DOCUMENT RESUME

ED 440 995 TM 030 793

AUTHOR	Kim, Seock-Ho; Cohen, Allan S.
TITLE	An Investigation of Ability Estimation in Gibbs Sampling.
PUB DATE	2000-04-25
NOTE	37p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Ability; *Bayesian Statistics; *Estimation (Mathematics); Intelligence Tests; Item Response Theory
IDENTIFIERS	BILOG Computer Program; *Gibbs Sampling

ABSTRACT

The ability estimates of Gibbs sampling and the magnitudes of the posterior standard deviations were investigated. Item parameters of the Q-E intelligence test (J. Fraenkel and N. Wallen, 2000) for 44 examinees were obtained using Gibbs sampling, marginal Bayesian estimation, and BILOG. Two normal priors were used in item parameter estimation. Ability estimates were obtained using Gibbs sampling, that is, jointly with item parameter estimates, and compared with estimates from the expected a posteriori method using item parameter estimates obtained from Gibbs sampling, marginal Bayesian estimation, and BILOG. Item parameter estimates were very similar, as were ability estimates, but the patterns of the magnitudes of the posterior standard deviations of ability estimates from Gibbs sampling were different from those based on the expected a posteriori method. (Contains 7 tables, 4 figures, and 45 references.) (Author/SLD)

An Investigation of Ability Estimation in Gibbs Sampling

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

April 25, 2000
Running Head: ABILITY ESTIMATION IN GIBBS SAMPLING

Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

An Investigation of Ability Estimation in Gibbs Sampling

Abstract

The ability estimates of Gibbs sampling and the magnitudes of the posterior standard deviations were investigated. Item parameters of the Q-E intelligence test were obtained using Gibbs sampling, marginal Bayesian estimation, and BILOG. Two normal priors were used in item parameter estimation. Ability estimates were obtained using Gibbs sampling, that is, jointly with item parameter estimates, and compared with estimates from the expected a posteriori method employing item parameter estimates obtained from Gibbs sampling, marginal Bayesian estimation, and BILOG. Item parameter estimates were very similar as were ability estimates but the patterns of the magnitudes of the posterior standard deviations of ability estimates from Gibbs sampling were different from those based on the expected a posteriori method.

Keywords: Bayesian inference, expected a posteriori, Gibbs sampling, item response theory, marginal Bayesian.

Introduction

Gibbs sampling is a member of a class of Markov Chain Monte Carlo (MCMC) methods that can be used for estimation of item and ability parameters in various item response theory (IRT) models. The advantages of Gibbs sampling include incorporation of standard errors of item parameter estimates into inferences of ability and an estimation process that remains straightforward even as model complexity increases (Patz & Junker, 1999). Although previous studies suggested Gibbs sampling could be a viable method in IRT, comparisons with existing estimation methods have only just begun to be presented, and characteristics of item and ability parameter estimates in Gibbs sampling are not well known. The purpose of this paper is to investigate ability estimation in Gibbs sampling under the two-parameter normal ogive model. Results are compared to those obtained using the expected a posteriori method.

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 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, 1999). The MCMC methods can be used to approximate the marginal density, and hence can be used in estimation of IRT parameters.

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. 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 to estimation of item parameters for the two-parameter normal ogive model and compared the estimates with those obtained using marginal maximum likelihood estimation. Baker (1998) investigated item parameter recovery characteristics of Albert's Gibbs sampling procedure for item parameter estimation via a simulation study. Patz and Junker (1999) developed a MCMC method based on the Metropolis-Hasting algorithm and presented an illustration using the two-parameter logistic model. Patz and Junker (1999) classified MCMC methods into several categories. In particular, Gibbs sampling of Albert (1992) was classified as Gibbs sampling with data augmentation (see Tanner & Wong, 1987; Tanner, 1996). Johnson and Albert (1999) presented a full Bayesian method of Gibbs sampling that can be seen as a modification of Albert (1992). This paper considers Gibbs sampling of Johnson and Albert (1999). Kim and Cohen (1999) investigated accuracy of item and ability parameter estimation in Gibbs sampling under the two-parameter logistic model.

MCMC computer programs in IRT have been developed largely only for specific applications. For example, Albert (1992) and Johnson and Albert (1999) used computer programs 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 which employs an adaptive rejection sampling algorithm (Gilks & Wild, 1992). Kim and Cohen (1999) used BUGS. In this paper, Baker's (1998) program was extended to implement the algorithm of Gibbs sampling in Johnson and Albert (1999).

The marginal maximum likelihood and marginal maximum a posteriori methods using the expectation and maximization algorithm, as implemented in the computer program BILOG (Mislevy & Bock, 1990), have become the standard estimation technique for obtaining IRT item parameter estimates (see Bock & Aitkin, 1981; Bock & Lieberman, 1970; Mislevy, 1986). Ability parameters are estimated in the marginalized solutions using either maximum likelihood, expected a posteriori, or maximum a posteriori estimation after obtaining the item parameter estimates and assuming the estimates are true values. Patz and Junker

(1999) indicated that these were based on the "divide and conquer" strategy because of the sequential nature of the marginal estimation procedures.

In Gibbs sampling ability parameters can be estimated either jointly with item parameters or sequentially after obtaining the item parameters. As Patz and Junker (1999) have indicated, one of the advantages of the MCMC method is incorporation of standard errors of item parameter estimates into inferences of ability. An issue to be examined in this study is the incorporation of instability of item parameter in small samples on estimation of ability parameters. All estimation methods should yield comparable item and ability parameter estimates, when comparable priors are used with large samples. This study was designed to investigate the comparability of item and ability parameter estimates for a small sample of data, when different strategies are used for ability estimation using the two-parameter normal ogive model. Specifically, Gibbs sampling of Johnson and Albert (1999) was examined and compared with other estimation methods. The differences between Gibbs sampling and other estimation methods are also discussed in the context of logic and underlying mathematics.

Background

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 $P(Y_{ij} = 1|\theta_i, \boldsymbol{\xi}_j) = P_{ij}$ and the probability of an incorrect response is given by $P(Y_{ij} = 0|\theta_i, \boldsymbol{\xi}_j) = 1 - P_{ij} = Q_{ij}$, where θ_i is ability and $\boldsymbol{\xi}_j$ is the vector of item parameters.

Birnbaum (1968) and Lord (1980) describe the estimation of the $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ by joint maximization of the likelihood function

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

where $\mathbf{Y}$ is an $N \times n$ matrix of observed responses, y_{ij}, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)' = \{\theta_i\}$, and $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n)'$. In joint maximum likelihood estimation (see Lord, 1986 for a comparison of marginalized and joint estimation methods), the item parameter estimation part for maximizing $l(\boldsymbol{\xi}|\mathbf{Y}, \hat{\boldsymbol{\theta}})$ and the ability parameter estimation part for maximizing $l(\boldsymbol{\theta}|\mathbf{Y}, \hat{\boldsymbol{\xi}})$ 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 $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ can be estimated by joint maximization with respect to the parameters of the posterior density

$$p(\boldsymbol{\theta}, \boldsymbol{\xi}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\boldsymbol{\theta}, \boldsymbol{\xi})p(\boldsymbol{\theta}, \boldsymbol{\xi})}{p(\mathbf{Y})} \propto l(\boldsymbol{\theta}, \boldsymbol{\xi}|\mathbf{Y})p(\boldsymbol{\theta}, \boldsymbol{\xi}), \tag{2}$$

where $p(\boldsymbol{\theta}, \boldsymbol{\xi})$ is the prior joint density of the item and ability parameters. This procedure is joint Bayesian estimation. Under the assumption that priors of $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$ are independently distributed with probability density functions $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\xi})$, the item parameter estimation part maximizing $l(\boldsymbol{\xi}|\mathbf{Y}, \hat{\boldsymbol{\theta}})p(\boldsymbol{\xi})$, and the ability parameter estimation part maximizing $l(\boldsymbol{\theta}|\mathbf{Y}, \hat{\boldsymbol{\xi}})p(\boldsymbol{\theta})$ are iterated to obtain stable Bayes modal estimates of item and ability parameters.

The main feature of MCMC methods is that of obtaining 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 usual IRT estimation procedures, however, the task is to obtain modes of either a likelihood function or a posterior distribution.

The Gibbs sampling algorithm proceeds as follows (Gelfand & Smith, 1990; Tanner, 1996). First, instead of using $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, let $\boldsymbol{\omega}$ be a vector of parameters with k elements. Suppose that the full or complete conditional distributions, $p(\omega_i|\omega_j, \mathbf{Y})$, where $i, j = 1(1)k$ 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:

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

The vectors $\boldsymbol{\omega}^{(0)}, \ldots, \boldsymbol{\omega}^{(t)}, \ldots$ are a realization of a Markov chain with a transition probability from $\boldsymbol{\omega}^{(t)}$ to $\boldsymbol{\omega}^{(t+1)}$ given by

$$p(\boldsymbol{\omega}^{(t)}, \boldsymbol{\omega}^{(t+1)}) = \prod_{l=1}^{k} p(\boldsymbol{\omega}_l^{(t+1)} | \boldsymbol{\omega}_j^{(t)}, j > l, \boldsymbol{\omega}_j^{(t+1)}, j < l, \mathbf{Y}). \tag{3}$$

The joint distribution of $\boldsymbol{\omega}^{(t)}$ converges geometrically to the posterior distribution $p(\boldsymbol{\omega}|\mathbf{Y})$ as $t \to \infty$ (Geman & Geman, 1984, Bernardo & Smith, 1994). In particular, $\boldsymbol{\omega}_i^{(t)}$ tends to be distributed as a random quantity whose density is $p(\boldsymbol{\omega}_i|\mathbf{Y})$. Now suppose that there exist m replications of the t iterations. For large t, the replicates $\boldsymbol{\omega}_{i1}^{(t)}, \ldots, \boldsymbol{\omega}_{im}^{(t)}$ are approximately a random sample from $p(\boldsymbol{\omega}_i|\mathbf{Y})$. If we make m reasonably large, then an estimate, $\hat{p}(\boldsymbol{\omega}_i|\mathbf{Y})$, can be obtained either as a kernel density estimate derived from the replicates or as

$$\hat{p}(\boldsymbol{\omega}_i|\mathbf{Y}) = \frac{1}{m} \sum_{l=1}^{m} p(\boldsymbol{\omega}_i | \boldsymbol{\omega}_{jl}^{(t)}, j \neq i, \mathbf{Y}). \tag{4}$$

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

Gibbs Sampling with Data Augmentation

Albert (1992) and Johnson and Albert (1999) presented programs for the Gibbs sampling procedure using the matrix language MATLAB (e.g., The MathWorks, Inc., 1996). The two-parameter normal ogive model was used for the item response function. The probit, without the addition of 5 (see Finney, 1971; i.e., the normal equivalent deviate), was given by

$$Z_{ij} = a_j \theta_i - b_j, \tag{5}$$

where a_j is the item discrimination parameter, and b_j is the negative intercept parameter for item j, that is $\boldsymbol{\xi}_j = (a_j, b_j)'$. Using the usual item response theory parameterization, we have

$$Z_{ij} = a_j \theta_i - b_j = \lambda_j \theta_i + \zeta_j = \alpha_j (\theta_i - \beta_j), \tag{6}$$

where $a_j = \lambda_j = \alpha_j$ is the slope parameter, ζ_j is the intercept parameter, and β_j is the item difficulty or threshold parameter (see Baker, 1992, p. 7, p. 26). Note that under the two-parameter logistic item response theory model,

$$Z_{ij} = \alpha_j D(\theta_i - \beta_j), \tag{7}$$

where $D = 1.7$ is the scaling factor (e.g., Mislevy & Bock, 1990).

Let $\mathbf{Z} = \{Z_{ij}\}$ be the matrix of independent random probits, that represents the matrix of the augmented data (Patz & Junker, 1999). A Gibbs sampling procedure can be used to sample from the joint posterior distribution over the entire collection of unknown parameters and latent data. The sampler is based on iteratively drawing values from three sets of conditional probability distributions, $p(\mathbf{Z}|\boldsymbol{\theta}, \boldsymbol{\xi}, \mathbf{Y})$, $p(\boldsymbol{\theta}|\mathbf{Z}, \boldsymbol{\xi}, \mathbf{Y})$, and $p(\boldsymbol{\xi}|\mathbf{Z}, \boldsymbol{\theta}, \mathbf{Y})$ (Johnson & Albert, 1999).

