Formulas and methods for Item Response Theory Models

Sergio Páez Moncaleano (SICS Research Group)

12 de abril de 2016

This document is dedicated to explain some methods for understand the process around Item Response Theory *IRT* models. First a theoretical review around EM algorithm is made, then a step by step for the EM are presented. Following a review of MIRT models at his whole theory and finally, the same for Multidimensional Polytomous models. Some specifications are made at each chapter around different kind of quadratures for EM chapter, general form of Unidimensional IRT models for MIRT chapter and Partial Credit and Generalized Partial Credit Models for Polytomous chapter.

1. Expectation Maximization (EM) Algorithm

The Expectation Maximization (EM) Algorithm is an iterative method to found Maximum Likelihood or Maximum a Posteriori estimates for parameters of statistical models which depends on latent variables. Broad strokes, the idea is that on a first step a function is created and in a second one a maximum for it is found. More specifically **E-step** creates a function for the Expectation (Expected Value) of the *log-likelihood* evaluated using some known or estimated parameters, and **M-step** recover parameters maximizing the *log-likelihood* previously formed.

Given the item parameters, the Fisher's identity to the marginal probability of the pattern $\boldsymbol{u}_l$ establishes that

$$\frac{\partial \log P^*(\boldsymbol{u}_l|\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \int \frac{\partial \log P^*(\boldsymbol{u}_l, \theta_l|\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} P^*(\theta_l|\boldsymbol{u}_l, \boldsymbol{\xi}) d\theta_l. \tag{1}$$

Where P^* is the probability for a given model.

Let us suppose that $\widetilde{\boldsymbol{\xi}}$ represents a provisional value of $\boldsymbol{\xi}$. Let $\widetilde{P}_{ik}(\theta)$ be $P^*_{ik}(\theta)$ evaluate at $\boldsymbol{\xi} = \widetilde{\boldsymbol{\xi}}$. Following the logic of the Fisher's identity the conditional expected complete data likelihood given provisional item parameters $\widetilde{\boldsymbol{\xi}}$ can be approximated by

$$\begin{aligned} Q(\boldsymbol{\xi}|\boldsymbol{u}, \widetilde{\boldsymbol{\xi}}) &= \sum_{l=1}^{s}\sum_{g=1}^{G}\sum_{i=1}^{p}\sum_{k=1}^{m_i} n_l x_{lik} \log P^*_{ik}(\theta_g) P^*(\theta_g|\boldsymbol{u}_l, \widetilde{\boldsymbol{\xi}}) && (2)\\ &= \sum_{g=1}^{G}\sum_{i=1}^{p}\sum_{k=1}^{m_i} \log P^*_{ik}(\theta_g) \left(\sum_{l=1}^{s} n_l x_{lik} P^*(\theta_g|\boldsymbol{u}_l, \widetilde{\boldsymbol{\xi}})\right) && (3)\\ &= \sum_{g=1}^{G}\sum_{i=1}^{p}\sum_{k=1}^{m_i} \widetilde{r}_{gik} \log P^*_{ik}(\theta_g), && (4) \end{aligned}$$

where

$$\widetilde{r}_{gik} = \sum_{l=1}^{s} n_l x_{lik} P^*(\theta_g|\boldsymbol{u}_l, \widetilde{\boldsymbol{\xi}}) = \sum_{l=1}^{s} n_l x_{lik} \widetilde{\pi}_{gl}, \tag{5}$$

and

$$\widetilde{\pi}_{gl} = \frac{\prod_{i=1}^{p} \prod_{k=1}^{m_i} [\widetilde{P}^*_{ik}(\theta_g)]^{x_{lik}} w_g}{\sum_{h=1}^{G} \prod_{i=1}^{p} \prod_{k=1}^{m_i} [\widetilde{P}^*_{ik}(\theta_h)]^{x_{lik}} w_h}. \tag{6}$$

2. Steps of the EM Algorithm

Obtain reasonable intitial values $\widetilde{\boldsymbol{\xi}}$ for the complete vector of item parameters.

2.1. E-step

Given provisional values $\widetilde{\boldsymbol{\xi}}$ of the complete vector of item parameters obtain the values $\widetilde{r}_{gik}$ using equations (5 and 6).

Let $\widetilde{\boldsymbol{\xi}}_i = (\widetilde{\boldsymbol{\alpha}}_i^t, \widetilde{\gamma}_{i1}, \ldots, \widetilde{\gamma}_{im_{i-1}})^t$

$$\widetilde{\pi}_{gl} = \frac{\prod_{i=1}^{p} \prod_{k=1}^{m_i} [\widetilde{P}^*_{ik}(\theta_g)]^{x_{lik}} w_g}{\sum_{h=1}^{G} \prod_{i=1}^{p} \prod_{k=1}^{m_i} [\widetilde{P}^*_{ik}(\theta_h)]^{x_{lik}} w_h}, \tag{7}$$

$g = 1, \ldots, G; l = 1, \ldots, s;$

$$\widetilde{r}_{gik} = \sum_{l=1}^{s} n_l x_{lik} \widetilde{\pi}_{gl}, \tag{8}$$

$g = 1, \ldots, G; i = 1, \ldots, p; k = 1, \ldots, m_i.$

2.2. M-step

Obtain new values $\widetilde{\boldsymbol{\xi}}_i$ to the parameters of item $i, i = 1, \ldots .p$ by maximizing

$$\widetilde{Q}_i(\boldsymbol{\xi}_i|\boldsymbol{u}, \widetilde{\boldsymbol{\xi}}) = \sum_{g=1}^{G} \sum_{k=1}^{m_i} \widetilde{r}_{gik} \log P^*_{ik}(\theta_g). \tag{9}$$

for $i = 1, \ldots, p,$

2.3. Clarifications

Note that:

1. For practical application step *1* and *2* are repeated until convergence.
2. This could be used as a general algorithm for any P^*, that is, for a specific problem an appropriated probability is selected and then used for the **EM** Algorithm.
3. θ_g is a vector of known parameters.

2.4. Quadratures

As it was formulated in previous subsections a set of known points θ_g with $g = 1, \ldots, G$ is used to estimate an integral.

Gaussian Hermite Quadrature

For integrals for the following kind:

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)dx \approx \sum_{g=1}^{G} w_i f(x_i) \tag{10}$$

with

$$w_i = \frac{2^{G-1}G!\sqrt{\pi}}{G^2\left[H_{G-1}(x_i)\right]}$$

With G the number of points to estimate the integral, and x_i the roots of the *Hermite Polynomial*$H_G x$. Form more dimensions, more integrals are included, then, more sums.

Quasi-Montecarlo (Low-discrepancy) Quadratures

The problem here is to approximate the integral of a function f that has domain in $[0, 1]$ as the average of the function evaluated at a set of points $theta_g$, $g = 1, \ldots, G$. As the

$$\int_{[0,1]} f(x)dx \approx \frac{1}{G}\sum_{i=1}^{G} f(x_i) \tag{11}$$

For more dimensions the integral is made at $[0, 1]^d$ (d-dimensional cube), the selected points are sequences of the kinds like *sobol*, *halton*, etc.

3. Multidimensional Item Response Theory (MIRT)

3.1. Introduction

In the dichotomous multidimensional item response theory (MIRT) models the data is organized in a $N \times p$ matrix $\boldsymbol{y}$ as

$$\boldsymbol{y} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \vdots & \vdots & y_{ij} & \vdots \\ 0 & 1 & \vdots & 1 \end{pmatrix}. \tag{12}$$

The ijth element represents the response of the person i to the item j. The value y_{ij} is the realization of a binary random variable $Y_{ij} \sim Ber(\pi_{ij})$. The value 1 is assigned to a correct response and 0 otherwise. The latent trait of person i is denoted θ_i and it is assumed to be an independent realization of a random d-vector θ, where θ has some multivariate distribution as $N_d(\mathbf{0}, \mathbf{\Sigma})$. The probability π_{ij} is given by

$$\pi_{ij} = c_j + \frac{1 - c_j}{1 + e^{-\eta_{ij}}}, \tag{13}$$

where

$$\eta_{ij} = \boldsymbol{a}_j^t \theta_i + d_j = a_{1j}\theta_{i1} + \cdots + a_{dj}\theta_{id} + d_j. \tag{14}$$

3.2. Log Likelihood

To build the likelihood it is necessary to state the following assumptions.

