Bias and the Effect of Priors in Bayesian Estimation of Parameters of Item Response Models

Janice A. Gifford
Mount Holyoke College

Hariharan Swaminathan
University of Massachusetts

The effectiveness of a Bayesian approach to the estimation problem in item response models has been sufficiently documented in recent years. Although research has indicated that Bayesian estimates, in general, are more accurate than joint maximum likelihood (JML) estimates, the effect of choice of priors on the Bayesian estimates is not well known. Moreover, the extent to which the Bayesian estimates are biased in comparison with JML estimates is not known. The effect of priors and the amount of bias in Bayesian estimates is examined in this paper through simulation studies. It is shown that different specifications of prior information have relatively modest effects on the Bayesian estimates. For small samples, it is shown that the Bayesian estimates are less biased than their JML counterparts. *Index terms: accuracy, Bayesian estimates, bias, item response models, joint maximum likelihood estimates, priors.*

Item response theory has the potential for significantly improving educational and psychological measurement. However, the successful application of this theory relies heavily on the existence of adequate estimation techniques. The effectiveness of a Bayesian approach to the estimation problem in the one-, two-, and three-parameter item response models has been demonstrated (Mislevy, 1986; Rigdon & Tsutakawa, 1983; Swaminathan & Gifford, 1982, 1985, 1986; Tsutakawa, 1984; Tsutakawa & Lin, 1986).

Swaminathan and Gifford demonstrated the feasibility of a joint Bayesian estimation procedure in item response models. They developed a hierarchical approach suggested by Lindley (1971) and showed that the Bayesian procedure produced more accurate estimates with short tests and small examinee samples than did the joint maximum likelihood (JML) procedures implemented by LOGIST IV (Wood, Wingersky, & Lord, 1976). However, because the Bayesian approach to estimation relies on the specification of priors, it is of interest to know how the various choices for the prior distributions placed on parameters affect the accuracy of estimation. A simulation study was implemented to study the effect of priors in the one-, two-, and three-parameter logistic models.

Although Swaminathan and Gifford showed that Bayesian estimates are more accurate than JML estimates, the Bayesian estimates theoretically would be expected to be biased. Although the extent of bias in JML estimators of item parameters is known when examinee abilities are known (Lord, 1983), the extent to which the JML estimators are biased is unknown. Hence it is of interest to further investigate the error of estimation and establish the relative sizes of bias and sampling error. A second simulation study was therefore conducted to investigate the bias in the Bayesian and JML estimates.

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 14, No. 1, March 1990, pp. 33–43

0146-6216/90/010033-11$1.80

The Model

For the three-parameter logistic model, the probability that examinee i responds correctly to item g is given as

$$P_g(\theta_i) = c_g + \frac{1-c_g}{1+\exp[-1.7a_g(\theta_i - b_g)]} \quad , \tag{1}$$

where θ_i is the ability level of examinee i,
b_g is the difficulty level of item g,
a_g is the discrimination of item g, and
c_g is the lower asymptote for item g.

The two-parameter logistic model is obtained by setting c equal to 0 for all items, while the additional restriction of fixing all $a_g = 1$ results in the one-parameter model. To obtain Bayesian estimates of the parameters of logistic models, it is necessary to specify the priors on the parameters. Following the procedures developed by Swaminathan and Gifford (1982, 1985, 1986), the priors for θ_i and b_g were specified in two stages. In the first stage, it is assumed that θ_i (and b_g) are independently and identically normally distributed:

$$\theta_i|\mu_\theta,\phi_\theta \sim \mathrm{N}(\mu_\theta,\phi_\theta) \quad , \tag{2}$$

$$b_g|\mu_b,\phi_b \sim \mathrm{N}(\mu_b,\phi_b) \quad . \tag{3}$$

In the second stage, the information on μ_θ, ϕ_θ, μ_b, and ϕ_b is specified. In setting the priors for μ_b and ϕ_b, it is assumed a priori that μ_b is distributed uniformly, while ϕ_b has the form of an inverse chi-square distribution with parameters ν_b and λ_b. That is,

$$\begin{aligned} f(\mu_b,\phi_b|\nu_b,\lambda_b) &= f(\mu_b)\, f(\phi_b|\nu_b,\lambda_b) \\ &\propto f(\phi_b|\nu_b,\lambda_b) \\ &\propto \phi_b^{-(1/2\nu_b+1)} \exp(-\lambda_b/2\phi_b) \quad . \end{aligned} \tag{4}$$

The same procedure was used in setting the priors for μ_θ and ϕ_θ.

