Research Report

On the Latent Regression Model of Item Response Theory

Tamás Antal

Research & Development

March 2007
RR-07-12

On the Latent Regression Model of Item Response Theory

Tamás Antal

ETS, Princeton, NJ

March 2007

As part of its educational and social mission and in fulfilling the organization's nonprofit charter and bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Copyright © 2007 by Educational Testing Service. All rights reserved.

ETS and the ETS logo are registered trademarks of
Educational Testing Service (ETS).

Abstract

Full account of the latent regression model for the National Assessment of Educational Progress is given. The treatment includes derivation of the EM algorithm, Newton-Raphson method, and the asymptotic standard errors. The paper also features the use of the adaptive Gauss-Hermite numerical integration method as a basic tool to evaluate expectations necessary to perform the parameter estimations.

Key words: Latent regression, EM algorithm, Newton-Raphson algorithm, Gauss-Hermite quadrature, NAEP

1 Introduction

Educational assessments such as the National Assessment of Educational Progress (NAEP) and Trends in International Mathematics and Science Studies contain broad content frameworks and consequently large item pools. Yet these assessments are governed by restrictions on the time and, subsequently, the number of items each student is requested to respond to. An efficient way to address this dilemma is to systematically distribute items among students in a so-called matrix design. In a matrix design, a given student is only administered a certain randomly assigned subset of items.

Not-so-pleasant implications of this matrix design are that only limited information is available about a specific individual and that comparisons between students are complicated. To further enhance the ability to produce the required set of estimates, this paper investigates a method to address these issues in which a large collection of *background* variables is created and responses corresponding to these variables are collected via several survey instruments. A regression of student's ability on these background variables is then incorporated into the existing marginal item response theory (IRT) model to create the main object of interest in this paper, the *latent regression item response theory model.* When only population subgroup relevant information is the concern of the assessment, as is the case with NAEP, this approach has been found to be satisfactory and became the workhorse of this sort of large scale assessments (Mislevy, 1984, 1985).

The structure of the paper is as follows. Some preliminary notes on mathematical notations are followed by the detailed account of the latent regression item response theory model. The next three sections present the usual way of parameter estimation via maximum likelihood method coupled with an application of the successive approximation technique (EM algorithm). Special emphasis to the adaptive Gauss-Hermite numerical integration method. There is a section about the common population IRT model (mainly for comparison purposes) followed by a section about the detailed account of the computation of asymptotic standard errors. Concrete derivations of the Newton method is given along with a discussion about its implementability in the latent regression framework complete the paper.

2 Notes on Notations

The definitions and notations introduced in this section are fairly standard in linear algebra. Nevertheless, they are included here in an effort to make the paper self-contained. The interested reader may find Halmos (1974), Harville (1997), and Magnus and Neudecker (1999) useful for a further study in linear algebra.

Tensor products of vectors are defined as

$$\otimes^K : \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_K} \to \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_K},$$

$$(a_1, \ldots, a_K) \mapsto a_1 \otimes \cdots \otimes a_K,$$

where

$$(a_1 \otimes \cdots \otimes a_K)_{i_1 \ldots i_K} = (a_1)_{i_1} \ldots (a_K)_{i_K}.$$

As an example, observe the map,

$$\mathbb{R}^n \otimes \mathbb{R}^n \to M_n(\mathbb{R}), \quad (a \otimes b)_{ij} = a_i b_j.$$

A general tensor, by definition, is a linear combination of elements of the form $a_1 \otimes \cdots \otimes a_K$. The extension of the tensor product to matrices and higher order tensors follow the same idea (e.g., $(A \otimes B)_{ijkl} = A_{ij} B_{kl}$, where A and B are ordinary square matrices). As a shorthand for the n-fold tensor product,

$$a^{\otimes n} = a \otimes \cdots \otimes a.$$

The symmetric tensor product of two tensors is defined as

$$A \otimes_s B = \frac{1}{2}(A \otimes B + B \otimes A).$$

The scalar product is a bilinear map,

$$\langle \mid \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \ (a, b) \mapsto \langle a \mid b \rangle := \sum_{i=1}^{n} a_i b_i.$$

It naturally extends to any tensor product,

$$\langle \mid \rangle : \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_K} \times \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_K} \to \mathbb{R},$$

$$(A, B) \mapsto \langle A \mid B \rangle := \sum_{i_1, \ldots, i_K} A_{i_1 \ldots i_K} B_{i_1 \ldots i_K}.$$

It will be convenient for the purposes of this paper to extend the scalar product for a specified set of indices of a tensor. The result is going to be a tensor of type determined by the unused indices. That is,

$$\langle \mid \rangle_{I_1 \ldots I_r, J_1 \ldots J_r} : \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_K} \times \mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_L} \to \mathbb{R}^N,$$

$$(A,B) \mapsto \langle A \mid B \rangle_{I_1 \dots I_r, J_1 \dots J_r} := \sum_{i_1,\dots,i_r} A_{\dots i_1 \dots i_r \dots} B_{\dots i_1 \dots i_r \dots},$$

where the summation indices of A appear at the positions $I_1, \dots, I_r$ and the that of B appear at the positions $J_1, \dots, J_r$. Moreover, $N = \frac{\prod_{i=1}^{K} \prod_{j=1}^{L} n_i n_j}{n_{I_1} \dots n_{I_r} n_{J_1} \dots n_{J_r}}$. An example may illuminate the idea better:

$$\langle A \mid B \rangle_{12,14} = \sum_{i,j} A_{ij..} B_{i..j}, \quad \text{that is} \quad (\langle A \mid B \rangle_{12,14})_{klmn} = \sum_{i,j} A_{ijkl} B_{imnj}.$$

Even the usual matrix multiplication can be expressed in this way:

$$Aa = \langle A \mid a \rangle_{2,1} = \sum_i A_{.i} a_i \in \mathbb{R}^n, \qquad A \in M_n(\mathbb{R}),\ a \in \mathbb{R}^n,$$
$$AB = \langle A \mid B \rangle_{2,1} = \sum_i A_{.i} B_{i.} \in M_n(\mathbb{R}), \qquad A, B \in M_n(\mathbb{R}).$$

The quadratic form is defined as,

$$\langle\ \mid\ \mid\ \rangle : \mathbb{R}^n \times M_n(\mathbb{R}) \times \mathbb{R}^n \to \mathbb{R},\ (a, A, b) \mapsto \langle a \mid A \mid b \rangle := \sum_{i,j=1}^{n} a_i A_{ij} b_j,$$

which can also be written as a scalar product of two tensors:

$$\langle a \mid A \mid b \rangle = \langle a \otimes b \mid A \rangle.$$

3 Likelihood of Latent Regression

The *marginal latent regression item response theory* (MLR-IRT) model is given by the log-likelihood (Mislevy, 1984, 1985; von Davier, Sinharay, Oranje, & Beaton, 2007),

$$\boxed{L = \sum_{i=1}^{N} \log \int_{\mathbb{R}^K} P(y_i \mid \theta) \varphi(\theta; \Gamma x_i, \Sigma) \mathrm{d}^K \theta.} \tag{1}$$

where

$$L_i(\theta) := P(y_i \mid \theta, \mathcal{B}) = \prod_{j=1}^{J} P^{\mathrm{3pl,pc}}(y_{ij} | \theta, \beta_j) \tag{2}$$

is the likelihood of the response vector y_i of student i given the ability $\theta \in \mathbb{R}^K$ (where K is the number of dimensions or subscales) and item parameters $\mathcal{B} = (\beta_1, \dots, \beta_J)$. The number of response categories m_j for item j splits the item pool into two disjoint subsets: dichotomous items ($m_j = 2$), and polytomous items ($m_j > 2$). For dichotomous items $\beta_j = (a_j, b_j, c_j)$ and for polytomous ones $\beta_j = (a_j, b_{j1}, ..., b_{jm_j})$. The *design* vector $q_j \in \{0,1\}^K$ for item j is also introduced so that $q_{j,k} = 1$ if item j is associated with subscale k; otherwise $q_{j,k} = 0$.

