Paper 266

A GENERAL GIBBS SAMPLING ALGORITHM FOR ANALYZING LINEAR MODELS USING THE SAS® SYSTEM

Jayawant Mandrekar, Daniel J. Sargent, Paul J. Novotny, Jeff A. Sloan
Mayo Clinic, Rochester, MN 55905

ABSTRACT

A general Gibbs sampling algorithm for analyzing a broad class of linear models under a Bayesian framework is presented using Markov Chain Monte Carlo (MCMC) methodology in the SAS system. The analysis of a North Central Cancer Treatment Group (NCCTG) oncology clinical trial involving a two-period two-treatment crossover design is presented as an example. Results for the Bayesian model are compared to standard linear models analysis of variance procedures.

INTRODUCTION

The past decade has seen an explosion of interest and research into new methods for statistical computing. At the forefront of this explosion has been the discovery by statisticians of methods grouped under the heading of Markov chain Monte Carlo (MCMC). Using MCMC methods, statisticians have been able to propose, fit, and make inferences on a wide variety of complex models that were previously computationally infeasible.

One area that MCMC computing has made great inroads into standard statistical practice is hierarchical linear models. In a hierarchical model, the model is specified in stages. For example, in one model for analyzing data from a cross-over clinical trial, described in detail by Sloan et al (1997), with repeated measurements per person, the first level of the model relates the observations in each period to a number of fixed effects, plus a subject specific random effect. In the second level of the model, the subject specific random effects are assumed to come from some distribution. If y_{ijk} is the response of patient i in group j in period k, we can specify the model as

$$y_{ijk} = \mu + \tau_j + \pi_k + \lambda_{jk} + s_i + e_{ijk}$$

where $e_i \sim N(0, \sigma^2)$ and $s_i \sim N\left(0, \nu\right)$.

In this model τ is the treatment effect, π is the period effect, λ is the differential carryover effect, and the s_i are patient specific random effects. To fit this model using MCMC methods, a prior distribution for the second level variance ν is necessary. This prior distribution is essential for model computational stability, and to ensure a proper posterior distribution.

Given a model specification, with an associated posterior distribution, MCMC methods are based on generating a sample from a Markov chain that has the true posterior as its stationary distribution. The most common MCMC algorithm, Gibbs sampling (Gelfand et al, 1991), is based on using the "full conditional" distributions for the model parameters, that is, the distribution of each model parameter conditional on all of the others. Under broad regularity conditions, it can be shown that a sequence of draws generated from these full conditionals, after algorithm convergence, is a Markov chain of draws from the full conditional distribution.

Until recently, there has been a lack of general purpose software for fitting models using MCMC methods. The software package BUGS (Spiegelhalter et al, 1996) uses an S-like syntax to specify the model and the model fitting. While this package is rather general, it requires installation of a separate software package. Cantell (1997) proposed a Gibbs sampler using the SAS system to fit a simple linear regression with one independent variable. It was our goal to create a SAS Macro that could fit a much broader class of models using MCMC methods. In particular, the macro should have the ability to handle multiple dependent variables, as well as a limited capacity to incorporate random effects into the model. In this paper we describe such a macro, including descriptions of its level of flexibility and the interpretation of the output.

MATHEMATICAL DETAILS

In this section we describe the key mathematical details necessary to understand the algorithm. The basis for all that will follow is a class of algorithms known as Structured Markov chain Monte Carlo (SMCMC), as introduced by Sargent et al (1998), who demonstrate how an arbitrarily complex hierarchical linear model can be fit using a straightforward Gibbs sampler. SMCMC uses the technique of Hodges (1998) to reformulate the hierarchical linear model into a standard linear model. Once we have reformulated the model into a standard linear model, the model can be expressed as

$$Y = X\Theta + E$$

where $Y, X, \Theta,$ and E are properly augmented matrices (see Sargent et al, 1998), and $Cov(E) = \Gamma$. The two components of the model that we may wish to draw inferences on are Θ, the vector of regression parameters and random effects, and Γ, which is composed of the error variance σ^2 and the random effect variance ν. Given a model specified in this fashion, the full conditional for the regression parameter Θ (which includes both the fixed and the random effects) can be written as

