The Use of Prior Distributions in Marginalized Bayesian Item Parameter Estimation: A Didactic

Michael R. Harwell, University of Pittsburgh

Frank B. Baker, University of Wisconsin

The marginal maximum likelihood estimation (MMLE) procedure (Bock & Lieberman, 1970; Bock & Aitkin, 1981) has led to advances in the estimation of item parameters in item response theory. Mislevy (1986) extended this approach by employing the hierarchical Bayesian estimation model of Lindley and Smith (1972). Mislevy's procedure posits prior probability distributions for both ability and item parameters, and is implemented in the PC-BILOG computer program. This paper extends the work of Harwell, Baker, and Zwarts (1988), who provided the mathematical and implementation details of MMLE in an earlier didactic paper, by encompassing Mislevy's marginalized Bayesian estimation of item parameters. The purpose was to communicate the essential conceptual and mathematical details of Mislevy's procedure to practitioners and to users of PC-BILOG, thus making it more accessible. *Index terms: Bayesian estimation, BILOG, item parameter estimation, item response theory.*

The joint maximum likelihood estimation (JMLE) paradigm of Birnbaum (1968), as implemented in LOGIST (Wingersky, Barton, & Lord, 1982), has been the standard method for estimating parameters in item response theory (IRT). In this approach, the true values of the item and ability parameters are unknown and must be estimated from an examinee's dichotomously scored item responses. Statistical inferences are confined to the items and examinees actually used.

JMLE has several problems. For example, item parameter estimates may assume unreasonable values. A second problem is that it might not be possible to estimate ability (θ) for unusual examinee response patterns, such as when all items are answered correctly or incorrectly (Mislevy & Stocking, 1989). In addition, for settings in which item parameter estimation is of interest, such as a calibration study, the use of JMLE for tests of finite length may result in estimates that are not statistically consistent. This is the problem of estimating structural (item) parameters in the presence of incidental (θ) parameters, first pointed out by Neyman and Scott (1948) in another context. Anderson (1972) and Wright (1977) have discussed this problem in the context of IRT.

Bock and Lieberman (1970) and Bock and Aitkin (1981) used marginalized maximum likelihood estimation (MMLE) to address the problem of inconsistent item parameter estimates. MMLE is distinguished from JMLE by the assumption that examinee abilities have a distribution in a population, which is sensible both substantively and statistically. When this distribution is known, or can be estimated, it permits θ to be integrated out of the likelihood function. This frees item parameters from their dependence on the θ parameters of individual examinees. Assuming that the item response model and the θ distribution are correct, the resulting item parameter estimates are maximum likelihood estimates (MLEs) that are consistent for tests of finite length (Bock & Aitkin, 1981). It should be noted that the Rasch model resolves the problem of incidental parameters by replacing an examinee's θ parameter with its sufficient statistic—the number-correct score. This solution is only possible under the Rasch model (see Anderson, 1972; Mislevy, 1988).

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 15, No. 4, December 1991, pp. 375-389
© Copyright 1991 Applied Psychological Measurement Inc.
0146-6216/91/040375-15$2.00

Mislevy and Bock (1986) have implemented MMLE on microcomputers in the computer program PC-BILOG. However, some estimation difficulties persist because item parameter estimates obtained through MMLE can assume unreasonable values in some datasets (Mislevy, 1986). The use of estimation procedures based on a Bayesian approach may prevent such values from occurring.

Bayesian approaches in IRT can be distinguished by whether parameter estimation takes place after marginalization (i.e., integration) over incidental parameters (Mislevy, 1986; Tsutakawa & Lin, 1986) or without any marginalization (Swaminathan & Gifford, 1982, 1985, 1986). This paper focuses on Mislevy's (1986) marginalized procedure because it is a direct extension of the MMLE approach of Bock and Aitkin (1981) and because O'Hagan (1976) provides numerical evidence that marginalized solutions are superior to unmarginalized solutions. Mislevy's estimation procedure inherits the properties of MMLE but constrains the item parameter estimates.

The availability of these Bayesian methods in PC-BILOG (Mislevy & Bock, 1986) provides a powerful yet flexible set of procedures for estimating item parameters in IRT. However, for a practitioner to take full advantage of the method, a complete understanding of the mathematical underpinnings of these procedures—especially the role of prior distributions—is necessary.

In a didactic paper, Harwell et al. (1988) provided the mathematical and implementation details of the MMLE procedure. This paper extends that presentation to encompass Mislevy's (1986) marginalized Bayesian procedure for estimating item parameters. The purpose was to supplement the work of Mislevy by communicating the essential conceptual and mathematical details of Mislevy's (1986) procedure to users of PC-BILOG. This should aid users of PC-BILOG and others in understanding the fundamentals on which the program is based. This paper does not explicate how to use PC-BILOG in specific testing situations. For an overview and comparison of LOGIST and BILOG, see Mislevy and Stocking (1989) and Yen (1987). The present discussion centers on the three-parameter logistic IRT model because the one- and two-parameter IRT models are special cases of this model and, with a few exceptions, the current presentation is applicable to the simpler models.

The Bayesian approach to parameter estimation is first outlined. An explanation of the marginalized procedure (Mislevy, 1986) is presented next, and the use of prior information in the Mislevy procedure, as implemented in PC-BILOG, is described. The role of prior probability distributions in estimating item parameters is emphasized.