The probability of the actual response of student i to item j as a function of $\theta \in \mathbb{R}^K$ is given by

$$P^{\mathrm{3pl}}(y_{ij} \mid \theta, \beta_j) = \begin{cases} c_j + \frac{(1-c_j)}{1+e^{-a_j(\langle q_j \mid \theta \rangle - b_j)}}, & \text{if} \quad y_{ij} = 1, \\ (1-c_j)\left(1 - \frac{1}{1+e^{-a_j(\langle q_j \mid \theta \rangle - b_j)}}\right), & \text{if} \quad y_{ij} = 0, \end{cases} \tag{3}$$

for dichotomous items and by

$$P^{\mathrm{pc}}(y_{ij} \mid \theta, \beta_j) = \frac{e^{\sum\limits_{v=1}^{h} a_j(\langle q_j \mid \theta \rangle - b_{jv})}}{\sum\limits_{u=1}^{m_j} e^{\sum\limits_{w=1}^{u} a_j(\langle q_j \mid \theta \rangle - b_{jw})}}, \qquad \text{if} \quad y_{ij} = h \in \{1, \ldots, m_j\}, \tag{4}$$

for polytomous items (Birnbaum, 1968; Muraki, 1992).

The population distribution is multivariate normal

$$\varphi(\theta; \Gamma x_i, \Sigma) = \frac{1}{(2\pi)^{k/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}\langle \theta - \Gamma x_i \mid \Sigma^{-1} \mid \theta - \Gamma x_i \rangle}. \tag{5}$$

Here, $x_i \in \mathbb{R}^L$ is the vector of background variables, $\Gamma \in M_{K,L}(\mathbb{R})$ is the matrix of regression coefficients, while $\Sigma \in M_K(\mathbb{R})$ is the covariance matrix of the subscales. Note, that while Σ is common across the population, the mean Γx_i is governed by the background variables and can be different for each student. Also introduced is

$$\mathcal{N}(L_i) = \int_{\mathbb{R}^K} P(y_i \mid \theta) \varphi(\theta; \Gamma x_i, \Sigma) \mathrm{d}^K \theta, \tag{6}$$

for the normalization of the likelihood L_i.

4 Maximum Likelihood Solution for Latent Regression

Useful references for the calculus involved in the rest of the paper may be Harville (1997) and Magnus and Neudecker (1999). The goal is to maximize the log-likelihood (1) with respect to Γ and Σ assuming that item parameters are known. To this end, first compute the derivative of L with respect to the matrix element γ_{kl} of Γ:

$$\begin{aligned} \frac{\partial L}{\partial \gamma_{kl}} &= -\frac{1}{2} \sum_{i=1}^{N} \frac{P(y_i \mid \theta)}{\mathcal{N}(L_i)} \int_{\mathbb{R}^K} \frac{\partial \left\langle\!\left\langle \theta - \Gamma x_i \,\middle|\, \Sigma^{-1} \,\middle|\, \theta - \Gamma x_i \right\rangle\!\right\rangle}{\partial \gamma_{kl}} \varphi(\theta; \Gamma x_i, \Sigma) \mathrm{d}^K \theta \\ &= -\frac{1}{2} \sum_{i=1}^{N} \int_{\mathbb{R}^K} \frac{\partial \left\langle\!\left\langle \theta - \Gamma x_i \,\middle|\, \Sigma^{-1} \,\middle|\, \theta - \Gamma x_i \right\rangle\!\right\rangle}{\partial \gamma_{kl}} \mathrm{d}\mu_i, \end{aligned} \tag{7}$$

where μ_i is the probability measure for student i: $\mathrm{d}\mu_i = \frac{P(y_i|\theta)}{\mathcal{N}(L_i)}\varphi(\theta;\Gamma x_i,\Sigma)\mathrm{d}^K\theta$. Now,

$$\begin{aligned}
&\frac{\partial}{\partial\gamma_{kl}}\langle\!\langle \theta-\Gamma x_i \,|\, \Sigma^{-1} \,|\, \theta-\Gamma x_i\rangle\!\rangle \\
= \;& -\langle\!\langle (0,\ldots,\overset{(k)}{x}_{il},\ldots,0) \,|\, \Sigma^{-1} \,|\, \theta-\Gamma x_i\rangle\!\rangle \\
& -\langle\!\langle \theta-\Gamma x_i \,|\, \Sigma^{-1} \,|\, (0,\ldots,\overset{(k)}{x}_{il},\ldots,0)\rangle\!\rangle \\
= \;& x_{il}(\Sigma^{-1}\Gamma x_i)_k + (\Gamma x_i\cdot\Sigma^{-1})_k x_{il} - x_{il}(\Sigma^{-1}\theta)_k - (\theta\cdot\Sigma^{-1})_k x_{il}. \qquad (8)
\end{aligned}$$

Here $\overset{(k)}{x}$ means that x appears at the kth position. $a\cdot B$ denotes the usual right action of matrix B on the vector a:

$$(a\cdot B)_k = \sum_r a_r B_{rk}.$$

Using (8), the equation $\frac{\partial L}{\partial\gamma_{kl}} = 0$ can be written as

$$\sum_{i=1}^{N} x_{il}(\Sigma^{-1}\Gamma x_i)_k + (\Gamma x_i\cdot\Sigma^{-1})_k x_{il} = \sum_{i=1}^{N} x_{il}(\Sigma^{-1}\tilde{\theta}_i)_k + (\tilde{\theta}_i\cdot\Sigma^{-1})_k x_{il} \qquad (9)$$

which, using that Σ is symmetric, becomes

$$\sum_{i=1}^{N} x_{il}(\Sigma^{-1}\Gamma x_i)_k = \sum_{i=1}^{N} x_{il}(\Sigma^{-1}\tilde{\theta}_i)_k. \qquad (10)$$

Here $\tilde{\theta}_i$ is the expectation of θ for student i:

$$\tilde{\theta}_i := \int_{\mathbb{R}^K} \theta\mathrm{d}\mu_i. \qquad (11)$$

Multiplying with Σ yields

$$\sum_{i=1}^{N} x_{il}(\Gamma x_i)_k = \sum_{i=1}^{N} x_{il}\tilde{\theta}_{ik} \qquad (12)$$

This is then rewritten as

$$\langle\!\langle \sum_{i=1}^{N} x_i\otimes x_i \,|\, \Gamma^t\rangle\!\rangle_{2,1} = \sum_{i=1}^{N} x_i\otimes\tilde{\theta}_i \qquad (13)$$

Therefore, after multiplying (13) by $\left(\sum_{i=1}^{N} x_i\otimes x_i\right)^{-1}$, the result is

$$\boxed{\Gamma^t = \left(\sum_{i=1}^{N} x_i\otimes x_i\right)^{-1}\left(\sum_{i=1}^{N} x_i\otimes\tilde{\theta}_i\right).} \qquad (14)$$

Note that this is an implicit equation, since for $\tilde{\theta}_i$ the knowledge of Γ and Σ would be required.