To implement the Gibbs sampler, suppose at iteration $(t-1)$ the current values of the model parameters are denoted by $\left\{Z_{ij}^{(t-1)}\right\}$, $\left\{\theta_i^{(t-1)}\right\}$, $\left\{a_j^{(t-1)}\right\}$, and $\left\{b_j^{(t-1)}\right\}$. Then one complete cycle of the Gibbs sampler can be described as follows:

First, values of the latent data $\left\{Z_{ij}^{(t)}\right\}$ are simulated conditional on the current values of the latent traits and item parameters and on the item response data. In other words, Z_{ij} is randomly sampled from $p(\mathbf{Z}^{(t)}|\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\xi}^{(t-1)}, \mathbf{Y})$. The conditional posterior distribution of Z_{ij} is a truncated normal distribution with mean

$$m_{ij} = a_j^{(t-1)}\theta_i^{(t-1)} - b_j^{(t-1)} \tag{8}$$

and variance 1. The truncation of the posterior distribution depends on the value of the corresponding observation y_{ij}. If $y_{ij} = 1$, the truncation of Z_{ij} is from the left at 0 and Z_{ij} is sampled from the part of the conditional posterior distribution above 0. If $y_{ij} = 0$, the truncation of Z_{ij} is from the right at 0 and Z_{ij} is sampled from the part below 0. Let the new latent data value simulated from this truncated normal distribution be denoted by $\left\{Z_{ij}^{(t)}\right\}$.

Second, latent traits $\left\{\theta_i^{(t)}\right\}$ are simulated from their posterior distribution conditional on current values of the latent data and item parameters, where the posterior can be denoted by $p(\boldsymbol{\theta}^{(t)}|\mathbf{Z}^{(t)}, \boldsymbol{\xi}^{(t-1)}, \mathbf{Y})$. Using the latent data representation, the item response model can be written as

$$Z_{ij}^{(t)} + b_j^{(t-1)} = a_j^{(t-1)}\theta_i + \epsilon_{ij}, \tag{9}$$

where the error term ϵ_{ij} are independent normal with mean 0 and variance 1. For a given value of i, this is a special case of the linear regression model with unknown parameter θ_i.

The likelihood function for θ_i is of the normal form with mean

$$\bar{\theta}_i = \frac{\sum_{j=1}^{J} a_j^{(t-1)} \left(Z_{ij}^{(t)} + b_j^{(t-1)} \right)}{\sum_{j=1}^{J} \left(a_j^{(t-1)} \right)^2} \tag{10}$$

and variance

$$\sigma_{\theta_i}^2 = \frac{1}{\sum_{j=1}^{J} \left(a_j^{(t-1)} \right)^2}. \tag{11}$$

Combining the sampling model with the $N(\mu_\theta, \tau_\theta^2)$ prior, it follows that the conditional posterior density of θ_i is normally distributed with mean

$$m_{\theta_i} = \frac{\bar{\theta}_i / \sigma_{\theta_i}^2 + \mu_\theta / \tau_\theta^2}{1/\sigma_{\theta_i}^2 + 1/\tau_\theta^2} \tag{12}$$

and variance

$$v_{\theta_i} = \frac{1}{1/\sigma_{\theta_i}^2 + 1/\tau_\theta^2} \tag{13}$$

Let $\left\{ \theta_i^{(t)} \right\}$ denote the vector of latent traits randomly drawn from the conditional posterior density. Specifically,

$$\theta_i^{(t)} = R\sqrt{v_{\theta_i}} + m_{\theta_i}, \tag{14}$$

where R is a random Gaussian deviate (Press, Teukolsky, Vetterling, & Flannery, 1992, p. 280).

Third, the item parameters $\{a_j, b_j\}$ are simulated from their joint posterior density, $p(\boldsymbol{\xi}^{(t)} | \mathbf{Z}^{(t)}, \boldsymbol{\theta}^{(t)}, \mathbf{Y})$, conditionally on the current values of the latent data and the latent traits. To determine the conditional distribution, the latent data model can be written as

$$Z_{ij}^{(t)} = a_j \theta_i^{(t)} - b_j + \epsilon_{ij}. \tag{15}$$

Since the values of the latent data $\left\{ Z_{ij}^{(t)} \right\}$ and the latent traits $\left\{ \theta_i^{(t)} \right\}$ are fixed, the model can be seen as a linear regression model with unknown parameters a_j and b_j for a fixed value of j. Using matrix notations, the model can be written as

$$\mathbf{z}_j = \mathbf{X} \boldsymbol{\xi}_j + \boldsymbol{\epsilon}_j, \tag{16}$$

where $\mathbf{z}_j = \begin{bmatrix} Z_{1j}^{(t)} \\ Z_{2j}^{(t)} \\ \vdots \\ Z_{Ij}^{(t)} \end{bmatrix}$, the design matrix $\mathbf{X} = \begin{bmatrix} \theta_1^{(t)} & -1 \\ \theta_2^{(t)} & -1 \\ \vdots & \vdots \\ \theta_I^{(t)} & -1 \end{bmatrix}$, $\boldsymbol{\xi}_j = \begin{bmatrix} a_j \\ b_j \end{bmatrix}$, and $\boldsymbol{\epsilon}_j = \begin{bmatrix} \epsilon_{1j}^{(t)} \\ \epsilon_{2j}^{(t)} \\ \vdots \\ \epsilon_{Ij}^{(t)} \end{bmatrix}$.

Let $\boldsymbol{\mu}_\xi$ denote the prior mean vector

$$\boldsymbol{\mu}_\xi = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix} \tag{17}$$

and let $\boldsymbol{\Sigma}_\xi$ denote the prior covariance matrix

$$\boldsymbol{\Sigma}_\xi = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix}. \tag{18}$$

It follows that the conditional posterior density of $\boldsymbol{\xi}_j$ is multivariate normal with mean vector

$$\mathbf{m}_j = \left(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_\xi^{-1}\right)^{-1} \left(\mathbf{X}'\mathbf{z}_j + \boldsymbol{\Sigma}_\xi^{-1}\boldsymbol{\mu}_\xi\right) \tag{19}$$

and covariance matrix

$$\mathbf{V}_j = \left(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_\xi^{-1}\right)^{-1}. \tag{20}$$

The Cholesky decomposition of $\mathbf{V}_j$ is performed and let

$$\mathbf{A} = \text{chol}(\mathbf{V}_j) = \text{chol}(\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_\xi^{-1})^{-1}. \tag{21}$$

Two random normal deviates R_1 and R_2 are generated and the item parameter values for item j are given by

$$\boldsymbol{\xi}_j^{(t)} = \mathbf{A}' \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} + \mathbf{m}_j. \tag{22}$$

According to Johnson and Albert (1999), given suitable starting values for the parameter values, these steps define one cycle in Gibbs sampling scheme that can be used to obtain samples from the posterior distribution over all model parameters. Convergence of the algorithm is typically obtained within several hundred observations and is usually not very sensitive to the choice of starting values.

Analyses of the Q-E Intelligence Test Data

Data

The data for this example are from the administration of the ten-item Q-E intelligence test (Fraenkel & Wallen, 2000, p. 181) to 44 examinees. Although there are 1,024 possible response patterns for this ten-item test, only 37 patterns occurred in this sample. Examinees 8 and 9, 10 and 11, 33 and 34, 40 to 44, respectively, had the same response patterns. The data are given in Table 1.

Insert Table 1 about here

Gibbs Sampling

Gibbs sampling typically employs the following four basic steps (cf. Kim & Cohen, 1999):

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.

A discussion of the four steps involved is briefly presented below using the Q-E intelligence test data.

To complete the specification of a full probability model, prior distributions of the parameters (i.e., θ_i, a_j, and b_j) need to be specified. Following Albert (1992) and Johnson and Albert (1999), the prior distribution of ability was specified as normal with mean 0 and variance 1, $p(\theta_j) = N(0,1)$. Priors for item parameters can be specified in several different ways. For example, we can impose priors directly on a_j and b_j or hierarchically using additional hyperparameters (e.g., Swaminathan & Gifford, 1985; Kim, Cohen, Baker, Subkoviak, & Leonard, 1994). Uninformative priors can be imposed, if it is preferred that the priors not be too influential. Alternatively, it may also be useful to include external information in the form of fairly informative prior distributions. For this study, two prior distributions were chosen for the item parameters: (1) $a_j \sim N(1,1)$ and $b_j \sim N(0,1)$, and (2) $a_j \sim N(1,.5^2)$ and $b_j \sim N(0,2^2)$. Two rather informative prior distributions were chosen for the item parameters due to the small sample size (see Harwell & Janosky, 1991).

The choice of starting values is not generally that critical as the Gibbs sampler should be run long enough to be sufficiently updated from its initial states. Good initial values certainly do not hurt (Johnson & Albert, 1999). Starting values suggested in Johnson and Albert (1999) were used in Gibbs sampling. In their strategy, the initial values of θ_i are set to 0 which is the value of their prior mean. The initial values of all item discrimination parameters are set to μ_a in accordance with their prior means. For both prior conditions $\mu_a = 0$ in this study. Using the initial values of the discrimination parameters $\{a_j\}$, the negative item intercept parameters $\{b_j\}$ can be initialized by solving for b_j in the equation,

$$p_j = \Phi\left(\frac{-b_j}{\sqrt{1+a_j^2}}\right). \tag{23}$$

In other words, the initial values of b_j are set to be

$$b_j = -\sqrt{1+\mu_a^2}\,\Phi^{-1}(\hat{p}_j), \tag{24}$$

where $\Phi^{-1}()$ denotes that standard normal quantile function and $\hat{p}_j$ denotes the proportion of individuals responding correctly to item j defined as

$$\hat{p}_j = \frac{\sum_{i=1}^{N} Y_{ij} + .5}{N+1}. \tag{25}$$

A critical issue in 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. The values for the unknown quantities generated by the Gibbs sampler can be graphically and statistically summarized to check convergence (see Best, Cowles, & Vines, 1997; Johnson & Albert, 1999). Following Albert (1992) and Baker (1998), the sampled values of $\boldsymbol{\omega}$ (i.e., $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$) were recorded at the 10th cycle of the Gibbs sampler, and at every 5th cycle thereafter. A total of 10,000 recorded sample values were used to obtain the density estimates for the ability and item parameters.

The posterior mean of the Gibbs sampler was obtained for each parameter using the 10,000 sampled values. In addition, the 95% posterior interval and the posterior standard deviation were obtained for each parameter using the sampled values. Each 95% posterior interval was not obtained from the normal approximation using the posterior standard deviation. Rather, the posterior interval was obtained from the ordered 10,000 sampled values.

Marginal Estimation Methods

For comparison purposes, estimates from the two marginalized methods, denoted as marginal Bayesian estimation and BILOG, were obtained. In marginal Bayesian estimation, the marginalized posterior distribution,

$$p(\boldsymbol{\xi}|\mathrm{Y}) = \int p(\boldsymbol{\theta}, \boldsymbol{\xi}|\mathrm{Y})p(\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{26}$$

was maximized to obtain item parameter estimates under the normal ogive model (Bock & Aitkin, 1981; Bock & Lieberman, 1970; Kim, 1994). The same priors were employed as in Gibbs sampling. After obtaining the item parameter estimates, expected a posteriori estimation (Bock & Mislevy, 1982) was used to obtain ability parameter estimates in marginal Bayesian estimation. The computer program BILOG (Mislevy & Bock, 1990) was also used to obtain item and ability parameter estimates for the Q-E intelligence test. For the BILOG runs, two different item priors were used: (1) $p(\log \alpha_j) = N(0,1)$ and $p(b_j) = N(0,1)$, and (2) $p(\log \alpha_j) = N(0,.5^2)$ and $p(b_j) = N(0,2^2)$. Otherwise, default options were used in the BILOG runs. Expected a posteriori estimation was used to obtain ability parameter estimates. A standard normal prior was used for ability.

For the marginal Bayesian estimation, the standard errors (i.e., posterior variances) of the item parameter estimates were not obtained by inverting the empirical information matrix in the Fisher-scoring solution. Instead the block-diagonal approximation to the information matrix was used. This implementation may underestimate the large-sample posterior standard deviations. The model analyzed in the BILOG runs was the two-parameter logistic model with $D = 1.7$.

Results

Item parameter estimates from Gibbs sampling are presented in Table 2 for the prior-1 and prior-2 conditions (see also Figure 1). In addition, the 95% posterior intervals and the approximated values of the posterior standard deviation are also presented in Table 2. For the approximation, each 95% posterior interval was divided by 3.92, but for brevity it is denoted as PI/4.

