DOCUMENT RESUME

ED 430 012 TM 029 725

AUTHOR	Kim, Seock-Ho; Cohen, Allan S.
TITLE	Accuracy of Parameter Estimation in Gibbs Sampling under the Two-Parameter Logistic Model.
PUB DATE	1999-04-02
NOTE	55p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Bayesian Statistics; *Estimation (Mathematics); Item Response Theory; Markov Processes; Monte Carlo Methods; Simulation
IDENTIFIERS	*Gibbs Sampling; Parameter Identification

ABSTRACT

The accuracy of Gibbs sampling, a Markov chain Monte Carlo procedure, was considered for estimation of item and ability parameters under the two-parameter logistic model. Memory test data were analyzed to illustrate the Gibbs sampling procedure. Simulated data sets were analyzed using Gibbs sampling and the marginal Bayesian method. The marginal Bayesian method combined with the expected a posteriori estimation of ability yielded consistently smaller root mean square errors and better bias results than Gibbs sampling. (Contains 12 figures, 29 tables, and 56 references.) (Author)

ED 430 012

Accuracy of Parameter Estimation in Gibbs Sampling Under the Two-Parameter Logistic Model

Seock-Ho Kim
The University of Georgia
Allan S. Cohen
University of Wisconsin–Madison

April 22, 1998
Running Head: GIBBS SAMPLING FOR 2PL

Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Accuracy of Parameter Estimation in Gibbs Sampling Under the Two-Parameter Logistic Model

Abstract

The accuracy of Gibbs sampling, a Markov chain Monte Carlo procedure, was considered for estimation of item and ability parameters under the two-parameter logistic model. Memory test data were analyzed to illustrate the Gibbs sampling procedure. Simulated data sets were analyzed using Gibbs sampling and the marginal Bayesian method. The marginal Bayesian method combined with the expected a posteriori estimation of ability yielded consistently smaller root mean square errors and better bias results than Gibbs sampling.

Keywords: *Bayesian inference, Gibbs sampling, item response theory, Markov chain Monte Carlo, marginal Bayesian.*

Introduction

For models with several parameters, statistical inference sometimes requires integration over high-dimensional probability distributions in order to estimate any parameter of interest or to obtain any particular function of the parameters. One such case 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 the Bayesian approach), the required integrations may be analytically nontractable. As is true for many cases in statistics, the marginal density can be approximated using various techniques (e.g., standard numerical integration, Laplacian approximation, Edgeworth expansion, importance sampling, Metropolis algorithm; see Bernardo & Smith, 1994; Leonard & Hsu, 1994). In this paper, we examine the accuracy of Gibbs sampling, one of the Markov Chain Monte Carlo (MCMC) methods for marginal density estimation, for estimation of IRT parameters. In particular, we focus on the accuracy of Gibbs sampling (Geman & Geman, 1984) for estimation of item and ability parameters under the two-parameter logistic (2PL) model when sample sizes are small.

A number of ways exist for implementing the MCMC method. [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. [See Gilks, Richardson, and Spiegelhalter (1996) for a recent survey of applications.] Basically Gibbs sampling is applicable for obtaining parameter estimates for the complicated joint posterior distribution in Bayesian estimation under IRT (e.g., Mislevy, 1986; Swaminathan & Gifford, 1985; Tsutakawa & Lin, 1986).

A few studies have examined the use of Gibbs sampling under IRT. 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 2PL model.

MCMC computer programs in the context of 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.

The marginal maximum likelihood (MML) and marginal Bayesian (MB) methods using the expectation and maximization (EM) algorithm, as implemented in the computer program BILOG (Mislevy & Bock, 1990), have become the standard estimation technique for obtaining item parameter estimates of IRT. Ability parameters are estimated in those marginalized solutions using either maximum likelihood (ML), expected a posteriori (EAP), or maximum a posteriori (MAP) estimation after obtaining the item parameter estimates and assuming the estimates are true values. The Gibbs sampling procedure approaches the estimation of item parameters using the joint posterior distribution rather than the marginal distribution. In Gibbs sampling ability parameters can be estimated either jointly with item parameters or after obtaining the item parameters. All of the estimation methods should yield comparable item and ability parameter estimates, when comparable priors are used or when ignorance or locally uniform priors are used when sample sizes are large. This study was designed to evaluate the comparability of item and ability parameter estimates using the 2PL model. Specifically, estimation methods implemented in the two computer programs, BUGS and BILOG, were examined and compared.

Theoretical Framework

Marginalized Solutions

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 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 marginal probability of obtaining the response vector Y_i for examinee i sampled from a given population is

$$p(Y_i|\xi) = \int p(Y_i|\theta_i, \xi)p(\theta_i)d\theta_i, \tag{2}$$

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

$$p(Y|\xi) = \prod_{i=1}^{N} p(Y_i|\xi) = l(\xi|Y), \tag{3}$$

where $l(\xi|Y)$ can be regarded as a function of ξ given the data Y. In MML, the marginal likelihood is maximized to obtain maximum likelihood estimates of item parameters (Bock & Aitkin, 1981; Bock & Lieberman, 1970).

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

$$p(\xi|Y) = \frac{p(Y|\xi)p(\xi)}{p(Y)} \propto l(\xi|Y)p(\xi), \tag{4}$$

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

Point estimates of ability parameters do not arise during the course of the marginalized estimation of item parameters. They are calculated after the item parameters are estimated assuming the obtained item parameters are true values. Three methods are generally available; ML, EAP (i.e., posterior mean), and MAP (i.e., posterior mode) (Bock & Aitkin, 1981; Bock & Mislevy, 1982).

Joint Estimation Procedures

Birnbaum (1968) and Lord (1980) describe the estimation of the θ and ξ by joint maximization of the likelihood function

$$p(Y|\theta,\xi) = \prod_{i=1}^{N}\prod_{j=1}^{n} P_j(\theta_i)^{Y_{ij}} Q_j(\theta_i)^{1-Y_{ij}} = l(\theta,\xi|Y), \tag{5}$$

where $\theta = (\theta_1, \ldots, \theta_N)'$. In implementation of joint maximum likelihood (JML) estimation (see Lord, 1986 for a comparison of marginalized and joint estimation methods), 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 are 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

$$p(\theta,\xi|Y) = \frac{p(Y|\theta,\xi)p(\theta,\xi)}{p(Y)} \propto l(\theta,\xi|Y)p(\theta,\xi), \tag{6}$$

where $p(\theta,\xi)$ is the prior density of the parameters θ and ξ. This procedure is joint Bayesian (JB) estimation. 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 stable Bayes modal estimates of item and ability parameters.

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 the IRT estimation procedures via MML, MB, JML, or JB noted above, however, 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 follows:

$$
\begin{aligned}
&\text{Draw } \omega_1^{(1)} \text{ from } p(\omega_1|\omega_2^{(0)}, \ldots, \omega_k^{(0)}, Y),\\
&\text{Draw } \omega_2^{(1)} \text{ from } p(\omega_2|\omega_1^{(1)}, \omega_3^{(0)}, \ldots, \omega_k^{(0)}, Y),\\
&\vdots\\
&\text{Draw } \omega_k^{(1)} \text{ from } p(\omega_k|\omega_1^{(1)}, \ldots, \omega_{k-1}^{(1)}, Y),\\
&\text{Draw } \omega_1^{(2)} \text{ from } p(\omega_1|\omega_2^{(1)}, \ldots, \omega_k^{(1)}, Y),\\
&\text{Draw } \omega_2^{(2)} \text{ from } p(\omega_2|\omega_1^{(2)}, \omega_3^{(1)}, \ldots, \omega_k^{(1)}, Y),\\
&\vdots\\
&\text{Draw } \omega_k^{(2)} \text{ from } p(\omega_k|\omega_1^{(2)}, \ldots, \omega_{k-1}^{(2)}, Y),\\
&\vdots\\
&\text{Draw } \omega_1^{(t+1)} \text{ from } p(\omega_1|\omega_2^{(t)}, \ldots, \omega_k^{(t)}, Y),\\
&\text{Draw } \omega_2^{(t+1)} \text{ from } p(\omega_2|\omega_1^{(t+1)}, \omega_3^{(t)}, \ldots, \omega_k^{(t)}, Y),\\
&\vdots\\
&\text{Draw } \omega_k^{(t+1)} \text{ from } p(\omega_k|\omega_1^{(t+1)}, \ldots, \omega_{k-1}^{(t+1)}, Y),\\
&\vdots
\end{aligned}
$$

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

$$p(\omega^{(t)}, \omega^{(t+1)}) = \prod_{l=1}^{k} p(\omega_l^{(t+1)}|\omega_j^{(t)}, j > l, \omega_j^{(t+1)}, j < l, Y). \tag{7}$$

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

$$\hat{p}(\omega_i|Y) = \frac{1}{m}\sum_{l=1}^{m} p(\omega_i|\omega_{jl}^{(t)}, j \neq i, Y). \tag{8}$$

In the context of IRT, Gibbs sampling attempts to 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.

An Example

Steps for Gibbs Sampling

The following example is presented using the 10-item memory test data for 40 examinees from Thissen (1982) (see Table 1). Model parameters were estimated by Gibbs sampling using the computer program BUGS (Spiegelhalter et al., 1997). These same data were analyzed under the Rasch model in Thissen (1982).

Insert Table 1 about here

Gibbs sampling uses the following four basic steps (cf. Spiegelhalter, Best, et al., 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. In addition, comparisons with the results from the marginalized methods (e.g., MB and MML) as implemented in the computer program BILOG (Mislevy & Bock, 1990) are presented.

Model Specifications

The model specifications are used as input to the BUGS computer program. In the memory test 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 item discrimination parameter α_j, and the item difficulty parameter β_j under the 2PL. The θ_i are assumed to be independently drawn from a standard normal distribution for scaling purposes. Figure 1 shows a directed acyclic graph (see Lauritzen, Dawid, Larsen, & Leimer, 1990; Whittaker, 1990; Spiegelhalter, Dawid, Lauritzen, & Cowell, 1993) based on these assumptions. λ_j and ζ_j are used in Figure 1 instead of α_j and β_j (see Equation 11). 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. 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.

Insert Figure 1 about here

We 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. 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 1 that, 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,

$$P(V) = \prod_{v \in V} P(v|\text{parents}[v]), \tag{9}$$

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 1 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

$$\begin{aligned} P(v|V-v) &\propto P(v, V-v) \\ &\propto P(v|\text{parent}[v]) \prod_{w \in \text{children}[v]} P(w|\text{parents}[w]). \end{aligned} \tag{10}$$

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 memory test data, we begin by specifying the forms of the parent and child relationships in Figure 1. Under the 2PL model, the probability that examinee i responds correctly to item j is assumed to follow a logistic function parameterized by the examinee's latent ability θ_i, the item discrimination parameter, α_j, and the item difficulty parameter, β_j. For estimation purposes, we use the form $\alpha_j(\theta_i - \beta_j) = \lambda_j\theta_i + \zeta_j$, where the slope parameter $\lambda_j = \alpha_j$ and the intercept parameter $\zeta_j = -\alpha_j\beta_j$. Hence,

$$P_{ij} = \frac{1}{1 + \exp[-\alpha_j(\theta_i - \beta_j)]} = \frac{1}{1 + \exp[-(\lambda_j\theta_i + \zeta_j)]}. \tag{11}$$

Since Y_{ij} are Bernoulli with parameter P_{ij}, we can define

$$Y_{ij} \sim \text{Bernoulli}(P_{ij}) \tag{12}$$

and

$$\text{logit}(P_{ij}) = \lambda_j\theta_i + \zeta_j. \tag{13}$$

To complete the specification of a full probability model for the BUGS computer program, prior distributions of the nodes without parents (i.e., θ_i, λ_j, and ζ_j) also need to be specified. We can define these priors in several different ways. We can impose priors on λ_j and ζ_j using a hierarchical Bayes approach (e.g., Swaminathan & Gifford, 1985; Kim, Cohen, Baker, Subkoviak, & Leonard, 1994) or, 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), two prior distributions were chosen for the memory test analyses: (1) $\lambda_j \sim N(0,1)$ with $\lambda_j > 0$ and $\zeta_j \sim N(0,100^2)$ and (2) $\lambda_j \sim N(0,10^2)$ with $\lambda_j > 0$ and $\zeta_j \sim N(0,100^2)$. An example input file for BUGS is given in the Appendix.

Starting Values

The choice of starting values (e.g., $\omega^{(0)}$) is not generally that critical as the Gibbs sampler (and most other MCMC algorithms as well) 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.

In this example, three runs were performed using the memory test data with three sets of starting values for λ_j and ζ_j, $j = 1(1)10$. The starting values for the item parameters are given 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 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 highly discriminating, and the third run represented an opposite assumption. The priors used in the three runs were the same; $\lambda_j \sim N(0,1)$ with $\lambda_j > 0$ and $\zeta_j \sim N(0,100^2)$.

Insert Table 2 about here

Each of the three runs consisted of 10,000 iterations. Results for λ_1 and ζ_1 are presented in Figure 2. The computer program CODA (Best, Cowles, & Vines, 1997) was used to obtain these graphs. The top two plots in Figure 2 contain the graphical summaries of the Gibbs sampler for λ_1. The top left plot shows the trace of the sampled values of λ_1 for the three runs. Results for all three runs show that the λ_1 generated by the Gibbs sampler quickly settled down regardless of the starting values. The top right graph shows the kernel density plot of the three pooled runs of 30,000 values for λ_1. The variability among the λ_1 values generated by the Gibbs sampler seems to be large, possibly due to the small sample size. The distribution looks like a truncated normal form due to the positive constraints on λ_j.

Insert Figure 2 about here

The bottom two plots contain graphical summaries of the Gibbs sampler for ζ_1. The bottom left plot shows the trace of the sampled values of ζ_1 for all three runs. The ζ_1 generated by the Gibbs sampler quickly settled down regardless of the starting values. The bottom right graph shows the kernel density plot of the three pooled runs of 30,000 values for ζ_1. The variability of the λ_1 values seems to be large. The sampled values seem to be concentrated around -2, and the sample values seem to follow a normal distribution.

The results for other item parameter estimates were very similar to those for λ_1 and ζ_1. Overall, the starting values appear to not have affected the final results. Useful starting values for IRT problems can be found from the noniterative minimum logit chi-square estimation solution (Baker, 1987) or from values based on Jensema (1976) and Urry (1974) as employed in BILOG. Use of "good" starting values, such as from the above methods, can avoid the time delay required by a lengthy starting period. Our experience with these starting values indicates $\lambda_j = 1$ and $\zeta_j = 0$ will work sufficiently well for applications under the 2PL. In subsequent analyses, therefore, the values, $\lambda_j = 1$ and $\zeta_j = 0$, were used as starting values.

Output Monitoring

A critical issue for MCMC methods including Gibbs sampling 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 MCMC algorithms.]

We illustrate here the use of Gelman and Rubin (1992) statistics on three 10,000 iteration runs. Details of the Gelman and Rubin method are given by Gelman (1996). Each 10,000 iteration run required about 10 minutes on a Pentium 90 megahertz computer. Monitoring was done using the suite of S-functions called CODA (Best et al., 1997). Figure 3a shows the trace lines of the sampled values of λ_1 and ζ_1 for the two runs. The plots in Figure 3a indicate that the three runs yielded similar values. Gelman-Rubin statistics (i.e., shrink factors) are plotted in Figure 3b for λ_1 and ζ_1. For both parameters, the medians were stabilized after roughly 500 iterations and definitely after about 5,000 iterations.

