PSYCHOMETRIKA—VOL. 51, NO. 4, 589–601
DECEMBER 1986

BAYESIAN ESTIMATION IN THE THREE-PARAMETER LOGISTIC MODEL

HARIHARAN SWAMINATHAN

UNIVERSITY OF MASSACHUSETTS

JANICE A. GIFFORD

MOUNT HOLYOKE COLLEGE

A joint Bayesian estimation procedure for the estimation of parameters in the three-parameter logistic model is developed in this paper. Procedures for specifying prior beliefs for the parameters are given. It is shown through simulation studies that the Bayesian procedure (i) ensures that the estimates stay in the parameter space, and (ii) produces better estimates than the joint maximum likelihood procedure as judged by such criteria as mean squared differences between estimates and true values.

Key words: Bayes estimation, three-parameter logistic model, item response theory.

Introduction

The three-parameter item response model (Hambleton, Swaminathan, Cook, Eignor, & Gifford, 1978; Lord, 1980) has received considerable attention recently because of its applicability to a variety of testing situations where the one- and two-parameter item response models may not be completely adequate (Loyd & Hoover, 1980; Slinde & Linn, 1979). However, in spite of its flexibility and advantages over the one- and the two-parameter models, the three-parameter model suffers from a serious drawback. The problem of estimating the parameters does not appear to have been completely solved.

The procedures currently available for estimating parameters in the three-parameter logistic model are either based on the heuristic estimation procedure advocated by Urry (1976), or the maximum likelihood procedure described by Lord (1980). The estimation procedure of Urry and the maximum likelihood procedure as implemented in the LOGIST program (Wood, Wingersky, & Lord, 1978) have certain deficiencies. It appears that the discrimination and pseudo-chance level parameters are not estimated well unless limits are imposed on the range of values taken by the estimates and to some degree this limitation extends to the estimation of difficulty and ability parameters. The problems with the estimation of the discrimination and chance-level parameters can be partially overcome if marginal maximum likelihood estimators of the parameters are employed (Bock & Aitkin, 1981). However, the marginal maximum likelihood procedure does not ensure that estimates remain in the parameter space.

Bayesian procedures for estimating parameters have been successfully applied in numerous situations. In particular, the approach suggested by Lindley and Smith (1972) has been successfully applied to estimate simultaneously the item and ability parameters in

The research reported here was performed pursuant to Grant No. N0014-79-C-0039 with the Office of Naval Research.

A related article by Robert J. Mislevy (1986) appeared when the present paper was in the printing stage.

Requests for reprints should be sent to H. Swaminathan, School of Education, University of Massachusetts, Amherst, MA 01003.

0033-3123/86/1200-6035$00.75/0

the Rasch model (Swaminathan & Gifford, 1982) and in the two-parameter model (Swaminathan & Gifford, 1985). These authors have found that the Bayesian estimators of the item and ability parameters are superior to the maximum likelihood estimators in that the Bayesian estimators are less biased and exhibit a smaller mean squared difference between the true values and the estimates. In the case of the two-parameter model, Swaminathan and Gifford (1985) have shown that the Bayesian estimates of the discrimination parameter, in contrast to the joint maximum likelihood estimates, do not drift out of bounds and, moreover, have a positive effect on the estimates of the ability and difficulty parameters.

The encouraging results obtained with the Bayesian estimation procedure with the one- and the two-parameter models suggest that the Bayesian approach may provide a meaningful solution to the problem of estimation in the three-parameter logistic model. Hence, a Bayesian procedure based on the hierarchical model suggested by the above authors is developed in this paper.