Notation

examinees	$i = 1, 2, 3, \ldots, n$
items	$j = 1, 2, 3, \ldots, J$
$y_{ji} = 0,1$	binary response of examinee i to item j
$\mathbf{Y}_i = (y_{1i}, y_{2i}, \ldots, y_{Ji})'$	vector of item responses of the ith examinee to the J items
$\boldsymbol{\theta}_i$	ability of examinee i
$\boldsymbol{\theta}$	$n \times 1$ vector containing the θ parameters of all n examinees
$\boldsymbol{X}_k$	known θ value (quadrature point) of the kth ($k = 1, \ldots, q$) ability group
$\boldsymbol{a}_j$	item discrimination parameter
$\boldsymbol{b}_j$	item difficulty parameter
c_j	item "guessing" parameter or lower asymptote
$P_j(\theta_i) = P(y_{ji} = 1 \mid \theta_i)$	probability of a correct response to item j for an examinee with θ_i, defined for a given IRT model

$$= c_j + \frac{(1 - c_j)\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]} \tag{1}$$

three-parameter logistic IRT model for individual item response data

$$Q_j(\theta_i) = 1 - P_j(\theta_i) \tag{2}$$

$$P_j^*(\theta_i) = \frac{\exp[(\theta_i - b_j)e_j^{\alpha}]}{1 + \exp[(\theta_i - b_j)e_j^{\alpha}]} \tag{3}$$

where

$$\alpha_j = \log a_j,\ a_j = e_j^{\alpha} \tag{4}$$

$$Q_j^*(\theta_i) = 1 - P_j^*(\theta_i) \tag{5}$$

ξ	vector of item parameters
τ	vector of parameters of examinee population distribution
η	vector of parameters of item parameter population distribution
$g(\theta_i \vert \tau)$	probability distribution for the ith examinee's θ, conditional on the population parameters in the vector τ
$g(\tau)$	unconditional probability distribution for τ
$g(\xi_j \vert \eta)$	probability distribution for the parameters of item j contained in the vector ξ_j, conditional on the population parameters in the vector η
$g(\xi \vert \mathbf{Y},\theta,\eta)$	posterior probability distribution for item parameters, conditional on the observed item response data, θ parameters, and parameters of the item parameter population distribution
$g(\xi)$	unconditional probability distribution for the jth item's parameters, ξ_j, assumed to have a common form across items
$g(\eta)$	unconditional probability distribution for η
$\propto$	"is proportional to"

The functions $g(\theta_i|\tau)$, $g(\tau)$, and so forth, are distinguished by their arguments. Following Mislevy (1986), a_j, b_j, c_j, and θ_i are referred to as parameters, and τ and η as vectors containing population parameters (hyperparameters).

The Bayesian Approach To Parameter Estimation

General introductions to Bayesian statistical methodology include Cornfield (1969), de Finetti (1974), Edwards, Lindman, and Savage (1963), Lindley (1970a, 1970b, 1971), and Novick and Jackson (1974). Examples of the use of these methods in educational settings can be found in Novick and Jackson (1974), Novick, Jackson, Thayer, and Cole (1972), and Rubin (1980). Lord (1986) compared maximum likelihood and Bayesian estimation methods in IRT.

Bayes' Theorem provides a way of expressing conditional probability. It combines probabilities obtained from a likelihood function that uses sample data with probabilities obtained using prior information about the distribution of the set of unknown parameters. An application of Bayes' Theorem produces a posterior probability distribution, which is proportional to the product of the likelihood function and the prior probability distribution. The posterior probability distribution is used to make inferences about the unknown parameters (Lindley, 1971, p. 36).

An Illustration of Bayesian Methods in Item Parameter Estimation

The use of Bayes' Theorem can be illustrated in an IRT setting in which item parameters are to be estimated when ability parameters are known. Let $g(\boldsymbol{\xi}|\boldsymbol{\eta})$ represent the probability distribution reflecting the prior belief about the distribution of the possible values of the jth item's parameters, conditional on the parameters in $\boldsymbol{\eta}$. Assume that each of the J items has prior distributions of the same form. Let $\mathbf{Y}$ be an $n \times J$ matrix of item responses and $L(\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\theta})$ be the likelihood function of the sample item responses conditional on the parameters in $\boldsymbol{\xi}$ and $\boldsymbol{\theta}$. The posterior probability distribution across items and examinees can be expressed as

$$g(\boldsymbol{\xi}|\mathbf{Y},\boldsymbol{\theta},\boldsymbol{\eta}) \propto L(\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\theta})\ g(\boldsymbol{\xi}|\boldsymbol{\eta}) \quad . \tag{6}$$

The likelihood function $L(\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\theta})$ is defined as

$$L(\mathbf{Y}|\boldsymbol{\theta},\xi) = \prod_{i}^{n} \prod_{j}^{J} P_j(\theta_i)^{y_{ji}}\ Q_j(\theta_i)^{1-y_{ji}} = \prod_{i}^{n} P(\mathbf{Y}_i|\theta_i,\xi) \quad . \tag{7}$$

$P(\mathbf{Y}_i|\theta_i,\xi)$ is the probability of an examinee's response vector $\mathbf{Y}_i$, conditional on a known value of θ_i and the item parameters in $\boldsymbol{\xi}$.

Inferences about the unknown item parameters in $\boldsymbol{\xi}$ typically take the form of point estimates that maximize the posterior probability $g(\boldsymbol{\xi}|\mathbf{Y},\boldsymbol{\theta},\boldsymbol{\eta})$ with respect to the unknown parameters. In Bayesian estimation, this is often the estimated mode of the posterior probability distribution, and the resulting estimates are typically known as Bayes modal estimates (BMEs) (Mislevy & Stocking, 1989).

The relative contribution of the prior distribution $g(\boldsymbol{\xi}|\boldsymbol{\eta})$ and the likelihood $L(\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\theta})$ to $g(\boldsymbol{\xi}|\mathbf{Y},\boldsymbol{\theta},\boldsymbol{\eta})$, and thus to the estimates, is an important issue. If the number of examinees is large and the contribution of the prior probability distribution is small, the likelihood will dominate the posterior probability distribution. That is, the item parameter estimates obtained from $g(\boldsymbol{\xi}|\mathbf{Y},\boldsymbol{\theta},\boldsymbol{\eta})$ in Equation 6 will depend almost entirely on the observed item response data represented through $L(\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\theta})$; $g(\boldsymbol{\xi}|\boldsymbol{\eta})$ will have little effect on the estimates. In this case, the BMEs are likely to be almost identical to the estimates obtained from maximum likelihood estimation. If the contribution of the prior probability distribution is substantial, the BMEs will differ from the MLEs.

In a Bayesian approach, it is typical (although not necessary) to assume that the parameters of a given type (e.g., a_j) in the IRT model are independently and identically distributed. It is also assumed that such parameters are exchangeable, meaning that the prior probability distribution for a particular parameter (e.g., a_j for item j) is no different from that of any other parameter of the same type (Swaminathan & Gifford, 1985; see Mislevy, 1988 for an alternative conceptualization of exchangeability). With these assumptions, an appropriate distributional form for the prior distribution is specified.

