Revista Colombiana de Estadística
Volumen 27 N° 1. Págs. 27 *a* 41. Junio 2004

Modeling Abilities in 3-IRT Models

Edilberto Cepeda Cuervo*
José Manuel Peláez Andrade**

Abstract

This paper considers situations where regression models are proposed to model abilities on three parameter logistic models. After a summary of the classical approach to estimate the item parameters and the abilities, we provide an exposition of the maximum likelihood method to estimate the regression parameters. We analyze how good these estimations are, using a simulated study, and we include an application.

Key words: *Logistic models, Ability modeling, Item Response Theory*

1. Introduction

In Item Response Theory (IRT) we have an estimation problem including two types of parameters: the item parameters and the subject's ability parameters. The difficulty in the general model is that the subject's ability, which appears as a nuisance parameter, cannot be eliminated from the likelihood function by conditioning a sufficient statistic in the way proposed for the one parameter logistic model by Rasch (see Andersen, 1980, chap 6). Moreover, joint maximum likelihood estimation of the subject's abilities and item parameters is not generally possible because the number of parameters increases with the number of subjects and then, standard limit theorems do not apply (Bock & Aitken, 1981).

A better approach to estimation in the presence of a random nuisance parameter is to integrate over the parameter distribution and to estimate the

*E-mail: ecepeda@unal.edu.co. Departamento de Matemáticas. Universidad de los Andes.
**E-mail: jos-pela@uniandes.edu.co. Departamento de Matemáticas. Universidad de los Andes.

structural parameters by maximum likelihood in the marginal distribution. Working with two parameter normal ogive model, Bock & Lieberman (1970) have already taken this approach and have estimated item stress holds and factor loadings based on the assumption that subjects are random samples from a $N(0,1)$ distribution of the ability.

Bock & Aitken (1981) show that, by a simple reformulation of the Bock–Lieberman likelihood equation, a computational solution is possible to find for both, small and large number of items. This reformulation of the likelihood equation clearly states that the form of the ability distribution does not need to be known. Instead, it can be estimated as a discrete distribution on a finite number of points (i.e. a histogram). The item parameters can be estimated by integrating over the empirical distribution; thus freeing the method from arbitrary assumptions about the ability distribution in the population effectively sampled.

Subject's abilities can be explained by associated factors such as habits, gender, number of hours of practice, socioeconomic status and parents' education level, among others. If $x_j^t = (1, x_{ij}, ..., x_{Rj})$, $j = 1, 2, ..., n$, is a vector of ability explanatory variables, we can model the abilities using $\theta_j = x_j^t\beta + \epsilon_j$, where $\beta^t = (\beta_0, \beta_1, ..., \beta_R)$ is a regression parameters vector and $\epsilon_j \sim N(0, \sigma^2)$. This is not a new idea. Verhelst & Eggen (1989) and Zwinderman (1991) formulated a structural model for the Rasch model assuming that the item parameters are known.

In this paper we model the abilities as a regression model on a 3-IRT model. Section 2 presents some topics on three parameter logistic models. Section 3 reviews the marginal maximum likelihood estimate of item parameters. Section 4 reviews the maximum likelihood estimate of the abilities. Section 5 presents the ability regression models and the ML method for the parameter estimation. Section 6 includes a simulated study. Section 7 presents an application and section 8 draws concluding remarks.

2. Three Parameter Logistic Models

Let u_{ij}, $i = 1, ..., I$ and $j = 1, ..., n$, be $I \times n$ binary random variables, where i indicates an item and j a subject. $u_{ij} = 1$ if subject j solves item i, otherwise $u_{ij} = 0$. The probability that a subject j with ability parameter θ_j solves item i with difficulty parameter b_i is given by:

$$p(u_{ij} = 1|\theta_j, \xi_i) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \quad i = 1, ..., I; \; j = 1, ..., n, \quad (1)$$

where $\xi_i = (a_i, b_i, c_i)$ are the parameters of item i, $0 \leq c_i < 1$ and $a_i > 0$ (Birnbaun, 1968). θ_j and b_i assume values between $-\infty$ and ∞. The latent continuous θ is called the ability, and θ_j is the ability of j-th subject.