1. The response patterns of the people are independent.
2. Given the latent trait θ_i, the responses $Y_{ij}, j = 1, \ldots, p$ are independent (conditional independence).

Since the probability function of each variable Y_{ij} is given by $f(y_{ij}|\pi_{ij}) = \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{1-y_{ij}}$, the log likelihood is given by

$$\begin{aligned} L(\boldsymbol{y}|\boldsymbol{\xi}, \theta) &= \sum_{i=1}^{N}\sum_{j=1}^{p} y_{ij}\log\pi_{ij} + (1 - y_{ij})\log(1 - \pi_{ij}) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{p} y_{ij}[\log\pi_{ij} - \log(1 - \pi_{ij})] + \log(1 - \pi_{ij}). \end{aligned} \tag{15}$$

The symbol $\boldsymbol{y}$ is the $N \times p$ matrix of binary responses. The set of item parameters of the test is denoted by $\boldsymbol{\xi}$. Thus, $\boldsymbol{\xi}$ is a $(d+2) \times p$ matrix. The set of latent traits of the sample is represented by θ. Then θ is a $N \times d$ matrix. The following notation is adopted. Let $\boldsymbol{\beta}_j = (a_{1j}, \ldots, a_{dj}, d_j)^t$. $\boldsymbol{\beta}_j$ will be called the latent regression parameter of the item j, $j = 1, \ldots, p$. Let $\boldsymbol{a}_j = (a_{1j}, \ldots, a_{dj})^t$. Then $\boldsymbol{\beta}_j = (\boldsymbol{a}_j^t, d_j)^t$. Let $c = (c_1, \ldots, c_p)^t$ be the vector of guessing parameters of the test. Then, $\boldsymbol{\xi}_j = (\boldsymbol{\beta}_j^t, c_j)^t, j = 1, \ldots, p$

If α_j is the Euclidean norm of the vector $\boldsymbol{a}_j$, that is if

$$\alpha_j = \sqrt{a_{1j}^2 + \ldots + a_{dj}^2}, \tag{16}$$

then α_j is called the multidimensional item discrimination (MDISC) of item j. The vector $\boldsymbol{\nu}_j = \boldsymbol{a}_j/\alpha_j$ is the direction of the item j and $b_j = -d_j/\alpha_j$ is called the multidimensional item difficulty (MDIFF) of item j. See for example Reckase **??**, **??**, **??**.

3.3. Derivatives and information matrix of the Log Likelihood

Let l_{ij} be the ijth term in the log likelihood. Thus,

$$l_{ij} = y_{ij}[\log \pi_{ij} - \log(1-\pi_{ij})] + \log(1-\pi_{ij}). \tag{17}$$

Gradient equations

The derivatives of $L(\boldsymbol{y}|\boldsymbol{\xi}, \theta)$ and the terms of the information matrix are computed from sums and products of the derivatives of the l_{ij} terms. It is clear that

$$\frac{\partial l_{ij}}{\partial \boldsymbol{\beta}_j} = \frac{\partial l_{ij}}{\partial \pi_{ij}} \frac{\partial \pi_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}_j}, \tag{18}$$

$$\frac{\partial l_{ij}}{\partial c_j} = \frac{\partial l_{ij}}{\partial \pi_{ij}} \frac{\partial \pi_{ij}}{\partial c_j}, \tag{19}$$

and,

$$\frac{\partial l_{ij}}{\partial \theta_i} = \frac{\partial l_{ij}}{\partial \pi_{ij}} \frac{\partial \pi_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial \theta_i}. \tag{20}$$

On the other hand we have that,

$$\frac{\partial l_{ij}}{\partial \pi_{ij}} = \frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1-\pi_{ij})} \tag{21}$$

$$\frac{\partial \pi_{ij}}{\partial \eta_{ij}} = \frac{\pi_{ij} - c_j}{1 + e^{\eta_{ij}}}, \tag{22}$$

and,

$$\frac{\partial \pi_{ij}}{\partial \boldsymbol{c}_j} = \frac{1}{1 + e^{\eta_{ij}}} \tag{23}$$

$$\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}_j} = \begin{bmatrix} \theta_i \\ 1 \end{bmatrix} \tag{24}$$

$$\frac{\partial \eta_{ij}}{\partial \theta_i} = \boldsymbol{a}_j. \tag{25}$$

Therefore,

$$\frac{\partial l_{ij}}{\partial \boldsymbol{\beta}_j} = \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1-\pi_{ij})} \times \frac{\pi_{ij} - c_j}{1 + e^{\eta_{ij}}} \right] \begin{bmatrix} \theta_i \\ 1 \end{bmatrix}, \tag{26}$$

$$\frac{\partial l_{ij}}{\partial c_j} = \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1-\pi_{ij})} \right] \left[\frac{1}{1 + e^{\eta_{ij}}} \right] \tag{27}$$

and,

$$\frac{\partial l_{ij}}{\partial \theta_i} = \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})} \times \frac{\pi_{ij} - c_j}{1 + e^{\eta_{ij}}}\right] \boldsymbol{a}_j. \tag{28}$$

Gradient equations

The gradient equations are given by

$$\begin{bmatrix} \frac{\partial L}{\partial \boldsymbol{\beta}_j} \\ \frac{\partial L}{\partial c_j} \end{bmatrix} = \sum_{i=1}^{N} \begin{bmatrix} \frac{\partial l_{ij}}{\partial \boldsymbol{\beta}_j} \\ \frac{\partial l_{ij}}{\partial c_j} \end{bmatrix}, \quad j = 1, \ldots, p. \tag{29}$$

$$\begin{bmatrix} \frac{\partial L}{\partial \theta_i} \end{bmatrix} = \sum_{j=1}^{p} \begin{bmatrix} \frac{\partial l_{ij}}{\partial \theta_i} \end{bmatrix}, \quad i = 1, \ldots, N. \tag{30}$$

Information equations

To derive the information matrices it is necessary to remember that

$$E\left[y_{ij} - \pi_{ij}\right]^2 = Var[y_{ij}] = \pi_{ij}(1 - \pi_{ij}). \tag{31}$$

Using this fact, the independence between pattern responses, and the conditional independence of the responses given a specific latent trait, it is straightforward the derivation of the following results. Only the nonzero expectations are shown.

$$E\left[\frac{\partial l_{ij}}{\partial \boldsymbol{\beta}_j} \frac{\partial l_{ij}}{\partial \boldsymbol{\beta}_j^t}\right] = \frac{1}{\pi_{ij}(1 - \pi_{ij})} \times \left[\frac{\pi_{ij} - c_j}{1 + e^{\eta_{ij}}}\right]^2 \begin{bmatrix} \theta_i \\ 1 \end{bmatrix} \begin{bmatrix} \theta_i \\ 1 \end{bmatrix}^t \tag{32}$$

$$E\left[\frac{\partial l_{ij}}{\partial \boldsymbol{\beta}_j} \frac{\partial l_{ij}}{\partial c_j}\right] = \frac{1}{\pi_{ij}(1 - \pi_{ij})} \times \frac{\pi_{ij} - c_j}{\left[1 + e^{\eta_{ij}}\right]^2} \begin{bmatrix} \theta_i \\ 1 \end{bmatrix} \tag{33}$$

$$E\left[\frac{\partial l_{ij}}{\partial c_j} \frac{\partial l_{ij}}{\partial c_j}\right] = \frac{1}{\pi_{ij}(1 - \pi_{ij})} \times \frac{1}{\left[1 + e^{\eta_{ij}}\right]^2} \tag{34}$$

$$E\left[\frac{\partial l_{ij}}{\partial \theta_i} \frac{\partial l_{ij}}{\partial \theta_i^t}\right] = \frac{1}{\pi_{ij}(1 - \pi_{ij})} \times \left[\frac{\pi_{ij} - c_j}{1 + e^{\eta_{ij}}}\right]^2 \boldsymbol{a}_j \boldsymbol{a}_j^t \tag{35}$$

$$E\left[\frac{\partial l_{ij}}{\partial \boldsymbol{\beta}_j} \frac{\partial l_{ij}}{\partial \theta_i^t}\right] = \frac{1}{\pi_{ij}(1 - \pi_{ij})} \times \left[\frac{\pi_{ij} - c_j}{1 + e^{\eta_{ij}}}\right]^2 \begin{bmatrix} \theta_i \\ 1 \end{bmatrix} \boldsymbol{a}_j^t \tag{36}$$

$$E\left[\frac{\partial l_{ij}}{\partial \theta_i} \frac{\partial l_{ij}}{\partial c_j}\right] = \frac{1}{\pi_{ij}(1 - \pi_{ij})} \times \frac{\pi_{ij} - c_j}{\left[1 + e^{\eta_{ij}}\right]^2} \boldsymbol{a}_j$$