Insert Table 2 and Figure 1 about here

The prior-1 condition employed two normal priors, $p(a_j) = N(1,1)$ and $p(b_j) = N(0,1)$. The prior-2 condition also employed two normal, but slightly different, priors from the prior-1 condition, $p(a_j) = N(1,.5^2)$ and $p(b_j) = N(0,2^2)$. Because the prior variance of a_j in the prior-2 condition was smaller than that of the prior-1 condition, the shrinkage of $\hat{a}_j$ toward the prior mean was greater for the prior-2 condition. In addition, the 95% posterior intervals of $\hat{a}_j$ were narrower for the prior-2 condition. This pattern can also be seen in the PI/4 values. The prior variance of b_j was smaller in the prior-1 condition. Consequently, the shrinkage of $\hat{b}_j$ toward the prior mean was greater for the items with large estimates and the 95% posterior intervals of $\hat{b}_j$ were generally narrower for the prior-1 condition than for the prior-2.

The posterior standard deviations of the sampled values of Gibbs sampling are reported in Table 3 for the prior-1 and prior-2 conditions. The 95% posterior intervals can be calculated from the posterior standard deviations. Table 4 contains the normal approximated 95% posterior intervals. In essence the posterior standard deviations and the normal approximated 95% posterior intervals have the same information with regard to the uncertainty of parameter estimates. Note that the estimates $\hat{a}_j$ and $\hat{b}_j$ of Gibbs sampling are the same for both Tables 2 and 3. When the PI/4 values and the posterior standard deviations are compared in Gibbs sampling, it can be seen in Table 3 that the magnitudes of the PI/4 values were uniformly smaller than those of the posterior standard deviations in Table 4. Hence, the posterior intervals of the sampled parameters were consistently narrower than those from the normal approximation.

Insert Tables 3 and 4 about here

Item parameter estimates from the marginal Bayesian estimation and the computer program BILOG are reported in Table 3 for both prior-1 and prior-2 conditions. Before comparing the item parameter estimates, it should be noted that the priors used in the BILOG runs were different from those used in Gibbs sampling and marginal Bayesian estimation. The priors used in BILOG were on the log item discrimination parameter and directly on the item difficulty parameter. The prior-1 condition of BILOG used $p(\log \alpha_j) = N(0,1)$ and $p(\beta_j) = N(0,1)$. The prior-2 condition of BILOG used $p(\log \alpha_j) = N(0,.5^2)$ and $p(\beta_j) = N(0,2^2)$. If we consider only the first two moments of the distribution for

$\log \alpha_j \sim N(\mu, \sigma^2)$, the expected value and the variance of α_j are defined as

$$E(\alpha_j) = \exp\left(\mu + \frac{1}{2}\sigma^2\right) \tag{27}$$

and

$$\mathrm{Var}(\alpha_j) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1] \tag{28}$$

(Hogg & Craig, 1978). For example, the specification $p(\log \alpha_j) = N(0,1)$ yields $E(\alpha_j) = \exp(.5) = 1.649$ and $\mathrm{Var}(\alpha_j) = \exp(1)[\exp(1) - 1] = 4.671$; the specification $p(\log \alpha_j) = N(0, .5^2)$ yields $E(\alpha_j) = 1.133$ and $\mathrm{Var}(\alpha_j) = .365$. These specifications are not the same as the prior specifications used in Gibbs sampling and marginal Bayesian estimation. Moreover, if $p(a_j) = N(\mu_1, \sigma_1^2)$ and $p(\beta_j) = N(\mu_2, \sigma_2^2)$, then $b_j = a_j\beta_j$ and will be distributed as normal with mean

$$\mu_{b_j} = \mu_1\mu_2 \tag{29}$$

and variance

$$\sigma_{b_j}^2 = \sigma_1^2\sigma_2^2 + \mu_1\sigma_2^2 + \mu_2\sigma_1^2 \tag{30}$$

(Hogg & Craig, 1978). For the prior-1 condition $p(a_j) = N(1,1)$ and $p(b_j) = N(0,1)$, and consequently $p(\beta_j) = N(0, \sqrt{2}^{-2}) = N(0, .707^2)$. The specifications in the prior-2 condition, $p(a_j) = N(1, .5^2)$ and $p(b_j) = N(0, 2^2)$, yield $p(\beta_j) = N(0, \sqrt{3.2}^2) = N(0, 1.789^2)$. Again, it should be noted that the default priors used in BILOG were slightly different from those used in Gibbs sampling and in marginal Bayesian estimation.

The values of $\hat{a}_j$ from the marginal Bayesian estimation were consistently smaller than the corresponding values from Gibbs sampling for both prior conditions (see Table 3 and Figure 1). The values of $\hat{b}_j$ from the marginal Bayesian estimation were uniformly larger and shifted in a positive direction than the corresponding values from Gibbs sampling for both prior conditions. Prior-2 yielded a narrow range for $\hat{a}_j$ due to the shrinkage effect.

The general patterns of the item parameter estimates of BILOG were very similar to those of Gibbs sampling. As was the case for both Gibbs sampling and marginal Bayesian estimation, the prior-2 condition yielded $\hat{a}_j$ that had shrunken closer to the prior mean than did the prior-1 condition. The values of $\hat{a}_j$ were very similar but those which had smaller $\hat{a}_j$ in Gibbs sampling became somewhat inflated in BILOG for both prior conditions. As was the case for marginal Bayesian estimation, the values of $\hat{b}_j$ from BILOG were shifted to the positive direction when compared to the corresponding values from Gibbs sampling (see Table 2 and Figure 1).

The posterior intervals (as well as PI/4) and the posterior standard deviations can also be used to compare different estimation procedures. There are two ways we obtain the posterior intervals for Gibbs sampling. One way is to use the actual sampled parameter values and the other way is to use the posterior standard deviations to construct the interval using such an equation, for example, as $\hat{a}_j \pm 1.96 \times \text{PSD}$. The two sets of the 95% posterior intervals of Gibbs sampling were reported in Tables 2 and 4, respectively.

Figure 2 presents the magnitudes of the PI/4 values and the posterior standard deviations of item parameter estimates. The y-axis in Figure 2 contains either PI/4 or posterior standard deviation. Marginal Bayesian estimation yielded smaller posterior standard deviations whereas BILOG yielded larger posterior standard deviations than did other estimation methods for both prior conditions. Item parameter estimates from Gibbs sampling, marginal Bayesian estimation, and BILOG are different due to the prior specification and the implementations of the estimation methods. Even so, the item parameter estimates obtained from all estimation methods differed very little as did the sizes of the posterior standard deviations and the PI/4 values.

Insert Figure 2 about here

The ability estimates of Gibbs sampling and the 95% posterior intervals are reported in Table 5. Since the same standard normal distribution was used here in both priors for ability estimation, the impact of prior on ability estimates could not be evaluated. Instead, any differential effects were a function of the priors employed for item parameters. The prior-1 condition yielded generally higher $\hat{\theta}_i$ than did the prior-2 condition except for some ability estimates in the middle of the scale and except for those examinees who had a perfect score. Overall, prior-1 yielded a slightly more shrunken scale than did the prior-2 condition.

Insert Table 5 about here

The 95% posterior intervals were wide and the corresponding PI/4 values large reflecting the small number of items (i.e., only ten) used to obtain the ability estimates. The posterior intervals and the PI/4 values show the estimation errors to be smaller when the ability distribution was more closely matched to the difficulty level of the test. The extreme $\hat{\theta}_i$ yielded relatively wide posterior intervals and consequently larger PI/4 values. The same $\hat{\theta}_i$

of Gibbs sampling are reported in Table 6 for the two prior conditions, but the posterior standard deviations are reported instead of the 95% posterior intervals from the actual sampled parameters. The normal approximated 95% posterior intervals from Gibbs sampling are reported in Table 7 for comparison purposes. Table 6 contains the ability estimates of the expected a posteriori method using item parameter estimates from Gibbs sampling, marginal Bayesian estimation, and BILOG. All item parameters were assumed to be known in the expected a posteriori method. Although all of these ability estimates were in fact expected a posteriori estimates, we will use Gibbs sampling/EAP, marginal Bayesian estimation, and BILOG in this discussion.

Insert Tables 6 and 7 about here

Figure 3 presents the relationship among the estimation procedures. The top six triangle plots were from the prior-1 condition and the remaining bottom triangle plots were from the prior-2 condition. The ability estimates of Gibbs sampling and Gibbs sampling/EAP were similar for the prior-1 condition except in the middle of the ability scale. Marginal Bayesian estimation and BILOG yielded similar but slightly larger positive ability estimates than did Gibbs sampling. For the prior-2 condition, all estimation methods appeared to yield the same ability estimates. Ability estimates were plotted on the 45 degree line for the prior-2 condition.

Insert Figure 3 about here

The posterior intervals are reported in Table 5 for Gibbs sampling and in Table 7 for Gibbs sampling using the posterior standard deviations, Gibbs sampling/EAP, marginal Bayesian estimation, and BILOG. The results of the posterior intervals can also be summarized in terms of the PI/4 values and the posterior standard deviations as in Figure 4. The patterns of the PI/4 values and the posterior standard deviations of Gibbs sampling, presented in the top two rows, were different from those obtained using the expected a posteriori method. For the Gibbs sampling, the PI/4 values were smaller than the corresponding posterior standard deviations. Among the three expected a posteriori methods using the item parameter estimates of Gibbs sampling, marginal Bayesian estimation, BILOG, respectively, Gibbs sampling/EAP yielded smaller posterior standard deviations whereas marginal Bayesian estimation yielded larger posterior standard deviations of ability estimates. When the expected a posteriori method was used to obtain ability estimates, a set of items with high $\hat{a}_j$ seem to reduce the sizes of the posterior standard deviations.

Insert Figure 4 about here

Discussion

Previous work using Gibbs sampling and MCMC suggests this method may provide a useful alternative for estimation of IRT parameters when small sample sizes and small numbers of items are used. Even though implementation of Gibbs sampling in IRT is available in several computer programs, the accuracy of the resulting estimates has not been thoroughly studied. The results of the analyses of the Q-E intelligence test presented in this study indicate that item parameter and ability estimates were similar but the magnitudes of the posterior standard deviations were different compared to estimates from the marginalized methods studied.

The main difference in item parameter estimation between Gibbs sampling and the marginalized methods of marginal Bayesian estimation and BILOG, is in the way these methods obtain the estimates. Gibbs sampling uses the sample of item parameter values to estimate the mean and variance of the posterior density of the item parameter. Under either the marginal Bayesian estimation or marginal maximum a posteriori estimation implemented in BILOG, the marginalized posterior distribution is maximized to obtain the marginal modes to be used as the item parameter estimates. Of course, estimation of the ability parameters does not arise during the course of item parameter estimation under these 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 or sequentially after obtaining item parameter estimates as seen in this paper. In the latter case the obtained item parameter estimates were assumed to be the true values. Comparisons of the two sets of ability estimates indicated that both estimates were similar, but the forms as well as the sizes of the posterior standard deviations were different.

As indicated in Patz and Junker (1997), Gibbs sampling and general MCMC methods are likely to be more useful for situations where complicated models are employed. Gibbs sampling in this paper was compared with marginalized methods for a two-parameter IRT model. In addition, Patz and Junker (1999) suggest that one of the potential advantages of using Gibbs sampling and other MCMC methods is incorporation of the uncertainty in item parameter estimates into estimation of ability parameters. The analysis of the Q-E intelligence test data suggests that Gibbs sampling, marginal Bayesian estimation, and BILOG yield comparable item and ability estimates. For the item parameter estimates, the magnitudes of the posterior standard deviations depend on the implementation of the estimation. The sizes of the posterior standard deviations (or PI/4) of Gibbs sampling do not seem to depend upon some items with deviant item discrimination estimates.

Although the results for the Q-E intelligence test presented above are informative, they do 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 available, however, based on studies of parameter recovery employing simulated data (e.g., Hulin, Lissak, & Drasgow, 1982; Yen, 1983). What is needed at this point, with respect to Gibbs sampling, is for more analysis with data simulated under various sample size, test length, and test design conditions (e.g., broad range test, selection test) to understand the comparative characteristics.

Finally, note that the examinees with the same response patterns may have different ability estimates when the ability parameters are estimated jointly with item parameters in Gibbs sampling. This is not acceptable in practice. The sequential estimation of ability using previously obtained item parameter estimates from Gibbs sampling may remove such oddity.

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. (1992). *Item response theory: Parameter estimation techniques.* New York: Marcel Dekker.

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

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.

Finney, D. J. (1971). *Probit analysis* (3rd ed.). Cambridge, Great Britain: Cambridge University Press.

Fraenkel, J. R., & Wallen, N. E. (2000). *How to design & evaluate research in education* (4th ed.). Boston: McGraw-Hill.

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

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., 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.

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.). New York: Macmillan.

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.

Johnson, V. E., & Albert, J. H. (1999). *Ordinal data modeling.* New York: Springer-Verlag.

Kim, S.-H. (1994, April). *Hierarchical Bayesian estimation of item parameters.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Kim, S.-H., & Cohen, A. S. (1999, April). *Accuracy of parameter estimation in Gibbs sampling.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

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.