The following considerations provide a basis for the interpretation of the parameters involved in the IRT model:

1. Since $c_i = \lim_{\theta_j \to -\infty} p(u_i = 1|\theta_j, \xi_i)$, c_i can be interpreted as the probability of random correct answers.
2. If $\theta_j = b_i$, equation (1) reduces to:
$$p(u_{ij} = 1|\theta_j, \xi_i) = \frac{1 + c_i}{2} \ ;$$
then b_i may be interpreted as the subject's ability necessary for the probability of solving the item equals $\frac{1 + c_i}{2}$; thus, higher values of b_i correspond to higher difficulty levels for the item being considered.
3. Viewing $p(u_{ij} = 1|\theta_j, \xi_i)$ as a function f of θ_j, $f'(b_i) = \frac{1}{2}(1 - c_i)Da_i$.
This fact justifies to take a_i as a discrimination measure. If for items i and j, $c_i = c_j$ and $a_i < a_j$, then the item with discrimination measure a_j discriminates more than the item with discrimination measure a_i.

Given an item with known parameters, the curve which describes the relation between subject's ability parameter and agreement probability is called Item Characteristic Curve (ICC) (See figure 1). If in equation (1) $c_i = 0$, then there is no possibility of having random correct answers and the resulting model is called the unidimensional two parameter logistic model or Birnbaum's model. Setting $a_i = a$ and $c = 0$ in equation (1), we obtain the unidimensional one parameter logistic model, also called the *Rasch Model* (1961). The Rasch model describes how the probability of correct answers depends on the subject's overall ability and on the level of difficulty of the questions.

One of the most important theoretical merits of the Rasch models is called by Fisher (1995) the "specific objectivity", consisting in that item parameters do not depend on the characteristics of the subjects answering the test, and subject's parameters do not depend on the items chosen from a given set. Consequently, the three parameters are independent from the sample location and dispersion of the ability.

As the ability scale does not have practical interpretation in pedagogical terms, it is necessary to define an ability scale characterized by sets of items

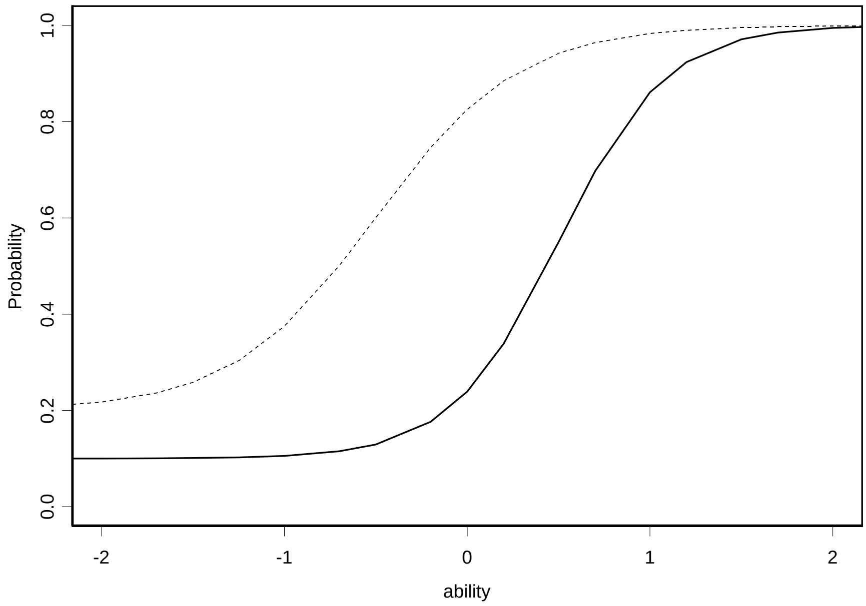

Figure 1: Item Characteristic Curve: Probability of agreement with (a) $c = 0.1$, $a = 2$ and $b = 0.5$ (full line), (b) $c = 0.2$, $a = 1.5$ and $b = -0.59$ (dashed line).

with pedagogical interpretation in the test theoretical frame. The ability scale is defined by a set of values, $\theta_0 < \theta_1 < ... < \theta_P$, called "anchor levels", selected by the analyst. The pedagogical interpretation of the scale is possible through the pedagogical interpretation of the set of items associated with each level.