The prior for a_g was taken as the chi distribution, that is,

$$f(a_g|\nu_g,\omega_g) \propto a_g^{\nu_g-1} \exp(-a_g^2/2\omega_g) \quad . \tag{5}$$

The prior distribution for c_g was taken as the beta distribution:

$$f(c_g|s_g,t_g) \propto c_g^{s_g}(1-c_g)^{t_g} \quad . \tag{6}$$

Justifications for these choices of prior distributions are given in Swaminathan and Gifford (1982, 1985, 1986) and hence will not be repeated here. The joint posterior probability, after integration with respect to the nuisance parameters μ and ϕ, is given by Swaminathan and Gifford (1986).

Study 1: The Effect of Priors on Accuracy

The Bayesian approach has often been criticized because of the subjective nature of the specification of priors. An important issue to investigate, therefore, is the susceptibility of the Bayesian procedure to changes in the specification of the prior distribution for different sample sizes and test lengths. Do large fluctuations in estimates occur when small changes in priors are made, or is the procedure reasonably robust or stable with respect to these changes?

The One-Parameter Model

To investigate the effect of priors in the one-parameter model, test lengths of 15, 25, and 50 items were completely crossed with sample sizes of 25, 50, 150, and 500 examinees. The true θs and item difficulties were distributed uniformly in the interval $[-1.73,1.73]$. These distributions were chosen to differ from the distributions selected for priors. Note that as tests were lengthened, the original items were retained and combined with newly selected items. The same holds true for examinees. That is, as sample size was increased, it was as if the same examinees, along with additional examinees, were administered the item set. This provided for more clarity of interpretation of trends, because the variability due to sampling from the population of true values was minimized. Data were generated through the use of a modification of the DATAGEN program (Hambleton & Rovinelli, 1973).

The $\chi^{-2}(\nu,\lambda)$ prior was placed on both ϕ_b and ϕ_θ. To obtain a range of priors, the scale parameter λ was fixed at 10.0 while the degrees of freedom ν were set at 5, 8, 15, 25, and 50. (See Swaminathan & Gifford, 1985, for justification of the choice of the value for λ.) Table 1 contains a description of the nature of the priors. When ν is small, the distribution is skewed with extremely large variance. As ν increases, the χ^{-2} approaches

Table 1
Characteristics of the Inverse Chi-Square Distribution for $\lambda = 10$ and Selected Values of ν

ν	Mode	Median	Mean	Variance	99% Credibility Interval
5	1.43	2.30	3.33	22.178	----
8	1.00	1.36	1.67	1.394	.32 to 6.10
15	.59	.70	.77	.108	.25 to 1.93
25	.37	.41	.43	.018	.19 to .88
50	.19	.20	.21	.002	.11 to .32

normality and the credibility interval becomes smaller.

Estimates and true values were scaled by standardizing on θ. Accuracy of estimation was examined in two ways. First, the correlations of the estimates with the true values were calculated for each testing situation. Second, the squared differences between each estimate and true value were calculated. These were averaged across all items (or θs) to yield an index for accuracy, denoted mean squared difference (MSD). Because MSD is in squared units, its square root, RMSD, was used for clarity of interpretation.

The results of the effect of varying ν are presented in Table 2. A single correlation is reported for each test situation because the correlations were virtually unaffected by the choice of ν. However, RMSD steadily increases as ν increases. This occurs because as ν increases, the distribution of ϕ becomes concentrated, reflecting increasingly stronger beliefs about the value of ϕ. Furthermore, the value of ϕ itself decreases as ν increases. Consequently, with large ν, firm beliefs about small variances for the θ and b parameters are expressed. These result in greater regression toward the mean, hence larger RMSD.