The Model

Let U_{ij} denote a random variable that represents the binary response of examinee i ($i = 1, \ldots, N$) on item j ($j = 1, \ldots, n$), where for a correct response

$$U_{ij} = 1,$$

while for an incorrect response

$$U_{ij} = 0.$$

Further, assume that the complete latent space is unidimensional and that the probability of a correct response

$$P_{ij} = P[U_{ij} = 1 \mid \theta_i, a_j, b_j, c_j] = c_j + (1 - c_j)\Psi_{ij}, \tag{1}$$

where

$$\Psi_{ij} = [1 + \exp\{-a_j(\theta_i - b_j)\}]^{-1}. \tag{2}$$

Clearly, the probability of an incorrect response is given by

$$Q_{ij} = 1 - P[U_{ij} = 1 \mid \theta_i, a_j, b_j, c_j]. \tag{3}$$

The parameters that define the functions in (1) and (2) are:

θ_i—ability of examinee i,
a_j—discrimination parameter for item j,
b_j—difficulty of item j,
c_j—pseudo-chance level parameter for item j or the lower asymptote of the item characteristic curve, P_{ij}.

It follows from the assumption of local independence that the joint probability of the response vector

$$P[U_{11} \quad U_{12} \quad \cdots \quad U_{Nn} \mid \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}]$$

of the N examinees on n items can be expressed as

$$P[\mathbf{U} \mid \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}] = \prod_{i=1}^{N} \prod_{j=1}^{n} P_{ij}^{U_{ij}} Q_{ij}^{1-U_{ij}}, \tag{4}$$

where

$$\boldsymbol{\theta} = [\theta_1 \quad \theta_2 \quad \cdots \quad \theta_N]$$
$$\mathbf{a} = [a_1 \quad a_2 \quad \cdots \quad a_n]$$
$$\mathbf{b} = [b_1 \quad b_2 \quad \cdots \quad b_n]$$

and

$$\mathbf{c} = [c_1 \quad c_2 \quad \cdots \quad c_n].$$

Once the vector of responses **U** has been observed, the likelihood function is given as:

$$L[\mathbf{u} \mid \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}] = \prod_{i=1}^{N} \prod_{j=1}^{n} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}. \tag{5}$$

Here $\mathbf{u} = [u_{11} \quad u_{12} \quad \cdots \quad \cdots \quad u_{Nn}]$ is the vector of observed responses.

Let the joint prior density of the parameters $\boldsymbol{\theta}$, **a**, **b**, and **c** be denoted as $f(\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c})$. It follows from Bayes' Theorem that the joint density, $f(\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c} \mid \mathbf{u})$, of the parameters given the vector of observations is

$$f(\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c} \mid \mathbf{u}) \propto L(\mathbf{u} \mid \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}) f(\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}).$$

This joint posterior density of the parameters is a revised expression of the belief one has about the parameters once the data have been collected. It contains all the information necessary for making probability statements regarding the parameters of interest.

Prior and Posterior Distribution

In order to define the posterior density or distribution precisely, it is first necessary to specify the prior belief about the parameters. We shall assume that

$$f(\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}) = f(\boldsymbol{\theta}) f(\mathbf{a}) f(\mathbf{b}) f(\mathbf{c}), \tag{6}$$

that is, a priori the parameter vectors $\boldsymbol{\theta}$, **a**, **b**, **c** are independent.

Following Lindley and Smith (1972) and consequently, Swaminathan and Gifford (1982, 1985), we shall specify prior beliefs for the parameter vectors $\boldsymbol{\theta}$ and **b** in two stages. In the first stage, we shall assume that the parameters θ_i are independently and identically distributed, that is, the information on these parameters is exchangeable (Novick, Lewis, & Jackson, 1973) and further that they are normally distributed. We thus have

$$\theta_i \mid \mu_\theta, \sigma_\theta^2 \sim N(\mu_\theta, \sigma_\theta^2). \tag{7}$$

similar iid and normality assumptions are made for the parameters b_j:

$$b_j \mid \mu_b, \sigma_b^2 \sim N(\mu_b, \sigma_b^2). \tag{8}$$

However, as a result of the hierarchical model, this assumption of normality has little effect on the outcomes (Swaminathan & Gifford, 1982).