Pertinent anchor levels, θ_p, depend of the conditional probabilities of correct answer $p(u = 1|\theta = \theta_p)$ and $p(u = 1|\theta = \theta_{p-1})$. Specifically, an item u is one anchor item of a level θ_p if and only if $p(u = 1|\theta = \theta_p) \geq 0.65$, $p(u = 1|\theta = \theta_{p-1}) < 0.5$, and $p(u = 1|\theta = \theta_p) - p(u = 1|\theta = \theta_{p-1}) \geq 0.30$ (see Andrade, 2000). Thus, the scale is defined only at the end of the statistical analysis of the data resulting from the test application.

3. Item parameter estimation

In this section we present the marginal maximum likelihood method to estimate item parameters, based on Andrade (2000). To this end, we consider two hypothesis:

1. Answers of different subjects are independent from each other.

2. Different items are solved independently by each subject, given their ability.

To gain computational advantage it is convenient to work with different response patterns (Bock & Lieberman, 1970). Assuming the ability having a known distribution $g(\theta|\eta)$, if $\tilde{u}_l$ is a response pattern,

$$p(\tilde{u}_l|\xi,\eta) = \int_R p(\tilde{u}_l|\theta,\xi)g(\theta|\eta)d\theta, \tag{2}$$

where $\xi = (\xi_i, ..., \xi_n)$ and $\xi_i = (a_i, b_i, c_i) = (\xi_{i1}, \xi_{i2}, \xi_{i3})$.

Given the independence between different subject's answers (hypothesis 1), if γ_l, is the number of occurrences of response pattern $\tilde{u}_l$, $l = 1, ..., s$, where s is the number of different response patterns with $\gamma_l > 0$, then the likelihood function is:

$$L(\xi,\eta) = \frac{n!}{\prod_{l=1}^{s}\gamma_l!}\prod_{l=1}^{s} p(\tilde{u}_l|\xi,\eta)^{\gamma_l}.$$

Thus, the log likelihood function is:

$$\ell = \log\left(\frac{n!}{\prod_{l=1}^{s}\gamma_l!}\right) + \sum_{l=1}^{s}\gamma_l \log\Big[p(\tilde{u}_l|\xi,\eta)\Big],$$

and

$$\frac{\partial \ell}{\partial \xi_{ir}} = \sum_{l=1}^{s}\frac{\gamma_l}{p(\tilde{u}_l|\xi,\eta)}\frac{\partial}{\partial \xi_{ir}}p(\tilde{u}_l|\xi,\eta),\ i = 1, ..., I;\ r = 1, 2, 3. \tag{3}$$

By the independence between responses, we have:

$$\begin{aligned}\frac{\partial p(\tilde{u}_l|\xi,\eta)}{\partial \xi_{ir}} &= \frac{\partial}{\partial \xi_{ir}} \prod_{j=1}^{I} p(u_{jl}|\xi_j,\eta)\\ &= \frac{\partial}{\partial \xi_{ir}} \prod_{j=1}^{I} \int_R p(u_{jl}|\theta,\xi_j) g(\theta|\eta) d\theta\\ &= \int_R \Big[\prod_{j\neq i}^{I} p(u_{jl}|\theta,\xi_j)\Big] \Big[\frac{\partial}{\partial \xi_{ir}} p(u_{il}|\theta,\xi_j) g(\theta|\eta)\Big] d\theta,\end{aligned}$$

where $u_{jl}=1$ if in the l-th pattern of response, item j is answered.

Since $\frac{\partial}{\partial \xi_{ir}} p(u_{il}|\theta,\xi_i) = \frac{\partial}{\partial \xi_{ir}} \Big[p_i^{u_{il}} q_i^{1-u_{il}}\Big] = (-1)^{u_{il}+1} \frac{\partial p_i}{\partial \xi_{ir}}$, $i = 1,...,I$; $l = 1, 2, .., s$, where $p_i = c_i + (1-c_i)\frac{1}{1+e^{-Da_i(\theta-b_i)}}$, we have:

$$\begin{aligned}\frac{\partial p(\tilde{u}_l|\xi,\eta)}{\partial \xi_{ir}} &= \int_R \Big[\frac{(-1)^{u_{il}+1}}{p_i^{u_{il}} q_i^{1-u_{il}}}\Big] \Big[\frac{\partial}{\partial \xi_{ir}} p(\tilde{u}_l|\theta,\xi) g(\theta|\eta)\Big] d\theta\\ &= \int_R \Big[\Big(\frac{u_{il}-p_i}{p_i q_i}\Big)\Big(\frac{\partial p_i}{\partial \xi_{ir}}\Big)\Big] p(\tilde{u}_l|\theta,\xi) g(\theta|\eta) d\theta. \end{aligned} \tag{4}$$

