Y, \{\theta\}, \mu_c, \sigma_c^2 = W\left\{\left[\sum_i (\theta_i - \mu_c)(\theta_i - \mu_c)^T + \rho R\right]^{-1}, k + \rho\right\}$$

$$\sigma_c^2 \mid Y, \{\theta\}, \mu_c, \Sigma_c^{-1} = \Gamma^{-1}\left(\frac{n + \nu_0}{2}, \frac{1}{2}\left[\sum_i (Y_i - X_i\theta_i)^T(Y_i - X_i\theta_i) + \nu_0\tau_0^2\right]\right), \text{ where } n = \sum_{i=1}^{k} n_i$$

Uninformative prior information relative to the one provided by the data is given by the hyperparameters :

$$C^{-1} = 0, \quad \nu_0 = 0, \quad \rho = 2, \quad R = \begin{pmatrix} 100 & 0 \\ 0 & 0.1 \end{pmatrix}$$

2.2 Convergence

The convergence of the simulations was monitored using the BOA package in R, specifically using the Geweke Z-score analysis. Geweke scores are based on the concept that if a z-score for a parameter is much bigger than 2 in absolute value, then the mean level of the time series is still drifting. In the interest of space, we only show the Z-score graphs for 2 of the β parameters of rat 1 and rat 10 in figure 1 below.

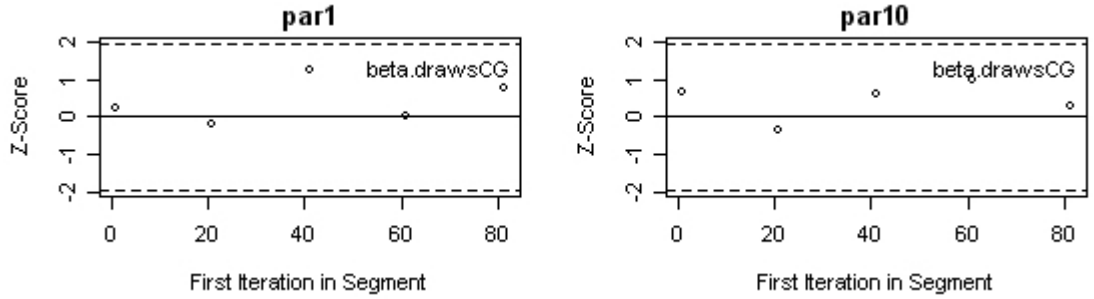

Figure 1. Geweke Z-Scores for beta parameter of rat 1 and rat 10.

2.3 Marginal Posterior Densities for the Initial Weight and Growth Rate of the Control Group

Using all 150 observations, we produce simulation estimates for the marginal posterior densities for the initial weight of rats in the control group, α , and for the growth rate of rats in the control group, β. These are shown in figure 2 below. The estimated mean and SD of α are 107 and 5.42, and the mean and SD of β are 6.17 and 0.22, respectively.

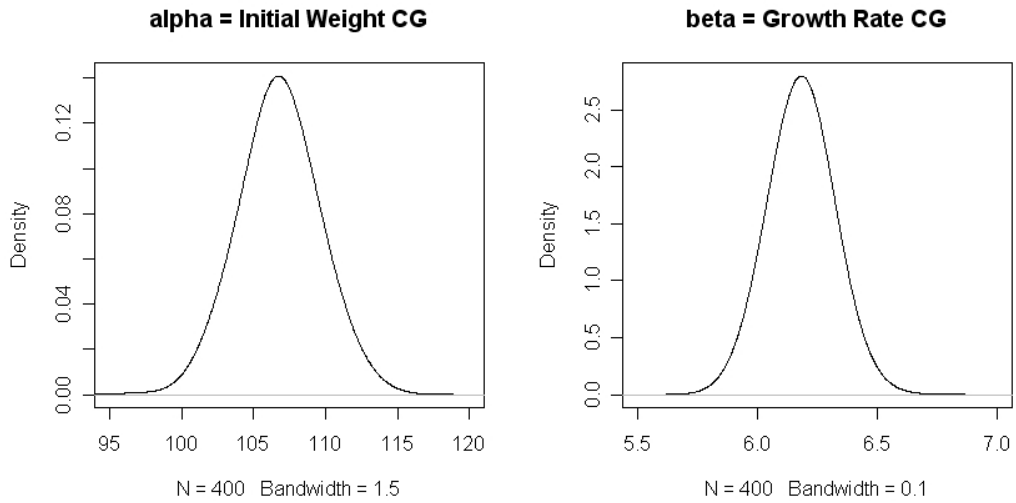

Figure 2. Marginal Posterior Density Estimates for Initial Weight and Growth Rate (150 observations).