$$\Theta \mid X, Y, \Gamma \sim N((X^T\Gamma^{-1}X)^{-1}X^T\Gamma^{-1/2}Y, \\ (X^T\Gamma^{-1}X)^{-1}),$$

which is a multivariate normal that can easily be sampled from. The full conditional for Γ can be represented through the full conditionals for σ^2 and ν. It is standard to assume a gamma prior distribution for $1/\nu = h_S$, with parameters α and β (using the parameterization that has mean α/β and variance α/β^2). For future reference, a convenient functional for σ^2 is defined by $h_e = 1/\sigma^2$. The gamma prior distribution for h_S is congugate to the likelihood, implying that the full conditional distribution for h_S will be a gamma distribution as well. A flat prior distribution is assumed for h_e.

With this set-up, the Gibbs sampler cycles through the 3 full conditionals $f(\Theta|Y, X, \Gamma)$, $f(h_e|Y, X, \Theta, h_S)$, and $f(h_S|Y, X, \Theta, h_e)$ to generate a sequence of draws from the joint posterior distribution $f(\Theta, \Gamma|Y, X)$. The result of the Gibbs sampler is a sequence of samples drawn from the posterior distribution. These samples can be used to generate any desired summary statistics for the parameters of interest. In our application, quantities of interest include posterior means, standard deviations, and univariate probabilities. These summary statistics allow for standard inferences to be drawn for the model parameters (treatment and covariate effects). Using the MCMC samples to generate more sophisticated summaries is straightforward, including joint confidence regions and probabilities of arbitrary intervals. To calculate probabilities of an arbitrary interval, for example, the user can count the number of iterations that the realization of a parameter is in the interval and divide by the total number of iterations. A further advantage of the Bayesian approach is the ability to make probabilistic statements regarding the relative likelihood of alternative values for the parameters. Such inference is possible only indirectly through standard frequentist methodology. This approach allows us therefore to verify standard approaches as well as allowing more complex modeling strategies.

SAS CODE

The SAS code developed to carry out the Gibbs sampler for linear models makes extensive use of PROC IML. The algorithm can handle linear models involving up to 10 covariates, and a single random effect. The macro expects data to be sorted by identification number (so that observations from the same patient are adjacent to each other). The default model will create a design matrix for 4 parameters that may be appropriate in analysis of the two-period, two-treatment cross-over clinical trial, namely μ, τ, π, and λ, which were appropriate for the case that provided the initial impetus for the algorithm creation. The user can select the desired parameters from these according to the particular model that is required. If any of these four parameters are to be included in the model, the first additional covariate included must be the patient's arm assignment.

To call the algorithm, the user has to specify the name of the response variable, whether each of the four default parameters should be in the model (coded as 1=term in the model, 0=term excluded from the model), the names of additional covariates (whose values for each patient must be defined in the input dataset), the number of observations per patient, the number of iterations, and the parameters for the prior distribution of h_S. The macro has the ability to accommodate any number of observations per patient, but each patient must have the same number of observations. Unbalanced designs could potentially be handled with further modification to the algorithm or by using established imputation methods.

Based on the variables specified in the macro call, the code first creates the usual design matrix based on the fixed effects. This design matrix is then augmented as described by Hodges (1998) so that the entire model specification can be written in the form (1). Once the model is written in this form, (2) specifies the multivariate normal distribution from which the value of Θ at the next iteration in the chain is drawn.

By default the macro assumes a gamma prior distribution for the parameter h_S with α = 1.0 and β =0.1, having mean and standard deviation 10. These can be modified in the macro call to be more appropriate to the particular problem.

Each iteration continues by updating the full conditional distributions. The code generates a random sample of size one from each conditional distribution in succession. This iterative procedure continues for a user-specified number of iterations. Finally, this entire process is repeated three times to generate three independent Markov chains that can be used to provide posterior summaries. This process generates two output files that contain listings of the realizations of the parameters at each iteration. The first file (alPARAMS) contains the model parameters Θ, in the order: parameter, iteration, chain. Thus if there are nparam parameters and nreps number of iterations, the first nparam elements of alPARAMS are the value of each parameter at iteration 1, the next nparam elements are the values at iteration 2, etc, until the nreps number of iterations in the first chain are presented. The second and third chains follow in the same order. The second output file (alhehs) contains the same data for the variance components.