To prevent extreme biasing when $\lambda = 10$, relatively small values of ν must be specified. These values of ν and λ result in a large value of ϕ together with a diffuse belief about its value. Clearly, for values of ν between 5 and 15, the prior distributions produce similar, if not identical, results.

Table 2
RMSDs and Correlation (ρ) of Estimated and True Parameter Values for Prior Distributions With $\nu = 5$, 8, 15, 25, and 50 for Tests of 15, 25, and 50 Items and Sample Sizes of 25, 50, 150, and 500 Examinees

	15 Items				25 Items				50 Items			
Statistic	25	50	150	500	25	50	150	500	25	50	150	500
Difficulty												
RMSD												
$\nu = 5$	.28	.27	.15	.11	.27	.21	.13	.08	.26	.22	.15	.08
$\nu = 8$	.29	.27	.15	.11	.28	.21	.13	.08	.26	.22	.15	.08
$\nu = 15$	.30	.28	.15	.12	.30	.22	.14	.08	.27	.22	.15	.08
$\nu = 25$	.33	.29	.16	.12	.33	.23	.15	.08	.28	.23	.15	.08
$\nu = 50$	.40	.34	.19	.14	.40	.28	.17	.09	.34	.26	.16	.08
ρ	.944	.952	.987	.997	.956	.970	.993	.997	.966	.975	.990	.998
Ability (θ)												
RMSD												
$\nu = 5$	.42	.39	.39	.39	.26	.28	.30	.30	.21	.20	.22	.24
$\nu = 8$	.43	.40	.39	.39	.26	.28	.30	.30	.21	.20	.22	.24
$\nu = 15$	.46	.42	.40	.39	.29	.28	.30	.30	.22	.20	.23	.24
$\nu = 25$	.50	.45	.42	.40	.32	.29	.30	.30	.24	.21	.23	.24
$\nu = 50$	.58	.53	.46	.41	.40	.33	.32	.30	.29	.23	.23	.24
ρ	.918	.929	.941	.935	.972	.957	.958	.958	.978	.980	.975	.973

The accuracy of estimation of b and θ parameters does not seem to be affected by such values of ν. Only large values of ν (>15) seem to have a detrimental effect on estimation.

The effects of sample size and test length on RMSD are clear. RMSD decreases as sample size increases and as test length increases. These trends indicate that as the amount of information obtained from data increases, the accuracy of estimation increases.

The Two-Parameter Model

For the two-parameter model, because the effect of prior specification on the a parameter was of primary concern, noninformative priors were placed on the b and θ parameters. The chi distribution was chosen to indicate prior belief on a, and priors were selected through the use of the normal approximation to the chi distribution (Swaminathan & Gifford, 1985). Identical priors were placed on each a_g. Table 3 contains descriptive information about the various chi distributions chosen for study.

The effects of the 12 prior distributions on the estimates were compared through analyses of two testing situations: (1) test length $n = 25$, sample size $N = 100$ and (2) $n = 35$, $N = 200$. For data generation, the bs and θs were distributed normally with mean 0 and variance 1. These distributions were chosen to differ from the distributions selected for priors. The a_g were distributed uniformly in the interval [.6,1.9].

Results are presented in Table 4. Certain specifications of prior information on the a parameter resulted in nonconvergence of the numerical procedure. This nonconvergence occurred when extreme values for the priors were specified, that is, when the mean was set at a large value together with a small standard deviation. This does not seem surprising in light of the fact that the true a_g were generated from a uniform distribution in the interval [.6,1.9]. Although the problem of nonconvergence occurred frequently for $n = 25$ and $N = 100$, it was less of a problem in the second situation, when $n = 35$ and $N = 200$.