Given (4), equation (3) can be written as:

$$\frac{\partial \ell}{\partial \xi_{ir}} = \sum_{l=1}^{s} \gamma_l \int_R \Big[(u_{il}-p_i)\frac{\partial p_i}{\partial \xi_{ir}} \frac{W_i}{p_i^* q_i^*}\Big] g_l^*(\theta) d\theta, \tag{5}$$

where $p_i^* = \{1+e^{-Da_i(\theta-b_i)}\}^{-1}$, $q_i^* = 1-p_i^*$, $W_i = \frac{p_i^* q_i^*}{p_i q_i}$ and

$$g_l^*(\theta) = g(\theta|u_l,\eta,\xi) = \frac{p(\tilde{u}_l|\theta,\xi) g(\theta|\eta)}{p(\tilde{u}_l|\xi,\eta)}.$$

The distribution $g_l^*(\theta)$ is the conditional distribution of θ_l given η.

From equation (5) we obtain:

$$\frac{\partial \ell}{\partial a_i} = D(1-c_i) \sum_{l=1}^{s} \gamma_l \int_R \Big[(u_{il}-p_i)(\theta-b_i) W_i\Big] g_l^*(\theta) d\theta,$$

$$\frac{\partial \ell}{\partial b_i} = -Da_i(1-c_i) \sum_{l=1}^{s} \gamma_l \int_R \Big[(u_{il}-p_i) W_i\Big] g_l^*(\theta) d\theta,$$

$$\frac{\partial \ell}{\partial c_i} = \sum_{l=1}^{s} \gamma_l \int_R \Big[(u_{il} - p_i) \frac{W_i}{p_i^*} \Big] g_l^*(\theta) d\theta.$$

Then the equations for the estimation of item parameters are: $\frac{\partial \ell}{\partial a_i} = 0$, $\frac{\partial \ell}{\partial b_i} = 0$, $\frac{\partial \ell}{\partial c_i} = 0$, i=1,...,I.

In order to apply the Newton-Raphson algorithm or Fisher scoring algorithm, we need the second derivative of $\ell(\xi, \eta)$. Given the independence between items (Bock & Aitkin, 1981), item parameters can be estimated individually, since from the hypothesis, $\frac{\partial^2 \log \ell}{\partial \xi_{ir} \partial \xi_{jk}} = 0$, for $i \neq j$, $r, k = 1, 2, 3$. For $i = 1, 2, ...I$ and $j = i$,

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \xi_i \partial \xi_i^t} &= \frac{\partial}{\partial \xi_i} \left(\frac{\partial \ell}{\partial \xi_i} \right)^t = \frac{\partial}{\partial \xi_i} \left[\frac{\log L(\xi, \eta)}{\partial \xi_i} \right]^t \\ &= \frac{\partial}{\partial \xi_i} \left[\sum_{l=1}^{s} \gamma_j \frac{1}{p(\tilde{u}_l | \xi, \eta))} \frac{p(\tilde{u}_l | \xi, \eta)}{\partial \xi_i} \right]^t \\ &= \sum_{l=1}^{s} \gamma_j \Bigg\{ \frac{\partial^2 p(\tilde{u}_l | \xi, \eta) / \partial \xi_i \partial \xi_i^t}{p(\tilde{u}_l | \xi, \eta)} \\ &\quad - \left(\frac{\partial p(\tilde{u}_l | \xi, \eta) / \xi_i}{p(\tilde{u}_j | \xi, \eta)} \right) \left(\frac{\partial p(\tilde{u}_l | \xi, \eta) / \partial \xi_i}{p(\tilde{u}_l | \xi, \eta)} \right)^t \Bigg\}. \end{aligned} \tag{6}$$

Thus the Newton-Raphson algorithm equation is:

$$\hat{\xi}^{(k+1)} = \xi^{(k)} - (H^{(k)})^{-1} q^{(k)}, \tag{7}$$

where $\hat{q}^{(k)} = q(\hat{\xi}^{(k)})$, $q = (q_1, q_2, ..., q_7)$ with $q_i = \frac{\partial \log L(\xi, \eta)}{\partial \xi_i}$,