The second stage of the hierarchical model requires the specification of prior information on μ_θ, σ_θ^2, μ_b and σ_b^2. It should be noted that the model given in (1) is not identified. In the Bayesian analysis, identifying restrictions can be incorporated directly into the prior. We therefore set $\mu_\theta = 0$ and $\sigma_\theta^2 = 1$, so that

$$f(\theta_i \mid \mu_\theta, \sigma_\theta^2) \propto \exp\left(-\tfrac{1}{2}\theta_i^2\right). \tag{9}$$

Priors for μ_b and σ_b^2 are specified by assuming that μ_b is uniform and that σ_b^2 has an inverse chi-square distribution with parameters ν_b and λ_b (Novick & Jackson, 1974, p.

109). The specification of these parameters is described in detail by Swaminathan and Gifford (1985) and hence is not repeated here.

Swaminathan and Gifford (1985) have argued that the discrimination parameter corresponds to the reciprocal of the standard deviation, and therefore that the prior distribution of a_j can be taken to be the chi distribution, that is,

$$f(a_j \mid v_j, \omega_j)da_j \propto a_j^{v_j - 1} \exp\left(\frac{-a_j^2}{2\omega_j}\right)da_j. \tag{10}$$

Here v_j corresponds to the degrees of freedom and ω_j may be interpreted as a scale parameter. We further assume, a priori, that

$$f(a_1, a_2, \ldots, a_n) = f(a_1)f(a_2) \cdots f(a_n). \tag{11}$$

The pseudo-chance level parameter, c_j, is clearly bounded above by one and below by zero. The prior distribution for c_j may be taken as the beta distribution with parameters s_j and t_j. Assuming that $c_1, c_2, \ldots, c_n$ are a priori independent, we have,

$$f(c_1, c_2, \ldots, c_n) \propto \prod_{i=1}^{n} c_j^{s_j}(1 - c_j^{t_j}). \tag{12}$$

The joint posterior distribution of the parameters is given by

$$\begin{aligned} f(\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mu_b, \sigma_b^2 \mid \mathbf{u}, \mu_\theta, \sigma_\theta^2, v_b, \lambda_b, \mathbf{v}, \boldsymbol{\omega}, \mathbf{s}, \mathbf{t}) &= L(\mathbf{u} \mid \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c})\left\{\prod_{i=1}^{N} f(\theta_i \mid \mu_\theta, \sigma_\theta^2)\right\} \\ &\cdot \left\{\prod_{j=1}^{n} f(b_j \mid \mu_b, \sigma_b^2)\right\}\{f(\mu_b, \sigma_b^2 \mid v_b, \lambda_b)\} \\ &\cdot \left\{\prod_{j=1}^{n} f(a_j \mid v_j, \omega_j)\right\}\left\{\prod_{j=1}^{n} f(c_j \mid s_j, t_j)\right\}. \end{aligned} \tag{13}$$

The likelihood function $L(\mathbf{u} \mid \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c})$ is given by (5). The RHS of (13) can be simplified further by noting that

$$\left\{\prod_{i=1}^{N} f(\theta_i \mid \mu_\theta, \sigma_\theta^2)\right\} \propto \exp -\frac{1}{2}\left\{\sum_{i=1}^{N} \theta_i^2\right\}, \tag{14}$$

and

$$\left\{\prod_{j=1}^{n} f(b_j \mid \mu_b, \sigma_b^2)db_j\right\} f(\mu_b, \sigma_b^2 \mid v_b, \lambda_b) \propto (\sigma_b^2)^{-(N+v_b+2)/2} \exp - \frac{\{\lambda_b + \sum_{j=1}^{n}(b_j - \mu_b)^2\}}{2\sigma_b^2} \tag{15}$$

Integrating with respect to the nuisance parameters μ_b, σ_b^2, the following joint posterior distribution is obtained:

$$\begin{aligned} f(\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c} \mid \mathbf{u}, \mu_\theta, \sigma_\theta^2, v_b, \lambda_b, \mathbf{v}, \boldsymbol{\omega}, \mathbf{s}, \mathbf{t}) &\propto L(\mathbf{u} \mid \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}) \exp -\left\{\frac{1}{2}\sum_{i=1}^{N} \theta_i^2\right\} \\ &\cdot \left[\left\{\lambda_b + \sum_{j=1}^{n}(b_j - b_.)^2\right\}^{-(n+v_b-1)/2}\right] \end{aligned}$$

$$\cdot \left[\prod_{j=1}^{n} a_j^{v_j - 1} \exp\left(-a_j^2/2\omega_j\right) c_j^{s_j} (1 - c_j)^{t_j} \right]. \tag{16}$$

Here b denotes the mean of $b_1, b_2, \ldots, b_n$.