Prior Distributions for Item Parameters

Fortunately, tests designed to assess specific traits, constructs, or abilities, and the IRT models used, often provide the necessary theoretical and empirical basis for selecting appropriate prior distributions. For example, b_j parameters are often between –4 and +4 standard deviations in value, suggesting that a unimodal, symmetric distribution such as the normal distribution can serve as a prior distribution for difficulty. If the variance of the prior distribution of b_j is small, the prior is considered to be informative and its contribution to parameter estimation is likely to be substantial. A small variance means that the values of b_j will be tightly clustered about the mean of the prior distribution, with some values of b_j more likely than others. A large variance, on the other hand, would result in an uninformative prior because the values of b_j are more variable and not clustered about the mean of the prior (Novick & Jackson, 1974, p. 154). Other things being equal, a prior distribution with a large variance would have less impact on parameter estimation than a prior with a small variance. Hence, the variance of a prior distribution plays a key role in estimating item parameters.

The primary effect of an informative prior on parameter estimation is to "shrink" the estimate toward the mean of the parameter's prior distribution by an amount that is proportional to the information contained in the prior distribution of that parameter (Mislevy & Stocking, 1989). The more informative the prior distribution, the more the parameter estimate tends to be pulled toward the mean of its prior distribution. Pulling estimates toward the prior mean helps prevent the estimates from moving toward unreasonable values during the estimation process. If the mean of the prior is close to the true value of the parameter, an informative prior may also lead to BMEs that are closer to the true parameter value than are their maximum likelihood counterparts (Swaminathan & Gifford, 1985).

Mislevy's Marginalized, Two-Stage Bayesian Procedure

Recall that a Bayesian approach produces a posterior distribution that depends on the contribution of both prior information about parameters and information obtained from the sample item response data. Inferences about unknown parameters are then based on the posterior distribution. Mislevy (1986) presented a Bayesian procedure for estimating item parameters in IRT that is a generalized form of Equation 6 and represents an extension of the marginalized solution of Bock and Aitkin (1981). Mislevy employed the two-stage, classical Bayesian estimation procedure attributed to Lindley and Smith (1972) in which prior information is specified in a hierarchical fashion.

The General Form of Mislevy's Model

Following Mislevy (1986), the process begins with the joint density of all of the parameters prior to data collection. These parameters are assumed to be independent and continuous random variables with specified probability distributions:

$$g(\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{\xi},\boldsymbol{\eta}) = \prod_{i}^{n} g(\theta_i|\boldsymbol{\tau}) \prod_{j}^{J} g(\xi_j|\boldsymbol{\eta})\, g(\boldsymbol{\tau})\, g(\boldsymbol{\eta}) \quad . \qquad \textbf{(8)}$$

Recall that $g(\theta_i|\tau)$ is the probability distribution of an examinee's θ parameter and is conditional on the population parameters of the θ distribution in the vector $\boldsymbol{\tau}$. Typically, the prior distribution of θ is posited to be normally distributed. Because abilities are assumed to be independently and identically distributed, $\boldsymbol{\tau}$ contains the common mean (μ_θ) and variance (σ_θ^2) of these prior θ distributions. In the Lindley and Smith (1972) hierarchical model, the population parameters μ_θ and σ_θ^2 are known as hyperparameters. The hyperparameters can also be treated as random variables having a probability distribution [e.g., $g(\boldsymbol{\tau})$].

To make inferences about all the unknown parameters—in this case $\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{\xi},\boldsymbol{\eta}$—after data collection, the posterior distribution across all items and examinees obtained from an application of Bayes' Theorem is:

$$g(\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{\xi},\boldsymbol{\eta}|\mathbf{Y}) \propto L(\mathbf{Y}|\boldsymbol{\theta},\boldsymbol{\xi})\, g(\boldsymbol{\theta}|\boldsymbol{\tau})\, g(\boldsymbol{\tau})\, g(\boldsymbol{\xi}|\boldsymbol{\eta})\, g(\boldsymbol{\eta}) \quad . \qquad \textbf{(9)}$$

In this context, specification of the prior information is accomplished through the parameters (hyperparameters) of the prior probability distributions, $g(\boldsymbol{\theta}|\boldsymbol{\tau})$ and $g(\boldsymbol{\xi}|\boldsymbol{\eta})$. This constitutes the first stage in the Lindley and Smith (1972) model. Specifying the parameters of the probability distributions of the hyperparameters, in this case the parameters characterizing $g(\boldsymbol{\tau})$ and $g(\boldsymbol{\eta})$, constitutes the second stage.

Equation 9 contains all the information available about the unknown parameters (Mislevy, 1986). To estimate the item parameters in Equation 9, the marginalization approach of Bock and Lieberman (1970) can be used. Under this approach, the likelihood function is marginalized with respect to θ. If the IRT model and the prior distribution of θ are correct, the resulting item parameter estimates (for tests of finite length) approach their true values as the number of examinees increases.

In general, the choice of variables to marginalize over is dictated by the parameters to be estimated and those of no interest (i.e., incidental or nuisance parameters). Mislevy suggested that in many testing settings the distributions of $\boldsymbol{\eta}$ are not especially interesting; therefore, treating the parameters in $\boldsymbol{\eta}$ as nuisance parameters is appropriate. The distribution of the nuisance parameters can be removed from Equation 9 by integrating over their probability distributions. In the presence of moderate numbers of items and examinees, this will have little effect on the item parameter estimates. In contrast, the hyperparameters contained in $\boldsymbol{\tau}$ are often of interest in θ estimation and their probability distribution is retained.

