DOCUMENT RESUME

ED 424 283		TM 029 163

AUTHOR	Kim, Seock-Ho
TITLE	An Evaluation of a Markov Chain Monte Carlo Method for the Rasch Model.
PUB DATE	1998-06-19
NOTE	61p.; Paper presented at the joint Annual Meeting of the Psychometric Society and the Classification Society of North America (Urbana, IL, June 19, 1998).
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Ability; Estimation (Mathematics); *Item Response Theory; *Markov Processes; Maximum Likelihood Statistics; *Monte Carlo Methods; *Sampling; Tables (Data)
IDENTIFIERS	*Gibbs Sampling; *Rasch Model

ABSTRACT

The accuracy of the Markov chain Monte Carlo procedure, Gibbs sampling, was considered for estimation of item and ability parameters of the one-parameter logistic model. Four data sets were analyzed to evaluate the Gibbs sampling procedure. Data sets were also analyzed using methods of conditional maximum likelihood, marginal maximum likelihood, and joint maximum likelihood. Two different ability estimation methods, maximum likelihood and expected a posteriori, were employed under the marginal maximum likelihood estimation of item parameters. Item parameter estimates from Gibbs sampling were similar to those obtained from the expected a posteriori method. (Contains 12 figures, 23 tables, and 60 references.) (Author)

An Evaluation of a Markov Chain Monte Carlo Method for the Rasch Model

Seock-Ho Kim
The University of Georgia

June 19, 1998
Running Head: MCMC METHOD FOR 1PL

Paper presented at the joint annual meeting of the Psychometric Society and the Classification Society of North America, Urbana, Illinois

An Evaluation of a Markov Chain Monte Carlo Method for the Rasch Model

Abstract

The accuracy of the Markov chain Monte Carlo procedure, Gibbs sampling, was considered for estimation of item and ability parameters of the one-parameter logistic model. Four data sets were analyzed to evaluate the Gibbs sampling procedure. Data sets were also analyzed using methods of conditional maximum likelihood, marginal maximum likelihood, and joint maximum likelihood. Two different ability estimation methods, maximum likelihood and expected a posteriori, were employed under the marginal maximum likelihood estimation of item parameters. Item parameter estimates from the four methods were almost identical. Ability estimates from Gibbs sampling were similar to those obtained from the expected a posteriori method.

Index terms: Bayesian inference, conditional maximum likelihood, Gibbs sampling, item response theory, joint maximum likelihood, Markov chain Monte Carlo, marginal maximum likelihood, Rasch model.

Introduction

Some problems in statistical inference require integration over possibly high-dimensional probability distributions in order to estimate model parameters of interest or to obtain characteristics of model parameters. One such problem is estimation of item and ability parameters in the context of item response theory (IRT). Except for certain rather simple problems with highly structured frameworks (e.g., an exponential family together with conjugate priors in Bayesian inference), the required integrations may not be analytically feasible. Many efficient numerical approximation strategies have been recently developed for complicated integrations. In this paper, we examine the accuracy of one of the efficient numerical approximation strategies, a Markov Chain Monte Carlo (MCMC) method, for estimation of IRT item and ability parameters. We focus on the accuracy of a particular MCMC procedure, Gibbs sampling (Geman & Geman, 1984), for estimation of item and ability parameters under the one-parameter logistic (1PL) model (Rasch, 1960/1980).

A number of ways exist for implementing the MCMC methods. For a review, refer to Bernardo and Smith (1994), Carlin and Louis (1996), and Gelman, Carlin, Stern, and Rubin (1995). Metropolis and Ulam (1949), Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953), and Hasting (1970) present a general framework within which Gibbs sampling (Geman & Geman, 1984) can be considered as a special case. In this regard, Gelfand and Smith (1990) discuss several different Monte Carlo-based approaches, including Gibbs sampling, for calculating marginal densities. Gilks, Richardson, and Spiegelhalter (1996) contains a recent survey of applications of Gibbs sampling. Basically Gibbs sampling is applicable for obtaining parameter estimates from the complicated joint posterior distribution in Bayesian estimation under IRT (e.g., Mislevy, 1986; Swaminathan & Gifford, 1982, 1985, 1986; Tsutakawa & Lin, 1986).

Albert (1992) applied Gibbs sampling in the context of IRT to estimate item parameters for the two-parameter normal ogive model and compared these estimates with those obtained using maximum likelihood estimation. Baker (1998) has also investigated item parameter recovery characteristics of Albert's Gibbs sampling method for item parameter estimation via a simulation study. Patz and Junker (1997) developed a MCMC method based on the Metropolis-Hasting algorithm and presented an illustration using the two-parameter logistic model.

MCMC computer programs in IRT have been developed largely only for specific applications. For example, Albert (1992) used a computer program written in MATLAB (The MathWorks, Inc., 1996). Baker (1998) developed a specialized FORTRAN version of Albert's Gibbs sampling program to estimate item parameters of the two parameter normal ogive model. Patz and Junker (1997) developed an S-PLUS code (MathSoft, Inc., 1995). Spiegelhalter, Thomas, Best, and Gilks (1997) have also developed a general Gibbs sampling computer program BUGS for Bayesian estimation, using the adaptive rejection sampling algorithm (Gilks & Wild, 1992). The computer program BUGS requires specification of the complete conditional distributions.

For the Rasch model (Rasch, 1960/1980; Fischer & Molenaar, 1995; Wright & Stone, 1979) many estimation methods can be used to obtain item and ability parameter estimates (Molenaar, 1995; Hoijtnik & Boomsma, 1995). Item and person parameters can be estimated jointly by maximizing the joint likelihood function (i.e., JML, Wright & Stone, 1979). Conditional maximum likelihood (CML) seems to be the standard estimation method under the Rasch model for estimation of item parameters (e.g., Molenaar, 1995). Also, Marginal maximum likelihood (MML) estimation using the expectation and maximization algorithm can be used to obtain item parameter estimates (Thissen, 1982). In addition, joint Bayesian estimation and marginal Bayesian estimation can be employed to obtain parameter estimates under the Rasch model (e.g., Swaminathan & Gifford, 1982). The Gibbs sampling procedure approaches the estimation of item and ability parameters using the joint posterior distribution rather than the marginal distribution. Even so, all methods should yield comparable item parameter estimates, especially when comparable priors are used or when ignorance or locally-uniform priors are used. This paper was designed to investigate this issue using the 1PL model. Specifically, item and ability estimates from the methods of Gibbs sampling, CML, MML, and JML, were examined and compared.

Theoretical Framework

Joint Estimation Procedures

Consider binary responses to a test with n items by each of N examinees. A response of examinee i to item j is represented by a random variable Y_{ij}, where $i = 1(1)N$ and $j = 1(1)n$. The probability of a correct response of examinee i to item j is given by $P(Y_{ij} = 1|\theta_i, \xi_j) = P_{ij}$ and the probability of an incorrect response is given by $P(Y_{ij} = 0|\theta_i, \xi_j) = 1 - P_{ij} = Q_{ij}$, where θ_i is ability and ξ_j is the item parameter or possibly the vector of item parameters.

For examinee i, there is an observed vector of dichotomously scored item responses of length n, $Y_i = (Y_{i1}, \ldots, Y_{in})'$. Under the assumption of conditional independence, the probability of Y_i given θ_i and the vector of all item parameters, $\xi = (\xi_1, \ldots, \xi_n)'$, is

$$p(Y_i|\theta_i, \xi) = \prod_{j=1}^{n} P_{ij}^{Y_{ij}} Q_{ij}^{1-Y_{ij}}. \tag{1}$$

The probability of obtaining the $N \times n$ response matrix Y is given by

$$p(Y|\theta, \xi) = \prod_{i=1}^{N} \prod_{j=1}^{n} P_j(\theta_i)^{Y_{ij}} Q_j(\theta_i)^{1-Y_{ij}} = l(\theta, \xi|Y), \tag{2}$$

where $\theta = (\theta_1, \ldots, \theta_N)'$. Note that $l(\theta, \xi|Y)$ can be regarded as a joint function of θ and ξ given the data Y. Wright and Stone (1979) describe the joint estimation of θ and ξ (cf. Birnbaum, 1968; Lord, 1980, 1986). In implementation of JML, the item parameter estimation part for maximizing $l(\xi|Y, \hat{\theta})$ and the ability parameter estimation part for maximizing $l(\theta|Y, \hat{\xi})$ are iterated until a stable set of maximum likelihood estimates of item and ability parameters is obtained.

Extending the idea of joint maximization, Swaminathan and Gifford (1982, 1985, 1986) suggested that θ and ξ can be estimated by joint maximization with respect to the parameters of the posterior density

$$p(\theta, \xi|Y) = \frac{p(Y|\theta, \xi)p(\theta, \xi)}{p(Y)} \propto l(\theta, \xi|Y)p(\theta, \xi), \tag{3}$$

where $\propto$ denotes proportionality and $p(\theta, \xi)$ is the prior density of the parameters θ and ξ. This procedure is called joint Bayesian estimation. A prior distribution represents what is known about unknown parameters before the data are obtained. Prior knowledge or even relative ignorance can be represented by such a distribution. Under the assumption that priors of θ and ξ are independently distributed with probability density functions $p(\theta)$ and $p(\xi)$, the item parameter estimation part maximizing $l(\xi|Y, \hat{\theta})p(\xi)$, and the ability parameter estimation part maximizing $l(\theta|Y, \hat{\xi})p(\theta)$ are iterated to obtain the Bayes modal estimates of item and ability parameters.

Conditional Maximum Likelihood

Andersen (1970, 1972) showed that consistent estimates of item parameters can be obtained using the conditional estimation procedure. The conditional estimation procedure is based on the availability of sufficient statistics for the ability parameters. Under the Rasch model, the number correct score, $R_i = \sum_j Y_{ij}$, is the sufficient statistics for θ_i and, consequently, R is the sufficient statistics for θ. For a given examinee with θ_i, the conditional probability of Y_i given R_i can be written as

$$\frac{p(Y_i|\theta_i,\xi)}{p(R_i|\theta_i,\xi)} = p(Y_i|R_i,\xi) \tag{4}$$

which does not contain θ_i. Hence, the entire likelihood can be expressed in terms of R instead of θ, that is,

$$l(\xi|Y,R). \tag{5}$$

The CML estimates of item parameters can be obtained by maximizing the conditional likelihood function without any reference to the ability parameters. The ability parameters are estimated separately under CML, in general, using the maximum likelihood method. The conditional likelihood function involves computing elementary symmetric functions (see Baker & Harwell, 1996).

Marginal Estimation Procedures

In marginal solutions, ability will be integrated out from either the likelihood function or the posterior distribution. The marginal probability of obtaining the response vector Y_i for examinee i sampled from a given population is

$$p(Y_i|\xi) = \int p(Y_i|\theta_i,\xi)p(\theta_i)d\theta_i, \tag{6}$$

where $p(\theta_i)$ is the population distribution of θ_i. Without loss of generality, we can assume that the θ_i are independent and identically distributed as standard normal, $\theta_i \sim N(0,1)$. This assumption may be relaxed as the ability distribution can also be empirically characterized (Bock & Aitkin, 1981). The marginal probability of Y_i can be approximated with any specified degree of precision by Gaussian quadrature formulas (Stroud & Secrest, 1966).

The marginal probability of obtaining the $N \times n$ response matrix Y is given by

$$p(Y|\xi) = \prod_{i=1}^{N} p(Y_i|\xi) = l(\xi|Y), \tag{7}$$

where $l(\xi|Y)$ can be regarded as a function of ξ given the data Y. In MML, this marginal likelihood is maximized to obtain maximum likelihood estimates of item parameters (Bock & Aitkin, 1980; Thissen, 1982). Ability parameters are estimated after obtaining the item parameter estimates assuming the estimates are the true parameter values.

Bayes' theorem tells us that the marginal posterior probability distribution for ξ given Y is proportional to the product of the marginal likelihood for ξ given Y and the prior distribution of ξ. That is,

$$p(\xi|Y) = \frac{p(Y|\xi)p(\xi)}{p(Y)} \propto l(\xi|Y)p(\xi). \tag{8}$$

The marginal likelihood function represents the information obtained about ξ from the data. In this way, the data modify our prior knowledge of ξ. In marginal Bayesian estimation of item parameters, the marginal posterior is maximized to obtain Bayes modal estimates of item parameters (Mislevy, 1986).

Gibbs Sampling

The main feature of MCMC methods is to obtain a sample of parameter values from the posterior density (Tanner, 1996). The sample of parameter values then can be used to estimate some functions or moments (e.g., mean and variance) of the posterior density of the parameter of interest. In comparison, in the above IRT estimation procedures via JML, CML, or MML, the task is to obtain modes of the likelihood function or of the posterior distribution.

The Gibbs sampling algorithm is as follows (Gelfand & Smith, 1990; Tanner, 1996). First, instead of using θ and ξ, let ω be a vector of parameters with k elements. Suppose that the full or complete conditional distributions, $p(\omega_i|\omega_j, Y)$, where $i = 1(1)k$ and $j \neq i$, are available for sampling. That is, samples may be generated by some method given values of the appropriate conditioning random variables. Then given an arbitrary set of starting values, $\omega_1^{(0)}, \ldots, \omega_k^{(0)}$, the algorithm proceeds as in Figure 1. The vectors $\omega^{(0)}, \ldots, \omega^{(t)}, \ldots$ are a realization of a Markov chain with a transition probability from $\omega^{(t)}$ to $\omega^{(t+1)}$ given by

$$p(\omega^{(t)}, \omega^{(t+1)}) = \prod_{l=1}^{k} p(\omega_l^{(t+1)}|\omega_j^{(t)}, j > l, \omega_j^{(t+1)}, j < l, Y). \tag{9}$$

Insert Figure 1 about here

The joint distribution of $\omega^{(t)}$ converges geometrically to the posterior distribution $p(\omega|Y)$ as $t \to \infty$ (Geman & Geman, 1984; Bernardo & Smith, 1994). In particular, $\omega_i^{(t)}$ tends to be distributed as a random quantity whose density is $p(\omega_i|Y)$. Now suppose that there exist

m replications of the t iterations. For large t, the replicates $\omega_{i1}^{(t)}, \ldots, \omega_{im}^{(t)}$ are approximately a random sample from $p(\omega_i|Y)$. If we make m reasonably large, then an estimate, $\hat{p}(\omega_i|Y)$, can be obtained either as a kernel density estimate derived from the replicates or as

$$\hat{p}(\omega_i|Y) = \frac{1}{m}\sum_{l=1}^{m} p(\omega_i|\omega_{jl}^{(t)}, j \neq i, Y). \tag{10}$$

In the context of IRT, Gibbs sampling tries to obtain or sample sets of parameters from the joint posterior density $p(\theta, \xi|Y)$. Inferences with regard to parameters can then be made using the sampled parameters. Note that inference for both θ and ξ can be made from the Gibbs sampling procedure.

Steps of Gibbs Sampling

Gibbs sampling uses the following four basic steps (cf. Spiegelhalter, Best, Gilks, & Inskip, 1996):

1. Full conditional distributions and sampling methods for unobserved parameters must be specified.
2. Starting values must be provided.
3. Output must be monitored.
4. Summary statistics (e.g., estimates and standard errors) for quantities of interest must be calculated.

Discussion of the four steps involved are presented in detail below using four data sets (i.e., Examples 1 to 4), especially in Example 1. In addition, comparisons with the results from CML, MML, and JML as implemented in the computer programs, PML (Molenaar, 1990), BILOG (Mislevy & Bock, 1990) and BIGSCALE (Wright, Linacre, & Schultz, 1989), are presented. The four data sets analyzed in Examples 1 to 4 represent different calibration situations under the Rasch model, ranging from an extremely small number of items/examinees to a relatively large number of items/examinees.

Example 1

Data

The first example is presented using the familiar Law School Admission Test Section 6 (LSAT6) data from Bock and Lieberman (1970) (see also Andersen, 1980; Bock & Aitkin, 1981). The LSAT6 data are given in Table 1. Model parameters were estimated by Gibbs sampling using the computer program BUGS (Spiegelhalter et al., 1997). These same LSAT6 data have been analyzed under the 1PL model and under the two-parameter normal ogive (i.e., probit) model in Spiegelhalter, Thomas, Best, and Gilks (1996). Spiegelhalter, Thomas, et al. (1996) also compared the BUGS results with those from Bock and Aitkin (1981).