Finally, the terms of the information matrices are given by

$$E\left[\frac{\partial L}{\partial \boldsymbol{\beta}_j}\frac{\partial L}{\partial \boldsymbol{\beta}_j^t}\right] = \sum_{i=1}^{N}\frac{1}{\pi_{ij}(1-\pi_{ij})}\times\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\begin{bmatrix}\theta_i\\1\end{bmatrix}\begin{bmatrix}\theta_i\\1\end{bmatrix}^t, \tag{37}$$

$$E\left[\frac{\partial L}{\partial \boldsymbol{\beta}_j}\frac{\partial L}{\partial c_j}\right] = \sum_{i=1}^{N}\frac{1}{\pi_{ij}(1-\pi_{ij})}\times\frac{\pi_{ij}-c_j}{[1+e^{\eta_{ij}}]^2}\begin{bmatrix}\theta_i\\1\end{bmatrix} \tag{38}$$

$$E\left[\frac{\partial L}{\partial c_j}\frac{\partial L}{\partial c_j}\right] = \sum_{i=1}^{N}\frac{1}{\pi_{ij}(1-\pi_{ij})}\times\frac{1}{[1+e^{\eta_{ij}}]^2} \tag{39}$$

$$E\left[\frac{\partial L}{\partial \theta_i}\frac{\partial L}{\partial \theta_i^t}\right] = \sum_{j=1}^{p}\frac{1}{\pi_{ij}(1-\pi_{ij})}\times\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\boldsymbol{a}_j\boldsymbol{a}_j^t \tag{40}$$

$$E\left[\frac{\partial L}{\partial \boldsymbol{\beta}_j}\frac{\partial L}{\partial \theta_i^t}\right] = \frac{1}{\pi_{ij}(1-\pi_{ij})}\times\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\begin{bmatrix}\theta_i\\1\end{bmatrix}\boldsymbol{a}_j^t \tag{41}$$

$$E\left[\frac{\partial L}{\partial \theta_i}\frac{\partial L}{\partial c_j}\right] = \frac{1}{\pi_{ij}(1-\pi_{ij})}\times\frac{\pi_{ij}-c_j}{[1+e^{\eta_{ij}}]^2}\boldsymbol{a}_j$$

3.4. Reparameterization of the guessing parameter

In this section we use a logistic reparameterization of the guessing parameter c_j as follows:

$$c_j = \frac{1}{1+e^{-\gamma_j}}. \tag{42}$$

Note that $\gamma_j \in \Re$, and $c_j \in (0,1)$. The probability of a correct response is

$$\pi_{ij} = \frac{1}{1+e^{-\gamma_j}} + \frac{1}{1+e^{\gamma_j}}\cdot\frac{1}{1+e^{-\eta_{ij}}}, \tag{43}$$

The partial derivative of c_j with respect to γ_j is given by

$$\frac{\partial c_j}{\partial \gamma_j} = \frac{e^{-\gamma_j}}{(1+e^{-\gamma_j})^2} = c_j(1-c_j). \tag{44}$$

Then, partial derivatives of l_{ij} with respect to γ_j are given by

$$\frac{\partial l_{ij}}{\partial \gamma_j} = \frac{\partial l_{ij}}{\partial \pi_{ij}}\frac{\partial \pi_{ij}}{\partial c_j}\frac{\partial c_j}{\partial \gamma_j} = \left[\frac{y_{ij}-\pi_{ij}}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{c_j(1-c_j)}{1+e^{\eta_{ij}}}\right].$$

Gradient Equations

Let $\boldsymbol{\xi}_j = (\boldsymbol{\beta}_j^t, \gamma_j)^t$.

$$\frac{\partial l_{ij}}{\partial \gamma_j} = \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})}\right] \left[\frac{e^{-\gamma_j}}{(1 + e^{-\gamma_j})^2}\right] \left[\frac{1}{1 + e^{\eta_{ij}}}\right], \tag{45}$$

$$= \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})}\right] \left[\frac{c_j(1 - c_j)}{1 + e^{\eta_{ij}}}\right], \tag{46}$$

$$\frac{\partial l_{ij}}{\partial \boldsymbol{\beta}_j} = \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})}\right] \left[\frac{1}{1 + e^{\gamma_j}}\right] \left[\frac{e^{\eta_{ij}}}{(1 + e^{\eta_{ij}})^2}\right] \begin{bmatrix}\theta_i \\ 1\end{bmatrix}, \tag{47}$$

$$= \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})} \times \frac{\pi_{ij} - c_j}{1 + e^{\eta_{ij}}}\right] \begin{bmatrix}\theta_i \\ 1\end{bmatrix}, \tag{48}$$

$$\frac{\partial l_{ij}}{\partial \boldsymbol{\xi}_j} = \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})}\right] \left[\frac{1}{1 + e^{\gamma_j}}\right] \left[\frac{e^{\eta_{ij}}}{(1 + e^{\eta_{ij}})^2}\right] \begin{bmatrix}\theta_i \\ 1 \\ \left[\frac{1+e^{-\eta_{ij}}}{1+e^{-\gamma_j}}\right]\end{bmatrix}, \tag{49}$$

$$= \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})}\right] \left[\frac{\pi_{ij} - c_j}{1 + e^{\eta_{ij}}}\right] \begin{bmatrix}\theta_i \\ 1 \\ \left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix}, \tag{50}$$

$$\frac{\partial l_{ij}}{\partial \theta_i} = \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})}\right] \left[\frac{1}{1 + e^{\gamma_j}}\right] \left[\frac{e^{\eta_{ij}}}{(1 + e^{\eta_{ij}})^2}\right] \boldsymbol{a}_j \tag{51}$$

$$= \left[\frac{y_{ij} - \pi_{ij}}{\pi_{ij}(1 - \pi_{ij})} \times \frac{\pi_{ij} - c_j}{1 + e^{\eta_{ij}}}\right] \boldsymbol{a}_j. \tag{52}$$