Integrating over the probability distributions of θ parameters $g(\boldsymbol{\theta}|\boldsymbol{\tau})$ and item population parameters $g(\boldsymbol{\eta})$ leads to a marginalized posterior distribution of the form:

$$g(\boldsymbol{\xi},\boldsymbol{\tau}|\mathbf{Y}) \propto \int_{\boldsymbol{\eta}}\int_{\boldsymbol{\theta}} L(\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\theta})\, g(\boldsymbol{\theta}|\boldsymbol{\tau})\, g(\boldsymbol{\xi}|\boldsymbol{\eta})\, g(\boldsymbol{\tau})\, g(\boldsymbol{\eta})\, d\boldsymbol{\theta}\, d\boldsymbol{\eta} \propto L(\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\tau})\, g(\boldsymbol{\xi})\, g(\boldsymbol{\tau}) \quad . \tag{10}$$

$L(\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\tau})$ is the marginal likelihood resulting from the integration of $L(\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\theta})$ with respect to θ. Equation 10 is the posterior probability distribution that is appropriate for making inferences about item parameters and about the parameters of the θ distribution in the population of examinees. Note that integrating over the population distribution of θ has eliminated the dependence of item parameter estimates on θ estimates of individual examinees. However, the marginal likelihood is still conditional on the hyperparameters, μ_θ and σ^2_θ, of the population θ distribution. Thus, these hyperparameter values and $g(\boldsymbol{\tau})$ must be specified. Similarly, integrating over the population distribution of item parameters in Equation 10 has not eliminated the need to specify values for the hyperparameters in $\boldsymbol{\eta}$ and $g(\boldsymbol{\xi})$.

Marginalized Bayesian Item Parameter Estimation in PC-BILOG

To estimate the unknown item parameters, the partial derivatives of Equation 10 are taken with respect to these parameters and set to 0. The resulting estimation equations are solved one item at a time in the M (maximization) step of the EM (expectation maximization) algorithm (see Mislevy, 1986; Harwell et al., 1988).

For convenience, logarithms of the quantities in Equation 10 can be used. The resulting system of Bayesian estimation equations is given by

$$\frac{\partial}{\partial u_j}\{\log[L(\mathbf{Y}|\boldsymbol{\xi},\boldsymbol{\tau})]\} + \frac{\partial}{\partial u_j}\{\log[g(\boldsymbol{\xi})]\} + \frac{\partial}{\partial u_j}\{\log[g(\boldsymbol{\tau})]\} = 0 \quad , \tag{11}$$

where the u_j represents a parameter associated with the jth item (e.g., a_j). Note that specifying prior distributions for the item parameters and the θ hyperparameters appends terms for the prior distributions to the expression for the derivative of the marginal likelihood. The estimation equations in Equation 11 are marginalized Bayes modal equations in which the BMEs represent the joint mode of the posterior distribution. The estimation of population θ parameters is not considered here (see Mislevy, 1984 for the estimation equations for the θ hyperparameters). Because the $g(\tau)$ distribution does not contain item parameters, its derivative with respect to u_j will be 0. Thus, $g(\tau)$ may be eliminated from Equation 11, resulting in the following system of equations for obtaining BMEs of the item parameters:

$$\frac{\partial}{\partial u_j}\{\log[L(\mathbf{Y}|\xi,\tau)]\} + \frac{\partial}{\partial u_j}\{\log[g(\xi)]\} = 0 \quad . \tag{12}$$

Because $g(\boldsymbol{\theta}|\tau)$ and $g(\boldsymbol{\eta})$ were integrated out in Equation 10 and $g(\tau)$ was eliminated for Equation 12, the Bayes modal estimation equations only involve the first stage of the Lindley and Smith (1972) two-stage hierarchical model.

Marginalized Maximum Likelihood Estimation Equations for Item Parameters

To simplify the presentation of the details of the estimation equations under Mislevy's procedure, the expressions associated with the marginalized likelihood component in Equation 12 are obtained. Then, the derivatives associated with the second component in Equation 12 are presented.

Let

$$L = L(\mathbf{Y}|\xi,\tau) = \prod_{i=1}^{n} P(\mathbf{Y}_i|\xi,\tau) \quad . \tag{13}$$

$P(\mathbf{Y}_i|\xi,\tau) = \int P(\mathbf{Y}|\boldsymbol{\theta},\xi)\, g(\boldsymbol{\theta}|\tau)d\boldsymbol{\theta}$ is the marginalized probability of a vector of item responses $\mathbf{Y}_i$ with respect to the item parameters ξ. After taking the logarithm of L, the first component in Equation 12 can be written as

$$\frac{\partial}{\partial u_j}\left[\sum_i^n \log P(\mathbf{Y}_i|\xi,\tau)\right] \quad . \tag{14}$$

Before estimating the discrimination parameters, PC-BILOG employs the transformation $\alpha_j = \log a_j$, from which it follows that $a_j = e_j^\alpha$. (The justification for employing this transformation is discussed below.) For a three-parameter logistic IRT model,

$$P_j(\theta_i) = c_j + (1 - c_j)\left\{\frac{\exp[(\theta_i - b_j)e_j^\alpha]}{1 + \exp[(\theta_i - b_j)e_j^\alpha]}\right\} = [c_j + (1 - c_j)]\, P_j^*(\theta_i) \quad . \tag{15}$$

The partial derivative of Equation 14 with respect to the transformed discrimination parameter is obtained first:

$$\frac{\partial}{\partial \alpha_j}[\log L] = e_j^\alpha(1 - c_j)\sum_i^n \int_{\theta} \frac{[y_{ji} - P_j(\theta_i)]}{P_j(\theta_i)Q_j(\theta_i)}(\theta_i - b_j)\, P_j^*(\theta_i)Q_j^*(\theta_i)\ [P(\theta_i|\mathbf{Y}_i,\xi,\tau)]\, d\boldsymbol{\theta} \quad . \tag{16}$$

Let $w_{ji} = P_j^*(\theta_i)Q_j^*(\theta_i)/P_j(\theta_i)Q_j(\theta_i)$. The marginalized likelihood component in Equation 12 corresponding to α_j is

$$\alpha_j : e_j^\alpha(1 - c_j)\sum_i^n \int_{\theta} [y_{ji} - P_j(\theta_i)]\, w_{ji}\, (\theta_i - b_j)\, [P(\theta_i|\mathbf{Y}_i,\xi,\tau)]\, d\boldsymbol{\theta} \quad . \tag{17}$$

The marginalized likelihood component in Equation 12 for the difficulty and "guessing" parameters are

$$b_j : -e_j^{\alpha}(1 - c_j)\sum_i^n \int_\theta [y_{ji} - P_j(\theta_i)]\, w_{ji}\, [P(\theta_i|\mathbf{Y}_i,\xi,\tau)]\, d\theta \tag{18}$$

and

$$c_j : (1 - c_j)^{-1}\sum_i^n \int_\theta \left[\frac{y_{ji} - P_j(\theta_i)}{P_j(\theta_i)}\right]\{P(\theta_i|Y_i,\xi,\tau)\}\, d\theta \quad . \tag{19}$$

These equations are identical to those that would result from an application of Equation 20 in Mislevy (1986). In addition, these equations are the same as those given by Harwell et al. (1988) except for the e_j^{α} multipliers.