Leonard, T., & Hsu, J. S. J. (1999). *Bayesian methods.* New York: Cambridge University Press.

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). *Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses* (Tech. Rep. No. 670). Pittsburgh, PA: Carnegie Mellon University, Department of Statistics.

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics, 24,* 146–178.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in Fortran 77: The art of scientific computing* (2nd ed.). Cambridge, Great Britain: Cambridge University Press.

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.

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.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association, 82,* 528–540.

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

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

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

Table 1
The Q-E Intelligence Test Data

	Item										
Examinee	1	2	3	4	5	6	7	8	9	10	Score
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	1	1
3	0	1	0	0	0	0	0	0	0	1	2
4	1	1	0	0	0	0	0	0	0	0	2
5	1	1	0	0	0	0	0	0	0	1	3
6	0	1	1	0	0	0	0	0	0	1	3
7	0	1	0	0	0	1	0	0	0	1	3
8	1	1	0	0	0	0	1	0	0	0	3
9	1	1	0	0	0	0	1	0	0	0	3
10	1	1	0	0	0	0	1	0	1	0	4
11	1	1	0	0	0	0	1	0	1	0	4
12	1	1	0	0	0	0	1	1	0	0	4
13	1	1	0	0	0	1	1	0	0	0	4
14	1	1	1	0	0	0	1	0	0	1	5
15	1	1	1	0	0	0	1	0	1	0	5
16	1	1	0	0	1	0	0	1	1	0	5
17	0	1	1	1	0	1	1	0	0	0	5
18	1	1	1	0	0	0	1	1	0	1	6
19	1	1	0	1	0	0	1	0	1	1	6
20	1	1	1	1	0	0	0	1	0	1	6
21	1	1	1	0	0	0	1	1	1	0	6
22	1	1	0	0	1	0	1	1	1	0	6
23	1	1	1	0	1	0	1	1	0	0	6
24	1	1	1	1	0	1	1	0	0	0	6
25	1	1	0	1	0	1	1	1	0	0	6
26	0	1	0	1	1	1	1	1	0	0	6
27	1	1	1	0	0	0	1	1	1	1	7
28	1	1	1	0	1	0	1	0	1	1	7
29	1	1	1	1	0	0	1	1	0	1	7
30	1	1	0	0	1	1	1	1	0	1	7
31	0	1	0	1	1	1	1	1	1	0	7
32	1	1	1	1	0	0	1	1	1	1	8
33	1	1	1	1	1	1	1	0	0	1	8
34	1	1	1	1	1	1	1	0	0	1	8
35	1	0	1	1	1	1	1	1	0	1	8
36	1	1	1	1	1	1	1	1	0	0	8
37	1	1	1	1	1	1	1	1	0	1	9
38	1	0	1	1	1	1	1	1	1	1	9
39	1	1	1	1	1	1	1	1	1	0	9
40	1	1	1	1	1	1	1	1	1	1	10
41	1	1	1	1	1	1	1	1	1	1	10
42	1	1	1	1	1	1	1	1	1	1	10
43	1	1	1	1	1	1	1	1	1	1	10
44	1	1	1	1	1	1	1	1	1	1	10

Table 2

Estimated Item Parameters and 95% Posterior Intervals of the Q-E Intelligence Test Items

	Gibbs Sampling							
	Prior-1[a]				Prior-2[b]			
Item	$\hat{a}_j$ (PI/4)[c]	(Post. Interval)	$-\hat{b}_j^d$ (PI/4)	(Post. Interval)	$\hat{a}_j$ (PI/4)	(Post. Interval)	$-\hat{b}_j$ (PI/4)	(Post. Interval)
1	.729 (.297)	(.201, 1.366)	1.042 (.229)	(.617, 1.513)	.829 (.265)	(.351, 1.389)	1.167 (.258)	(.688, 1.699)
2	.233 (.248)	(−.237, .734)	1.316 (.220)	(.895, 1.757)	.395 (.231)	(−.043, .862)	1.435 (.242)	(.975, 1.923)
3	1.176 (.366)	(.556, 1.990)	.242 (.223)	(−.185, .689)	1.112 (.275)	(.616, 1.693)	.280 (.235)	(−.170, .753)
4	1.935 (.531)	(1.038, 3.120)	−.109 (.298)	(−.703, .465)	1.444 (.317)	(.876, 2.117)	−.071 (.277)	(−.616, .470)
5	1.872 (.503)	(1.020, 2.992)	−.304 (.285)	(−.870, .248)	1.435 (.302)	(.878, 2.063)	−.255 (.272)	(−.794, .273)
6	1.517 (.454)	(.736, 2.515)	−.155 (.246)	(−.638, .326)	1.190 (.288)	(.658, 1.788)	−.123 (.236)	(−.588, .338)
7	1.396 (.408)	(.676, 2.277)	1.240 (.300)	(.693, 1.869)	1.277 (.300)	(.711, 1.887)	1.320 (.312)	(.749, 1.971)
8	1.333 (.405)	(.646, 2.232)	.143 (.234)	(−.319, .599)	1.171 (.290)	(.640, 1.776)	.176 (.238)	(−.288, .643)
9	.633 (.244)	(.186, 1.144)	−.267 (.183)	(−.629, .088)	.694 (.220)	(.283, 1.144)	−.266 (.193)	(−.648, .109)
10	.360 (.214)	(−.034, .805)	.127 (.168)	(−.200, .457)	.441 (.198)	(.072, .847)	.147 (.178)	(−.199, .499)

[a]Priors were $p(a_j) = N(1,1)$ and $p(b_j) = N(0,1)$.

[b]Priors were $p(a_j) = N(1, .5^2)$ and $p(b_j) = N(0, 2^2)$.

[c]PI/4 = Posterior Interval/3.92

[d]The values are $\hat{\zeta}_j = -\hat{b}_j$ instead of $\hat{b}_j$.

Table 3

Estimated Item Parameters and Posterior Standard Deviations (PSDs) of the Q-E Intelligence Test Items

	Gibbs Sampling			
	Prior-1		Prior-2	
Item	$\hat{a}_j$ (PSD)	$-\hat{b}_j$ (PSD)	$\hat{a}_j$ (PSD)	$-\hat{b}_j$ (PSD)
1	.729 (.355)	1.042 (.276)	.829 (.316)	1.167 (.313)
2	.233 (.297)	1.316 (.265)	.395 (.277)	1.435 (.294)
3	1.176 (.453)	.242 (.268)	1.112 (.334)	.280 (.282)
4	1.935 (.643)	−.109 (.353)	1.444 (.376)	−.071 (.328)
5	1.872 (.603)	−.304 (.345)	1.435 (.364)	−.255 (.325)
6	1.517 (.546)	−.155 (.293)	1.190 (.342)	−.123 (.280)
7	1.396 (.501)	1.240 (.364)	1.277 (.359)	1.320 (.376)
8	1.333 (.491)	.143 (.278)	1.171 (.342)	.176 (.282)
9	.633 (.294)	−.267 (.217)	.694 (.265)	−.266 (.230)
10	.360 (.258)	.127 (.201)	.441 (.238)	.147 (.212)

	Marginal Bayesian			
	Prior-1		Prior-2	
Item	$\hat{a}_j$ (PSD)	$-\hat{b}_j$ (PSD)	$\hat{a}_j$ (PSD)	$-\hat{b}_j$ (PSD)
1	.580 (.246)	.938 (.235)	.711 (.243)	1.028 (.255)
2	.186 (.243)	1.251 (.250)	.333 (.236)	1.332 (.271)
3	.940 (.269)	.127 (.212)	.970 (.248)	.149 (.217)
4	1.552 (.387)	−.273 (.248)	1.318 (.292)	−.222 (.237)
5	1.437 (.365)	−.440 (.251)	1.299 (.292)	−.403 (.243)
6	1.206 (.316)	−.285 (.230)	1.083 (.262)	−.252 (.226)
7	1.194 (.352)	1.091 (.293)	1.189 (.303)	1.168 (.300)
8	1.077 (.292)	.035 (.218)	1.072 (.260)	.056 (.221)
9	.503 (.208)	−.305 (.200)	.605 (.206)	−.325 (.207)
10	.298 (.189)	.085 (.190)	.379 (.185)	.091 (.194)

	BILOG			
	Prior-1[a]		Prior-2[b]	
Item	$\hat{a}_j$ (PSD)	$-\hat{b}_j$ (PSD)	$\hat{a}_j$ (PSD)	$-\hat{b}_j$ (PSD)
1	.645 (.338)	.862 (.333)	.809 (.295)	1.098 (.381)
2	.359 (.188)	.885 (.254)	.607 (.191)	1.460 (.370)
3	.968 (.466)	.079 (.278)	.984 (.373)	.188 (.281)
4	1.856(1.301)	−.405 (.541)	1.414 (.659)	−.168 (.392)
5	1.525(1.023)	−.531 (.461)	1.363 (.669)	−.353 (.388)
6	1.265 (.588)	−.387 (.363)	1.116 (.424)	−.218 (.378)
7	1.333 (.797)	1.128 (.495)	1.300 (.601)	1.274 (.517)
8	1.122 (.533)	−.014 (.293)	1.093 (.418)	.114 (.300)
9	.534 (.259)	−.302 (.223)	.672 (.224)	−.298 (.245)
10	.379 (.190)	.048 (.206)	.520 (.163)	.099 (.241)

[a]Priors were $p(\log \alpha_j) = N(0,1)$ and $p(\beta_j) = N(0,1)$.
[b]Priors were $p(\log \alpha_j) = N(0, .5^2)$ and $p(\beta_j) = N(0, 2^2)$.

Table 4

Estimated Item Parameters and Normal-Approximated 95% Posterior Intervals of the Q-E Intelligence Test Items

	Gibbs Sampling							
	Prior-1				Prior-2			
Item	$\hat{a}_j$	(Post. Interval)	$-\hat{b}_j$	(Post. Interval)	$\hat{a}_j$	(Post. Interval)	$-\hat{b}_j$	(Post. Interval)
1	.729	(.033, 1.425)	1.042	(.501, 1.583)	.829	(.210, 1.448)	1.167	(.554, 1.780)
2	.233	(−.349, .815)	1.316	(.797, 1.835)	.395	(−.148, .938)	1.435	(.859, 2.011)
3	1.176	(.288, 2.064)	.242	(−.283, .767)	1.112	(.457, 1.767)	.280	(−.273, .833)
4	1.935	(.675, 3.195)	−.109	(−.801, .583)	1.444	(.707, 2.181)	−.071	(−.714, .572)
5	1.872	(.690, 3.054)	−.304	(−.980, .372)	1.435	(.722, 2.148)	−.255	(−.892, .382)
6	1.517	(.447, 2.587)	−.155	(−.729, .419)	1.190	(.520, 1.860)	−.123	(−.672, .426)
7	1.396	(.414, 2.378)	1.240	(.527, 1.953)	1.277	(.573, 1.981)	1.320	(.583, 2.057)
8	1.333	(.371, 2.295)	.143	(−.402, .688)	1.171	(.501, 1.841)	.176	(−.377, .729)
9	.633	(.057, 1.209)	−.267	(−.692, .158)	.694	(.175, 1.213)	−.266	(−.717, .185)
10	.360	(−.146, .866)	.127	(−.267, .521)	.441	(−.025, .907)	.147	(−.269, .563)

	Marginal Bayesian							
	Prior-1				Prior-2			
Item	$\hat{a}_j$	(Post. Interval)	$-\hat{b}_j$	(Post. Interval)	$\hat{a}_j$	(Post. Interval)	$-\hat{b}_j$	(Post. Interval)
1	.580	(.098, 1.062)	.938	(.477, 1.399)	.711	(.235, 1.187)	1.028	(.528, 1.528)
2	.186	(−.290, .662)	1.251	(.761, 1.741)	.333	(−.130, .796)	1.332	(.801, 1.863)
3	.940	(.413, 1.467)	.127	(−.289, .543)	.970	(.484, 1.456)	.149	(−.276, .574)
4	1.552	(.793, 2.311)	−.273	(−.759, .213)	1.318	(.746, 1.890)	−.222	(−.687, .243)
5	1.437	(.722, 2.152)	−.440	(−.932, .052)	1.299	(.727, 1.871)	−.403	(−.879, .073)
6	1.206	(.587, 1.825)	−.285	(−.736, .166)	1.083	(.569, 1.597)	−.252	(−.695, .191)
7	1.194	(.504, 1.884)	1.091	(.517, 1.665)	1.189	(.595, 1.783)	1.168	(.580, 1.756)
8	1.077	(.505, 1.649)	.035	(−.392, .462)	1.072	(.562, 1.582)	.056	(−.377, .489)
9	.503	(.095, .911)	−.305	(−.697, .087)	.605	(.201, 1.009)	−.325	(−.731, .081)
10	.298	(−.072, .668)	.085	(−.287, .457)	.379	(.016, .742)	.091	(−.289, .471)