Insert Table 1 about here

Model Specifications

The model specifications are used as input to the BUGS computer program. In the LSAT6 data set, the item responses Y_{ij} are independent, conditional on their parameters P_{ij}. For examinee i and item j, each P_{ij} is a function of the ability parameter θ_i, the location parameter β_j, and the slope parameter α under the 1PL (cf. Thissen, 1982). The θ_i are assumed to be independently drawn from a standard normal distribution for scaling purposes. Figure 2 is adopted from Spiegelhalter, Thomas, et al. (1996) and shows a directed acyclic graph (see Lauritzen, Dawid, Larsen, & Leimer, 1990; Whittaker, 1990; Spiegelhalter, Dawid, Lauritzen, & Cowell, 1993) based on these assumptions. It is only possible to proceed by following the directions of the arrows. Each variable or quantity in the model appears as a node in the graph, and directed links correspond to direct dependencies as specified above. The solid arrow denotes the probabilistic dependency, while dashed arrows indicate functional or deterministic relationships. The rectangle designates observed data, and circles represent unknown quantities. The model can be seen as directed because each link between nodes is represented as an arrow. The model can also be seen as acyclic because it is impossible to return to a node after leaving.

Insert Figure 2 about here

It may be helpful to use the following definitions: Let v be a node in the graph, and V be the set of all nodes. A parent of v is defined as any node with an arrow extending from it and pointing to v, and a descendant of v is defined as any node on a direct path beginning from v. For identifying parents and descendants, deterministic links should be combined so that, for example, the parent of Y_{ij} is P_{ij}. It is assumed in Figure 2 for any node v, if we know the value of its parents, then no other nodes would be informative concerning v except descendants of v.

Lauritzen et al. (1990) indicated that, in a full probability model, the directed acyclic graph model is equivalent to assuming that the joint distribution of all the random quantities is fully specified in terms of the conditional distribution of each node given its parents. That is,

$$P(V) = \prod_{v \in V} P(v|\text{parents}[v]), \tag{11}$$

where $P(\cdot)$ denotes a probability distribution. This factorization not only allows extremely complex models to be built up from local components, but also provides an efficient basis for the implementation of MCMC methods (Spiegelhalter, Best, et al., 1996).

Gibbs sampling via the BUGS computer program works by iteratively drawing samples from the full conditional distributions of unobserved nodes in Figure 2 using the adaptive rejection sampling algorithm (Gilks, 1996; Gilks & Wild, 1992). For any node v, the remaining nodes are denoted by $V - v$. It follows that the full conditional distribution, $P(v|V - v)$, has the form

$$\begin{aligned} P(v|V-v) &\propto P(v, V-v) \\ &\propto P(v|\text{parent}[v]) \prod_{w \in \text{children}[v]} P(w|\text{parents}[w]). \end{aligned} \tag{12}$$

The proportionality constant, which is a function of the remaining nodes, ensures that the distribution is a probability function that integrates to unity.

To analyze the LSAT6 data, we begin by specifying the forms of the parent and child relationships in Figure 2. Under the 1PL model, the probability that examinee i responds correctly to item j is assumed to follow a logistic function

$$P_{ij} = \frac{\exp(\alpha\theta_i - \beta_j)}{1 + \exp(\alpha\theta_i - \beta_j)} = \frac{1}{1 + \exp[-(\alpha\theta_i - \beta_j)]}. \tag{13}$$

For scaling purposes, we may use the form

$$\theta_i' - b_j = \alpha\theta_i - \beta_j, \tag{14}$$

where θ_i' is the usual Rasch ability parameter and b_j is the Rasch item difficulty parameter defined as $\theta_i' = \alpha\theta_i - \bar{\beta}$ and $b_j = \beta_j - \bar{\beta}$, where $\bar{\beta}$ is the mean of the location parameters, $\bar{\beta} = \sum_j \beta_j / n$. Since Y_{ij} are Bernoulli with parameter P_{ij}, we can define

$$Y_{ij} \sim \text{Bernoulli}(P_{ij}) \tag{15}$$

and

$$\text{logit}(P_{ij}) = \alpha\theta_i - \beta_j. \tag{16}$$

To complete the specification of a full probability model in for the BUGS computer program, prior distributions of the nodes without parents (i.e., θ_i, β_j, and α) also need to be specified. We can define these priors in several different ways. We can impose priors on β_j and α using a hierarchical Bayes approach (e.g., Swaminathan & Gifford, 1982, 1985; Kim, Cohen, Baker, Subkoviak, & Leonard, 1994). If it is preferred that the priors not be too influential, uninformative priors could be imposed. Alternatively, it may also be useful to include external information in the form of fairly informative prior distributions. According to Spiegelhalter, Best, et al. (1996), it is important to avoid causal use of standard improper priors in MCMC modeling, since these may result in improper posterior distributions. Following Spiegelhalter, Thomas, et al. (1996), the uninformative prior distributions were chosen for the LSAT6 analyses to make comparisons with other estimation methods. The prior of β_j was $N(0, 100^2)$ and the prior of α was $N(0, 100^2)$ with the range restriction, $\alpha > 0$, to yield only positive values of the Gibbs sampler for the slope parameter. The prior distribution for α can be seen as a half normal distribution or the singly truncated normal distribution (Johnson, Kotz, & Balakrishnan, 1994). These prior distributions were similar to uninformative uniform distributions defined on the entire real line for β_j and on the positive real number line for α. An example input file for BUGS is given in Appendix.

Starting Values

The choice of starting values (e.g., $\omega^{(0)}$) is not generally that critical as the Gibbs sampler should be run long enough to be sufficiently updated from its initial states. It is useful, however, to perform a number of runs using different starting values to verify that the final results are not sensitive to the choice of starting values (Gelman, 1996). Raftery (1996) indicated that extreme starting values could lead to a very long burn-in or stabilization process.

To check the sensitivity of the starting values, three separate runs were performed using the LSAT6 data with three sets of starting values for β_j, $j = 1(1)5$, and α. The three sets of starting values are summarized in Table 2. The first run started at values considered plausible in the light of the usual range of item parameters. The second run and the third run represented substantial deviations in initial values. In particular, the second run was intended to represent a situation in which there was a possibility that items were difficult, and the third run represented an opposite assumption.

Insert Table 2 about here

Each of the three runs consisted of 3,000 iterations. The results for β_1 are presented in Figure 3. The computer program CODA (Best, Cowles, & Vines, 1997) was used to obtain these graphs. The plots in Figure 3 contain the graphical summaries of the Gibbs sampler for β_1. The left plot shows the trace of the sampled values of β_1 for the three runs. In the legend, '0' indicates the initial values for β_j and α were 0 and 1, respectively; '5' indicates the initial values were 5 and 5, respectively; and, '-5' indicates initial values were -5 and .01, respectively. Results for all three runs show that the β_1 generated by the Gibbs sampler quickly settled down regardless of the starting values. The right graph shows the kernel density plot of the three pooled runs of 9,000 values for β_1. The variability among the β_1 values generated by the Gibbs sampler seems not to be too great. The sampled values seem to be concentrated around -2.5. The kernel density plot looks like a normal distribution.

Insert Figure 3 about here

The results for other item parameters were very similar to those from β_1. Overall, the starting values do not appear to affect the final results for the LSAT6 data. Useful starting values for the Rasch model can be found in Molenaar (1995), Gustafsson (1977), and Wright and Stone (1979). Also methods by Baker (1987), Jensema (1976), and Urry (1974) can be used to obtain starting values. Use of good starting values, such as from the above methods, can avoid the time delay required by a lengthy burn in. Our experience with these starting values indicates $\beta_j = 0$ and $\alpha = 1$ will work sufficiently well for applications under the 1PL. In subsequent analyses, therefore, the values, $\beta_j = 0$ and $\alpha = 1$, were used as starting values for LSAT6.

Output Monitoring

A critical issue for MCMC methods is how to determine when one can safely stop sampling and use the results to estimate characteristics of the distributions of the parameters of interest. In this regard, the values for the unknown quantities generated by the Gibbs sampler can be graphically and statistically summarized to check mixing and convergence. The method proposed by Gelman and Rubin (1992) is one of the most popular for monitoring Gibbs sampling. Cowles and Carlin (1996) presented a comparative review of convergence diagnostics for the MCMC algorithms.

We illustrate, here, the use of Gelman and Rubin (1992) statistics on two 3,000 iteration runs. Details of the Gelman and Rubin method are also given in Gelman (1996). Each 3,000 iteration run required about 50 minutes on a Pentium 90 megahertz computer. Monitoring was done using the suite of S-functions called CODA (Best et al., 1997). Gelman-Rubin statistics (i.e., shrink factors) are plotted on Figure 4 for $\beta_1, \ldots, \beta_5$, and α, respectively. For all parameters, the medians were stabilized after about 1,000 iterations.

Insert Figure 4 about here

For each parameter, the Gelman-Rubin statistics estimate the reduction in the pooled estimate of variance if the runs were continued indefinitely. The Gelman-Rubin statistics can be calculated sequentially as the runs proceed. The Gelman-Rubin statistics should be near 1 in order to be reasonably assured that convergence has occurred. Table 3 contains the Gelman-Rubin statistics for LSAT6. The median for β_1, for example, was 1.00 and the 97.5 percentage point was 1.01. The median for α was 1.00 and the 97.5 percentage point was 1.02. These values were very close to 1 indicating that reasonable convergence was realized for all parameters. It is important to notice that the results in Table 3 and the plots in Figure 4 suggest the first 1,000 iterations of each run be discarded and the remaining samples be pooled. We used 1,000 iterations as burn-in and the subsequent 2,000 iterations for the estimation purpose.

Insert Table 3 about here

Item Parameter Estimates

The last step of Gibbs sampling is to obtain summary statistics for the quantities of interest. The posterior mean of the Gibbs sampler can be obtained for each item parameter. The posterior interval as well as the posterior standard deviation can also be obtained for each item parameter from the results of Gibbs sampling. In order to compare item parameter estimates of the LSAT6 items, data were first analyzed via the computer program BUGS (Spiegelhalter et al., 1997) for Gibbs sampling using uninformative prior distributions for item parameters, $\beta_j \sim N(0, 100^2)$ and $\alpha \sim N(0, 100^2)$ with $\alpha > 0$. The starting values of the Gibbs sampler were $\beta_j = 0$ and $\alpha = 1$. There were the first 1,000 burn-in iterations. The subsequent 2,000 iterations were used to obtain posterior means and intervals of the item parameters. All of these were, of course, based on the results of the previous analyses presented earlier in this section. The trace lines of the sampled values and the kernel density plots for LSAT6 item parameters b_j, $j = 1(1)5$, and α are presented in Figure 5. All of the kernel density plots seem to follow the normal distributions.

Insert Figure 5 about here

Table 4 contains the Rasch item parameter estimates (i.e., posterior means for Gibbs sampling) and the 95% posterior intervals for the LSAT6 items. Table 4 also contains the item parameter estimates of the LSAT6 items from the methods of CML, MML, and JML using the computer programs PML (Molenaar, 1990), BILOG (Mislevy & Bock, 1990), and BIGSCALE (Wright, Linacre, & Schultz, 1989), respectively. All default options were used in running the programs. Note that the item parameter estimates from BILOG under MML were initially expressed in terms of the posterior ability metric. The item parameter estimates were transformed onto the usual Rasch model metric (i.e., the metric of either CML or JML with the restriction, $\sum_j b_j = 0$) in order to make the comparison possible.

All in all the item parameter estimates are the same. We also obtained correlations and root mean squared differences between sets of estimates for comparison purposes (see Table 5). The differences occurred mostly in the second or third decimal places. Considering the sizes of the confidence and posterior intervals of the estimates, there seem to be no practical differences in using the item parameter estimates for applications. In terms of confidence intervals, both MML and CML yielded relatively wider intervals than either Gibbs sampling or JML did.

Insert Tables 4 and 5 about here

Ability Parameter Estimates

The Rasch ability estimates and the posterior intervals of the LSAT6 data are reported in Table 6. It is important to notice that Gibbs sampling might yield different posterior means for examinees who have the same response pattern. For example, there were three examinees with the response pattern $(0,0,0,0,1)$ for LSAT6. If we obtain the posterior means for the three examinees, the values will be different (but obviously very similar). In this sense, estimates of the ability parameter from the Gibbs sampling are not unique if we try to obtain them jointly with item parameters. The ability estimates and the posterior intervals reported in Table 5 are, in fact, the average values based on the same raw scores.

Insert Tables 6 and 7 about here

The ability estimates from the methods of CML, MML, and JML can also be found in Table 6. Under MML, ability parameters are estimated after obtaining item parameter estimates and assuming the estimates are the true values. Three estimation methods, maximum likelihood (ML), expected a posteriori (EAP), and maximum a posteriori (MAP), can be used to obtain ability parameter estimates. Since the EAP methods is default in BILOG and since under CML and JML the ability estimates are based on the method of maximum likelihood, both EAP and ML methods were employed under MML. Note that the ability estimates were expressed in the same metric of the item parameter estimates.

In Table 7 it can be noticed that ability estimates from Gibbs sampling and EAP are about the same as both were based on the normal prior (i.e., Bayes methods). CML, MML/ML, and JML yielded very similar ability estimates. Especially, ability estimates and the confidence intervals from CML and MML/ML seem to be more similar each other than those from JML. Clearly, the Bayes ability estimates from both Gibbs sampling and EAP were different from those based on the maximum likelihood methods.

Example 2

Preliminary Analyses

The second example is based on the Memory Test data from Thissen (1982) (see Table 8). The Memory Test data contained 40 examinees responses to the ten items. This example may represent a situation where a small number of examinees' responses to a smaller number of items are to be analyzed under the Rasch model. Model parameters were estimated by Gibbs sampling using the computer program BUGS (Spiegelhalter et al., 1997) under the 1PL model with the same sets of the prior distributions used in the LSAT6 analyses. That is, $\theta_i \sim N(0,1)$, $\beta_j \sim N(0,100^2)$, and $\alpha \sim N(0,100^2)$ with $\alpha > 0$.

Insert Table 8 about here

To check the sensitivity of the starting values for the Memory Test data, three separate runs were performed with three sets of starting values as in Table 2 for $\beta_j, j = 1(1)10$, and α. The three sets of starting values reflected such situations as we have items matched with ability, we have difficult items, and we have easy items, respectively. Each of the three runs consisted of 3,000 iterations.

The results for β_1 are presented in Figure 6. The left plot shows the trace of the sampled values of β_1 from the three runs. Results for all three runs indicated that the β_1 generated by the Gibbs sampler quickly settled down without any visible dependency on the starting values. The right graph shows the kernel density plot of the three pooled runs of 9,000 values for β_1. Variability among the β_1 values generated by the Gibbs sampler was very large, and it might reflect the fact that only 40 examinees were used to estimate parameters. The sampled values were concentrated around -1.5. The distribution did not reveal any bimodality or trimodality. The kernel density seemed to be a normal distribution indicating all three runs yielded similar sets of generated values that equally represented the underlying parameter β_1.

Insert Figure 6 about here

The results for other item parameters were almost the same as those for β_1. Overall the starting values do not appear to affect the final results for the Memory Test. In subsequent analyses for the Memory Test, therefore, $\beta_j = 0$ and $\alpha = 1$ were used as starting values.

The Gelman and Rubin (1992) statistics on two separate 3,000 iteration runs with different random number seeds were used to check mixing and convergence. Gelman-Rubin statistics are plotted on Figure 7 for $\beta_1, \ldots, \beta_{10}$, and α, respectively. In general, the medians were stabilized after about 1,000 iterations for all parameters.

Insert Figure 7 and Table 9 about here

Table 9 contains the Gelman-Rubin statistics of the Memory test. The median for α was 1.00 and the 97.5 percentage point was 1.02. The medians and the 97.5 percentage points for all β_j were 1.00. Reasonable convergence was achieved for all parameters. Note that Figure 7 suggests the first 1,000 iterations of each run be removed and the remaining samples be pooled. The first 1,000 iterations were treated as burn-in and the subsequent 2,000 iterations were used for making inferences.

Item Parameter Estimates

In order to compare item parameter estimates of the Memory Test, data were analyzed via the computer program BUGS (Spiegelhalter et al., 1997) for Gibbs sampling using the uninformative prior distributions for item parameters. Again $\beta_j \sim N(0, 100^2)$ and $\alpha \sim N(0, 100^2)$ with $\alpha > 0$ were used as priors. The starting values for Gibbs sampling were $\beta_j = 0$ and $\alpha = 1$. The last 2,000 iterations were used to obtain posterior means and posterior intervals of the item parameters. The trace lines of the sampled values and the kernel density plots for the Memory Test items parameters are presented in Figure 8. All of the kernel density plots seemed to follow the normal distributions.

Insert Figure 8 and Tables 10 and 11 about here

Table 10 contains the Rasch item parameter estimates and the 95% confidence and posterior intervals for the Memory Test items. Item parameter estimates were expressed in terms of the usual Rasch model metric. The item parameter estimates were very similar (see Table 11). The difference occurred mostly in the second decimal places and, sometimes, in the first decimal places. It can be noticed that the sizes of the confidence and posterior intervals were very large. This might not be surprising because there were only 40 examinees in the data. No practical differences, however, may occur in using these item parameter estimates. In terms of confidence intervals, MML and CML yielded relatively wider intervals than did the other two methods.