One effective method for examining the output from a MCMC algorithm is the use of graphical techniques. We have written a plotting macro that displays the output of the Gibbs macro. Required arguments are the names of the datasets for the covariates and the variance parameters, the number of covariates, the number of variance parameters, the number of iterations at the beginning of the chain to exclude from the summary statistics (the "burn-in" period) and the number of iterations per chain. The results from the MCMC algorithm described in the example below are shown in Figure 2. This complex figure shows trace plots for a representative sample of the parameters in the model. The trace plots use the overlay option in PROC GPLOT to plot the iteration number on the x-axis versus the value at that iteration on the y-axis for three independent MCMC chains. Shown below the plot for each parameter are estimated means and standard deviations for that parameter, as well as the estimated probability that each parameter is less than zero, which gives an indication as to how strongly the parameter differs from zero. Shown above each plot is an estimate of the within-chain single lag auto-correlation, which is a measure of how well the chain is spanning the sample space. Values near 1 indicate a slowly moving chain, whereas values near 0 indicate a chain that is moving well through the sample space. Formal convergence tests or diagnostics are not included as part of the macro, any standard technique can be applied to the output (see Cowles and Carlin, 1996).

EXAMPLE

We illustrate the Gibbs sampling algorithm with the analysis of a two-treatment two period crossover design. The study was an NCCTG oncology clinical trial with the goal of determining if post-mastectomy pain could be alleviated by the application of capsaicin cream. The endpoint of interest was the area under the curve (AUC) of patient-reported pain while using either a placebo or capsaicin cream. The extensive alternative analyses produced by SAS code presented previously (Sloan et al, 1997) indicated that effects due to treatment, period and patient were all present in the experimental layout.

Figure 1 (taken from Ellison et al, 1997), plots the results by week, showing a significant decrease in pain in patients using the capsaicin cream compared to those on placebo.

The call to the macro for this experimental layout (assuming two observations per patient, 100 iterations per chain, and a prior distribution with α = 1.0 and β =0.1 is as follows:

```
%gibbs(response=y, mu=1, tau=1, pi=1, lambda=1, covar=arm, nmult=2, nreps=100, alpha=1.0, beta=0.1);
```

Figures 2 and 3 present the results of three chains run for 100 iterations each for this model. These plots are created using the plotting macro, with the call:

```
%gibbsplot(alPARAMS, alhehs, ncovpar=14, nvarpar=2, nthrow=25, nreps=100).
```

The first four parameters shown in Figure 2 are μ, τ, π , and λ, the last ten parameters are a sample of the individual patient random effects. Figure 3 displays trace plots for the two variance components h_e and h_s. Based on the figures, the model parameters stabilize quickly after a few iterations. This sort of pattern is necessary for the subsequent model analysis to be credible.

Table 1 provides estimates, standard errors, and p-values produced for the various effects involved in the Bayes and Fisherian analysis of variance models. Both approaches provide comparable inference regarding the experimental layout and validate the results published previously. The standard errors from the MCMC algorithm are approximately 1/3 larger than those resulting from the standard code. This is likely due to the Bayesian procedure’s ability to account for additional sources of variability, such as the variability in h_e, when estimating the main effect parameters.

Table 1				
	Proc Mixed		Gibbs Macro	
Parameter	Mean	S.E.	Mean	S.E.
μ	1.137	(0.014)	1.135	(0.025)
τ	-0.001	(0.021)	0.000	(0.032)
π	0.010	(0.021)	0.011	(0.032)
λ	-0.006	(0.041)	-0.007	(0.063)
s_1	-0.039	(0.028)	-0.040	(0.033)
s_{11}	0.105	(0.028)	0.110	(0.035)
s_{21}	-0.072	(0.028)	-0.076	(0.034)
s_{31}	-0.002	(0.028)	0.000	(0.037)
s_{41}	0.120	(0.028)	0.128	(0.035)
s_{51}	-0.085	(0.028)	-0.090	(0.035)
s_{61}	0.065	(0.028)	0.071	(0.032)
s_{71}	-0.043	(0.028)	-0.042	(0.030)
s_{81}	-0.111	(0.028)	-0.115	(0.033)
s_{91}	-0.090	(0.028)	-0.091	(0.032)

A variety of prior distributions were used here to examine the sensitivity of the results to the specific prior distributions used. If prior information was known about the variability inherent in any of the model process, this information could be incorporated through prior distributions.