The expressions given in Equations 17 through 19 are difficult to evaluate by computer because of the required integration over θ. Bock and Lieberman (1970) suggested using numerical quadrature as an approximation for the integration. In numerical quadrature, the area of a continuous distribution with finite moments is approximated by erecting rectangles over specified values, finding the area of these rectangles, and summing them. The midpoint of each rectangle is its "node," X_k ($k = 1, 2, \ldots, q$), and each node has an associated quadrature weight, $A(X_k)$.

In PC-BILOG, a simple histogram approach is employed to implement numerical quadrature. The program assumes that $g(\theta|\tau)$ is standard-normal with $q = 10$ or 20, equally-spaced, standard-normal deviates (X_k) used over the range −4 to +4. The $A(X_k)$ are found by multiplying the width of each rectangle by its height. The width is the difference between the values of adjacent X_ks and the height is given by the ordinate of the standard-normal density at that X_k. At each cycle of the item parameter estimation process, adjusted quadrature weights $A(X_k)$ are computed using the equations given by Mislevy and Bock (1985). An undocumented algorithm is used to standardize the histogram so that the constraints

$$\sum_k^q A(X_k)X_k = 0 \ , \ \sum A(X_k)X_k^2 = 1 \tag{20}$$

are met for the set of nodes employed.

The use of numerical quadrature involves a change from working with individual examinee data to using "artificial" data at each of the q quadrature points (see Bock & Aitkin, 1981; Harwell et al., 1988). The "artificial" data consists of the expected number of examinees, $\bar{n}_{jk}$, and the expected number of correct responses, $\bar{r}_{jk}$, at each node X_k (see Harwell et al., 1988). The $\bar{n}_{jk}$ are defined as

$$\bar{n}_{jk} = \sum_i^n \frac{L(X_k)\, A(X_k)}{\sum_s^q L(X_k)\ A(X_k)} = \sum_i^n P(X_k|\mathbf{Y}_i,\xi,\tau) \ , \ s = 1,2, \ldots, q \tag{21}$$

where

$$L(X_k) = \prod_j^J P_j(X_k)^{y_{ji}} Q_j(X_k)^{1-y_{ji}} \tag{22}$$

is the quadrature form of the likelihood of $\mathbf{Y}_i$ conditional on $\theta_i = X_k$ and the item parameters. In simple terms, Equation 21 is based on computing the expectation (probability) of each examinee having ability X_k for all values of X_k. Then the $\bar{n}_{jk}$ are obtained by aggregating these probabilities separately at each X_k. The $\bar{r}_{jk}$ is the expected number of correct responses to item j made by the $\bar{n}_{jk}$ "artificial examinees" at ability level X_k, and is defined as

$$\bar{r}_{jk} = \sum_{i}^{n} \frac{y_{ji} L(X_k)\, A(X_k)}{\sum_{s}^{q} L(X_k)\, A(X_k)} = \sum_{i}^{n} y_{ji} P(X_k|\mathbf{Y}_i,\xi,\tau) \quad . \tag{23}$$

The $\bar{r}_{jk}$ and $\bar{n}_{jk}$ terms are used in PC-BILOG to estimate item parameters.

Using numerical quadrature results in replacing the $[P(\theta_i|\mathbf{Y}_i,\xi,\tau)]$ term in Equations 17 through 19 with $[P(X_k|\mathbf{Y}_i,\xi,\tau)]$, and replacing the integral over θ by a summation over the X_ks. It can be seen in Equations 21 and 23 that the $[P(X_k|\mathbf{Y}_i,\xi,\tau)]$ terms are then absorbed in the $\bar{r}_{jk}$ and $\bar{n}_{jk}$.

Rewriting Equations 17, 18, and 19 in their numerical quadrature form using the "artificial data" results in the following MMLE equations for item parameters in PC-BILOG:

$$\alpha_j : e_j^{\alpha}(1 - c_j) \sum_{k}^{q} [\bar{r}_{jk} - \bar{n}_{jk} P_j(X_k)]\, w_{jk}\, (X_k - b_j) \quad , \tag{24}$$

$$b_j : -e_j^{\alpha}(1 - c_j) \sum_{k}^{q} [\bar{r}_{jk} - \bar{n}_{jk} P_j(X_k)]\, w_{jk} \quad , \tag{25}$$

and

$$c_j :\ (1 - c_j)^{-1} \sum_{k}^{q} \frac{[\bar{r}_{jk} - \bar{n}_{jk} P_j(X_k)]}{P_j(X_k)} \quad . \tag{26}$$

These equations are set to 0 and the parameters for a single item estimated simultaneously by the Fisher scoring-for-parameters procedure within the context of an EM algorithm (see Harwell et al., 1988, for the implementation details). However, in a Bayesian solution, the components corresponding to the prior probability distributions (the second term in Equation 12) must be appended to the marginal likelihood components. These terms are derived below.

The Role of Prior Distributions in PC-BILOG in Estimating Item Parameters

A number of authors have recommended that informative priors be used in estimating certain item parameters so that reasonable estimates may be obtained (Mislevy, 1986; Mislevy & Stocking, 1989; Swaminathan & Gifford, 1985, 1986). The prior distributions and default hyperparameter values used in PC-BILOG are:

α_j: normal $(\mu_\alpha = 0,\ \sigma_\alpha = .5)$,

b_j: normal $(\mu_b = 0,\ \sigma_b = 2)$,

c_j: beta $(\alpha_\beta = 5,\ \beta_\beta = 17)$. (27)

For c_j, the values assume a dichotomously scored item with five multiple-choice options; for other scoring schemes the default α_β and β_β values of the beta distribution will change. The prior distributions for the discrimination and "guessing" parameters are discussed below. Expressions for the difficulty parameter follow directly from those for discrimination.