The derivative of L with respect to the matrix element $\sigma_{kk'}$ of Σ is

$$\begin{aligned} \frac{\partial L}{\partial \sigma_{kk'}} &= \sum_{i=1}^{N} \int_{\mathbb{R}^K} \frac{\partial \det(\Sigma)^{-\frac{1}{2}}}{\partial \sigma_{kk'}} \det(\Sigma)^{\frac{1}{2}} \mathrm{d}\mu_i \\ &\quad -\frac{1}{2} \sum_{i=1}^{N} \int_{\mathbb{R}^K} \frac{\partial \langle\!\langle \theta - \Gamma x_i \,\big|\, \Sigma^{-1} \,\big|\, \theta - \Gamma x_i \rangle\!\rangle}{\partial \sigma_{kk'}} \mathrm{d}\mu_i. \end{aligned} \tag{15}$$

To proceed, use the following computations.

$$\det(\Sigma)^{\frac{1}{2}} \frac{\partial \det(\Sigma)^{-\frac{1}{2}}}{\partial \sigma_{kk'}} = -\frac{1}{2} \det(\Sigma)^{-1} (2 - \delta_{kk'}) \xi_{kk'} = -\frac{1}{2} (2 - \delta_{kk'}) \Sigma^{-1}_{kk'}, \tag{16}$$

where $\delta_{kk'} = 1$ if $k = k'$; otherwise $\delta_{kk'} = 0$. Also, ξ denotes the adjoint matrix of Σ.

The derivative of the quadratic form is computed using

$$\frac{\partial \Sigma^{-1}}{\partial \sigma_{kk'}} = \frac{1}{2} (\delta_{kk'} - 2) \Sigma^{-1} (e_k \otimes e_{k'} + e_{k'} \otimes e_k) \Sigma^{-1} = (\delta_{kk'} - 2) (\Sigma^{-1} \otimes_s \Sigma^{-1})_{kk'}, \tag{17}$$

where for $e_k \in \mathbb{R}^K$, $e_{k\kappa} = \delta_{k\kappa}$ and

$$(\Sigma^{-1} \otimes_s \Sigma^{-1})_{kk'} = \frac{1}{2} \Sigma^{-1} (e_k \otimes e_{k'} + e_{k'} \otimes e_k) \Sigma^{-1}.$$

Then,

$$\begin{aligned} & \frac{\partial \langle\!\langle \theta - \Gamma x_i \,\big|\, \Sigma^{-1} \;\big|\; \theta - \Gamma x_i \rangle\!\rangle}{\partial \sigma_{kk'}} \\ = \;& (\delta_{kk'} - 2) \langle\!\langle \theta - \Gamma x_i \;\big|\; (\Sigma^{-1} \otimes_s \Sigma^{-1})_{kk'} \;\big|\; \theta - \Gamma x_i \rangle\!\rangle, \end{aligned} \tag{18}$$

Using $\theta - \Gamma x_i = (\theta - \tilde{\theta}_i) + (\tilde{\theta}_i - \Gamma x_i)$,

$$\begin{aligned} \frac{\partial L}{\partial \sigma_{kk'}} &= -\frac{1}{2} (2 - \delta_{kk'}) \Bigg(\sum_{i=1}^{N} \Sigma^{-1}_{kk'} \\ &\quad - \sum_{i=1}^{N} \mathcal{E} \langle\!\langle \theta - \tilde{\theta}_i \,\big|\, (\Sigma^{-1} \otimes_s \Sigma^{-1})_{kk'} \;\big|\; \theta - \tilde{\theta}_i \rangle\!\rangle \\ &\quad - \sum_{i=1}^{N} \langle\!\langle \tilde{\theta}_i - \Gamma x_i \,\big|\, (\Sigma^{-1} \otimes_s \Sigma^{-1})_{kk'} \;\big|\; \tilde{\theta}_i - \Gamma x_i \rangle\!\rangle \Bigg). \end{aligned} \tag{19}$$

That is

$$\begin{aligned} (2 - \mathbb{I})^{-1} \frac{\partial L}{\partial \Sigma} &= -\frac{1}{2} \sum_{i=1}^{N} \Sigma^{-1} \\ &\quad + \frac{1}{2} \sum_{i=1}^{N} \langle\!\langle \Sigma^{-1} \otimes_s \Sigma^{-1} \,\big|\, \mathcal{E} \left((\theta - \tilde{\theta}_i) \otimes (\theta - \tilde{\theta}_i) \right) \rangle\!\rangle_{34} \\ &\quad + \frac{1}{2} \sum_{i=1}^{N} \langle\!\langle \Sigma^{-1} \otimes_s \Sigma^{-1} \,\big|\, (\tilde{\theta}_i - \Gamma x_i) \otimes (\tilde{\theta}_i - \Gamma x_i) \rangle\!\rangle_{34}. \end{aligned} \tag{20}$$

Setting $\frac{\partial L}{\partial \Sigma} = 0$ and multiplying it with $\Sigma \otimes_s \Sigma$, the result is

$$\boxed{\Sigma = \frac{1}{N} \sum_{i=1}^{N} \left[(\tilde{\theta}_i - \Gamma x_i) \otimes (\tilde{\theta}_i - \Gamma x_i) + \mathcal{E} \left((\theta - \tilde{\theta}_i) \otimes (\theta - \tilde{\theta}_i) \right) \right].} \tag{21}$$

Here is introduced

$$\begin{aligned} \check{\Sigma}^{(i)} &:= \mathcal{E} \left((\theta - \tilde{\theta}_i) \otimes (\theta - \tilde{\theta}_i) \right) \\ &= \int_{\mathbb{R}^K} (\theta - \tilde{\theta}_i) \otimes (\theta - \tilde{\theta}_i) \, \frac{P(y_i|\theta)}{\mathcal{N}(L_i)} \varphi(\theta; \Gamma x_i, \Sigma) \mathrm{d}^K \theta. \end{aligned} \tag{22}$$

5 Computing Expectations

To find the maximum place of the likelihood of the latent regression, several integrals need to be computed: one for the normalization constant of individual's likelihood

$$\boxed{\mathcal{N}(L_i) = \int_{\mathbb{R}^K} P(y_i \mid \theta) \varphi(\theta; \Gamma x_i, \Sigma) \mathrm{d}^K \theta,} \tag{23}$$

another for the expectation of the latent trait

$$\boxed{\tilde{\theta}_i := \int_{\mathbb{R}^K} \theta \, \frac{P(y_i|\theta)}{\mathcal{N}(L_i)} \varphi(\theta; \Gamma x_i, \Sigma) \mathrm{d}^K \theta,} \tag{24}$$

and one for the expectation of the variance of the latent trait

$$\boxed{\check{\Sigma}^{(i)} := \int_{\mathbb{R}^K} (\theta - \tilde{\theta}_i) \otimes (\theta - \tilde{\theta}_i) \, \frac{P(y_i|\theta)}{\mathcal{N}(L_i)} \varphi(\theta; \Gamma x_i, \Sigma) \mathrm{d}^K \theta.} \tag{25}$$

The computations in NAEP are carried out using numerical quadrature for $K < 3$ and Laplace approximation for $K > 2$.

The detailed account of adaptive numerical integration is given in (Antal & Oranje, 2007), here we only provide a short description of the major tools. In normal population marginal IRT the implemented numerical integration procedures aim to compute multidimensional integrals of the form

$$I := \int_{\mathbb{R}^K} f(x) e^{-\langle x \mid x \rangle} \mathrm{d}^K x, \tag{26}$$

where f is a smooth function. The rectangle rule approximates the integral by

$$I \cong \sum_{q \in \mathcal{QP}} f(q) e^{-\langle q \mid q \rangle} \Delta q, \tag{27}$$

where $\mathcal{QP} = \{q_1, \ldots, q_Q\}^K$ with $q_i = q_{min} + \frac{q_{max}-q_{min}}{Q-1}(i-1), \;\; (i = 1, \ldots, Q)$. $\Delta q = \left(\frac{q_{max}-q_{min}}{Q-1}\right)^K$. An example would be $[-4, 4]$ or $[-5, 5]$ with $Q = 41$.