Information Equations

$$E\left[\frac{\partial L}{\partial \boldsymbol{\beta}_j}\frac{\partial L}{\partial \boldsymbol{\beta}_j^t}\right] = \sum_{i=1}^{N}\left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{1}{1+e^{\gamma_j}}\right]^2 \frac{(e^{\eta_{ij}})^2}{[1+e^{\eta_{ij}}]^4}\begin{bmatrix}\theta_i\\1\end{bmatrix}\begin{bmatrix}\theta_i\\1\end{bmatrix}^t, \tag{53}$$

$$= \sum_{i=1}^{N}\frac{1}{\pi_{ij}(1-\pi_{ij})}\times\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\begin{bmatrix}\theta_i\\1\end{bmatrix}\begin{bmatrix}\theta_i\\1\end{bmatrix}^t \tag{54}$$

$$E\left[\frac{\partial L}{\partial \boldsymbol{\beta}_j}\frac{\partial L}{\partial \gamma_j}\right] = \sum_{i=1}^{N}\left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{(e^{-\gamma_j})^2}{(1+e^{-\gamma_j})^3}\right]\frac{e^{\eta_{ij}}}{[1+e^{\eta_{ij}}]^3}\begin{bmatrix}\theta_i\\1\end{bmatrix} \tag{55}$$

$$E\left[\frac{\partial L}{\partial \gamma_j}\frac{\partial L}{\partial \gamma_j}\right] = \sum_{i=1}^{N}\left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\times\frac{1}{[1+e^{\eta_{ij}}]^2} \tag{56}$$

$$E\left[\frac{\partial L}{\partial \boldsymbol{\xi}_j}\frac{\partial L}{\partial \boldsymbol{\xi}_j^t}\right] = \sum_{i=1}^{N}\left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{1}{1+e^{\gamma_j}}\right]^2\frac{(e^{\eta_{ij}})^2}{[1+e^{\eta_{ij}}]^4}\begin{bmatrix}\theta_i\\1\\\left[\frac{1+e^{-\eta_{ij}}}{1+e^{-\gamma_j}}\right]\end{bmatrix}\begin{bmatrix}\theta_i\\1\\\left[\frac{1+e^{-\eta_{ij}}}{1+e^{-\gamma_j}}\right]\end{bmatrix}^t, \tag{57}$$

$$= \sum_{i=1}^{N}\left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\begin{bmatrix}\theta_i\\1\\\left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix}\begin{bmatrix}\theta_i\\1\\\left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix}^t, \tag{58}$$

$$E\left[\frac{\partial L}{\partial \theta_i}\frac{\partial L}{\partial \theta_i^t}\right] = \sum_{j=1}^{p}\left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\times\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\boldsymbol{a}_j\boldsymbol{a}_j^t \tag{59}$$

$$E\left[\frac{\partial L}{\partial \boldsymbol{\beta}_j}\frac{\partial L}{\partial \theta_i^t}\right] = \left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\times\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\begin{bmatrix}\theta_i\\1\end{bmatrix}\boldsymbol{a}_j^t \tag{60}$$

$$E\left[\frac{\partial L}{\partial \boldsymbol{\xi}_j}\frac{\partial L}{\partial \theta_i^t}\right] = \left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\times\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\begin{bmatrix}\theta_i\\1\\\left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix}\boldsymbol{a}_j^t \tag{61}$$

The final formulas

$$c_j = \frac{1}{1+e^{-\gamma_j}} \tag{62}$$

$$\eta_{ij} = \boldsymbol{a}_j^t \theta_i + d_j \tag{63}$$

$$\pi_{ij} = c_j + \frac{1-c_j}{1+e^{-\eta_{ij}}} \tag{64}$$

$$\boldsymbol{\xi}_j = (\boldsymbol{a}_j^t, d_j, \gamma_j)^t \tag{65}$$

$$L(\boldsymbol{y}|\boldsymbol{\xi},\theta) = \sum_{i=1}^{N}\sum_{j=1}^{p} y_{ij}\log\pi_{ij} + (1-y_{ij})\log(1-\pi_{ij}) \tag{66}$$

$$\frac{\partial L}{\partial \boldsymbol{\xi}_j} = \sum_{i=1}^{N}\left[\frac{y_{ij}-\pi_{ij}}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]\begin{bmatrix}\theta_i \\ 1 \\ \left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix} \tag{67}$$

$$\frac{\partial L}{\partial \theta_i} = \sum_{j=1}^{p}\left[\frac{y_{ij}-\pi_{ij}}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]\boldsymbol{a}_j \tag{68}$$

$$E\left[\frac{\partial L}{\partial \boldsymbol{\xi}_j}\frac{\partial L}{\partial \boldsymbol{\xi}_j^t}\right] = \sum_{i=1}^{N}\left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\begin{bmatrix}\theta_i \\ 1 \\ \left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix}\begin{bmatrix}\theta_i \\ 1 \\ \left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix}^t \tag{69}$$

$$E\left[\frac{\partial L}{\partial \theta_i}\frac{\partial L}{\partial \theta_i^t}\right] = \sum_{j=1}^{p}\left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\boldsymbol{a}_j\boldsymbol{a}_j^t \tag{70}$$

$$E\left[\frac{\partial L}{\partial \boldsymbol{\xi}_j}\frac{\partial L}{\partial \theta_i^t}\right] = \left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\begin{bmatrix}\theta_i \\ 1 \\ \left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix}\boldsymbol{a}_j^t \tag{71}$$

3.5. Unidimensional Item Response Theory model

As a special case, when the dimension is 1, the final formulas are as follows:

$$c_j = \frac{1}{1+e^{-\gamma_j}} \tag{72}$$

$$\eta_{ij} = a_j\theta_i + d_j \tag{73}$$

$$\pi_{ij} = c_j + \frac{1-c_j}{1+e^{-\eta_{ij}}} \tag{74}$$

$$\boldsymbol{\xi}_j = (a_j, d_j, \gamma_j)^t \tag{75}$$

$$L(\boldsymbol{y}|\boldsymbol{\xi},\theta) = \sum_{i=1}^{N}\sum_{j=1}^{p} y_{ij}\log\pi_{ij} + (1-y_{ij})\log(1-\pi_{ij}) \tag{76}$$

$$\frac{\partial L}{\partial \boldsymbol{\xi}_j} = \sum_{i=1}^{N}\left[\frac{y_{ij}-\pi_{ij}}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]\begin{bmatrix}\theta_i\\ 1\\ \left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix} \tag{77}$$

$$\frac{\partial L}{\partial \theta_i} = \sum_{j=1}^{p}\left[\frac{y_{ij}-\pi_{ij}}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]a_j \tag{78}$$

$$E\left[\frac{\partial L}{\partial \boldsymbol{\xi}_j}\frac{\partial L}{\partial \boldsymbol{\xi}_j^t}\right] = \sum_{i=1}^{N}\left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\begin{bmatrix}\theta_i\\ 1\\ \left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix}\begin{bmatrix}\theta_i\\ 1\\ \left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix}^t \tag{79}$$

$$E\left[\frac{\partial L}{\partial \theta_i}\frac{\partial L}{\partial \theta_i^t}\right] = \sum_{j=1}^{p}\left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2 a^2 \tag{80}$$

$$E\left[\frac{\partial L}{\partial \boldsymbol{\xi}_j}\frac{\partial L}{\partial \theta_i}\right] = \left[\frac{1}{\pi_{ij}(1-\pi_{ij})}\right]\left[\frac{\pi_{ij}-c_j}{1+e^{\eta_{ij}}}\right]^2\begin{bmatrix}\theta_i\\ 1\\ \left[\frac{c_j(1-c_j)}{\pi_{ij}-c_j}\right]\end{bmatrix}a_j \tag{81}$$

4. Polytomous MIRT Models

4.1. Basic Definitions

In this work we will use the following notation.

1. It is assumed that the sample size, i.e., the number of examinees is N, and, that there is $s \leq N$ different response patterns. Each observed response pattern is denoted $\boldsymbol{u}_l, l = 1,\ldots,s$. It is assumed that the response pattern $\boldsymbol{u}_l$ has frequency n_l, and $\sum_{l=1}^{s} n_l = N$.

2. Expansion factors. If there a sampling scheme from which the examinees were selected to response the test, the values n_l can be computed as follows. Let υ_i the factor expansion for the examinee $i, i = 1,\ldots,N$ and

define the variables n_{il} as :

$$n_{il} = \begin{cases} 1, \text{ if examinee } i \text{ has the pattern } u_l \\ 0, \text{ otherwise.} \end{cases}$$

Thus, $n_l = \sum_{i=1}^{N} \upsilon_i n_{il}$. We only take integer values, so υ_i must be an integer number. In this case, $\sum_{l=1}^{s} n_l = M$, where M is the population size.

3. It is assumed that the test size, i.e.,the number of items is p. It is assumed that the item $i, i = 1, \ldots, p$ has m_i categories, denoted respectively $k = 1, \ldots, m_i$. Each response pattern has the form $\boldsymbol{u}_l = (u_{l1}, \ldots, u_{lp})^t$, and $u_{li} = k \; k \in \{1, 2, \ldots, m_i\}, i = 1, \ldots, p$.