Bayes Modal Estimation Equations for Discrimination and Difficulty Parameters

PC-BILOG assumes that each a_j has a lognormal prior distribution over the range $0 \leq a_j \leq \infty$. Theoretical justification for the use of a lognormal prior rests on the fact that in most testing settings the a_j are typically greater than 0, suggesting that the distribution of the a_j can be modeled by a unimodal and positively skewed distribution such as the lognormal (Mislevy, 1986). The transformation $\alpha_j = \log a_j$ results in each α_j having a normal prior distribution with a density that is proportional to $\exp\{-.5[(\alpha_j - \mu_\alpha)/\sigma_\alpha]^2\}$. The normal prior distribution of each α_j is defined by its hyperparameters, μ_α and σ_α, which are assigned default values of 0 and .5, respectively, by PC-BILOG. A program user can specify the values of μ_α and σ_α to be employed. Alternatively, the sample item response data may be used to estimate μ_α through the FLOAT option; however, σ_α cannot be estimated from the data.

Because the antilog of the mean of α_j is not the mean of a_j, it is useful to consider the mean and variance of the lognormal distribution of a_j expressed in terms of the transformed discrimination indices, α_j. Aitchison and Brown (1957, p. 8) provide the following:

$$\mu_a = \exp(\mu_\alpha + .5\,\sigma_\alpha^2) \tag{28}$$

and

$$\sigma_a^2 = \exp(2\mu_\alpha + \sigma_\alpha^2)[\exp(\sigma_\alpha^2) - 1] \quad . \tag{29}$$

For example, the default values of $\mu_\alpha = 0$ and $\sigma_\alpha = .5$ in PC-BILOG result in $\mu_a = 1.13$ and $\sigma_a = .6$.

To examine the effect of a normal prior on estimating α_j, recall that, under a Bayesian approach, Equation 12 results in appending the likelihood component for α_j (Equation 17) with a partial derivative term that is associated with the joint prior distribution (across items) of α_j. This means that the prior information associated with all J items is used in estimating α_j. Under the assumption of independently and identically distributed parameters, the joint prior distribution of each type of item parameter can be examined (see Swaminathan & Gifford, 1986).

Let $\boldsymbol{\alpha}$ represent a $J \times 1$ vector of transformed discrimination parameters. Then, the portion of the expression $\{\log[g(\xi)]\}$ in Equation 12 that pertains to discrimination parameters can be written as $\{\log[g(\boldsymbol{\alpha})]\}$, which is the logarithm of the joint prior probability distribution of J transformed discrimination parameters. The appended term in Equation 12 is the partial derivative of this expression taken with respect to the discrimination parameter for the jth item. Because α_j is normally distributed, the derivative is

$$\frac{\partial}{\partial \alpha_j}\{\log[g(\boldsymbol{\alpha})]\} = \frac{\partial}{\partial \alpha_j}\left[-.5\left(\frac{\alpha_j - \mu_\alpha}{\sigma_\alpha}\right)^2\right] = -(\alpha_j - \mu_\alpha)/\sigma_\alpha^2 \tag{30}$$

[see Equation 24 in Mislevy (1986)].

Appending the likelihood component in Equation 24 with that of Equation 30 results in the Bayes modal estimation equation for α_j used in PC-BILOG:

$$\frac{\partial}{\partial \alpha_j}\{\log[L(\mathbf{Y}|\xi,\boldsymbol{\tau})\,g(\xi)]\} = e_j^{\alpha}(1 - c_j)\sum_{k}^{q}[\bar{r}_{jk} - \bar{n}_{jk}P_j(X_k)]\,w_{jk}(X_k - b_j) - (\alpha_j - \mu_\alpha)/\sigma_\alpha^2 = 0 \quad . \tag{31}$$

The transformation $\alpha_j = \log a_j$ was used earlier without justification. This transformation is convenient because it keeps the metric of the discrimination parameter the same in both components of the Bayes modal estimation equations. Because a normal prior is used for the item difficulty parameter, b_j, the Bayes modal estimation equation for b_j can be written from Equation 31:

$$b_j : -e_j^{\alpha}(1 - c_j)\sum_{k}^{q}[\bar{r}_{jk} - \bar{n}_{jk}P_j(X_k)]w_{jk} - (b_j - \mu_b)/\sigma_b^2 = 0 \quad . \tag{32}$$

Prior distributions supplement the information contained in the sample data; therefore, if the prior distribution is informative, the appending term should have an impact on item parameter estimation. As noted above, this is accomplished through the Bayesian concept of "shrinkage." The contribution of a prior distribution to the solution equation depends on the amount of "shrinkage" of the estimated item parameter toward the mean, say μ_α, of its prior distribution. The amount of "shrinkage" is associated with the loss function $(\alpha_j - \mu_\alpha)$ and with the size of σ_α (see Novick & Jackson, 1974, pp. 3–15).

Other things being equal, the more similar α_j and μ_α are, the smaller the loss and the less the shrinkage. Greater shrinkage occurs with estimates of α_j that differ substantially from μ_α. This tends to restrain estimates from assuming unreasonable values. Suppose, for example, that $\sigma_\alpha = .5$ and $\mu_\alpha = 0$. If the estimated $\alpha_j = 5$, the contribution of the appending term in Equation 31 (i.e., the weight of the prior) is –20; if $\alpha_j = 2$, the contribution is –8; for $\alpha_j = \mu_\alpha = 0$, the appending term makes no contribution and there is no shrinkage.

These examples illustrate that, other things being equal, the closer the estimated α_j is to the mean of its prior distribution, the less impact the prior has on the estimation process. These examples also help illustrate the key role that σ_α plays. A noninformative prior distribution would tend to have a large variance and would reduce the value of $(\alpha_j - \mu_\alpha)/\sigma_\alpha^2$ to a small contribution, whereas an informative prior would tend to have a small variance and, other things being equal, the $(\alpha_j - \mu_\alpha)/\sigma_\alpha^2$ term would make a larger contribution to estimating the item parameter.