Ability Parameter Estimates

The Rasch ability estimates and the confidence and posterior intervals of the Memory Test data are reported in Table 12. Note that most of ability estimates and the posterior intervals from Gibbs sampling reported in Table 12 were the average values based on the same raw scores. The ability estimates from Gibbs sampling and EAP were very similar (see Table 13). CML, MML/ML, and JML yielded very similar ability estimates. The ability estimates and the confidence intervals from CML and MML/ML seemed to be more similar each other than those from JML. The Bayes ability estimates from Gibbs sampling and EAP were different from those based on the maximum likelihood methods.

Insert Tables 12 and 13 about here

Example 3

Preliminary Analyses

The third example used data from Patz and Junker (1997). The data consisted of 3,000 examinees' responses to six short constructed-response items from the 1992 Trial State Assessment in Reading of the National Assessment of Educational Progress (NAEP). According to Patz and Junker (1997), the sample of 3,000 examinees could be considered as a representative random sample of the population of the fourth grade students in the United States. The data provided a situation where a relatively short test was calibrated using a large number of examinees. Item response patterns of the six NAEP items and the numbers of examinees for the respective response patterns are displayed in Table 15. All 64 possible patterns were observed. The calibration was performed using BUGS (Spiegelhalter et al., 1997) under the 1PL. Prior distributions employed in calibration were $\theta_i \sim N(0,1)$, $\beta_j \sim N(0,100^2)$, and $\alpha \sim N(0,100^2)$ with $\alpha > 0$. It was expected that the relatively large sample size of the data would yield item parameter estimates that were not sensitive to the prior specifications because of the dominant effect of the likelihood in the posterior.

Insert Table 14 about here

Before making comparisons of calibration results, the effect of the starting values on the final parameter estimates was investigated for the NAEP data using three sets of starting values as in Table 2. Each of the three runs consisted of 3,000 iterations. Figure 9 illustrates the convergence results of β_1 based on the three calibration runs. Each of the 3,000 iterations yielded very similar results. Regardless of starting values, the trace lines from the left plot were stabilized after just a few iterations. The kernel density plot of the combined 9,000 sampled values of β_1 is also presented in Figure 9. The kernel density plot shows all three starting values yielded the same pattern of sampled values. The density plot seems to follow a normal distribution. The results from other item parameters were very similar to the results of β_1. Overall the starting values did not appear to affect the final results for the NAEP items. The starting values, $\beta_j = 0$ and $\alpha = 1$, were used in the analyses.

Insert Figure 9 about here

In order to check mixing and convergence, Gelman and Rubin (1992) statistics were obtained from the two separate 3,000 iteration runs. Gelman-Rubin statistics are plotted in Figure 10. The medians were stabilized after about 1,000 iterations for all parameters. Hence, the first 1,000 iterations were treated as burn-in and the subsequent 2,000 iterations were used for estimating.

Insert Figure 10 and Table 15 about here

The Gelman-Rubin statistics for the parameters of the six NAEP items are presented in Table 15. The median for α was 1.00 and the 97.5 percentage point was 1.01. The medians for all β_j were 1.00. Three β_j yielded the 97.5 percentage points of 1.00. Two β_j (i.e., β_3 and β_5) yielded the 97.5 percentage points of 1.01. Reasonable convergence was realized for all parameters.

Item Parameter Estimates

In order to compare item parameter estimates of the NAEP items, data were analyzed via the computer program BUGS (Spiegelhalter et al., 1997) for Gibbs sampling using uninformative prior distributions for item parameters. The priors were $\beta_j \sim N(0, 100^2)$ and $\alpha \sim N(0, 100^2)$ with $\alpha > 0$. The starting values for Gibbs sampling were $\beta_j = 0$ and $\alpha = 1$. The last 2,000 iterations were used to obtain posterior means and posterior intervals of the item parameters.

The trace lines of the sampled values and the kernel density plots for the six NAEP items are presented in Figure 11. All of the kernel density plots seem to follow normal distributions.

Insert Figure 11 and Tables 16 and 17 about here

Table 16 contains the Rasch item parameter estimates and the 95% confidence and posterior intervals for the NAEP items from Gibbs sampling, CML, MML, and JML. Item parameter estimates were expressed in terms of the usual Rasch model metric. All four methods yielded almost the same item parameter estimates (see also Table 17). Gibbs sampling and MML yielded an identical set of item parameter estimates. The differences among item parameter estimates across estimation methods occurred mostly in the second decimal places. Gibbs sampling yielded relatively shorter posterior intervals than the other methods. The confidence and posterior intervals of the estimates were very short reflecting the fact that a total of 3,000 examinees were used to calibrate items. MML and CML yielded relatively wider confidence intervals than did the other two methods.

Ability Parameter Estimates

The Rasch ability estimates and the confidence and posterior intervals of the NAEP data are reported in Table 18. Note that the ability estimates and the posterior intervals for Gibbs sampling reported in Table 18 are the average values based on the same raw scores. The ability estimates from Gibbs sampling and EAP were very similar (see Table 19). CML, MML/ML, and JML yielded also very similar ability estimates. Among the maximum likelihood methods, the results from CML and MML/ML were more similar each other than those from JML. JML yielded relatively wider confidence intervals. Gibbs sampling yielded wider posterior intervals than EAP did except for scores 0 and 6. The Bayes ability estimates from Gibbs sampling and EAP were obviously quite different from those obtained from the other three maximum likelihood methods.

Insert Tables 18 and 19 about here

Example 4

Preliminary Analyses

The last Example represented a typical data set for 1PL that contained item responses from 365 examinees for the 31-item English Usage Test. The 1PL model with the prior distributions, $\theta_i \sim N(0, 1)$, $\beta_j \sim N(0, 100^2)$, and $\alpha \sim N(0, 100^2)$ with $\alpha > 0$, was used in Gibbs sampling to estimate item and ability parameters.

To check the sensitivity of the starting values for the Usage Test data, three separate runs were performed with three sets of starting values for $\beta_j, j = 1(1)31$, and α, that were used in the previous Examples (see Table 2). The first starting values reflected a plausible set in the light of the usual range of item parameters. The second set represented a situation we have difficult items. The third implied that we have easy items. Each of the three runs consisted of 3,000 iterations. The results for β_1 are presented in Figure 12. The left plot shows that trace lines of the sampled values of β_1 for the three runs. The values of β_1 generated by the Gibbs sampler quickly settled down without having any visible effects of the starting values. The right graph shows the kernel density plot of the three pooled runs of 9,000 values of β_1. The shape of the kernel density was that of a normal distribution indicating all three runs yielded vary comparable sets that equally represented the underlying parameter β_1.

Insert Figure 12 about here

As the starting values did not appear to affect the final results for the Usage Test, the starting values, $\beta_j = 0$ and $\alpha = 1$, were used in the analyses. In addition, based on the results from the earlier Examples, the first 1,000 iterations were treated as burn-in and the next 2,000 iterations were used to obtain the posterior means and the posterior intervals of the item and ability parameters for Gibbs sampling.

Item Parameter Estimates

Table 20 contains the Rasch item parameter estimates and the 95% confidence and posterior intervals for the Usage Test items. Item parameter estimates were expressed in terms of the usual Rasch model metric. All item parameter estimates were very similar (see also Table 21). The differences among estimates occurred mostly in the second decimal places. In terms of confidence intervals, MML and CML yielded relatively wider intervals than the other two methods did.

Insert Tables 20 and 21 about here

Ability Parameter Estimates

The Rasch ability estimates and the confidence and posterior intervals of the Usage Test are reported in Table 22. Note that the ability estimates and the posterior intervals of Gibbs sampling reported were the average values based on the same raw scores. The ability estimates from Gibbs sampling and EAP were very similar (see Table 23). CML, MML/ML, and JML yielded very similar ability estimates. Bayes ability estimates from Gibbs sampling and EAP were clearly different from those obtained from the maximum likelihood methods.

Insert Tables 22 and 23 about here

Discussion

Previous work with the MCMC method using Gibbs sampling suggests this method may provide a useful alternative method for estimation when small sample sizes and small numbers of items are used. Even though implementation of the Gibbs sampling method in IRT is available in several computer programs, the accuracy of the resulting estimates have not been thoroughly studied. More simulation results should be reported.

The main difference between the Gibbs sampling method and the other estimation methods lies in the way these methods obtain parameter estimates. The Gibbs sampling method uses the sample of parameter values to estimate the mean and variance of the posterior density of the parameter. Under CML and MML, the conditional likelihood function and the marginalized likelihood function are maximized to obtain modes of item parameters. Estimates of the ability parameters do not arise during the course of item parameter estimation under CML and MML. Instead, ability parameters are typically estimated after obtaining the item parameter estimates, assuming the obtained estimates are true values. For the Gibbs sampling method, ability parameters can be estimated jointly with item parameters, similar, in this sense, to JML or joint Bayesian. It is important to know that the ability parameters can also be estimated in Gibbs sampling after obtaining item parameter estimates as in CML or MML, assuming the estimates are true values.

In the above context, one other difference between Gibbs sampling and the other estimation methods is that persons with the same response pattern may produce different ability estimates under Gibbs sampling. Clearly, it is not acceptable. Note that this will occur in a usual case of Gibbs sampling where both item and ability parameters are obtained jointly. We may perform Gibbs sampling initially only to estimate item parameters. After obtaining item parameter estimates, ability parameters can be obtained using a maximum likelihood or Bayesian method. It will remove such an awkward situation where examinees with the same response pattern have different ability estimates.

The estimation of item and ability parameters using Gibbs sampling requires a considerable amount of computing time. This was particularly true for the computer program BUGS used in this study. For example, as noted earlier, one computer run for Gibbs sampling using the LSAT6 data took about 50 minutes, whereas each of the other three estimation methods, MML, CML, and JML, took definitely less than a minute. The computer programs for MML, CML, and JML used in this study are extremely efficient, of course, in comparison to BUGS. One alternative solution may be implementing the Gibbs sampling method using lower level computer languages (e.g., FORTRAN or C++). The iterative nature of Gibbs sampling, however, may prohibit us from seeing a noticeable reduction of computing time.

The Gibbs sampling and general MCMC methods are likely to be more useful for situations where complicated models are employed. For example, Gibbs sampling can be applicable to the estimation of item and ability parameters in the hierarchical Bayes approach (Mislevy, 1986; Swaminathan & Gifford, 1982, 1985, 1986). In this study the priors were imposed directly on the parameters. Accuracy of the Gibbs sampling method with different kinds of priors, perhaps more informative in a Bayesian sense, should be investigated. This kind of research may be particularly valuable for small samples and short tests.

One of the possible advantages of using Gibbs sampling or general MCMC methods, and something to consider in future research on these methods, is incorporation of uncertainly in item parameter estimates into estimation of ability parameters (e.g. Patz & Junker, 1997; Tsutakawa & Johnson, 1990). The data sets used in the four Examples did not clearly exhibit any pronounced effects of errors in item parameter estimates on the ability estimates. This type of investigation can be performed in the context of simulation (e.g., Hulin, Lissak, & Drasgow, 1982). Additional simulation studies may reveal whether such incorporation is, in fact, valuable in the context of the Rasch model.

In this paper, The Rasch model was used without addressing the problem of model selection and criticism (e.g., the choice of the linking function, model fit). The model criticism for Gibbs sampling seems to be an important topic to investigate in future research. Also the evaluation of the Gibbs sampling method to other IRT models, for example, other logistic or probit models for binary items, the partial credit model, the graded response model, and the linear logistic test model, may provide guidelines for using the method under IRT.

Finally, it should be noted that the computer programs BUGS (Spiegelhalter et al., 1997) and CODA (Best et al., 1997) as well as the accompanying manuals are freely available over the Web. The uniform resource locator (URL) of the Medical Research Council Biostatistics Unit at the University of Cambridge is:

`http://www.mrc-bsu.cam.ac.uk/bugs/`

References

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics, 17,* 251–269.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistics Society, Series B, 32,* 283–301.

Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society, Series B, 34,* 42-54.

Andersen, E. B. (1980). *Discrete statistical models with social science applications.* Amsterdam: North-Holland.

Baker, F. B. (1987). Item parameter estimation via minimum logit chi-square. *British Journal of Mathematical and Statistical Psychology, 40,* 50–60.

Baker, F. B. (1998). An investigation of the item parameter recovery characteristics of a Gibbs sampling approach. *Applied Psychological Measurement, 22,* 153–169.

Baker, F. B., & Harwell, M. R. (1996). Computing elementary symmetric functions and their derivatives: A didactic. *Applied Psychological Measurement, 20,* 169–192.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian theory.* Chichester, England: Wiley.

Best, N. G., Cowles, M. K., & Vines, S. K. (1997). CODA: Convergence diagnosis and output analysis software for Gibbs sampling output (Version 0.4) [Computer software]. Cambridge, UK: University of Cambridge, Institute of Public Health, Medical Research Council Biostatistics Unit.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 395–479). Reading, MA: Addison-Wesley.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Applications of an EM algorithm. *Psychometrika, 46,* 443–459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35,* 179–197.

Carlin, B. P., & Louis, T. A. (1996). *Bayes and empirical Bayes methods for data analysis.* London: Chapman & Hall.

Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association, 91,* 883–904.

Fischer, G. H., & Molenaar, I. W. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications.* New York: Springer-Verlag.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85,* 398–409.

Gelman, A. (1996). Inference and monitoring convergence. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 131–143). London: Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* London: Chapman & Hall.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science, 7,* 457–511.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6,* 721–741.

Gilks, W. R. (1996). Full conditional distribution. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 75–88). London: Chapman & Hall.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice.* London: Chapman & Hall.

Gilks, W. R., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics, 41,* 337–348.

Gustafsson, J.-E. (1977). *The Rasch model for dichotomous items: Theory, applications and a computer program.* Sweden: Göteborg University, Institute of Education.

Hasting, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57,* 97–109.

Hoijtink, H., & Boomsma, A. (1995). On person parameter estimation in the dichotomous Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 53–68). New York: Springer-Verlag.

Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement, 6,* 249–260.

Jensema, C. (1976). A simple technique for estimating latent trait mental test parameters. *Educational and Psychological Measurement, 36,* 705–715.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions* (2nd ed., Vol. 1). New York: Wiley.

Kim, S.-H., Cohen, A. S., Baker, F. B., Subkoviak, M. J., & Leonard, T. (1994). An investigation of hierarchical Bayes procedures in item response theory. *Psychometrika, 59,* 405–421.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N., & Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks, 20,* 491–505.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Lord, F. M. (1986). Maximum likelihood and Bayesian parameter estimation in item response theory. *Journal of Educational Measurement, 23,* 157–162.

MathSoft, Inc. (1995). S-PLUS (Version 3.3 for Windows) [Computer software]. Seattle, WA: Author.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics, 21,* 1087–1092.

Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association, 44,* 335–341.

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51,* 177–195.

Mislevy, R. J., & Bock, R. D. (1990). BILOG 3: Item analysis and test scoring with binary logistic models [Computer software]. Mooresville, IN: Scientific Software.

Molenaar, I. W. (1990). PML user's manual, PC version [Computer software]. Groningen, The Netherlands: ProGAMMA.

Molenaar, I. W. (1995). Estimation of item parameters. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 39–51). New York: Springer-Verlag.

Patz, R. J., & Junker, B. W. (1997). *A straightforward approach to Markov chain Monte Carlo methods for item response models* (Tech. Rep. No. 658). Pittsburgh, PA: Carnegie Mellon University, Department of Statistics.

Raftery, A. E. (1996). Hypothesis testing and model selection. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 163–187). London: Chapman & Hall.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment test.* Chicago: The University of Chicago Press. (Original work published 1960)

Spiegelhalter, D. J., Best, N. G., Gilks, W. R., & Inskip, H. (1996). Hepatitis B: a case study in MCMC methods. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice* (pp. 21–43). London: Chapman & Hall.

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L., & Cowell, R. G. (1993). Bayesian analysis in expert systems (with discussion). *Statistical Science, 8,* 219–283.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (1996). *BUGS 0.5 examples* (Vol. 1, Version i). Cambridge, UK: University of Cambridge, Institute of Public Health, Medical Research Council Biostatistics Unit.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (1997). BUGS: Bayesian inference using Gibbs sampling (Version 0.6) [Computer software]. Cambridge, UK: University of Cambridge, Institute of Public Health, Medical Research Council Biostatistics Unit.

Stroud, A. H., & Secrest, D. (1966). *Gaussian quadrature formulas.* Englewood Cliff, NJ: Prentice-Hall.

Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics, 7,* 175–191.

Swaminathan, H., & Gifford, J. A. (1985). Bayesian estimation in the two-parameter logistic model. *Psychometrika, 50,* 349–364.

Swaminathan, H., & Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika, 51,* 581–601.

Tanner, M. A. (1996). *Tools for statistical inference: Methods for the exploration of posterior distributions and likelihood functions* (2nd ed.). New York: Springer-Verlag.

The MathWorks, Inc. (1996). MATLAB: The language of technical computing [Computer software]. Natick, MA: Author.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika, 47,* 175–186.

Tsutakawa, R. K., & Lin, H. Y. (1986). Bayesian estimation of item response curves. *Psychometrika, 51,* 251–267.

Tsutakawa, R. K., & Johnson, J. C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika, 55,* 371–390.

Urry, V. W. (1974). Approximations to item parameters of mental test models and their uses. *Educational and Psychological Measurement, 34,* 253–269.

Whittaker, J. (1990). *Graphical models in applied multivariate analysis.* Chichester: Wiley.

Wright, B. D., Linacre, J. M., & Schultz, M. (1989). A user's guide to BIGSCALE: Rasch-model rating scale analysis computer program [Computer software]. Chicago: MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago: MESA Press.

Table 1
LSAT6 Data of Bock and Lieberman (1970) with 32 Response Patterns

Index	Item Pattern 1	2	3	4	5	Observed Freuqency
1	0	0	0	0	0	3
2	0	0	0	0	1	6
3	0	0	0	1	0	2
4	0	0	0	1	1	11
5	0	0	1	0	0	1
6	0	0	1	0	1	1
7	0	0	1	1	0	3
8	0	0	1	1	1	4
9	0	1	0	0	0	1
10	0	1	0	0	1	8
11	0	1	0	1	0	0
12	0	1	0	1	1	16
13	0	1	1	0	0	0
14	0	1	1	0	1	3
15	0	1	1	1	0	2
16	0	1	1	1	1	15
17	1	0	0	0	0	10
18	1	0	0	0	1	29
19	1	0	0	1	0	14
20	1	0	0	1	1	81
21	1	0	1	0	0	3
22	1	0	1	0	1	28
23	1	0	1	1	0	15
24	1	0	1	1	1	80
25	1	1	0	0	0	16
26	1	1	0	0	1	56
27	1	1	0	1	0	21
28	1	1	0	1	1	173
29	1	1	1	0	0	11
30	1	1	1	0	1	61
31	1	1	1	1	0	28
32	1	1	1	1	1	298

Table 2
Starting Values for Item Parameters in the Three Runs of the Gibbs Sampler

Parameter	Run First	Second	Third
$\beta_j, j = 1(1)5$	0	5	−5
α	1	5	.01

Table 3
Gelman-Rubin Statistics for the Parameters of the LSAT6 Items

Parameter	Shrink Factor Estimate	97.5 Percentile
β_1	1.00	1.01
β_2	1.00	1.01
β_3	1.00	1.01
β_4	1.00	1.01
β_5	1.00	1.01
α	1.00	1.02

Table 4

Estimated Item Parameters and 95% Confidence/Posterior Intervals of the LSAT6 Items from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML), and Joint Maximum Likelihood (JML)

Item	Gibbs Sampling[a]		CML		MML[a]		JML	
	Difficulty	Conf. Interval	Difficulty	Conf. Interval	Difficulty	Conf. Interval	Difficulty	Conf. Interval
1	−1.26	(−1.47, −1.05)	−1.26	(−1.49, −1.02)	−1.26	(−1.51, −1.00)	−1.24	(−1.46, −1.02)
2	.48	(.34, .62)	.47	(.31, .64)	.48	(.32, .63)	.45	(.31, .59)
3	1.24	(1.11, 1.37)	1.24	(1.08, 1.40)	1.24	(1.09, 1.38)	1.30	(1.16, 1.44)
4	.17	(.02, .31)	.17	(.00, .34)	.17	(.00, .34)	.13	(−.01, .27)
5	−.63	(−.79, −.47)	−.62	(−.82, −.43)	−.62	(−.83, −.42)	−.64	(−.80, −.48)

[a]The restriction, $\Sigma_j b_j = 0$, has been applied.

Table 5

Correlations (Lower Triangle) and Root Mean Squared Differences (Upper Triangle) of the LSAT6 Item Parameter Estimates from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML), and Joint Maximum Likelihood (JML)

Method	Gibbs Sampling	CML	MML	JML
Gibbs Sampling		.006	.004	.036
CML	1.000		.004	.036
MML	1.000	1.000		.037
JML	.999	.999	.999	

Table 6

Ability Estimates and 95% Confidence/Posterior Intervals of the LSAT6 Data from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML) with Maximum Likelihood (ML) and Expected A Posteriori (EAP), and Joint Maximum Likelihood (JML)

Score	Gibbs Sampling[a]		CML		MML[a] ML		MML[a] EAP		JML	
	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval
0	.02	(−1.15, 1.18)					.03	(−1.14, 1.21)		
1	.39	(−.79, 1.60)	−1.60	(−3.92, .71)	−1.55	(***, ***[b])	.40	(−.78, 1.58)	−1.72	(−4.08, .65)
2	.76	(−.42, 1.98)	−.47	(−2.41, 1.47)	−.47	(−2.41, 1.47)	.76	(−.43, 1.96)	−.52	(−2.54, 1.50)
3	1.15	(−.06, 2.39)	.48	(−1.45, 2.42)	.48	(−1.45, 2.41)	1.14	(−.07, 2.36)	.52	(−1.50, 2.54)
4	1.54	(.31, 2.81)	1.60	(−.71, 3.91)	1.60	(−.71, 3.91)	1.54	(.29, 2.78)	1.72	(−.65, 4.09)
5	1.96	(.69, 3.26)					1.95	(.67, 3.23)		

[a]The restriction, $\Sigma_j b_j = 0$, has been applied.

[b]Improper values were obtained.

Table 7

Correlations (Lower Triangle) and Root Mean Squared Differences (Upper Triangle) of the LSAT6 Ability Estimates from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML) with Maximum Likelihood (ML) and Expected A Posteriori (EAP), and Joint Maximum Likelihood (JML)

Method	Gibbs Sampling	CML	MML ML	MML EAP	JML
Gibbs Sampling		1.217	1.197	.008	1.277
CML	.999		.025	1.220	.091
MML/ML	.927	.941		1.199	.109
MML/EAP	1.000	.999	.926		1.280
JML	.999	1.000	.939	.999	

BEST COPY AVAILABLE

Table 8
Ten-Item Memory Test Data from Thissen (1982) with 31 Response Patterns

	Item Pattern										Observed
Index	1	2	3	4	5	6	7	8	9	10	Freuqency
1	0	0	0	0	0	0	0	0	0	0	5
2	0	0	0	0	0	0	0	0	0	1	1
3	0	0	0	0	0	0	0	0	1	1	3
4	0	0	0	0	0	0	0	1	0	1	2
5	0	0	0	0	0	1	0	0	0	1	1
6	0	0	0	0	1	0	0	0	0	1	1
7	0	0	0	0	1	0	0	0	1	0	1
8	0	0	1	0	0	0	0	0	0	1	1
9	0	0	0	0	0	0	0	1	1	1	2
10	0	0	0	0	0	0	1	0	1	1	1
11	0	0	1	0	0	0	0	1	0	1	1
12	0	0	1	0	0	0	1	0	0	1	1
13	0	1	0	0	0	1	0	1	0	0	1
14	1	0	0	0	0	0	0	0	1	1	1
15	1	0	0	0	0	0	1	0	0	1	1
16	1	0	0	1	0	0	0	0	1	0	1
17	0	0	0	0	0	0	1	1	1	1	1
18	0	0	0	0	0	1	0	1	1	1	2
19	0	0	0	0	1	0	1	0	1	1	1
20	0	0	0	1	0	0	1	0	1	1	1
21	0	0	0	1	0	0	1	1	0	1	1
22	0	1	0	0	0	0	0	1	1	1	1
23	0	1	0	0	0	1	0	0	1	1	1
24	0	1	0	0	1	0	0	1	1	0	1
25	0	1	0	0	0	0	1	1	1	1	1
26	1	0	0	0	0	1	1	1	0	1	1
27	1	0	0	1	1	0	1	1	0	0	1
28	1	1	0	0	1	0	0	1	0	1	1
29	0	1	0	0	0	1	1	1	1	1	1
30	1	1	0	0	1	1	0	1	0	1	1
31	0	1	1	1	1	0	0	1	1	1	1

Table 9
Gelman-Rubin Statistics for the Parameters of the Memory Test Items

	Shrink Factor	
Parameter	Estimate	97.5 Percentile
β_1	1.00	1.00
β_2	1.00	1.00
β_3	1.00	1.00
β_4	1.00	1.00
β_5	1.00	1.00
β_6	1.00	1.00
β_7	1.00	1.00
β_8	1.00	1.00
β_9	1.00	1.00
β_{10}	1.00	1.00
α	1.00	1.02

BEST COPY AVAILABLE

Table 10

Estimated Item Parameters and 95% Confidence/Posterior Intervals of the Memory Test Items from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML), and Joint Maximum Likelihood (JML)

	Gibbs Sampling[a]		CML		MML[a]		JML	
Item	Difficulty	Conf. Interval	Difficulty	Conf. Interval	Difficulty	Conf. Interval	Difficulty	Conf. Interval
1	.70	(−.08, 1.51)	.66	(−.19, 1.51)	.65	(−.37, 1.67)	.68	(−.10, 1.46)
2	.31	(−.40, 1.09)	.33	(−.46, 1.12)	.31	(−.67, 1.30)	.35	(−.38, 1.08)
3	1.44	(.51, 2.61)	1.33	(.30, 2.36)	1.34	(.08, 2.60)	1.34	(.36, 2.32)
4	1.12	(.21, 2.11)	1.07	(.12, 2.02)	1.07	(−.05, 2.20)	1.09	(.19, 1.99)
5	.50	(−.25, 1.33)	.49	(−.33, 1.31)	.47	(−.48, 1.43)	.51	(−.23, 1.25)
6	.49	(−.25, 1.28)	.49	(−.33, 1.31)	.47	(−.50, 1.45)	.51	(−.23, 1.25)
7	.01	(−.68, .73)	.05	(−.71, .80)	.02	(−.82, .85)	.06	(−.63, .75)
8	−1.01	(−1.68, −.34)	−.91	(−1.62, −.20)	−.96	(−1.75, −.16)	−.93	(−1.58, −.28)
9	−1.13	(−1.79, −.44)	−1.02	(−1.74, −.31)	−1.07	(−1.93, −.21)	−1.05	(−1.70, −.40)
10	−2.43	(−3.26, −1.69)	−2.49	(−3.38, −1.59)	−2.31	(−3.25, −1.36)	−2.58	(−3.46, −1.70)

[a]The restriction, $\Sigma_j b_j = 0$, has been applied.

Table 11

Correlations (Lower Triangle) and Root Mean Squared Differences (Upper Triangle) of the Memory Test Item Parameter Estimates from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML), and Joint Maximum Likelihood (JML)

Method	Gibbs Sampling	CML	MML	JML
Gibbs Sampling		.066	.061	.072
CML	.999		.063	.034
MML	1.000	.999		.090
JML	.998	1.000	.998	

Table 12

Ability Estimates and 95% Confidence/Posterior Intervals of the Memory Test Data from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML) with Maximum Likelihood (ML) and Expected A Posteriori (EAP), and Joint Maximum Likelihood (JML)

					MML[a]					
	Gibbs Sampling[a]		CML		ML		EAP		JML	
Score	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval
0	−2.24	(−3.63, −.89)					−2.05	(−3.23, −.88)		
1	−1.84	(−3.17, −.58)	−2.71	(−4.98, −.44)	−2.67	(−4.91, −.44)	−1.70	(−2.84, −.56)	−2.78	(−5.05, −.51)
2	−1.47	(−2.74, −.23)	−1.69	(−3.45, .06)	−1.69	(−3.42, .05)	−1.37	(−2.48, −.27)	−1.72	(−3.48, .04)
3	−1.11	(−2.33, .07)	−1.00	(−2.54, .54)	−1.00	(−2.53, .52)	−1.06	(−2.14, .02)	−.99	(−2.54, .56)
4	−.78	(−2.01, .39)	−.43	(−1.86, 1.00)	−.44	(−1.86, .98)	−.76	(−1.82, .30)	−.41	(−1.84, 1.02)
5	−.46	(−1.68, .71)	.08	(−1.30, 1.45)	.06	(−1.31, 1.44)	−.48	(−1.52, .57)	.11	(−1.26, 1.48)
6	−.15	(−1.31, .99)	.57	(−.81, 1.95)	.56	(−.83, 1.94)	−.20	(−1.23, .83)	.60	(−.77, 1.97)
7	.17	(−.99, 1.35)	1.09	(−.36, 2.53)	1.07	(−.38, 2.52)	.08	(−.95, 1.10)	1.12	(−.31, 2.55)

[a]The restriction, $\Sigma_j b_j = 0$, has been applied.

Table 13

Correlations (Lower Triangle) and Root Mean Squared Differences (Upper Triangle) of the Memory Test Ability Estimates from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML) with Maximum Likelihood (ML) and Expected A Posteriori (EAP), and Joint Maximum Likelihood (JML)

			MML		
Method	Gibbs Sampling	CML	ML	EAP	JML
Gibbs Sampling		.609	.592	.100	.642
CML	.995		.019	.672	.036
MML/ML	.995	1.000		.654	.054
MML/EAP	1.000	.994	.995		.223
JML	.994	1.000	1.000	.993	

Table 14

NAEP Data in Patz and Junker (1997) with 64 Response Patterns

Index	Item Pattern 1	2	3	4	5	6	Observed Frequency	Index	Item Pattern 1	2	3	4	5	6	Observed Freuqency
1	0	0	0	0	0	0	145	33	1	0	0	0	0	0	13
2	0	0	0	0	0	1	44	34	1	0	0	0	0	1	8
3	0	0	0	0	1	0	49	35	1	0	0	0	1	0	7
4	0	0	0	0	1	1	16	36	1	0	0	0	1	1	5
5	0	0	0	1	0	0	13	37	1	0	0	1	0	0	4
6	0	0	0	1	0	1	2	38	1	0	0	1	0	1	1
7	0	0	0	1	1	0	17	39	1	0	0	1	1	0	3
8	0	0	0	1	1	1	6	40	1	0	0	1	1	1	5
9	0	0	1	0	0	0	141	41	1	0	1	0	0	0	22
10	0	0	1	0	0	1	49	42	1	0	1	0	0	1	9
11	0	0	1	0	1	0	79	43	1	0	1	0	1	0	20
12	0	0	1	0	1	1	45	44	1	0	1	0	1	1	16
13	0	0	1	1	0	0	22	45	1	0	1	1	0	0	3
14	0	0	1	1	0	1	14	46	1	0	1	1	0	1	1
15	0	0	1	1	1	0	21	47	1	0	1	1	1	0	10
16	0	0	1	1	1	1	18	48	1	0	1	1	1	1	11
17	0	1	0	0	0	0	157	49	1	1	0	0	0	0	34
18	0	1	0	0	0	1	47	50	1	1	0	0	0	1	16
19	0	1	0	0	1	0	104	51	1	1	0	0	1	0	36
20	0	1	0	0	1	1	65	52	1	1	0	0	1	1	33
21	0	1	0	1	0	0	37	53	1	1	0	1	0	0	6
22	0	1	0	1	0	1	28	54	1	1	0	1	0	1	3
23	0	1	0	1	1	0	32	55	1	1	0	1	1	0	20
24	0	1	0	1	1	1	40	56	1	1	0	1	1	1	30
25	0	1	1	0	0	0	265	57	1	1	1	0	0	0	40
26	0	1	1	0	0	1	106	58	1	1	1	0	0	1	33
27	0	1	1	0	1	0	202	59	1	1	1	0	1	0	60
28	0	1	1	0	1	1	177	60	1	1	1	0	1	1	98
29	0	1	1	1	0	0	64	61	1	1	1	1	0	0	19
30	0	1	1	1	0	1	46	62	1	1	1	1	0	1	26
31	0	1	1	1	1	0	107	63	1	1	1	1	1	0	50
32	0	1	1	1	1	1	93	64	1	1	1	1	1	1	107

Table 15

Gelman-Rubin Statistics for the Parameters of the NAEP Items

Parameter	Shrink Factor Estimate	97.5 Percentile
β_1	1.00	1.00
β_2	1.00	1.00
β_3	1.00	1.01
β_4	1.00	1.00
β_5	1.00	1.01
β_6	1.00	1.00
α	1.00	1.01

Table 16

Estimated Item Parameters and 95% Confidence/Posterior Intervals of the NAEP Items from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML), and Joint Maximum Likelihood (JML)

	Gibbs Sampling[a]		CML		MML[a]		JML	
Item	Difficulty	Conf. Interval	Difficulty	Conf. Interval	Difficulty	Conf. Interval	Difficulty	Conf. Interval
1	1.14	(1.07, 1.23)	1.15	(1.05, 1.24)	1.14	(1.05, 1.24)	1.13	(1.05, 1.24)
2	−1.27	(−1.34, −1.19)	−1.26	(−1.35, −1.17)	−1.27	(−1.36, −1.17)	−1.25	(−1.33, −1.17)
3	−.89	(−.97, −.82)	−.89	(−.98, −.81)	−.89	(−.98, −.81)	−.88	(−.96, −.81)
4	.93	(.85, 1.00)	.93	(.84, 1.02)	.93	(.84, 1.02)	.92	(.84, 1.00)
5	−.26	(−.33, −.19)	−.26	(−.35, −.18)	−.26	(−.34, −.17)	−.25	(−.33, −.17)
6	.35	(.28, .42)	.34	(.26, .43)	.35	(.26, .43)	.34	(.26, .42)

[a]The restriction, $\Sigma_j b_j = 0$, has been applied.