In the joint maximum-likelihood estimation of the item and ability parameters, the discrimination and pseudo-chance level parameters are often poorly estimated. Swaminathan and Gifford (1985) have observed that in the two-parameter model the imposition of an informative prior distribution on the discrimination parameter (with noninformative priors on the ability and difficulty parameters) often yields valid estimates for the discrimination, difficulty, and ability parameters. Since this approach reduces the complexity of the estimation procedure, it appears that it may be employed in the three-parameter model when appropriate. In order to implement this procedure, we merely assume that

$$f(\boldsymbol{\theta})d\boldsymbol{\theta} \propto d\boldsymbol{\theta}, \tag{17}$$

and

$$f(\mathbf{b})d\mathbf{b} \propto d\mathbf{b}. \tag{18}$$

The joint posterior given by (16) reduces to

$$f(\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c} \mid \mathbf{u}, \mathbf{v}, \boldsymbol{\omega}, \mathbf{s}, \mathbf{t}) = L(\mathbf{u} \mid \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}) \cdot \left[\prod_{j=1}^{n} a_j^{v_j - 1} \exp\left(\frac{-a_j^2}{2\omega_j}\right) c_j^{s_j} (1 - c_j)^{t_j} \right]. \tag{19}$$

The posterior distributions given by (16) and (19) contain all the information required for making probability statements about the item and ability parameters. However, these expressions are unwieldy and hence point estimates of the parameters have to be obtained.

Estimation of Parameters

Following Lindley and Smith (1972) and Novick et al. (1973), we shall take the joint modal estimators of the parameters as the relevant estimators. These are defined as those values that maximize the joint posterior distribution given by (16) and (19), or more conveniently, as those values that maximize the logarithm of the joint posterior distribution. Now,

$$\begin{aligned} \log f(\boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c} \mid \mathbf{u}, \mu_\theta, \sigma_\theta^2, v_b, \lambda_b, \mathbf{v}, \boldsymbol{\omega}, \mathbf{s}, \mathbf{t}) &= \log L(\mathbf{u} \mid \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}) + \log f_1(\boldsymbol{\theta}) \\ &\quad + \log f_2(\mathbf{b}) + \log f_3(\mathbf{a}) + \log f_4(\mathbf{c}) + \text{constant}, \end{aligned} \tag{20}$$

where

$$\log L(\mathbf{u} \mid \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{c}) = \sum_{i=1}^{N} \sum_{j=1}^{n} \left[u_{ij} \log P_{ij} + (1 - u_{ij}) \log (1 - P_{ij})\right], \tag{21}$$

$$\log f_1(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{i=1}^{N} \theta_i^2, \tag{22}$$

$$\log f_2(\mathbf{b}) = -\tfrac{1}{2}(n + v_b - 1) \log \left\{\lambda_b + \sum_{j=1}^{n} (b_j - b_.)^2\right\}, \tag{23}$$

$$\log f_3(\mathbf{a}) = \sum_{j=1}^{n} (v_j - 1) \log a_j - \frac{a_j^2}{2\omega_j}\bigg\}, \tag{24}$$

and

$$\log f_4(\mathbf{c}) = \sum_{j=1}^{n} \{s_j \log c_j + t_j \log (1 - c_j)\}. \tag{25}$$

For the posterior distribution given by (19), the terms involving $\log f_1(\boldsymbol{\theta})$ and $\log f_2(\mathbf{b})$ are omitted.