	BILOG							
	Prior-1*				Prior-2*			
Item	$\hat{a}_j$	(Post. Interval)	$-\hat{b}_j$	(Post. Interval)	$\hat{a}_j$	(Post. Interval)	$-\hat{b}_j$	(Post. Interval)
1	.645	(−.017, 1.307)	.862	(.209, 1.515)	.809	(.231, 1.387)	1.098	(.351, 1.845)
2	.359	(−.009, .727)	.885	(.387, 1.383)	.607	(.233, .981)	1.460	(.735, 2.185)
3	.968	(.055, 1.881)	.079	(−.466, .624)	.984	(.253, 1.715)	.188	(−.363, .739)
4	1.856	(−.694, 4.406)	−.405	(−1.465, .655)	1.414	(.122, 2.706)	−.168	(−.936, .600)
5	1.525	(−.480, 3.530)	−.531	(−1.435, .373)	1.363	(.052, 2.674)	−.353	(−1.113, .407)
6	1.265	(.113, 2.417)	−.387	(−1.098, .324)	1.116	(.285, 1.947)	−.218	(−.959, .523)
7	1.333	(−.229, 2.895)	1.128	(.158, 2.098)	1.300	(.122, 2.478)	1.274	(.261, 2.287)
8	1.122	(.077, 2.167)	−.014	(−.588, .560)	1.093	(.274, 1.912)	.114	(−.474, .702)
9	.534	(.026, 1.042)	−.302	(−.739, .135)	.672	(.233, 1.111)	−.298	(−.778, .182)
10	.379	(.007, .751)	.048	(−.356, .452)	.520	(.201, .839)	.099	(−.373, .571)

Table 5

Ability Estimates and 95% Posterior Intervals of the Q-E Intelligence Test

	Gibbs Sampling			
	Prior-1		Prior-2	
Examinee	θ_i (PI/4)[a]	(Post. Interval)	θ_i (PI/4)	(Post. Interval)
1	−1.815 (.528)	(−2.946, −.875)	−1.992 (.526)	(−3.084, −1.024)
2	−1.570 (.521)	(−2.688, −.646)	−1.679 (.503)	(−2.733, −.761)
3	−1.488 (.472)	(−2.473, −.624)	−1.512 (.455)	(−2.442, −.658)
4	−1.265 (.438)	(−2.181, −.466)	−1.304 (.431)	(−2.191, −.501)
5	−1.144 (.420)	(−2.028, −.383)	−1.153 (.416)	(−2.005, −.376)
6	−.982 (.401)	(−1.832, −.261)	−1.042 (.398)	(−1.866, −.304)
7	−.773 (.372)	(−1.559, −.102)	−.933 (.385)	(−1.722, −.211)
8	−.813 (.378)	(−1.604, −.123)	−.874 (.394)	(−1.674, −.129)
9	−.838 (.389)	(−1.661, −.138)	−.894 (.395)	(−1.706, −.157)
10	−.645 (.361)	(−1.389, .025)	−.663 (.377)	(−1.416, .063)
11	−.620 (.345)	(−1.322, .031)	−.640 (.366)	(−1.365, .070)
12	−.476 (.340)	(−1.179, .155)	−.532 (.368)	(−1.271, .173)
13	−.414 (.333)	(−1.095, .211)	−.506 (.363)	(−1.234, .190)
14	−.437 (.330)	(−1.108, .185)	−.436 (.354)	(−1.135, .254)
15	−.382 (.321)	(−1.035, .222)	−.373 (.346)	(−1.061, .297)
16	−.241 (.307)	(−.861, .342)	−.309 (.340)	(−.985, .348)
17	−.008 (.296)	(−.586, .575)	−.120 (.335)	(−.775, .537)
18	−.175 (.310)	(−.794, .423)	−.150 (.345)	(−.834, .519)
19	−.149 (.294)	(−.745, .409)	−.171 (.335)	(−.827, .486)
20	−.105 (.297)	(−.688, .476)	−.149 (.332)	(−.801, .502)
21	−.120 (.318)	(−.748, .498)	−.082 (.355)	(−.773, .617)
22	.014 (.305)	(−.587, .609)	.006 (.346)	(−.677, .679)
23	.090 (.303)	(−.499, .687)	.084 (.345)	(−.594, .760)
24	.135 (.302)	(−.445, .738)	.098 (.340)	(−.562, .771)
25	.156 (.303)	(−.419, .768)	.102 (.345)	(−.567, .785)
26	.331 (.324)	(−.276, .996)	.198 (.355)	(−.480, .913)
27	−.066 (.314)	(−.678, .554)	.009 (.350)	(−.676, .695)
28	.031 (.298)	(−.549, .620)	.070 (.343)	(−.603, .740)
29	.155 (.313)	(−.450, .778)	.180 (.358)	(−.510, .893)
30	.205 (.301)	(−.362, .816)	.203 (.346)	(−.452, .903)
31	.471 (.353)	(−.194, 1.190)	.372 (.374)	(−.346, 1.119)
32	.274 (.310)	(−.310, .904)	.360 (.354)	(−.312, 1.076)
33	.573 (.357)	(−.070, 1.330)	.568 (.385)	(−.148, 1.361)
34	.574 (.363)	(−.067, 1.354)	.572 (.390)	(−.135, 1.393)
35	.951 (.440)	(.172, 1.898)	.841 (.440)	(.035, 1.759)
36	.888 (.408)	(.141, 1.739)	.841 (.428)	(.038, 1.716)
37	1.048 (.439)	(.254, 1.975)	1.038 (.450)	(.205, 1.968)
38	1.173 (.479)	(.330, 2.207)	1.090 (.473)	(.229, 2.082)
39	1.214 (.493)	(.354, 2.287)	1.196 (.492)	(.305, 2.232)
40	1.430 (.531)	(.492, 2.573)	1.466 (.527)	(.512, 2.577)
41	1.421 (.546)	(.468, 2.610)	1.462 (.550)	(.470, 2.626)
42	1.376 (.512)	(.471, 2.479)	1.426 (.522)	(.486, 2.534)
43	1.365 (.499)	(.475, 2.430)	1.407 (.501)	(.486, 2.451)
44	1.341 (.484)	(.472, 2.370)	1.399 (.500)	(.486, 2.446)

[c]PI/4 = Posterior Interval/3.92

BEST COPY AVAILABLE

Table 6

Ability Estimates and Posterior Standard Deviations (PSDs) of the Q-E Intelligence Test

	Gibbs Sampling		Gibbs Sampling/EAP		Marginal Bayesian		BILOG	
	Prior-1	Prior-2	Prior-1	Prior-2	Prior-1	Prior-2	Prior-1*	Prior-2*
Examinee	θ_i (PSD)	θ_i (PSD)	θ_i (PSD)	θ_i (PSD)	θ_i (PSD)	θ_i (PSD)	θ_i (PSD)	θ_i (PSD)
1	−1.815 (.633)	−1.992 (.626)	−1.823 (.607)	−1.955 (.613)	−1.801 (.631)	−1.903 (.627)	−1.789 (.627)	−1.972 (.609)
2	−1.570 (.632)	−1.679 (.603)	−1.622 (.550)	−1.705 (.539)	−1.616 (.591)	−1.674 (.572)	−1.556 (.578)	−1.674 (.549)
3	−1.488 (.563)	−1.512 (.544)	−1.504 (.529)	−1.526 (.497)	−1.506 (.576)	−1.498 (.544)	−1.364 (.546)	−1.399 (.491)
4	−1.265 (.528)	−1.304 (.517)	−1.342 (.516)	−1.376 (.485)	−1.358 (.565)	−1.344 (.532)	−1.233 (.533)	−1.283 (.482)
5	−1.144 (.507)	−1.153 (.501)	−1.180 (.524)	−1.196 (.505)	−1.202 (.558)	−1.164 (.536)	−1.054 (.519)	−1.077 (.487)
6	−.982 (.481)	−1.042 (.477)	−.922 (.498)	−1.032 (.503)	−.988 (.521)	−1.002 (.516)	−.911 (.500)	−1.006 (.487)
7	−.773 (.448)	−.933 (.466)	−.674 (.393)	−.925 (.499)	−.765 (.462)	−.877 (.499)	−.792 (.471)	−.953 (.485)
8	−.813 (.452)	−.874 (.469)	−.747 (.441)	−.841 (.484)	−.782 (.480)	−.799 (.486)	−.677 (.428)	−.773 (.456)
9	−.838 (.464)	−.894 (.475)	−.747 (.441)	−.841 (.484)	−.782 (.480)	−.799 (.486)	−.677 (.428)	−.773 (.456)
10	−.645 (.432)	−.663 (.449)	−.588 (.317)	−.615 (.367)	−.615 (.403)	−.593 (.399)	−.537 (.359)	−.574 (.375)
11	−.620 (.415)	−.640 (.439)	−.588 (.317)	−.615 (.367)	−.615 (.403)	−.593 (.399)	−.537 (.359)	−.574 (.375)
12	−.476 (.408)	−.532 (.439)	−.495 (.208)	−.520 (.296)	−.473 (.346)	−.475 (.359)	−.421 (.335)	−.484 (.338)
13	−.414 (.396)	−.506 (.432)	−.475 (.188)	−.506 (.277)	−.433 (.332)	−.461 (.345)	−.393 (.340)	−.480 (.337)
14	−.437 (.391)	−.436 (.423)	−.486 (.213)	−.471 (.291)	−.449 (.356)	−.422 (.373)	−.377 (.346)	−.406 (.337)
15	−.382 (.382)	−.373 (.413)	−.464 (.219)	−.432 (.304)	−.400 (.373)	−.361 (.397)	−.344 (.358)	−.376 (.346)
16	−.241 (.367)	−.309 (.407)	−.422 (.244)	−.423 (.263)	−.327 (.375)	−.336 (.372)	−.261 (.396)	−.340 (.361)
17	−.008 (.354)	−.120 (.402)	.001 (.485)	−.201 (.444)	.159 (.464)	−.015 (.490)	.244 (.389)	−.090 (.449)
18	−.175 (.374)	−.150 (.414)	−.322 (.368)	−.229 (.436)	−.148 (.473)	−.091 (.487)	−.068 (.450)	−.126 (.442)
19	−.149 (.350)	−.171 (.401)	−.304 (.380)	−.270 (.410)	−.119 (.476)	−.126 (.477)	.036 (.451)	−.123 (.443)
20	−.105 (.356)	−.149 (.398)	−.252 (.415)	−.263 (.411)	−.070 (.483)	−.116 (.476)	.111 (.439)	−.087 (.450)
21	−.120 (.377)	−.082 (.420)	−.251 (.417)	−.142 (.468)	−.060 (.488)	.009 (.493)	−.015 (.453)	−.075 (.452)
22	.014 (.364)	.006 (.412)	.008 (.485)	−.016 (.487)	.143 (.468)	.142 (.472)	.171 (.421)	.059 (.455)
23	.090 (.365)	.084 (.413)	.207 (.439)	.146 (.467)	.279 (.410)	.262 (.431)	.286 (.366)	.165 (.435)
24	.135 (.360)	.098 (.408)	.275 (.401)	.155 (.464)	.341 (.373)	.269 (.427)	.377 (.308)	.189 (.428)
25	.156 (.365)	.102 (.411)	.315 (.371)	.176 (.457)	.372 (.353)	.300 (.413)	.400 (.296)	.222 (.418)
26	.331 (.391)	.198 (.424)	.460 (.203)	.317 (.383)	.509 (.304)	.417 (.339)	.531 (.322)	.364 (.363)
27	−.066 (.375)	.009 (.420)	−.140 (.465)	.020 (.487)	.051 (.488)	.149 (.472)	.115 (.438)	.108 (.448)
28	.031 (.356)	.070 (.407)	.078 (.479)	.124 (.472)	.191 (.452)	.231 (.441)	.235 (.394)	.196 (.426)
29	.155 (.373)	.180 (.423)	.314 (.372)	.284 (.408)	.375 (.353)	.367 (.382)	.412 (.291)	.326 (.377)
30	.205 (.362)	.203 (.411)	.372 (.315)	.301 (.393)	.401 (.324)	.384 (.362)	.407 (.293)	.345 (.370)
31	.471 (.421)	.372 (.449)	.497 (.193)	.441 (.287)	.595 (.354)	.522 (.327)	.648 (.400)	.507 (.361)
32	.274 (.372)	.360 (.425)	.416 (.259)	.427 (.309)	.462 (.302)	.493 (.343)	.488 (.297)	.471 (.352)
33	.573 (.434)	.568 (.465)	.536 (.249)	.552 (.326)	.692 (.426)	.662 (.421)	.752 (.452)	.648 (.424)
34	.574 (.438)	.572 (.470)	.536 (.249)	.552 (.326)	.692 (.426)	.662 (.421)	.752 (.452)	.648 (.424)
35	.951 (.530)	.841 (.530)	.874 (.511)	.775 (.476)	1.110 (.558)	.957 (.535)	1.070 (.525)	.822 (.491)
36	.888 (.493)	.841 (.507)	.851 (.503)	.823 (.496)	1.079 (.556)	.996 (.544)	1.060 (.524)	.859 (.501)
37	1.048 (.528)	1.038 (.540)	1.022 (.555)	1.029 (.552)	1.237 (.570)	1.192 (.564)	1.244 (.544)	1.104 (.548)
38	1.173 (.570)	1.090 (.565)	1.202 (.588)	1.103 (.570)	1.376 (.583)	1.264 (.571)	1.334 (.556)	1.137 (.552)
39	1.214 (.596)	1.196 (.595)	1.174 (.586)	1.168 (.576)	1.344 (.583)	1.310 (.577)	1.324 (.555)	1.182 (.558)
40	1.430 (.636)	1.466 (.638)	1.394 (.620)	1.435 (.613)	1.520 (.606)	1.533 (.605)	1.535 (.595)	1.480 (.606)
41	1.421 (.651)	1.462 (.650)	1.394 (.620)	1.435 (.613)	1.520 (.606)	1.533 (.605)	1.535 (.595)	1.480 (.606)
42	1.376 (.615)	1.426 (.628)	1.394 (.620)	1.435 (.613)	1.520 (.606)	1.533 (.605)	1.535 (.595)	1.480 (.606)
43	1.365 (.602)	1.407 (.606)	1.394 (.620)	1.435 (.613)	1.520 (.606)	1.533 (.605)	1.535 (.595)	1.480 (.606)
44	1.341 (.583)	1.399 (.601)	1.394 (.620)	1.435 (.613)	1.520 (.606)	1.533 (.605)	1.535 (.595)	1.480 (.606)

