STATISTICAL INFERENCE FOR MULTIPLE CHOICE TESTS

JOHN S.J. HSU

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY
THE UNIVERSITY OF CALIFORNIA, SANTA BARBARA

TOM LEONARD

KAM-WAH TSUI

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF WISCONSIN, MADISON

Finite sample inference procedures are considered for analyzing the observed scores on a multiple choice test with several items, where, for example, the items are dissimilar, or the item responses are correlated. A discrete p -parameter exponential family model leads to a generalized linear model framework and, in a special case, a convenient regression of true score upon observed score. Techniques based upon the likelihood function, Akaike's information criteria (AIC), an approximate Bayesian marginalization procedure based on conditional maximization (BCM), and simulations for exact posterior densities (importance sampling) are used to facilitate finite sample investigations of the average true score, individual true scores, and various probabilities of interest. A simulation study suggests that, when the examinees come from two different populations, the exponential family can adequately generalize Duncan's beta-binomial model. Extensions to regression models, the classical test theory model, and empirical Bayes estimation problems are mentioned. The Duncan, Keats, and Matsumura data sets are used to illustrate potential advantages and flexibility of the exponential family model, and the BCM technique.

Key words: multiple choice test, exponential family, likelihood, Akaike's information criterion, generalized linear model, Bayesian marginalization, importance sampling, regression of true score upon observed score, classical test theory model.

1. Introduction

We consider a problem highlighted by Lord and Novick (1968, p. 524), Duncan (1974), and Morrison and Brockway (1979). Suppose that an examinee completes a multiple choice test with m items and obtains y correct responses out of m . Then a number of appealing choices, such as the compound binomial error model (Lord & Novick, p. 384), are available for modeling the random variability of y . Note that

$$y = x_1 + \cdots + x_m, \quad (1)$$

where $x_1, \dots, x_m$ are binary responses, with $x_i = 1$ if the examinee correctly answers the i -th item, and $x_i = 0$ otherwise. One approach is to first model the joint probability mass function $p(x_1, x_2, \dots, x_m)$ of $x_1, \dots, x_m$. Several possibilities are

i. *The beta-binomial model* (Duncan, 1974; Leonard & Novick, 1986; Prentice &

The authors wish to thank Ella Mae Matsumura for her data set and helpful comments, Frank Baker for his advice on item response theory, Hirotugu Akaike and Taskin Atilgan, for helpful discussions regarding AIC, Graham Wood for his advice concerning the class of all binomial mixture models, Yiu Ming Chiu for providing useful references and information on tetrachoric models, and the Editor and two referees for suggesting several references and alternative approaches.

Requests for reprints should be sent to John S.J. Hsu, Department of Statistics and Applied Probability, University of California-Santa Barbara, Santa Barbara, CA 93106.

Barlow, 1988). Suppose that the items are similar or parallel and that $x_1, \dots, x_m$ are permutable. Given ξ , take $x_1, \dots, x_m$ to be independent and binary distributed with common probability ξ . Then, if ξ possesses a beta distribution, y in (1) possesses a beta-binomial distribution. Therefore, a mixture of binomial distributions for y can always be justified by a permutable (i.e., exchangeable) distribution for $x_1, \dots, x_m$. The beta-binomial model may also be reasonable for y in some situations where $x_1, \dots, x_m$ are not permutable.

ii. *A modified beta-binomial model* (Morrison & Brockway, 1979). In model i, let $\xi = \xi^* + c(1 - \xi^*)$, where c is a guessing parameter and ξ^* possesses a beta distribution. Then the distribution of y may be expressed as a finite series.

iii. *A multivariate logistic/binary distribution* (Anderson & Aitken, 1985; Leonard, 1972). Suppose that, given $\xi_1, \dots, \xi_m$, the random variables $x_1, \dots, x_m$ possess binary distributions with respective probabilities $\xi_1, \dots, \xi_m$, and consider the logits $\zeta_i = \log \xi_i - \log(1 - \xi_i)$. Take $\zeta_1, \dots, \zeta_m$ to possess a multivariate normal distribution. Then appropriate choices of the covariances introduce flexible interdependencies between $x_1, \dots, x_m$. Also, unequal means or unequal variances permit the modeling of dissimilar items. The true score may be defined as $\xi_1 + \dots + \xi_m$, and it is in principle possible to predict the true score, for any particular examinee. However, it is difficult to express the distribution of y in closed form, and any analysis would require heavy computations.

iv. *General mixtures* (Bock & Aitken, 1981; Hsu, 1990; Leonard, 1984; Lord, 1969; Lord & Novick, 1988, p. 512; Lord & Stocking, 1976). In Model i, take ξ to possess cumulative distribution function G where, for example, G may be estimated as a discrete distribution, using generalized maximum likelihood procedures.

If the user is less interested in modeling the joint distribution of $x_1, \dots, x_m$, a more direct approach may be developed by assuming the distribution of y to belong to a family that is broad enough to approximate, as special cases, the compound binomial error model, the beta-binomial model, the modified beta-binomial model, the generalized hypergeometric model (Keats, 1964), the gamma-Poisson model (Wilcox, 1981a, 1981b), the four-parameter beta-binomial model (Lord, 1965), the beta-binomial mixture model (Bell, 1990; Dalal & Hall, 1983), and the distributions of y in Models iii and iv. Other generalizations of the simple binomial model were discussed by Consul (1974, 1975), Altham (1978), Lord (1965), and Carter and Williford (1975). For a review, see Wilcox (1981a). A very general specification of discrete distributions is provided by the discrete p -parameter exponential family (e.g., Leonard, Hsu, & Tsui, 1989; Takane, Bozdogan, & Shibayama, 1987), which assumes that $\phi_j \equiv p(y = j)$ satisfies

$$\phi_j = \frac{e^{\gamma_j}}{\sum_{h=0}^m e^{\gamma_h}}, \quad (j = 0, 1, \dots, m), \quad (2)$$

where the multivariate logits $\gamma_0, \gamma_1, \dots, \gamma_m$ are linear functions,

$$\gamma_j = a_j + \theta_1 b_1(j) + \theta_2 b_2(j) + \dots + \theta_p b_p(j), \quad (3)$$

of specified basic functions $a_j, b_1(j), b_2(j), \dots, b_p(j)$; $\theta_1, \dots, \theta_p$ are unknown parameters, and $p < m$ is either specified or unknown. We will proceed under the particular assumptions

$$a_j = \log^m C_j, \quad (j = 0, 1, \dots, m), \quad (4)$$

where ${}^m C_j = m!j!(m-j)!$, and

$$b_k(j) = \left(\frac{j}{m}\right)^k, \quad (k = 1, \dots, p; j = 0, 1, \dots, m), \quad (5)$$

so that we have a polynomial model for the logits, in the spirit of Bock (1972) and others. When $p = 1$, y possesses a binomial distribution with probability $e^{\theta_1}/(1 + e^{\theta_1})$ and sample size m . However, a larger value of p permits substantial deviations from the binomial distribution. When $p = 2$, a two parameter distribution, similar in quality to the beta-binomial distribution is obtained. However, another reasonable choice is $p = 4$, since the quartic function in (3) then permits bimodality of the probability mass function of y , and hence, the potential for a reasonable fit to many data sets. An appropriately large choice of p will enable us to approximate many previously proposed procedures for modeling the distribution of y .

The polynomials in (5) can be replaced by a variety of other convenient choices of basis function, such as B-splines or terms from a Fourier Series. The usual p -dimensional minimal sufficient statistic, when p is fixed, can always be expressed as the vector of the corresponding sample quantities, and we can refer to standard exponential family theory for estimation (e.g., Lehmann, 1983, p. 80). For example, for fixed p , there always exist sufficient statistics that are uniform minimum variance unbiased estimators of their expectations, a finite sample property.

The model in (2) through (5) will permit the prediction of an average true score, across a set of examinees. As it is not described in two stages, it does not in general facilitate the prediction of individual true scores. When the items are dissimilar, a model of the complexity of Model iii or the tetrachoric correlation model (Lord & Novick, p. 345), would be required to express individual true scores as sums of m components corresponding to the different items. Nevertheless, in the special case where the model in (2) through (5) approximates a model of the form of Model iv for some cumulative distribution function G , it is possible to predict individual true scores. This key property is described below in comment (iv) and (12).

Suppose that n examinees respectively score $y_1, \dots, y_n$ correct responses out of m . Then, if $y_1, \dots, y_n$ are independent, the likelihood of $\theta' = (\theta_1, \dots, \theta_p)$, given $y' = (y_1, \dots, y_n)$, under assumptions (2) through (5), is

$$l(\theta_1, \dots, \theta_p; p|y) = \exp \{t_0 + \theta_1 t_1 + \dots + \theta_p t_p - n\Lambda(\theta)\}, \quad (6)$$

where

$$t_0 = \sum_{j=0}^m a_j n_j, \quad (7)$$

and the (fixed p) minimal sufficient statistics $t_1, \dots, t_p$ satisfy

$$t_k = \sum_{j=1}^m n_j \left(\frac{j}{m}\right)^k, \quad (k = 1, \dots, p), \quad (8)$$

with n_j representing the number of examinees with score j , and

$$\Lambda(\theta) = \log \sum_{j=0}^m \exp \{a_j + m^{-1}\theta_1 j + m^{-2}\theta_2 j^2 + \dots + m^{-p}\theta_p j^p\}. \quad (9)$$

The explicit parameters in the exponential family model are not always meaningful in their own right, since the model is motivated toward a reasonable semi-parametric fit to the data. However, many parameters of interest can be expressed as functions of the model parameters, and our objective is to make finite sample inferences about the parameters of interest. In section 2 we will consider maximum likelihood procedures for $\theta_1, \dots, \theta_p$ together with an information criterion (e.g., Akaike, 1978; Atilgan & Leonard, 1988). In section 4 we will discuss more precise finite sample inference for parameters of interest that are arbitrary functions of $\theta_1, \dots, \theta_p$, for example:

(i). *The sample mean or average true score.*

$$\begin{aligned}\eta &= \frac{\sum_{j=1}^m j\phi_j}{\sum_{j=1}^m \phi_j} \\ &= \frac{\sum_{j=1}^m j \exp \{a_j + m^{-1}\theta_1 j + \dots + m^{-p}\theta_p j^p - \Lambda(\boldsymbol{\theta})\}}{\sum_{j=1}^m \exp \{a_j + m^{-1}\theta_1 j + \dots + m^{-p}\theta_p j^p - \Lambda(\boldsymbol{\theta})\}}.\end{aligned}\quad (10)$$

(ii). *The probability of passing.*

$$\begin{aligned}\eta &= p(y \geq a) \\ &= \sum_{j=a}^m \phi_j \\ &= \sum_{j=a}^m \exp \{a_j + m^{-1}\theta_1 j + \dots + m^{-p}\theta_p j^p - \Lambda(\boldsymbol{\theta})\},\end{aligned}\quad (11)$$

where a denotes the passing score.

(iii). *Comparison of two groups.* We will consider the probability that the score y_1 of a randomly chosen examinee from one group exceeds the score y_2 of a randomly chosen person from another group.

(iv). *Individual true scores.* Whenever the model in (2) through (5) is equivalent to a general mixture of the form described in Model iv (the results by Lord, 1969, tell us that this may sometimes, but not always, be exactly true or approximately true), the conditional expectation of the logit $\gamma = \log \xi - \log(1 - \xi)$, the binomial mixture model, given y , is

$$\begin{aligned}\gamma^* &= E(\gamma|y) = \frac{\partial \log \phi_y}{\partial y} + \psi(y+1) - \psi(m-y+1) \\ &= \theta_1 m^{-1} + 2\theta_2 m^{-2}y + 3\theta_3 m^{-3}y^2 + \dots + p\theta_p m^{-p}y^{p-1},\end{aligned}\quad (12)$$

where $\psi(y) = \partial \log \Gamma(y)/\partial y$ denotes the digamma function. The digamma terms conveniently vanish because of assumption (4). Hence, a particular examinee's true ability, given observed score, can be predicted by $\xi^* = e^{\gamma^*}/(1 + e^{\gamma^*})$, where γ^* is the linear function of $\theta_1, \dots, \theta_p$ defined in (12). Then ξ^* is much more general than the linear regression of ξ on y predicted by the beta-binomial model. Although Lord (1969) indicates that, for any observed score distribution, the true score distribution is not uniquely defined, the value of ξ^* will be unaffected by this choice. Equation (12) provides an alternative to the regression of true score on observed score, considered by

Lord and Novick (1968, p. 513). The predictive variance of y , given y , is just the first derivative with respect to y of (12) and this will be positive whenever the two models are identical. In this case, the first p predictive moments may be obtained by further differentiations, and predictive central moments higher than the first p are zero. The full result in (12) seems new; see Appendix 1 for details of our derivation, and section 8 for an analogue within the framework of the classical test theory model.

Until recently, the nonlinear functions in (10) and (11) would have created severe technical difficulties. However, these will be circumvented via recent developments in Bayesian marginalization (e.g., Geweke, 1988, 1989; Leonard, 1982; and Leonard, Hsu, & Tsui, 1989), which permit a precise finite sample analysis of complicated nonlinear models with several parameters.

2. Likelihood Techniques

When p is specified, the likelihood (6) may be analyzed within the standard framework of generalized linear models (McCullagh & Nelder, 1985, p. 142). The maximum likelihood estimates $\hat{\theta}_1, \dots, \hat{\theta}_p$ of $\theta_1, \dots, \theta_p$, for fixed p , satisfy the equations

$$t_k = n \sum_{j=1}^m \hat{\phi}_j \left(\frac{j}{m} \right)^k, \quad (k = 1, \dots, p), \quad (13)$$

where the t_k satisfy (8), and

$$\hat{\phi}_j = \exp \{ (a_j + m^{-1} \hat{\theta}_1 j + \dots + m^{-p} \hat{\theta}_p j^p - \Lambda(\hat{\theta})) \}, \quad (14)$$

and the likelihood information matrix $\mathbf{R}$ of $\theta_1, \dots, \theta_p$ has elements r_{kk^*} satisfying

$$r_{kk^*} = n \sum_{j=1}^m \left(\frac{j}{m} \right)^{k+k^*} \hat{\phi}_j (1 - \hat{\phi}_j), \quad (k = 1, \dots, p; k^* = 1, \dots, p). \quad (15)$$

The matrix $\mathbf{R}$ also provides the Hessian in the Newton-Raphson iterations for the maximization of the log of the likelihood (6) and the existence of a finite solution is guaranteed for nonzero $t_1, \dots, t_p$. Then $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)'$ yields consistent point estimation for $\theta = (\theta_1, \dots, \theta_p)'$ as $n \rightarrow \infty$, with p fixed, and for finite n , (13) provides uniform minimum variance unbiased estimators for p linear functions of $\phi_1, \dots, \phi_m$. Also, as $n \rightarrow \infty$, with p fixed, $\mathbf{R}^{1/2}(\hat{\theta} - \theta)$ converges in distribution to a standard multivariate normal distribution. Hence, the likelihood dispersion matrix $\mathbf{R}^{-1}$ may be regarded as a large sample estimated covariance matrix for $\hat{\theta}$, and approximate confidence intervals for linear combinations of θ are available.

When p is unspecified, one possibility is to choose p to maximize the general information criterion (GIC)

$$\text{GIC} = L_p - \frac{1}{2} \alpha p, \quad (16)$$

where L_p is the log of the likelihood (6), evaluated at the maximum likelihood estimates, and α denotes a penalty per parameter included in the model. Various choices of α have been proposed (e.g., Akaike, 1978, Atilgan, 1983, and Schwarz, 1978). For example, $\alpha = 2$, and $\alpha = \log n$ respectively lead to Akaike's information criterion

$$\text{AIC} = L_p - p, \quad (17)$$

and Schwarz's information criterion

$$\text{BIC} = L_p - \frac{1}{2}p \log n. \quad (18)$$

Alternatively, if L_p is plotted as a function of p , then the graph typically rises rapidly for small p , and then flattens into a slightly increasing ridge. The problem is to estimate a sensible value for p near the beginning of the ridge. One possibility is Atilgan's empirical information criterion (EIC) that considers the vertices $p_1, p_2, \dots$ of the convex boundary of the plot of L_p against p , and then chooses the vertex minimizing $\gamma_i p_i$, where γ_i is the slope of the convex boundary after the i -th vertex. Atilgan performs simulations to show that AIC and the data dependent EIC typically perform equally well, and somewhat better than other choices of α . As a preliminary to the Bayesian analysis of Section 4, we recommend evaluating p by AIC, since a mixed model analysis is much more complicated. Akaike (1978) provides a Bayesian justification for this preliminary step, by assuming a particular prior distribution for p .

These ideas are easy to extend to further regression situations, thus providing an alternative to the beta-binomial developments of Prentice and Barlow (1988). Suppose we have n observed scores $y_1, \dots, y_n$ with corresponding design points $u_1, \dots, u_n$. Then let the i -th observation possess the distribution defined in (2) through (5), but with θ_1 replaced by, say $\beta_0 + \beta_1 u_i$. The $p + 1$ parameters $\beta_0, \beta_1, \theta_2, \dots, \theta_p$ may be estimated within the generalized linear model framework, and $\theta_2, \dots, \theta_p$ measure the deviations from the standard linear logistic model for binomial data. The term $\theta_2 j^2 + \dots + \theta_p j^p$ can be interpreted as a "parametric residual" in the sense of Leonard and Novick (1986). A more complicated regression component, or dependence of $\theta_2, \dots, \theta_p$ on the design points, can also be considered.

3. Simulation Results

Five hundred simulations were first performed, where on each simulation $n = 200$ observations were generated from a mixture $\frac{1}{3} B(0.4, 10) + \frac{2}{3} B(0.6, 10)$ of two binomial distributions, with mixing probabilities $\frac{1}{3}$ and $\frac{2}{3}$, binomial probabilities 0.4 and 0.6, and common sample size $m = 10$. On each simulation, we fitted a beta-binomial distribution, with parameters estimated by maximum likelihood, and the exponential family distribution in (2) through (5), with $p = 2$. The exponential family distribution fitted the data, with a lower value of the chi-squared goodness-of-fit statistic χ^2_8 , with eight degrees of freedom, on 349 out of the 500 simulations. The average value of χ^2_8 , under the exponential family model, was $\bar{\chi}^2_8 = 8.623$, compared with $\bar{\chi}^2_8 = 8.828$ under the beta-binomial model. The fitted exponential family model was closer to the true model, when compared with the beta-binomial model, on 424 out of the 500 simulations. Here, we measured the distance from the true model by

$$\text{DEVIANCE} = \sum_{j=0}^m \frac{(f_j - e_j)^2}{e_j}, \quad (19)$$

where e_j is the fitted frequency for the j -th cell, under the fitted model, and f_j family is the expected frequency under the true model. The average deviance of 2.517 for the exponential family model was less than the corresponding value of 2.761 for the beta-binomial model.

A further 500 simulations were performed, each generating $n = 200$ observations from the $0.4B(0.3, 10) + 0.6B(0.8, 10)$ mixture. On each simulation, a beta-binomial distribution was compared with the exponential family distribution, with $p = 4$. The latter yielded a much lower value for the normalized chi-squared statistic $\chi^2_6/6$, when compared with the statistic $\chi^2_8/8$, for the beta-binomial distribution, on each of the 500

simulations. The average $\bar{\chi}_6^2 = 8.604$ compared with an average $\bar{\chi}_8^2 = 41.501$ for the beta-binomial model.

The exponential family model was closer to the true model on each of the 500 simulations, when an average deviance of 6.376 compared with an average deviance of 34.486 for the beta-binomial model. Hence, there do exist situations where the generality of (2) through (5) may prove useful when compared with other models.

4. Bayesian Marginalization

In some situations it may be possible to incorporate prior information regarding $\theta = (\theta_1, \dots, \theta_p)'$ via a multivariate normal prior distribution, say with mean vector μ_θ and covariance matrix C_θ . For example, C_θ could be taken to represent the covariance structure of an autoregressive process in the spirit of Young (1977) and Leonard (1973). It may be valuable to incorporate prior information when p is large, to obtain smoother estimates, but this is less important when p is small, when the parameters are better estimated by the data. This formulation provides a flexible alternative to the prior mixture of beta distributions recommended by Dalal and Hall (1983) and Bell (1990) for binomial models. The posterior density of θ for fixed p , is, using the notation of section 1,

$$\pi(\theta|y) \propto \exp \{ \theta_1 t_1 + \dots + \theta_p t_p - n\Lambda(\theta) - \frac{1}{2}(\theta - \mu_\theta)' C_\theta^{-1}(\theta - \mu_\theta) \}, \quad (\theta \in R^p). \quad (20)$$

However, as $|C_\theta| \rightarrow \infty$, the prior information becomes vague, and quadratic term, within the exponential of (20), vanishes, and (20) becomes proportional to the likelihood (6). In general, we consider a parameter of interest

$$\eta = g(\theta) = g(\theta_1, \dots, \theta_p), \quad (\eta \in \Omega), \quad (21)$$

without further regularity conditions concerning g , where Ω denotes the set of possible values of η . Special cases include the average true score (10) and probability of success (11). The marginal posterior density of η is

$$\pi(\eta|y) = \lim_{\delta \rightarrow 0} \delta^{-1} \int_D \pi(\theta|y) d\theta, \quad (\eta \in \Omega), \quad (22)$$

where D denotes the region

$$D = D(\eta, \delta) = \{ \theta : \eta \leq g(\theta) \leq \eta + \delta \}, \quad (23)$$

and $\pi(\theta|y)$ satisfies (20). The integration in (22) is difficult to exactly perform, unless p is small. However, an approximation fully described by Leonard, Hsu, and Tsui (1989) referred to here as BCM (Bayesian conditional maximization) typically gives precise results (see Appendix 2 for details). BCM involves three components, including the maximized posterior density

$$\pi_M(\eta|y) = \sup_{\theta: g(\theta) = \eta} \pi(\theta|y), \quad (\eta \in \Omega), \quad (24)$$

which conditionally maximizes (20) subject to the constraint (21). For the functions in (10) and (11), it is straightforward to compute (24), and the conditional maximum θ_η , of θ , given η , using standard Newton-Raphson/Lagrange multiplier methods. However (24) does not generally approximate (22) as well as possible. The adjustment terms described by Leonard et al. (1989) each merit inclusion. A key adjustment term, for

finite samples, depends upon the density $f(\eta|\mu, C)$ of $\eta = g(\theta)$ when θ possesses a multivariate normal distribution with mean vector μ and covariance matrix C , where μ and C are particular functions of θ_η . For the functions in (10) and (11), further approximations to $f(\eta|\mu, C)$ are required.

For (10), f was approximated by taking η/m to possess a beta distribution with correct first two moments, obtained by computer simulation. For (11), a similar beta approximation was used for η . If instead $\eta = \phi_j$ for some fixed j , we have

$$\eta = \phi_j = \frac{1}{1 + \varepsilon_j(\theta)}, \quad (25)$$

where

$$\varepsilon_j(\theta) = \sum_{h \neq j} \exp \{a_h - a_j + m^{-1}\theta_1(h - j) + m^{-2}\theta_2(h^2 - j^2) + \cdots + m^{-p}\theta_p(h^p - j^p)\}. \quad (26)$$

In this case, the density of $\varepsilon_j(\theta)$ was first approximated by a log-normal density, with correct first moments, and the transformation (25) used to obtain our f contribution for η . Note that the moments of (26) can be obtained analytically if θ possesses a multivariate normal distribution. Secondly, a scale multiple of F -approximation was used, with correct first three moments, for $\varepsilon_j(\theta)$ and again referred to the transformation (25).

As an alternative to the BCM procedure, it is possible to simulate the exact posterior density of η in (22) using importance sampling (e.g., Geweke, 1988, 1989; Leonard et al., 1989; Rubinstein, 1981). In the present context we recommend simulating θ vectors from a multivariate normal distribution with mean vector $\bar{\theta}$ and covariance matrix D , where $\bar{\theta}$ unconditionally maximizes (20) and D^{-1} denotes the Hessian in the Newton-Raphson iterations for the unconditional maximum. As $|C_\theta| \rightarrow \infty$, $\bar{\theta}$ is just the maximum likelihood vector satisfying (13), and D^{-1} is just the likelihood information matrix R , whose elements satisfy (15). Then importance sampling (see Rubinstein, 1981, for details) permits simulation of the exact posterior density (22), even though the simulations for θ are from an approximate distribution. The Gibbs sampler (e.g., Gelfand & Smith, 1990) could be used to try to speed up the convergence, but this is not particularly necessary for the current straightforward model.

Hsu (1990) considers a discrete mixture of binomials as an alternative to our exponential family model. Unfortunately, the regularity conditions for BCM breakdown for the discrete binomial mixture model owing to multimodality of the likelihood function. Hsu develops a generalization of importance sampling (referred to as permutable Bayesian marginalization, PBM) for handling discrete mixtures. However, BCM for our exponential family model is much simpler when compared with PBM for the binomial mixture model. Note that BCM is also applicable to the models considered by Mislevy (1986), as an alternative to procedures based upon joint posterior modes.

5. The Duncan Data

We first analyze the Duncan data (Duncan, 1974, p. 55; Morrison & Brockway, 1979, p. 439) that consists of two random samples, from a discrete distribution for the observed number of successful responses on $m = 20$ items. For the likelihood analysis of section 2, AIC was minimized in both cases when $p = 2$, and EIC provided similar results. The maximum likelihood estimates of θ_1 and θ_2 , together with their approximate standard errors (in brackets), obtained via (15), are reported in Table 1, for the

TABLE 1
Maximum Likelihood Analysis for Duncan Data

	$\hat{\theta}_1$	$\hat{\theta}_2$	χ^2_{10}	AIC
Group 1	-17.78 (2.94)	22.94 (2.39)	2.35	-2545.93
Group 2	-18.90 (3.22)	23.77 (2.64)	12.76	-1960.03

two groups of examinees. Note that $\hat{\theta}_1$ and $\hat{\theta}_2$ are more than five standard errors from zero in each case, refuting a simple binomial model ($\theta_2 = 0$).

Our calculations of the chi-squared statistic on ten degrees of freedom is based on a similar grouping of the cells to the grouping employed by Morrison and Brockway. The values $\chi^2_{10} = 2.35$ and 12.76 are similar to the values of 2.01 and 12.56 obtained by Morrison and Brockway for Duncan's beta-binomial model, and the values $\chi^2_9 = 1.82$ and 12.82 obtained for their modified beta-binomial model. However, under our model, it is easier to obtain approximate confidence intervals by reference to the likelihood dispersion matrix $\mathbf{R}^{-1}$, where $\mathbf{R}$ has elements defined in (15). As our parameter space is a subset of logit space, there is good potential for approximate normality. For the two groups of the Duncan data, the likelihood dispersion matrices are

$$\mathbf{R}^{-1} = \begin{pmatrix} 8.616 & -6.925 \\ -6.925 & 5.714 \end{pmatrix}, \text{ and } \mathbf{R}^{-1} = \begin{pmatrix} 10.378 & -8.380 \\ -8.380 & 6.954 \end{pmatrix}.$$

In this case, (12) reduces to a linear regression of true score on observed score. Consider the predicted true score, in logit space,

$$\gamma^* = \frac{\theta_1 + 1.2\theta_2}{20}, \quad (27)$$

for an examinee who correctly answers $y = 12$ items out of 20. Then, in the two groups, γ^* has maximum likelihood estimates 0.4874 and 0.4812 , with approximate standard errors 0.0236 and 0.0264 , yielding approximate 95% confidence intervals $(0.4401, 0.5337)$ and $(0.4295, 0.5329)$ for γ^* . The corresponding maximum likelihood estimator of $\xi^* = e^{\gamma^*}$ are 0.6195 and 0.6180 , quite close to the observed score of 0.6 , with reassuringly narrow 95% intervals $(0.6083, 0.6303)$ and $(0.6058, 0.6301)$.

In this case, the predictive variance of γ given y is $\theta_2/200$ and this has maximum likelihood estimates 0.1147 and 0.1189 in the two groups, with standard errors 0.012 and 0.0131 . More precise and detailed inferences are available under the Bayesian paradigm (see Figures 1–6). Only figures for the first group of examinees are reported here; similar results are available for the second group of examinees. Figure 1 describes the posterior density of θ_2 , for the first group of examinees, and under the uniform prior for $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ obtained by letting $|\mathbf{C}_\theta| \rightarrow \infty$. Histogram (a) used 50,000 simulations for the

FIGURE 1
Posterior density of θ_2 (group 1). (a) histogram; (b) maximized posterior density; (c) BCM.

importance sampling technique discussed in section 4. Curve (b) is the maximized posterior density (24), but renormalized to integrate to unity. Curve (c) was obtained by the BCM techniques of section 4. When $\eta = \theta_j$ for some j , BCM reduces to Laplace's method (Leonard, 1982) which does not involve an f -contribution. Note the reassuringly close correspondence between the approximate and simulated exact results. As the posterior density of θ_2 is strongly concentrated on the region $(0, \infty)$, this confirms that a binomial model ($\theta_2 = 0$) is inappropriate. As the posterior density is slightly skew to the left, a normal approximation is not quite adequate. In Figure 2, we describe similar curves but for the predicted true score ξ^* described above.

In Figure 3, we consider inferences for $\eta = \phi_{20}$, the probability of correctly

FIGURE 2

Posterior density of an examinee predicted true score when she/he correctly answered 12 items (group 1). (a) histogram; (b) maximized posterior density; (c) BCM.

answering all 20 items, which possesses maximum likelihood estimates $\hat{\phi}_{20} = 0.00185$ and $\hat{\phi}_{20} = 0.00198$ in the two groups. The two BCM curves (c) and (d) compare favorably with the normalized maximized posterior density (b) in relation to the simulated histogram (a). Curves (c) and (d) follow the two different suggestions made in section 4 for the ε -contribution to (25). Curve (c) uses a log-normal, and curve (d) uses a scale multiple of F -contribution. The posterior mean of ϕ_{20} is equal to 0.00190; this is our Bayes estimate under squared error loss and also our predictive probability that a further student answers all 20 items correctly.

In Figure 4, we assume that a student passes the test if he correctly answers 12

FIGURE 3

Posterior density of ϕ_{20} (group 1). (a) histogram; (b) maximized posterior density; (c) BCM with lognormal contribution; (d) BCM with a scale multiple of F contribution.

items, so that $\eta = \phi_{12} + \dots + \phi_{20}$ is the probability of passing. The BCM curve (c), for η in (11) includes an f-contribution corresponding to a beta density with correct first two moments, and improves upon the normalized maximized posterior density (b), when compared with Histogram (a). The posterior means/predictive probabilities of success, for the two groups, were 0.646 and 0.635, compared with maximum likelihood estimates of 0.650 and 0.642.

In Figure 5, we consider the posterior densities of the average true score (10). In this case the BCM curve (c) was very close to the normalized maximized posterior density (b). The posterior means of the average true score were respectively 12.48 and

FIGURE 4

Posterior density of probability of passing (group 1). (a) histogram; (b) maximized posterior density; (c) BCM.

12.28, providing optimal predictions under squared error loss, and comparing with maximum likelihood estimate of 12.56 and 12.51.

Finally, Figure 6 permits formal comparisons of the results for the two groups by considering $\eta_1 = p(y_1 < y_2)$ and $\eta_2 = p(y_1 > y_2)$, again with $p = 2$, where y_1 and y_2 are the observed scores, for randomly selected examinees from groups 1 and 2. Histograms (a) and (b) describe our simulated posterior densities for η_1 and η_2 , while (c) and (d) providing corresponding smooth curves obtained via a version of BCM (technical details omitted). The maximum likelihood estimates $\hat{\eta}_1 = 0.450$ and $\hat{\eta}_2 = 0.458$ compare with posterior means of 0.451 and 0.458. Since $\hat{\eta}_1 < \hat{\eta}_2$, this suggests that examinees from Group 1 are slightly more likely to perform better than examinees from Group 2.

FIGURE 5

Posterior density of (average) true score (group 1). (a) histogram; (b) maximized posterior density; (c) BCM.

6. The Matsumura Data

In Table 2 we report the observed scores for 145 University of Wisconsin Accounting students, each of whom completed 4 tests, the first three tests containing 25 dissimilar items, and the last test containing 30 dissimilar items. We fitted the exponential family model, using the techniques described in section 2. The AIC criterion suggested $p = 3$ for Test A, and $p = 2$ for each of tests B, C, and D. For Test A, our value of chi-squared was $\chi^2_6 = 7.09$, with six degrees of freedom (based on the 10 groupings 0–16, 17, 18, . . . , 25 of the items) and this was comparable with a beta-binomial fit that provided $\chi^2_7 = 7.37$, with seven degrees of freedom.

FIGURE 6

Comparison of two groups. (a) histogram for $p(y_1 < y_2)$; (b) histogram for $p(y_1 > y_2)$; (c) BCM for $p(y_1 < y_2)$; (d) BCM for $p(y_1 > y_2)$.

The maximum likelihood estimates $\hat{\theta}_1 = 167.82$, $\hat{\theta}_2 = -215.73$ and $\hat{\theta}_3 = 112.41$, for Test A, possessed respective approximate standard errors 100.66, 125.83, and 51.9. This yields, from (12), the quadratic regression, $\gamma^* = 6.713 - 0.690y + .0215y^2$. The deviation of θ_3 from zero can be more accurately judged by reference to Figure 7. Histogram (a) describes our simulated posterior density for θ_3 , under uniform priors for θ_1 , θ_2 , and θ_3 , and curve (c) describes our BCM approximation. The posterior density is skew to the right and the value $\theta_3 = 0$ lies in the extreme left tail. Therefore, the Bayesian analysis makes it quite evident that the cubic term should be included in the sampling model. Standard techniques for generalized linear models do not permit such precise judgments concerning individual parameters.

TABLE 2
Frequencies of Observed Scores for 145 Students on 4 Multiple Choice Tests

Observed Score	Test A	Test B	Test C	Test D
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	1
7	0	0	0	1
8	0	1	2	0
9	0	1	0	6
10	0	1	0	6
11	0	2	1	11
12	0	4	0	11
13	1	1	3	11
14	0	7	3	14
15	0	8	3	12
16	5	8	10	12
17	6	21	10	10
18	7	17	11	20
19	10	15	13	8
20	19	18	15	6
21	15	18	7	8
22	28	11	13	1
23	22	7	13	0
24	15	2	18	0
25	17	3	7	0
26				10
27				4
28				0
29				0
30				1

FIGURE 7

Posterior density of θ_3 (Matsumura's EXAM A). (a) histogram; (b) maximized posterior density; (c) BCM.

The remainder of our analysis of the Matsumura data is omitted for brevity of presentation. We however, finally mention the data described by Leonard et al. (1989, sec. 7) where $n = 200$ and $m = 10$, with $n_0, n_1, \dots, n_{10}$ respectively equalling 8, 12, 17, 18, 12, 23, 27, 34, 31, 14, and 4. In this case, our model with $p = 3$ fitted well with $\chi^2_7 = 5.59$, comparing favorably with the beta-binomial fit, $\chi^2_8 = 19.23$.

7. The Keats Data

Columns 1 and 2 of Table 3 present the observed frequency distribution reported by Keats (1964). We fitted the exponential family model, using the techniques described in section 2. The AIC criterion suggested $p = 5$. The maximum likelihood estimates

TABLE 3

Comparison of Observed Distribution with Four Estimated Distributions

Raw Score	Observed Frequency	Four- Parameter		Five- Parameter	
		Beta- Binomial	Quadratic Regression	Gamma- Poisson	Exponential Family
0	0	0.26	2.43	0.79	0.06
1	0	0.63	2.05	0.97	0.37
2	1	1.10	2.10	1.18	1.06
3	2	1.66	2.29	1.44	2.07
4	4	2.30	2.56	1.75	3.15
5	6	3.02	2.91	2.14	4.11
6	7	3.82	3.34	2.60	4.87
7	2	4.71	3.87	3.17	5.44
8	3	5.69	4.52	3.85	5.92
9	4	6.77	5.29	4.67	6.40
10	5	7.96	6.22	5.67	6.95
11	14	9.26	7.34	6.87	7.67
12	10	10.70	8.68	8.32	8.61
13	12	12.28	10.29	10.05	9.87
14	10	14.03	12.21	12.13	11.51
15	14	15.98	14.51	14.62	13.63
16	17	18.16	17.25	17.58	16.31
17	16	20.61	20.52	21.10	19.64
18	22	23.40	24.41	25.25	23.69
19	27	26.61	29.02	30.14	28.50
20	37	30.37	34.46	35.85	34.11
21	48	34.86	40.84	42.47	40.53
22	40	40.40	48.25	50.06	47.80
23	50	47.58	56.77	58.64	56.01
24	74	57.54	66.39	68.14	65.36
25	78	72.30	76.96	78.37	76.09
26	85	94.17	88.07	88.84	88.39
27	103	121.46	198.76	98.58	101.79
28	112	139.28	107.05	105.68	113.48
29	114	119.00	108.58	106.09	114.25
30	83	54.05	92.04	89.44	82.38

were $\hat{\theta}_1 = 48.56$, $\hat{\theta}_2 = 14.68$, $\hat{\theta}_3 = 217.30$, $\hat{\theta}_4 = 259.14$, and $\hat{\theta}_5 = 112.26$. Our value of chi-squared was $\chi^2_{18} = 14.91$, with 18 degrees of freedom (based on 24 groupings 0–5, 6–8, 9, 10, . . . , 30 of the items) and the chi-square level of significance was 0.67, and these were comparable with the four-parameter beta-binomial model, the quadratic regression model suggested by Keats (1964) and the gamma-Poisson model suggested by Wilcox (1981a, 1981b) provided chi-squared values $\chi^2_{19} = 45.91$, $\chi^2_{19} = 21.30$, and $\chi^2_{21} = 26.87$ with degrees of freedom 19, 19 and 21 and chi-square levels of significance were 0.0005, 0.32, and 0.18, respectively. The corresponding estimated frequency distributions are presented in columns 3–6 in Table 3. The exponential family model produced larger chi-square level of significance than the other three models in this example. A Bayesian analysis of the Keats data is also available from the authors.

8. The Classical Test Theory Model

Consider the classical test theory model (Lord & Novick, 1968, p. 56),

$$X = T + E, \quad (28)$$

where X is the observed score and T is the true score. Suppose that T and E are independent and the E is normally distributed with mean zero and known variance σ_E^2 . Then the regression of true score upon observed score is

$$E(T|X) = X + \sigma_E^2 \frac{\partial \log p(X)}{\partial X}, \quad (29)$$

where $p(x)$ denotes the unconditional density of x . Suppose also that the density of X belongs to the continuous p -parameter exponential family

$$p(x) = \exp \{ \theta_1 t_1(x) + \theta_2 t_2(x) + \cdots + \theta_p t_p(x) - D(\boldsymbol{\theta}) \}, \quad (-\infty < x < \infty), \quad (30)$$

where

$$D(\boldsymbol{\theta}) = \log \int_{-\infty}^{\infty} \exp \{ \theta_1 t_1(x) + \theta_2 t_2(x) + \cdots + \theta_p t_p(x) \} dx. \quad (31)$$

Then (29) reduces to the regression

$$E(T|X) = X + \sigma_E^2 (\theta_1 t_1^{(1)}(x) + 2\theta_2 t_2^{(1)}(x) + \cdots + p\theta_p t_p^{(1)}(x)), \quad (32)$$

where

$$t_j^{(1)}(x) = \frac{\partial t_j(x)}{\partial x}, \quad \text{for } j = 1, 2, \dots, p. \quad (33)$$

Most of the techniques described in our paper are readily generalizable to this model. There are obvious applications of (12) and (32) to empirical Bayes estimation problems for binomial and normal data (e.g., Leonard, 1972, 1984; Lord, 1969).

9. Concluding Remarks

Our exponential family model provides a framework for handling dissimilar items and inhomogeneous populations of students (e.g., the Keats data) and will often be more flexible than the simple beta-binomial model and its generalizations. It explains

many data sets without needing the construction of a guessing parameter, depending upon the item, rather than upon the student's abilities and knowledge. Note that the simple beta-binomial model possesses a coefficient of variation larger than does a binomial model with the same mean. On many data sets where the items are dissimilar (e.g., the Keats data), the coefficient of variation is smaller than for a binomial model with the same mean. Trying to explain this by an extra guessing parameter, rather than a more appropriate sampling model might be quite misleading. Of course, if the guessing probability for any particular item is constant and non-zero across all students, then Morrison and Brockway (1979) introduce the guessing aspect into the standard beta-binomial model, in a simple, appealing and estimable manner, providing an advantage when compared with the exponential family formulation. However, Wilcox (1981a, p. 22) discusses three further problems involving guessing parameters.

With appropriate choices of p and the basis functions $b_1, b_2, \dots, b_p$ in (3), most practical data sets, requiring generalizations of random sampling from the binomial distribution, can be modelled by a member of the p -parameter exponential family, defined by (2) and (3), and our approach may therefore be regarded as global in nature. It might be reasonable to use our approach to explore the main features of the data, but, to use our analysis to suggest a specialized model from the literature, for example, the Morrison-Brockway model, or one of the alternatives cited in section 1, for confirmatory purposes.

Appendix 1

We give fuller details of property (iv) of section 1, and the derivation of (12). Under the beta-binomial model, $\phi_j = p(y = j)$ takes the form

$$\begin{aligned}\phi_j &= {}^m C_j E_{\xi} [\xi^j (1 - \xi)^{m-j}] \\ &= {}^m C_j E_{\gamma} \left[\frac{\exp(j\gamma)}{(1 + e^{\gamma})^m} \right], \quad (j = 0, 1, \dots, m),\end{aligned}$$

where the first expectation is with respect to some distribution for ξ , and the second expectation is with respect to a corresponding distribution of $\gamma = \log \xi - \log(1 - \xi)$. Therefore,

$$\begin{aligned}\frac{\partial \log \phi_y}{\partial y} &= -\psi(y+1) + \psi(m-y+1) + \frac{E[\gamma(1 + e^{\gamma})^{-m} \exp(y\gamma)]}{E[(1 + e^{\gamma})^{-m} \exp(y\gamma)]} \\ &= -\psi(y+1) + \psi(m-y+1) + E(\gamma|y),\end{aligned}$$

yielding

$$E(\gamma|y) = \frac{\partial \log \phi_y}{\partial y} + \psi(y+1) - \psi(m-y+1).$$

The result so far is reasonably standard. However, when ϕ_j also satisfies (2) through (5), for some $\theta_1, \dots, \theta_p$, appropriate differentiations of $\log \phi_y$ yield (12) together with the further results stated in property (iv) of section 1.

Appendix 2

We now describe fuller details of the BCM procedure discussed in section 4. Leonard, Hsu, and Tsui (1989) obtain this via a Taylor Series expansion of the log posterior density $\log \pi(\boldsymbol{\theta}|\mathbf{y})$ about the conditional maximum $\boldsymbol{\theta}_\eta$. If the cubic and higher terms are neglected, then the following apparently extremely accurate approximation to (22) can be obtained analytically:

$$\pi^*(\eta|\mathbf{y}) \propto \pi_M(\eta|\mathbf{y}) f(\eta|\boldsymbol{\theta}_\eta^*, \mathbf{R}_\eta^{-1}) f^*(\eta), \quad (n \in \Omega),$$

where

1. $\pi_M(\eta|\mathbf{y})$ is the maximized posterior density satisfying (24);
2. $f(\eta|\boldsymbol{\mu}, \mathbf{C})$ denotes the density of $\eta = g(\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$;
3. $f^*(\eta) = |\mathbf{R}_\eta|^{-1/2} \exp\left(\frac{1}{2} \boldsymbol{\ell}_\eta' \mathbf{R}_\eta \boldsymbol{\ell}_\eta\right)$,

with

$$\text{a. } \mathbf{R}_\eta = -[(\partial^2 \log \pi(\boldsymbol{\theta}|\mathbf{y})) / \partial(\boldsymbol{\theta}\boldsymbol{\theta}')]_{\boldsymbol{\theta}=\boldsymbol{\theta}_\eta},$$

denoting the posterior information matrix of $\boldsymbol{\theta}$, evaluated at the conditional maximum $\boldsymbol{\theta}_\eta$;

$$\text{b. } \boldsymbol{\ell}_\eta = [(\partial \log \pi(\boldsymbol{\theta}|\mathbf{y})) / \partial \boldsymbol{\theta}]_{\boldsymbol{\theta}=\boldsymbol{\theta}_\eta};$$

$$\text{c. } \boldsymbol{\theta}_\eta^* = \boldsymbol{\theta}_\eta + \mathbf{R}_\eta^{-1} \boldsymbol{\ell}_\eta.$$

Further computations or approximations for the f contribution $f(\eta|\boldsymbol{\mu}, \mathbf{C})$ are typically required. The results in the current paper suggest that any sensible approximation, capturing the first two moments, will probably suffice.

References

- Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, 30(A), 9–14.
- Altham, P. M. E. (1978). Two generalizations of the binomial distribution. *Applied Statistics*, 27, 162–167.
- Anderson, D. A., & Aitken, M. (1985). Marginal maximum likelihood estimation of item parameters: Application of an algorithm. *Journal of Royal Statistical Society, Series B*, 26, 203–210.
- Atilgan, T. (1983). *Parameter parsimony, model selection, and smooth density estimation*. Unpublished doctoral dissertation, University of Wisconsin-Madison.
- Atilgan, T., & Leonard, T. (1988). On the application of AIC to bivariate density estimation, non-parametric regression, and discrimination. In H. Bozadogan & A. K. Gupta (Eds.), *Multivariate statistical modeling and data analysis* (pp. 1–16). Dordrecht, Holland: Reidel.
- Bell, S. S. (1990). Empirical Bayes alternatives to the beta-binomial model. Unpublished doctoral dissertation, Columbia University.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29–51.
- Bock, R. D., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, 46, 443–454.
- Carter, M. C., & Williford, W. O. (1975). Estimation in a modified binomial distribution. *Applied Statistics*, 24, 319–328.
- Consul, P. C. (1974). A simple urn model dependent upon predetermined strategy. *Sankhya, Series B*, 36, 391–399.
- Consul, P. C. (1975). On a characterization of Lagrangian Poisson and quasi-binomial distributions. *Communications in Statistics*, 4, 555–563.
- Dalal, S. R., & Hall, W. J. (1983). Approximating priors by mixtures of natural conjugate priors. *Journal of Royal Statistical Society, Series B*, 45, 278–286.
- Duncan, G. T. (1974). An empirical Bayes approach to scoring multiple-choice tests in the misinformation model. *Journal of the American Statistical Association*, 69, 50–57.
- Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398–409.
- Geweke, J. (1988). Antithetic acceleration of Monte-Carlo integration in Bayesian inference. *Journal of Econometrics*, 38, 73–89.

- Geweke, J. (1989). Exact predictive density for linear models with arch distributions. *Journal of Econometrics*, 40, 63–86.
- Hsu, J. S.J. (1990). *Bayesian inference and marginalization*. Unpublished doctoral dissertation, University of Wisconsin-Madison.
- Keats, J. A. (1964). Some generalizations of a theoretical distribution of mental test scores. *Psychometrika*, 29, 215–231.
- Lehmann, E. L. (1983). *Theory of point estimation*. New York: John Wiley & Sons.
- Leonard, T. (1972). Bayesian methods for binomial data. *Biometrika*, 59, 581–589.
- Leonard, T. (1973). A Bayesian method for histograms. *Biometrika*, 60, 297–308.
- Leonard, T. (1982). Comment on the paper by Lejeune and Faulkenberry. *Journal of the American Statistical Association*, 77, 657–658.
- Leonard, T. (1984). Some data-analytic modifications to Bayes-Stein estimation. *Annals of the Institute of Statistical Mathematics*, 36, 11–21.
- Leonard, T., Hsu, J. S.J., & Tsui, K. (1989). Bayesian marginal inference. *Journal of the American Statistical Association*, 84, 1051–1058.
- Leonard, T., & Novick, J. B. (1986). Bayesian full rank marginalization for two-way contingency tables. *Journal of Educational Statistics*, 11, 33–56.
- Lord, F. M. (1965). A strong true-score theory, with applications. *Psychometrika*, 30, 239–270.
- Lord, F. M. (1969). Estimating true-score distributions in psychological testing: An empirical Bayes estimation problem. *Psychometrika*, 34, 259–299.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores (with contributions by Allen Birnbaum)*. Reading, MA: Addison-Wiley.
- Lord, F. M., & Stocking, M. L. (1976). An interval estimate for making statistical inference about true scores. *Psychometrika*, 41, 79–87.
- McCullagh, P., & Nelder, J. A. (1985). *Generalized linear models*. New York: Chapman and Hall.
- Mislevy, R. J. (1986). Bayes modal estimation in item response. *Psychometrika*, 51, 177–195.
- Morrison, D. G., & Brockway, G. (1979). A modified beta-binomial model with applications to multiple choice and taste tests. *Psychometrika*, 44, 427–442.
- Prentice, R. L., & Barlow, W. E. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44, 1033–48.
- Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo method*. New York: John Wiley and Sons.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Mathematical Statistics*, 6, 461–464.
- Takane, Y., Bozdogan, H., & Shibayama, T. (1987). Ideal point discriminant analysis. *Psychometrika*, 52, 371–392.
- Wilcox, R. R. (1981a). A review of the beta-binomial model and its extensions. *Journal of Educational Statistics*, 6, 3–32.
- Wilcox, R. R. (1981b). A cautionary note on estimating the reliability of a mastery test with the beta-binomial model. *Applied Psychological Measurement*, 5, 531–537.
- Young, A. S. (1977). A Bayesian approach to prediction using polynomials. *Biometrika*, 64, 309–318.

Manuscript received 11/6/89

Final version received 6/6/90