The modal equations are:

$$\partial \frac{\log L + \log f_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0} \tag{26}$$

$$\partial \frac{\log L + \log f_2(\mathbf{b})}{\partial \mathbf{b}} = \mathbf{0} \tag{27}$$

$$\partial \frac{\log L + \log f_3(\mathbf{a})}{\partial \mathbf{a}} = \mathbf{0} \tag{28}$$

and

$$\partial \frac{\log L + \log f_4(\mathbf{c})}{\partial \mathbf{c}} = \mathbf{0}. \tag{29}$$

Since closed form solutions to these equations do not appear to be available, the Newton-Raphson procedure may be employed to obtain the joint modal estimates. In general, if we let

$$\boldsymbol{\tau}' = [\boldsymbol{\theta}'\mathbf{b}'\mathbf{a}'\mathbf{c}'], \tag{30}$$

and denote $\boldsymbol{\tau}^{(k)}$ as the kth approximation to the vector of values that maximizes the logarithm of the joint posterior distribution, $\log f$, then a better approximation $\boldsymbol{\tau}^{(k+1)}$ is given by

$$\boldsymbol{\tau}^{(k+1)} = \boldsymbol{\tau}^{(k)} - H^{-1}\{\boldsymbol{\tau}^{(k)}\}\mathbf{g}\{\boldsymbol{\tau}^{(k)}\}, \tag{31}$$

where $H^{-1}\{\boldsymbol{\tau}^{(k)}\}$ and $\mathbf{g}\{\boldsymbol{\tau}^{(k)}\}$ are the matrix of second derivatives and the vector of first derivatives of $\log f$, respectively, evaluated at $\boldsymbol{\tau}^{(k)}$. Since H is a square matrix of order $3n + N$, the inversion of H raises difficulties when N and n are large. For ease of implementation, therefore, the off-diagonal elements of H may be ignored, in which case,

$$\tau_m^{(k+1)} = \tau_m^{(k)} - \frac{g\{\tau_m^{(k)}\}}{h\{\tau_m^{(k)}\}} \tag{32}$$

where τ_m is the m-th parameter in the vector $\boldsymbol{\tau}$ and $g(\tau)$ and $h(\tau)$ are first and second derivatives of $\log f$. It is easily seen that $g(\tau)$ and $h(\tau)$ have two additive components, one resulting from the likelihood and the other from the prior distribution. These components are summarized in Table 1.

The elements in Table 1 can be easily adjusted to reflect various situations or the various prior distributions that may be employed. If, for example, noninformative or uniform priors are specified for ability and difficulty parameters, the first and second derivatives will have contributions only from the likelihood. If it is assumed that the pseudo chance level parameter c_j is common to all items, then c_j is replaced by c, the common value. Accordingly, s_j and t_j are replaced by the common values s and t. Similar modifications can be made for the discrimination parameter.

Table 1
First and Second Derivatives of the Logarithm of the Joint Posterior Distribution

Parameter	Contribution	First* Derivative	Second Derivative
θ_i	Likelihood	$\sum_{j=1}^{n} a_j(P_{ij}-c_j)(u_{ij}-P_{ij})/P_{ij}(1-c_j)$	$\sum_{j=1}^{n} a_j^2(P_{ij}-c_j)(u_{ij}c_j-P_{ij}^2)Q_{ij}/P_{ij}^2(1-c_j)^2$
	Prior	$-\theta_i$	-1
a_j	Likelihood	$\sum_{i=1}^{N} (\theta_i-b_j)(P_{ij}-c_j)(u_{ij}-P_{ij})/P_{ij}(1-c_j)$	$\sum_{i=1}^{N} (\theta_i-b_j)^2(P_{ij}-c_j)(u_{ij}c_j-P_{ij}^2)Q_{ij}/P_{ij}(1-c_j)^2$
	Prior	$+(\nu_j-1)/a_j - a_j/\omega_j$	$-(\nu_j-1)/a_j^2 - 1/\omega_j$
b_j	Likelihood	$-\sum_{i=1}^{N} a_j(P_{ij}-c_j)(u_{ij}-P_{ij})/P_{ij}(1-c_j)$	$\sum_{i=1}^{N} a_j^2(P_{ij}-c_j)(u_{ij}c_j-P_{ij}^2)Q_{ij}^2/P_{ij}(1-c_j)^2$
	Prior	$-(b_j-b.)^2/\sigma_b^2$	$-[\sigma_b^2(1-\frac{1}{n}) - [2(b_i-b.)^2/(\nu_b+n-1]/(\sigma_b^2)^2$
c_j	Likelihood	$\sum_{i=1}^{N} (u_{ij}-P_{ij})/P_{ij}(1-c_j)$	$\sum_{i=1}^{N} [u_{ij}(2P_{ij}-1)-P_{ij}^2]/P_{ij}^2(1-c_j)^2$
	Prior	$+s_j/c_j - t_j/(1-c_j)$	$-s_j/c_j^2 + t_j/(1-c_j)^2$