Table 17

Correlations (Lower Triangle) and Root Mean Squared Differences (Upper Triangle) of the NAEP Item Parameter Estimates from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML), and Joint Maximum Likelihood (JML)

Method	Gibbs Sampling	CML	MML	JML
Gibbs Sampling		.007	.000	.012
CML	1.000		.007	.011
MML	1.000	1.000		.012
JML	1.000	1.000	1.000	

Table 18

Ability Estimates and 95% Confidence/Posterior Intervals of the NAEP Data from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML) with Maximum Likelihood (ML) and Expected A Posteriori (EAP), and Joint Maximum Likelihood (JML)

					MML[a]					
	Gibbs Sampling[a]		CML		ML		EAP		JML	
Score	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval
0	−1.29	(−2.65, −.03)					−1.33	(−2.65, −.01)		
1	−.88	(−2.20, .40)	−1.86	(−4.12, .39)	−1.87	(−4.12, .39)	−.89	(−2.17, .39)	−1.96	(−4.25, .33)
2	−.49	(−1.77, .75)	−.82	(−2.66, 1.02)	−.82	(−2.66, 1.02)	−.47	(−1.73, .78)	−.87	(−2.77, 1.03)
3	−.06	(−1.44, 1.17)	.01	(−1.75, 1.76)	.01	(−1.75, 1.76)	−.07	(−1.31, 1.18)	.01	(−1.79, 1.81)
4	.35	(−.92, 1.64)	.83	(−1.01, 2.66)	.83	(−1.01, 2.66)	.34	(−.91, 1.59)	.88	(−1.00, 2.76)
5	.74	(−.56, 2.07)	1.86	(−.38, 4.10)	1.86	(−.38, 4.11)	.75	(−.52, 2.02)	1.96	(−.31, 4.23)
6	1.18	(−.07, 2.45)					1.18	(−.12, 2.48)		

[a]The restriction, $\Sigma_j b_j = 0$, has been applied.

Table 19

Correlations (Lower Triangle) and Root Mean Squared Differences (Upper Triangle) of the NAEP Ability Estimates from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML) with Maximum Likelihood (ML) and Expected A Posteriori (EAP), and Joint Maximum Likelihood (JML)

			MML		
Method	Gibbs Sampling	CML	ML	EAP	JML
Gibbs Sampling		.715	.718	.018	.785
CML	.998		.004	.713	.071
MML/ML	.998	1.000		.716	.068
MML/EAP	1.000	.999	.999		.783
JML	.998	1.000	1.000	.999	

BEST COPY AVAILABLE

Table 20

Estimated Item Parameters and 95% Confidence/Posterior Intervals of the English Usage Items from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML), and Joint Maximum Likelihood (JML)

	Gibbs Sampling[a]		CML		MML[a]		JML	
Item	Difficulty	Conf. Interval	Difficulty	Conf. Interval	Difficulty	Conf. Interval	Difficulty	Conf. Interval
1	−2.67	(−3.12, −2.24)	−2.68	(−3.13, −2.23)	−2.65	(−3.13, −2.17)	−2.68	(−3.13, −2.23)
2	1.15	(.92, 1.38)	1.14	(.90, 1.38)	1.14	(.90, 1.38)	1.14	(.90, 1.38)
3	1.20	(.96, 1.44)	1.20	(.96, 1.43)	1.20	(.97, 1.43)	1.20	(.96, 1.44)
4	1.92	(1.67, 2.17)	1.91	(1.64, 2.18)	1.91	(1.66, 2.16)	1.92	(1.67, 2.17)
5	−.97	(−1.23, −.73)	−.96	(−1.23, −.70)	−.97	(−1.24, −.70)	−.97	(−1.22, −.72)
6	−.62	(−.87, −.37)	−.61	(−.86, −.36)	−.62	(−.87, −.36)	−.62	(−.86, −.38)
7	−.65	(−.91, −.41)	−.64	(−.89, −.39)	−.65	(−.90, −.39)	−.65	(−.89, −.41)
8	.40	(.17, .62)	.39	(.16, .62)	.39	(.17, .62)	.39	(.17, .61)
9	.82	(.59, 1.04)	.81	(.58, 1.04)	.81	(.59, 1.04)	.81	(.59, 1.03)
10	.53	(.30, .76)	.52	(.29, .75)	.52	(.29, .76)	.52	(.30, .74)
11	.17	(−.05, .39)	.17	(−.06, .40)	.17	(−.06, .39)	.17	(−.05, .39)
12	−.37	(−.61, −.13)	−.37	(−.61, −.13)	−.37	(−.61, −.14)	−.37	(−.61, −.13)
13	1.00	(.77, 1.22)	.99	(.75, 1.22)	.99	(.77, 1.21)	.99	(.75, 1.23)
14	−1.55	(−1.85, −1.27)	−1.55	(−1.86, −1.24)	−1.55	(−1.85, −1.25)	−1.55	(−1.84, −1.26)
15	1.29	(1.05, 1.52)	1.28	(1.04, 1.52)	1.29	(1.03, 1.54)	1.29	(1.05, 1.53)
16	1.09	(.88, 1.30)	1.08	(.85, 1.32)	1.09	(.86, 1.31)	1.09	(.85, 1.33)
17	−.87	(−1.12, −.62)	−.86	(−1.12, −.60)	−.86	(−1.13, −.60)	−.86	(−1.11, −.61)
18	−.59	(−.83, −.34)	−.58	(−.83, −.33)	−.58	(−.82, −.35)	−.58	(−.82, −.34)
19	−.64	(−.89, −.40)	−.64	(−.89, −.39)	−.65	(−.90, −.39)	−.65	(−.89, −.41)
20	−1.28	(−1.57, −1.00)	−1.28	(−1.56, −.99)	−1.28	(−1.58, −.98)	−1.28	(−1.55, −1.01)
21	.17	(−.05, .39)	.17	(−.06, .40)	.17	(−.06, .40)	.17	(−.05, .39)
22	.77	(.55, .98)	.76	(.53, .99)	.76	(.53, .99)	.76	(.54, .98)
23	.54	(.32, .76)	.55	(.32, .78)	.55	(.33, .77)	.55	(.33, .77)
24	.76	(.53, .98)	.76	(.53, .99)	.76	(.52, 1.00)	.76	(.54, .98)
25	−1.92	(−2.26, −1.59)	−1.91	(−2.25, −1.56)	−1.90	(−2.26, −1.55)	−1.91	(−2.24, −1.58)
26	−.54	(−.78, −.31)	−.53	(−.78, −.29)	−.54	(−.79, −.28)	−.54	(−.78, −.30)
27	.06	(−.18, .29)	.06	(−.17, .29)	.06	(−.17, .29)	.06	(−.16, .28)
28	−.46	(−.70, −.23)	−.46	(−.70, −.22)	−.46	(−.70, −.22)	−.46	(−.70, −.22)
29	2.17	(1.88, 2.44)	2.16	(1.88, 2.44)	2.15	(1.87, 2.43)	2.17	(1.90, 2.44)
30	−.08	(−.32, .14)	−.07	(−.31, .16)	−.08	(−.31, .16)	−.08	(−.32, .16)
31	−.80	(−1.06, −.54)	−.79	(−1.05, −.53)	−.79	(−1.06, −.53)	−.79	(−1.04, −.54)

[a]The restriction, $\Sigma_j b_j = 0$, has been applied.

Table 21

Correlations (Lower Triangle) and Root Mean Squared Differences (Upper Triangle) of the English Usage Item Parameter Estimates from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML), and Joint Maximum Likelihood (JML)

Method	Gibbs Sampling	CML	MML	JML
Gibbs Sampling		.008	.009	.006
CML	1.000		.008	.005
MML	1.000	1.000		.007
JML	1.000	1.000	1.000	

BEST COPY AVAILABLE

Table 22

Ability Estimates and 95% Confidence/Posterior Intervals of the English Usage Test Data from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML) with Maximum Likelihood (ML) and Expected A Posteriori (EAP), and Joint Maximum Likelihood (JML)

	Gibbs Sampling[a]		CML		MML[a]				JML	
					ML		EAP			
Score	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval	Ability	Conf. Interval
0	−2.69	(−3.90, −1.68)					−2.65	(−3.64, −1.66)		
2	−2.17	(−3.16, −1.29)	−3.17	(−4.68, −1.67)	−3.17	(−4.67, −1.67)	−2.14	(−3.11, −1.16)	−3.20	(−4.71, −1.69)
4	−1.74	(−2.66, −.93)	−2.31	(−3.44, −1.18)	−2.31	(−3.44, −1.18)	−1.73	(−2.53, −.92)	−2.33	(−3.47, −1.19)
5	−1.55	(−2.45, −.71)	−2.00	(−3.04, −.97)	−2.00	(−3.04, −.97)	−1.55	(−2.39, −.72)	−2.03	(−3.07, −.99)
6	−1.37	(−2.27, −.56)	−1.74	(−2.72, −.77)	−1.74	(−2.71, −.77)	−1.36	(−2.26, −.46)	−1.76	(−2.74, −.78)
7	−1.20	(−2.04, −.39)	−1.51	(−2.43, −.58)	−1.51	(−2.43, −.58)	−1.15	(−2.02, −.28)	−1.52	(−2.44, −.60)
8	−1.03	(−1.87, −.26)	−1.29	(−2.18, −.41)	−1.30	(−2.18, −.41)	−.98	(−1.73, −.23)	−1.31	(−2.19, −.43)
9	−.86	(−1.63, −.10)	−1.10	(−1.96, −.24)	−1.10	(−1.96, −.24)	−.85	(−1.50, −.21)	−1.11	(−1.97, −.25)
10	−.71	(−1.48, .07)	−.91	(−1.75, −.07)	−.91	(−1.75, −.07)	−.74	(−1.41, −.07)	−.92	(−1.76, −.08)
11	−.58	(−1.37, .19)	−.73	(−1.55, .09)	−.73	(−1.55, .09)	−.61	(−1.40, .18)	−.74	(−1.56, .08)
12	−.42	(−1.20, .32)	−.56	(−1.37, .25)	−.56	(−1.37, .25)	−.42	(−1.30, .45)	−.57	(−1.37, .23)
13	−.27	(−1.01, .48)	−.39	(−1.19, .40)	−.39	(−1.19, .40)	−.23	(−1.06, .60)	−.40	(−1.20, .40)
14	−.13	(−.88, .58)	−.23	(−1.02, .56)	−.23	(−1.02, .56)	−.08	(−.77, .62)	−.23	(−1.01, .55)
15	.02	(−.73, .77)	−.07	(−.85, .72)	−.07	(−.86, .72)	.03	(−.56, .62)	−.07	(−.85, .71)
16	.15	(−.57, .89)	.09	(−.69, .88)	.09	(−.69, .88)	.12	(−.49, .73)	.09	(−.69, .87)
17	.30	(−.44, 1.09)	.26	(−.53, 1.04)	.26	(−.53, 1.04)	.24	(−.51, .98)	.26	(−.52, 1.04)
18	.45	(−.34, 1.21)	.42	(−.38, 1.21)	.42	(−.38, 1.21)	.41	(−.45, 1.27)	.42	(−.38, 1.22)
19	.57	(−.20, 1.33)	.58	(−.22, 1.39)	.58	(−.22, 1.39)	.61	(−.25, 1.46)	.59	(−.21, 1.39)
20	.74	(−.01, 1.47)	.75	(−.06, 1.57)	.75	(−.06, 1.57)	.78	(.03, 1.52)	.76	(−.06, 1.58)
21	.89	(.10, 1.64)	.93	(.10, 1.76)	.93	(.10, 1.76)	.90	(.25, 1.55)	.94	(.10, 1.78)
22	1.04	(.28, 1.83)	1.11	(.26, 1.97)	1.11	(.26, 1.97)	1.01	(.33, 1.69)	1.13	(.27, 1.99)
23	1.20	(.42, 1.97)	1.31	(.43, 2.19)	1.31	(.43, 2.19)	1.16	(.35, 1.96)	1.32	(.44, 2.20)
24	1.37	(.56, 2.20)	1.52	(.60, 2.43)	1.52	(.60, 2.43)	1.35	(.45, 2.25)	1.53	(.61, 2.45)
25	1.56	(.71, 2.42)	1.75	(.79, 2.71)	1.75	(.79, 2.71)	1.56	(.68, 2.44)	1.76	(.80, 2.72)
26	1.72	(.88, 2.61)	2.00	(.98, 3.02)	2.00	(.98, 3.02)	1.75	(.92, 2.57)	2.02	(1.00, 3.04)
27	1.95	(1.05, 2.90)	2.30	(1.18, 3.41)	2.30	(1.18, 3.41)	1.92	(1.08, 2.77)	2.32	(1.20, 3.44)
28	2.16	(1.26, 3.15)	2.66	(1.41, 3.90)	2.66	(1.41, 3.90)	2.13	(1.19, 3.08)	2.68	(1.43, 3.93)
29	2.40	(1.45, 3.45)	3.13	(1.65, 4.61)	3.13	(1.65, 4.61)	2.39	(1.37, 3.40)	3.16	(1.67, 4.65)
30	2.68	(1.66, 3.83)	3.90	(1.87, 5.93)	3.90	(1.87, 5.92)	2.66	(1.62, 3.71)	3.92	(1.90, 5.94)
31	2.98	(1.88, 4.18)					2.97	(1.85, 4.08)		

[a]The restriction, $\Sigma_j b_j = 0$, has been applied.

Table 23

Correlations (Lower Triangle) and Root Mean Squared Differences (Upper Triangle) of the English Usage Ability Estimates from Gibbs Sampling, Conditional Maximum Likelihood (CML), Marginal Maxium Likelihood (MML) with Maximum Likelihood (ML) and Expected A Posteriori (EAP), and Joint Maximum Likelihood (JML)

			MML		
Method	Gibbs Sampling	CML	ML	EAP	JML
Gibbs Sampling		.403	.404	.031	.417
CML	.996		.000	.415	.016
MML/ML	.996	1.000		.416	.015
MML/EAP	1.000	.996	.996		.429
JML	.996	1.000	1.000	.996	

Figure Captions

Figure 1. The Gibbs sampling algorithm.

Figure 2. A directed acyclic graph for LSAT6 data.

Figure 3. Convergence with starting values for LSAT6 item 1.

Figure 4. Gelman and Rubin shrink factors for LSAT6 items.

Figure 5. Trace lines of the sampled values and kernel density plots for LSAT6 items.

Figure 6. Convergence with starting values for Memory Test item 1.

Figure 7. Gelman and Rubin shrink factors for Memory Test items.

Figure 8. Trace lines of the sampled values and kernel density plots for Memory Test items.

Figure 9. Convergence with starting values for NAEP item 1.

Figure 10. Gelman and Rubin shrink factors for NAEP items.

Figure 11. Trace lines of the sampled values and kernel density plots for NAEP items.

Figure 12. Convergence with starting values for Usage item 1.