2.4 Missing Values

To get a sense of how well this model performs, we decided to remove large chunks of the data, and see how the estimated posterior densities compared to their counterparts when using all 150 observations. We repeated the simulation with a 90 observation data set, which consisted of removing the final data point from rats 6-10, the final 2 from 11-20, the final three from 21-25, and the final four from 26-30. The 75 observation set was obtained from the set of 90 by removing all but one observation from rats 16-30. The plots of the marginal posterior densities for β in both scenarios are shown in figures 3 and 4 below. Comparing to the density for β in figure 2, one can see that the densities with missing values are not as tight, which is expected. Nevertheless, they match quite well, indicating the robustness of the model.

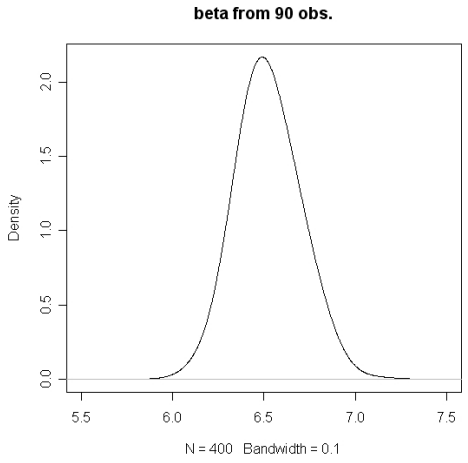

Figure 3. Marginal Posterior beta (90 obs).

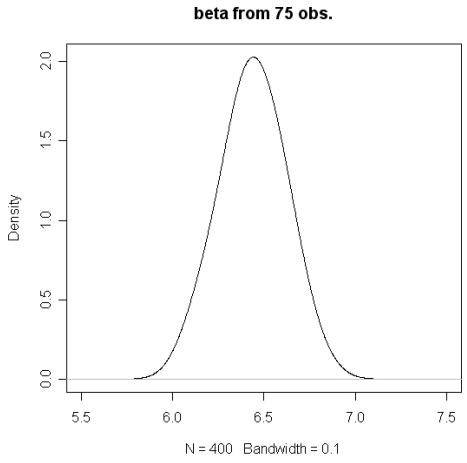

Figure 4. Marginal Posterior beta (75 obs).

2.5 Data Replicates

Following the example in [1], we ran a predictive model check on one of the rats in the control group, rat #26. The goal is to compare how the replicated data from the model compares to the actual observations. Figure 5 shows the results of this experiment. For each week $X_{26,j}$, we draw 1000 samples from the predictive distribution $Y_{26,j}$ based on the posterior means of α_{26} and β_{26}. We plot the 95% posterior predictive interval as two boundary lines, and the actual observed value as a black dot. As one can see, the actual values always lies in the 95% interval, once again indicating the stability of the model in replicating the data.

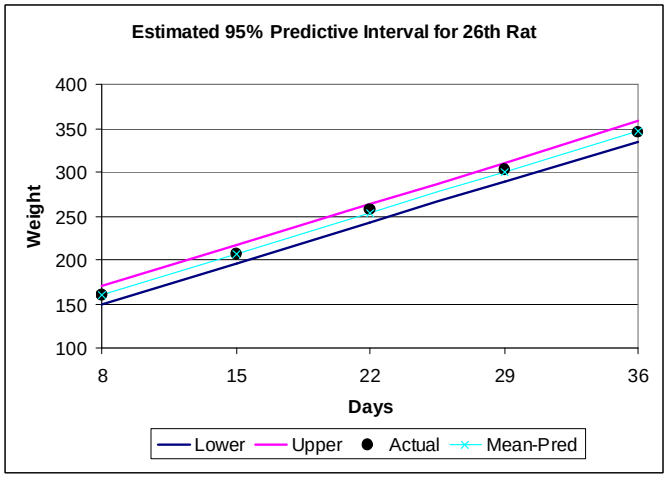

Figure 5. Estimated 95% Predictive Intervals for 26th Rat in Control Group Based on Data Replicates. Actual observations shown as black dots.

2.6 Data Prediction

To further test our model, we perform the following experiment to see how the model extrapolates to new unseen data. We build a posterior model based on all observations from weeks 1-4 only, and use it to predict the weights of all 30 rats at week 5. Figure 6 shows the results in the form of a histogram of the actual values at week 5 compared to the means of the predicted values at week 5. The mean difference between the 30 actual and predicted values is -10.8, with a standard deviation of 8.3, which is quite good given the scale of the data. It would seem that the model tends to over-estimate the future weights.