4. It is assumed that each response pattern $\boldsymbol{u}$ can be linked to a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^t$ in the latent trait space. In this work we will assume that the latent trait space is $\Re^d$; d will be called the dimension of the latent trait space, and $d < p$. A vector $\boldsymbol{\theta}$ will be called a latent trait vector. In the case of unidimensional models we have that $d = 1$.

5. It is assumed that the latent trait vectors $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_s$, linked respectively to the response patterns $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_s$, can be grouped into G homogeneous groups. $\boldsymbol{\theta}_g$ will denote the centroid point of the group g, $g = 1, 2, \ldots, G$, and will be used as the representat of the group g in the estimation process.

6. The observed response patterns will be grouped in the same way that their linked trait vectors. So we group the response patterns in the same G groups, according to the groups of the latent trait vectors.

7. The conditional expected number of response patterns in each of the G groups are denoted respectively $f_1, \ldots, f_G$, and $\sum_{g=1}^{G} f_g = s$.

8. Within a given group $g, g = 1, \ldots, G$, the conditional expected number of response patterns choosing the category $k, k = 1, \ldots, m_i$ to item $i, i = 1, \ldots, p$ is denoted by r_{gik}, and $\sum_{k=1}^{m_i} r_{gik} = f_g$.

9. The conditional expected proportions of response patterns choosing the category $k, k = 1, \ldots, m_i$, to item $i, i = 1, \ldots, p$, into the group $g, g = 1, \ldots, G$, is given by

$$p_{gik} = \frac{r_{gik}}{f_g}, \quad k = 1, \ldots, m_i,$$

and, $\sum_{k=1}^{m_i} p_{gik} = 1$.

10. $\boldsymbol{Y}$, $s \times p$, matrix of pattern responses. Each element is $y_{ji} = k, k = 1, \ldots, m_i$ if the pattern j obtain the category k to item i.

11. The probability that a response pattern belonging to group g has item response category $k, k = 1, \ldots, m_i$, to item $i, 1 = 1, \ldots, p$, is $P_{gki} = E(p_{gik})$, and, $\sum_{k=1}^{m_i} P_{gik} = 1$.

12. $\boldsymbol{R}_i, i = 1, \ldots, p$, is the $G \times m_i$ matrix, whose elements are just $(r_{gik}), \quad g = 1, \ldots, G, \quad k = 1, \ldots, m_i$.

13. $\boldsymbol{P}_i, i = 1, \ldots, p$, is the $G \times m_i$ matrix, whose elements are just $(P_{gik}), \quad g = 1, \ldots, G, \quad k = 1, \ldots, m_i$.

14. $\boldsymbol{X}_i, i = 1, \ldots, p$, is the $s \times m_i$ response matrix of dichotomized responses to item i, whose elements are just $(x_{lik}), \quad l = 1, \ldots, s, \quad k = 1, \ldots, m_i$, where

$$x_{lik} = \begin{cases} 1, & \text{if the response pattern } \boldsymbol{u}_l \text{ has the category } k \text{ to item } i, \\ 0, & \text{otherwise.} \end{cases}$$

4.2. Response probabilities for the graded models

The probability that a response pattern belonging to group g obtain one of the categories $k, k+1, \ldots, m_i - 1$, is denoted P^*_{gik} and is approximately given by

$$P^*_{gik} = \frac{1}{1 + e^{-\eta_{gik}}}, \tag{82}$$

where,

$$\eta_{gik} = \boldsymbol{\alpha}_i^t \boldsymbol{\theta}_g + \gamma_{ik} \tag{83}$$

$$\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{id})^t, \quad i = 1, \ldots, p \tag{84}$$

$$\boldsymbol{\theta}_g = (\theta_{g1}, \ldots, \theta_{gd})^t, \quad g = 1, \ldots, G. \tag{85}$$

For completation we define $P^*_{gi0} = 1$, and, $P^*_{gim_i} = 0$

The probability that a response pattern belonging to group g obtain the category $k = 2, \ldots, m_i - 1$, given by

$$P_{gik} = P^*_{gi,k-1} - P^*_{gi,k} = \frac{1}{1 + e^{-\eta_{gi,k-1}}} - \frac{1}{1 + e^{-\eta_{gi,k}}}. \tag{86}$$

Additionally we have that

$$P_{gi1} = 1 - \frac{1}{1 + e^{-\eta_{gi1}}} \tag{87}$$

$$P_{gim_i} = \frac{1}{1 + e^{-\eta_{gi,m_i-1}}}. \tag{88}$$

Remark 1 *1. In the case of binary items, the category 1 corresponds to correct response.*

2. The probability that a response pattern $\boldsymbol{U}_l$ has the category k to item i, given that its latent trait vector is $\boldsymbol{\theta}$ is given by

$$P_{ik}(\boldsymbol{\theta}) = \frac{1}{1 + e^{-(\boldsymbol{\alpha}_i^t \boldsymbol{\theta} + \gamma_{i,k-1})}} - \frac{1}{1 + e^{-(\boldsymbol{\alpha}_i^t \boldsymbol{\theta} + \gamma_{ik})}}. \tag{89}$$

On the other hand, to ease the estimation process we assume that the latent traits are grouped into homogeneous groups, such that $\boldsymbol{\theta}_g$ is a good approximation of the latent trait vector linked to each response pattern belonging to the cluster. This justifies the probabilities P^*_{gik} and P_{gik} given in the equations (82) and (86).

3. *η_{gik} is a latent linear predictor to the probability that a response pattern belonging to group g obtain at least the category k to item i.*

4. *Classical parametrizations to polytomous multidimensional models can be obtained as follows. Let $a_i = ||\boldsymbol{\alpha}_i||$, and $\boldsymbol{\beta}_i = \boldsymbol{\alpha}_i/a_i$. In this case the latent linear predictor can be rewritten as*

$$\eta_{gik} = a_i(\boldsymbol{\beta}_i^t \boldsymbol{\theta}_g - b_{ik}), \tag{90}$$

   where $b_{ik} = -\gamma_{ik}/a_i$. The unitary vector $\boldsymbol{\beta}_i$ is called the direction of the item i; a_i is the multidimensional discrimination (MDISC) paramter of item i, and b_{ik} the multidimensional difficulty (MDIFF) parameter of the category k of item i.

5. *Classical parametrizations to polytomous unidimensional models can be obtained as follows. In this case $\boldsymbol{\beta}_i$ and $\boldsymbol{\theta}_g$ are numbers. In particular $\boldsymbol{\beta}_i = 1$, and*

$$\eta_{gik} = a_i(\theta_g - b_{ik}), \tag{91}$$

6. *In the classical unidimensional dichotomous model the latent linear predictor has the form*

$$\eta_{gik} = a_i(\theta_g - b_i), \tag{92}$$

4.3. Guessing parameters

In the case of the dichotomous models, it is common to introduce the so called *guessing parameters*. The role of the guessing parameters is to improve the modeling of response patterns with linked latent trait vectors that have small probability of a *correct response*. In the dichotomous models, the categories are usually denote 0 and 1, where 1 means a correct response to item i. The probabity that the pattern l has the category 1 (correct response) to item i given the latent trait vector $\boldsymbol{\theta}_l$ is given by

$$P_{li1} = P[Y_{li} = 1|\boldsymbol{\theta}_l, \boldsymbol{\alpha}_i, \boldsymbol{\gamma}_i, \tau_i] = \tau_i + \frac{1-\tau_i}{1-e^{-(\boldsymbol{\alpha}_i^t\boldsymbol{\theta}_l+\boldsymbol{\gamma}_i)}}, \tag{93}$$

where $\tau_i \in (0,1)$ is the guessing parameter. Note that $P_{li0} = 1 - P_{li1}$. So, with the corresponding changes in the probability of the response patterns the rest of the theory is the same.