1st Iteration:	Draw $\omega_1^{(1)}$ from $p(\omega_1 \mid \omega_2^{(0)}, \ldots, \omega_k^{(0)}, Y)$
	Draw $\omega_2^{(1)}$ from $p(\omega_2 \mid \omega_1^{(1)}, \omega_3^{(0)}, \ldots, \omega_k^{(0)}, Y)$
	$\vdots$
	Draw $\omega_k^{(1)}$ from $p(\omega_k \mid \omega_1^{(1)}, \ldots, \omega_{k-1}^{(1)}, Y)$

$$\downarrow$$

2nd Iteration:	Draw $\omega_1^{(2)}$ from $p(\omega_1 \mid \omega_2^{(1)}, \ldots, \omega_k^{(1)}, Y)$
	Draw $\omega_2^{(2)}$ from $p(\omega_2 \mid \omega_1^{(2)}, \omega_3^{(1)}, \ldots, \omega_k^{(1)}, Y)$
	$\vdots$
	Draw $\omega_k^{(2)}$ from $p(\omega_k \mid \omega_1^{(2)}, \ldots, \omega_{k-1}^{(2)}, Y)$

$$\downarrow$$

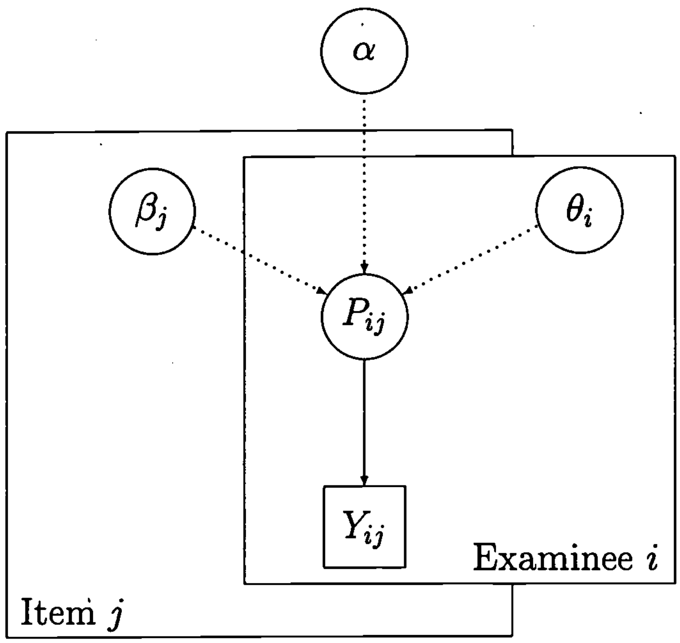
$$\vdots$$

$$\downarrow$$

tth Iteration:	Draw $\omega_1^{(t)}$ from $p(\omega_1 \mid \omega_2^{(t-1)}, \ldots, \omega_k^{(t-1)}, Y)$
	Draw $\omega_2^{(t)}$ from $p(\omega_2 \mid \omega_1^{(t)}, \omega_3^{(t-1)}, \ldots, \omega_k^{(t-1)}, Y)$
	$\vdots$
	Draw $\omega_k^{(t)}$ from $p(\omega_k \mid \omega_1^{(t)}, \ldots, \omega_{k-1}^{(t)}, Y)$

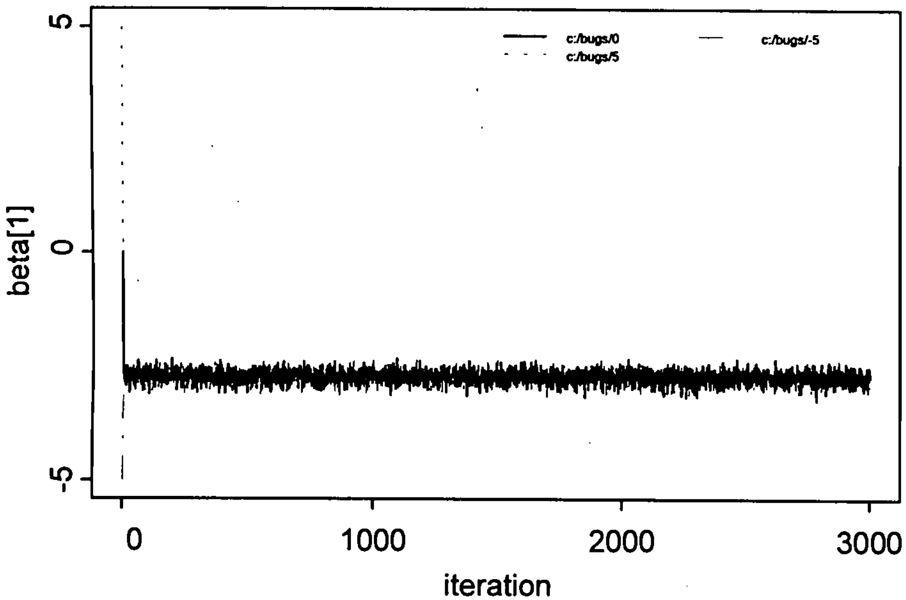
$$\downarrow$$

$$\vdots$$

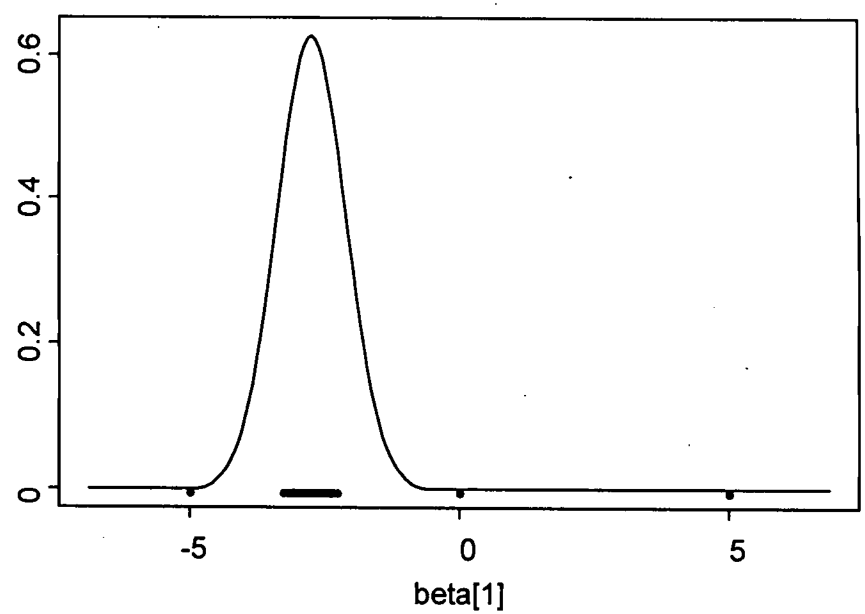

α
β_j
θ_i
P_ij
Y_ij
Examinee i
Item j

Convergence with Starting Values for LSAT6 Item-1

(3000 values per trace)

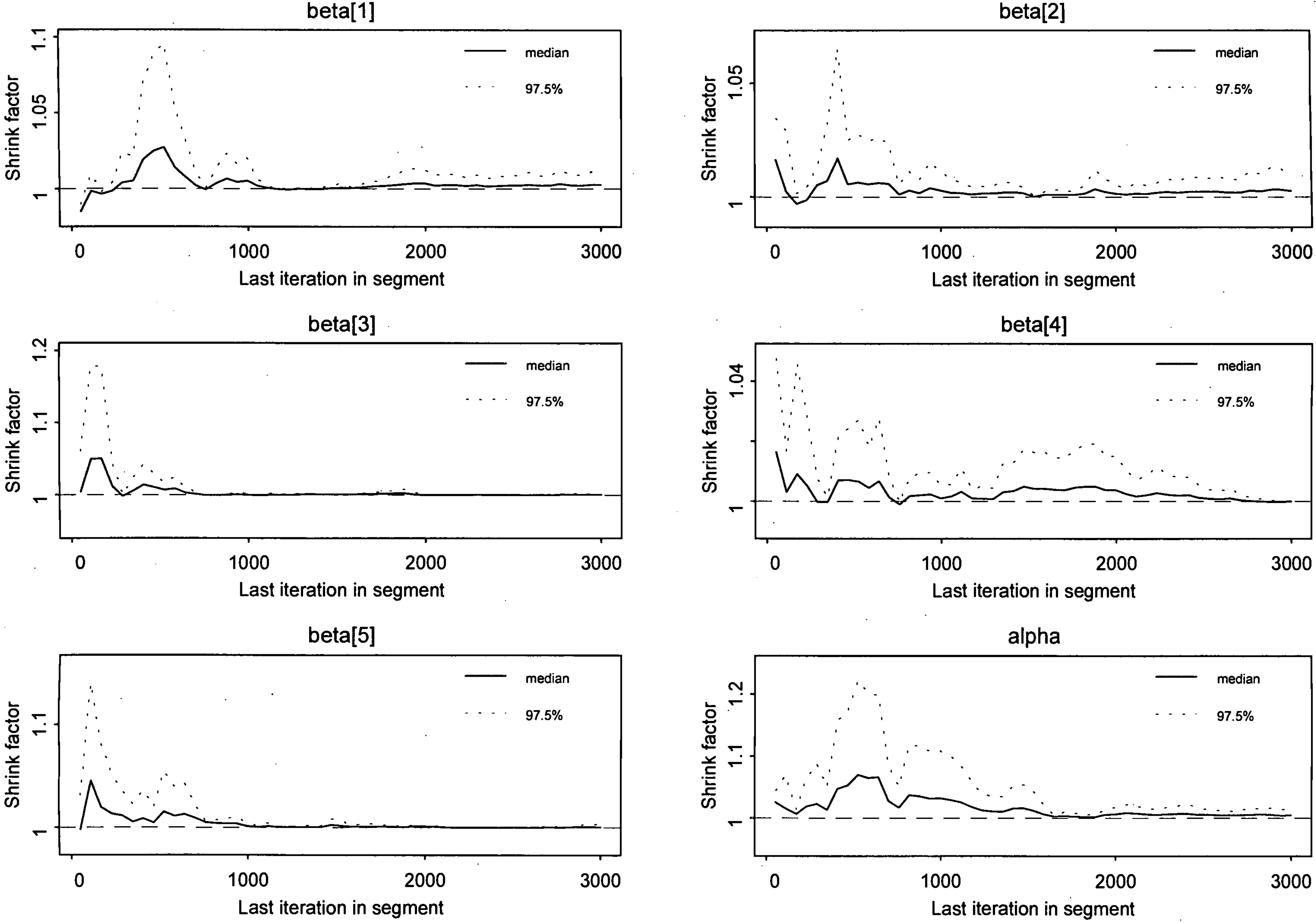


(9000 values)

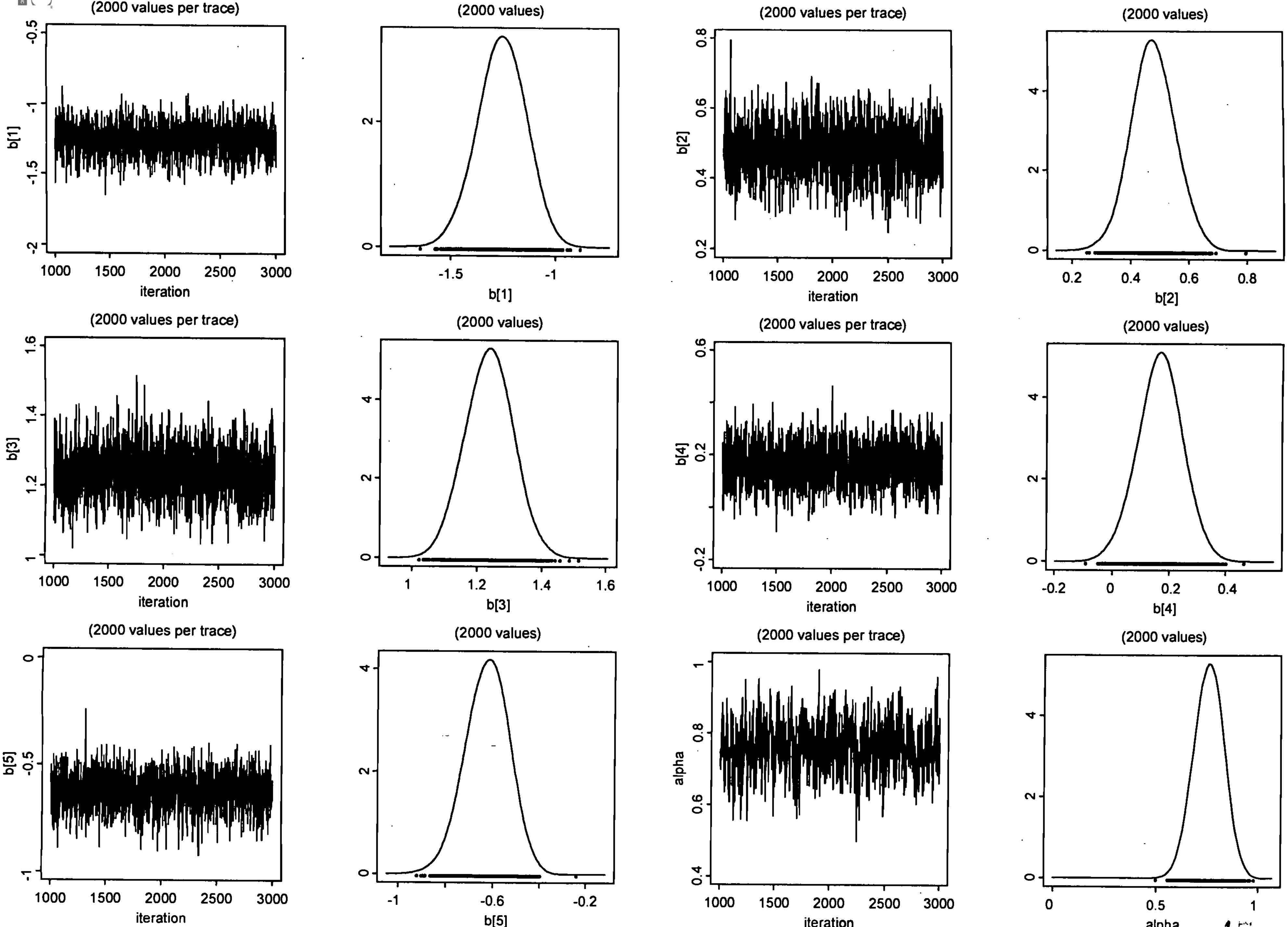


LSAT6 Items

Gelman & Rubin Shrink Factors

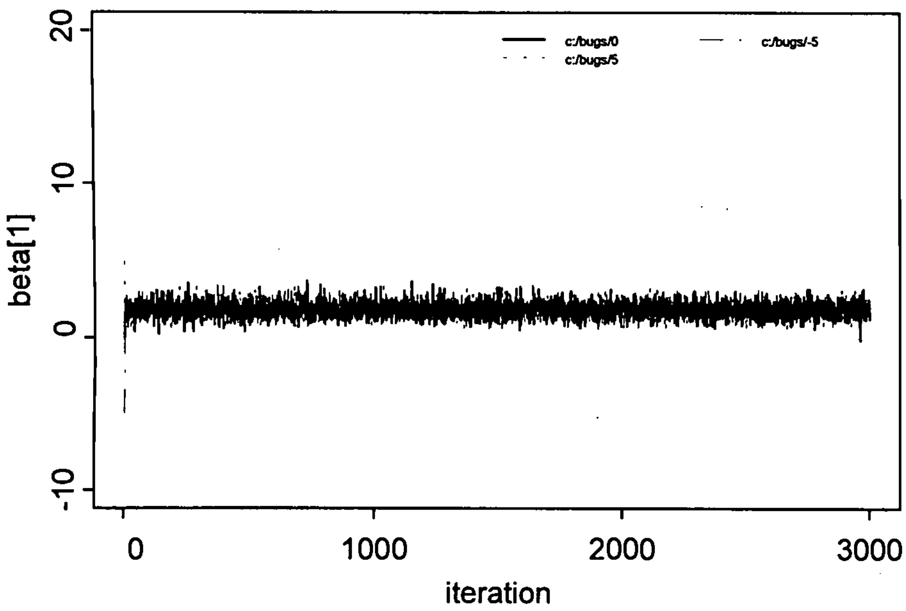


LSAT6 Items

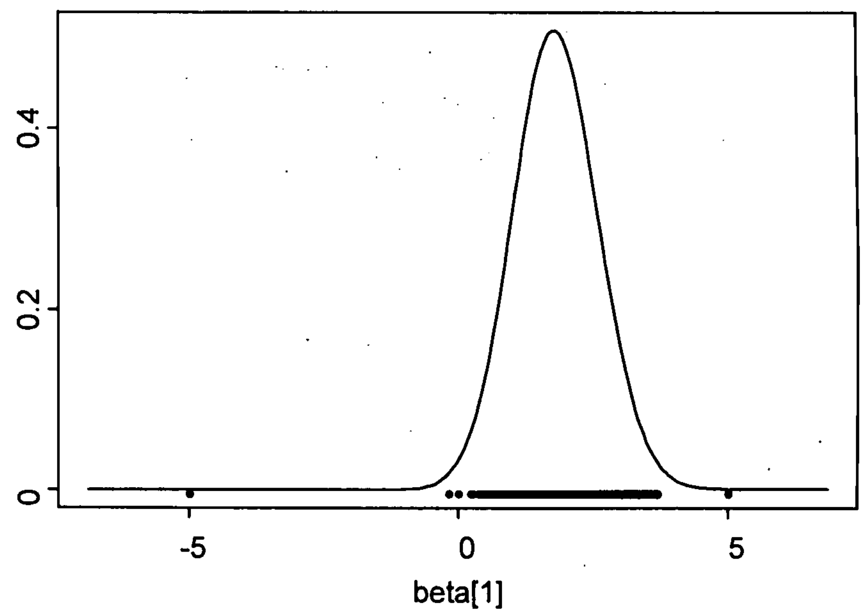

(2000 values per trace)
b[1]
iteration
(2000 values)
b[1]
(2000 values per trace)
b[2]
iteration
(2000 values)
b[2]
(2000 values per trace)
b[3]
iteration
(2000 values)
b[3]
(2000 values per trace)
b[4]
iteration
(2000 values)
b[4]
(2000 values per trace)
b[5]
iteration
(2000 values)
b[5]
(2000 values per trace)
alpha
iteration
(2000 values)
alpha