* $P_{ij} = [1+\exp\{-a_j(\theta_i-b_j)\}]^{-1}$ $\qquad$ $\sigma_b^2 = [\lambda_b + \sum_{j=1}^{n}(b_j-b.)^2]/(\nu_b+n-1)$

$b. = \Sigma b_j/n$

Initial values for starting the iterative process defined by (32) may be taken as follows:

$$\theta_i^{(0)} = \log\left(\frac{p_i}{1 - p_i}\right) \tag{33}$$

$$b_j^{(0)} = \Phi^{-1}\frac{q_j}{r_j} \tag{34}$$

$$a_j^{(0)} = \frac{r_j}{(1 - r_j^2)^{1/2}} \tag{35}$$

and

$$c_j^{(0)} = \frac{1}{m_j}. \tag{36}$$

Here

$$p_i = \sum_j \frac{u_{ij}}{n},$$

$$q_j = \sum_i \frac{u_{ij}}{N},$$

where m_j is the number of choices for answering item j, $\Phi^{-1}(q_j)$ is the normal deviate that cuts off an area q_j to its right, and r_j is the point biserial correlation between item score and the total score. These initial values are not valid when $\sum_j u_{ij}$ is 0 or n. In this case $\sum_j u_{ij}$ may be set at 0.5 and $(n - 0.5)$ respectively. The same consideration applies when $\sum_i u_{ij} = 0$ or N. The same adjustment should be made in this case also. The starting values given here for b_j and a_j are based on the relationships that exist between conventional item statistics and the item parameters (Lord, 1980, pp. 33–34). In the strict sense, these relationships hold only for the model with $c_j = 0$. They are employed here merely for convenience and do not seem to affect the convergence of the iterative process adversely.

Specification of Prior Belief

The specification of the prior distributions of b_j and a_j are described by Swaminathan and Gifford (1985), and hence, will not be described here.

The descriptive statistics of the beta distribution with parameters (s_j, t_j) are (Novick & Jackson, 1974, p. 113):

$$\text{mean} = \frac{s_j + 1}{s_j + t_j + 2}$$

$$\text{variance} = \frac{(s_j + 1)(t_j + 1)}{(s_j + t_j + 2)^2(s_j + t_j + 3)}$$

and

$$\text{mode} = \frac{s_j}{s_j + t_j}, \qquad s_j, t_j > 0.$$

This information together with the $(1 - \alpha)$-th highest density credibility intervals as tabulated in Novick and Jackson (1974, p. 402) provide guidelines for selecting values for s_j

and t_j. For example, if we feel that the probability that an examinee with low ability will respond correctly to an item is .15, then this may be taken as the mean value, M. Following Novick et al. (1973), an interrogative process may be employed. This requires the specification of m, the number of observations that the prior information is worth, that is, the degree of confidence the investigator has about the prior. Since we can take