4.4. Identifiability of the polytomous MIRT models

To solve the identifiability problem of the models, some constraints are requiered. For the strict multidimensional case, i.e., when $d \geq 2$ we will assume that latent traits $\boldsymbol{\theta}_l$ linked to the observed response patterns $\boldsymbol{u}_l$ are realizations of a random variable $\boldsymbol{\Theta}$ which has a d-dimensional standard normal distribution $N(\mathbf{0}, \boldsymbol{I}_d)$.

To avoid rotation effects and to fix an orientation of the latent traits in the latent trait space we fix for example the first d distrimination paramater in such a way that the matrix $d \times d$ of those $\boldsymbol{a}$ parameters is $\boldsymbol{I}_d$. To fix an origin, one of the corresponding intermediate dificulty parameters, say b_{ik} will be defined as 0. Of course, other restrictions are possible.

In the unidimensional models, it is sufficient to assume the standard normal distribution $N(0, 1)$ for the latent variable Θ.

4.5. ML Estimation of the Item Parameters

Let us suppose that the response patterns are represented by the random variables $\boldsymbol{U}_l, \quad l = 1, \ldots, s$. An observed pattern is denoted $\boldsymbol{u}_l$. Let n_l be the frequency of the observed pattern u_l in the sample. Thus $\sum_{l=1}^{s} n_l = N$, where N is the sample size. Each response pattern has the form $\boldsymbol{U}_l = (U_{l1}, \ldots, U_{lp})^t$, where p is the test size, i.e., the number of items. An observed response pattern is given by $\boldsymbol{u}_l = (u_{l1}, \ldots, u_{lp})^t$, where $u_{li} \in \{1, \ldots, m_i\}$.

Let $\boldsymbol{\xi}_i = (\boldsymbol{\alpha}_i^t, \gamma i1, \ldots, \gamma i, m_i - 1)^t$ be the vector parameter of item i ,and let $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^t, \ldots, \boldsymbol{\xi}_p^t)^t$ be the complete vector of item parameters.

Conditional Joint Probability of the Sample

According to the local independence assumption, the conditional probability that a response pattern $\boldsymbol{U}_l$, given the latent trait vector $\boldsymbol{\theta}_l$, and the complete parameter vector $\boldsymbol{\xi}$, has the observed response pattern $\boldsymbol{u}_l$ is given by

$$P(\boldsymbol{U}_l = \boldsymbol{u}_l | \boldsymbol{\theta}_l, \boldsymbol{\xi}) = \prod_{i=1}^{p} P(U_{li} = u_{li} | \boldsymbol{\theta}_l, \boldsymbol{\xi}_i), \tag{94}$$

where

$$P(U_{li} = u_{li} | \boldsymbol{\theta}_l, \boldsymbol{\xi}_i) = \prod_{k=1}^{m_i} [P_{ik}(\boldsymbol{\theta}_l)]^{x_{lik}}, \tag{95}$$

with $x_{lik} = 1$ if $u_{il} = k$, 0 otherwise, and,

$$P_{ik}(\boldsymbol{\theta}_l) = \frac{1}{1 + e^{-(\boldsymbol{\alpha}_i^t \boldsymbol{\theta}_l + \gamma_{i,k-1})}} - \frac{1}{1 + e^{-(\boldsymbol{\alpha}_i^t \boldsymbol{\theta}_l + \gamma_{ik})}}. \tag{96}$$

Hence, we have that

$$P(\boldsymbol{U}_l = \boldsymbol{u}_l|\boldsymbol{\theta}_l, \boldsymbol{\xi}) = \prod_{i=1}^{p}\prod_{k=1}^{m_i}[P_{ik}(\boldsymbol{\theta}_l)]^{x_{lik}}. \tag{97}$$

Let $\phi(\cdot)$ denotes the density of the standard multivariate normal distribution. The joint probability of the observed pattern $\boldsymbol{u}_l$ and the latent trait vector $\boldsymbol{\theta}_l$ given the item parameters is given by

$$P(\boldsymbol{U}_l = \boldsymbol{u}_l, \boldsymbol{\theta}_l, |\boldsymbol{\xi}) = \prod_{i=1}^{p}\prod_{k=1}^{m_i}[P_{ik}(\boldsymbol{\theta}_l)]^{x_{lik}}\phi(\boldsymbol{\theta}_l), \tag{98}$$

and the marginal probability of $\boldsymbol{U}_l$ given $\boldsymbol{\xi}$ is

$$\bar{P}_l = P(\boldsymbol{U}_l = \boldsymbol{u}_l|\boldsymbol{\xi}) = \int \prod_{i=1}^{p}\prod_{k=1}^{m_i}[P_{ik}(\boldsymbol{\theta}_l)]^{x_{lik}}\phi(\boldsymbol{\theta}_l)d\boldsymbol{\theta}_l. \tag{99}$$

If the centroid points $\boldsymbol{\theta}_g$ are interpreted as quadrature points with weights $w_g, g = 1, \ldots, G$ respectively, the probability $P(\boldsymbol{U}_l = \boldsymbol{u}_l|\boldsymbol{\xi})$ can be approximated by

$$\bar{P}_l = P(\boldsymbol{U}_l = \boldsymbol{u}_l|\boldsymbol{\xi}) \approx \sum_{g=1}^{G}\prod_{i=1}^{p}\prod_{k=1}^{m_i}[P_{ik}(\boldsymbol{\theta}_g)]^{x_{lik}}w_g. \tag{100}$$

Let $\boldsymbol{U} = (\boldsymbol{U}_1, \ldots, \boldsymbol{U}_s)^t$ be the random vector of the complete sample and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_s)^t$ be the complete vector of the unobserved latent traits. The marginal joint probability of $\boldsymbol{U}$ given all the item parameters $\boldsymbol{\xi}$ is

$$P(\boldsymbol{U} = \boldsymbol{u}|\boldsymbol{\xi}) = \prod_{l=1}^{s}\left[\int \prod_{i=1}^{p}\prod_{k=1}^{m_i}[P_{ik}(\boldsymbol{\theta}_l)]^{x_{lik}}\phi(\boldsymbol{\theta}_l)d\boldsymbol{\theta}_l\right]^{n_l}. \tag{101}$$

Prediction of the r_{gik} values

Remember that r_{gik} denote the conditional expected value of the examinees which belonging to group g obtain the category k to item i. By the Bayes theorem we have that

$$\pi_{gl} = P(\boldsymbol{\theta}_g|\boldsymbol{u}_l, \boldsymbol{\xi}) = \frac{P(\boldsymbol{U}_l = \boldsymbol{u}_l|\boldsymbol{\xi}, \boldsymbol{\theta}_g)\phi(\boldsymbol{\theta}_g)}{\bar{P}_l}, \tag{102}$$

which can be approximated using the quadrature points as

$$\pi_{gl} \approx \pi^*_{gl} = \frac{\prod_{i=1}^{p}\prod_{k=1}^{m_i}[P_{ik}(\boldsymbol{\theta}_g)]^{x_{lik}}w_g}{\sum_{h=1}^{G}\prod_{i=1}^{p}\prod_{k=1}^{m_i}[P_{ik}(\boldsymbol{\theta}_h)]^{x_{lik}}w_h}. \tag{103}$$

The value π_{gl} is interpreted as the probability that the group g contais the pattern $\boldsymbol{u}_l$. Consequently, the conditional expected value of examinees belonging to the group g is given by

$$f_g = \sum_{l=1}^{s} n_l \pi_{gl}. \tag{104}$$

and approximated by

$$f_g \approx \sum_{l=1}^{s} n_l \pi^*_{gl}. \tag{105}$$

On the other hand, the conditional expected value of examinees belonging to group g wich obtain the category k to item i is given by

$$r_{gik} = \sum_{l=1}^{s} n_l \pi_{gl} x_{lik}. \tag{106}$$

and approximated by

$$r_{gik} \approx \sum_{l=1}^{s} n_l \pi^*_{gl} x_{lik}. \tag{107}$$

A key observation is that given the quadrature points, the item parameters and the observed scores the values π^*_{gl} can be computed.