Table 7

Ability Estimates and Normal Approximated 95% Posterior Intervals of the Q-E Intelligence Test

	Gibbs Sampling				Gibbs Sampling/EAP			
	Prior-1		Prior-2		Prior-1		Prior-2	
Examinee	$\hat{\theta}_i$	(Post. Interval)	$\hat{\theta}_i$	(Post. Interval)	$\hat{\theta}_i$	(Post. Interval)	$\hat{\theta}_i$	(Post. Interval)
1	−1.815	(−3.056, −.574)	−1.992	(−3.219, −.765)	−1.823	(−3.013, −.633)	−1.955	(−3.156, −.754)
2	−1.570	(−2.809, −.331)	−1.679	(−2.861, −.497)	−1.622	(−2.700, −.544)	−1.705	(−2.761, −.649)
3	−1.488	(−2.591, −.385)	−1.512	(−2.578, −.446)	−1.504	(−2.541, −.467)	−1.526	(−2.500, −.552)
4	−1.265	(−2.300, −.230)	−1.304	(−2.317, −.291)	−1.342	(−2.353, −.331)	−1.376	(−2.327, −.425)
5	−1.144	(−2.138, −.150)	−1.153	(−2.135, −.171)	−1.180	(−2.207, −.153)	−1.196	(−2.186, −.206)
6	−.982	(−1.925, −.039)	−1.042	(−1.977, −.107)	−.922	(−1.898, .054)	−1.032	(−2.018, −.046)
7	−.773	(−1.651, .105)	−.933	(−1.846, −.020)	−.674	(−1.444, .096)	−.925	(−1.903, .053)
8	−.813	(−1.699, .073)	−.874	(−1.793, .045)	−.747	(−1.611, .117)	−.841	(−1.790, .108)
9	−.838	(−1.747, .071)	−.894	(−1.825, .037)	−.747	(−1.611, .117)	−.841	(−1.790, .108)
10	−.645	(−1.492, .202)	−.663	(−1.543, .217)	−.588	(−1.209, .033)	−.615	(−1.334, .104)
11	−.620	(−1.433, .193)	−.640	(−1.500, .220)	−.588	(−1.209, .033)	−.615	(−1.334, .104)
12	−.476	(−1.276, .324)	−.532	(−1.392, .328)	−.495	(−.903, −.087)	−.520	(−1.100, .060)
13	−.414	(−1.190, .362)	−.506	(−1.353, .341)	−.475	(−.843, −.107)	−.506	(−1.049, .037)
14	−.437	(−1.203, .329)	−.436	(−1.265, .393)	−.486	(−.903, −.069)	−.471	(−1.041, .099)
15	−.382	(−1.131, .367)	−.373	(−1.182, .436)	−.464	(−.893, −.035)	−.432	(−1.028, .164)
16	−.241	(−.960, .478)	−.309	(−1.107, .489)	−.422	(−.900, .056)	−.423	(−.938, .092)
17	−.008	(−.702, .686)	−.120	(−.908, .668)	.001	(−.950, .952)	−.201	(−1.071, .669)
18	−.175	(−.908, .558)	−.150	(−.961, .661)	−.322	(−1.043, .399)	−.229	(−1.084, .626)
19	−.149	(−.835, .537)	−.171	(−.957, .615)	−.304	(−1.049, .441)	−.270	(−1.074, .534)
20	−.105	(−.803, .593)	−.149	(−.929, .631)	−.252	(−1.065, .561)	−.263	(−1.069, .543)
21	−.120	(−.859, .619)	−.082	(−.905, .741)	−.251	(−1.068, .566)	−.142	(−1.059, .775)
22	.014	(−.699, .727)	.006	(−.802, .814)	.008	(−.943, .959)	−.016	(−.971, .939)
23	.090	(−.625, .805)	.084	(−.725, .893)	.207	(−.653, 1.067)	.146	(−.769, 1.061)
24	.135	(−.571, .841)	.098	(−.702, .898)	.275	(−.511, 1.061)	.155	(−.754, 1.064)
25	.156	(−.559, .871)	.102	(−.704, .908)	.315	(−.412, 1.042)	.176	(−.720, 1.072)
26	.331	(−.435, 1.097)	.198	(−.633, 1.029)	.460	(.062, .858)	.317	(−.434, 1.068)
27	−.066	(−.801, .669)	.009	(−.814, .832)	−.140	(−1.051, .771)	.020	(−.935, .975)
28	.031	(−.667, .729)	.070	(−.728, .868)	.078	(−.861, 1.017)	.124	(−.801, 1.049)
29	.155	(−.576, .886)	.180	(−.649, 1.009)	.314	(−.415, 1.043)	.284	(−.516, 1.084)
30	.205	(−.505, .915)	.203	(−.603, 1.009)	.372	(−.245, .989)	.301	(−.469, 1.071)
31	.471	(−.354, 1.296)	.372	(−.508, 1.252)	.497	(.119, .875)	.441	(−.122, 1.004)
32	.274	(−.455, 1.003)	.360	(−.473, 1.193)	.416	(−.092, .924)	.427	(−.179, 1.033)
33	.573	(−.278, 1.424)	.568	(−.343, 1.479)	.536	(.048, 1.024)	.552	(−.087, 1.191)
34	.574	(−.284, 1.432)	.572	(−.349, 1.493)	.536	(.048, 1.024)	.552	(−.087, 1.191)
35	.951	(−.088, 1.990)	.841	(−.198, 1.880)	.874	(−.128, 1.876)	.775	(−.158, 1.708)
36	.888	(−.078, 1.854)	.841	(−.153, 1.835)	.851	(−.135, 1.837)	.823	(−.149, 1.795)
37	1.048	(.013, 2.083)	1.038	(−.020, 2.096)	1.022	(−.066, 2.110)	1.029	(−.053, 2.111)
38	1.173	(.056, 2.290)	1.090	(−.017, 2.197)	1.202	(.050, 2.354)	1.103	(−.014, 2.220)
39	1.214	(.046, 2.382)	1.196	(.030, 2.362)	1.174	(.025, 2.323)	1.168	(.039, 2.297)
40	1.430	(.183, 2.677)	1.466	(.216, 2.716)	1.394	(.179, 2.609)	1.435	(.234, 2.636)
41	1.421	(.145, 2.697)	1.462	(.188, 2.736)	1.394	(.179, 2.609)	1.435	(.234, 2.636)
42	1.376	(.171, 2.581)	1.426	(.195, 2.657)	1.394	(.179, 2.609)	1.435	(.234, 2.636)
43	1.365	(.185, 2.545)	1.407	(.219, 2.595)	1.394	(.179, 2.609)	1.435	(.234, 2.636)
44	1.341	(.198, 2.484)	1.399	(.221, 2.577)	1.394	(.179, 2.609)	1.435	(.234, 2.636)

Table 7—Continued

Ability Estimates and Normal Approximated 95% Posterior Intervals of the Q-E Intelligence Test

	Marginal Bayesian				BILOG			
	Prior-1		Prior-2		Prior-1		Prior-2	
Examinee	θ_i	(Post. Interval)	θ_i	(Post. Interval)	θ_i	(Post. Interval)	θ_i	(Post. Interval)
1	−1.801	(−3.038, −.564)	−1.903	(−3.132, −.674)	−1.801	(−3.018, −.560)	−1.903	(−3.166, −.778)
2	−1.616	(−2.774, −.458)	−1.674	(−2.795, −.553)	−1.616	(−2.689, −.423)	−1.674	(−2.750, −.598)
3	−1.506	(−2.635, −.377)	−1.498	(−2.564, −.432)	−1.506	(−2.434, −.294)	−1.498	(−2.361, −.437)
4	−1.358	(−2.465, −.251)	−1.344	(−2.387, −.301)	−1.358	(−2.278, −.188)	−1.344	(−2.228, −.338)
5	−1.202	(−2.296, −.108)	−1.164	(−2.215, −.113)	−1.202	(−2.071, −.037)	−1.164	(−2.032, −.122)
6	−.988	(−2.009, .033)	−1.002	(−2.013, .009)	−.988	(−1.891, .069)	−1.002	(−1.961, −.051)
7	−.765	(−1.671, .141)	−.877	(−1.855, .101)	−.765	(−1.715, .131)	−.877	(−1.904, −.002)
8	−.782	(−1.723, .159)	−.799	(−1.752, .154)	−.782	(−1.516, .162)	−.799	(−1.667, .121)
9	−.782	(−1.723, .159)	−.799	(−1.752, .154)	−.782	(−1.516, .162)	−.799	(−1.667, .121)
10	−.615	(−1.405, .175)	−.593	(−1.375, .189)	−.615	(−1.241, .167)	−.593	(−1.309, .161)
11	−.615	(−1.405, .175)	−.593	(−1.375, .189)	−.615	(−1.241, .167)	−.593	(−1.309, .161)
12	−.473	(−1.151, .205)	−.475	(−1.179, .229)	−.473	(−1.078, .236)	−.475	(−1.146, .178)
13	−.433	(−1.084, .218)	−.461	(−1.137, .215)	−.433	(−1.059, .273)	−.461	(−1.141, .181)
14	−.449	(−1.147, .249)	−.422	(−1.153, .309)	−.449	(−1.055, .301)	−.422	(−1.067, .255)
15	−.400	(−1.131, .331)	−.361	(−1.139, .417)	−.400	(−1.046, .358)	−.361	(−1.054, .302)
16	−.327	(−1.062, .408)	−.336	(−1.065, .393)	−.327	(−1.037, .515)	−.336	(−1.048, .368)
17	.159	(−.750, 1.068)	−.015	(−.975, .945)	.159	(−.518, 1.006)	−.015	(−.970, .790)
18	−.148	(−1.075, .779)	−.091	(−1.046, .864)	−.148	(−.950, .814)	−.091	(−.992, .740)
19	−.119	(−1.052, .814)	−.126	(−1.061, .809)	−.119	(−.848, .920)	−.126	(−.991, .745)
20	−.070	(−1.017, .877)	−.116	(−1.049, .817)	−.070	(−.749, .971)	−.116	(−.969, .795)
21	−.060	(−1.016, .896)	.009	(−.957, .975)	−.060	(−.903, .873)	.009	(−.961, .811)
22	.143	(−.774, 1.060)	.142	(−.783, 1.067)	.143	(−.654, .996)	.142	(−.833, .951)
23	.279	(−.525, 1.083)	.262	(−.583, 1.107)	.279	(−.431, 1.003)	.262	(−.688, 1.018)
24	.341	(−.390, 1.072)	.269	(−.568, 1.106)	.341	(−.227, .981)	.269	(−.650, 1.028)
25	.372	(−.320, 1.064)	.300	(−.509, 1.109)	.372	(−.180, .980)	.300	(−.597, 1.041)
26	.509	(−.087, 1.105)	.417	(−.247, 1.081)	.509	(−.100, 1.162)	.417	(−.347, 1.075)
27	.051	(−.905, 1.007)	.149	(−.776, 1.074)	.051	(−.743, .973)	.149	(−.770, .986)
28	.191	(−.695, 1.077)	.231	(−.633, 1.095)	.191	(−.537, 1.007)	.231	(−.639, 1.031)
29	.375	(−.317, 1.067)	.367	(−.382, 1.116)	.375	(−.158, .982)	.367	(−.413, 1.065)
30	.401	(−.234, 1.036)	.384	(−.326, 1.094)	.401	(−.167, .981)	.384	(−.380, 1.070)
31	.595	(−.099, 1.289)	.522	(−.119, 1.163)	.595	(−.136, 1.432)	.522	(−.201, 1.215)
32	.462	(−.130, 1.054)	.493	(−.179, 1.165)	.462	(−.094, 1.070)	.493	(−.219, 1.161)
33	.692	(−.143, 1.527)	.662	(−.163, 1.487)	.692	(−.134, 1.638)	.662	(−.183, 1.479)
34	.692	(−.143, 1.527)	.662	(−.163, 1.487)	.692	(−.134, 1.638)	.662	(−.183, 1.479)
35	1.110	(.016, 2.204)	.957	(−.092, 2.006)	1.110	(.041, 2.099)	.957	(−.140, 1.784)
36	1.079	(−.011, 2.169)	.996	(−.070, 2.062)	1.079	(.033, 2.087)	.996	(−.123, 1.841)
37	1.237	(.120, 2.354)	1.192	(.087, 2.297)	1.237	(.178, 2.310)	1.192	(.030, 2.178)
38	1.376	(.233, 2.519)	1.264	(.145, 2.383)	1.376	(.244, 2.424)	1.264	(.055, 2.219)
39	1.344	(.201, 2.487)	1.310	(.179, 2.441)	1.344	(.236, 2.412)	1.310	(.088, 2.276)
40	1.520	(.332, 2.708)	1.533	(.347, 2.719)	1.520	(.369, 2.701)	1.533	(.292, 2.668)
41	1.520	(.332, 2.708)	1.533	(.347, 2.719)	1.520	(.369, 2.701)	1.533	(.292, 2.668)
42	1.520	(.332, 2.708)	1.533	(.347, 2.719)	1.520	(.369, 2.701)	1.533	(.292, 2.668)
43	1.520	(.332, 2.708)	1.533	(.347, 2.719)	1.520	(.369, 2.701)	1.533	(.292, 2.668)
44	1.520	(.332, 2.708)	1.533	(.347, 2.719)	1.520	(.369, 2.701)	1.533	(.292, 2.668)