$$s_j + t_j + 2 = m,$$

and

$$\frac{s_j + 1}{s_j + t_j + 2} = M,$$

we have

$$s_j = mM,$$

and

$$t_j = m(1 - M) - 2.$$

Suppose that the number of observations the prior information is worth is 15. Then, with $M = .15$ and $m = 15$, we have

$$s_j = 2.25,$$

and

$$t_j = 10.75.$$

Rounding off, we may take $s_j = 2.0$ and $t_j = 11.0$. This yields the 90% credibility interval (.04, .36), the 95% credibility interval (.03, .40), and the 99% credibility interval (.01, .49) for c_j. If these intervals are considered too wide, the value of m may be increased. Empirical investigations suggest that in general a prior that is not too vague and at the same time not too precise is desirable. Thus, a conservative estimate of m such as the one employed in the illustration is reasonable in most situations.

Simulation Studies

In order to demonstrate the Bayesian procedure, a simulation study was carried out. Item responses for N individuals ($N = 100, 200, 400$) and n items ($n = 25, 35$) were artificially generated according to the three-parameter logistic model using the computer program DATGEN (Hambleton & Rovinelli, 1973). Ability and difficulty parameters were generated as samples from a normal distribution with zero mean and unit variance. The values for a_j were generated as samples from a uniform distribution on the interval (.5, 2.0). The values for c_j were also generated as samples from a uniform distribution, but on the interval (.04, .22).

The priors for the parameters θ_i, b_j, a_j, and c_j were set in such a way that they did not duplicate the generated data. Since θ_i and b_j were generated from a normal distribution, noninformative or uniform priors were specified for these parameters. (This was true except in one situation where the hierarchical Bayes model was employed with $v_b = 10$ to coax convergence.) A chi distribution with parameters $v_j = 10$ and $\omega_j = .5$ was chosen for each a_j (these in turn yielded the 99% credibility interval with upper and lower limits 2.45 and .63 respectively). In order to simplify the setting of priors, s_j and t_j were set at 2 and 12 for each c_j.

Item responses generated were analyzed using the LOGIST program (Wood, Wingersky, & Lord, 1978) and also using the Bayes procedure with a program developed for this purpose. The maximum likelihood estimators (ML) and the Bayes estimators were compared with respect to: (a) the mean squared differences between the true values and the estimated values; (b) the correlations between the true values and the estimated values. Mean squared deviation of estimates from the mean value of the estimates was not employed as a criterion since this would clearly favor the Bayes estimator.

The results of the simulation study are summarized in Table 2a through Table 2d. Clearly, the trend favors the Bayes estimator. The most important improvement that results with the Bayes estimator is that the outward drift of the estimates is arrested. The LOGIST program, as mentioned earlier, requires that upper limits be placed on the discrimination and pseudo-chance level estimates. For the purpose of this study, an upper limit of 10 was imposed on the estimates of a_j. The number in parentheses in Table 2 indicates the number of estimates that reached the maximum value of 5 (in separate runs when the limit was set at 20, these estimates reached this value). Similar outward drift was experienced in the maximum likelihood estimation of difficulty and ability parameters. These drifts were clearly arrested in the Bayes procedure.

An examination of the mean squared differences reveals another interesting trend. The mean squared differences are consistently smaller for the Bayes estimators than for the maximum likelihood estimators indicating that the Bayes estimators are more accurate than the ML estimators. It is interesting to note that this improvement is observed for the estimates of difficulty and ability even though non-informative priors were used for these parameters. The reason for this is that the discrimination and chance level parameters are estimated more accurately with the Bayes procedure than the ML procedure. These in turn yield more accurate estimates of ability and difficulty parameters.

Conclusion

The Bayes procedure outlined in this paper for the estimation of parameters in the three-parameter logistic model appears to be an attractive alternative to the maximum likelihood procedure as implemented by the LOGIST program. The most important fea-

Table 2a

Accuracy of Estimation in the Three-Parameter Model

		Ability Parameter			
n	N	Correlation		MSD	
		Bayes	ML	Bayes	ML
	100	.930	.898	.139	.201
25	200	.936	.906	.128	.186
	400	.935	--	.130	--
	100	.956	.947	.087	.105
35	200	.953	.954	.094	.091
	400	.950	.947	.100	.106

-- No convergence.