$H^{(k)} = diag(H_i^{(k)})$ and $H_i^{(k)} = \hat{H}_i^{(k)})$, where

$$H_i = \frac{\partial^2 \ell(\xi, \eta)}{\partial \xi_i \partial \xi_i'}.$$

Given the independence between items, the computational process is simple. However, when a theoretical structure for test elaboration exists, this can induce a linked structure in the items set. There will be groups formed by independent items and groups with non independent items. In this case, the probability of some response vector is given by the product of the group probabilities instead of the product of probabilities of individual item responses. For detail and comments about estimation methods see Mark & Raymond(1995) and Adams & Wilson(1992).

4. Ability estimation

Now we know the item parameters. By hypothesis (1) and (2), the logarithm of likelihood function can be written as:

$$\ell(\theta) = \log L(\theta) = \sum_{j=1}^{n}\sum_{i=1}^{I}\big\{u_{ij}\log(p_{ij}) + (1-u_{ij})\log(q_{ij})\big\},$$

where $p_{ij} = p(u_{ij} = 1|\theta_j, \xi_i)$ and $q_{ij} = 1 - p_{ij}$.

Then for $j = 1, 2, ..., n$,

$$\begin{aligned}\frac{\partial\ell(\theta)}{\partial\theta_j} &= \sum_{i=1}^{I}\left\{\frac{u_{ij}-p_{ij}}{p_{ij}q_{ij}}\right\}\left[\frac{\partial p_{ij}}{\partial\theta_j}\right]\\ &= D\sum_{i=1}^{I}a_i(1-c_i)(u_{ij}-p_{ij})W_{ij},\end{aligned}$$

where

$$W_{ij} = \frac{p^*_{ij}q^*_{ij}}{p_{ij}q_{ij}},$$

$p^*_{ij} = \big\{1 + e^{-Da_i(x_j^t\beta - b_i)}\big\}^{-1}$ and $q^*_{ij} = 1 - p^*_{ij}$.

In order to apply Newton-Raphson algorithm or Fisher scoring algorithm, we need to calculate the second derivatives, $\dfrac{\partial\ell^2(\theta)}{\partial\theta_j^2}$, since by hypothesis (2), section (3), $\dfrac{\partial^2\ell(\theta)}{\partial\theta_j\partial\theta_k} = 0$ for $k \neq j$.

$$\frac{\partial^2\ell(\theta)}{\partial\theta_j^2} = \sum_{i=1}^{I}\left\{\left(\frac{u_{ij}-p_{ij}}{p_{ij}q_{ij}}\right)\left(\frac{\partial^2 p_{ij}}{\partial\theta_j^2}\right) - \left(\frac{u_{ij}-p_{ij}}{p_{ij}q_{ij}}\right)^2\left(\frac{\partial p_{ij}}{\partial\theta_j}\right)^2\right\} \tag{8}$$

where

$$\begin{aligned}\frac{\partial p_{ij}}{\partial \theta_j} &= Da_i(1-c_i)p^*_{ij}q^*_{ij},\\ \frac{\partial^2 p_{ij}}{\partial \theta_j^2} &= D^2a_i^2(1-c_i)(1-2p^*_{ij})(p^*_{ij}q^*_{ij})^2.\end{aligned}$$

As in (7) we can use the Newton-Raphson algorithm to estimate subject's abilities.

5. Modeling the abilities

Subject's abilities may be explained by associated factors such as hours of practice, socioeconomic status or education level, with the model $\theta = x^t\beta + \epsilon$. In this model $x^t = (1, x_1, ..., x_R)$, is a explanatory variables vector; $\beta = (\beta_0, \beta_1, ..., \beta_R)$ is a parameter vector and ϵ is a random variable associated with error. Thus, model (1) can be written as:

$$p(u_{ij} = 1|\beta, \xi_i, \epsilon_j) = c_i + (1-c_i)\frac{1}{1 + e^{-Da_i(x_j^t\beta + \epsilon_j - b_i)}} := p_{ij}, \qquad (9)$$

for $i = 1, ..., I,\ j = 1, ..., n$.