Insert Figures 3a and 3b 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 should be near 1 in order to be reasonably assured that convergence has occurred. The median for λ_1 in the example was 1.00 and the 97.5 percentage point was 1.00. The median for ζ_1 was 1.00 and the 97.5 percentage point was 1.00. These values indicated that reasonable convergence was realized for these parameters.

The Gelman-Rubin statistics can be calculated sequentially as the runs proceed, and plotted as in Figure 3b. These plots as well as other plots for λ_j and ζ_j suggest the first 1,000 iterations of each run be discarded and the remaining samples be pooled. We used 5,000 iterations as burn-in and the subsequent 5,000 iterations for estimating.

BUGS and BILOG Parameter Estimates

The posterior mean of the Gibbs sampler was obtained for each parameter. Two different sets of prior distributions for item parameters were employed in the BUGS runs. The first set employed an informative prior on $\lambda_j \sim N(0,1)$ and an uninformative prior on $\zeta_j \sim N(0,100^2)$. In addition, a constraint was imposed on the ranges of λ_j to allow only positive values (i.e., $\lambda_j > 0$). The prior distribution for λ_j limits possible values. Gibbs sampling-informative (GS-I) indicates this informative prior for λ_j. The second set employed two uninformative prior distributions, $\lambda_j \sim N(0,10^2)$ with the constraint $\lambda_j > 0$ and $\zeta_j \sim N(0,100^2)$. This second set of priors is Gibbs sampling-uninformative (GS-U).

For BILOG runs, two procedures were used: MB/EAP (i.e, marginal Bayesian item parameter estimation with expected a posteriori ability estimation) and MML/ML (i.e, marginal maximum likelihood item parameter estimation with maximum likelihood ability estimation). The default prior in BILOG for the estimation of item parameters in the 2PL is only on the item discrimination parameter as $p(\log \alpha_j) = N(\mu_{\log \alpha_j}, \sigma^2_{\log \alpha_j}) = N(0,.5^2)$. Default options of BILOG yield MB/EAP. For MML/ML, no prior distributions were used (although, technically speaking, the marginalization required the standard normal prior for ability).

Insert Tables 3 and 4 about here

The information in Table 3 indicates that the four estimation methods yielded somewhat different item parameter estimates. Differences between estimates from Gibbs sampling with informative priors and marginal Bayesian were relatively small, indicating the estimates from the methods were comparable. Both Gibbs sampling with uninformative priors and marginal maximum likelihood yielded very unstable item parameter estimates.

The ability estimates and the standard errors from the memory test are presented in Table 4. The maximum likelihood method after MML estimation of item parameters yielded several unstable estimates. GS-I, GS-U, and MB/EAP yielded relatively similar results. Recall that normal priors were used in those three Bayes methods of ability estimation.

It is important to note that the posterior interval from Gibbs sampling can be constructed not from the normal based method using the standard errors but from the sampled values. Figure 4 shows the trace lines of the 5,000 sampled values of λ_1 and ζ_1 for the Gibbs sampling-informative. The kernel density plots can also be found in Figure 4. Since the distribution of the sampled values of λ_1 looks like a truncated normal form, it is also of interest to obtain the posterior interval directly from the sampled values. The 95% posterior intervals of the GS-I and MB are presented in Table 5. Table 6 presents the ability estimates and the 95% posterior intervals. It is important to notice that GS-I may yield different ability estimates for examinees who had the same response pattern (e.g., examinees 1 to 5).

Insert Figure 4 and Tables 5 and 6 about here

Method

Simulation Conditions

Although the example presented above is informative, it does not provide enough information with regard to comparative characteristics of item and ability parameter estimates of Gibbs sampling. A standard method for examining such characteristics is based on studies of parameter recovery employing simulated data (e.g., Hulin, Lissak, & Drasgow, 1982; Yen, 1983). Hence, data were simulated under the following conditions; the number of examinees ($N = 50, 100, 200$) and the number of items ($n = 10, 20, 40$). Due to the small sample sizes, informative priors were employed in the two estimation methods. The sample sizes and the test lengths were selected to emulate a situation in which estimation procedures and priors might have some impact upon item parameter estimates (e.g., Harwell & Janosky, 1991). Sample size and test length were completely crossed to yield nine conditions.

For the Gibbs sampling procedure, an informative prior was used: $\lambda_j \sim N(0, 1)$ with the constraint $\lambda_j > 0$ and $\zeta_j \sim N(0, 100^2)$. For MB estimation via BILOG the default priors were used with EAP estimation of ability. We denote these two methods as Gibbs sampling and marginal Bayesian (MB) estimation.

Data Generation

Item response vectors were generated via the computer program GENIRV (Baker, 1982) for the 2PL model. The generating parameters for item discrimination were distributed with mean 1.00 and variance .09 (i.e., standard deviation .3), and the underlying item difficulty parameters were distributed normal with mean 0 and variance 1. Item discrimination and item difficulty parameters for the 10-, 20-, and 40-item tests are presented in Tables 7, 8, and 9, respectively. Item discrimination and difficulty parameters were not correlated. The distribution of the underlying ability parameters distribution was normal (0, 1) and, consequently, matched to the distribution of item difficulty. One hundred replications were generated for each of the sample size and test length conditions. Nine hundred GENIRV runs were needed to obtain the data sets for the study.

Insert Tables 7, 8, and 9 about here

Item Parameter Estimation

Each of the generated data sets was analyzed via the computer program BILOG (Mislevy & Bock, 1990) for MB, and via the computer program BUGS (Spiegelhalter et al., 1997) for Gibbs sampling. For example, the generated item response data set for the first replication of sample size 50 and test length 10 was analyzed by two different computer runs, on each for the MB and Gibbs sampling procedures.

For MB, a lognormal prior on item discrimination with mean 0 and variance .25 [i.e., $\log \alpha_j \sim N(0, .5^2)$] was used. This is the default prior specification in BILOG for estimation of item parameters in the 2PL model. The ability estimates were obtained by EAP estimation.

For the Gibbs sampling, an informative prior was used for λ_j and an uninformative prior for ζ_j. The prior distribution for λ_j was set to have a normal distribution with mean 0 and variance 1 [i.e., $\lambda_j \sim N(0,1)$] with range restricted to yield positive values of λ_j (i.e., $\lambda_j > 0$). The prior distribution for ζ_j was $\sim N(0, 100^2)$. The prior distribution for λ_j can be seen as a half normal distribution or the singly truncated normal distribution (Johnson, Kotz, & Balakrishnan, 1994). Since λ_j, without the range restriction, was sampled from a unit normal distribution, then $E(\lambda_j) = .798$ and $\mathrm{Var}(\lambda_j) = .363$ (standard deviation .603). The prior distribution for ζ_j, however, was similar to the uniform distribution defined on the entire real line. The priors for MB and Gibbs sampling were similar but not exactly the same.

Metric Transformation

In parameter recovery studies, such as the present one, comparisons between estimates and the underlying parameters require that the item parameter estimates obtained from different calibration runs be placed on a common metric with their underlying parameters (Baker & Al-Karni, 1991; Yen, 1987). Parameter estimation procedures under IRT yield metrics which are unique up to a linear transformation. To link both sets of estimates and parameters, it is necessary to determine the slope and intercept of the equating coefficients required for the transformation.

The estimates of the item parameters for each of the estimation procedures were placed on the scale of the true parameters before comparisons were made. The test characteristic curve method by Stocking and Lord (1983) as implemented in the computer program EQUATE (Baker, 1993) was used.

Evaluation Criteria

The evaluation of accuracy in this study involved three criteria: root mean square error (RMSE), bias, and correlation between estimates and parameters. The RMSE is the square root of the average of the squared differences between estimated and true values. For item discrimination, for example, the RMSE of item j is $\left\{(1/R)\sum_{k=1}^{R}(\hat{\alpha}_{jk} - \alpha_j)^2\right\}^{1/2}$, where R is the total number of replications (i.e, $R = 100$).

It is also useful to examine the bias, B, between the expected value of the estimates and the corresponding parameter. The bias of the item discrimination estimates, for example, is given as $B_{\alpha_j} = E(\hat{\alpha}_{jk}) - \alpha_j$, where the expectation is with regard to $k = 1(1)R$. This estimate of bias was obtained for both parameters in the model across the 100 replications.

Results

RMSEs for Item Parameters

RMSEs for item parameters of the 10-, 20-, and 40-item tests are reported in Tables 10, 11, and 12, respectively. As sample size increased, RMSEs for both item parameters decreased.

Insert Tables 10, 11, and 12 about here

The average RMSEs of the 10-, 20-, and 40-item tests are reported in Tables 13, 14, and 15, respectively. The patterns of the RMSE results were consistent across all tables. RMSE results are also presented graphically in Figures 5, 6, and 7.

Insert Tables 13, 14, and 15, and Figures 5, 6, and 7 about here

---

In Gibbs sampling, the RMSEs for item discrimination increased as the values of discrimination parameters increased. For MB, items with $\alpha_j = .73$ and $\alpha_j = 1.00$ yielded somewhat smaller RMSEs. Overall, MB consistently yielded smaller RMDSs than did Gibbs sampling. For item difficulty, the two extreme item difficulties $\beta_j = -1.83$ and $\beta_j = 1.83$ yielded larger RMSEs for both MB and Gibbs sampling. MB also yielded consistently smaller RMSEs for item difficulty for all conditions.

Bias Results for Item Parameters

The bias statistics for item discrimination and difficulty, presented in Tables 16, 17, and 18 for the 10-, 20-, and 40-item tests, appear to decrease as sample size increases.

Insert Tables 16, 17, and 18 about here

Tables 19, 20, and 21 summarize the average sizes of bias for different test lengths. Figures 8, 9, and 10 also present the bias results of the respective tests. Bias statistics decreased with an increase in sample size for item discrimination. When priors of item discriminations were used, it was expected that positive bias would be observed for the smaller item discrimination parameters (i.e., $\alpha_j = .45$ or $\alpha_j = .73$) and negative bias for the larger item discrimination parameters (i.e., $\alpha_j = 1.27$ and $\alpha_j = 1.55$). This shrinkage effect was observed mainly for MB and for Gibbs sampling, only for sample size 50.

Insert Tables 19, 20, and 21, and Figures 8, 9, and 10 about here

The bias patterns for item difficulty was somewhat different from the patterns for item discrimination. Items with negative difficulty parameters had negative bias whereas positive bias was observed for items with positive difficulty parameters. The same pattern was observed across the three test lengths. MB consistently yielded better bias results than did Gibbs sampling. The difference between the two methods decreased as the sample sizes increased.

Correlation Results for Item Parameters

The average correlations between true and estimated values of both item discrimination and item difficulty across 100 replications are given in Table 22. As sample sizes increased, the average correlations increased. Only minor differences occurred between the two estimation methods: Gibbs sampling yielded better results for item discrimination whereas MB yielded better results for item difficulty.

Insert Table 22 about here

RMSEs for Ability Parameters

The average RMSEs for ability parameters for 50, 100, and 200 examinees are reported in Tables 23, 24, and 25, respectively. As test length increased, RMSEs for ability parameters decreased.

Insert Tables 23, 24, and 25, and Figure 11 about here

Figure 11 summarizes the results from Tables 23, 24, and 25. When ability parameters were close to zero, Gibbs sampling yielded smaller RMSEs. For extreme ability parameters, MB yielded smaller RMSEs. RMSEs decreased around zero, that is, they were smaller around the mean of item difficulty parameters. RMSEs increased when ability parameters were not well matched with the mean of the item difficulty parameters.

Bias Results for Ability Parameters

Tables 26, 27, and 28 summarize the average sizes of bias from 50, 100, and 200 examinees. Figure 12 presents the bias results for the three sample sizes. For all sample sizes, an increase in test length was associated with a decrease in bias. Recall that both ability estimation used in Gibbs sampling and MB (i.e., EAP) employed priors for ability. It was expected that positive bias would be observed for the larger negative ability parameters and negative bias for the larger positive ability parameters. This shrinkage effect was observed, in fact, for all conditions. Increasing test length reduced the shrinkage effect. MB consistently yielded smaller bias across all conditions.

Insert Tables 26, 27, and 28, and Figure 12 about here

Correlation Results for Ability

The average correlations between true and estimated values of ability parameters over 100 replications are given in Table 29. As test lengths increased, average correlations increased. Differences in correlations were not associated with sample size. Gibbs sampling and MB yielded the same results.

Insert Table 29 about here

Discussion

Previous work using Gibbs sampling and MCMC methods suggests this method may provide a useful alternative method for estimation of IRT parameters 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 has not been thoroughly studied. The simulation results of this study indicate that MB via BILOG yielded better item and ability parameter estimates than Gibbs sampling. This is consistent with the results reported by Baker (1998).

The main difference between Gibbs sampling and the marginalized methods, MMLE and MBE, is in the way these methods obtain parameter estimates. Gibbs sampling uses the sample of parameter values to estimate the mean and variance of the posterior density of the parameter. Under MML and MB, the marginalized likelihood function and the marginalized posterior distribution, respectively, are maximized to obtain the marginal modes. Estimates of the ability parameters do not arise during the course of item parameter estimation under the marginalized methods. Instead, ability parameters are typically estimated after obtaining the item parameter estimates, under the assumption that the obtained estimates are true values. In the Gibbs sampling approach, ability parameters can be estimated jointly with item parameters as in this paper, and the method is similar, in this sense, to JML or JB. Note that ability can be obtained not jointly but after estimating item parameters in Gibbs sampling.

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/`

Gibbs sampling and general MCMC methods are likely to be more useful for situations where complicated models are employed. For example, Gibbs sampling could be usefully applied to the estimation of item and ability parameters in the hierarchical Bayes approach (Mislevy, 1986; Swaminathan & Gifford, 1982, 1985, 1986). In this study, priors were imposed directly on the parameters and the priors used for the Gibbs sampling and MB were not precisely the same. Accuracy of Gibbs sampling with different kinds of priors has not been investigated. This kind of research may be particularly valuable for small samples and short tests.

The focus in this paper was estimation of item and ability parameters in terms of RMSE and bias. In addition to RMSE and bias, future studies may also consider accuracy with respect to the posterior intervals of the estimates. This is because of the fact that one of the possible advantages of using Gibbs sampling or other MCMC methods is incorporation of uncertainly in item parameter estimates into estimation of ability parameters (e.g. Patz & Junker, 1997).

In this paper, we employed the 2PL model in the example and in the simulation section without addressing the problem of model selection and criticism. The model criticism for Gibbs sampling seems to be an important topic to investigate in future research. Also the evaluation of Gibbs sampling for other models including the three-parameter logistic model, the partial credit model, and the graded response model may provide guidelines for using the method under IRT.

References

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

Baker, F. B. (1982). *GENIRV: A program to generate item response vectors* [Computer program]. Madison, University of Wisconsin, Department of Educational Psychology, Laboratory of Experimental Design.

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. (1993). EQUATE 2.0: A computer program for the characteristic curve method of IRT equating. *Applied Psychological Measurement, 17,* 20.

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., & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measurement, 28,* 147–162.

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.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 4,* 431–444.

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.

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.