Gaussian quadrature (Stoer & Bulirsch, 2002, pp. 171–180) provides a tool that makes it possible to compute higher dimensional integrals without compromising computational precision. With the Rth Gauss-Hermite quadrature, the integral is approximated as

$$I \cong \sum_{q \in \mathcal{QP}^K_{GH_R}} f(q) w_q, \tag{28}$$

where $\mathcal{QP}_{GH_R}$ is the set of the zeros of the Rth Hermite polynomial H_R and $\mathcal{QP}^K_{GH_R}$ is the Kth Cartesian power of $\mathcal{QP}_{GH_R}$. The *weights* are given by $w_q = w_{q_1} w_{q_2} \ldots w_{q_K}$, where

$$w_{q_i} = \frac{2^{R-1} R! \sqrt{\pi}}{R^2 H_{R-1}(q_i)^2}. \tag{29}$$

If the number of items is relatively large, it is possible that the response likelihood $P(y_i \mid \theta, \beta)$ has a sharp peak at a location depending on the item parameters β and the item responses y_i. It is then possible that an integration technique based on finite number of function evaluations fails to sufficiently capture the behavior of the response likelihood. While this is very uncommon in NAEP, where the number of items per subscale rarely exceeds 10, this issue is addressed here for the sake of completeness.[1] In addition, it is expected that a method more cognizant of the actual behavior of the response likelihood may be computationally more efficient even for tamer response likelihoods.

One way of taking the peak of the response likelihood into consideration finds the modal multivariate normal approximation

$$P(y_i \mid \theta, \beta) \cong \varphi(\theta; \theta_i^m, \Sigma_i^m), \tag{30}$$

where θ_i^m is the mode of $P(y_i \mid \theta, \beta)$ and Σ_i^m is the modal covariance matrix of $P(y_i \mid \theta, \beta)$. More precisely, θ_i^m is obtained as the solution of

$$\frac{\partial P(y_i \mid \theta, \beta)}{\partial \theta} = 0, \qquad (\theta = ?), \tag{31}$$

and the modal covariance is defined by

$$\Sigma_i^m = \left. \left(- \frac{\partial^2 \log P(y_i \mid \theta, \beta)}{\partial \theta^2} \right)^{-1} \right|_{\theta = \theta_i^m}. \tag{32}$$

For an arbitrary smooth function $g(\theta)$, proceed with the integration as follows:

$$\begin{aligned}
\mathcal{E}(g)_i &= \int_{\mathbb{R}^K} g(\theta)\frac{P(y_i \mid \theta,\beta)}{\mathcal{N}(L_i)}\varphi(\theta;\Gamma x_i,\Sigma)\mathrm{d}^K\theta \\
&= \int_{\mathbb{R}^K} g(\theta)\frac{P(y_i \mid \theta,\beta)}{\mathcal{N}(L_i)\varphi(\theta,\theta_i^m,\Sigma_i^m)}\varphi(\theta;\theta_i^m,\Sigma_i^m)\varphi(\theta;\Gamma x_i,\Sigma)\mathrm{d}^K\theta \\
&= \int_{\mathbb{R}^K} g(\theta)\frac{P(y_i \mid \theta,\beta)}{\mathcal{N}(L_i)\varphi(\theta;\theta_i^m,\Sigma_i^m)}C_i\varphi(\theta;\theta_i^p,\Sigma_i^p)\mathrm{d}^K\theta,
\end{aligned}$$

where

$$\Sigma_i^p = \left(\Sigma^{-1} + (\Sigma_i^m)^{-1}\right)^{-1}, \tag{33}$$

$$\theta_i^p = \Sigma_i^p(\Sigma^{-1}\Gamma x_i + (\Sigma_i^m)^{-1}\theta_i^m), \tag{34}$$

and

$$C_i = \frac{\sqrt{|\Sigma_i^p|}}{(2\pi)^{K/2}\sqrt{|\Sigma_i^m||\Sigma|}}. \tag{35}$$

Then, one finds the Cholesky decomposition $T_iT_i^t = 2\Sigma_i^p$ and performs the change of variables

$$z = T_i^{-1}(\theta - \theta_i^p), \quad \theta = T_iz + \theta_i^p \tag{36}$$

to obtain the Gauss-Hermite rule

$$\mathcal{E}(g)_i \cong C_i \sum_{q\in\mathcal{QP}_{GH_R}^K} g(T_iq + \theta_i^p)\frac{P(y_i|T_iq + \theta_i^p,\beta)}{\mathcal{N}(L_i)\varphi(T_iq + \theta_i^p,\theta_i^m,\Sigma_i^m)}w_q. \tag{37}$$

When the approximation (30) is good, then the function $\frac{P(y_i \mid \theta,\beta)}{\varphi(\theta;\theta_i^m,\Sigma_i^m)}$ is approximately constant in the range where the normal integration weight $\varphi(\theta;\theta_i^p,\Sigma_i^p)$ is not negligible.

Because this computation uses additional information about the integrand (i.e., the method adapts itself to the integrand), the technique is sometimes referred to as adaptive numerical integration.

6 EM Algorithm

The previous two sections presented five computational steps towards maximizing the likelihood (1) in terms of the latent regression coefficients Γ and Σ—(14), (21), and three expectations (23, (24), and (25). All of these formulae are implicit though, as they depend on parameters to be estimated. One way of solving this system of integral equations is to construct an iterative scheme out of the the five building blocks. This is usually termed as an expectation-maximization (EM) algorithm; in this case, however, the maximization steps (14) and

(21) are trivial. This particular solution can be thought of as a *successive approximation* scheme. The parameters gain an extra index to indicate the iteration status. That is, the initial value for Γ is $\Gamma^{(0)}$. By choosing $\Gamma^{(0)} = 0$ and $\Sigma^{(0)} = Id_K$ it is possible to compute the expectations as follows

$$\mathcal{N}(L_i)^{(0)} = \int_{\mathbb{R}^K} P(y_i \mid \theta)\varphi(\theta; \Gamma^{(0)}x_i, \Sigma^{(0)})\mathrm{d}^K\theta, \tag{38}$$

$$\tilde{\theta}_i^{(0)} := \int_{\mathbb{R}^K} \theta \, \frac{P(y_i|\theta)}{\mathcal{N}(L_i)^{(0)}}\varphi(\theta; \Gamma^{(0)}x_i, \Sigma^{(0)})\mathrm{d}^K\theta, \tag{39}$$

and

$$\mathrm{Var}(\theta_i)^{(0)} = \int_{\mathbb{R}^K} (\theta - \tilde{\theta}_i^{(0)}) \otimes (\theta - \tilde{\theta}_i^{(0)}) \, \frac{P(y_i|\theta)}{\mathcal{N}(L_i)^{(0)}}\varphi(\theta; \Gamma^{(0)}x_i, \Sigma^{(0)})\mathrm{d}^K\theta. \tag{40}$$

Using these initial estimates, it is now possible to obtain updates for Γ and Σ via the so-called M-steps (14) and (21).

The iteration stops when the prescribed convergence threshold is reached.