CONCLUSIONS

The construction of this general Gibbs sampling algorithm allows for the routine application of Bayes methodology of the analysis of linear models using the SAS system. The Bayes approach may provide more efficient estimation than standard analysis of variance procedures if informative prior information on the experimental layout is available.

The SAS code is applicable to a wide variety of linear models involving multiple treatment and covariate effects. A specific subset of the Gibbs sampling code has been added to the previous work of Sloan et al (1997) for specific and complete analysis of the two-period two-treatment crossover design.

REFERENCES

1. Cantell B. (1997). Using Linear Regression Analysis and the Gibbs Sampler to Estimate the Probability of a Part Being Within Specification. *Proceedings of the 22nd Annual SAS Users Group International Conference*, 22, 1220-1225.

2. Cowles MK, and Carlin BP. (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*, 91:883-904.

3. Ellison NE, Loprinzi CL, Kugler J, et al. (1997). Phase III Placebo Controlled Trial of Capsaicin Cream in the Management of Surgical Neuropathic Pain in Cancer Patients. *Journal of Clinical Oncology* 15, 2974-2980.

4. Gelfand AE, and Smith AFM. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* 85, 398-409.

5. Hodges JS. (1998). Some Algebra and Geometry for Hierarchical Models, Applied to Diagnostis (with discussion). *Journal of the Royal Staistical. Society, Series B*, 60:497-536.

6. Sargent DJ, Hodges JS, Carlin BP. (1998). Structured Markov Chain Monte Carlo. Submitted to *Journal of Computational & Graphical Statistics*.

7. Sloan JA, Novotny PJ, Loprinzi CL, Nair S. (1997). Graphical and Analytical Tools for the Analysis of Two-Period Crossover Clinical Trials. *Proceedings of the 22nd Annual SAS Users Group International Conference*, 22, 1312-1317.

8. Spiegelhalter DJ, Thomas A, Best N, Gilks WR. (1995a). BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50. Technical Report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.

CONTACT INFORMATION

For further information, please contact:

Dr. Daniel Sargent
Mayo Clinic
200 1st ST SW
Rochester, MN 55905
Phone: (507) 284-5380, Fax: (507) 284-1902
Email: sargent.daniel@mayo.edu