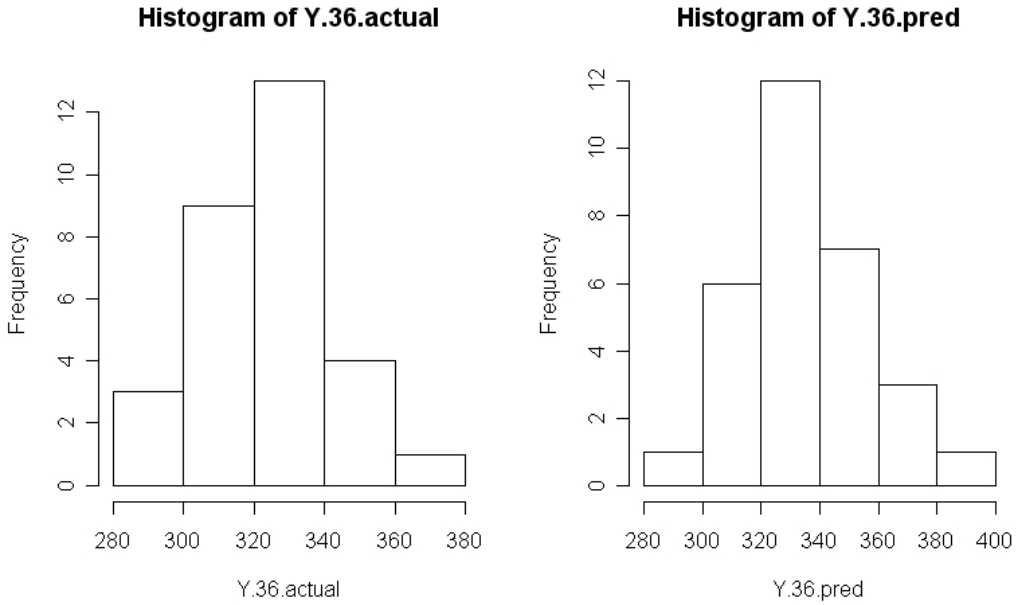

Figure 6. Predicting Weights at Week 5 Based on Model From Weeks 1-4.

2.7 Treatment Effect

Finally we compare the control group to a treatment group to determine the effect of a supposed treatment on the growth rate of the rats (possibly a dietary drug). Figure 7 below shows the estimated density of the difference of growth rates between the treatment group and the control group, as well as the ratio of the density of variance of the growth rates. One can clearly see that the treatment definitely retards the growth rate; the mean of the difference density is -1.32 with an SD of 0.15.

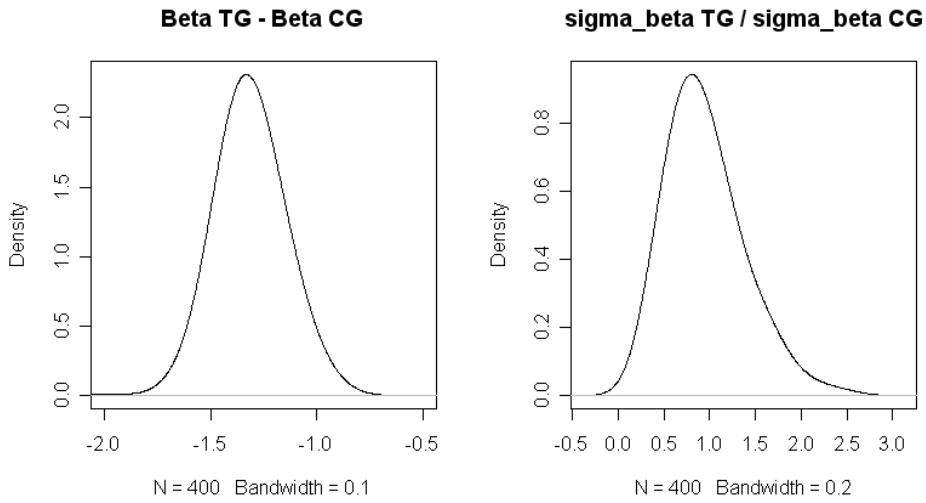

Figure 7. Difference of the growth rate (beta) and the ratio of the variances of the growth rates between the control group (CG) and treatment group (TG).

3. Conclusions

We draw a number of conclusions from this exercise. First, in spite of the complexity of the model, once we were able to derive the full conditional distributions, deriving posterior densities for any parameter or function of parameters becomes trivial using Gibbs sampling. For example, see Figure 7, where we were able to estimate the posterior density of $\sigma_{\beta\_TG} / \sigma_{\beta\_CG}$, which are the individual components of the full $\sum$ matrices. This is something that is not easily done using other methods. Second, using Gibbs sampling we were able to show that the population growth rate is lower in the treatment group, i.e. that whatever treatment is being applied has an effect.

However, in our experiments, the treatment group displays seemingly the same variation around its population growth rate as the control group (mean of the ratio density in figure 7 is 0.98), whereas in [1], the posterior density of the ratio implies a tigher density for the growth rate for the treatment group. Unfortunately, the specifics of how the authors arrived at their plots is not available, thus not enabling us to do any further substantive comparisons. Nevertheless, all other results match very well.

4. References

[1] A. Gelfand, S. Hills, A. Racine-Poon, A. Smith. “Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling”. JASA, Vol. 95, No. 412, Dec 1990, 972-985.