Complete Likelihood of the Item Parameters

The joint probability of a response patterns $\boldsymbol{u}_l$ and and the corresponding unobserved latent trait vector $\boldsymbol{\theta}_l$, given the vector of item parameters $\boldsymbol{\xi}$ is

Once the item response are observed we treat them as fixed. Therefore, the marginal likelihood function for all the item parameters $\boldsymbol{\xi}$ can be written as

$$L(\boldsymbol{\xi}|\boldsymbol{u}) = \prod_{l=1}^{s} \left[\int \prod_{i=1}^{p} \prod_{k=1}^{m_i} [P_{ik}(\boldsymbol{\theta}_l)]^{x_{lik}} \phi(\boldsymbol{\theta}_l) d\boldsymbol{\theta}_l \right]^{n_l}. \tag{108}$$

$L(\boldsymbol{\xi}|\boldsymbol{u})$ is named the *observed* data likelihood.

In the spirit of the EM algorithm, the complete sample is given by $\{(\boldsymbol{u}_l, \boldsymbol{\theta}_l),\ \text{with frequency } n_l, l = 1, \ldots, s\}$, where the $\boldsymbol{\theta}$'s are missing data. If we treat the item responses as fixed once they are observed, and also suppose that the latent trait vector were observed, the so called *complete* likelihood function of the vector of item parameters is given by

$$L(\boldsymbol{\xi}|\boldsymbol{u}, \boldsymbol{\theta}) = \prod_{l=1}^{s} [\phi(\boldsymbol{\theta}_l)]^{n_l} \prod_{l=1}^{s} \prod_{i=1}^{p} \prod_{k=1}^{m_i} [P_{ik}(\boldsymbol{\theta}_l)]^{n_l x_{lik}} \tag{109}$$

$$\propto \prod_{l=1}^{s} \prod_{i=1}^{p} \prod_{k=1}^{m_i} [P_{ik}(\boldsymbol{\theta}_l)]^{n_l x_{lik}}. \tag{110}$$

The proportionality holds because $\phi(\boldsymbol{\theta}_l)$ does not depend on $\boldsymbol{\xi}$. The likelihood of the item parameters of item i is

$$L_i(\boldsymbol{\xi}_i|\boldsymbol{u},\boldsymbol{\theta}) \propto \prod_{l=1}^{s}\prod_{k=1}^{m_i}[P_{ik}(\boldsymbol{\theta}_l)]^{n_l x_{lik}}. \tag{111}$$

Ignoring the constants involving $\phi(\boldsymbol{\theta}_l)$ we will write the complete data log-likelihood for the item parameters as

$$l(\boldsymbol{\xi}|\boldsymbol{u},\boldsymbol{\theta}) = \log L(\boldsymbol{\xi}|\boldsymbol{u},\boldsymbol{\theta}) = \sum_{l=1}^{s}\sum_{i=1}^{p}\sum_{k=1}^{m_i} n_l x_{lik}\log P_{ik}(\boldsymbol{\theta}_l). \tag{112}$$

and the complete data log-likelihood for the parameters of item i is

$$l_i(\boldsymbol{\xi}_i|\boldsymbol{u},\boldsymbol{\theta}) = \sum_{l=1}^{s}\sum_{k=1}^{m_i} n_l x_{lik}\log P_{ik}(\boldsymbol{\theta}_l). \tag{113}$$

4.6. The EM Algortihm

Given the item parameters, the Fisher's identity to the marginal probability of the pattern $\boldsymbol{u}_l$ establishes that

$$\frac{\partial \log P(\boldsymbol{u}_l|\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \int \frac{\partial \log P(\boldsymbol{u}_l,\boldsymbol{\theta}_l|\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} P(\boldsymbol{\theta}_l|\boldsymbol{u}_l,\boldsymbol{\xi})d\boldsymbol{\theta}_l. \tag{114}$$

The interpretation of the equation (114) is that the gradient of the observed data log-likelihood $\log(\boldsymbol{\xi}|\boldsymbol{U})$ is equal to the conditional expectation of the gradient of the complete data log-likelihood $\log(\boldsymbol{\xi}|\boldsymbol{U},\boldsymbol{\theta})$ over the posterior distribution of the latent trait vectors given the observed variables.

Let us suppose that $\widetilde{\boldsymbol{\xi}}$ represents a provisional value of $\boldsymbol{\xi}$. Let $\widetilde{P}_{ik}(\boldsymbol{\theta})$ be $P_{ik}(\boldsymbol{\theta})$ evaluate at $\boldsymbol{\xi} = \widetilde{\boldsymbol{\xi}}$. Following the logic of the Fisher's identity the conditional expected complete data likelihood given provisional item parameters $\widetilde{\boldsymbol{\xi}}$ can be approximated by

$$\begin{aligned}
Q(\boldsymbol{\xi}|\boldsymbol{u},\widetilde{\boldsymbol{\xi}}) &= \sum_{l=1}^{s}\sum_{g=1}^{G}\sum_{i=1}^{p}\sum_{k=1}^{m_i} n_l x_{lik}\log P_{ik}(\boldsymbol{\theta}_g)P(\boldsymbol{\theta}_g|\boldsymbol{u}_l,\widetilde{\boldsymbol{\xi}}) && (115)\\
&= \sum_{g=1}^{G}\sum_{i=1}^{p}\sum_{k=1}^{m_i}\log P_{ik}(\boldsymbol{\theta}_g)\left(\sum_{l=1}^{s} n_l x_{lik}P(\boldsymbol{\theta}_g|\boldsymbol{u}_l,\widetilde{\boldsymbol{\xi}})\right) && (116)\\
&= \sum_{g=1}^{G}\sum_{i=1}^{p}\sum_{k=1}^{m_i}\widetilde{r}_{gik}\log P_{ik}(\boldsymbol{\theta}_g), && (117)
\end{aligned}$$

where

$$\widetilde{r}_{gik} = \sum_{l=1}^{s} n_l x_{lik}P(\boldsymbol{\theta}_g|\boldsymbol{u}_l,\widetilde{\boldsymbol{\xi}}) = \sum_{l=1}^{s} n_l x_{lik}\widetilde{\pi}_{gl}, \tag{118}$$

and

$$\widetilde{\pi}_{gl} = \frac{\prod_{i=1}^{p}\prod_{k=1}^{m_i}[\widetilde{P}_{ik}(\boldsymbol{\theta}_g)]^{x_{lik}} w_g}{\sum_{h=1}^{G}\prod_{i=1}^{p}\prod_{k=1}^{m_i}[\widetilde{P}_{ik}(\boldsymbol{\theta}_h)]^{x_{lik}} w_h}. \tag{119}$$

4.7. Steps of the EM Algorithm

1. *Step 0.* Obtain reasonable intitial values $\widetilde{\boldsymbol{\xi}}$ for the complete vector of item parameters.

2. *Step E.* Given provisional values $\widetilde{\boldsymbol{\xi}}$ of the complete vector of item parameters obtain the values $\widetilde{r}_{gik}$ using equations (**??**).

   Let $\widetilde{\boldsymbol{\xi}}_i = (\widetilde{\boldsymbol{\alpha}}_i^t, \widetilde{\gamma}_{i1}, \ldots, \widetilde{\gamma}_{im_{i-1}})^t$

$$\widetilde{P}_{i1}(\boldsymbol{\theta}_g) = 1 - \frac{1}{1+e^{-(\widetilde{\boldsymbol{\alpha}}_i^t\boldsymbol{\theta}_g+\widetilde{\gamma}_{i1})}}, \tag{120}$$

$$\widetilde{P}_{ik}(\boldsymbol{\theta}_g) = \frac{1}{1+e^{-(\widetilde{\boldsymbol{\alpha}}_i^t\boldsymbol{\theta}_g+\widetilde{\gamma}_{i,k-1})}} - \frac{1}{1+e^{-(\widetilde{\boldsymbol{\alpha}}_i^t\boldsymbol{\theta}_g+\widetilde{\gamma}_{ik})}}, k = 2, \ldots, m_i - 1 \tag{121}$$

$$\widetilde{P}_{im_i}(\boldsymbol{\theta}_g) = \frac{1}{1+e^{-(\widetilde{\boldsymbol{\alpha}}_i^t\boldsymbol{\theta}_g+\widetilde{\gamma}_{i,m_i-1})}}; \tag{122}$$

$$\widetilde{\pi}_{gl} = \frac{\prod_{i=1}^{p}\prod_{k=1}^{m_i}[\widetilde{P}_{ik}(\boldsymbol{\theta}_g)]^{x_{lik}} w_g}{\sum_{h=1}^{G}\prod_{i=1}^{p}\prod_{k=1}^{m_i}[\widetilde{P}_{ik}(\boldsymbol{\theta}_h)]^{x_{lik}} w_h}, \tag{123}$$