For the cases where convergence occurred, the correlations between true and estimated values for θ and b were unaffected by the specification of prior information. The correlations between estimated and true values for the a parameter, however, were affected by the specification of the prior distribution on the a_g. A similar trend was observed for RMSDs. In general, the priors with $\mu = 1.0$ or 1.5 and $\sigma = .25$ or .50 showed the best results. This result indicates that values for μ and σ that reflect the distribution of a_g reasonably well result

Table 3
Characteristics of the Chi Distribution for Selected Parameters

Mean	SD	ν	ω	99% Credibility Interval
1.0	1.00	1.000	2.000	.00 to 3.97
	.50	2.500	.500	.14 to 2.42
	.25	8.500	.125	.44 to 1.69
1.5	1.00	1.625	2.000	.00 to 4.38
	.50	5.000	.500	.45 to 2.89
	.25	18.500	.125	.90 to 2.18
2.0	1.00	2.500	2.000	.29 to 4.84
	.50	8.500	.500	.88 to 3.37
	.25	32.500	.125	1.39 to 2.67
2.5	1.00	3.625	2.000	.56 to 5.31
	.50	13.000	.500	1.34 to 3.86
	.25	50.500	.125	1.88 to 3.16

Table 4
Effect of Prior Distribution for Values of $\mu = [\omega(\nu-.5)]^{1/2}$ and $\sigma = (\omega/2)^{1/2}$ in the Two-Parameter Model for 25 Items and 100 Examinees and for 35 Items and 200 Examinees

		Difficulty		Discrimination		θ	
μ	σ	RMSD	ρ	RMSD	ρ	RMSD	ρ
25 Items and 100 Examinees							
1.0	1.00	.18	.976	.41	.777	.30	.955
	.50	.14	.986	.26	.807	.29	.957
	.25	.12	.991	.29	.813	.30	.955
1.5	1.00	.18	.977	.44	.772	.30	.955
	.50	.14	.987	.27	.799	.29	.956
	.25	*		*		*	
2.0	1.00	.18	.977	.47	.764	.30	.955
	.50	.13	.987	.32	.788	.29	.956
	.25	*		*		*	
2.5	1.00	.17	.978	.52	.755	.30	.955
	.50	*		*		*	
	.25	*		*		*	
35 Items and 200 Examinees							
1.0	1.00	.15	.985	.30	.880	.26	.965
	.50	.14	.988	.20	.904	.26	.966
	.25	.13	.991	.19	.914	.26	.967
1.5	1.00	.15	.985	.31	.878	.26	.966
	.50	.14	.988	.23	.903	.26	.967
	.25	.12	.991	.21	.912	.25	.967
2.0	1.00	.15	.985	.33	.877	.26	.966
	.50	.13	.989	.28	.902	.26	.967
	.25	.12	.991	.33	.909	.26	.967
2.5	1.00	.15	.986	.36	.874	.26	.966
	.50	.13	.990	.37	.900	.26	.967
	.25	*		*		*	

*No convergence.

in the most accurate estimation. The effects of the prior distribution on θ and b estimates are negligible for these values of μ and σ.

The Three-Parameter Model

To investigate the effect of specification of prior distributions on the c parameter, a single testing situation was selected: $n = 35$, $N = 200$. Values of θ and b were drawn from uniform distributions in the interval $[-1.73, 1.73]$. This was done to ensure an adequate number of low-θ examinees which, in turn, ensures reasonable estimates of the c parameter. To ensure that the priors were not consistent with the true distribution of parameters, a_g and c_g were selected from uniform populations in the intervals $[.6, 1.9]$ and $[.00, .22]$, respectively.

Noninformative priors were placed on both θ and b parameters. Identical chi priors were placed on each a_g. Because the specification $\mu = 1.5$ and $\sigma = .5$ produced consistently good results for the two-parameter model, it was employed here. To investigate the effect of varying the priors on the c_g, nine beta distributions were selected in the following manner. Because the c_g were drawn from a uniform distribution in the interval $[.00, .22]$, three modal values—.06, .11, and .16—were chosen to evenly span the interval. These were crossed with three levels of dispersion. The widths of the 99% credibility intervals were chosen to be .12, .15, and .22. These represent varying strengths in belief about the value of c_g. The descriptive information about the beta distributions for selected values of s and t is presented in Table 5.

The results are summarized in Table 6. Values of b and θ show very little change as the priors on c_g are varied, with the exception of the analysis with $s = 1.0$, $t = 15.7$ (which resulted in an inappropriate solution). For b, the correlations maintain the value of .986, while RMSDs are in the range of .21 to .23. For θ, correlations vary from .953 to .956 while RMSDs stay the same.