7 Common Population Estimation

The common population approach aims to maximize the log-likelihood (1) with respect to the population mean and and population covariance without the regression effect. That is, one assumes that the log likelihood is of the form

$$L = \sum_{i=1}^{N} \log \int_{\mathbb{R}^K} P(y_i \mid \theta)\varphi(\theta; \mu, \Sigma)\mathrm{d}^K\theta. \tag{41}$$

For the derivatives of L with respect to the population parameters μ and Σ,

$$\begin{aligned} \frac{\partial L}{\partial \mu} &= -\frac{1}{2}\sum_{i=1}^{N} \frac{1}{\mathcal{N}(L_i)} \int_{\mathbb{R}^K} P(y_i \mid \theta)\frac{\partial \langle\!\langle \theta - \mu \,|\, \Sigma^{-1} \,|\, \theta - \mu \rangle\!\rangle}{\partial \mu}\varphi(\theta; \mu, \Sigma)\mathrm{d}^K\theta \\ &= -\frac{1}{2}\sum_{i=1}^{N} \frac{1}{\mathcal{N}(L_i)} \int_{\mathbb{R}^K} P(y_i \mid \theta)\Sigma^{-1}(\theta - \mu)\varphi(\theta; \mu, \Sigma)\mathrm{d}^K\theta. \end{aligned} \tag{42}$$

Setting $\frac{\partial L}{\partial \mu} = 0$, the implicit equation for the estimate of μ is:

$$\boxed{\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} \int_{\mathbb{R}^K} \theta \frac{P(y_i \mid \theta)}{\mathcal{N}(L_i)}\varphi(\theta; \hat{\mu}, \hat{\Sigma})\mathrm{d}^K\theta.} \tag{43}$$

The computation for the derivative with respect to the population covariance is almost identical to the corresponding computation performed for the M-step. It yields the following implicit equation for the covariance estimate

$$\boxed{\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} \int_{\mathbb{R}^K} (\theta - \hat{\mu}) \otimes (\theta - \hat{\mu})\frac{P(y_i \mid \theta)}{\mathcal{N}(L_i)}\varphi(\theta; \hat{\mu}, \hat{\Sigma})\mathrm{d}^K\theta.} \tag{44}$$

8 Standard Error

Assume asymptotic consistency in that the estimates follow a normal distribution around the true value, with covariance being the inverse of the information matrix of the likelihood, so that

$$I(L) = -\begin{pmatrix} \frac{\partial^2 L}{\partial \Gamma^2} & \frac{\partial^2 L}{\partial \Gamma \partial \Sigma} \\ \\ \frac{\partial^2 L}{\partial \Sigma \partial \Gamma} & \frac{\partial^2 L}{\partial \Sigma^2} \end{pmatrix}, \qquad \operatorname{Cov}(\hat{\Gamma}, \hat{\Sigma}) = I(L)^{-1}. \tag{45}$$

Then compute the standard error. Recall that the first derivative of L with respect to Γ is given by

$$\frac{\partial L}{\partial \gamma_{kl}} = -\sum_{i=1}^{N} \int_{\mathbb{R}^K} \frac{P(y_i|\theta)}{\mathcal{N}(L_i)} x_{il} (\Sigma^{-1}(\Gamma x_i - \theta))_k \varphi(\theta; \Gamma x_i, \Sigma) \mathrm{d}^K \theta. \tag{46}$$

Then,

$$\begin{aligned} & \frac{\partial^2 L}{\partial \gamma_{k'l'} \partial \gamma_{kl}} \\ = & \sum_{i=1}^{N} \int_{\mathbb{R}^K} \frac{1}{\mathcal{N}(L_i)} \frac{\partial \mathcal{N}(L_i)}{\partial \gamma_{k'l'}} x_{il} (\Sigma^{-1}(\Gamma x_i - \theta))_k \mathrm{d}\mu_i \\ & - \sum_{i=1}^{N} \int_{\mathbb{R}^K} x_{il} \frac{\partial (\Sigma^{-1}(\Gamma x_i - \theta))_k}{\partial \gamma_{k'l'}} \mathrm{d}\mu_i \\ & - \sum_{i=1}^{N} \int_{\mathbb{R}^K} \frac{P(y_i|\theta)}{\mathcal{N}(L_i)} x_{il} (\Sigma^{-1}(\Gamma x_i - \theta))_k \frac{\partial \varphi(\theta; \Gamma x_i, \Sigma)}{\partial \gamma_{k'l'}} \mathrm{d}^K \theta. \end{aligned}$$

Performing the prescribed differentiations yields

$$\begin{aligned} \frac{\partial^2 L}{\partial \gamma_{k'l'} \partial \gamma_{kl}} = & \sum_{i=1}^{N} \sum_{\kappa,\kappa'=1}^{K} x_{il} x_{il'} \Sigma^{-1}_{k\kappa} \Sigma^{-1}_{k'\kappa'} (\Gamma x_i - \tilde{\theta}_i)_\kappa (\Gamma x_i - \tilde{\theta}_i)_{\kappa'} \\ & - \sum_{i=1}^{N} x_{il} x_{il'} \Sigma^{-1}_{kk'} \\ & - \sum_{i=1}^{N} \sum_{\kappa,\kappa'=1}^{K} x_{il} x_{il'} \Sigma^{-1}_{k\kappa} \Sigma^{-1}_{k'\kappa'} \mathcal{E} \left((\Gamma x_i - \theta)_\kappa (\Gamma x_i - \theta)_{\kappa'} \right). \end{aligned} \tag{47}$$

Finally,

$$\begin{aligned} \frac{\partial^2 L}{\partial \gamma_{k'l'} \partial \gamma_{kl}} = & -\sum_{i=1}^{N} (x_i \otimes x_i \otimes \Sigma^{-1})_{ll'kk'} \\ & - \sum_{i=1}^{N} \sum_{\kappa,\kappa'=1}^{K} (x_i \otimes x_i \otimes \check{\Sigma}^{(i)} \otimes \Sigma^{-1} \otimes \Sigma^{-1})_{ll'\kappa\kappa' k\kappa k'\kappa'}. \end{aligned} \tag{48}$$

The terms $\frac{\partial^2 L}{\partial \sigma_{k'k''}\partial \gamma_{kl}}$ are computed as follows:

$$\begin{aligned}
&\frac{\partial^2 L}{\partial \sigma_{k'k''}\partial \gamma_{kl}} \\
= \quad & \sum_{i=1}^{N}\int_{\mathbb{R}^K}\frac{1}{\mathcal{N}(L_i)}\frac{\partial \mathcal{N}(L_i)}{\partial \sigma_{k'k''}} x_{il}(\Sigma^{-1}(\Gamma x_i-\theta))_k \mathrm{d}\mu_i \qquad (49)\\
& -\sum_{i=1}^{N}\int_{\mathbb{R}^K} x_{il}\frac{\partial(\Sigma^{-1}(\Gamma x_i-\theta))_k}{\partial \sigma_{k'k''}}\mathrm{d}\mu_i \qquad (50)\\
& -\sum_{i=1}^{N}\int_{\mathbb{R}^K}\frac{P(y_i|\theta)}{\mathcal{N}(L_i)} x_{il}(\Sigma^{-1}(\Gamma x_i-\theta))_k\frac{\partial \varphi(\theta;\Gamma x_i,\Sigma)}{\partial \sigma_{k'k''}}\mathrm{d}^K\theta. \qquad (51)
\end{aligned}$$

First,

$$\frac{1}{\mathcal{N}(L_i)}\frac{\partial \mathcal{N}(L_i)}{\partial \sigma_{k'k''}} \quad = \quad \int_{\mathbb{R}^K}\frac{P(y_i \mid \theta)}{\mathcal{N}(L_i)}\frac{\partial \varphi(\theta;\Gamma x_i,\Sigma)}{\partial \sigma_{k'k''}}\mathrm{d}^K\theta, \qquad (52)$$

that is

$$(49) = \sum_{i=1}^{N} x_{il}(\Sigma^{-1}(\Gamma x_i-\tilde{\theta}_i))_k\int_{\mathbb{R}^K}\frac{P(y_i \mid \theta)}{\mathcal{N}(L_i)}\frac{\partial \varphi(\theta;\Gamma x_i,\Sigma)}{\partial \sigma_{k'k''}}\mathrm{d}^K\theta. \qquad (53)$$