   $g = 1, \ldots, G; l = 1, \ldots, s;$

$$\widetilde{r}_{gik} = \sum_{l=1}^{s} n_l x_{lik} \widetilde{\pi}_{gl}, \tag{124}$$

   $g = 1, \ldots, G; i = 1, \ldots, p; k = 1, \ldots, m_i.$

3. *Step M.* Obtain new values $\widetilde{\boldsymbol{\xi}}_i$ to the parameters of item $i, i = 1, \ldots .p$ by maximizing

$$\widetilde{Q}_i(\boldsymbol{\xi}_i|\boldsymbol{u}, \widetilde{\boldsymbol{\xi}}) = \sum_{g=1}^{G}\sum_{k=1}^{m_i} \widetilde{r}_{gik} \log P_{ik}(\boldsymbol{\theta}_g). \tag{125}$$

   for $i = 1, \ldots, p,$

where

$$P_{i1}(\boldsymbol{\theta}_g) = 1 - \frac{1}{1+e^{-(\boldsymbol{\alpha}_i^t\boldsymbol{\theta}_g+\gamma_{i1})}}, \tag{126}$$

$$P_{ik}(\boldsymbol{\theta}_g) = \frac{1}{1+e^{-(\boldsymbol{\alpha}_i^t\boldsymbol{\theta}_g+\gamma_{i,k-1})}} - \frac{1}{1+e^{-(\boldsymbol{\alpha}_i^t\boldsymbol{\theta}_g+\gamma_{ik})}}, k = 2, \ldots, m_i - 1 \tag{127}$$

$$P_{im_i}(\boldsymbol{\theta}_g) = \frac{1}{1+e^{-(\boldsymbol{\alpha}_i^t\boldsymbol{\theta}_g+\gamma_{i,m_i-1})}}. \tag{128}$$

4.8. Warming Period, Montenegro

In this section we procedure to obtain the quadrature points, in a way that they follows the structure of the data. First, we porposse to select a representative sample of the observed patterns ando use this patterns and the correspondig factors of expansion. We propose the following sampling scheme,

1. Group the pattern according to its classical score.

2. Select a sampling of patterns in each group porportional to the size of the group (a simple random sample within each group). Let N_0 be the total sample size.

3. Compute the expasion factors of each selected pattern, to obtain the value n_l of each selected pattern, see in section 4.1.

The warming period is as follows.

1. Let d be the dimension of the latent trait space. Obtain a sample from the d-dimensional standard multivariate normal distribution $N(\boldsymbol{O}, \boldsymbol{I})$, of size of G_0, say $G_0 = 200$. The values of this sample will be the initial quadrature points $\boldsymbol{\theta}_g$. The initial weigths w_g will be computed as $w_g = \frac{\phi(\boldsymbol{\theta}_g)}{\sum_{h=1}^{G_0} \phi(\boldsymbol{\theta}_h)}$.

2. Obtain initial reasonable values $\boldsymbol{\xi}^{(0)}$ to all the item parameters $\boldsymbol{\xi}$.

3. Step t. Compute $P(\boldsymbol{\theta}_g|\boldsymbol{u}_l, \boldsymbol{\xi}^{(t)}), g = 1, \ldots, G_0, \quad l = 1, \ldots, N_0$. Assign the pattern $\boldsymbol{u}_l$ to the group with the greatest probability $P(\boldsymbol{\theta}_g|\boldsymbol{u}_l, \boldsymbol{\xi}^{(t)})$.

4. Remove those nodes which group have less size than 0,5 % of the sample, and update G_0.

5. Reasign the patterns of the removed nodes to the group with the greatest probability, after removing the nodes.

6. Update the weigths w_g as n_g/N_0, where n_g is the current size of the group g.

7. Obtain new values to the latent traits $\boldsymbol{\theta}_l$ by maximizing their log-likehood given by

$$T_l = \sum_{i=1}^{p} \sum_{k=1}^{m_i} x_{lik} \log P_{ik}^{(t-1)}(\boldsymbol{\theta}_l), \quad l = 1, \ldots, N_0. \tag{129}$$

8. Update the nodes. The new nodes are the centroids of the current groups. These points will be denoted $\boldsymbol{\theta}_g^{(t)}$

9. *Step EM.* Compute the values $r_{gik}^{(t)}$ as

$$r_{gik}^{(t)} = \sum_{l=1}^{s} n_l x_{lik} P(\boldsymbol{\theta}_g^{(t)}|\boldsymbol{u}_l, \boldsymbol{\xi}^{(t)}) = \sum_{l=1}^{s} n_l x_{lik} \pi_{gl}^{(t)}, \tag{130}$$

and

$$\pi_{gl}^{(t)} = \frac{\prod_{i=1}^{p} \prod_{k=1}^{m_i} [P_{ik}^{(t)}(\boldsymbol{\theta}_g^{(t)})]^{x_{lik}} w_g^{(t)}}{\sum_{h=1}^{G} \prod_{i=1}^{p} \prod_{k=1}^{m_i} [P_{ik}^{(t)}(\boldsymbol{\theta}_h^{(t)})]^{x_{lik}} w_h^{(t)}}. \tag{131}$$

Obtain new values $\boldsymbol{\xi}_i^{(k+1)}$ to the parameters of item $i, i = 1, \ldots . p$ by maximizing

$$\widetilde{Q}_i(\boldsymbol{\xi}_i|\boldsymbol{u}, \boldsymbol{\xi}^{(t)}) = \sum_{g=1}^{G} \sum_{k=1}^{m_i} r_{gik}^{(t)} \log P_{ik}(\boldsymbol{\theta}_g^{(t)}). \tag{132}$$

for $i = 1, \ldots, p$.

10. Stop when $||\boldsymbol{\xi}^{(k-1)} - \boldsymbol{\xi}^{(k)}|| < \epsilon$, for a given ϵ.

The final value of $\boldsymbol{\xi}$ will be the initial value to the full EM algorithm. The final values of $\boldsymbol{\theta}_g$ and w_g will the the quadrature points with the corresponding weigths.

5. Partial Credit Model

For this and the following section as they are so similar to the philosophy and implementation of the polytomous model (see section 4) only some details are presented, this with coherence to Section 1 when we say that just probability must be changed for the estimation process.

$$P_{ik}^{*}(\boldsymbol{\theta}) = \frac{P_{ik}(\boldsymbol{\theta})}{P_{i(k-1)}(\boldsymbol{\theta}) + P_{ik}(\boldsymbol{\theta})} \tag{133}$$

$$= \frac{1}{1 + e^{-\boldsymbol{\theta} - b_{i1}}} \tag{134}$$

with $i = 1, \ldots, p$, (p the number of items) and $k = 1, \ldots, m_i$ (m_i the number of categories of the item i) Note that in this model, all items are assumed to have uniform discriminating power.

6. Generalized Partial Credit Model

For the generalized partial credit model which is a generalization of the partial credit model we have

$$P^*_{ik}(\boldsymbol{\theta}) = \frac{P_{ik}(\boldsymbol{\theta})}{P_{i(k-1)}(\boldsymbol{\theta}) + P_{ik}(\boldsymbol{\theta})} \tag{135}$$

$$= \frac{1}{1 + e^{-\boldsymbol{a}_i^{\boldsymbol{t}}\boldsymbol{\theta} - b_{ik}}} \tag{136}$$

with $i = 1, \ldots, p$, (p the number of items) and $k = 1, \ldots, m_i$ (m_i the number of categories of the item i)