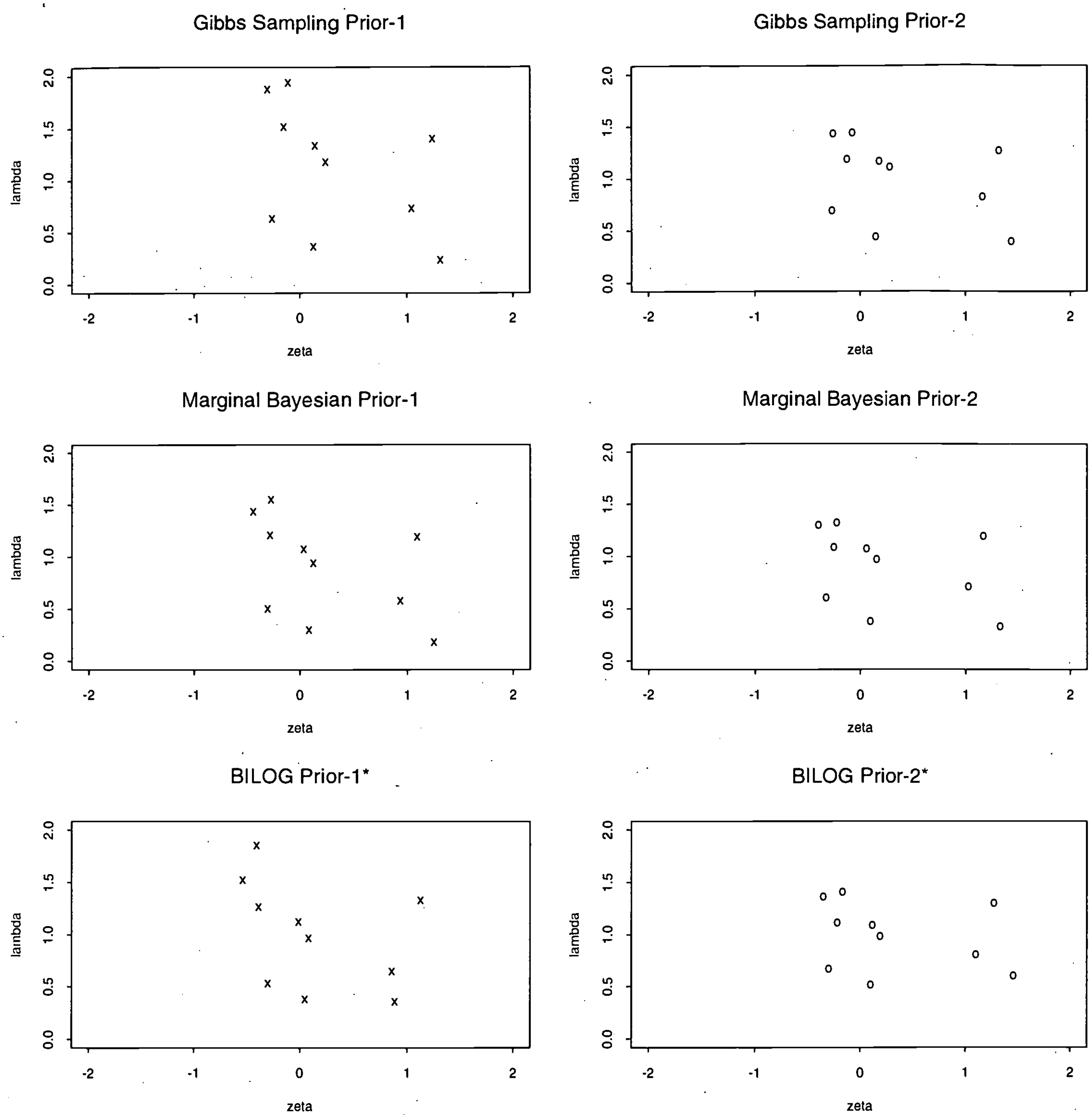
Figure Captions

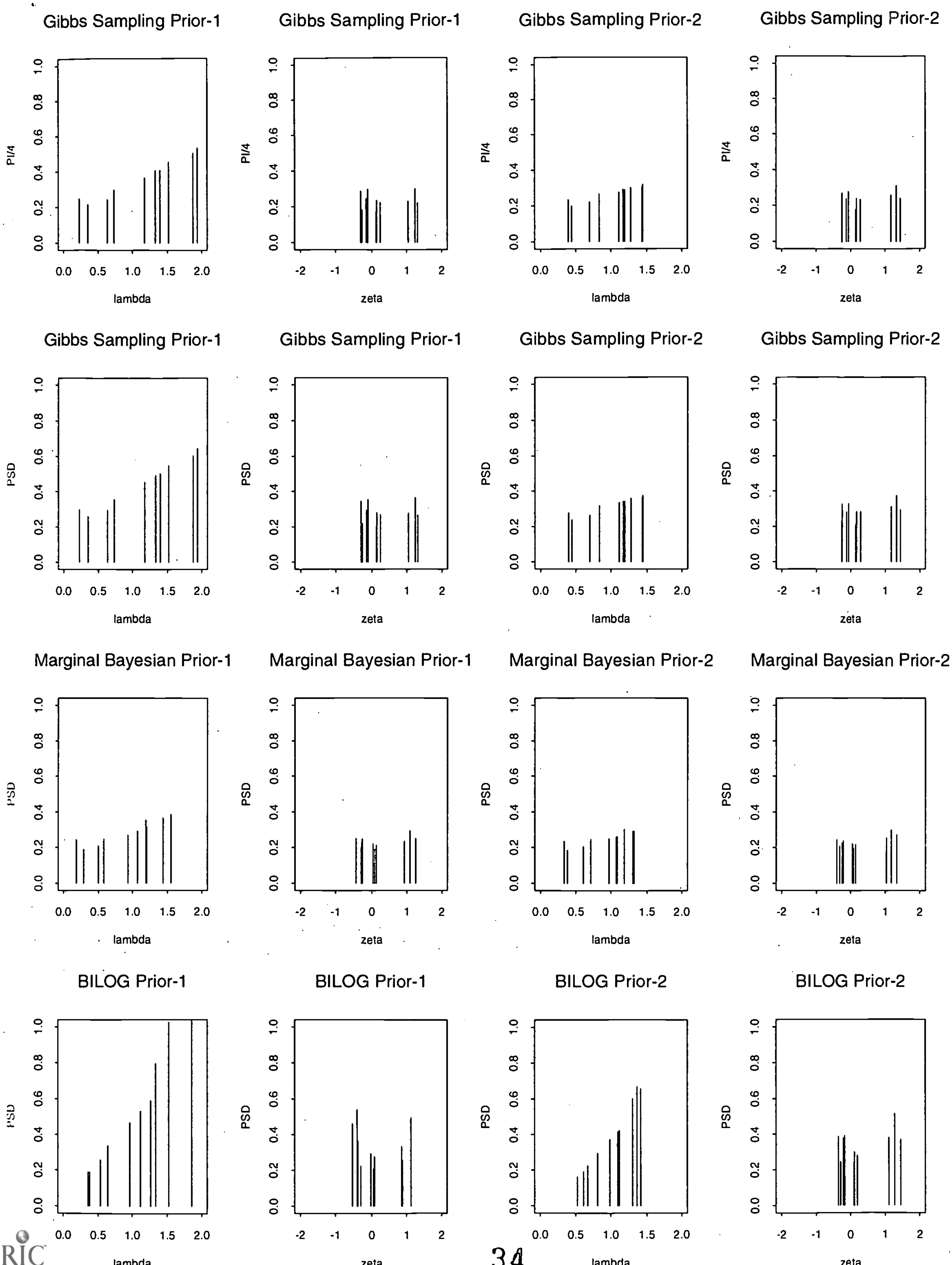
Figure 1. Plots of item parameter estimates (lambda = $\hat{\lambda}_j = \hat{a}_j$, zeta = $\hat{\zeta}_j = -\hat{b}_j$).