Assuming that the distribution of the random variable ϵ_j, $g(\epsilon_j|\eta)$, is known,

$$P(u_{ij} = 1|\beta, \xi_i, \eta) = \int_R p(u_{ij} = 1|\beta, \xi_i, \epsilon_j)g(\epsilon_j|\eta)d\epsilon_j := P_{ij}, \qquad (10)$$

where $\beta = (\beta_0, \beta_1, ..., \beta_R)^t$, $\xi = (\xi_1, ..., \xi_n)^t$, $\xi_i = (a_i, b_i, c_i)^t = (\xi_{i1}, \xi_{i2}, \xi_{i3})^t$ and $g(\epsilon_j|\eta) = N(0, \sigma^2)$.

Given the independence between answers from different subjects (hypothesis 1) and that different items are solved independently by each subject (hypothesis 2), the logarithm of the likelihood function is:

$$\ell = \sum_{j=1}^{n}\sum_{i=1}^{I}\Big\{u_{ij}\log(P_{ij}) + (1-u_{ij})\log(Q_{ij})\Big\}.$$

Given that

$$\begin{aligned}\frac{\partial P(u_{ij}|\beta, \xi_i, \eta)}{\partial \beta_r} &= \frac{\partial}{\partial \beta_r}\int_R p(u_{ij}|\beta, \xi_i, \epsilon_j)g(\epsilon_j|\eta)d\epsilon_j\\ &= \int_R \Big[\frac{\partial}{\partial \beta_r}p(u_{ij}|\beta, \xi_i, \epsilon_j)g(\epsilon_j|\eta)\Big]d\epsilon_j,\end{aligned}$$

for $r = 1, 2, ..., R$, we obtain:

$$\frac{\partial \ell}{\partial \beta_r} = \sum_{j=1}^{n}\sum_{i=1}^{I}\int_R \Big[\Big(\frac{u_{ij}-P_{ij}}{P_{ij}Q_{ij}}\Big)\Big(\frac{\partial p_{ij}}{\partial \beta_r}\Big)\Big] g(\epsilon_j|\eta)d\epsilon_j, \tag{11}$$

where $p_{ij} = p(u_{ij} = 1|\beta, \xi_i, \eta, \epsilon_j)$ as in (9), $P_{ij} = p(u_{ij} = 1|\beta, \xi_i, \eta)$ as in (10) and $Q_{ij} = 1 - P_{ij} = p(u_{ij} = 0|\beta, \xi_i, \eta)$. So equation (11) can be written as:

$$\frac{\partial \ell(\theta)}{\partial \beta_r} = D\sum_{j=1}^{n}\sum_{i=1}^{I}(1-c_i)a_i x_{jr}\int_{\mathbf{R}}(u_{ij}-P_{ij})W_{ij}g(\epsilon_j|\eta)d\epsilon_j,$$

where $W_{ij} = \dfrac{p^*_{ij} * q^*_{ij}}{P_{ij} * Q_{ij}}$, $p^*_{ij} = \big(1 + e^{-Da_i(x_i^t\beta+\epsilon_j-b_i)}\big)^{-1}$ and $q^*_{ij} = 1 - p_{ij}$, $j = 1, ..., R$.

Finally, the estimates are obtained by solving the equations $\dfrac{\partial \ell}{\partial \beta_r} = 0$, $r = 1, 2, ..., R$.

The marginal maximum likelihood estimation of regression parameters, applying the Fisher scoring algorithm, requires the second derivative matrix, given by:

$$\begin{aligned}\frac{\partial^2 \ell(\theta)}{\partial \beta_r \beta_s} = {} & \sum_{j=1}^{J}\sum_{i=1}^{I}\int_R \Big\{\Big(\frac{u_{ij}-P_{ij}}{P_{ij}Q_{ij}}\Big)\Big(\frac{\partial^2 p_{ij}}{\partial \beta_s \beta_r}\Big) - \\ & \Big(\frac{u_{ij}-P_{ij}}{P_{ij}Q_{ij}}\Big)^2\Big(\frac{\partial P_{ij}}{\partial \beta_s}\Big)\Big(\frac{\partial p_{ij}}{\partial \beta_r}\Big)\Big\} g(\epsilon_j|\eta)d\epsilon_j,\end{aligned} \tag{12}$$

where $$\frac{\partial^2 p_{ij}}{\partial \beta_r \beta_s} = (1-c_i)D^2 a_i^2 x_{js}x_{jr}(1-2p^*_{ij})p^*_{ij} * q^*_{ij}$$

and $$\frac{\partial P_{ij}}{\partial \beta_s} = (1-c_i)Da_i x_{jr}\textstyle\int_{\mathbf{R}} p^*_{ij}q^*_{ij}g(\epsilon_j|\eta).$$