Harwell, M. R., & Janosky, J. E. (1991). An empirical study of the effects of small datasets and varying prior variance on item parameter estimation in BILOG. *Applied Psychological Measurement, 15,* 279–291.

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

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.

Leonard, T., & Hsu, J. S. J. (1994). *Bayesian methods for research specialists.* Unpublished manuscript.

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.

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.

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.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201–210.

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.

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.

Yen, W. M. (1983). Using simulation results to choose a latent trait model. *Applied Psychological Measurement, 5,* 245–262.

Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika, 52,* 275–291.

Table 1

Memory Test Data from Thissen (1982)

Examinee	Item 1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	1
7	0	0	0	0	0	0	0	0	1	1
8	0	0	0	0	0	0	0	0	1	1
9	0	0	0	0	0	0	0	0	1	1
10	0	0	0	0	0	0	0	1	0	1
11	0	0	0	0	0	0	0	1	0	1
12	0	0	0	0	0	1	0	0	0	1
13	0	0	0	0	1	0	0	0	0	1
14	0	0	0	0	1	0	0	0	1	0
15	0	0	1	0	0	0	0	0	0	1
16	0	0	0	0	0	0	0	1	1	1
17	0	0	0	0	0	0	0	1	1	1
18	0	0	0	0	0	0	1	0	1	1
19	0	0	1	0	0	0	0	1	0	1
20	0	0	1	0	0	0	1	0	0	1
21	0	1	0	0	0	1	0	1	0	0
22	1	0	0	0	0	0	0	0	1	1
23	1	0	0	0	0	0	1	0	0	1
24	1	0	0	1	0	0	0	0	1	0
25	0	0	0	0	0	0	1	1	1	1
26	0	0	0	0	0	1	0	1	1	1
27	0	0	0	0	0	1	0	1	1	1
28	0	0	0	0	1	0	1	0	1	1
29	0	0	0	1	0	0	1	0	1	1
30	0	0	0	1	0	0	1	1	0	1
31	0	1	0	0	0	0	0	1	1	1
32	0	1	0	0	0	1	0	0	1	1
33	0	1	0	0	1	0	0	1	1	0
34	0	1	0	0	0	0	1	1	1	1
35	1	0	0	0	0	1	1	1	0	1
36	1	0	0	1	1	0	1	1	0	0
37	1	1	0	0	1	0	0	1	0	1
38	0	1	0	0	0	1	1	1	1	1
39	1	1	0	0	1	1	0	1	0	1
40	0	1	1	1	1	0	0	1	1	1

Table 2

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

Run	Parameter λ_j	ζ_j
First	1	0
Second	10	5
Third	.1	−5

BEST COPY AVAILABLE

Table 3

Estimated Item Parameters and Standard Errors (s.e.) of the Memory Test Items

Item	BUGS: Gibbs Sampling-Informative λ_j (s.e.)	BUGS: Gibbs Sampling-Informative ζ_j (s.e.)	BUGS: Gibbs Sampling-Uninformative λ_j (s.e.)	BUGS: Gibbs Sampling-Uninformative ζ_j (s.e.)	BILOG: Marginal Bayesian λ_j (s.e.)	BILOG: Marginal Bayesian ζ_j (s.e.)	BILOG: Margianal Maximum Likelihood λ_j (s.e.)	BILOG: Margianal Maximum Likelihood ζ_j (s.e.)
1	.671 (.463)	−1.775 (.510)	.793 (.615)	−1.768 (.522)	.869 (.382)	−1.760 (.559)	2.344 (1.550)	−.525 (.938)
2	1.416 (.662)	−1.753 (.617)	27.800(22.320)	−16.860(14.660)	1.413 (.793)	−1.655 (.737)	6.066(30.895)	−5.595(13.719)
3	.521 (.419)	−2.484 (.614)	.728 (.604)	−2.488 (.630)	.769 (.323)	−2.403 (.659)	.255 (1.932)	−2.072 (1.730)
4	.700 (.511)	−2.264 (.617)	.843 (.667)	−2.275 (.622)	.906 (.409)	−2.208 (.635)	1.395 (3.164)	−1.619 (.863)
5	.782 (.512)	−1.640 (.504)	1.256 (.858)	−1.741 (.612)	.932 (.398)	−1.606 (.534)	1.153 (1.519)	−1.979 (.951)
6	.827 (.536)	−1.669 (.524)	1.733 (1.124)	−1.968 (.799)	.933 (.404)	−1.606 (.537)	.465 (.814)	−1.719 (.520)
7	.595 (.421)	−1.103 (.405)	.598 (.437)	−1.058 (.402)	.834 (.356)	−1.105 (.449)	.177 (.849)	−1.138 (.525)
8	1.380 (.633)	−.163 (.459)	14.520 (1.932)	−1.629 (4.836)	1.355 (.690)	−.153 (.472)	.761 (.985)	−.647 (.588)
9	.517 (.367)	−.007 (.345)	.701 (.480)	.006 (.361)	.747 (.301)	−.004 (.424)	2.168 (1.415)	1.105 (.922)
10	.727 (.477)	1.270 (.436)	1.040 (.647)	1.353 (.494)	.914 (.365)	1.270 (.505)	.624 (.910)	1.046 (1.049)

Table 4

Ability Estimates and Standard Errors (s.e.) of the Memory Test

Examinee	BUGS: GS-I θ_i	BUGS: GS-I (s.e.)	BUGS: GS-U θ_i	BUGS: GS-U (s.e.)	BILOG: MB/EAP θ_i	BILOG: MB/EAP (s.e.)	BILOG: MML/ML θ_i	BILOG: MML/ML (s.e.)
1	−1.167	(.788)	−1.198	(.728)	−1.309	(.738)	−3.968	(2.549)
2	−1.148	(.793)	−1.194	(.718)	−1.309	(.738)	−3.968	(2.549)
3	−1.148	(.779)	−1.189	(.723)	−1.309	(.738)	−3.968	(2.549)
4	−1.160	(.776)	−1.196	(.703)	−1.309	(.738)	−3.968	(2.549)
5	−1.144	(.780)	−1.187	(.722)	−1.309	(.738)	−3.968	(2.549)
6	−.773	(.751)	−.779	(.631)	−.840	(.695)	−1.873	(1.434)
7	−.509	(.734)	−.557	(.577)	−.495	(.666)	−.348	(.622)
8	−.516	(.737)	−.560	(.575)	−.495	(.666)	−.348	(.622)
9	−.516	(.754)	−.566	(.582)	−.495	(.666)	−.348	(.622)
10	−.129	(.712)	.121	(.448)	−.234	(.646)	−1.029	(.822)
11	−.135	(.709)	.114	(.461)	−.234	(.646)	−1.029	(.822)
12	−.366	(.752)	−.331	(.550)	−.414	(.659)	−1.259	(.948)
13	−.379	(.753)	−.432	(.563)	−.414	(.659)	−.797	(.727)
14	−.489	(.770)	−.520	(.598)	−.487	(.665)	−.152	(.597)
15	−.515	(.772)	−.557	(.596)	−.485	(.665)	−1.476	(1.097)
16	.066	(.702)	.203	(.408)	.069	(.625)	−.070	(.589)
17	.080	(.700)	.212	(.405)	.069	(.625)	−.070	(.589)
18	−.222	(.734)	−.399	(.529)	−.140	(.640)	−.281	(.612)
19	.116	(.714)	.200	(.415)	.077	(.625)	−.872	(.754)
20	−.241	(.737)	−.401	(.547)	−.131	(.639)	−1.289	(.967)
21	.478	(.746)	.890	(.396)	.329	(.609)	.753	(.328)
22	−.195	(.731)	−.366	(.525)	−.126	(.639)	.411	(.491)
23	−.157	(.731)	−.398	(.550)	−.090	(.636)	−.215	(.604)
24	−.195	(.782)	−.416	(.560)	−.129	(.639)	.568	(.412)
25	.330	(.687)	.260	(.385)	.385	(.607)	−.010	(.583)
26	.416	(.706)	.358	(.371)	.421	(.605)	.087	(.572)
27	.419	(.699)	.358	(.375)	.421	(.605)	.087	(.572)
28	.100	(.726)	−.176	(.477)	.227	(.615)	.120	(.568)
29	.066	(.744)	−.247	(.495)	.217	(.616)	.197	(.556)
30	.403	(.700)	.269	(.410)	.443	(.605)	−.285	(.613)
31	.641	(.707)	.884	(.377)	.595	(.601)	.971	(.303)
32	.430	(.701)	.556	(.522)	.442	(.605)	.944	(.301)
33	.659	(.722)	.905	(.397)	.602	(.601)	1.021	(.313)
34	.853	(.671)	.940	(.415)	.894	(.597)	.988	(.306)
35	.687	(.693)	.416	(.380)	.766	(.599)	.199	(.556)
36	.690	(.750)	.368	(.391)	.763	(.599)	.555	(.420)
37	.982	(.694)	1.024	(.437)	.972	(.596)	1.106	(.342)
38	1.189	(.683)	1.175	(.489)	1.223	(.592)	1.033	(.316)
39	1.302	(.716)	1.308	(.524)	1.300	(.592)	1.165	(.372)
40	1.415	(.711)	1.277	(.540)	1.519	(.597)	1.354	(.514)

Table 5

Estimated Item Parameters and 95% Posterior Intervals of the Memory Test Items

	Gibbs Sampling-Informative				Marginal Bayesian			
Item	λ_j	(Post. Interval)	ζ_j	(Post. Interval)	λ_j	(Post. Interval)	ζ_j	(Post. Interval)
1	.671	(.035, 1.759)	−1.775	(−2.881, −.883)	.869	(.120, 1.621)	−1.760	(−2.856, −.664)
2	1.416	(.219, 2.803)	−1.753	(−3.153, −.733)	1.413	(−.141, 2.974)	−1.655	(−3.100, −.210)
3	.521	(.019, 1.551)	−2.484	(−3.826, −1.434)	.769	(.136, 1.405)	−2.403	(−3.695, −1.111)
4	.700	(.033, 1.894)	−2.264	(−3.597, −1.186)	.906	(.104, 1.711)	−2.208	(−3.453, −.963)
5	.782	(.045, 1.936)	−1.640	(−2.740, −.752)	.932	(.152, 1.716)	−1.606	(−2.653, −.559)
6	.827	(.050, 2.086)	−1.669	(−2.842, −.757)	.933	(.141, 1.728)	−1.606	(−2.659, −.553)
7	.595	(.029, 1.613)	−1.103	(−1.947, −.371)	.834	(.136, 1.535)	−1.105	(−1.985, −.225)
8	1.380	(.272, 2.765)	−.163	(−1.089, .739)	1.355	(.003, 2.714)	−.153	(−1.078, .772)
9	.517	(.027, 1.405)	−.007	(−.694, .670)	.747	(.157, 1.340)	−.004	(−.835, .827)
10	.727	(.045, 1.819)	1.270	(.492, 2.182)	.914	(.199, 1.633)	1.270	(.280, 2.260)

Table 6

Ability Estimates and 95% Posterior Intervals of the Memory Test

	Gibbs Sampling-Informative		MML/Expected A Posteriori	
Examinee	θ_i	Posterior Interval	θ_i	Posterior Interval
1	−1.167	(−2.736, .339)	−1.309	(−2.755, .138)
2	−1.148	(−2.788, .334)	−1.309	(−2.755, .138)
3	−1.148	(−2.716, .324)	−1.309	(−2.755, .138)
4	−1.160	(−2.772, .290)	−1.309	(−2.755, .138)
5	−1.144	(−2.732, .324)	−1.309	(−2.755, .138)
6	−.773	(−2.366, .610)	−.840	(−2.202, .522)
7	−.509	(−2.027, .883)	−.495	(−1.799, .809)
8	−.516	(−2.037, .859)	−.495	(−1.799, .809)
9	−.516	(−2.075, .870)	−.495	(−1.799, .809)
10	−.129	(−1.589, 1.216)	−.234	(−1.500, 1.033)
11	−.135	(−1.630, 1.141)	−.234	(−1.500, 1.033)
12	−.366	(−1.943, 1.003)	−.414	(−1.706, .879)
13	−.379	(−1.917, 1.071)	−.414	(−1.706, .878)
14	−.489	(−2.081, .975)	−.487	(−1.790, .816)
15	−.515	(−2.089, .960)	−.485	(−1.788, .818)
16	.066	(−1.420, 1.408)	.069	(−1.157, 1.294)
17	.080	(−1.359, 1.440)	.069	(−1.157, 1.294)
18	−.222	(−1.716, 1.197)	−.140	(−1.394, 1.114)
19	.116	(−1.339, 1.533)	.077	(−1.148, 1.302)
20	−.241	(−1.734, 1.167)	−.131	(−1.384, 1.122)
21	.478	(−1.084, 1.854)	.329	(−.865, 1.524)
22	−.195	(−1.695, 1.187)	−.126	(−1.378, 1.126)
23	−.157	(−1.620, 1.277)	−.090	(−1.338, 1.157)
24	−.195	(−1.765, 1.309)	−.129	(−1.382, 1.124)
25	.330	(−1.093, 1.616)	.385	(−.805, 1.574)
26	.416	(−1.034, 1.781)	.421	(−.766, 1.607)
27	.419	(−.966, 1.763)	.421	(−.766, 1.607)
28	.100	(−1.393, 1.508)	.227	(−.979, 1.432)
29	.066	(−1.419, 1.509)	.217	(−.990, 1.423)
30	.403	(−.970, 1.800)	.443	(−.742, 1.628)
31	.641	(−.747, 2.018)	.595	(−.582, 1.772)
32	.430	(−.974, 1.789)	.442	(−.743, 1.627)
33	.659	(−.839, 2.045)	.602	(−.576, 1.779)
34	.853	(−.486, 2.154)	.894	(−.276, 2.064)
35	.687	(−.681, 2.007)	.766	(−.407, 1.939)
36	.690	(−.813, 2.139)	.763	(−.410, 1.936)
37	.982	(−.379, 2.322)	.972	(−.195, 2.139)
38	1.189	(−.138, 2.545)	1.223	(.063, 2.384)
39	1.302	(−.094, 2.722)	1.300	(.140, 2.460)
40	1.415	(.033, 2.826)	1.519	(.349, 2.689)

Table 7
Item Parameters of the 10 Item Test

Item	Parameter	
	α_j	β_j
1	.45	.00
2	.73	−.91
3	.73	.91
4	1.00	−1.83
5	1.00	.00
6	1.00	.00
7	1.00	1.83
8	1.27	−.91
9	1.27	.91
10	1.55	.00

Table 8
Item Parameters of the 20 Item Test

Item	Parameter	
	α_j	β_j
1	.45	−.91
2	.45	.91
3	.73	−1.83
4	.73	.00
5	.73	.00
6	.73	1.83
7	1.00	−.91
8	1.00	−.91
9	1.00	.00
10	1.00	.00
11	1.00	.00
12	1.00	.00
13	1.00	.91
14	1.00	.91
15	1.27	−1.83
16	1.27	.00
17	1.27	.00
18	1.27	1.83
19	1.55	−.91
20	1.55	.91