The estimation of the a parameter seems to be affected to a greater extent by the specification of priors on c_g. The worst estimation occurs for the highest mode (.16) and strongest prior (width = .12). Here the correlation and RMSD are .654 and .33 respectively. The best estimation of the a parameter occurs when a more diffuse prior (width = .15) is specified for c_g with a mode of .06. This results in a correlation of .698 and RMSD of .26.

The choice of priors has a discernible effect on the estimates of c_g. The priors with the mode of .11 result in better estimates. This could be expected because the distribution of true c_g goes from .00 to .22. The most accurate guess as to a typical value for c in the absence of any other information, on the average, could be obtained through the choice of a value in the middle of the range.

Study 2: Bias

Swaminathan and Gifford (1982, 1985, 1986) demonstrated that the Bayesian procedure consistently produced more accurate estimates than did the JML procedure as implemented by LOGIST IV (Wood et al., 1976). As before, accuracy was defined as the mean of the squared differences between estimates and true values. Although this index represents the overall accuracy of the estimation procedure, it does not provide any explanation of the nature of the differences between estimates and true values. The difference could be attributable to sampling error or to systematic bias.

In order to separate the error into these two components, a single test situation was selected and replicated for each of the three models. The replications were generated as follows: For a given testing situation, true values for item and θ parameters were generated according to specified distributions; these true values were held fixed and 20 item response data matrices were generated randomly. This simulated the responses to an item by an examinee with a given level of θ over a set of independent replications.

For any parameter τ, let m_k be the estimate for replication k. Then accuracy can be measured in terms of the discrepancy $(m_k - \tau)$. Now

$$(m_k-\tau) = (m_k-m.)+(m.-\tau) \quad , \tag{7}$$

Table 5
Characteristics of the Beta Distribution for Selected Parameters

s	t	Mode	Median	Mean	99% Credibility Interval	Width
1.0	15.7	.06	.093	.107	.021 to .242	.22
2.0	31.3	.06	.077	.085	.024 to .172	.15
3.0	47.0	.06	.072	.077	.027 to .145	.12
2.5	20.2	.11	.132	.142	.047 to .269	.22
5.5	44.5	.11	.120	.125	.059 to .207	.15
8.5	68.8	.11	.117	.120	.066 to .185	.12
4.5	23.6	.16	.176	.183	.082 to .307	.22
10.5	55.1	.16	.167	.170	.101 to .250	.15
16.5	86.6	.16	.164	.167	.111 to .229	.12

Table 6
Effect of Prior Distribution for Values of s and t in the Three-Parameter Model for 35 Items and 200 Examinees

				Difficulty		θ		Discrimination		Chance Level	
s	t	Mode	Width	ρ	RMSD	ρ	RMSD	ρ	RMSD	ρ	RMSD
1.0	15.7	.06	.22	.962	.37	.931	.37	.249	.44	*	*
2.0	31.3	.06	.15	.986	.23	.955	.30	.698	.26	.465	.073
3.0	47.0	.06	.12	.986	.23	.954	.30	.683	.26	.442	.075
2.5	20.2	.11	.22	.986	.21	.956	.30	.683	.29	.565	.053
5.5	44.5	.11	.15	.986	.21	.955	.30	.677	.29	.558	.054
8.5	68.8	.11	.12	.986	.21	.955	.30	.666	.29	.540	.056
4.5	23.6	.16	.22	.985	.22	.954	.30	.654	.33	.540	.062
10.5	55.1	.16	.15	.986	.21	.954	.30	.659	.33	.550	.062
16.5	86.6	.16	.12	.986	.21	.953	.30	.654	.33	.536	.065

*All cs were estimated to be 0.0.

where $m.$ represents the mean estimate of τ over the r replications. It follows that

$$\sum_{k=1}^{r}(m_k-\tau)^2 = r(m.-\tau)^2+\sum_{k=1}^{r}(m_k-m.)^2 \quad , \tag{8}$$

or equivalently,

$$\sum_{k=1}^{r}\frac{(m_k-\tau)^2}{r} = (m.-\tau)^2+\sum_{k=1}^{r}\frac{(m_k-m.)^2}{r} \quad . \tag{9}$$

This relationship demonstrates that the MSD across r replications is separable into two components. One component, the variance of the estimates, is given as

$$V(m) = \sum_{k=1}^{r}\frac{(m_k-m.)^2}{r} \tag{10}$$

while the second, bias, is given as

$$B(m) = (m.-\tau)^2 \quad . \tag{11}$$

For each parameter type in each model, these quantities were calculated for each item and each examinee. In order to summarize this information, MSD, $V(m)$, and $B(m)$ were averaged across items (or examinees) to give overall indices.