Then, the contribution of (50) is computed:

$$(50) \quad = \quad -\frac{1}{2}\sum_{i=1}^{N} x_{il}(\delta_{k'k''}-2)\left((\Sigma^{-1}\otimes_s\Sigma^{-1})_{k'k''}(\Gamma x_i-\tilde{\theta}_i)\right)_k \qquad (54)$$

Furthermore,

$$\begin{aligned}
(51) \quad = \quad & -\sum_{i=1}^{N}\int_{\mathbb{R}^K}\frac{P(y_i|\theta)}{\mathcal{N}(L_i)} x_{il}(\Sigma^{-1}(\Gamma x_i-\tilde{\theta}_i))_k\frac{\partial \varphi(\theta;\Gamma x_i,\Sigma)}{\partial \sigma_{k'k''}}\mathrm{d}^K\theta \\
& -\sum_{i=1}^{N}\int_{\mathbb{R}^K}\frac{P(y_i|\theta)}{\mathcal{N}(L_i)} x_{il}(\Sigma^{-1}(\theta-\tilde{\theta}_i))_k\frac{\partial \varphi(\theta;\Gamma x_i,\Sigma)}{\partial \sigma_{k'k''}}\mathrm{d}^K\theta, \qquad (55)
\end{aligned}$$

where the first summand in (55) cancels out the contribution of (53).

From previous computations,

$$\begin{aligned}
&\frac{\partial \varphi(\theta;\Gamma x_i,\Sigma)}{\partial \sigma_{k'k''}} \\
= \quad & -\frac{1}{2}(2-\delta_{k'k''})\Sigma^{-1}_{k'k''}\varphi(\theta;\Gamma x_i,\Sigma) \\
& +\frac{1}{2}(2-\delta_{k'k''})\left\langle\!\left\langle \Gamma x_i-\tilde{\theta}_i \,\middle|\, (\Sigma^{-1}\otimes_s\Sigma^{-1})_{k'k''} \,\middle|\, \Gamma x_i-\tilde{\theta}_i \right\rangle\!\right\rangle \varphi(\theta;\Gamma x_i,\Sigma) \\
& +\frac{1}{2}(2-\delta_{k'k''})\left\langle\!\left\langle \theta-\tilde{\theta}_i \,\middle|\, (\Sigma^{-1}\otimes_s\Sigma^{-1})_{k'k''} \,\middle|\, \theta-\tilde{\theta}_i \right\rangle\!\right\rangle \varphi(\theta;\Gamma x_i,\Sigma) \\
& -(2-\delta_{k'k''})\left\langle\!\left\langle \theta-\tilde{\theta}_i \,\middle|\, (\Sigma^{-1}\otimes_s\Sigma^{-1})_{k'k''} \,\middle|\, \tilde{\theta}_i-\Gamma x_i \right\rangle\!\right\rangle \varphi(\theta;\Gamma x_i,\Sigma). \qquad (56)
\end{aligned}$$

Finally,

$$
\begin{aligned}
&\frac{\partial^2 L}{\partial \sigma_{k'k''}\partial \gamma_{kl}} \\
= \;& \frac{1}{2}(\delta_{k'k''}-2)\Bigg(-\sum_{i=1}^{N} x_{il}\left((\Sigma^{-1}\otimes_s \Sigma^{-1})_{k'k''}(\Gamma x_i - \tilde{\theta}_i)\right)_k \\
& -\sum_{i=1}^{N}\int_{\mathbb{R}^K} x_{il}(\Sigma^{-1}(\theta-\tilde{\theta}_i))_k \langle\!\langle(\theta-\tilde{\theta}_i)\otimes(\theta-\tilde{\theta}_i) \,\big|\, (\Sigma^{-1}\otimes_s \Sigma^{-1})_{k'k''}\rangle\!\rangle \mathrm{d}\mu_i \\
& +2\sum_{i=1}^{N}\int_{\mathbb{R}^K} x_{il}(\Sigma^{-1}(\theta-\tilde{\theta}_i))_k \langle\!\langle(\theta-\tilde{\theta}_i)\otimes(\Gamma x_i-\tilde{\theta}_i) \,\big|\, (\Sigma^{-1}\otimes_s \Sigma^{-1})_{k'k''}\rangle\!\rangle \mathrm{d}\mu_i \Bigg). \qquad (57)
\end{aligned}
$$

The terms $\frac{\partial^2 L}{\partial\sigma_{\kappa\kappa'}\partial\sigma_{kk'}}$ are computed using (19) in several steps as follows. First, recall the derivatives for Σ^{-1} are

$$
\frac{\partial \Sigma^{-1}_{kk'}}{\partial \sigma_{\kappa\kappa'}} = -(2-\delta_{\kappa\kappa'})(\Sigma^{-1}\otimes_s \Sigma^{-1})_{\kappa\kappa' kk'}. \qquad (58)
$$

For $(\Sigma^{-1}\otimes_s \Sigma^{-1})$, they are

$$
\frac{\partial (\Sigma^{-1}\otimes_s \Sigma^{-1})_{kk'rr'}}{\partial \sigma_{\kappa\kappa'}} = -(2-\delta_{\kappa\kappa'})(\Sigma^{-1}\otimes_s \Sigma^{-1}\otimes_s \Sigma^{-1})_{\kappa\kappa' kk'rr'}. \qquad (59)
$$

Also, for the normalizations $\mathcal{N}(L_i)$

$$
\begin{aligned}
&\frac{1}{\mathcal{N}(L_i)}\frac{\partial \mathcal{N}(L_i)}{\partial \sigma_{\kappa\kappa'}} \\
= \;& -\frac{1}{2}(2-\delta_{\kappa\kappa'})\Sigma^{-1}_{\kappa\kappa'} - \frac{1}{2}\int_{\mathbb{R}^K}\frac{\partial\langle\!\langle \theta-\Gamma x_i \,\big|\, \Sigma^{-1} \,\big|\, \theta-\Gamma x_i\rangle\!\rangle}{\partial\sigma_{\kappa\kappa'}}\mathrm{d}\mu_i \\
= \;& -\frac{1}{2}(2-\delta_{\kappa\kappa'})\Sigma^{-1}_{\kappa\kappa'} + \frac{1}{2}(2-\delta_{\kappa\kappa'})\langle\!\langle \Sigma^{(i)} \;\big|\; (\Sigma^{-1}\otimes_s \Sigma^{-1})_{\kappa\kappa'}\rangle\!\rangle. \qquad (60)
\end{aligned}
$$