Table 9
Item Parameters of the 40 Item Test

Item	Parameter α_j	Parameter β_j
1	.45	−.91
2	.45	.00
3	.45	.00
4	.45	.91
5	.73	−1.83
6	.73	−.91
7	.73	−.91
8	.73	.00
9	.73	.00
10	.73	.91
11	.73	.91
12	.73	1.83
13	1.00	−1.83
14	1.00	−1.83
15	1.00	−.91
16	1.00	−.91
17	1.00	.00
18	1.00	.00
19	1.00	.00
20	1.00	.00
21	1.00	.00
22	1.00	.00
23	1.00	.00
24	1.00	.00
25	1.00	.91
26	1.00	.91
27	1.00	1.83
28	1.00	1.83
29	1.27	−1.83
30	1.27	−.91
31	1.27	−.91
32	1.27	.00
33	1.27	.00
34	1.27	.91
35	1.27	.91
36	1.27	1.83
37	1.55	−.91
38	1.55	.00
39	1.55	.00
40	1.55	.91

Table 10

Root Mean Square Errors of the 10 Item Test

	Gibbs Sampling						Marginal Bayesian					
	$N = 50$		$N = 100$		$N = 200$		$N = 50$		$N = 100$		$N = 200$	
Item	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j
1	.358	.585	.281	.491	.189	.382	.338	.433	.273	.322	.196	.248
2	.357	.573	.305	.418	.231	.298	.242	.404	.219	.294	.177	.239
3	.365	.507	.335	.426	.242	.300	.257	.383	.236	.312	.184	.217
4	.381	.861	.372	.679	.290	.524	.245	.487	.260	.422	.222	.375
5	.412	.271	.342	.198	.242	.141	.257	.273	.226	.200	.181	.144
6	.472	.343	.370	.206	.269	.163	.311	.337	.255	.208	.206	.165
7	.358	.827	.365	.603	.313	.529	.217	.438	.253	.391	.228	.332
8	.400	.428	.396	.276	.313	.218	.311	.384	.310	.264	.261	.207
9	.425	.452	.391	.293	.290	.194	.323	.367	.300	.281	.263	.196
10	.420	.260	.361	.149	.330	.124	.425	.266	.374	.161	.316	.130

Table 11

Root Mean Square Errors of the 20 Item Test

	Gibbs Sampling						Marginal Bayesian					
	$N = 50$		$N = 100$		$N = 200$		$N = 50$		$N = 100$		$N = 200$	
Item	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j
1	.396	.719	.233	.694	.161	.572	.358	.500	.236	.389	.166	.309
2	.344	.856	.260	.578	.170	.592	.320	.521	.255	.377	.175	.341
3	.377	.842	.299	.727	.186	.531	.281	.499	.220	.387	.141	.313
4	.389	.480	.341	.381	.202	.197	.269	.379	.254	.302	.164	.189
5	.369	.436	.314	.277	.219	.205	.247	.371	.234	.260	.180	.197
6	.429	1.016	.280	.831	.205	.697	.301	.529	.202	.405	.155	.396
7	.380	.460	.341	.331	.208	.235	.243	.376	.244	.286	.162	.220
8	.378	.388	.333	.326	.246	.239	.248	.356	.242	.291	.199	.209
9	.314	.330	.282	.214	.243	.169	.200	.324	.206	.212	.202	.172
10	.391	.327	.323	.234	.223	.139	.257	.327	.231	.232	.181	.143
11	.381	.308	.345	.234	.237	.163	.270	.305	.243	.233	.195	.167
12	.446	.348	.365	.254	.228	.152	.316	.343	.265	.254	.182	.157
13	.406	.483	.329	.274	.231	.240	.278	.418	.232	.228	.184	.219
14	.425	.716	.292	.354	.215	.226	.269	.432	.206	.299	.170	.213
15	.443	1.034	.432	.744	.292	.360	.336	.672	.353	.533	.258	.336
16	.438	.264	.344	.168	.240	.127	.327	.278	.273	.181	.197	.134
17	.409	.255	.311	.192	.275	.127	.325	.270	.265	.204	.237	.133
18	.403	.819	.394	.645	.274	.406	.312	.588	.314	.456	.237	.375
19	.426	.335	.442	.279	.340	.178	.436	.360	.408	.283	.314	.192
20	.382	.315	.368	.223	.361	.207	.374	.327	.337	.224	.333	.216

Table 12

Root Mean Square Errors of the 40 Item Test

	Gibbs Sampling						Marginal Bayesian					
	$N = 50$		$N = 100$		$N = 200$		$N = 50$		$N = 100$		$N = 200$	
Item	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j
1	.351	.800	.253	.665	.158	.427	.327	.535	.250	.398	.150	.288
2	.362	.642	.258	.461	.183	.325	.335	.489	.256	.339	.185	.264
3	.369	.648	.221	.494	.151	.294	.341	.462	.229	.366	.154	.240
4	.311	.838	.206	.646	.152	.511	.306	.564	.209	.400	.150	.352
5	.380	.956	.311	.903	.213	.598	.269	.530	.231	.459	.170	.369
6	.337	.556	.287	.425	.205	.283	.240	.399	.214	.300	.167	.242
7	.344	.639	.283	.659	.193	.321	.237	.487	.212	.393	.158	.269
8	.357	.531	.219	.303	.191	.240	.253	.436	.160	.287	.155	.231
9	.338	.429	.306	.308	.199	.203	.231	.386	.233	.285	.161	.195
10	.364	.572	.266	.566	.176	.280	.260	.422	.193	.355	.143	.237
11	.383	.588	.240	.573	.185	.358	.276	.471	.172	.320	.146	.275
12	.329	.980	.296	.824	.239	.628	.232	.536	.218	.465	.189	.388
13	.415	.717	.322	.685	.279	.446	.285	.465	.242	.464	.232	.361
14	.398	1.060	.307	.649	.253	.424	.253	.574	.221	.441	.203	.341
15	.413	.495	.316	.351	.229	.210	.281	.381	.231	.295	.182	.187
16	.426	.557	.304	.489	.259	.299	.298	.443	.226	.370	.215	.243
17	.382	.326	.311	.204	.184	.156	.251	.331	.218	.206	.154	.159
18	.356	.324	.292	.255	.212	.151	.229	.308	.228	.259	.178	.154
19	.397	.324	.259	.234	.215	.168	.291	.320	.195	.240	.176	.173
20	.401	.346	.326	.251	.200	.158	.254	.356	.254	.251	.169	.162
21	.370	.331	.293	.210	.233	.133	.251	.329	.217	.218	.187	.138
22	.365	.318	.317	.238	.191	.165	.242	.326	.243	.244	.155	.170
23	.363	.368	.267	.305	.199	.168	.250	.348	.207	.266	.172	.170
24	.372	.436	.318	.219	.233	.135	.242	.381	.241	.225	.190	.139
25	.412	.510	.364	.305	.233	.253	.288	.410	.274	.278	.187	.232
26	.343	.550	.304	.351	.207	.244	.229	.391	.225	.304	.173	.226
27	.429	.780	.337	.645	.242	.428	.299	.519	.243	.428	.195	.322
28	.402	.838	.291	.626	.218	.397	.268	.515	.208	.457	.173	.321
29	.433	1.056	.427	.691	.310	.506	.330	.719	.356	.521	.268	.430
30	.427	.362	.324	.231	.217	.158	.340	.336	.263	.231	.194	.166
31	.402	.414	.311	.269	.276	.172	.306	.382	.252	.269	.241	.173
32	.419	.342	.277	.210	.213	.143	.325	.343	.229	.226	.191	.150
33	.435	.262	.328	.186	.210	.138	.318	.278	.264	.198	.183	.146
34	.370	.398	.313	.257	.268	.175	.298	.384	.258	.271	.235	.177
35	.419	.371	.373	.320	.247	.179	.311	.375	.301	.285	.238	.190
36	.402	.787	.376	.609	.277	.313	.308	.627	.315	.492	.245	.302
37	.414	.365	.374	.230	.314	.157	.381	.391	.373	.252	.299	.168
38	.417	.234	.310	.162	.276	.114	.386	.258	.316	.175	.257	.119
39	.398	.234	.341	.150	.266	.111	.378	.254	.335	.160	.254	.118
40	.405	.293	.331	.218	.278	.154	.381	.318	.302	.240	.259	.181

BEST COPY AVAILABLE

Table 13
Average Root Mean Square Errors of the 10 Item Test

Parameter	Gibbs Sampling			Marginal Bayesian		
	$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$
$\alpha_j = .45$	.358	.281	.189	.338	.273	.196
.73	.361	.320	.237	.250	.228	.181
1.00	.406	.362	.279	.258	.249	.209
1.27	.413	.394	.302	.317	.305	.262
1.55	.420	.361	.330	.425	.374	.316
$\beta_j = -1.83$	.861	.679	.524	.487	.422	.375
−.91	.501	.347	.258	.394	.279	.223
.00	.365	.261	.203	.327	.223	.172
.91	.480	.360	.247	.375	.297	.207
1.83	.827	.603	.529	.438	.391	.332

Table 14
Average Root Mean Square Errors of the 20 Item Test

Parameter	Gibbs Sampling			Marginal Bayesian		
	$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$
$\alpha_j = .45$	.370	.247	.166	.339	.246	.171
.73	.391	.309	.203	.275	.228	.160
1.00	.390	.326	.229	.260	.234	.184
1.27	.423	.370	.270	.325	.301	.232
1.55	.404	.405	.351	.405	.373	.324
$\beta_j = -1.83$	.938	.736	.446	.586	.460	.325
−.91	.476	.408	.306	.398	.312	.233
.00	.344	.244	.160	.325	.235	.162
.91	.593	.357	.316	.425	.282	.247
1.83	.918	.738	.552	.559	.431	.386

Table 15
Average Root Mean Square Errors of the 40 Item Test

Parameter	Gibbs Sampling			Marginal Bayesian		
	$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$
$\alpha_j = .45$	.348	.235	.161	.327	.236	.160
.73	.354	.276	.200	.250	.204	.161
1.00	.390	.308	.224	.263	.230	.184
1.27	.413	.341	.252	.317	.280	.224
1.55	.409	.339	.284	.382	.332	.267
$\beta_j = -1.83$	.947	.732	.494	.572	.471	.375
−.91	.524	.415	.253	.419	.314	.217
.00	.381	.262	.175	.350	.247	.171
.91	.515	.405	.269	.417	.307	.234
1.83	.846	.676	.442	.549	.461	.333

Table 16

Bias Results of the 10 Item Test

	Gibbs Sampling						Marginal Bayesian					
	$N = 50$		$N = 100$		$N = 200$		$N = 50$		$N = 100$		$N = 200$	
Item	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j
1	.200	−.045	.107	−.026	.059	.005	.285	−.034	.214	−.024	.153	−.008
2	.135	−.029	.071	−.008	.065	.022	.136	.068	.091	.073	.075	.061
3	.105	.048	.094	.054	.055	.050	.124	−.059	.106	−.027	.070	−.003
4	.054	−.255	.046	−.212	.018	−.154	.001	−.143	.006	−.155	−.003	−.126
5	.148	.000	.105	.019	.080	.011	.044	−.002	.020	.015	.023	.010
6	.187	.019	.080	−.016	.048	−.009	.076	.012	.002	−.020	−.007	−.008
7	.073	.220	.103	.098	.091	.058	.005	.144	.041	.087	.045	.060
8	.039	−.083	.063	−.028	.021	−.036	−.106	−.136	−.079	−.096	−.074	−.084
9	−.005	.100	.075	.029	−.026	.050	−.136	.127	−.064	.092	−.110	.096
10	−.108	.026	−.033	.009	.010	−.018	−.290	.023	−.213	.010	−.116	−.021

Table 17

Bias Results of the 20 Item Test

	Gibbs Sampling						Marginal Bayesian					
	$N = 50$		$N = 100$		$N = 200$		$N = 50$		$N = 100$		$N = 200$	
Item	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j
1	.235	.048	.083	−.102	.034	−.136	.302	.237	.189	.164	.127	.101
2	.176	.015	.095	.087	.040	.094	.266	−.218	.198	−.154	.132	−.134
3	.134	−.144	.049	−.181	−.005	−.124	.153	.074	.086	.039	.033	.019
4	.162	.017	.103	−.008	.044	−.010	.154	.002	.106	−.013	.055	−.010
5	.132	.041	.100	.031	.057	.016	.133	.029	.105	.023	.066	.012
6	.128	.125	.054	.166	.012	.148	.149	−.182	.087	−.072	.046	−.015
7	.102	−.015	.126	.011	.063	−.016	.018	−.020	.048	−.017	.016	−.045
8	.107	−.029	.043	−.033	.030	.019	.025	−.048	−.011	−.047	−.010	.002
9	.052	.014	.043	.027	.051	.014	−.019	.011	−.020	.027	.008	.014
10	.132	.059	.090	−.022	.047	−.011	.038	.058	.021	−.023	.003	−.011
11	.095	.009	.101	.005	.046	.034	.012	.003	.025	.004	.002	.036
12	.100	.044	.059	.012	.056	−.021	.022	.043	−.004	.011	.008	−.021
13	.109	.055	.098	.024	.050	−.012	.029	.057	.026	.051	.008	.011
14	.081	.189	.034	.114	.042	.013	.009	.119	−.021	.126	−.001	.037
15	−.033	−.451	.043	−.232	.007	−.087	−.121	−.371	−.044	−.247	−.058	−.149
16	.108	.023	.105	.002	.079	.001	−.051	.024	−.032	.001	−.007	.002
17	.024	−.007	.034	.005	.060	−.012	−.114	−.004	−.086	.006	−.027	−.012
18	−.024	.240	.002	.135	−.004	.117	−.126	.235	−.089	.167	−.070	.177
19	−.100	−.099	.033	−.040	.027	−.019	−.264	−.180	−.117	−.111	−.081	−.072
20	−.026	.047	.025	−.026	.037	.021	−.215	.132	−.137	.047	−.070	.073