To further investigate the bias quantity $B(m)$, for each item (or θ) the mean of the estimates obtained from the 20 replications was calculated. The distribution of these means for the test (or examinee sample) enabled comparison with the distribution of true values. The distributions were compared with respect to the first four moments and the range.

The One-Parameter Model

A testing situation of $n = 25$, $N = 100$ was selected to be replicated for the investigation of bias. The true item and θ parameters were generated from a uniform distribution in the interval $[-1.73, 1.73]$. This ensured that the distributions had means of 0 and standard deviations of 1.

For the Bayesian estimation, χ^{-2} priors with $\nu = 8$, $\lambda = 10$ were placed on the variances of θ_i and b_g. In Table 7 a description of the distribution of true values is presented alongside the distribution of the mean of the estimates (across replications). In general, the two estimation procedures reproduce the true distribution reasonably well. The distributions of Bayesian estimates for b and θ have smaller standard deviations than the true distributions, while JML estimates have larger standard deviations.

From Table 8 it can be seen that the two procedures clearly differ in the decomposition of error into the two components. Although the Bayesian estimates have smaller MSD and smaller variance of the estimates, they tend to be more biased than the JML estimates. The bias is more evident with the θ estimates than with the b estimates.

Table 7
Distributions of True Values and Means of the Bayesian and ML Estimates in the One-Parameter Model

	Difficulty			θ		
Statistic	True	Bayesian	ML	True	Bayesian	ML
Mean	−.384	−.353	−.407	.000	.000	.000
SD	.851	.770	.910	1.012	.850	1.066
Skewness	.496	.491	.484	−.219	−.322	−.256
Kurtosis	−.655	−.566	−.576	−1.287	−1.338	−1.273
Minimum	−1.756	−1.657	−1.954	−1.757	−1.538	−1.966
Maximum	1.284	1.202	1.427	1.647	1.330	1.830

The Two-Parameter Model

To investigate bias in the two-parameter model, a testing situation of $n = 25, N = 200$ was selected for the 20 replications. The true values for b and θ parameters were drawn from normal distributions with means of 0 and standard deviations of 1, while the a parameters were drawn from a uniform distribution in the interval [.6,1.9]. For Bayesian estimation, uniform priors were placed on θ_i and b_g while a chi prior with $\mu = 1.5$ and $\sigma = .5$ was chosen for each a parameter. The resulting distribution of the mean of the estimates (across the 20 replications) is compared to the distribution of true values in Table 9.

As in the one-parameter model, only slight differences occur with respect to b and θ. With respect to a, however, the Bayesian procedure clearly reproduces the original distribution more closely than does JML.

The error components that combine to form the MSD are presented in Table 10. Again there are virtually no differences between the JML and Bayesian estimates for b and θ. It should be noted that the priors on b and θ were chosen to be uniform, hence differences would not be expected to occur. On the other hand, for the a parameters, the Bayesian procedure produces dramatically smaller error components than the JML procedure. The priors arrest the outward drift of the estimates and hence the resulting estimates have clearly less variance and bias.