Convergence with Starting Values for Memory Test Item-1

(3000 values per trace)

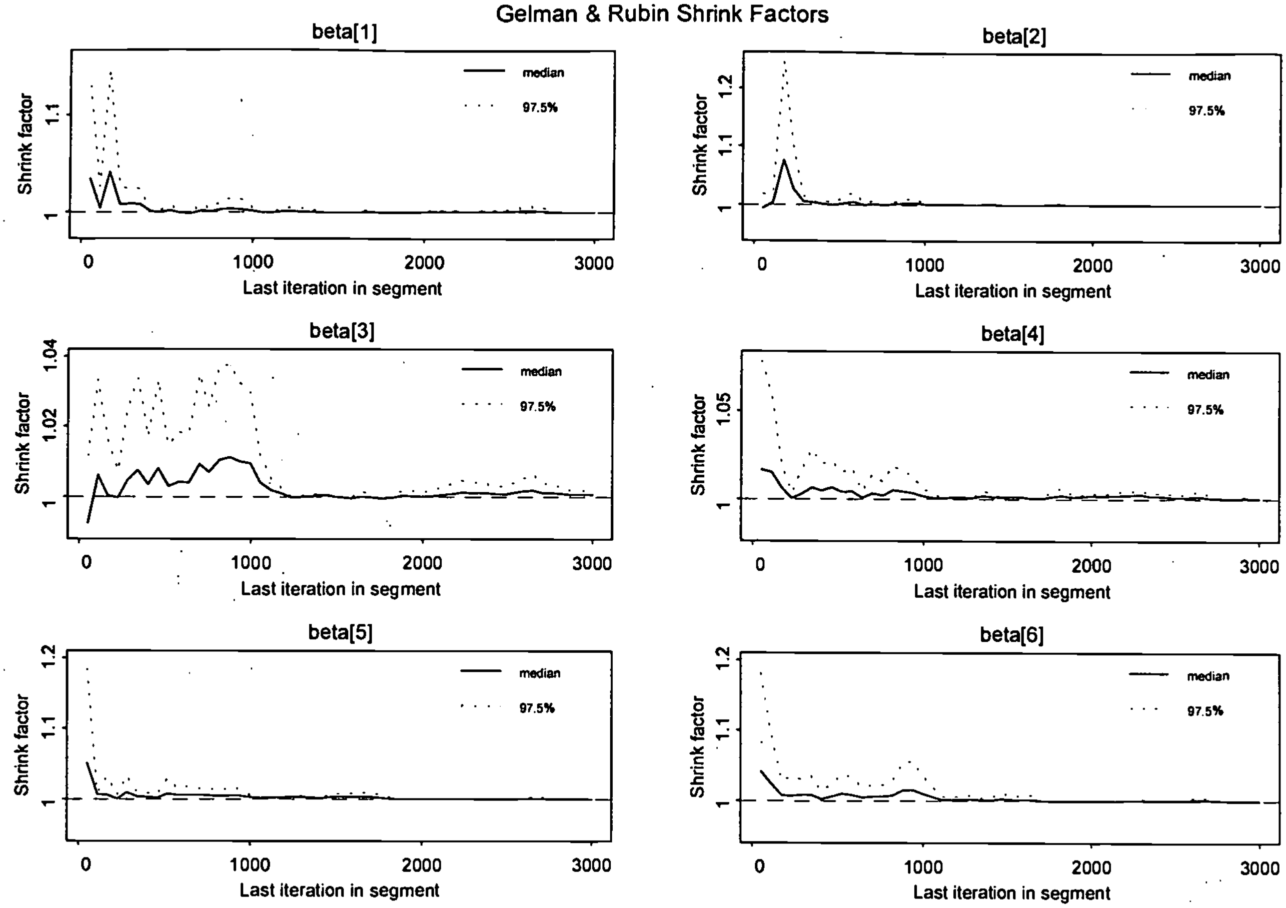


(9000 values)