Table 18

Bias Results of the 40 Item Test

	Gibbs Sampling						Marginal Bayesian					
	$N = 50$		$N = 100$		$N = 200$		$N = 50$		$N = 100$		$N = 200$	
Item	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j
1	.195	.028	.096	−.107	.009	−.103	.275	.230	.194	.153	.103	.114
2	.190	.054	.107	.005	.060	−.006	.276	.041	.200	.010	.137	.000
3	.183	.114	.098	.030	.030	.012	.274	.079	.189	.011	.115	.003
4	.168	−.098	.053	.063	.014	.041	.262	−.281	.165	−.196	.107	−.173
5	.161	−.146	.047	−.229	.022	−.126	.163	.065	.091	.044	.053	.016
6	.124	−.037	.085	−.028	.046	−.007	.131	.057	.094	.034	.058	.025
7	.082	−.016	.081	−.058	.055	−.008	.105	.090	.094	.040	.061	.019
8	.085	.103	.038	.046	.040	.022	.107	.099	.056	.051	.047	.021
9	.138	−.034	.115	−.027	.047	−.023	.133	−.028	.113	−.031	.053	−.020
10	.139	−.062	.048	.038	.020	.000	.137	−.143	.071	−.057	.037	−.034
11	.160	.001	.032	.154	.038	.053	.155	−.075	.058	.047	.051	.010
12	.104	.179	.065	.141	.041	.076	.125	−.071	.096	−.082	.065	−.058
13	.122	−.132	.057	−.167	.027	−.105	.050	−.091	.020	−.138	.005	−.107
14	.084	−.266	.047	−.118	.047	−.032	.025	−.142	.009	−.093	.018	−.046
15	.106	−.069	.101	−.013	.074	.020	.030	−.057	.032	−.036	.031	−.003
16	.133	−.097	.047	−.087	.023	−.033	.053	−.088	−.005	−.088	−.010	−.042
17	.121	.021	.109	−.002	.038	−.001	.029	.025	.032	−.006	−.003	−.002
18	.095	−.030	.042	−.012	.022	−.024	.015	−.027	−.014	−.018	−.021	−.023
19	.082	−.013	.063	.016	.051	.000	.010	−.001	.001	.017	.008	.000
20	.157	−.055	.049	−.002	.017	−.014	.048	−.056	−.010	−.001	−.021	−.013
21	.089	.011	.065	−.019	.066	−.007	.011	.008	.001	−.019	.014	−.008
22	.095	.024	.097	.003	.045	−.005	.006	.024	.025	.002	−.001	−.006
23	.004	−.002	.006	−.049	−.004	−.017	−.043	−.004	−.038	−.043	−.040	−.017
24	.085	.001	.075	.009	.049	−.005	.009	−.003	.012	.007	.004	−.006
25	.093	.107	.117	.012	.070	.015	.023	.095	.048	.038	.025	.039
26	−.035	.177	.041	.061	.009	.061	−.068	.125	−.010	.073	−.024	.081
27	.139	.140	.086	.064	.016	.100	.072	.083	.040	.041	−.004	.097
28	.102	.170	.032	.144	.053	.004	.040	.093	−.006	.125	.024	.025
29	−.065	−.438	−.066	−.273	−.003	−.123	−.146	−.367	−.118	−.261	−.059	−.162
30	.093	−.037	.076	.012	.031	.007	−.053	−.097	−.037	−.048	−.046	−.038
31	.051	−.055	.053	.006	.065	.030	−.085	−.104	−.062	−.054	−.012	−.014
32	.029	.013	.059	−.007	.038	.005	−.110	.013	−.061	−.008	−.037	.006
33	.119	.035	.084	.021	.043	.000	−.041	.040	−.035	.021	−.039	.000
34	.000	.101	.063	.032	.048	−.017	−.124	.154	−.063	.102	−.028	.026
35	.090	.030	.040	.023	.010	.017	−.073	.100	−.066	.067	−.058	.061
36	−.005	.310	.017	.181	.011	.060	−.101	.315	−.062	.223	−.047	.118
37	−.009	−.093	.007	−.021	.008	−.010	−.198	−.180	−.130	−.095	−.090	−.059
38	.037	−.022	−.013	.012	.042	.011	−.173	−.022	−.172	.012	−.062	.012
39	.000	−.015	−.014	.003	.048	.004	−.202	−.015	−.159	.004	−.060	.004
40	.026	.015	.063	.001	.031	.048	−.168	.107	−.096	.078	−.069	.100

Table 19

Average Bias Results of the 10 Item Test

Parameter	Gibbs Sampling $N = 50$	Gibbs Sampling $N = 100$	Gibbs Sampling $N = 200$	Marginal Bayesian $N = 50$	Marginal Bayesian $N = 100$	Marginal Bayesian $N = 200$
$\alpha_j = .45$	.200	.107	.059	.285	.214	.153
.73	.120	.083	.060	.130	.099	.073
1.00	.116	.084	.059	.032	.017	.015
1.27	.017	.069	−.003	−.121	−.072	−.092
1.55	−.108	−.033	.010	−.290	−.213	−.116
$\beta_j = -1.83$	−.255	−.212	−.154	−.143	−.155	−.126
−.91	−.056	−.018	−.007	−.034	−.012	−.012
.00	−.000	−.004	−.003	−.000	−.005	−.007
.91	.074	.042	.050	.034	.033	.047
1.83	.220	.098	.058	.144	.087	.060

Table 20

Average Bias Results of the 20 Item Test

Parameter	Gibbs Sampling $N = 50$	Gibbs Sampling $N = 100$	Gibbs Sampling $N = 200$	Marginal Bayesian $N = 50$	Marginal Bayesian $N = 100$	Marginal Bayesian $N = 200$
$\alpha_j = .45$	.206	.089	.037	.284	.194	.130
.73	.139	.077	.027	.147	.096	.050
1.00	.097	.074	.048	.017	.008	.004
1.27	.019	.046	.036	−.103	−.063	−.041
1.55	−.063	.029	.032	−.240	−.127	−.076
$\beta_j = -1.83$	−.298	−.207	−.106	−.149	−.104	−.065
−.91	−.024	−.041	−.038	−.003	−.003	−.004
.00	.025	.007	.001	.021	.005	.001
.91	.077	.050	.029	.023	.018	−.003
1.83	.183	.151	.133	.027	.048	.081

Table 21

Average Bias Results of the 40 Item Test

Parameter	Gibbs Sampling $N = 50$	Gibbs Sampling $N = 100$	Gibbs Sampling $N = 200$	Marginal Bayesian $N = 50$	Marginal Bayesian $N = 100$	Marginal Bayesian $N = 200$
$\alpha_j = .45$	.184	.089	.028	.272	.187	.116
.73	.124	.064	.039	.132	.084	.053
1.00	.092	.065	.038	.019	.009	.000
1.27	.039	.041	.030	−.092	−.063	−.041
1.55	.014	.011	.032	−.185	−.139	−.070
$\beta_j = -1.83$	−.246	−.197	−.097	−.134	−.112	−.075
−.91	−.047	−.037	−.013	−.019	−.012	.000
.00	.013	.002	−.003	.011	.001	−.003
.91	.034	.048	.027	.010	.019	.014
1.83	.200	.133	.060	.105	.077	.046

Table 22

Average Correlations Between Item Parameters and Estimates over 100 Replications

Test	Gibbs $N = 50$ $r_{\alpha\hat{\alpha}}$	Gibbs $N = 50$ $r_{\beta\hat{\beta}}$	Gibbs $N = 100$ $r_{\alpha\hat{\alpha}}$	Gibbs $N = 100$ $r_{\beta\hat{\beta}}$	Gibbs $N = 200$ $r_{\alpha\hat{\alpha}}$	Gibbs $N = 200$ $r_{\beta\hat{\beta}}$	Marginal Bayesian $N = 50$ $r_{\alpha\hat{\alpha}}$	Marginal Bayesian $N = 50$ $r_{\beta\hat{\beta}}$	Marginal Bayesian $N = 100$ $r_{\alpha\hat{\alpha}}$	Marginal Bayesian $N = 100$ $r_{\beta\hat{\beta}}$	Marginal Bayesian $N = 200$ $r_{\alpha\hat{\alpha}}$	Marginal Bayesian $N = 200$ $r_{\beta\hat{\beta}}$
10-Item	.503	.920	.624	.950	.737	.968	.499	.948	.615	.969	.738	.980
20-item	.521	.899	.658	.937	.788	.961	.520	.930	.653	.960	.782	.975
40-item	.561	.892	.686	.927	.801	.963	.554	.927	.679	.955	.797	.974

Table 23
Average Root Mean Square Errors of Ability for 50 Examinees

	Gibbs Sampling			Marginal Bayesian		
θ	$n = 10$	$n = 20$	$n = 40$	$n = 10$	$n = 20$	$n = 40$
−2.5	1.284	.962	.679	1.059	.745	.500
−2.0	.974	.730	.550	.812	.582	.433
−1.5	.726	.572	.434	.646	.508	.386
−1.0	.597	.469	.368	.586	.470	.381
−.5	.509	.437	.321	.559	.480	.355
.0	.507	.420	.309	.585	.478	.354
.5	.521	.441	.322	.579	.479	.353
1.0	.574	.493	.370	.566	.494	.371
1.5	.729	.529	.429	.635	.466	.366
2.0	.863	.691	.555	.697	.544	.437
2.5	1.248	.961	.696	1.022	.740	.519

Table 24
Average Root Mean Square Errors of Ability for 100 Examinees

	Gibbs Sampling			Marginal Bayesian		
θ	$n = 10$	$n = 20$	$n = 40$	$n = 10$	$n = 20$	$n = 40$
−2.5	1.265	.928	.651	1.086	.773	.523
−2.0	.963	.691	.543	.840	.590	.456
−1.5	.732	.558	.434	.664	.509	.404
−1.0	.589	.470	.366	.584	.475	.371
−.5	.509	.418	.319	.551	.448	.338
.0	.481	.408	.307	.536	.452	.338
.5	.524	.406	.327	.563	.434	.349
1.0	.588	.463	.372	.581	.463	.375
1.5	.737	.560	.428	.676	.511	.394
2.0	.950	.717	.467	.823	.616	.392
2.5	1.247	.937	.631	1.075	.776	.505

Table 25
Average Root Mean Square Errors of Ability for 200 Examinees

	Gibbs Sampling			Marginal Bayesian		
θ	$n = 10$	$n = 20$	$n = 40$	$n = 10$	$n = 20$	$n = 40$
−2.5	1.218	.885	.630	1.112	.795	.556
−2.0	.936	.669	.490	.859	.608	.444
−1.5	.703	.532	.407	.662	.508	.388
−1.0	.571	.451	.343	.570	.454	.343
−.5	.514	.419	.326	.540	.437	.339
.0	.502	.412	.317	.536	.440	.336
.5	.503	.421	.315	.529	.438	.328
1.0	.563	.465	.342	.560	.467	.345
1.5	.701	.542	.406	.663	.516	.386
2.0	.898	.647	.479	.824	.581	.434
2.5	1.192	.871	.604	1.091	.776	.527

Table 26

Average Bias Results of Ability for 50 Examinees

θ	Gibbs Sampling $n = 10$	Gibbs Sampling $n = 20$	Gibbs Sampling $n = 40$	Marginal Bayesian $n = 10$	Marginal Bayesian $n = 20$	Marginal Bayesian $n = 40$
−2.5	1.233	.892	.597	.987	.633	.353
−2.0	.913	.609	.428	.713	.393	.220
−1.5	.591	.392	.257	.427	.219	.086
−1.0	.390	.230	.129	.273	.112	.005
−.5	.182	.104	.059	.127	.039	−.006
.0	−.012	−.012	−.004	−.014	−.012	−.001
.5	−.147	−.135	−.068	−.090	−.077	.001
1.0	−.354	−.246	−.166	−.244	−.128	−.042
1.5	−.600	−.355	−.287	−.431	−.178	−.111
2.0	−.763	−.595	−.424	−.535	−.375	−.206
2.5	−1.191	−.890	−.589	−.942	−.625	−.334

Table 27

Average Bias Results of Ability for 100 Examinees

θ	Gibbs Sampling $n = 10$	Gibbs Sampling $n = 20$	Gibbs Sampling $n = 40$	Marginal Bayesian $n = 10$	Marginal Bayesian $n = 20$	Marginal Bayesian $n = 40$
−2.5	1.214	.844	.560	1.019	.657	.393
−2.0	.882	.565	.399	.722	.409	.254
−1.5	.595	.381	.231	.469	.257	.111
−1.0	.360	.211	.126	.274	.124	.040
−.5	.140	.090	.078	.092	.042	.036
.0	−.017	.000	−.008	−.019	−.000	−.009
.5	−.186	−.100	−.063	−.143	−.054	−.020
1.0	−.365	−.232	−.136	−.278	−.145	−.054
1.5	−.584	−.383	−.229	−.459	−.257	−.111
2.0	−.869	−.581	−.317	−.708	−.425	−.170
2.5	−1.194	−.869	−.531	−1.000	−.687	−.364

Table 28

Average Bias Results of Ability for 200 Examinees

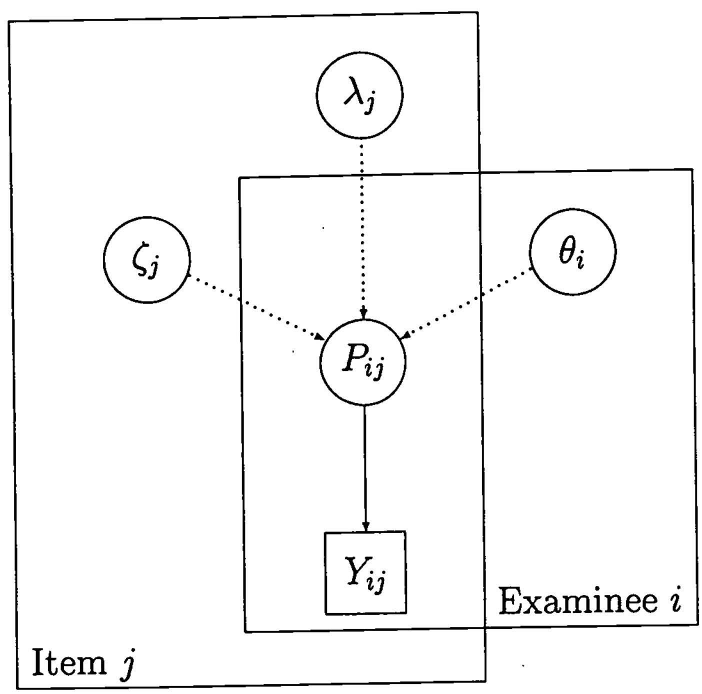
θ	Gibbs Sampling $n = 10$	Gibbs Sampling $n = 20$	Gibbs Sampling $n = 40$	Marginal Bayesian $n = 10$	Marginal Bayesian $n = 20$	Marginal Bayesian $n = 40$
−2.5	1.162	.812	.530	1.048	.703	.435
−2.0	.841	.537	.334	.743	.443	.249
−1.5	.551	.329	.201	.474	.254	.130
−1.0	.313	.190	.126	.258	.138	.076
−.5	.140	.092	.051	.110	.064	.025
.0	.009	−.010	−.000	.010	−.010	.000
.5	−.140	−.104	−.054	−.112	−.075	−.027
1.0	−.330	−.210	−.106	−.277	−.157	−.054
1.5	−.545	−.346	−.209	−.469	−.269	−.138
2.0	−.802	−.526	−.308	−.703	−.431	−.221
2.5	−1.138	−.796	−.521	−1.026	−.684	−.423