The Three-Parameter Model

A test situation of $n = 35, N = 200$ was replicated in order to examine the bias in the three-parameter model. As in Study 1, all true values were drawn from uniform distributions. The a parameters were generated in the interval [.6,1.9]; the chance-level parameters were in the interval [.00,.22]; θ and b parameters were in the interval [−1.73,1.73]. For Bayesian estimation, noninformative priors were placed on θ and b. The chi prior of $\mu = 1.5$, $\sigma = .5$ that was used previously was

Table 8
Error Components in the Estimates of the One-Parameter Model Based on 20 Replications ($n = 25$, $N = 100$)

Parameter	Estimate	MSD	V(m)	B(m)
Difficulty	Bayesian	.032	.022	.009
	ML	.038	.032	.006
θ	Bayesian	.100	.067	.032
	ML	.130	.121	.009

Table 9
Distributions of True Values and Means of the Bayesian and ML Estimates in the Two-Parameter Model

Statistic	Difficulty True	Difficulty Bayesian	Difficulty ML	Discrimination True	Discrimination Bayesian	Discrimination ML	θ True	θ Bayesian	θ ML
Mean	.073	.067	.065	1.399	1.547	1.937	.000	.000	.000
SD	.848	.831	.847	.405	.443	1.122	1.000	.967	.961
Skew	.088	.097	.047	-.473	-.482	1.830	-.086	-.076	-.111
Kurt	-.814	-.658	-.638	-1.047	-.894	4.599	-1.263	-1.254	-1.256
Min	-1.391	-1.399	-1.486	.642	.665	.581	-1.748	-1.693	-1.717
Max	1.631	1.683	1.685	1.918	2.132	5.686	1.631	1.605	1.583

again chosen for this study. The parameters chosen to define the beta prior for the c_g were $s = 3$ and $t = 22$.

Table 11 contains the comparison of the distribution of true values to the distribution of the estimates averaged across replications. As was the case in both the one- and two-parameter models, there are only slight differences between the distributions of JML and Bayesian estimates for b and θ. These parameters seem to be estimated with stability for both procedures.

As in the two-parameter model, the a parameter is recovered much better with the Bayesian procedure than with JML. The distribution of JML mean estimates has a standard deviation of 1.146 while the true standard deviation was .347. This is due to the tendency of JML estimates of a to drift upward.

With respect to the c parameter, both estimation procedures produce distributions that are tighter than the true distribution. The Bayesian and JML procedures result in standard deviations of .027 and .024, respectively, while the true standard deviation was .065. This is also demonstrated by the ranges. The JML estimates are in the interval [.045,.154], Bayesian estimates are in the interval [.064,.184], and the true values are in the interval [.014,.217].

The tight distribution for JML estimates is a result of the very controlled estimation procedure of LOGIST IV. Most analyses resulting from LOGIST yield the majority of the estimates of c placed at a single value, with a few c values falling above or below the common value. Although the distribution of Bayesian estimates is equally tight, the mean of the distribution is closer to the true value than is the JML mean.

The information pertaining to the error components is presented in Table 12. Over all entries in the table, the Bayesian values are smaller than the JML entries. As expected, the JML procedure results in larger $V(m)$ for all parameters (except for the c_g, where the procedures produce equivalent results). In addition the JML estimates are consistently

Table 10
Error Components in the Estimates of the Two-Parameter Model Based on 20 Replications
($n = 25$, $N = 200$)

Parameter	Estimate	MSD	V(m)	B(m)
Difficulty	Bayesian	.014	.013	.001
	ML	.020	.018	.002
Discrimination	Bayesian	.088	.059	.029
	ML	3.066	2.095	.971
θ	Bayesian	.070	.065	.004
	ML	.081	.076	.005

Table 11
Distributions of True Values and Means of the Bayesian and ML Estimates in the Three-Parameter Model

Statistic	Difficulty True	Difficulty Bayesian	Difficulty ML	θ True	θ Bayesian	θ ML	Discrimination True	Discrimination Bayesian	Discrimination ML	Chance Level True	Chance Level Bayesian	Chance Level ML
Mean	−.184	−.183	−.207	.000	.000	−.001	1.168	1.367	2.094	.117	.122	.099
SD	1.218	1.204	1.268	1.000	.952	.947	.347	.306	1.146	.065	.027	.024
Skewness	−.147	−.279	−.182	−.104	−.181	−.110	−.073	−.496	.608	.060	−.019	.385
Kurtosis	−1.522	−1.403	−1.324	−1.170	−1.070	−1.222	−1.276	−1.024	−.409	−1.437	.509	1.556
Minimum	−2.006	−2.183	−2.261	−1.865	−1.897	−1.796	.636	.794	.585	.014	.064	.045
Maximum	1.538	1.500	1.977	1.595	1.550	1.559	1.708	1.849	4.787	.217	.184	.154

more biased, as indicated by $B(m)$, than the Bayesian estimates in the three-parameter model.