Thus, the Fisher scoring algorithm is given by:

$$\hat\beta^{k+1} = \hat\beta^k + \big[I(\beta^k)\big]^{-1}\hat q^k,$$

where $\hat q^k = \hat q\big(\beta^{(k)}\big)$, $q = (q_1, q_2, ..., q_r)$ with $q_i = \dfrac{\partial \ell(\theta)}{\partial \beta_i}$. The second information matrix is $I^k = (h_{ij})$, with $h_{ij} = -E\Big[\dfrac{\partial^2 \ell(\theta)}{\partial \beta_i \partial \beta_r}\Big]$.

We may also consider models with two levels of explanatory variables. For example, school variables and student's variables. Then in the k-th school we can model the abilities as $\theta_j = \beta_{0k} + \beta_{1k}x_j + \epsilon_j$, where x_j is the ability explanatory variable corresponding to the j-th student. If the parameters β_{0k} and β_{1k} are not the same between schools, it can be modeled as a function of school variables z_k, given by $\beta_{0k} = \alpha_0 + \alpha_1 z_k$ and $\beta_{1k} = \gamma_0 + \gamma_1 z_k$ for $k = 1, ..., K$. So, $\theta_j = \alpha_0 + \alpha_1 z_k + (\gamma_0 + \gamma_1 z_k)x_j + \epsilon_j$.

6. Simulated study

A simulated study was conducted to examine how the estimates are similar to the original parameters when we model the ability in the IRT models as a function of some explanatory variables (Model 13). Initially, we considered twenty five items, with $a_i's$ and $b_i's$ values simulated from uniform distributions $U(1, 1.5)$ and $U(-1.5, 1.5)$, respectively. The c_i values were simulated from a discrete uniform distribution with mass value 0.5 on 0.1, and on 0.2. For the variables X_1, X_2 and X_3 were simulated n=100 (500) values: $x_{1i} = 1$ to obtain an intercept model, x_{2i} from a discrete distribution with $P(X_2 = 0) = 0.4$ and $P(X_2 = 1) = 0.6$, x_{3i} from an uniform distribution on the interval $(-10, 10)$. The values y_i of interest variables Y were generated from a Bernoulli distribution with parameters $(1, p_{ij})$, where p_{ij} is given by the deterministic model:

$$p(u_{ij} = 1|\beta, \xi_i) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i(1-0.5x_{1j}+0.02x_{2i}-b_i)}} , \qquad (13)$$

for $i = 1, ..., 25$ and $j = 1, ..., n$.

The estimates of regression parameters (with standard deviations) are given in table 1.

Table 1: Ability model

I	**n**		$\beta_{\mathbf{0}}$	$\beta_{\mathbf{1}}$	$\beta_{\mathbf{2}}$
25	100	Estimates	9.955×10^{-1}	-5.1749×10^{-1}	2.5479×10^{-2}
		S. Deviation	5.487×10^{-2}	6.817×10^{-2}	5.9945×10^{-3}
25	500	Estimates	9.819×10^{-1}	-4.837×10^{-1}	1.926×10^{-2}
		S. Deviation	2.357×10^{-2}	2.980×10^{-2}	2.592×10^{-3}
10	500	Estimates	9.292×10^{-1}	-4.177×10^{-1}	2.441×10^{-2}
		S. Deviation	4.220×10^{-2}	4.955×10^{-2}	4.038×10^{-3}

The simulation shows that using more data, better estimate values will be obtained, but with sets with fewer data, precision decreases quickly. In general, we need larger samples to obtain better estimates, closer to true values and with small variance. We can see better estimates of parameters and smaller standard deviations in the study with $I = 25$ and $n = 500$, as expected from the increasing likelihood information. For example, with twenty five items, a sample size of thirty or fifty students may be very small. The conclusions are the same for non deterministic models, where $\theta_j = \mu_j + \epsilon_j$.