and $\tilde{\theta}_i$

$$
\begin{aligned}
\frac{\partial \tilde{\theta}_i}{\partial \sigma_{\kappa\kappa'}} = \;& \int_{\mathbb{R}^K}\theta\frac{\partial\varphi(\theta;\Gamma x_i,\Sigma)}{\partial\sigma_{\kappa\kappa'}}\frac{1}{\varphi(\theta;\Gamma x_i,\Sigma)}\mathrm{d}\mu_i - \int_{\mathbb{R}^K}\theta\frac{1}{\mathcal{N}(L_i)}\frac{\partial\mathcal{N}(L_i)}{\partial\sigma_{\kappa\kappa'}}\mathrm{d}\mu_i \\
= \;& \int_{\mathbb{R}^K}(\theta-\tilde{\theta}_i)\frac{\partial\varphi(\theta;\Gamma x_i,\Sigma)}{\partial\sigma_{\kappa\kappa'}}\frac{1}{\varphi(\theta;\Gamma x_i,\Sigma)}\mathrm{d}\mu_i \\
= \;& -\frac{1}{2}\int_{\mathbb{R}^K}(\theta-\tilde{\theta}_i)\frac{\partial\langle\!\langle \theta-\Gamma x_i \,\big|\, \Sigma^{-1} \,\big|\, \theta-\Gamma x_i\rangle\!\rangle}{\partial\sigma_{\kappa\kappa'}}\mathrm{d}\mu_i \\
= \;& \frac{1}{2}(2-\delta_{\kappa\kappa'})\Bigg[\langle\!\langle \mathcal{E}\left((\theta-\tilde{\theta}_i)^{\otimes 3}\right) \;\big|\, (\Sigma^{-1}\otimes_s \Sigma^{-1})_{\kappa\kappa'}\rangle\!\rangle_{23} \\
& +2\langle\!\langle \check{\Sigma}^{(i)}\otimes(\tilde{\theta}_i-\Gamma x_i) \;\big|\; (\Sigma^{-1}\otimes_s \Sigma^{-1})_{\kappa\kappa'}\rangle\!\rangle_{23}\Bigg]. \qquad (61)
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\frac{\partial \mathcal{E}((\theta-\tilde{\theta}_i)\otimes(\theta-\tilde{\theta}_i))}{\partial \sigma_{\kappa\kappa'}}\\
=\;& 2\int_{\mathbb{R}^K}\left(\theta-\frac{\partial\tilde{\theta}_i}{\partial\sigma_{\kappa\kappa'}}\right)\otimes_s(\theta-\tilde{\theta}_i)\mathrm{d}\mu_i\\
&-\int_{\mathbb{R}^K}(\theta-\tilde{\theta}_i)\otimes(\theta-\tilde{\theta}_i)\frac{1}{\mathcal{N}(L_i)}\frac{\partial\mathcal{N}(L_i)}{\partial\sigma_{\kappa\kappa'}}\mathrm{d}\mu_i\\
&+\int_{\mathbb{R}^K}(\theta-\tilde{\theta}_i)\otimes(\theta-\tilde{\theta}_i)\frac{\partial\varphi(\theta;\Gamma x_i,\Sigma)}{\partial\sigma_{\kappa\kappa'}}\frac{1}{\varphi(\theta;\Gamma x_i,\Sigma)}\mathrm{d}\mu_i.
\end{aligned}
$$

This, using

$$
\left(\theta-\frac{\partial\tilde{\theta}_i}{\partial\sigma_{\kappa\kappa'}}\right)\otimes(\theta-\tilde{\theta}_i)=(\theta-\tilde{\theta}_i)^{\otimes 2}+\left(\tilde{\theta}_i-\frac{\partial\tilde{\theta}_i}{\partial\sigma_{\kappa\kappa'}}\right)\otimes(\theta-\tilde{\theta}_i),
$$

yields

$$
\begin{aligned}
&\frac{\partial \mathcal{E}((\theta-\tilde{\theta}_i)\otimes(\theta-\tilde{\theta}_i))}{\partial \sigma_{\kappa\kappa'}}\\
=\;& \check{\Sigma}^{(i)}+\int_{\mathbb{R}^K}\left((\theta-\tilde{\theta}_i)^{\otimes 2}-\check{\Sigma}^{(i)}\right)\frac{\partial\varphi(\theta;\Gamma x_i,\Sigma)}{\partial\sigma_{\kappa\kappa'}}\frac{1}{\varphi(\theta;\Gamma x_i,\Sigma)}\mathrm{d}\mu_i\\
=\;& \check{\Sigma}^{(i)}-\frac{1}{2}\int_{\mathbb{R}^K}\left((\theta-\tilde{\theta}_i)^{\otimes 2}-\check{\Sigma}^{(i)}\right)\frac{\partial\langle\!\langle\theta-\Gamma x_i\mid\Sigma^{-1}\mid\theta-\Gamma x_i\rangle\!\rangle}{\partial\sigma_{\kappa\kappa'}}\mathrm{d}\mu_i\\
=\;& \check{\Sigma}^{(i)}+\frac{1}{2}(2-\delta_{\kappa\kappa'})\Big[\langle\!\langle\mathcal{E}\left((\theta-\tilde{\theta})_i^{\otimes 4}\right)-\check{\Sigma}^{(i)\otimes 2}\mid(\Sigma^{-1}\otimes_s\Sigma^{-1})_{\kappa\kappa'}\rangle\!\rangle_{34}\\
&+\langle\!\langle\mathcal{E}\left((\theta-\tilde{\theta})_i^{\otimes 3}\right)\otimes(\tilde{\theta}_i-\Gamma x_i)\mid(\Sigma^{-1}\otimes_s\Sigma^{-1})_{\kappa\kappa'}\rangle\!\rangle_{34}\\
&+\langle\!\langle\mathcal{E}\left((\theta-\tilde{\theta})_i^{\otimes 2}\otimes(\tilde{\theta}_i-\Gamma x_i)\otimes(\theta-\tilde{\theta})_i\right)\mid(\Sigma^{-1}\otimes_s\Sigma^{-1})_{\kappa\kappa'}\rangle\!\rangle_{34}\Big].
\end{aligned}
\tag{62}
$$

Finally, putting all the above together yields

$$
\begin{aligned}
&\frac{\partial^2 L}{\partial\sigma_{\kappa\kappa'}\partial\sigma_{kk'}}\\
=\;& -\frac{N}{2}(2-\delta_{kk'})(2-\delta_{\kappa\kappa'})(\Sigma^{-1}\otimes_s\Sigma^{-1})_{kk'\kappa\kappa'}\\
&+\frac{1}{2}(2-\delta_{kk'})\sum_{i=1}^{N}\langle\!\langle\frac{\partial \mathcal{E}((\theta-\tilde{\theta}_i)\otimes(\theta-\tilde{\theta}_i))}{\partial \sigma_{\kappa\kappa'}}\mid(\Sigma^{-1}\otimes_s\Sigma^{-1})_{kk'}\rangle\!\rangle\\
&-\frac{1}{2}(2-\delta_{kk'})(2-\delta_{\kappa\kappa'})\sum_{i=1}^{N}\langle\!\langle\check{\Sigma}^{(i)}\mid(\Sigma^{-1}\otimes_s\Sigma^{-1}\otimes_s\Sigma^{-1})_{kk'\kappa\kappa'}\rangle\!\rangle\\
&+(2-\delta_{kk'})\sum_{i=1}^{N}\langle\!\langle\left(\frac{\partial\tilde{\theta}_i}{\partial\sigma_{\kappa\kappa'}}-\Gamma x_i\right)\otimes_s(\tilde{\theta}_i-\Gamma x_i)\mid(\Sigma^{-1}\otimes_s\Sigma^{-1})_{kk'}\rangle\!\rangle\\
&-\frac{1}{2}(2-\delta_{kk'})(2-\delta_{\kappa\kappa'})\sum_{i=1}^{N}\langle\!\langle(\tilde{\theta}_i-\Gamma x_i)^{\otimes 2}\mid(\Sigma^{-1}\otimes_s\Sigma^{-1}\otimes_s\Sigma^{-1})_{kk'\kappa\kappa'}\rangle\!\rangle.
\end{aligned}
\tag{63}
$$