Figure 1: Average pain score by treatment group sequence and week

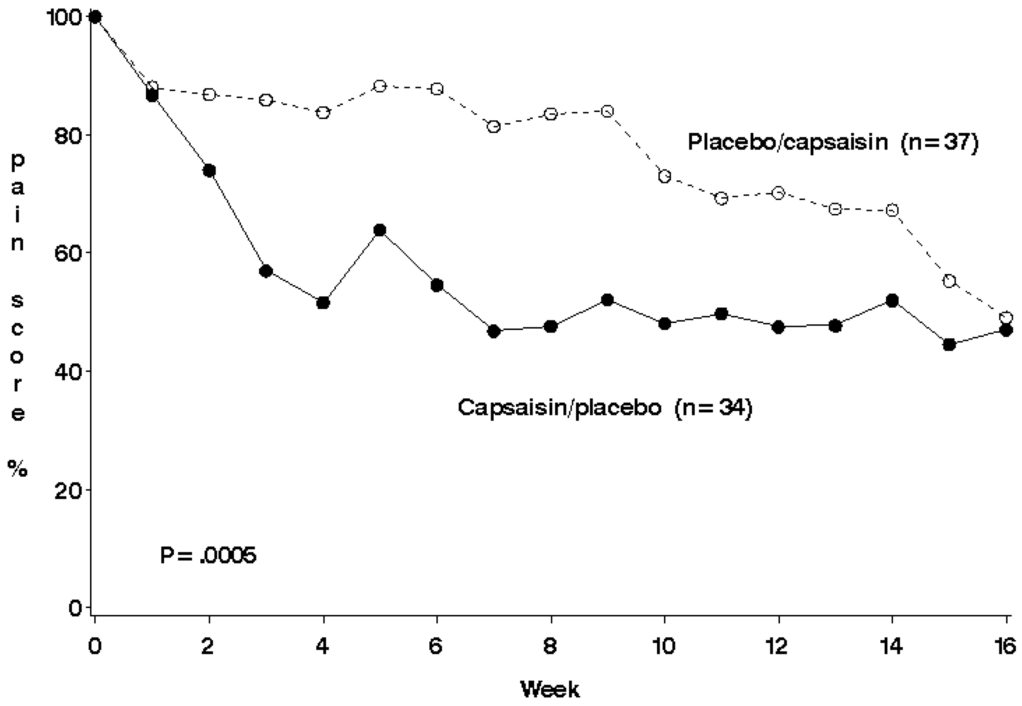


Figure 3: chainplots for two parameters

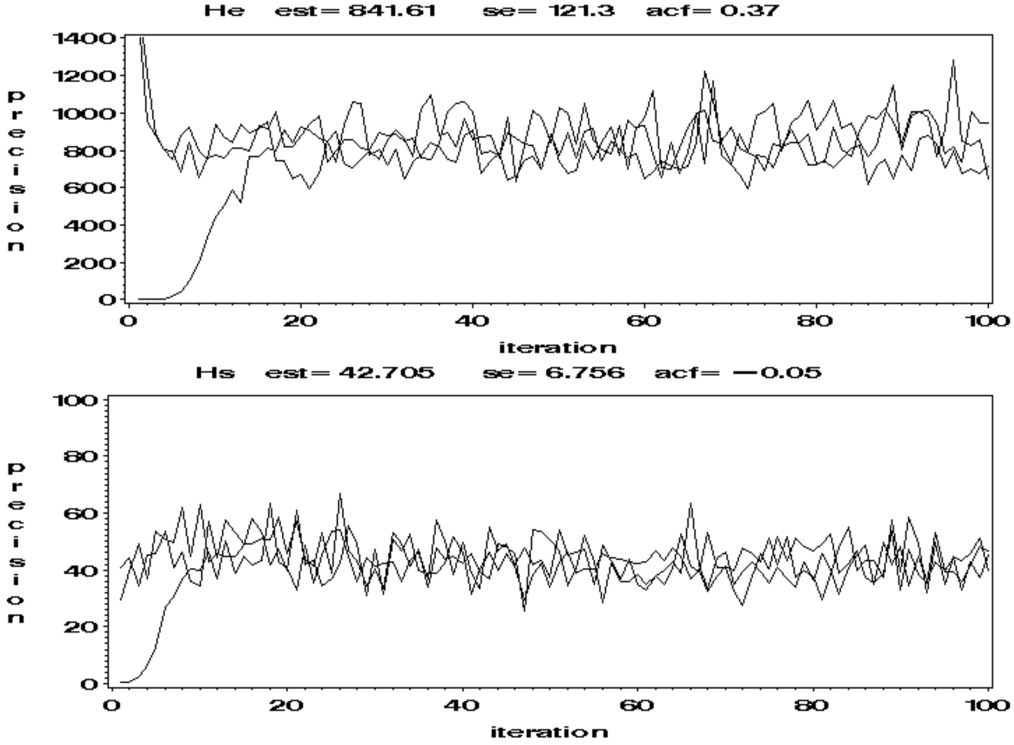
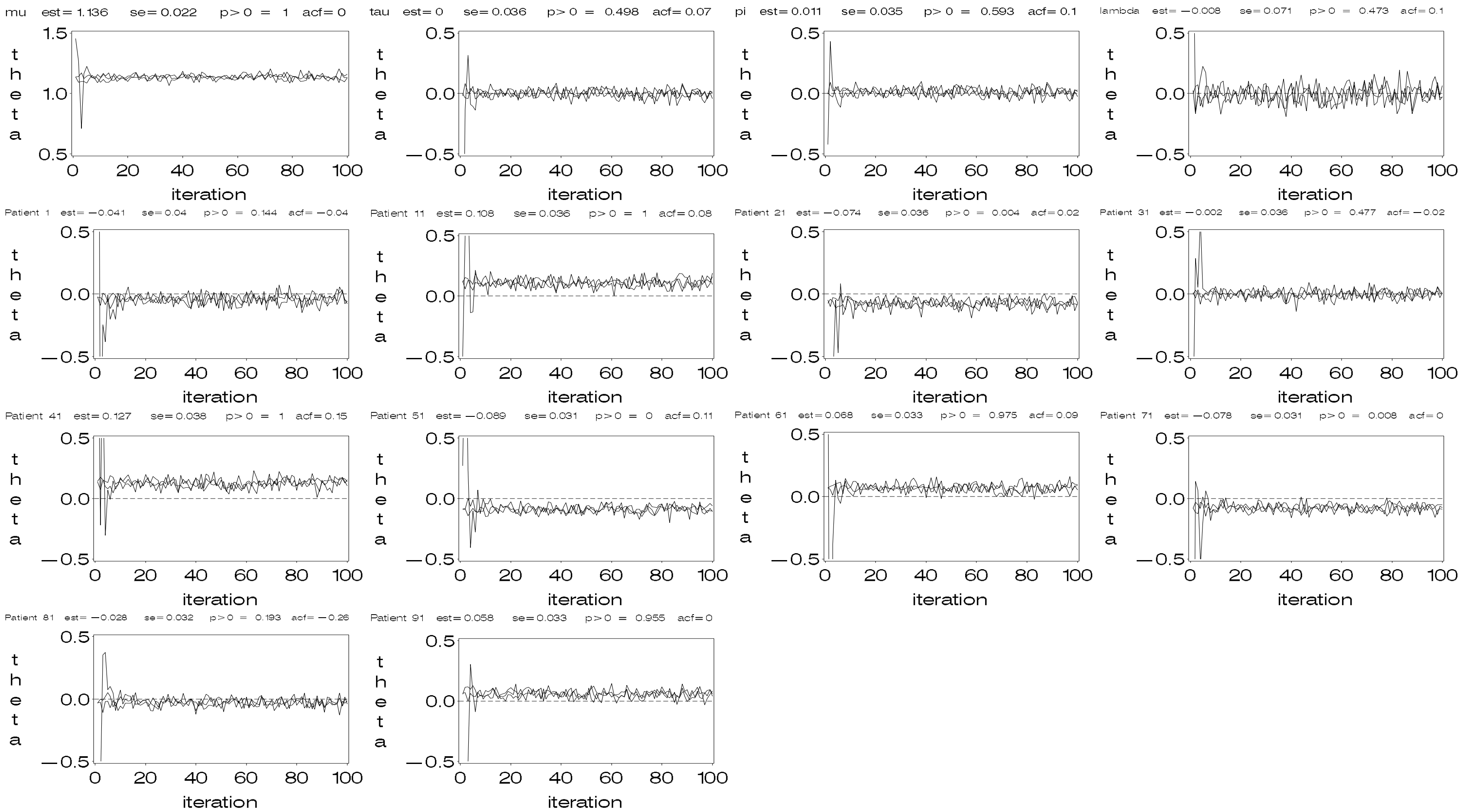