It is important to emphasize that an informative prior is not necessarily an appropriate prior. For example, a user-specified value of μ_α or a default value could differ considerably from the true underlying mean discrimination. When combined with a small σ_α, the result would be to pull the α_j estimate toward an inappropriate mean of the prior distribution of discrimination—clearly an undesirable effect. Mislevy (1986) pointed out that incorrectly specifying μ_α is likely to result in an "ensemble bias." This means that the estimated α_js will be biased, and statistical properties like consistency are unlikely to hold. The same holds true for estimates of other item parameters. Mislevy and Stocking (1989) urge PC-BILOG users to avail themselves of the diagnostic features of PC-BILOG to check on the correctness of the IRT model and the prior distributions to minimize the possibility of ensemble bias.

The possibility of incorrectly specifying the mean of a prior distribution can sometimes be lessened by using the FLOAT option in PC-BILOG. Under this option, the hyperparameter μ_α is estimated from the sample's item response data. The formula for estimating μ_α is obtained by finding the partial derivative of the densities $g(\boldsymbol{\alpha}|\mu_\alpha,\sigma_\alpha)$ and $g(\mu_\alpha|\lambda_\alpha,\varsigma_\alpha)$ with respect to μ_α, setting the resulting equation equal to 0, and solving for μ_α. The latter density arises when μ_α is treated as a continuous (normally distributed) random variable with mean λ_α and variance ς_α (see Anderson, 1984, p. 272). The expression for μ_α is a weighted average of the mean of the J estimated (transformed) discrimination parameters $(\bar{\alpha})$ and λ_α:

$$\mu_\alpha = \frac{J\bar{\alpha} + d\lambda_\alpha}{J + d} \quad . \tag{33}$$

The scalar d is the weight or believability of the prior information, expressed in terms of the number of items that the prior information is considered to be worth (see Equation 26 in Mislevy, 1986).

When the number of items in a test is large relative to the weight allocated to the mean of a prior distribution, the estimate of μ_α in Equation 33 will be close to the mean of the estimates of the α_j obtained in the previous cycle of the parameter estimation algorithm. When the number of items is not large and/or the prior information is weighted more heavily through d, the effect of λ_α in Equation 33 will be greater. In PC-BILOG, the value of d is set to 0 and cannot be changed by the user. Thus, the FLOAT option estimates μ_α as the average of the J sample item discriminations in a test. When the FLOAT option is selected, the estimate of μ_α obtained through Equation 33 is employed in the appended term of the Bayes modal estimation (Equation 31). The estimated value of μ_α is then used as the mean of the prior distribution of each item in the test. If the FLOAT option is not selected, the μ_α used in Equation 31 is either specified by the user or assigned the program default value.

An additional comment concerning the estimation of μ_α from the sample item response data is in order. Mislevy and Stocking (1989) recommended the use of the FLOAT option in PC-BILOG, unless the values of the prior distribution's hyperparameters are appropriate. However, the obtained value of μ_α may still be inappropriate because there is no guarantee that the sample data will accurately estimate μ_α. Also, the PC-BILOG manual recommends that the FLOAT option not be used for small datasets as it can induce "item parameter drift" and the EM algorithm will not result in convergence. Unfortunately, exhaustive comparisons of the effects of the FLOAT option on parameter estimation across varying number of items and examinees are not yet available. For some initial results comparing the use of the FLOAT option and the default value of μ_α, see Baker (1990). It seems safe to conclude that considerable computer simulation work needs to be done before questions of this nature can be adequately addressed.

Bayes Modal Estimation Equations for Guessing Parameters

Consider the role of a prior distribution when estimating "guessing" parameters. Because c_j is bounded by 0 and 1, a beta prior distribution was proposed by Swaminathan and Gifford (1986) and was incorporated into PC-BILOG. A beta distribution is bounded by 0 and 1 and, depending on the values of the parameters, can assume a variety of shapes. The two parameters of a beta distribution are given in PC-BILOG as α_β and β_β (the subscripts were added here to differentiate from the item parameters). These parameters are defined as $\alpha_\beta = mp + 1$ and $\beta_\beta = m(1 - p) + 1$, in which m is an a priori weight assigned to the prior information and p is the mean of the beta prior distribution. Swaminathan and Gifford (1986) indicated that weights of $m = 15$ to 20 are preferred—by default, PC-BILOG assigns $m = 20$.

The use of a beta prior for guessing parameters revolves around interpreting the mean p as the probability that a low θ examinee will respond correctly to an item. The idea is to specify α_β and β_β values that result in the desired p value. This is achieved through the relationship $p = (\alpha_\beta - 1)/(\alpha_\beta + \beta_\beta - 2)$.

Consider the three-parameter logistic IRT model. By default, PC-BILOG assumes that the number of alternatives (NALT) for each item is 5 and that $p = 1/\text{NALT} = 1/5 = .2$. The default values are then $\alpha_\beta = 20(.2) + 1 = 5$ and $\beta_\beta = 20(1 - .2) + 1 = 17$, from which $p = .2$ is obtained [i.e., $.2 = (5 - 1)/(5 + 17 - 2)$]. If $p = .15$ is specified with an a priori weight of $m = 26$, then $\alpha_\beta = (.26)(.15) + 1 = 4.90 = 5$ and $\beta_\beta = (26)(1 - .15) + 1 = 23.1 = 23$. Thus $p = (5 - 1)/(5 + 23 - 2) = .15$.

The assigned values of m also influence the credibility intervals (akin to confidence intervals) constructed about c_j (see Swaminathan & Gifford, 1986; Novick & Jackson, 1974, p. 119). These intervals take into account prior information and provide a measure of the strength of the belief that

c_j lies within a range of values. Other things being equal, larger a priori weights lead to narrower intervals. This suggests that credibility intervals can be used in lieu of specifying the variance of a prior distribution for "guessing" parameters. These intervals are tabled in Novick and Jackson (1974, pp. 402–409), which should be consulted for additional information.

Generation of the Bayes modal estimation equation for c_j requires that the likelihood component in Equation 26 be appended with a term based on the second term in Equation 12. Let **c** represent a $J \times 1$ vector of guessing parameters, and write the expression $\{\log[g(\xi)]\}$ in terms of **c** as $\{\log[g(\mathbf{c})]\}$. Assuming that the "guessing" parameters are independently and identically distributed, the partial derivative with respect to c_j is

$$\frac{\partial}{\partial c_j}\{\log[g(c)]\} = \frac{\partial}{\partial c_j}\sum_j^J[(\alpha_\beta - 2)\log c_j + (\beta_\beta - 2)\log(1 - c_j)] = (\alpha_\beta - 2)/c_j - (\beta_\beta - 2)/(1 - c_j) \qquad (34)$$

(see Swaminathan & Gifford, 1986). Appending Equation 26 with Equation 34 yields the Bayes modal estimation equation for c_j that is employed in PC-BILOG:

$$c_j : \ (1 - c_j)^{-1}\sum_k^q \frac{[\bar{r}_{jk} - \bar{n}_{jk}P_j(X_k)]}{P_j(X_k)} + [(\alpha_\beta - 2)/c_j - (\beta_\beta - 2)/(1 - c_j)] = 0 \quad . \qquad (35)$$

The role of the appending term in Equation 35 is similar to that in the estimation of discrimination and difficulty.

The set of marginalized Bayesian modal estimation equations defined by Equations 31, 32, and 35 are solved simultaneously for each item to obtain the item parameter estimates. This is accomplished within the framework of an EM algorithm (Bock & Aitkin, 1981). In the E step of the first cycle of the algorithm, the "artificial data" are computed using a priori values for the item parameters. In the M step, the Bayesian modal estimates of the item parameters are computed on a one-item-at-a-time basis using the "artificial data." The equations are nonlinear in the parameters and the Fisher scoring-for-parameters method is used within the M step of the EM algorithm in PC-BILOG to obtain the item parameter estimates. These item parameter estimates are used in the E phase of the second EM cycle. This process is repeated until a suitable convergence criterion is met. (An example of implementing Mislevy's marginalized Bayesian approach in BASIC for a two-parameter logistic model is available from the authors.)

Summary

The application of Bayesian statistical procedures in IRT provides a powerful yet flexible method for estimating item parameters. The Bayesian estimation procedure proposed by Mislevy (1986) permits prior probability distributions to be posited for item parameters, and under the assumption that the prior probability distributions and the IRT model are correct, produces estimates that possess desirable properties, such as consistency. These estimates are also likely to assume reasonable values. The availability of this estimation procedure in PC-BILOG makes it possible for users to take full advantage of the Bayesian approach. However, the complexity of the mathematics underlying Bayesian item parameter estimation makes understanding the procedure difficult. The key to understanding the procedure and its implementation in PC-BILOG is the role of prior distributions in estimating item parameters. Essentially, the original MMLE equations, which do not posit prior probability distributions for item parameters, are appended by expressions based on these priors. An examination of these appending terms clarifies the impact of informative versus noninformative prior distributions on item parameter estimation and the role that the hyperparameters of these prior distributions play.

Little is known about the quality of item parameter estimates produced by PC-BILOG for small datasets or the impact of prior distributions other than those used in PC-BILOG on item parameter estimation. Computer simulation studies of the effect of these factors on item parameter estimation in PC-BILOG are needed to provide additional guidelines for the use of the Mislevy estimation model.

References

Aitchison, S., & Brown, J. (1957). *The lognormal distribution.* Cambridge, England: Cambridge University Press.

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

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis (2nd ed.).* New York: Wiley.

Baker, F. B. (1990). Some observations on the metric of PC-BILOG results. *Applied Psychological Measurement, 14,* 139–150.

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. 397–479). Reading MA: Addison-Wesley.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application 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.

Cornfield, J. (1969). The Bayesian outlook and its applications. *Biometrics, 25,* 617–657.

de Finetti, B. (1974). Bayesianism: Its unifying role for both the foundations and applications of statistics. *International Statistical Review, 42,* 117–130.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242.

Harwell, M. R., Baker, F. B., & Zwarts, M. (1988). Item parameter estimation via marginal maximum likelihood and an EM algorithm: A didactic. *Journal of Educational Statistics, 13,* 243–271.

Lindley, D. V. (1970a). *Introduction to probability and statistics from a Bayesian viewpoint. Part 1. Probability.* Cambridge, England: Cambridge University Press.

Lindley, D. V. (1970b). *Introduction to probability and statistics from a Bayesian viewpoint. Part 2. Inference.* Cambridge, England: Cambridge University Press.

Lindley, D. V. (1971). *Bayesian statistics: A review.* Regional Conference Series in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes 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 NJ: Erlbaum.

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

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49,* 359–381.

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51,* 177–194.

Mislevy, R. J. (1988). Exploiting auxiliary information about items in the estimation of Rasch item difficulty parameters. *Applied Psychological Measurement, 12,* 281–296.

Mislevy, R. J., & Bock, R. D. (1985). Implementation of the EM algorithm in the estimation of item parameters: The BILOG computer program. In D. J. Weiss (Ed.), *Proceedings of the 1982 item response theory and computerized adaptive testing conference* (pp. 189–202). Minneapolis MN: University of Minnesota, Department of Psychology, Computerized Adaptive Testing Laboratory.

Mislevy, R. J., & Bock, R. D. (1986). *PC-BILOG: Item analysis and test scoring with binary logistic models.* Mooresville IN: Scientific Software Inc.

Mislevy, R. J., & Stocking, M. L. (1989). A consumer's guide to LOGIST and BILOG. *Applied Psychological Measurement, 13,* 57–75.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica, 16,* 1–32.

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

Novick, M. R., Jackson, P. H., Thayer, D. T., & Cole, N. S. (1972). Estimating multiple regressions in *m* groups: A cross-validational study. *British Journal of Mathematical and Statistical Psychology, 5,* 33–50.

O'Hagan, A. (1976). On posterior joint and marginal modes. *Biometrika, 63,* 329–333.

Rubin, D. B. (1980). Using empirical Bayes techniques in the law school validity studies. *Journal of the American Statistical Society, 75,* 801–827.

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.

Swaminathan, H., & Gifford, J. A. (1986). Bayesian estimation in the three-parameter logistic model. *Psychometrika, 51,* 589-601.

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

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). *LOGIST user's guide.* Princeton NJ: Educational Testing Service.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement, 14,* 97-116.

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

Acknowledgments

The authors are indebted to Seock-Ho Kim for assistance in preparing the paper and to two anonymous reviewers and the editor for their helpful comments.

Author's Address

Send requests for reprints or further information to Michael R. Harwell, 5B27 Forbes Quad, University of Pittsburgh, Pittsburgh PA 15260, U.S.A.