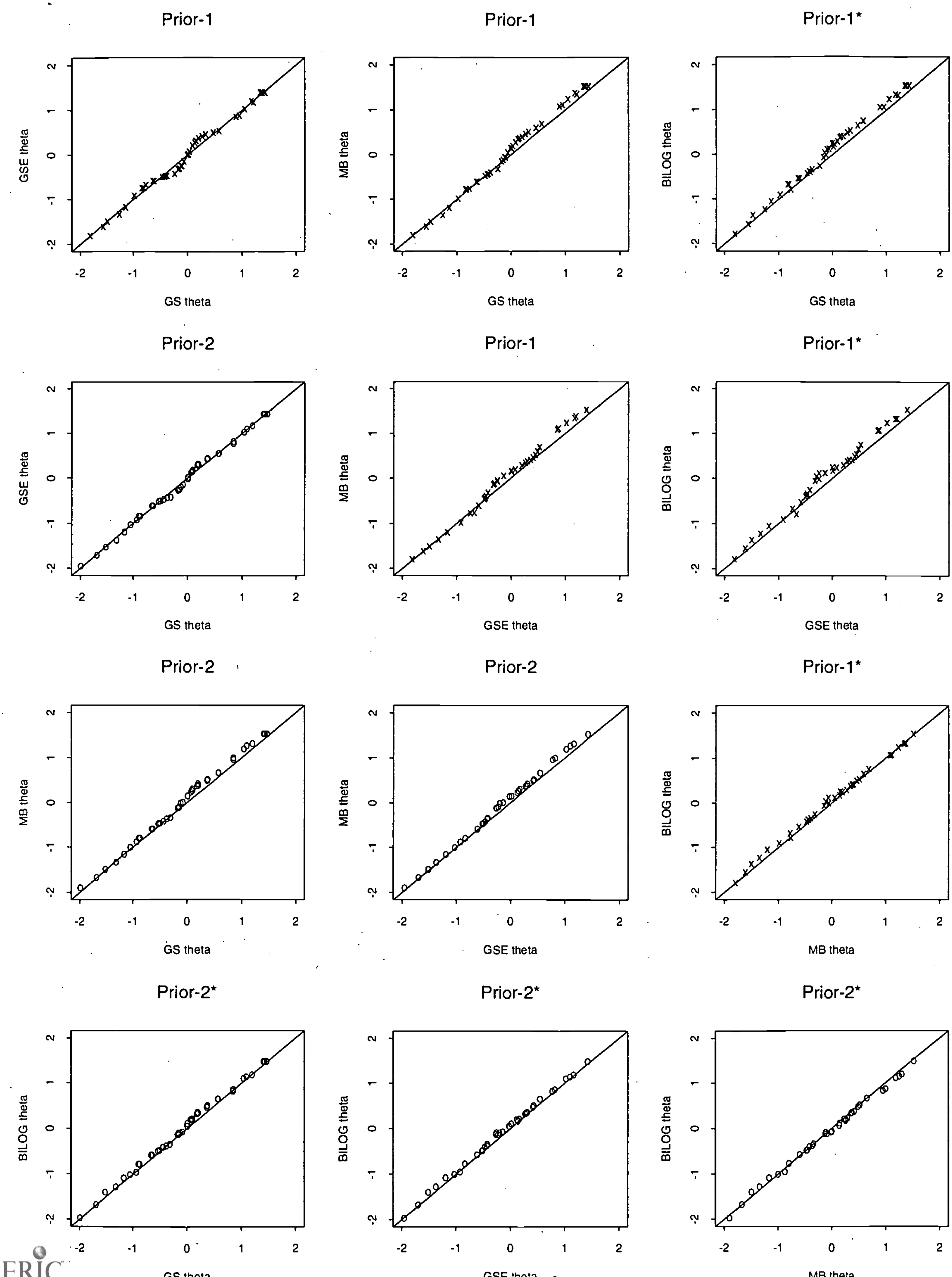
Figure 2. Magnitudes of posterior standard deviations of item parameter estimates.

Figure 3. Plots of ability estimates.

Figure 4. Magnitudes of posterior standard deviations of ability estimates.

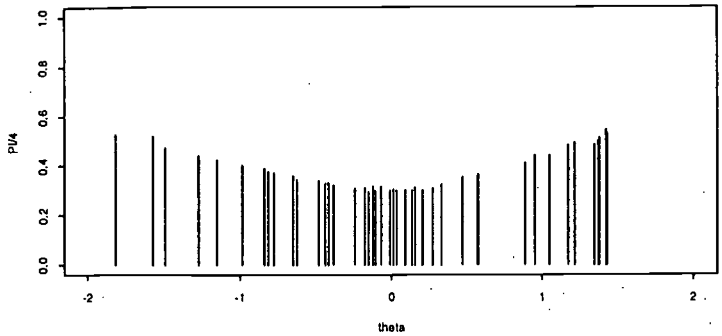

Gibbs Sampling Prior-1
Gibbs Sampling Prior-2
Marginal Bayesian Prior-1
Marginal Bayesian Prior-2
BILOG Prior-1*
BILOG Prior-2*
lambda
zeta
0.0
0.5
1.0
1.5
2.0
-2
-1
0
1
2


Gibbs Sampling Prior-1
PI/4
lambda
Gibbs Sampling Prior-1
PI/4
zeta
Gibbs Sampling Prior-2
PI/4
lambda
Gibbs Sampling Prior-2
PI/4
zeta
Gibbs Sampling Prior-1
PSD
lambda
Gibbs Sampling Prior-1
PSD
zeta
Gibbs Sampling Prior-2
PSD
lambda
Gibbs Sampling Prior-2
PSD
zeta
Marginal Bayesian Prior-1
PSD
lambda
Marginal Bayesian Prior-1
PSD
zeta
Marginal Bayesian Prior-2
PSD
lambda
Marginal Bayesian Prior-2
PSD
zeta
BILOG Prior-1
PSD
lambda
BILOG Prior-1
PSD
zeta
BILOG Prior-2
PSD
lambda
BILOG Prior-2
PSD
zeta

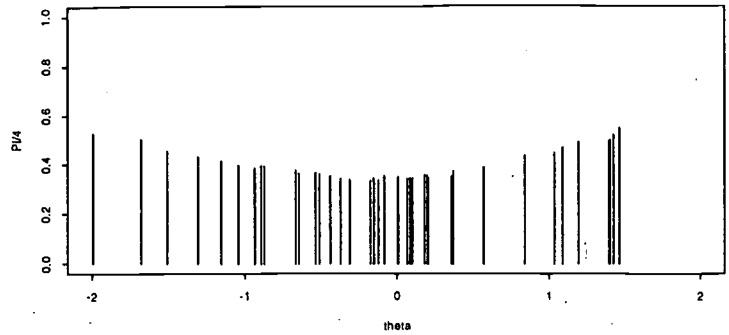


Prior-1
GSE theta
GS theta
Prior-1
MB theta
GS theta
Prior-1*
BILOG theta
GS theta
Prior-2
GSE theta
GS theta
Prior-1
MB theta
GSE theta
Prior-1*
BILOG theta
GSE theta
Prior-2
MB theta
GS theta
Prior-2
MB theta
GSE theta
Prior-1*
BILOG theta
MB theta
Prior-2*
BILOG theta
GS theta
Prior-2*
BILOG theta
GSE theta
Prior-2*
BILOG theta
MB theta

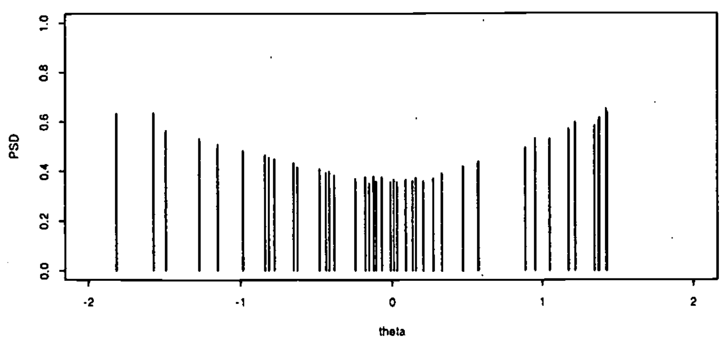
Gibbs Sampling Prior-1



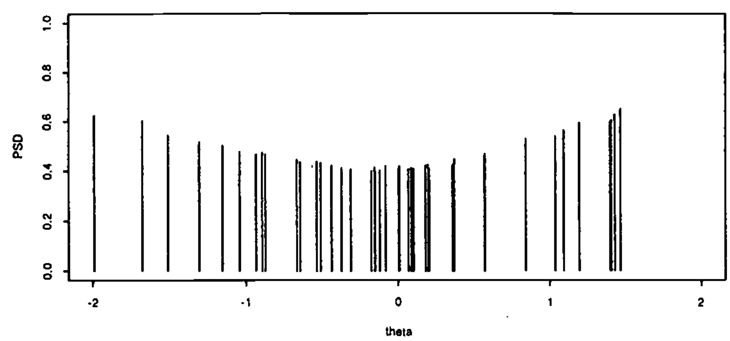
Gibbs Sampling Prior-2



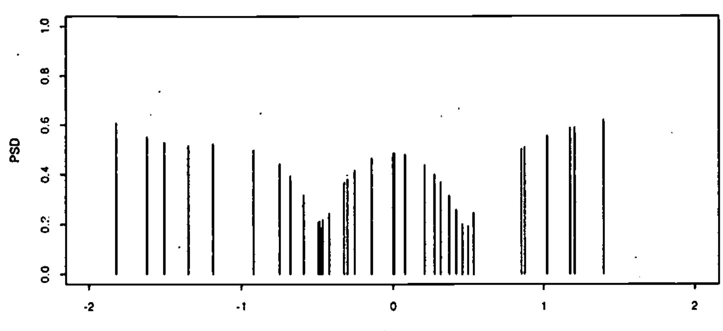
Gibbs Sampling Prior-1



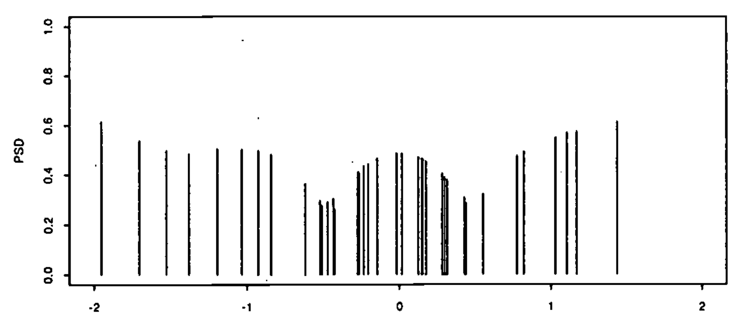
Gibbs Sampling Prior-2



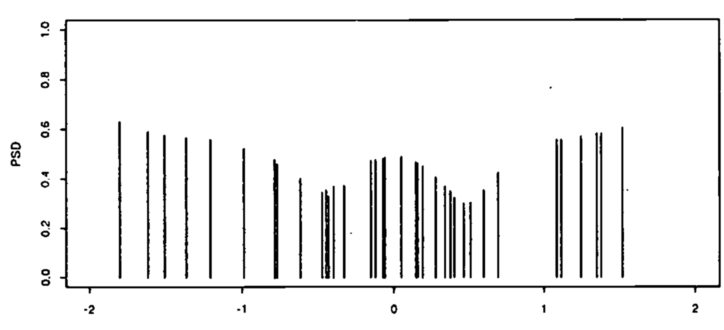
Gibbs Sampling/EAP Prior-1



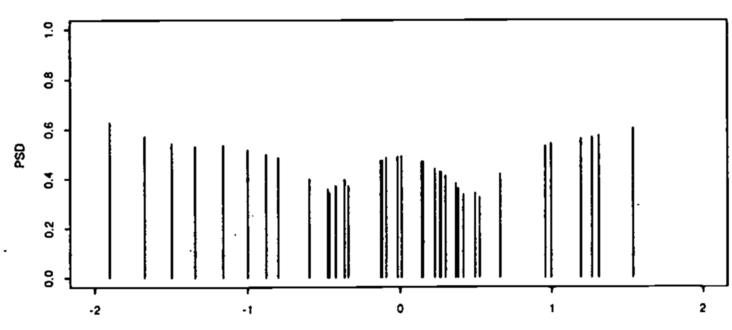
Gibbs Sampling/EAP Prior-2



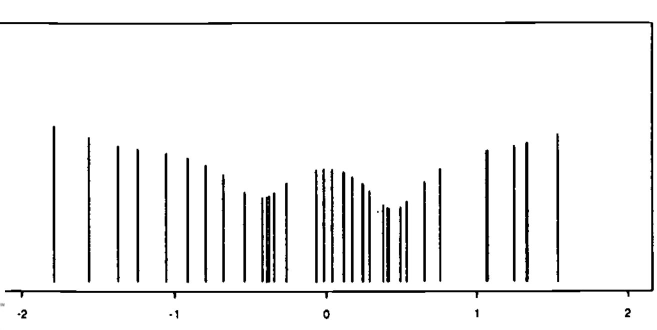
Marginal Bayesian Prior-1



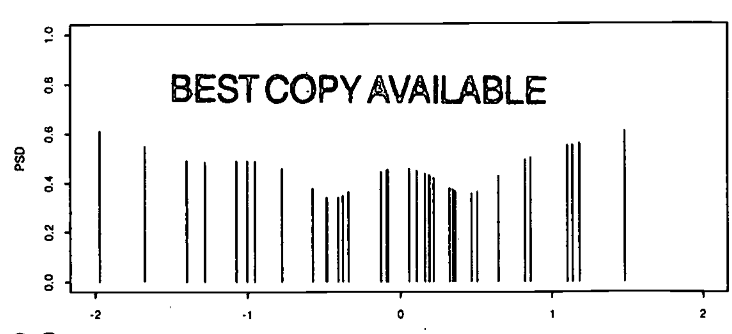
Marginal Bayesian Prior-2



BILOG Prior-1



BILOG Prior-2



Authors's Addresses

Send requests for copies or further information to Seock-Ho Kim, The University of Georgia, 325 Aderhold Hall, Athens, GA 30602-7143 (skim@coe.uga.edu), or Allan S. Cohen, University of Wisconsin–Madison, 1025 W. Johnson, Madison, WI 53706 (ascohen@facstaff.wisc.edu).

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: An Investigation of Ability Estimation in Gibbs Sampling

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

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

Publication Date: 2000

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

Telephone: (706) 542-4224

FAX: (706) 542-4240

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

Date: 4/12/00

(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:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20772
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: