4

The Mixed-Coefficients Multinomial Logit Model: A Generalized Form of the Rasch Model

Raymond J. Adams and Margaret L. Wu

University of Melbourne

4.1 Introduction

Since Rasch's introduction of his item response models (Rasch, 1960), there has been a proliferation of extensions and alternatives, each of which has a different name and different matching software package. As Adams, Wilson, & Wang (1997) pointed out, the proliferation of models has, in some ways, been a hindrance to practitioners. This paper presents a generalized item response model that provides a unifying framework for a large class of Rasch-type models. The advantages of a single framework include mathematical elegance, generality in a single software package, and a facilitation of the development, testing, and comparison of new models. The unified model is a multidimensional item response model, the specification of which is achieved through the use of design matrices chosen to represent the parametrization of the model. In the paper we discuss the estimation of the parameters of the model, the testing of model fit, and we illustrate how standard models (such as the simple logistic, the rating scale, and facets models) and alternative user-defined models are specified.

Over the past 30 years, a proliferation of item response models has emerged. In the logistic item response model family, notably, the simple logistic model (Rasch, 1980), the partial-credit model (Masters, 1982), the rating-scale model (Andrich, 1978), the facets model (Linacre, 1989), and the linear logistic model (Fischer, 1973) have all played an important role in the analysis of item response data. Typically, the development of the estimation procedures of parameters for each item response model was specific to the model, as was the development of dedicated software programs for each model. Surveying the family of RMs, Adams & Wilson (1996) developed a unified approach to specifying the models and then consequentially estimating the parameters. There are at least two advantages to developing one single framework to encompass a family of models. First, the development of the estimation procedures and associated software for the implementation of the models can be streamlined within a single framework of models. That is, one needs to develop only one set of estimation procedures and one software program to carry out the estimation of the parameters in the models. Second, a generalized framework provides an opportunity for the development of new models that fit in the framework. This allows for the flexible application of item response models to suit users' requirements.

This paper describes a generalized framework for specifying a family of logistic item response models through the specification of design matrices. The estimation procedures are also described. The idea of the use of design matrices is extended to the construction of a family of goodness-of-fit tests. Flexibility in the construction of fit tests allows the users to target specific hypotheses regarding the fit of the items to the model, such as the violation of local independence between subsets of items.

4.2 The Mixed-Coefficients Multinomial Logit Model

The mixed-coefficients multinomial logit model (MCML) is a categorical response model, and in most applications, the response patterns to a set of test items (the categorical outcomes) are modeled as the dependent variable. Under the model, the response patterns are predicted by logistic regression, where the independent variables are item difficulty and person abilities.[1]

The model is referred to as a mixed-coefficients model because items are described by a fixed set of unknown parameters, ξ, while the student ability (the latent variable), θ, is a random effect.

The model is specified as follows. Assume that there are I items and they are indexed $i = 1, \ldots, I$ with each item admitting $K_i + 1$ response categories indexed $k = 0, 1, \ldots, K_i$. That is, a response to item i by a student can be allocated to one of $K_i + 1$ response categories. The vector-valued random variable $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iK_i})^T$, where for $k = 1, \ldots, K_i$,

$$X_{ik} = \begin{cases} 1 \text{ if response to item } i \text{ is in category } k, \\ 0 \text{ otherwise,} \end{cases} \tag{4.1}$$

is used to indicate the $K_i + 1$ possible responses to item i. A vector of zeros denotes a response in category zero, making the zero category a reference category, which is necessary for model identification. Using this as the reference category is arbitrary, and does not affect the generality of the model.

Each $\mathbf{X}_i$ consists of a sequence of 0's and possibly one 1, indicating the student's response category for that item. For example, if the response category is 0 for an item with four categories (0, 1, 2, 3), then $\mathbf{X}_i^T = (0, 0, 0)$. If the response category is 2, then $\mathbf{X}_i^T = (0, 1, 0)$.

[1] Throughout this article the term "ability" is used as a generic placeholder to refer to the latent variable being measured. The term "difficulty" refers to parameters that characterize the items.

The $\mathbf{X}_i$ can also be collected together into the single vector $\mathbf{X}^T = \left(\mathbf{X}_1^T, \mathbf{X}_2^T, \ldots, \mathbf{X}_I^T\right)$, called the response vector. Particular instances of each of these random variables are indicated by their lowercase equivalents, $\mathbf{x}$, $\mathbf{x}_i$, and x_{ik}.

Items are described through a vector $\xi^T = (\xi_1, \xi_2, \ldots, \xi_p)$, of p parameters. Linear combinations of these are used in the response probability model to describe the empirical characteristics of the response categories of each item. These linear combinations are defined by design vectors $\mathbf{a}_{ik}$ $(i = 1, \ldots, I;\ k = 1, \ldots K_i)$, each of length p, which can be collected to form a design matrix $\mathbf{A}^T = (\mathbf{a}_{11}, \mathbf{a}_{12}, \ldots, \mathbf{a}_{1K_1}, \mathbf{a}_{21}, \ldots, \mathbf{a}_{2K_2}, \ldots, \mathbf{a}_{IK_I})$.

The multidimensional form of the model assumes that a set of D traits underlies the individuals' responses. The D latent traits define a D-dimensional latent space. The vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_D)^T$ represents an individual's position in the D-dimensional latent space.

The model also introduces a scoring function that allows the specification of the score or performance level assigned to each possible response category to each item. To do so, the notion of a response score b_{ikd} is introduced, which gives the performance level of an observed response in category k, item i, dimension d. The scores across D dimensions can be collected into a column vector $\mathbf{b}_{ik} = (b_{ik1}, b_{ik2}, \ldots, b_{ikD})^T$, and again collected into the scoring submatrix for item i, $\mathbf{B}_i = (\mathbf{b}_{i1}, \mathbf{b}_{i2}, \ldots, \mathbf{b}_{ik_i})^T$ and then into a scoring matrix $\mathbf{B} = \left(\mathbf{B}_1^T, \mathbf{B}_2^T, \ldots, \mathbf{B}_I^T\right)^T$ for the entire test. (The score for a response in the zero category is zero, but other responses may also be scored zero.)

The regression of the response vector on the item and person parameters is

$$f(\mathbf{x}; \xi | \theta) = \Psi(\theta, \xi) \exp\left[\mathbf{x}^T (\mathbf{B}\theta + \mathbf{A}\xi)\right], \tag{4.2}$$

with

$$\Psi(\theta, \xi) = \left\{ \sum_{\mathbf{z} \in \Omega} \exp\left[\mathbf{z}^T (\mathbf{B}\theta + \mathbf{A}\xi)\right] \right\}^{-1}, \tag{4.3}$$

where Ω is the set of all possible response vectors.

4.2.1 Simple Logistic Model (SLM) Example

Equations (4.2) and (4.3) can be illustrated with some simple cases. Consider a simple logistic model for dichotomous data. This model would normally be written (in the notation of Wright & Stone, 1979) as

$$\Pr(X_{i1} = 1, \delta_i | \theta) = \frac{\exp(\theta + \delta_i)}{1 + \exp(\theta + \delta_i)}. \tag{4.4}$$

For three dichotomous items, the probability of the response vector, $\mathbf{x}^T = (x_{11}, x_{21}, x_{31})$, is then

$$
\begin{aligned}
\Pr\left(\mathbf{X}=\mathbf{x},\delta_1,\delta_2,\delta_3|\theta\right) &= \prod_{i=1}^{3}\frac{\exp\left\{x_{i1}\left(\theta+\delta_i\right)\right\}}{1+\exp\left(\theta+\delta_i\right)} \\
&= \frac{\exp\left\{\sum_{i=1}^{3}x_{i1}\left(\theta+\delta_i\right)\right\}}{\prod_{i=1}^{3}\left\{1+\exp\left(\theta+\delta_i\right)\right\}} \\
&= \frac{\exp\left(r\theta+x_{11}\delta_1+x_{21}\delta_2+x_{31}\delta_3\right)}{D},
\end{aligned}
\tag{4.5}
$$

where

$$
\begin{aligned}
D = 1 &+ \exp\left(\theta+\delta_1\right)+\exp\left(\theta+\delta_2\right)+\exp\left(\theta+\delta_3\right) \\
&+\exp\left(2\theta+\delta_1+\delta_2\right)+\exp\left(2\theta+\delta_2+\delta_3\right) \\
&+\exp\left(2\theta+\delta_1+\delta_3\right)+\exp\left(3\theta+\delta_1+\delta_2+\delta_3\right),
\end{aligned}
$$

and

$$
r=\sum_{i=1}^{3}x_{i1}.
$$

To show how (4.2) and (4.5) can be made equivalent, consider the following choices of $\mathbf{A}$, $\mathbf{B}$ and ξ:

$$
\mathbf{A}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\quad \mathbf{B}=\begin{bmatrix}1\\1\\1\end{bmatrix},\quad \text{and}\quad \xi=\begin{bmatrix}\delta_1\\\delta_2\\\delta_3\end{bmatrix},
\tag{4.6}
$$

where the first row of $\mathbf{A}$ corresponds to item one category one; the second row corresponds to item two category one; the third row corresponds to item three category one. The rows of $\mathbf{B}$ correspond to the same item and category as for the rows of $\mathbf{A}$. The elements of ξ correspond to the item difficulty parameters of items one to three respectively. Note that with three dichotomous items there are eight different response patterns.

4.2.2 Partial-Credit Example

As a second example, consider a partial-credit item with three categories: 0, 1, and 2. Using the notation of Wright & Masters (1982), (4.2) and (4.3) can be written as

$$
\begin{aligned}
\Pr\left(X_i^T=(0,0);\delta_{i1},\delta_{i2}|\theta\right) &= \Pr\left(\text{category}\,0;\mathbf{A},\mathbf{B},\xi|\theta\right) \\
&= \frac{1}{1+\exp\left(\theta+\delta_{i1}\right)+\exp\left(2\theta+\delta_{i1}+\delta_{i2}\right)},
\end{aligned}
$$

$$\Pr\left(X_i^T = (1,0)\,;\delta_{i1},\delta_{i2}|\theta\right) = \Pr\left(\text{category 1}; \mathbf{A},\mathbf{B},\xi|\theta\right) = \frac{\exp\left(\theta+\delta_{i1}\right)}{1+\exp\left(\theta+\delta_{i1}\right)+\exp\left(2\theta+\delta_{i1}+\delta_{i2}\right)}, \quad (4.7)$$

$$\Pr\left(X_i^T = (0,1)\,;\delta_{i1},\delta_{i2}|\theta\right) = \Pr\left(\text{category 2}; \mathbf{A},\mathbf{B},\xi|\theta\right) = \frac{\exp\left(2\theta+\delta_{i1}+\delta_{i2}\right)}{1+\exp\left(\theta+\delta_{i1}\right)+\exp\left(2\theta+\delta_{i1}+\delta_{i2}\right)}.$$

For two three-category partial-credit items, the probability of the response vector $\mathbf{x}$ is then

$$\Pr\left(\mathbf{X}=\mathbf{x};\delta_{11},\delta_{12},\delta_{21},\delta_{22}|\theta\right) = \prod_{i=1}^{2} \frac{\exp\left(s_i\theta + \sum_{k=1}^{s_i}\delta_{ik}\right)}{1+\exp\left(\theta+\delta_{i1}\right)+\exp\left(2\theta+\delta_{i1}+\delta_{i2}\right)}, \quad (4.8)$$

where s_i is the observed response category for item i and $\sum_{i=1}^{0} u \equiv 0$ for all possible values of u.

To show how (4.2) and (4.8) can be made equivalent, consider the following choices of $\mathbf{A}$, $\mathbf{B}$, and ξ:

$$\mathbf{A} = \begin{bmatrix} 1&0&0&0 \\ 1&1&0&0 \\ 0&0&1&0 \\ 0&0&1&1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \quad \text{and} \quad \xi = \begin{bmatrix} \delta_{11} \\ \delta_{12} \\ \delta_{21} \\ \delta_{22} \end{bmatrix}, \quad (4.9)$$

where the first row of $\mathbf{A}$ corresponds to item one category one; the second row corresponds to item one category two; the third row corresponds to item two category one; and the fourth row corresponds to item two category two. The same row referencing applies to the matrix $\mathbf{B}$.

4.2.3 Facet Example

Consider an example of a facets model (Linacre, 1989) in which there are three raters, each rater rating the same two dichotomous items. This is modeled as six *generalized items*. A generalized item is defined for each of the possible combinations of a rater and an actual item. Generalized item one is the response category given by rater one on item one. Generalized item two is the response category given by rater one on item two, and so on. The following choices of $\mathbf{A}$, $\mathbf{B}$, and ξ will give this facets model:

$$\mathbf{A} = \begin{bmatrix} 1&0&0&1&0 \\ 1&0&0&0&1 \\ 0&1&0&1&0 \\ 0&1&0&0&1 \\ 0&0&1&1&0 \\ 0&0&1&0&1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad \xi = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \delta_1 \\ \delta_2 \end{bmatrix}, \quad (4.10)$$

where the first row of **A** corresponds to category one of generalized item one (rater one, item one); the second row corresponds to category one of generalized item two (rater one, item two); the third row corresponds to category one of generalized item three (rater two, item one); the fourth row corresponds to category one of generalized item four (rater two, item two), and so on. The same row referencing applies to the matrix **B**. The first three elements (ρ_1, ρ_2, ρ_3) of ξ are the severity parameters of raters one to three respectively. The fourth and fifth element, (δ_1, δ_2) of ξ are the item-difficulty parameters for the two dichotomous items.

4.2.4 Multidimensional Examples

Finally, Figure 4.1 shows two possible multidimensional models: a between-item multidimensionality model and a within-item multidimensionality model (Adams, Wilson, & Wang, 1997). In each case, a hypothetical nine-item test is considered. In the between-item multidimensional case (the left-hand side of Figure 4.1), each item is associated with a single dimension, but the collection of items covers three dimensions: three items are associated with each of the three latent dimensions. In the within-item case (the right-hand side of Figure 4.1), some items are associated with more than one dimension. For example, item two is associated with both dimensions one and two.

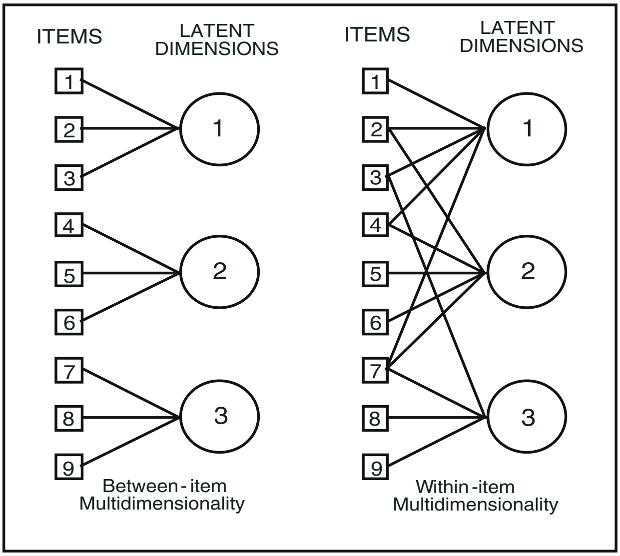

Fig. 4.1. Between- and within-item multidimensionality

If the items shown in Figure 4.1 are all dichotomous, then the matrices **A**, **B**, and ξ, as given in (4.11) and (4.12), if substituted into (4.2), will yield the between- and within-item multidimensional models respectively as shown in Figure 4.1.

Note that the only difference between (4.11) and (4.12) is the **B** matrix. This matrix is called the score matrix and is used to indicate the scores of the items on each of the three dimensions. Note also that the **B** matrices (4.11) and (4.12) have three columns, one for each of the three dimensions that are modeled:

$$\mathbf{A} = \begin{bmatrix} 1\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1\,0\,0 \\ 1\,0\,0 \\ 1\,0\,0 \\ 0\,1\,0 \\ 0\,1\,0 \\ 0\,1\,0 \\ 0\,0\,1 \\ 0\,0\,1 \\ 0\,0\,1 \end{bmatrix}, \quad \text{and} \quad \xi = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5\,, \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \end{bmatrix}, \tag{4.11}$$

$$\mathbf{A} = \begin{bmatrix} 1\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,1\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1\,0\,0 \\ 1\,1\,0 \\ 1\,0\,1 \\ 1\,1\,0 \\ 0\,1\,0 \\ 0\,1\,0 \\ 1\,1\,1 \\ 0\,0\,1 \\ 0\,0\,1 \end{bmatrix}, \quad \text{and} \quad \xi = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \end{bmatrix}. \tag{4.12}$$

4.2.5 The Population Model

The item response model (4.2) is a conditional model, in the sense that it describes the process of generating item responses conditional on the latent variable, $\boldsymbol{\theta}$. The complete definition of the model, therefore, requires the specification of a density, $f_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\alpha})$, for the latent variable, $\boldsymbol{\theta}$. Let $\boldsymbol{\alpha}$ symbolize a set of parameters that characterize the distribution of θ. The most common practice in specifying unidimensional marginal item response models is to assume that students have been sampled from a normal population with mean μ and variance σ^2. That is,

$$f_\theta(\theta;\alpha) \equiv f_\theta\left(\theta;\mu,\sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left[-\frac{(\theta-\mu)^2}{2\sigma^2}\right], \tag{4.13}$$

or equivalently $\theta = \mu + E$, where $E \sim N\left(0,\sigma^2\right)$.

Adams, Wilson, & Wu (1997) discuss how a natural extension of (4.13) is to replace the mean, μ, with the regression model, $\mathbf{Y}_n^T\beta$ where $\mathbf{Y}_n$ is a vector of u fixed and known values for student n, and β is the corresponding vector of regression coefficients. For example, $\mathbf{Y}_n$ could be constituted of student variables such as gender or socioeconomic status. Then the population model for student n becomes,

$$\boldsymbol{\theta}_n = \mathbf{Y}_n^T \beta + E_n, \tag{4.14}$$

where the E_n are assumed to be independently and identically normally distributed with mean zero and variance σ^2, so that (4.13) can be generalized to

$$f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_n; \mathbf{Y}_n, \beta, \boldsymbol{\sigma^2}\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(\boldsymbol{\theta}_n - \mathbf{Y}_n^T\beta\right)^T \left(\boldsymbol{\theta}_n - \mathbf{Y}_n^T\beta\right)\right], \tag{4.15}$$

a normal distribution with mean $\mathbf{Y}_n^T\beta$ and variance σ^2. The generalization needs to be taken one step further to apply it to the vector-valued $\boldsymbol{\theta}$ (of length d) rather than the scalar-valued $\boldsymbol{\theta}$. The extension results in the multivariate population model

$$f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_n; \mathbf{W}_n, \gamma, \Sigma\right) = (2\pi)^{-\frac{d}{2}} \left|\Sigma\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}_n - \gamma\mathbf{W}_n\right)^T \Sigma^{-1} \left(\boldsymbol{\theta}_n - \gamma\mathbf{W}_n\right)\right], \tag{4.16}$$

where γ is a $d \times u$ matrix of regression coefficients, a $d \times d$ variance–covariance matrix Σ, and $\mathbf{W}_n$ is a $u \times 1$ vector of fixed variables.

While in most cases, the multivariate normal distribution (4.16) is assumed as the population distribution, other forms of the population distribution can also be considered. For example, Adams, Wilson, & Wang (1997) considered a step distribution defined on a prespecified set of nodes. They argued that this could be used as an opportunity to approximate an arbitrary continuous-trait distribution.

4.2.6 Combined Model

The conditional item response model (4.2) and the population model (4.16) are combined to obtain the unconditional, or marginal, item response model:

$$f_{\mathrm{x}}\left(\mathbf{x}; \xi, \gamma, \Sigma\right) = \int_{\theta} f_{\mathrm{x}}\left(\mathbf{x}; \xi | \theta\right)\ f_{\theta}\left(\theta; \gamma, \Sigma\right)\ d\theta. \tag{4.17}$$

It is important to recognize that under this model, the locations of individuals on the latent variables are not estimated. The parameters of the model are γ, Σ, and ξ, where γ, Σ are the population parameters and ξ are the item parameters.

4.3 Identification

For the purposes of the identification of (4.17), certain constraints must be placed on the design matrices **A** and **B**.[2] Volodin & Adams (1995) show that the following are necessary and sufficient conditions for the identification of (4.17).

Proposition One: If D is the number of latent dimensions, P is the length of the parameter vector ξ, K_i+1 is the number of response categories for item i, and $K = \sum_{i\in\mathbf{I}} K_i$, then model (4.17), if applied to the set of items **I,** can be identified only if $P + D \leqslant K$.

Proposition Two: If D is the number of latent dimensions and P is the length of the parameter vector ξ, then model (4.17) can only be identified if $\text{rank}(\mathbf{A}) = P$, $\text{rank}(\mathbf{B}) = D$ and $\text{rank}([\mathbf{BA}]) = P + D$.

Proposition Three: If D is the number of latent dimensions, P is the length of the parameter vector ξ, K_i+1 is the number of response categories for item i, and $K = \sum_{i\in\mathbf{I}} K_i$, then model (4.17), if applied to the set of items I, can be identified only if and only if $\text{rank}([\mathbf{BA}]) = P + D \leqslant K$.

4.4 Estimation

In the following section, a maximum likelihood approach to estimating the parameters is sketched (Adams, Wilson, & Wu, 1997), and the possibility of using a conditional maximum likelihood (Andersen, 1970) approach is discussed.

4.4.1 Maximum Likelihood

The maximum likelihood approach to estimating the parameters of (4.17) proceeds as follows. Let $\mathbf{x}_n$ be the response pattern of person n and assume independent observations are made for $n = 1, \ldots, N$ persons.[3] It follows that the likelihood for the N sampled students is

$$\Lambda = \prod_{n=1}^{N} f_{\mathrm{x}}\left(\mathbf{x}_n; \xi, \gamma, \Sigma\right). \tag{4.18}$$

Differentiating with respect to each of the parameters and defining the marginal posterior as

$$h_\theta\left(\theta_n; \mathbf{W}_n, \xi, \gamma, \Sigma | \mathbf{x}_n\right) = \frac{f_{\mathbf{x}}\left(\mathbf{x}_n; \xi | \theta_n\right)\ f_\theta\left(\theta_n; \mathbf{W}_n, \gamma, \Sigma\right)}{f_{\mathbf{x}}\left(\mathbf{x}_n; \mathbf{W}_n, \xi, \gamma, \Sigma\right)}, \tag{4.19}$$

[2] In fact, the design matrices as used in the examples do not yield identified models.

[3] For notational convenience, the symbol $\boldsymbol{x}_n$ is used here to denote the full response pattern for person n, and not just the response vector for a particular item as defined in (4.1).

the following system of likelihood equations is derived (see Adams, Wilson, & Wu, 1997):

$$\mathbf{A^T} \sum_{n=1}^{N} \left[\mathbf{x}_n - \int_{\theta_n} E_{\mathbf{z}} \left(\mathbf{z} | \theta_n \right) \, h_\theta \left(\theta_n ; \mathbf{Y}_n, \xi, \gamma, \Sigma | \mathbf{x}_n \right) \, d\theta_n \right] = 0, \tag{4.20}$$

$$\hat{\gamma} = \left(\sum_{n=1}^{N} \overline{\theta_n} \mathbf{W}_n^T \right) \left(\sum_{n=1}^{N} \mathbf{W}_n \mathbf{W}_n^T \right)^{-1}, \tag{4.21}$$

$$\hat{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} \int_{\theta_n} \left(\theta_n - \gamma \mathbf{W}_n \right) \left(\theta_n - \gamma \mathbf{W}_n \right)^T h_\theta \left(\theta_n ; \mathbf{Y}_n, \xi, \gamma, \Sigma | \mathbf{x}_n \right) d\theta_n, \tag{4.22}$$

where

$$E_{\mathbf{z}} \left(\mathbf{z} | \theta_n \right) = \Psi \left(\theta_n, \xi \right) \sum_{\mathbf{z} \in \Omega} \mathbf{z} \exp \left[\mathbf{z}^T \left(\mathbf{b} \theta_n + \mathbf{A} \xi \right) \right] \tag{4.23}$$

and

$$\bar{\theta}_n = \int_{\theta_n} \theta_n h_\theta \left(\theta_n ; \mathbf{Y}_n, \xi, \gamma, \Sigma | \mathbf{x}_n \right) d\theta_n . \tag{4.24}$$

The system of equations is solved using an EM algorithm (Dempster et al., 1977) following the approach of Bock & Aitkin (1981).

Quadrature and Monte Carlo Approximations

The integrals in (4.20) to (4.24) are approximated numerically using either quadrature or Monte Carlo methods. Each case proceeds by defining (Θ_q), $q = 1, \ldots, Q$, a set of Q D-dimensional vectors (referred to as nodes), and for each node defining a corresponding weight $(W_q(\gamma, \Sigma))$. The vector response probability (4.17) is then approximated using

$$f_{\mathbf{x}} \left(\mathbf{x} ; \xi, \gamma, \Sigma \right) = \sum_{p=1}^{Q} f_{\mathbf{x}} \left(\mathbf{x} ; \xi | \Theta_p \right) W_p \left(\gamma, \Sigma \right), \tag{4.25}$$

and the marginal posterior (4.18) is approximated using

$$h_\Theta \left(\Theta_q ; \mathbf{W}_n, \xi, \gamma, \Sigma | \mathbf{x}_n \right) = \frac{f_{\mathbf{x}} \left(\mathbf{x}_n ; \xi | \Theta_q \right) \; W_q \left(\gamma, \Sigma \right)}{\sum_{p=1}^{Q} f_{\mathbf{x}} \left(\mathbf{x}_n ; \xi | \Theta_p \right) W_p \left(\gamma, \Sigma \right)} \tag{4.26}$$

for q=1,...,Q.

The EM algorithm then proceeds as follows:

Step 1. Prepare a set of nodes and weights depending upon $\gamma^{(t)}$ and $\Sigma^{(t)}$, which are the estimates of γ and Σ at iteration t.

Step 2. Calculate the discrete approximation of the marginal posterior density of θ_n, given $\mathbf{x}_n$ at iteration t, using

$$h_\Theta\left(\Theta_q;\mathbf{W}_n,\xi^{(t)},\gamma^{(t)},\Sigma^{(t)}|\mathbf{x}_n\right)=\frac{f_{\mathbf{x}}\left(\mathbf{x}_n;\xi^{(t)}|\Theta_q\right)\ W_q\left(\gamma^{(t)},\Sigma^{(t)}\right)}{\sum_{p=1}^{Q} f_{\mathbf{x}}\left(\mathbf{x}_n;\xi^{(t)}|\Theta_p\right)W_p\left(\gamma^{(t)},\Sigma^{(t)}\right)}, \tag{4.27}$$

where $\xi^{(t)}$, $\gamma^{(t)}$, and $\Sigma^{(t)}$ are estimates of $\xi^{(t)}$, $\gamma^{(t)}$, and $\Sigma^{(t)}$ at iteration t.

Step 3. Use the Newton–Raphson method to solve the following to produce estimates of $\hat{\xi}^{(t+1)}$:

$$\mathbf{A}'\sum_{n=1}^{N}\left[\mathbf{x}_n-\sum_{r=1}^{Q}E_{\mathbf{z}}\left(\mathbf{z}|\Theta_r\right)\ h_\Theta\left(\Theta_r;\mathbf{W}_n,\xi^{(t)},\gamma^{(t)},\Sigma^{(t)}|\mathbf{x}_n\right)\right]=\mathbf{0}\,. \tag{4.28}$$

Step 4. Estimate $\gamma^{(t+1)}$ and $\Sigma^{(t+1)}$, using

$$\hat{\gamma}^{(t+1)}=\left(\sum_{n=1}^{N}\overline{\Theta_n}\mathbf{W}_n^T\right)\left(\sum_{n=1}^{N}\mathbf{W}_n\mathbf{W}_n^T\right)^{-1} \tag{4.29}$$

and

$$\begin{aligned}\hat{\Sigma}^{(t+1)}=&\frac{1}{N}\sum_{n=1}^{N}\sum_{r=1}^{Q}\left(\Theta_r-\gamma^{(t+1)}\mathbf{W}_n\right)\\&\left(\Theta_r-\gamma^{(t+1)}\mathbf{W}_n\right)^T h_\Theta\left(\Theta_r;\mathbf{Y}_n,\xi^{(t)},\gamma^{(t)},\Sigma^{(t)}|\mathbf{x}_n\right),\end{aligned} \tag{4.30}$$

where

$$\bar{\Theta}_n=\quad\sum_{r=1}^{Q}\Theta_r\,h_\Theta\left(\Theta_r;\mathbf{W}_n,\xi^{(t)},\gamma^{(t)},\Sigma^{(t)}|\mathbf{x}_n\right). \tag{4.31}$$

Step 5. Return to step 1.

The difference between the quadrature and Monte Carlo methods lies in the way the nodes and weights are prepared. For the quadrature case, begin by choosing a fixed set of Q points, $(Q_{d1},Q_{d2},\ldots,Q_{dQ})$, for each latent dimension d and then define a set of Q^D nodes that are indexed $r=1,\ldots,Q^D$ and are given by the Cartesian coordinates

$$Q_r=(Q_{1j_1},Q_{2j_2},\ldots,Q_{dj_d})\text{ with }j_1=1,\ldots,Q;j_2=1,\ldots,Q;\ldots;j_d=1,\ldots,Q.$$

The weights are then chosen to approximate the continuous multivariate latent population density (4.16). That is,

$$W_r = K\left(2\pi\right)^{-d/2}\left|\Sigma\right|^{-1/2}\exp\left[-\frac{1}{2}\left(\Theta_r - \gamma\mathbf{W}_n\right)'\Sigma^{-1}\left(\Theta_r - \gamma\mathbf{W}_n\right)\right], \quad (4.32)$$

where K is a scaling factor to ensure that the sum of the weights is 1.

In the Monte Carlo case, the nodes are drawn at random from the standard multivariate normal distribution; and at each iteration, the nodes are rotated, using standard methods, so that they become random draws from a multivariate normal distribution with mean $\gamma\mathbf{W}_n$ and covariance Σ. In the Monte Carlo case, the weight for all nodes is $1/Q$.

For further information on the quadrature approach to estimating the model, see Adams, Wilson, & Wang (1997); and for further information on the Monte Carlo estimation method, see Volodin & Adams (1995).

4.4.2 Conditional Maximum Likelihood

The first step in the derivation of the conditional maximum likelihood (CML) estimators is to compute the probability of a response pattern conditional on that pattern yielding a specific score. More formally, let $\mathbf{R}$ be a vector-valued random variable that is the vector of scores of a response pattern. Then a realization of this variable is $\mathbf{r} = \mathbf{x}^T\mathbf{B}$, where $\mathbf{x}^T$ and $\mathbf{B}$ are as defined in (4.2), and the probability of a response pattern conditional on $\mathbf{R}$ taking the value $\mathbf{r}$ is given by

$$\begin{aligned}
f\left(\mathbf{x};\xi,\gamma,\mathbf{\Sigma}|\mathbf{R}=\mathbf{r}\right) &= \frac{f\left(\mathbf{x};\xi,\gamma,\mathbf{\Sigma},\mathbf{R}=\mathbf{r}\right)}{\sum\limits_{\mathbf{z}\in\Omega_\mathbf{r}} f\left(\mathbf{z};\xi,\gamma,\mathbf{\Sigma},\mathbf{R}=\mathbf{r}\right)} \\
&= \frac{\int f_\mathrm{x}\left(\mathbf{x};\xi,\mathbf{R}=\mathbf{r}|\theta\right) f_\theta\left(\theta;\gamma,\Sigma\right) d\theta}{\sum\limits_{\mathbf{z}\in\Omega_\mathbf{r}} \int f_\mathrm{x}\left(\mathbf{z};\xi,\mathbf{R}=\mathbf{r}|\theta\right) f_\theta\left(\theta;\gamma,\Sigma\right) d\theta} \\
&= \frac{\int \Psi\left(\theta,\xi\right)\exp\left(\mathbf{r}\theta+\mathbf{x}^T\mathbf{A}\xi\right) f_\theta\left(\theta;\gamma,\Sigma\right) d\theta}{\sum\limits_{\mathbf{z}\in\Omega_\mathbf{r}} \int \Psi\left(\theta,\xi\right)\exp\left(\mathbf{r}\theta+\mathbf{z}^T\mathbf{A}\xi\right) f_\theta\left(\theta;\gamma,\Sigma\right) d\theta} \\
&= \frac{\exp\left(\mathbf{x}^T\mathbf{A}\xi\right)\int \Psi\left(\theta,\xi\right)\exp\left(\mathbf{r}\theta\right) f_\theta\left(\theta;\gamma,\Sigma\right) d\theta}{\sum\limits_{\mathbf{z}\in\Omega_\mathbf{r}} \exp\left(\mathbf{z}^T\mathbf{A}\xi\right)\int \Psi\left(\theta,\xi\right)\exp\left(\mathbf{r}\theta\right) f_\theta\left(\theta;\gamma,\Sigma\right) d\theta} \\
&= \frac{\exp\left(\mathbf{x}^T\mathbf{A}\xi\right)}{\sum\limits_{\mathbf{z}\in\Omega_\mathbf{r}} \exp\left(\mathbf{z}^T\mathbf{A}\xi\right)},
\end{aligned} \quad (4.33)$$

where Ω_r is the set of response patterns where the vector of scores is $\mathbf{r}$.

Equation (4.33) shows that the probability of a response pattern conditional on $\mathbf{R}$ taking the value $\mathbf{r}$ is not dependent on the ability θ or its distribution. The consequential advantage of the CML approach is that it provides the same estimates for the item parameters regardless of the choice of the population distribution. As such, the CML item parameter estimator is not influenced by any assumption about the population distribution. The disadvantage is that the population parameters are not estimated. If, as is often the case, the population parameters are of interest, they must be estimated in a second step. The second step involves solving the system of equations (4.21) and (4.22) while assuming that the item parameters are known. Apart from underestimating the uncertainty in the population parameter estimates, the consequences of using the CML item-parameter estimates, in this second step, as if they were true values, are not clear.

In contrast, the maximum likelihood approach provides direct estimates of both item parameters and population parameters. However, it suffers from the risk that if the population distributional assumption is incorrect, the item parameters may be biased.

4.4.3 Estimating Standard Errors

Asymptotic standard errors for the parameter estimates are estimated using the observed Fisher's information. For the unidimensional case, a derivation of the formulae for the observed information is provided in Adams, Wilson, & Wu (1997).

The estimation of asymptotic standard errors using the observed information can be very time-consuming. The matrix that is computed is of dimension $p + r{+}2$, where p is the number of item parameters and r is the number of regression variables; and the computation of each element requires integration over the posterior distribution of each case. The time taken is therefore quadratic in the number of parameters and linear in the number of cases and nodes. Because the estimation of these errors can take considerable time, the following approximations for the error variances are often used:

$$\operatorname{var}\left(\hat{\xi}_i\right) = \sum_{n=1}^{N} \left\{ \operatorname{diag}\left[\mathbf{A}' \left(\int_{\theta_n} E_{\mathbf{z}}\left(\mathbf{z}\mathbf{z}'|\theta_n\right) h_\theta\left(\theta_n; \mathbf{Y}_n, \hat{\xi}, \hat{\beta}, \hat{\sigma}^2 | \mathbf{x}_n\right) d\theta_n \right.\right.\right.$$
$$\left.\left.\left. - \int_{\theta_n} E_{\mathbf{z}}\left(\mathbf{z}|\theta_n\right) E_{\mathbf{z}}\left(\mathbf{z}'|\theta_n\right) h_\theta\left(\theta_n; \mathbf{Y}_n, \hat{\xi}, \hat{\beta}, \hat{\sigma}^2 | \mathbf{x}_n\right) d\theta_n \right) \mathbf{A} \right] \right\}^{-1}, \tag{4.34}$$

$$\operatorname{var}\left(\hat{\beta}_i\right) = \hat{\sigma}^2 \left(\sum_{n=1}^{N} \mathbf{Y}_n \mathbf{Y}_n^T \right)^{-1}, \tag{4.35}$$

$$\text{var}\left(\hat{\sigma}^2\right) = \frac{2\hat{\sigma}^4}{N}. \tag{4.36}$$

These approximations ignore all of the covariances in the parameter estimates. The approximations of the item parameters will generally underestimate the sampling error, particularly for parameters associated with facets that have few levels for the step parameters in multicategory items. The accuracy of (4.35) and (4.36) depends on the magnitude of the measurement error, since it is reflected in the variances of the individual's posterior distributions.

4.4.4 Latent Ability Estimation and Prediction

The marginal item response model (4.17) does not include parameters for the latent values θ_n; and hence the estimation algorithm does not result in estimates of the latent values for persons. While this may not be of concern when the modeling is undertaken for the purposes of estimating population parameters, that is, the elements of γ and Σ, it does cause inconveniences when there is an interest in estimates of the latent values for individuals.

There are a number of standard approaches that can be applied to provide estimates, or perhaps, more accurately, predictions, of the latent values. Perhaps the most common approach is to use expectation of the posterior distribution of θ_n, the so-called expected a posteriori (EAP) (Bock & Aitkin, 1981). The EAP prediction of the latent quantity for case n is

$$\theta_n^{EAP} = \sum_{r=1}^{Q} \Theta_r \, h_\Theta \left(\Theta_r; \mathbf{W}_n, \hat{\xi}, \hat{\gamma}, \hat{\Sigma} | \mathbf{x}_n\right). \tag{4.37}$$

Variance estimates for these predictions can be estimated using

$$\text{var}\left(\theta_n^{EAP}\right) = \sum_{r=1}^{Q} \left(\Theta_r - \theta_n^{EAP}\right) \left(\Theta_r - \theta_n^{EAP}\right)' \; h_\Theta \left(\Theta_r; \mathbf{W}_n, \hat{\xi}, \hat{\gamma}, \hat{\Sigma} | \mathbf{x}_n\right). \tag{4.38}$$

An alternative to the EAP is the maximum a posteriori (MAP) (Bock & Aitkin, 1981), which requires finding the modes, rather than the expectations (means), of the posterior distributions.

A maximum likelihood approach to the estimation of the ability estimates can also be used. Following the weighted likelihood approach of Warm (1985, 1989), this is achieved by solving the equations

$$\sum_{i \in \Omega} \left(\left(\mathbf{b}_{ix_{ni}} + \frac{J_{ni}}{2I_{ni}} \right) - \sum_{j=1}^{K_i} \frac{\mathbf{b}_{ij} \exp\left(\mathbf{b}_{ij}\theta_n + \mathbf{a'}_{ij}\hat{\xi}\right)}{\sum_{k=1}^{K_i} \exp\left(\mathbf{b}_{ik}\theta_n + \mathbf{a'}_{ik}\hat{\xi}\right)} \right) = 0 \tag{4.39}$$

for each case, where $\hat{\xi}$ is the vector of item parameter estimates, I_{ni} is the information function evaluated for item i, and J_{ni} is the derivative of I_{ni} with respect to θ_n. These equations can be readily solved using a routine based on the Newton–Raphson method.

Drawing Plausible Values

The model presented in (4.17) provides estimates of the γ and Σ parameters of the population, but of course there are many other characteristics of the population that may be of interest. In most measurement applications, these parameters would be estimated from point estimates of the θ_n parameters. It is well known, however, that the use of point *estimates* such as the EAP, MLE, and WLE in a two-step approach to estimating population parameters is fraught with challenges.

As an alternative to using point estimates, Mislevy (see Mislevy, 1991, and Mislevy, Beaton, et al., 1992) proposed an approach based on the use of random draws from the marginal posterior, (4.19), for each student. These random draws have become widely known as plausible values.

The following describes a method for drawing plausible values from the posterior distributions. Unlike previously described methods for drawing plausible values (Beaton, 1987; Mislevy, Beaton, et al., 1992), the method described here does not assume normality of the marginal posterior distributions. Recall from (4.19) that the marginal posterior is given by

$$h_\theta\left(\theta_n; \mathbf{W}_n, \xi, \gamma, \Sigma | \mathbf{x}_n\right) = \frac{f_{\mathbf{x}}\left(\mathbf{x}_n; \xi | \theta_n\right)\ f_\theta\left(\theta_n; \mathbf{W}_n, \gamma, \Sigma\right)}{\int_\theta f_{\mathbf{x}}\left(\mathbf{x}; \xi | \theta\right) f_\theta\left(\theta; \mathbf{W}_n, \gamma, \Sigma\right) d\theta}. \tag{4.40}$$

First draw M vector-valued random deviates, $\{j_{mn}\}_{m=1}^{M}$, from the multivariate normal distribution, $f_\theta\left(\theta_n; \mathbf{W}_n, \gamma, \Sigma\right)$, for each case n. These vectors are used to compute an approximation to the integral in the denominator of (4.40), using the Monte Carlo integration

$$\int_\theta f_{\mathbf{x}}\left(\mathbf{x}; \xi | \theta\right) f_\theta\left(\theta, ; \mathbf{W}_n, \gamma, \Sigma\right) d\theta \approx \frac{1}{M} \sum_{m=1}^{M} f_{\mathbf{x}}(\mathbf{x}; \xi | \varphi_{mn}) \equiv \Im\,. \tag{4.41}$$

At the same time, the values

$$p_{mn} = f_{\mathbf{x}}\left(\mathbf{x}_n; \xi | \varphi_{mn}\right)\ f_\theta\left(\varphi_{mn}; \mathbf{W}_n, \gamma, \Sigma\right) \tag{4.42}$$

are calculated, and the set of pairs $\left(\varphi_{mn},\ {}^{p_{mn}}\!/_{\Im}\right)_{m=1}^{M}$ is obtained. This set of pairs can be used as an approximation of the posterior density (4.34); and the probability that φ_{nj} could be drawn from this density is given by

$$q_{nj} \quad = \quad \frac{p_{mn}}{\sum\limits_{m=1}^{M} p_{mn}}\,. \tag{4.43}$$

At this point, L uniformly distributed random numbers, $\{\eta_i\}_{i=1}^{L}$, are generated and for each random draw, the vector, φ_{ni_0}, that satisfies the condition

$$\sum_{s=1}^{i_0-1} q_{sn} < \eta_i \leqslant \sum_{s=1}^{i_0} q_{sn} \tag{4.44}$$

is selected as a plausible vector.

4.5 Generalized Fit Test

A convenient way to assess the fit of the model is to follow the residual-based approach of Wright & Stone (1979) and Wright & Masters (1982). Wu (1997) extended this approach for application with the marginal model used here, and more recently Adams & Wu (2004) generalized the approach so that a range of tests of specific hypotheses could be tested.

If $\mathbf{A}_p$ is used to indicate the pth column of the design matrix $\mathbf{A}$**,** the Wu fit statistic is based on the standardized residual

$$z_{np}(\theta_n) = \left(\mathbf{A}_p^T \mathbf{x}_n - E_{np}\right) \Big/ \sqrt{V_{np}}\ , \tag{4.45}$$

where $\mathbf{A}_p^T \mathbf{x}_n$ is the contribution of person n to the sufficient statistic for parameter p, and E_{np} and V_{np} are, respectively, the conditional expectation and the variance of $\mathbf{A}_p^T \mathbf{x}_n$.

To construct an unweighted fit statistic,[4] the square of this residual is averaged over the cases and then integrated over the posterior ability distributions to obtain

$$\begin{aligned} Fit_{out,p} = \int_{\theta_1}\int_{\theta_2}\cdots\int_{\theta_N} \left[\frac{1}{N}\sum_{n=1}^{N} \hat{z}_{np}^2(\theta_n)\right] \\ \prod_{n=1}^{N} h_\theta\left(\theta_n; \mathbf{Y}_n, \hat{x}, \hat{b}, \hat{\sigma}^2|\mathbf{x}_n\right) d\theta_N d\theta_{N-1}\cdots d\theta_1 . \end{aligned} \tag{4.46}$$

For the weighted fit,[5] a weighted average of the squared residuals is used as follows:

$$\begin{aligned} Fit_{in,p} = \int_{\theta_1}\int_{\theta_2}\cdots\int_{\theta_N} \left[\frac{\sum\limits_{n=1}^{N} \hat{z}_{np}^2(\theta_n)\, V_{np}(\theta_n)}{\sum\limits_{n=1}^{N} V_{np}(\theta_n)}\right] \\ \prod_{n=1}^{N} h_\theta\left(\theta_n; \mathbf{Y}_n, \hat{\xi}, \hat{\beta}, \hat{\sigma}^2|\mathbf{x}_n\right) d\theta_N d\theta_{N-1}\cdots d\theta_1 . \end{aligned} \tag{4.47}$$

[4] Often referred to as *outfit.*

[5] Often referred to as *infit.*

It is convenient to use the Monte Carlo method to approximate the integrals in (4.46) and (4.47). Wu (1997) has shown that the statistics produced by (4.46) and (4.47) have approximate scaled chi-squared distributions. These statistics are transformed to approximate normal deviates using the Wilson–Hilferty transformations

$$t_{\text{out},p} = \left(Fit_{\text{out},p}^{1/3} - 1 + 2/(9rN)\right) \Big/ \left(2/(9rN)\right)^{1/2} \tag{4.48}$$

and

$$t_{\text{in},p} = \left[Fit_{\text{in},p}^{1/3} - 1\right] \times \frac{3}{\sqrt{\text{Var}\left(\text{Fit}_{in,p}\right)}} + \frac{\sqrt{\text{Var}\left(\text{Fit}_{in,p}\right)}}{3}, \tag{4.49}$$

where r is the number of draws used in the Monte Carlo approximation of the integrals in (4.40) and (4.41) and

$$\text{Var}(\text{Fit}_i n, p) = \left(\frac{1}{\sum_n V_{np}}\right)^2 \left(\sum_n \left(E\left(\left(\mathbf{A}_p^T \mathbf{X}_n - E_{np}\right)^4\right) - V_{np}^2\right)\right). \tag{4.50}$$

The derivation and justification for these transformations is given in Wu (1997).

The fit-testing approach described here works at the parameter level; that is, it provides a fit statistic for each of the estimated item parameters. A more general approach was introduced by Adams & Wu (2004), who suggested that the matrix $\mathbf{A}$ that is used in (4.39) could be replaced with an alternative matrix, $\mathbf{F}$ which they called a fit matrix.

Since the derivation of the fit statistics described in the previous section is based on the comparison of a linear combination of item responses, $\mathbf{A}_p^T \mathbf{x}_n$, and its expectation and variance, the fit statistics can be generalized to include any linear combinations of the item responses, and not necessarily be limited to $\mathbf{A}_p^T \mathbf{x}_n$, where $\mathbf{A}_p$ is the design vector for the parameter ξ_p. If $\mathbf{F}_u$ is any vector of the same length as $\mathbf{A}_p$, then $\mathbf{F}_u^T \mathbf{x}_n$ is a linear combination of the item responses of person n. One can compute the expectation and variance of $\mathbf{F}_u^T \mathbf{x}_n$, and construct a fit statistic in exactly the same way as for $\mathbf{A}_p^T \mathbf{x}_n$. The following is an example for constructing user-defined fit tests for a simple dichotomous RM.

Consider a test consisting of 10 dichotomous items; the design matrix, $\mathbf{A}$, for the simple logistic model for such a test would be a 10 by 10 identity matrix.

Using the notation defined earlier, the first column of $\mathbf{A}$ is $\mathbf{A}_1^T = \left(1\;0\;0\;0\;0\;0\;0\;0\;0\;0\right)$. The product $\mathbf{A}_1^T \mathbf{x}_n$ gives the item response of person n on item 1. This is the contribution of person n to the sufficient statistic for the first item parameter.

Similarly, $\mathbf{A}_2^T = \left(0\;1\;0\;0\;0\;0\;0\;0\;0\;0\right)$, and $\mathbf{A}_2^T \mathbf{x}_n$ is the contribution of person n to the sufficient statistic for the second item parameter, and so on.

For a user-defined fit test, the design vector $\mathbf{A}_p$ can be replaced by any arbitrary vector $\mathbf{F}_u$. Consider the fit design matrix

$$\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} . \tag{4.51}$$

If $\mathbf{F}_1$ and $\mathbf{F}_2$ are the first and second columns of $\mathbf{F}$, then the product $\mathbf{F}_1^T \mathbf{x}_n$ gives the total score on the first five items for person n. Similarly, $\mathbf{F}_2^T \mathbf{x}_n$ gives the total score on the last five items for person n.

Adams & Wu (2004) showed how the fit statistics based on $\mathbf{F}_1$ and $\mathbf{F}_2$ worked well as a test of the hypothesis that the first and second five items were tapping into two different latent dimensions, whereas the fit tests given in (4.46) and (4.47) failed to identify the multidimensionality of the test items.

As a second possible set of fit tests, consider the matrix in (4.52). A fit test based on the first column of this matrix tests whether items one and six are both answered correctly as often as would be expected under the model. Similarly, the second column provides a test of whether items two and seven are both answered correctly as often as would be expected under the model. As such, these are tests of the local independence of items one and six, and two and seven respectively.

The third column compares the score on the first five items with its expectation, that is, whether the subtest consisting of the first five items fits with the rest of the items as predicted by the model:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} . \tag{4.52}$$

4.6 Conclusion

This paper has demonstrated the flexibility of using design matrices to specify a family of item response models. Not only can standard item response models such as the partial-credit, the rating-scale, and the facets models be included under one single framework of models, but many other models can be specified through user-defined design matrices.

The estimation procedures described in this paper allow for a joint (or one-step) calibration of both item parameters and population parameters, as opposed to a two-step process in which individual student abilities are first estimated and then aggregated to form population parameter estimates. The advantages of a joint calibration of parameters include more accurate standard errors for the estimates of the population parameters and less bias of some population parameter estimates.

Similarly, user-defined fit-design matrices allow for more focused testing of goodness-of-fit of the data to the model. In many cases, such focused fit tests are statistically more powerful in detecting misfit in the data.

However, the theoretical elegance of the use of design matrices can be overshadowed by the tediousness of the construction of these matrices in practice. A software package, ConQuest (Wu et al., 1997), has been developed in which users can specify various item response models through a command language. The design matrices are then automatically built by ConQuest. ConQuest also allows users to import a design matrix should the need arise. Thus the advantages of a unified framework of item response models can be easily implemented in practice for the analysis of a vast range of data sets.