Table 2b

Accuracy of Estimation in the Three-Parameter Model

		Difficulty Parameter			
n	N	Correlation		MSD	
		Bayes	ML	Bayes	ML
	100	.947	.893	.176	.620
25	200	.973	.954	.075	.160
	400	.984	--	.043	--
	100	.981	.914(1)*	.067	.520
35	200	.986	.968	.040	.124
	400	.992	.989	.024	.045

* The number in parentheses indicates the number of estimates that were outside [-5,5].

-- No convergence.

ture of the Bayes procedure is that the estimates of the parameters are meaningful in that they remain in the admissible parameter space. The outward drift experienced with the maximum likelihood procedure is arrested naturally with the specification of prior beliefs rather than through the use of artificially set limits.

Table 2c

Accuracy of Estimation in the Three-Parameter Model

		Discrimination Parameter			
n	N	Correlation		MSD	
		Bayes	ML	Bayes	ML
	100	.584	.462(3)*	.083	8.646
25	200	.623	.480(3)	.072	8.496
	400	.777	--	.051	--
	100	.523	.293(6)	.091	13.103
35	200	.585	.446(3)	.077	6.224
	400	.697	.382(1)	.068	2.293

* The number in parentheses indicates the number of estimates that were outside [0,10].

-- No convergence.

Table 2d

Accuracy of Estimation in the Three-Parameter Model

		Chance-level Parameter			
n	N	Correlation		MSD	
		Bayes	ML	Bayes	ML
	100	.637	.147	.0035	.0043
25	200	.336	.334	.0040	.0039
	400	.340	--	.0040	--
	100	.460	.155	.0036	.0064
35	200	.546	.447	.0030	.0075
	400	.730	.662	.0022	.0026

-- No convergence.

An alternative to the Bayes procedure described in this paper is the combination of informative priors on the discrimination and chance level parameters and non-informative priors on the ability and difficulty parameters. As demonstrated through the simulation study, it is effective. Moreover, this procedure is attractive in that it reduces the complexity of the estimation procedure and also the regression effect on the estimates, in particular, on the estimates of ability.

Specifications of priors on the discrimination and chance level parameters can be accomplished in a straightforward manner. The interrogative procedure suggested by Novick and Jackson (1974), and Novick et al. (1973), can be easily implemented in this situation.

References

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika, 46*, 443–460.

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

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eignor, D. R., & Gifford, J. A. (1978). Developments in latent train theory: Models, technical issues, and applications. *Review of Educational Research, 48*, 467–510.

Lindley, D. V., & Smith, A. F. (1972). Bayesian estimates for the linear model. *Journal of the Royal Statistical Society, 34*, (Series B), 1–41.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale: Lawrence Erlebaum.

Loyd, B. H., & Hoover, H. D. (1980) Vertical equating using the Rasch model. *Journal of Educational Measurement, 17*, 179–193.

Mislevy, R. J. (1986) Bayes Modal Estimation in Item Response Models. *Psychometrika, 51*, 177–195.

Novick, M. R., & Jackson, P. H. (1974). *Statistical methods for educational and psychological research.* New York: McGraw-Hill.

Novick, M. R., Lewis, C., & Jackson, P. H. (1973). The estimation of proportions in *m* groups. *Psychometrika, 38*, 19–46.

Slinde, J. A., & Linn, R. L. (1979). A note on vertical equating via the Rasch model for groups of quite different ability and tests of quite different difficulty. *Journal of Educational Measurement, 16*, 159–165.

Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics, 7*, 175–192.

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

Urry, V. W. (1976). Ancilliary estimators for the item parameters of mental tests. In W. A. Gorham (Chair), *Computerized testing: Steps toward the inevitable conquest* (PS-76-1). Washington, DC: US Civil Service Commission, Personnel Research and Development Center.

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