Figure 2: Chain plots for three independent Gibbs sampler chains


mu est= 1.136 se= 0.022 p> 0 = 1 acf= 0
tau est= 0 se= 0.036 p> 0 = 0.498 acf= 0.07
pi est= 0.011 se= 0.035 p> 0 = 0.593 acf= 0.1
lambda est= —0.008 se= 0.071 p> 0 = 0.473 acf= 0.1
Patient 1 est= —0.041 se= 0.04 p> 0 = 0.144 acf= —0.04
Patient 11 est= 0.108 se= 0.036 p> 0 = 1 acf= 0.08
Patient 21 est= —0.074 se= 0.036 p> 0 = 0.004 acf= 0.02
Patient 31 est= —0.002 se= 0.036 p> 0 = 0.477 acf= —0.02
Patient 41 est= 0.127 se= 0.038 p> 0 = 1 acf= 0.15
Patient 51 est= —0.089 se= 0.031 p> 0 = 0 acf= 0.11
Patient 61 est= 0.068 se= 0.033 p> 0 = 0.975 acf= 0.09
Patient 71 est= —0.078 se= 0.031 p> 0 = 0.008 acf= 0
Patient 81 est= —0.028 se= 0.032 p> 0 = 0.193 acf= —0.26
Patient 91 est= 0.058 se= 0.033 p> 0 = 0.955 acf= 0
theta
iteration
—0.5