9 Newton Method

9.1 Basic Computations

This section derives the formulae for the estimation based on Newton's method. First define

$$\begin{aligned} F : M_{K,L}(\mathbb{R}) \times M_K(\mathbb{R}) &\to M_{K,L}(\mathbb{R}) \times M_K(\mathbb{R}), \quad F = (F_\Gamma, F_\Sigma), \\ F_\Gamma : M_{K,L}(\mathbb{R}) \times M_K(\mathbb{R}) &\to M_{K,L}(\mathbb{R}), \\ F_\Sigma : M_{K,L}(\mathbb{R}) \times M_K(\mathbb{R}) &\to M_K(\mathbb{R}), \end{aligned}$$

where

$$F_\Gamma(\Gamma, \Sigma) = \Gamma^t - \left(\sum_{i=1}^N x_i \otimes x_i \right)^{-1} \left(\sum_{i=1}^N x_i \otimes \tilde{\theta}_i \right), \tag{64}$$

$$F_\Sigma(\Gamma, \Sigma) = \Sigma - \frac{1}{N} \sum_{i=1}^N \left[(\tilde{\theta}_i - \Gamma x_i) \otimes (\tilde{\theta}_i - \Gamma x_i) + \check{\Sigma}^{(i)} \right]. \tag{65}$$

By construction, the equation

$$F(\Gamma, \Sigma) = 0, \qquad ((\Gamma, \Sigma) =?) \tag{66}$$

is equivalent to the system of equations presented in (14) and in (21). The Newton method for (66) yields the following scheme for the iterative estimates $(\Gamma^{(n)}, \Sigma^{(n)})$:

$$(\Gamma^{(n+1)}, \Sigma^{(n+1)}) = (\Gamma^{(n)}, \Sigma^{(n)}) - \mathrm{D}F[\Gamma^{(n)}, \Sigma^{(n)}]^{-1} F(\Gamma^{(n)}, \Sigma^{(n)}). \tag{67}$$

The computation of the derivative

$$\mathrm{D}F = \begin{pmatrix} \frac{\partial F_\Gamma}{\partial \Gamma} & \frac{\partial F_\Gamma}{\partial \Sigma} \\ \\ \frac{\partial F_\Sigma}{\partial \Gamma} & \frac{\partial F_\Sigma}{\partial \Sigma} \end{pmatrix} \tag{68}$$

is as follows. First,

$$\begin{aligned} \frac{\partial F_\Gamma}{\partial \gamma_{kl}} &= \frac{\partial \Gamma^t}{\partial \gamma_{kl}} - \left(\sum_{i=1}^N x_i \otimes x_i \right)^{-1} \left(\sum_{i=1}^N x_i \otimes \frac{\partial \tilde{\theta}_i}{\partial \gamma_{kl}} \right) \\ &= \frac{\partial \Gamma^t}{\partial \gamma_{kl}} - \left(\sum_{i=1}^N x_i \otimes x_i \right)^{-1} \sum_{i=1}^N \sum_{\kappa=1}^K x_i \otimes (x_i \otimes \Sigma^{-1} \otimes \check{\Sigma}^{(i)})_{lk\kappa\kappa}. \end{aligned}$$

Then,

$$\frac{\partial F_\Gamma}{\partial \sigma_{kk'}} = - \left(\sum_{i=1}^N x_i \otimes x_i \right)^{-1} \left(\sum_{i=1}^N x_i \otimes \frac{\partial \tilde{\theta}_i}{\partial \sigma_{kk'}} \right). \tag{69}$$

The term $\frac{\partial \tilde{\theta}_i}{\partial \sigma_{kk'}}$ is computed in (61). Next,

$$\begin{aligned}\frac{\partial F_\Sigma}{\partial \gamma_{kl}} &= -\frac{2}{N}\sum_{i=1}^{N}\left[\left(\frac{\partial \tilde{\theta}_i}{\partial \gamma_{kl}} - \frac{\partial \Gamma}{\partial \gamma_{kl}} x_i\right) \otimes_s (\tilde{\theta}_i - \Gamma x_i) + \mathcal{E}\left(\left(\theta - \frac{\partial \tilde{\theta}_i}{\partial \gamma_{kl}}\right) \otimes_s (\theta - \tilde{\theta}_i)\right)\right] \\ &= -\frac{2}{N}\sum_{i=1}^{N}\left[\left(\frac{\partial \tilde{\theta}_i}{\partial \gamma_{kl}} - \frac{\partial \Gamma}{\partial \gamma_{kl}} x_i\right) \otimes_s (\tilde{\theta}_i - \Gamma x_i) + \frac{1}{2}\check{\Sigma}^{(i)}\right]. \qquad (70)\end{aligned}$$

Finally,

$$\frac{\partial F_\Sigma}{\partial \sigma_{kk'}} = \frac{\partial \Sigma}{\partial \sigma_{kk'}} - \frac{2}{N}\sum_{i=1}^{N}\left[\left(\frac{\partial \tilde{\theta}_i}{\partial \sigma_{kk'}} - \Gamma x_i\right) \otimes_s (\tilde{\theta}_i - \Gamma x_i) + \frac{1}{2}\check{\Sigma}^{(i)}\right]. \qquad (71)$$

9.2 Notes on Implementation

From the actual forms of the second derivatives—most notably (63)—it is clear that direct application should be done very carefully. The number of function evaluations in any numerical integration scheme with N_{qp} quadrature points in N_{dim} dimensions (N_{dim} = number of subscales) for the expectation

$$\mathcal{E}\left((\theta - \tilde{\theta})_i^{\otimes 4}\right)$$

is

$$N_{dim}^4 N_{qp}^{N_{dim}},$$

which is $5^4 \cdot 12^5 = 155,520,000$ for a reasonable choice of five dimensions and 12 quadrature points. Now, there is a relatively high degree of symmetry in the computation that can be used to decrease this number significantly. The point is that a careful analysis should be performed to find the best way to use this symmetry and to determine if it brings down the number of function evaluations to a manageable range.

With the Newton method, the computation might not worth the effort, since there already is a well-functioning (even if slow) estimation method (the EM-algorithm). For the standard error (where the exact same computations are needed to be performed), the effort is justified only if the added correction may significantly alter the existing approximate estimator.

In both cases, the discussions are postponed for a future research study.

10 Conclusion

This rather technical paper presented the fundamental computations underlying the marginal maximum likelihood estimation of the latent regression model as used in NAEP. Special emphasis was given to the numerical integration method, which, while an artifact of the calculation, can play a decisive role in the manageability of the computations required for the parameter estimation.

While the computations of the formulae for the asymptotic standard error and that of the second derivatives of the likelihood appearing in the Newton method pose no theoretical problems, practical implementations may encounter serious difficulties. This is mainly due to the appearance of the expectations of the third and fourth order tensors $(\theta - \tilde{\theta})^{\otimes 3}$ and $(\theta - \tilde{\theta})^{\otimes 4}$, since the integration should be performed for each matrix element separately. While the existing symmetry may be used to reduce the computational burden, a practically useful way is yet to be found.

Another direction for future study could be the comparison of the theoretical standard error formulae to both simulation study empirical standard error (mainly to check appropriateness of asymptotic approximation) and to the existing NAEP standard error (to ensure consistency of practice with theory).

References

Antal, T., & Oranje, A. (2007). *Adaptive numerical integration for item response theory models* (ETS Research Rep. No. RR-06-07). Princeton, NJ: ETS.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: MIT Press.

Halmos, P. R. (1974). *Finite-dimensional vector spaces* (2nd ed.). New York: Springer-Verlag.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective.* New York: Springer-Verlag.

Magnus, J. R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics.* New York: Wiley.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, *49*(3), 359–381.

Mislevy, R. J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association*, *80*(392), 993–997.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*, 159–176.

Stoer, J., & Bulirsch, R. (2002). *Introduction to numerical analysis.* New York: Springer.

von Davier, M., Sinharay, S., Oranje, A., & Beaton, A. (2007). The statistical procedures used in National Assessment of Educational Progress: Recent developments and future directions. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26. psychometrics* (pp. 1039–1056). New York: Elsevier.

Notes

[1] Note, that, in general, the more items, the sharper is the peak.