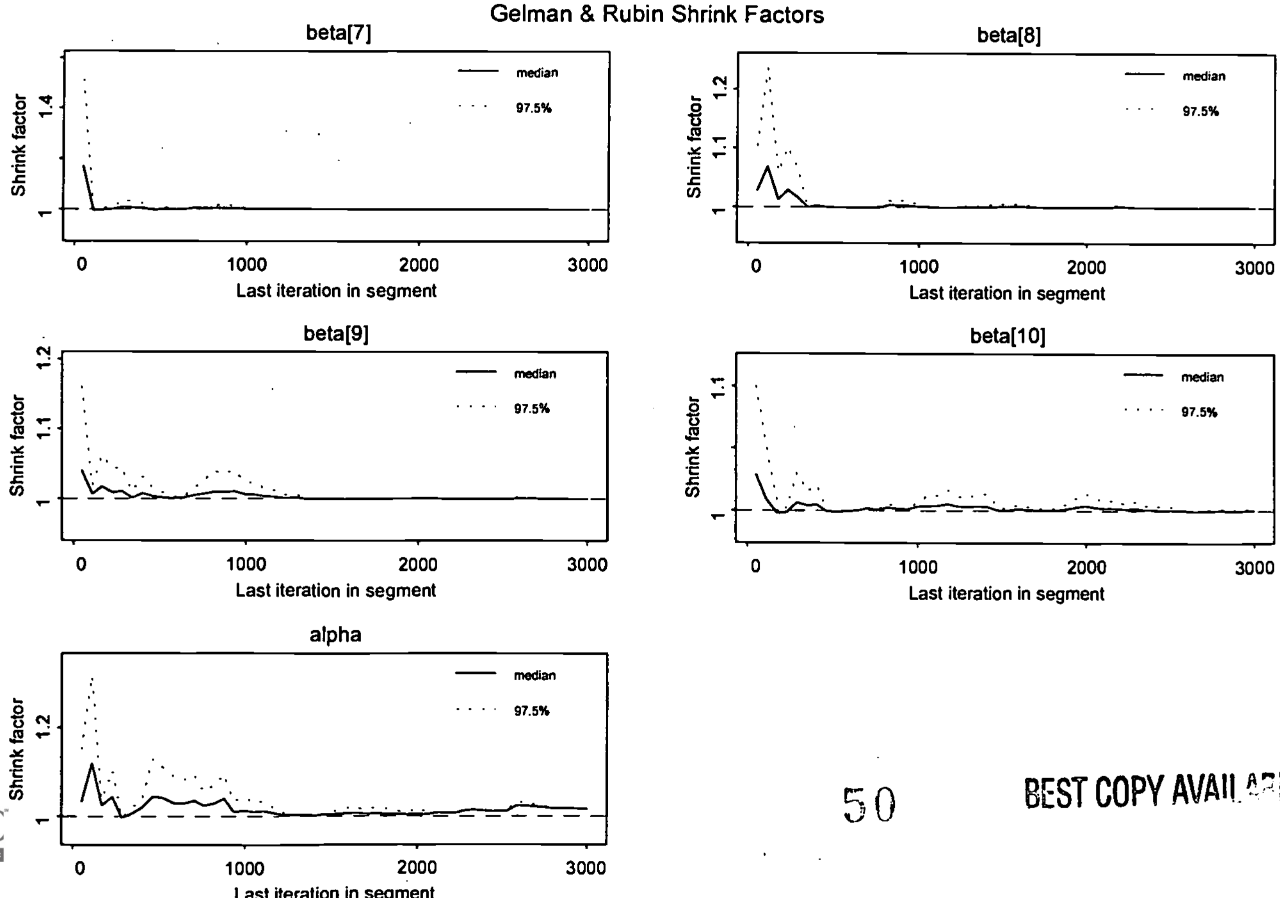
Memory Test Items

Gelman & Rubin Shrink Factors

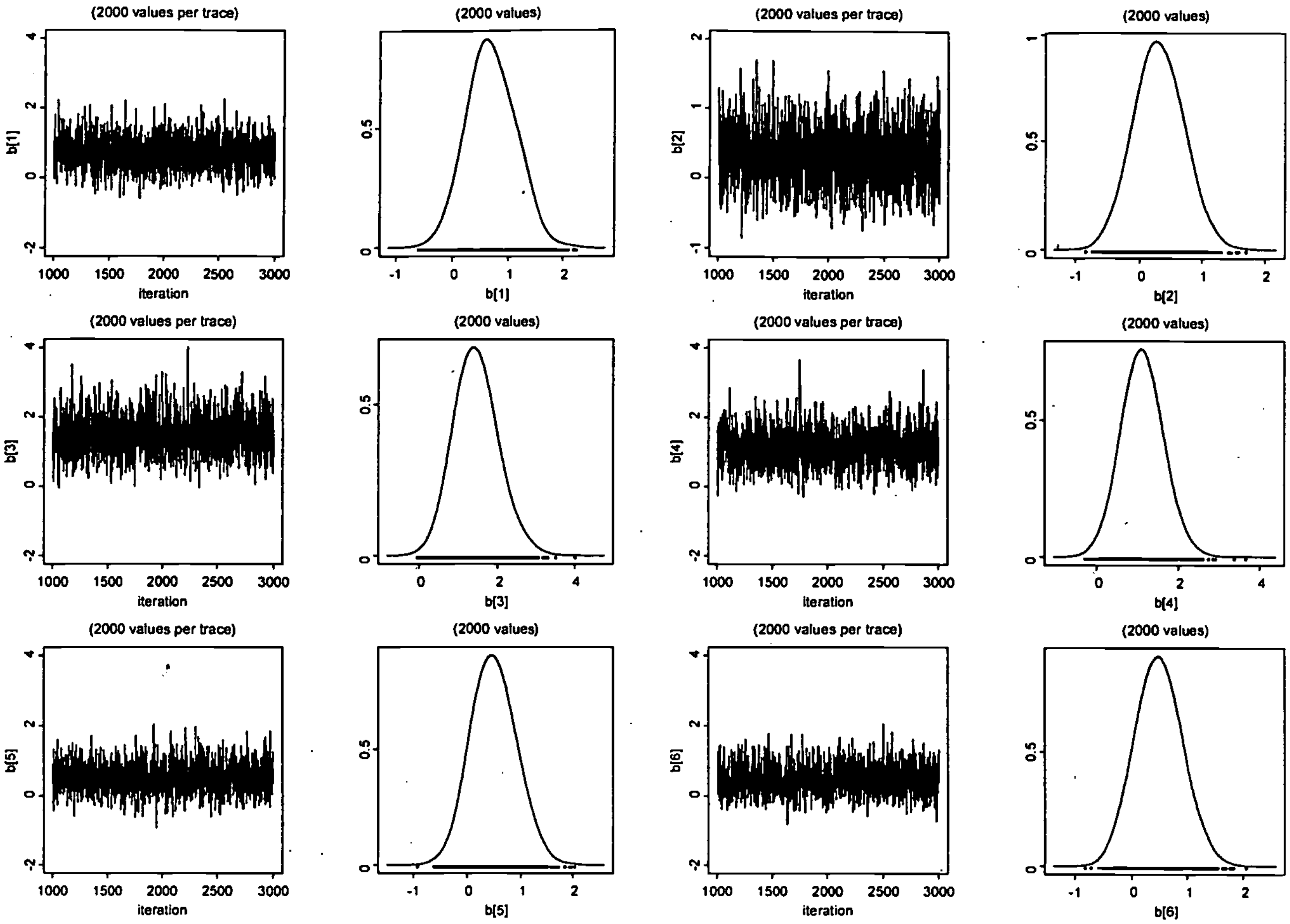


Memory Test Items

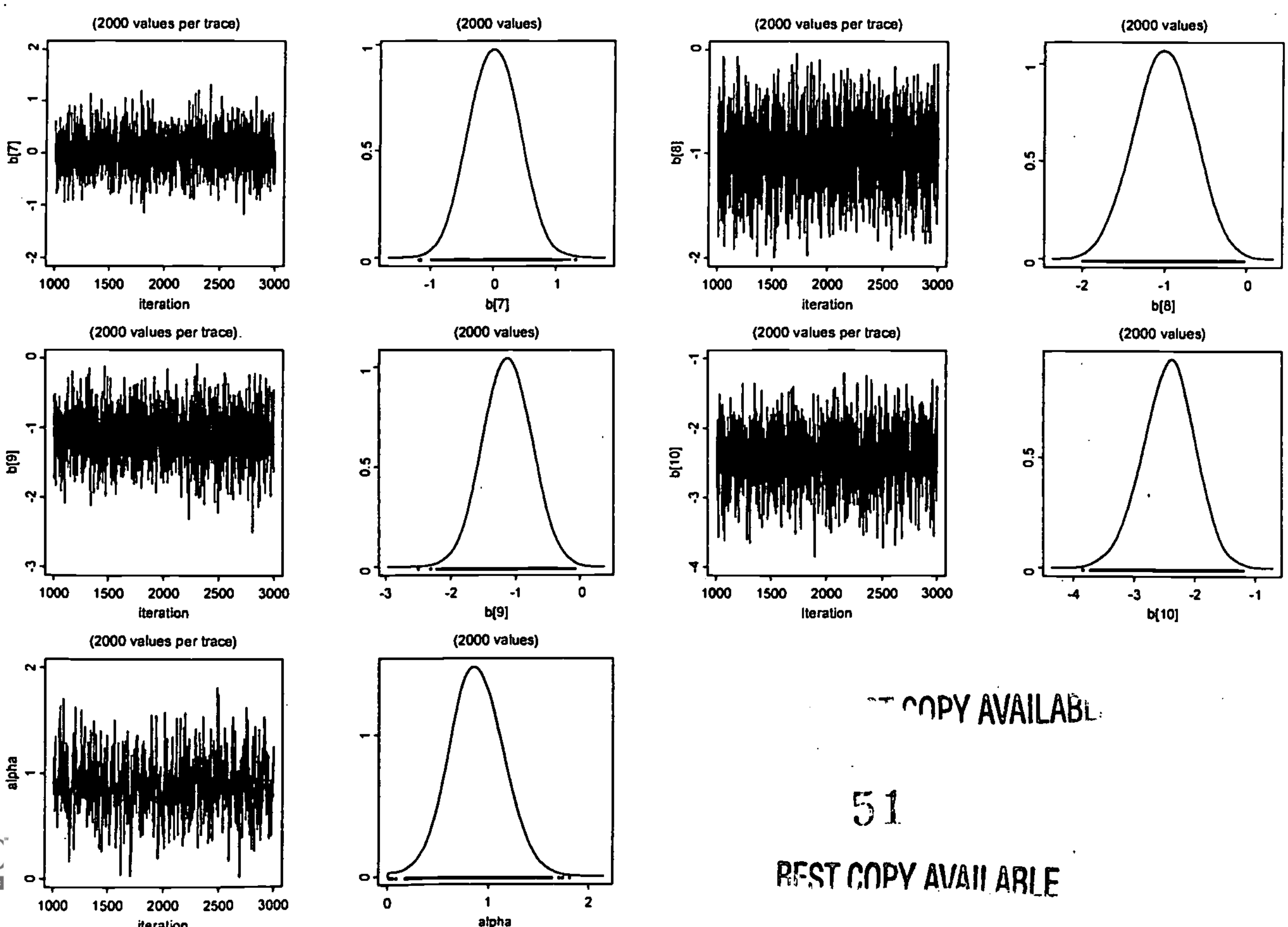
Gelman & Rubin Shrink Factors



Memory Test Items

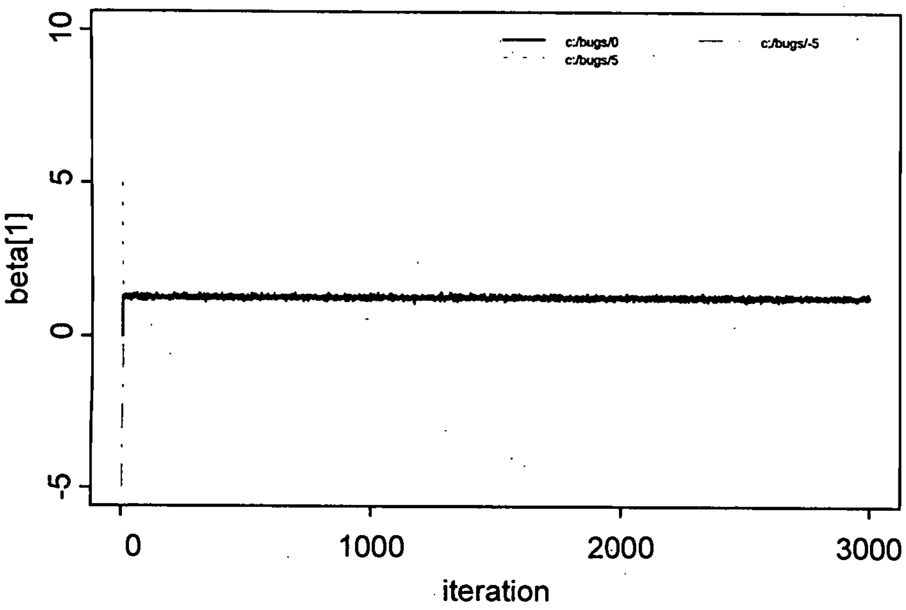


Memory Test Items

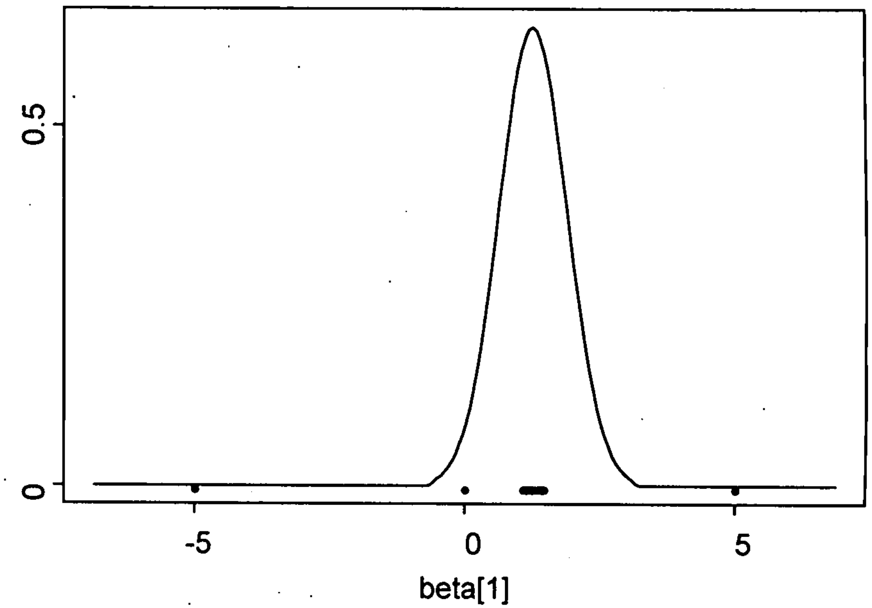


Convergence with Starting Values for NAEP Item-1

(3000 values per trace)

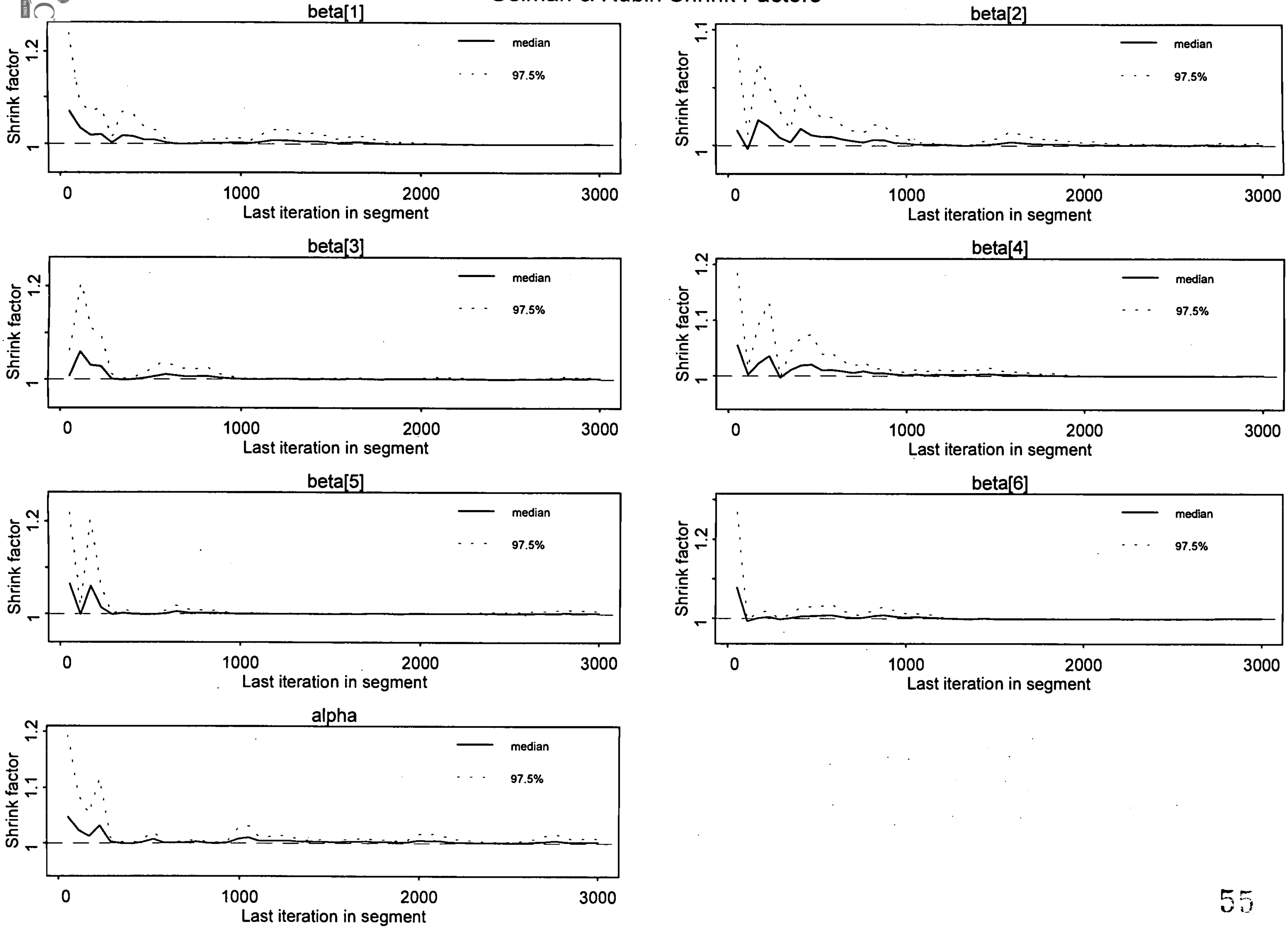


(9000 values)

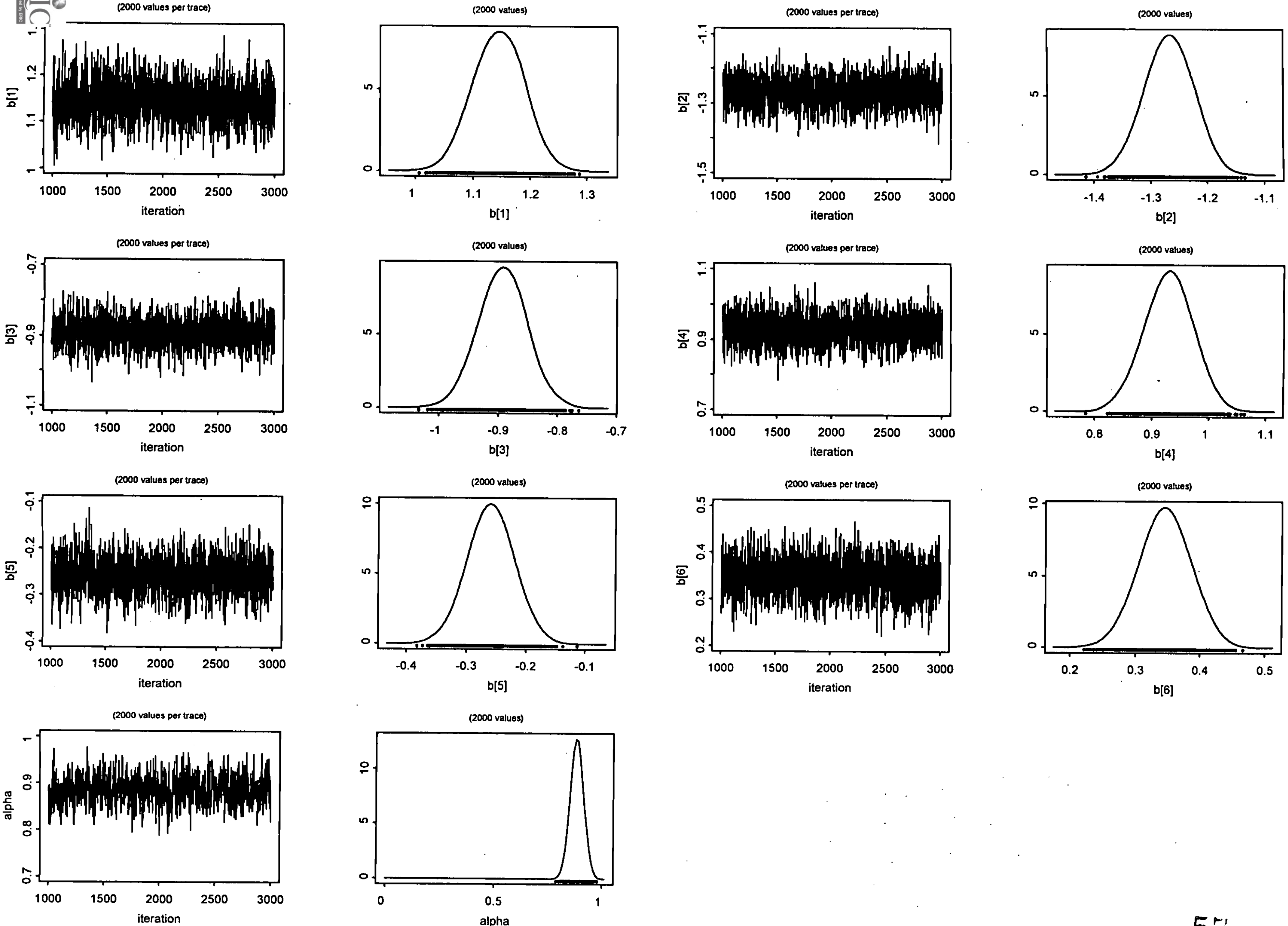


NAEP Items

Gelman & Rubin Shrink Factors



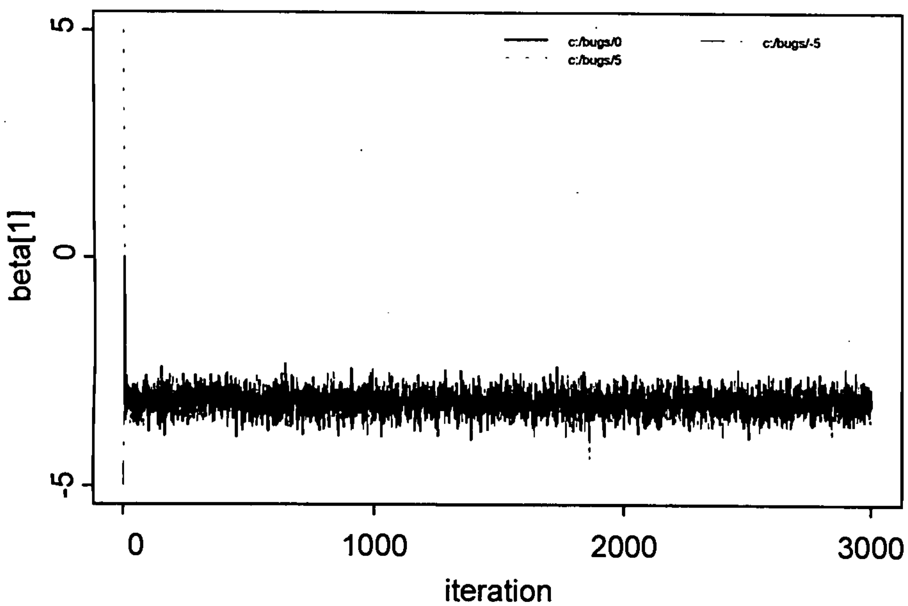
NAEP Items



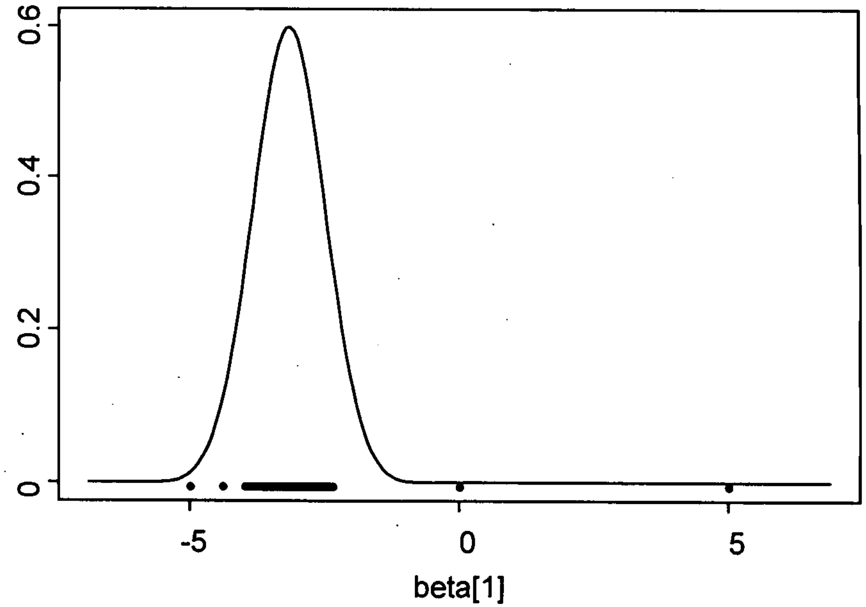
(2000 values per trace)
b[1]
iteration
(2000 values)
b[1]
(2000 values per trace)
b[2]
iteration
(2000 values)
b[2]
(2000 values per trace)
b[3]
iteration
(2000 values)
b[3]
(2000 values per trace)
b[4]
iteration
(2000 values)
b[4]
(2000 values per trace)
b[5]
iteration
(2000 values)
b[5]
(2000 values per trace)
b[6]
iteration
(2000 values)
b[6]
(2000 values per trace)
alpha
iteration
(2000 values)
alpha

Convergence with Starting Values for Usage Item-1

(3000 values per trace)



(9000 values)



Appendix

```
model lsat6;
const
   I = 1000,
   J = 5;
var
   y[I,J], p[I,J], theta[I], alpha, zeta[J], b[J];
data in "lsat6-s.dat";
inits in "rasch.in";
{
  for (i in 1:I) {
     for (j in 1:J) {
        logit(p[i,j]) <- alpha*theta[i] - beta[j];
        y[i,j] ~ dbern(p[i,j]);
     }
     theta[i] ~ dnorm(0,1);
  }
  for (j in 1:J) {
     beta[j] ~ dnorm(0,0.0001);
     b[j] <- - beta[j] - mean(beta[]);
  }
  alpha ~ dnorm(0,0.0001) I(0,);
}
```

Author's Address

Send correspondence to Seock-Ho Kim, Department of Educational Psychology, The University of Georgia, 325 Aderhold Hall, Athens, GA 30602-7143. Internet: skim@coe.uga.edu

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

TM029163

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: An Evaluation of a Markov Chain Monte Carlo Method for the Rasch Model

Author(s): Seock-Ho Kim

Corporate Source: The University of Georgia; Paper presented at the joint annual meeting of the Psychometric Society and the Classification Society of North America, Urbana, Illinois.

Publication Date: 6/19/98

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: Seock-Ho Kim

Printed Name/Position/Title: Seock-Ho Kim, Assistant Professor

Organization/Address: The University of Georgia, 325 Aderhold Hall, Athens, GA 30602

Telephone: (706) 542-4224

FAX: (706) 542-4240

E-Mail Address: skim@coe.uga.edu

Date: 8/7/98

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 9/97)
PREVIOUS VERSIONS OF THIS FORM ARE OBSOLETE.