Conclusions

The results of the simulation studies reported here show that different specifications of prior distributions have relatively modest effects on the Bayesian estimates. Exceptions to this finding occur with distributions that are "extreme" in nature. Prior information that reflects extreme values of parameters often results in nonconvergence of the numerical procedure. Vague or diffuse priors seem to provide less regressed estimates and improved correlations between true values and estimates.

Note that identical prior distributions were used for each a parameter and for each chance-level parameter. This approach was chosen to represent the situation where very little knowledge is available about individual items. Prior distributions that are not too tight are clearly preferable in these circumstances. However, in the event that information is available for individual items, different priors can be selected for each item and these can afford to be specific. Further improvement in estimation can be expected to occur with such specific priors.

Bias was present in estimates of b and θ in the one-parameter model, with the Bayesian estimates showing slightly more bias than the JML estimates. In the two- and three-parameter models the Bayesian estimates had very little bias, even less than the JML estimates. This is because in the two- and three-parameter models, noninformative priors were placed on the θ and b parameters, whereas in the one-parameter model informative priors are used. Consequently, the estimates in the one-parameter model were regressed and showed more bias.

Although bias was present for both Bayesian and JML estimates of a and chance-level parameters, the Bayesian procedure was shown to be less biased throughout the study, even though the priors chosen for a and c clearly differed from the generating distributions. This could be a result of the influence

Table 12
Error Components in the Estimates of the Three-Parameter Model Based on 20 Replications ($n = 35$, $N = 200$)

Parameter	Estimate	MSD	$V(m)$	$B(m)$
Difficulty	Bayesian	.044	.032	.012
	ML	.172	.144	.027
Discrimination	Bayesian	.139	.077	.063
	ML	7.250	5.663	1.587
Chance Level	Bayesian	.003	.001	.002
	ML	.004	.001	.003
θ	Bayesian	.100	.093	.007
	ML	.123	.115	.008

of highly unstable estimates of the *a* parameter on the estimates of other parameters in the JML procedure. Thus it appears that using prior information that is not very specific improves the quality of estimation.

In general, it appears that a Bayesian procedure that places priors on the *a* and *c* parameters improves the accuracy of estimation of parameters in item response models. Surprisingly, the Bayesian estimates are less biased than the JML estimates. Different choices for parameters of the prior distributions do not have any marked effect on the estimation as long as the prior distributions are not too extreme.

References

Hambleton, R. K., & Rovinelli, R. (1973). A FORTRAN IV program for generating examinee response data from logistic test models. *Behavioral Science, 18,* 74.

Lindley, D. V. (1971). The estimation of many parameters. In V. P. Godambe & D. A. Sprott (Eds.), *Foundations of statistical inference* (pp. 435–455). Toronto: Holt, Rinehart and Winston.

Lord, F. M. (1983). Statistical bias in maximum likelihood estimation of item parameters. *Psychometrika, 48,* 425–436.

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51,* 177–196.

Rigdon, S. E., & Tsutakawa, R. K. (1983). Estimation in latent trait models. *Psychometrika, 48,* 567–574.

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,* 589–601.

Tsutakawa, R. K. (1984). Estimation of two-parameter logistic item response curves. *Journal of Educational Statistics, 9,* 263–276.

Tsutakawa, R. K., & Lin, H. Y. (1986). Bayesian estimation of item response curves. *Psychometrika, 51,* 251–268.

Wood, R. L., Wingersky, M. S., & Lord, F. M. (1976). *A computer program for estimating examinee ability and item characteristic curve parameters* (Research Memorandum 76-6). Princeton NJ: Educational Testing Service. (Revised 1978)

Acknowledgments

This research was partially supported by grants from the Office of Naval Research (Contract No. N0014-79-C-0039) and National Evaluation Systems, Amherst MA.

Author's Address

Send requests for reprints or further information to Janice A. Gifford, Department of Mathematics, Statistics, and Computation, Mount Holyoke College, South Hadley MA 01075, U.S.A.