Table 29

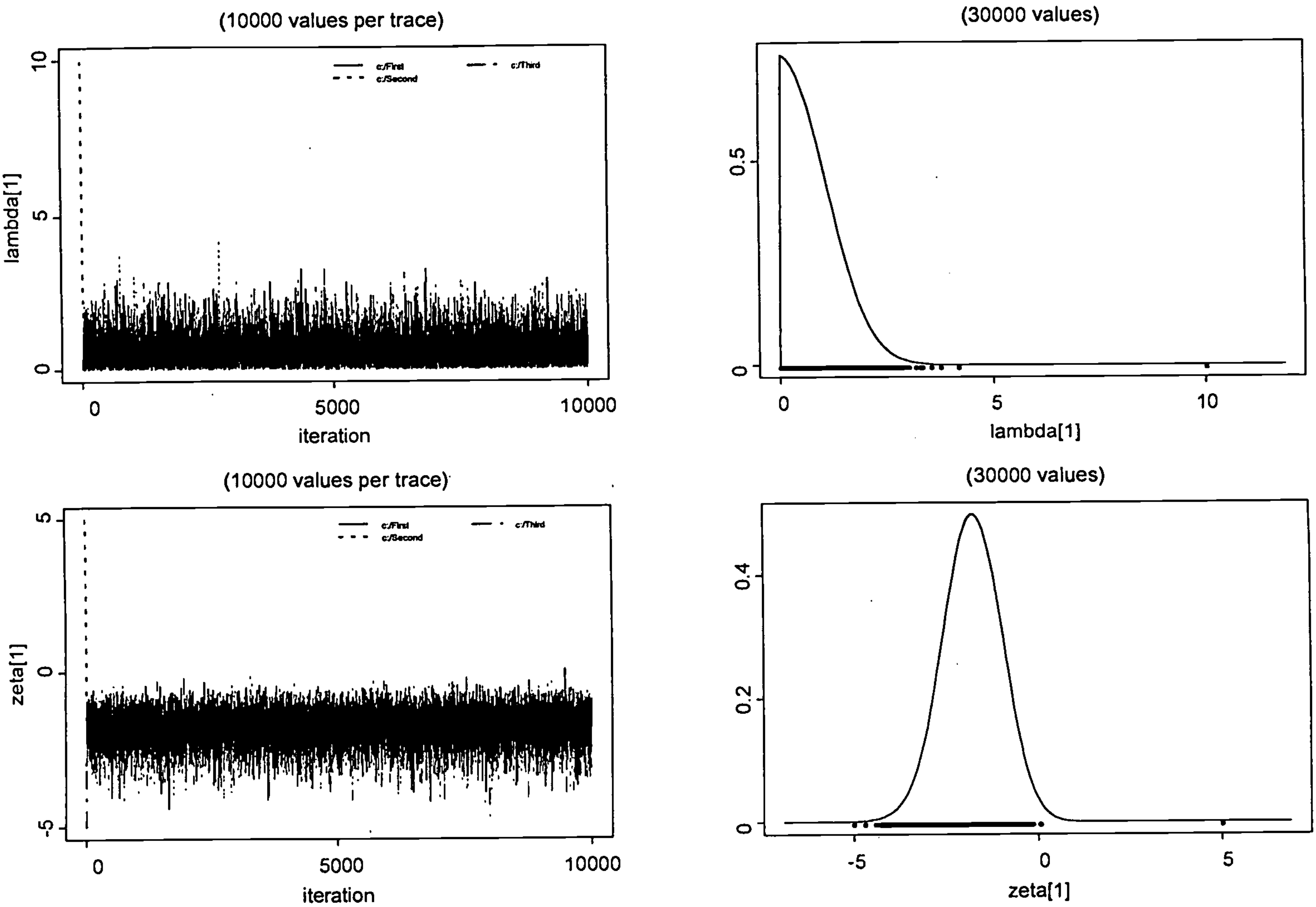
Average Correlations $r_{\theta\hat{\theta}}$ Between Ability Parameters and Estimates over 100 Replications

Examinee	Gibbs Sampling $n = 10$	Gibbs Sampling $n = 20$	Gibbs Sampling $n = 40$	Marginal Bayesian $n = 10$	Marginal Bayesian $n = 20$	Marginal Bayesian $n = 40$
50	.796	.875	.932	.802	.879	.933
100	.798	.880	.932	.802	.882	.933
200	.801	.880	.934	.803	.881	.935

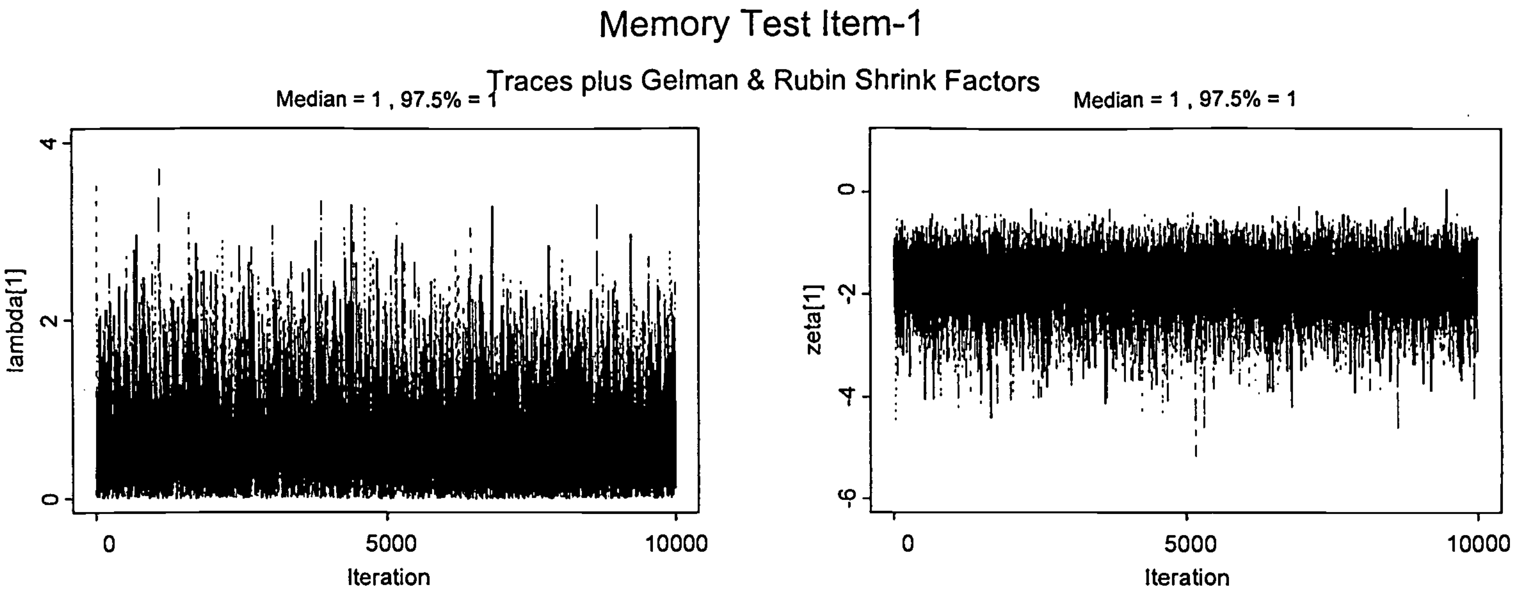
Figure Captions


λ_j
ζ_j
θ_i
P_{ij}
Y_{ij}
Examinee i
Item j

Convergence with Starting Values for Memory Test Item-1



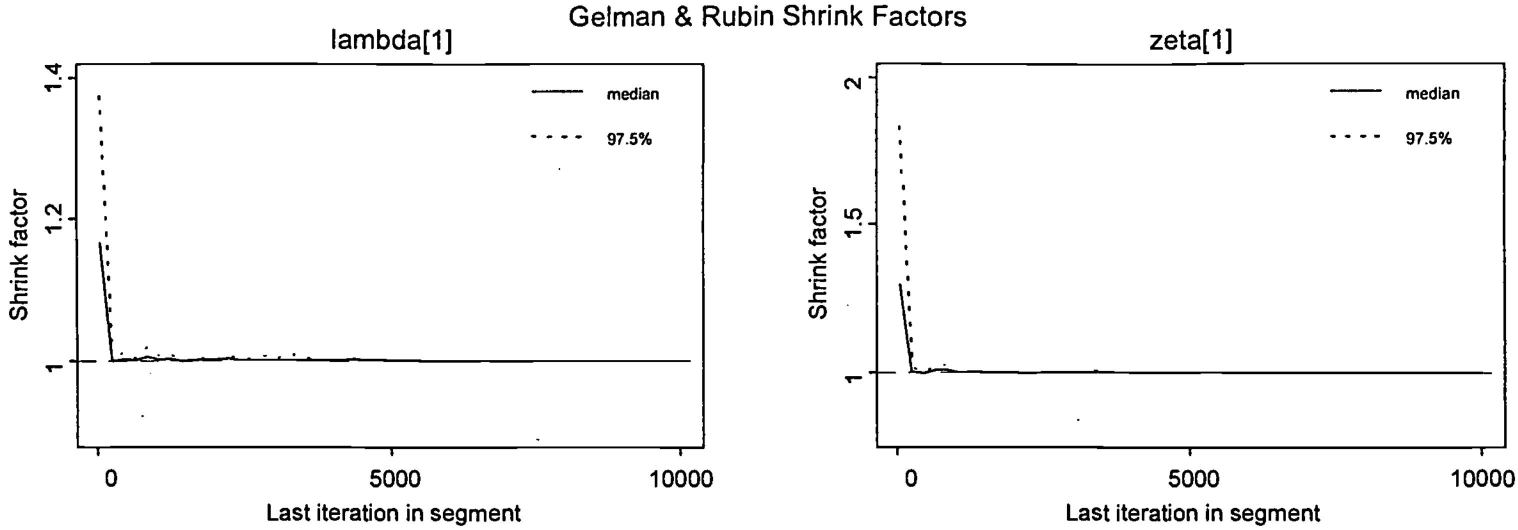
BEST COPY AVAILABLE



BEST COPY AVAILABLE

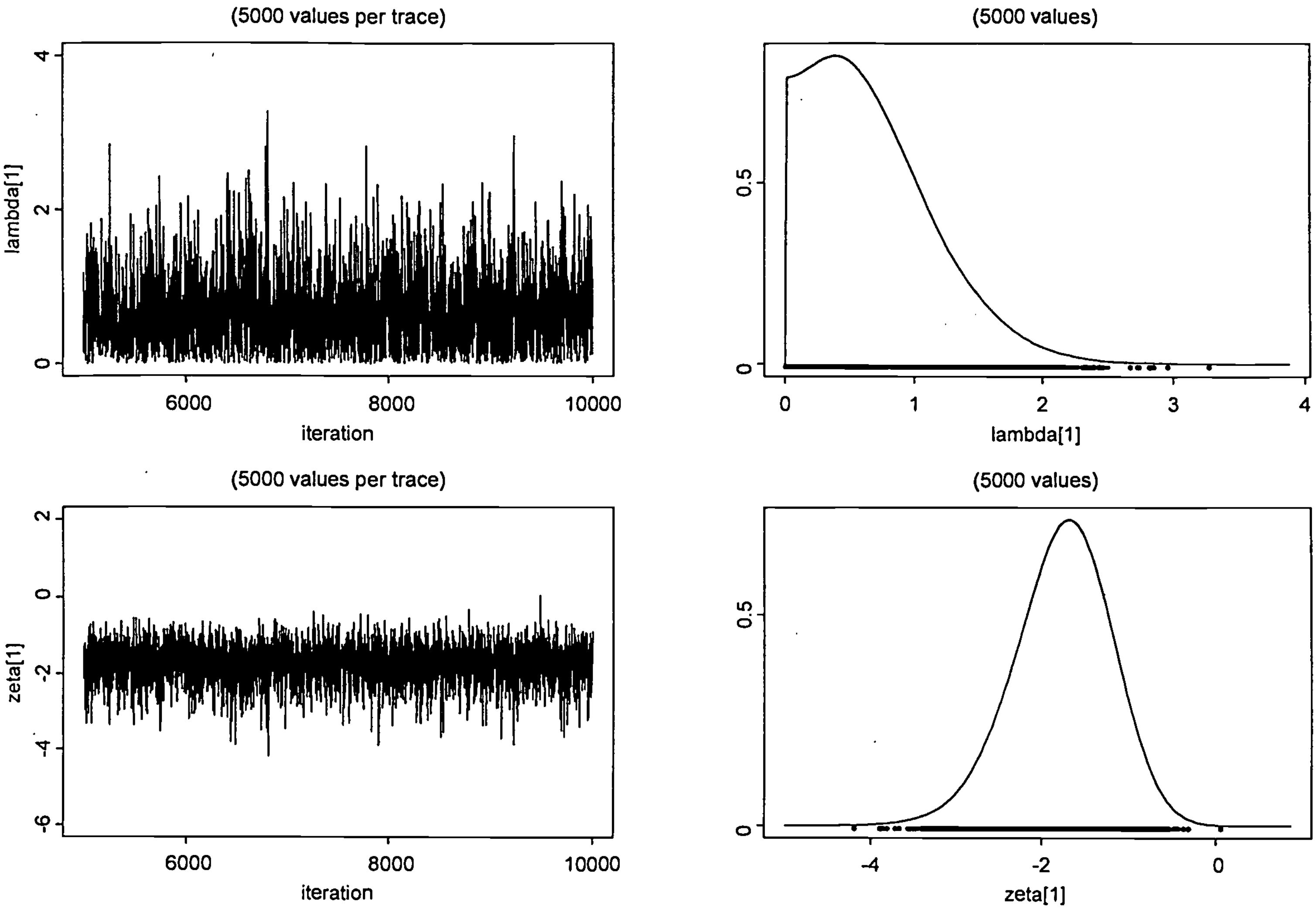
Memory Test Item-1

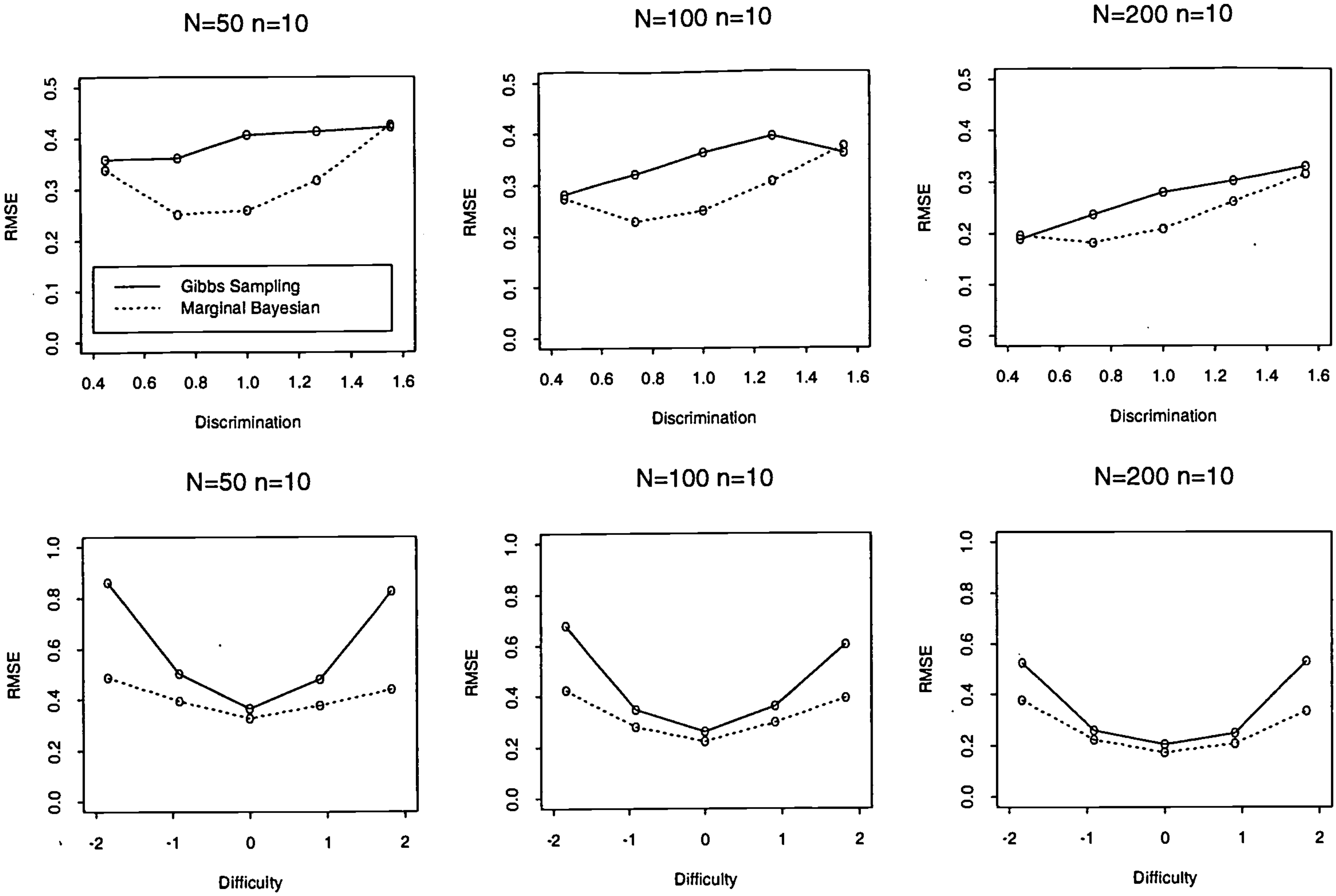
Gelman & Rubin Shrink Factors

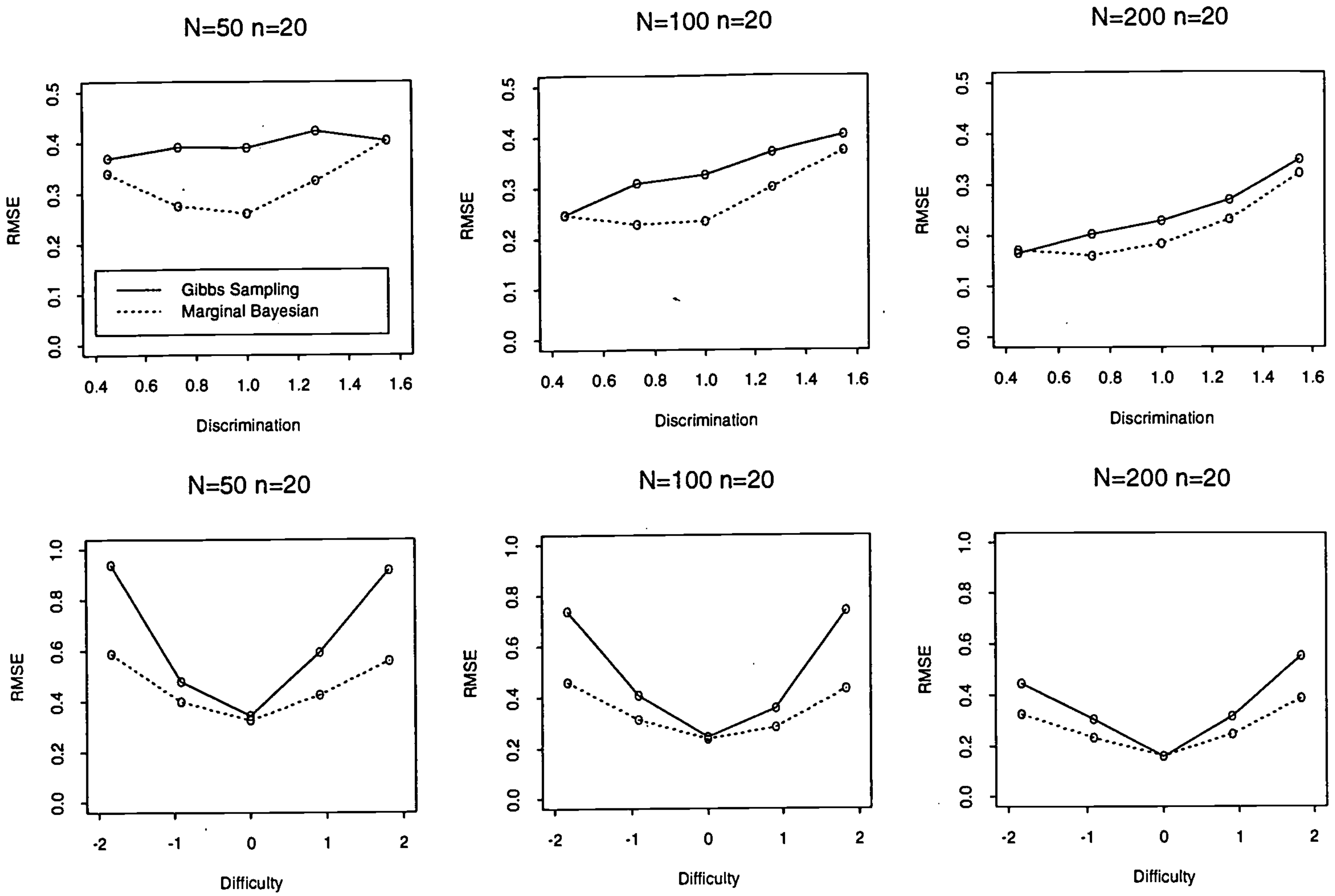


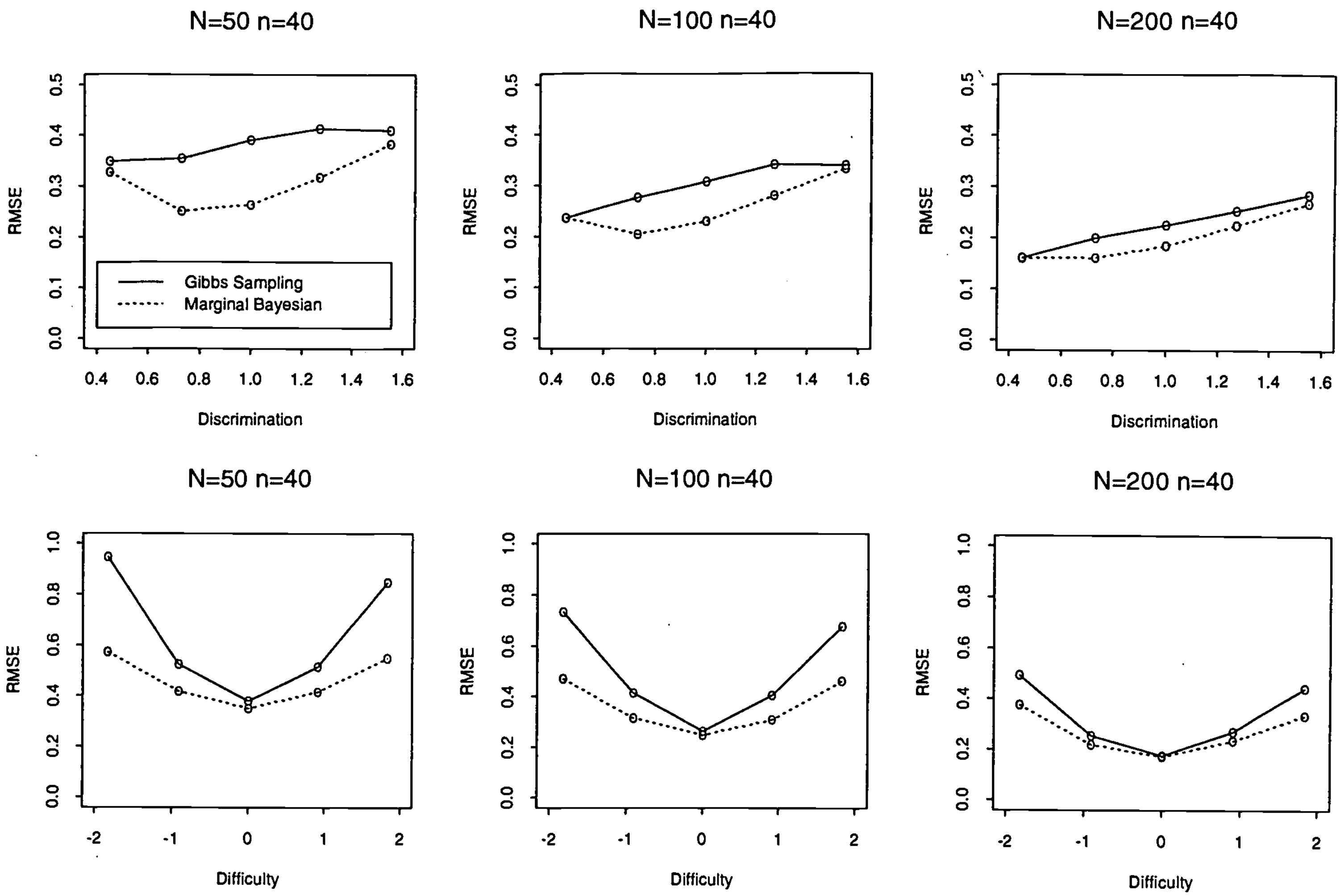
BEST COPY AVAILABLE

Memory Test Item-1

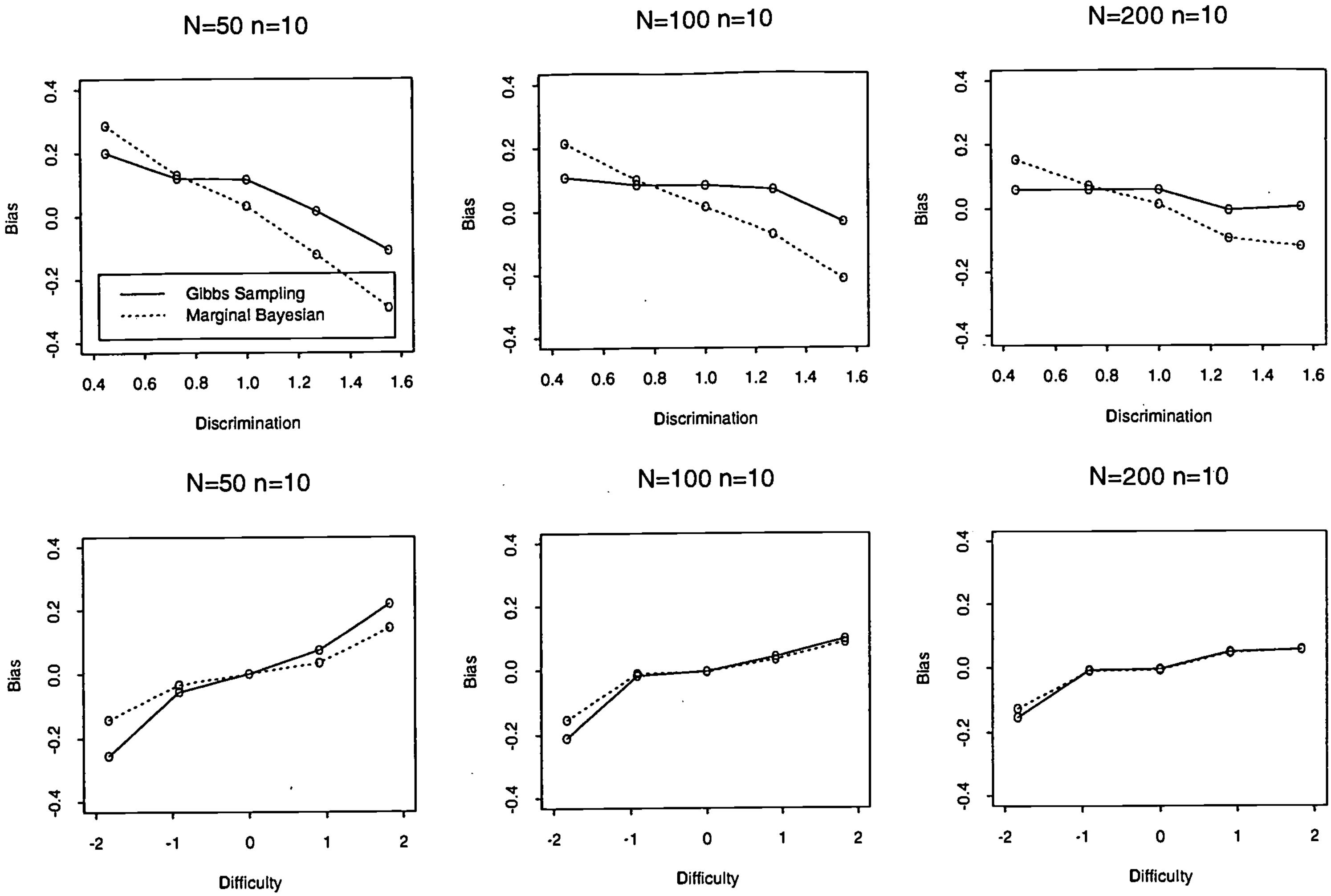



N=50 n=10
N=100 n=10
N=200 n=10
RMSE
Gibbs Sampling
Marginal Bayesian
0.0 0.1 0.2 0.3 0.4 0.5
0.4 0.6 0.8 1.0 1.2 1.4 1.6
Discrimination
N=50 n=10
N=100 n=10
N=200 n=10
RMSE
0.0 0.2 0.4 0.6 0.8 1.0
-2 -1 0 1 2
Difficulty

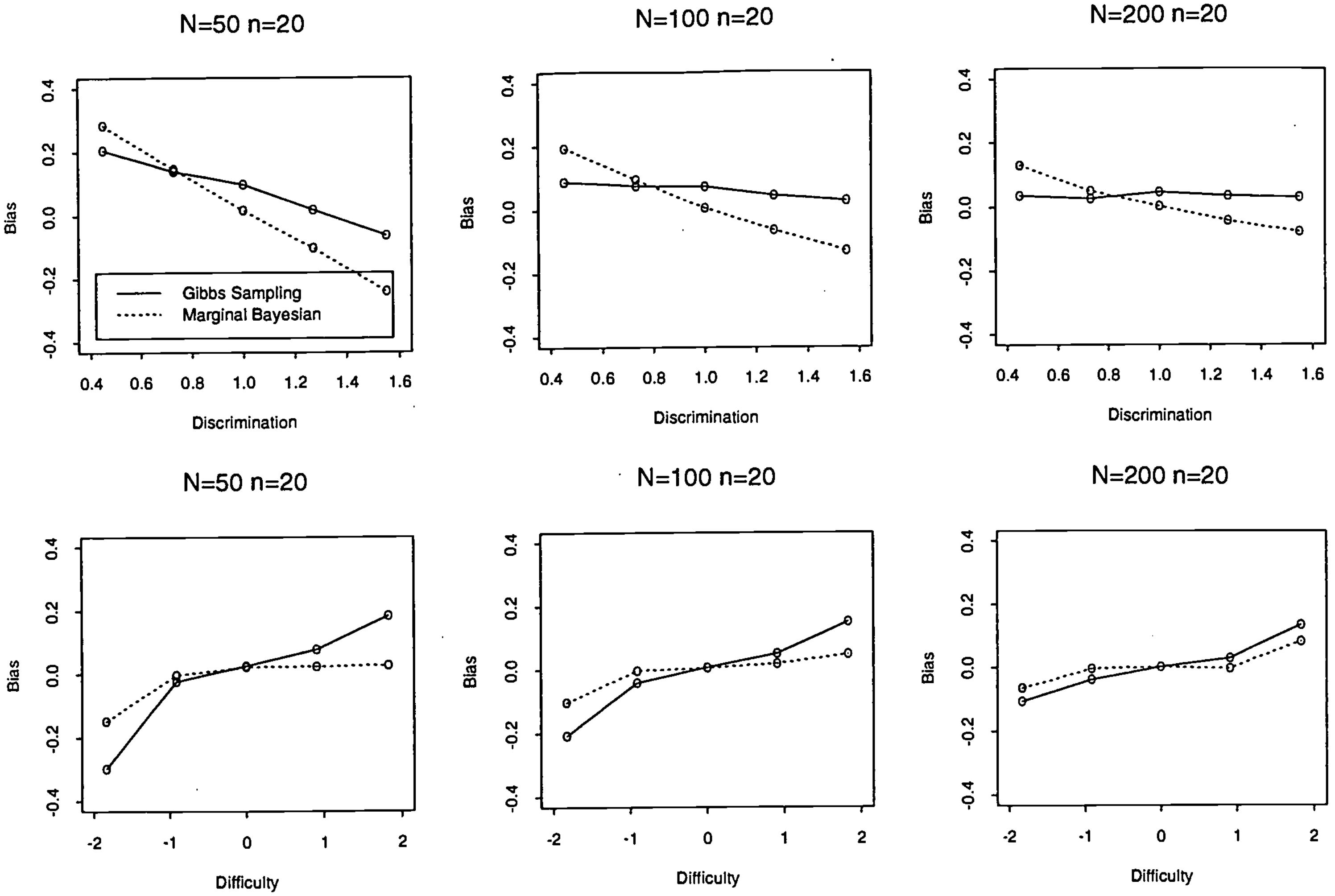

N=50 n=20
N=100 n=20
N=200 n=20
RMSE
Discrimination
Gibbs Sampling
Marginal Bayesian
Difficulty

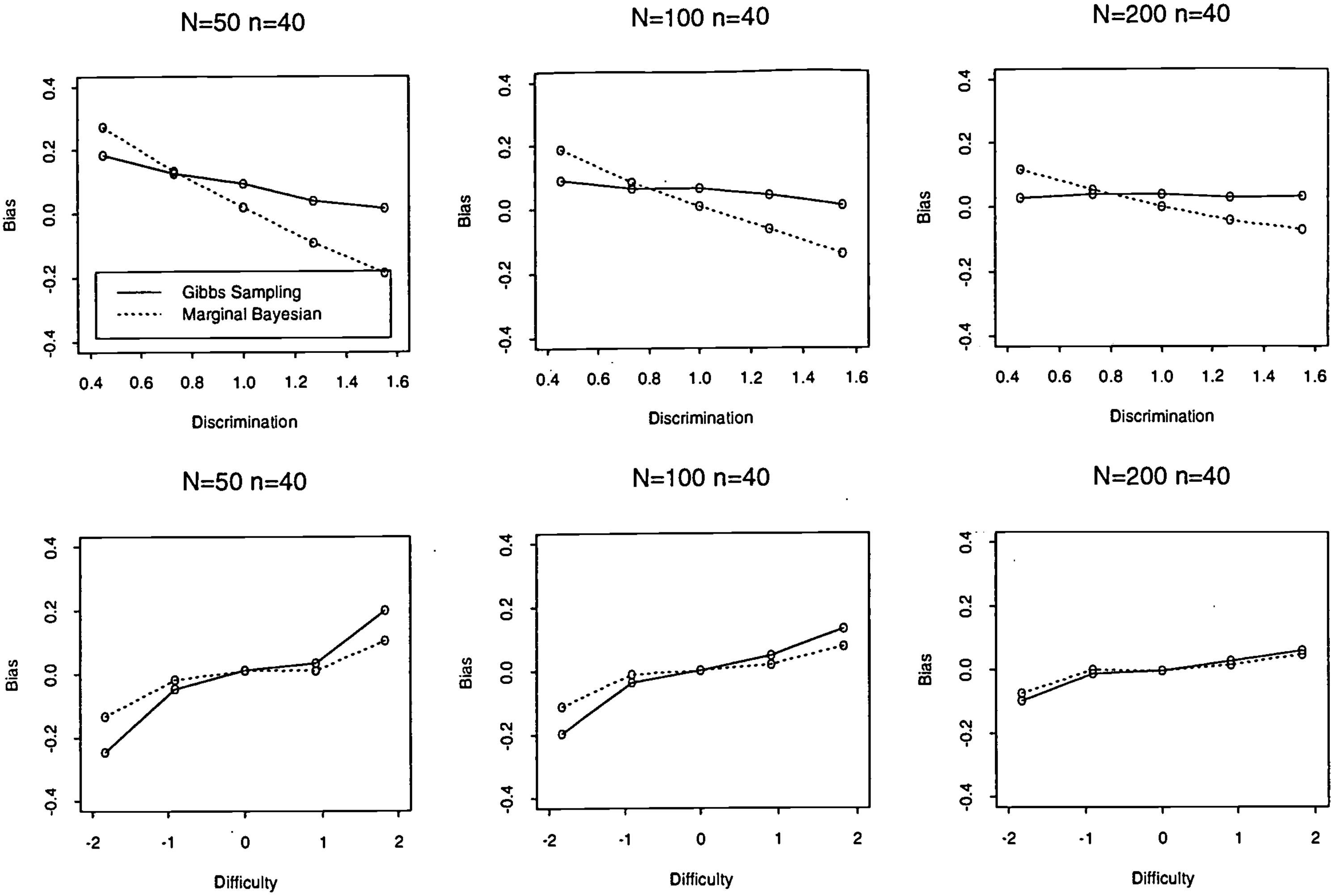


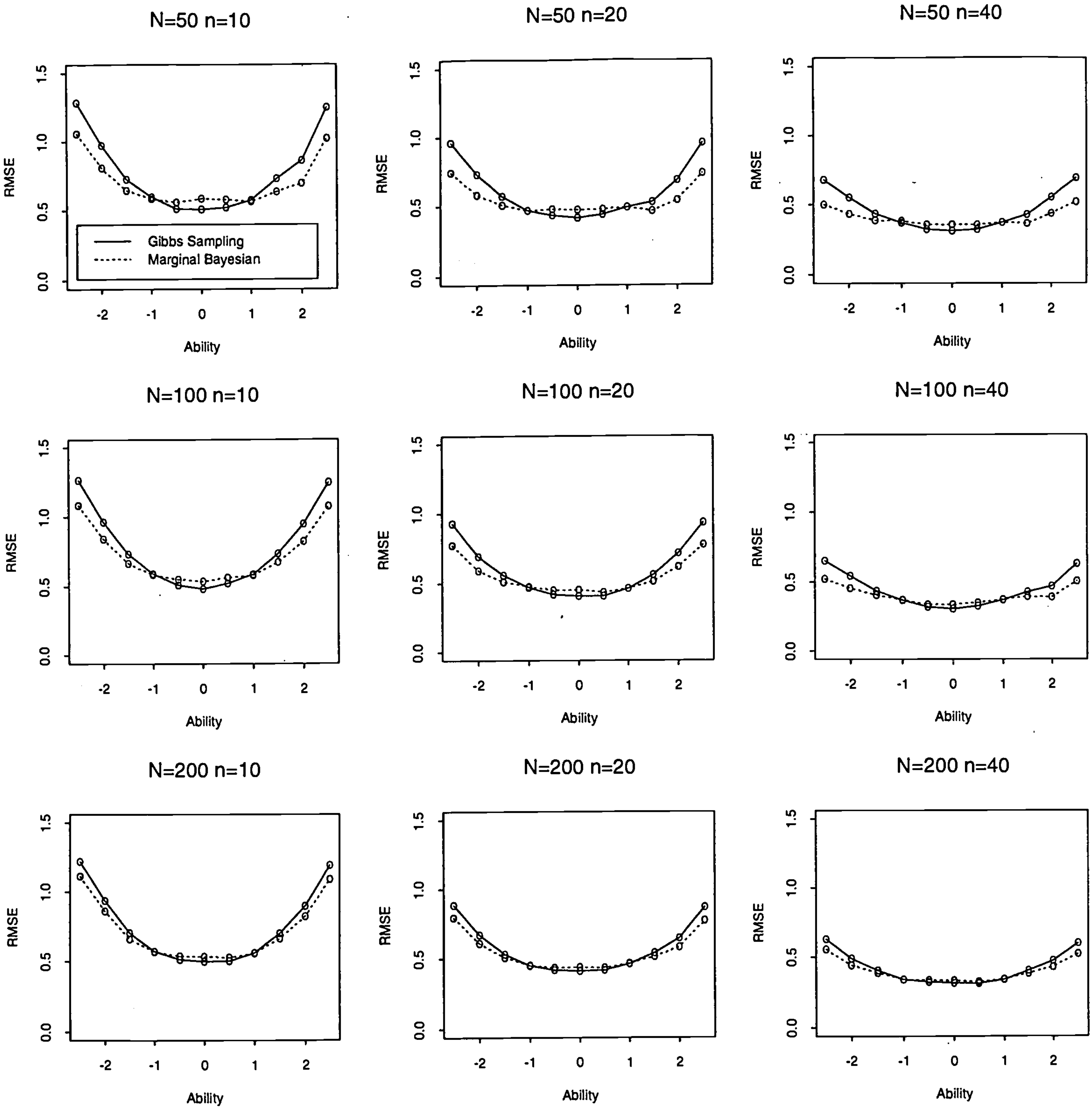
BEST COPY AVAILABLE

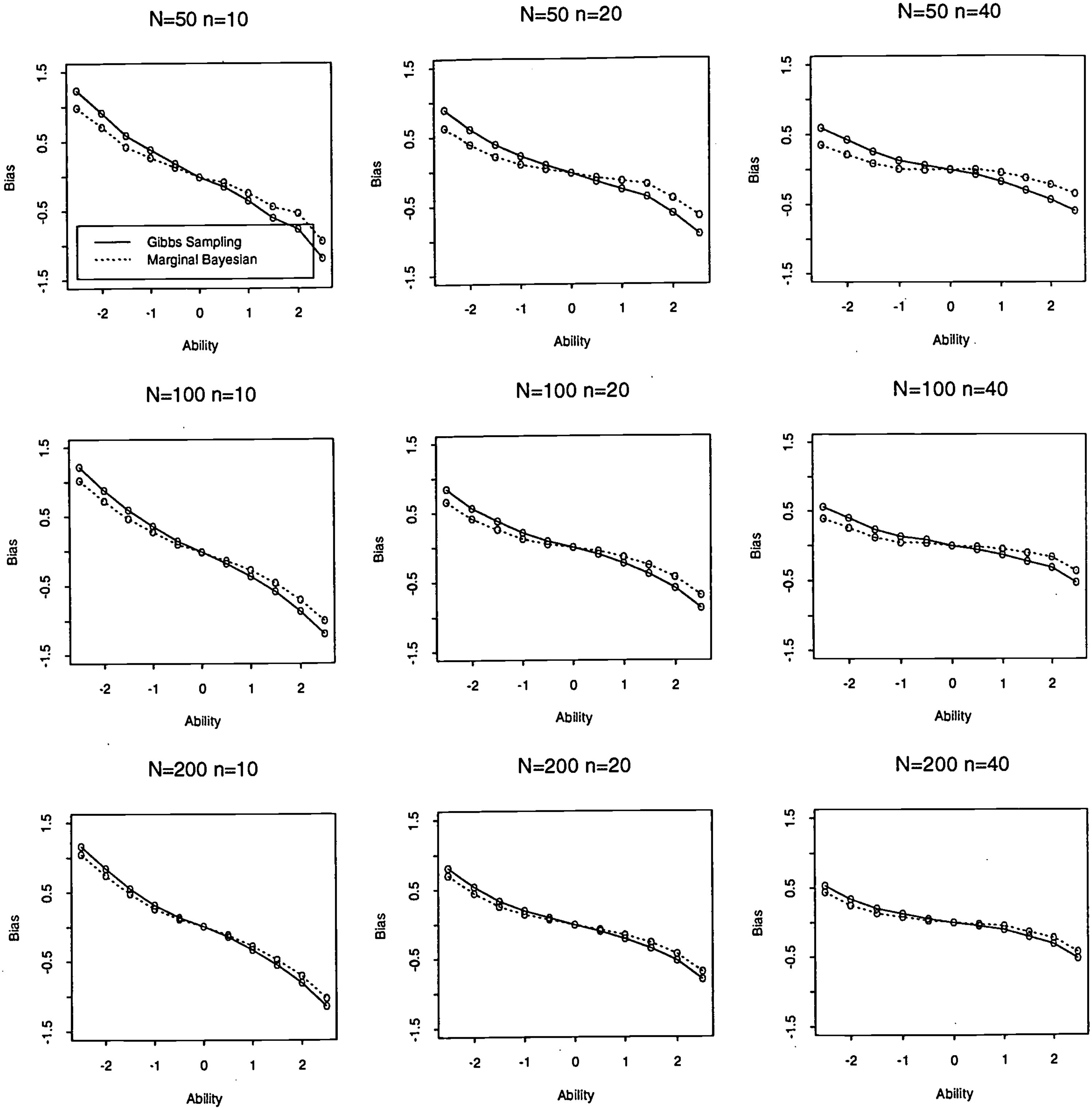


BEST COPY AVAILABLE


N=50 n=20
N=100 n=20
N=200 n=20
Bias
Discrimination
Gibbs Sampling
Marginal Bayesian
0.4
0.2
0.0
-0.2
-0.4
0.4 0.6 0.8 1.0 1.2 1.4 1.6
N=50 n=20
N=100 n=20
N=200 n=20
Bias
Difficulty
-2 -1 0 1 2


N=50 n=40
N=100 n=40
N=200 n=40
Bias
Discrimination
0.4 0.6 0.8 1.0 1.2 1.4 1.6
-0.4 -0.2 0.0 0.2 0.4
Gibbs Sampling
Marginal Bayesian
N=50 n=40
N=100 n=40
N=200 n=40
Bias
Difficulty
-2 -1 0 1 2


N=50 n=10
N=50 n=20
N=50 n=40
N=100 n=10
N=100 n=20
N=100 n=40
N=200 n=10
N=200 n=20
N=200 n=40
Gibbs Sampling
Marginal Bayesian
RMSE
0.0
0.5
1.0
1.5
Ability
-2
-1
0
1
2


N=50 n=10
N=50 n=20
N=50 n=40
N=100 n=10
N=100 n=20
N=100 n=40
N=200 n=10
N=200 n=20
N=200 n=40
Bias
Ability
1.5
0.5
-0.5
-1.5
-2
-1
0
1
2
Gibbs Sampling
Marginal Bayesian

Appendix

```
model memory;
const
   I = 40,
   J = 10;
var
   y[I,J], p[I,J], theta[I], lambda[J], zeta[J], b[J];
data in "memory.dat";
inits in "memory.in";
{
  for (i in 1:I) {
     for (j in 1:J) {
        logit(p[i,j]) <- lambda[j]*theta[i] + zeta[j];
        y[i,j] ~ dbern(p[i,j]);
     }
     theta[i] ~ dnorm(0,1);
  }
  for (j in 1:J) {
     lambda[j] ~ dnorm(0,1) I(0,);
     zeta[j] ~ dnorm(0,0.0001);
     b[j] <- - zeta[j]/lambda[j]
  }
}
```

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Accuracy of Parameter Estimation in Gibbs Sampling Under the Two-Parameter Logistic Model

Author(s): Seock-Ho Kim and Allan S. Cohen

Corporate Source: The University of Georgia and University of Wisconsin - Madison

Publication Date: 1999

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: 325 Aderhold Hall, Athens, GA 30602-7143
Telephone: 706-542-4224
FAX: 706-542-4240
E-Mail Address: skim@coe.uga.edu
Date: 4/9/99

(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