7. Application

In the universities at El Salvador, candidates applying for spanish teachers are evaluated by the Ministry of Education. With an evaluation of academical and pedagogical abilities of the candidates, the Ministry intends to improve the education quality. The academic and pedagogical abilities in spanish teaching are measured in the effectiveness of approaching capacity to language analysis, specifications of communicative phenomenon, universal and local literature, and pedagogical techniques which guarantee an efficient practice of the teacher's job in these topics in the first levels of the educational process.

The test to explore academic abilities included topics knowledge and disciplinary theories, as well as about the student's capacity of developing strategies to teach these concepts. The test, with one hundred items, includes the following topics: literature, language, communicative and didactic abilities. It was answered by 230 students. As predictor variables of academic abilities we used X_1=gender (0=male, 1=female) and X_2=age.

From the marginal likelihood function we found the item parameters. We selected twenty one items with discrimination parameters: a_i= 0.905, 0.833, 0.44, 0.526, 0.64, 0.74, 0.61, 1.09, 1.28, 1.25, 1.28, 1.40, 0.8, 0.96, 0.55, 0.889, 1.33,0.83, 0.549, 0.59, 0.68; difficulty parameters: b_i=0.42, 0.67, -1.32, 0.83, 0.06, -0.24, 0.917, -0.77, 0.204, 0.052, 0.47,-0.978, -0.93, 1.29, -0.025, 0.33, 2.26, 0.54, 0.98, -1.16, -0.24; and probability of random correct answers: c_i =0.16, 0.18, 0.19, 0.18, 0.18, 0.16, 0.18, 0.17, 0.19,0.14,0.12, 0.17, 0.18, 0.12, 0.18, 0.18, 0.11, 0.19, 0.2, 0.18, 0.179, i=1,2,...,21.

In this application σ^2 is unknown. $(\hat{\beta}, \hat{\sigma}^2)$ can be estimated using different algorithms, (some of them will be presented in a future paper) but here we estimate the regression parameters β for fixed values of σ^2, between 0 and 1. First, for each value of σ^2 the regression model parameters are estimated as in section 4. Next, we determine the likelihood value $L(\hat{\beta}, \sigma_i^2)$. Finally, we compare the likelihood values, to get the estimates as the values corresponding to the maximum of the likelihood function. Figure 2 is a plot of log likelihood versus variance. It shows an increasing behavior before 0.35 (± 0.01) and decreasing behavior afterwards. For $\sigma^2 = 0.35$ the maximum likelihood estimates (and standard deviation) are $\hat{\beta}_0 = 0.740(0.047)$, $\hat{\beta}_1 = -2.083 \times 10^{-2}(4.367 \times 10^{-3})$, and $\hat{\beta}_2 = 0.526(0.092)$. From these estimates we can see that gender and age

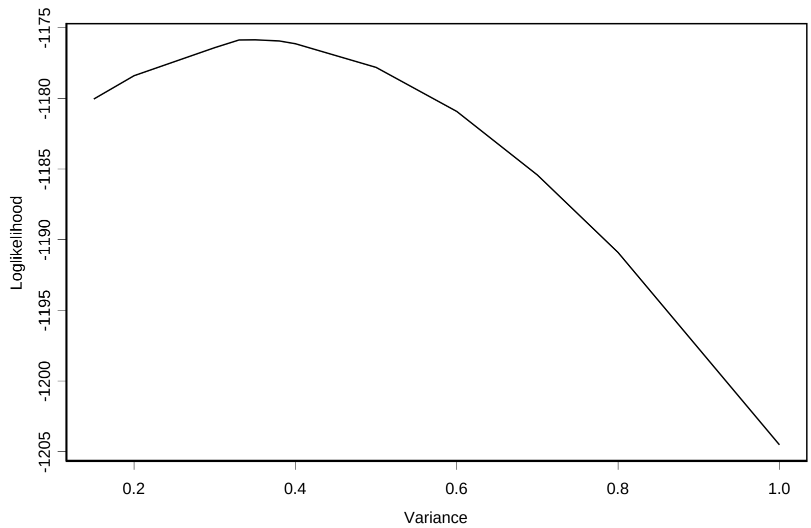

Figure 2: Log likelihood function of variance.

have contribution in the statistical explanation of differences in the abilities of university students. An interpretation is that female students have more developed language abilities than male students, while older students have less developed language abilities than younger students.

8. Conclusions

The subject's parameters estimates may be subject to considerable error and consequently these should not be considered as dependent variables in regression models (Zwindeman, 1991), although latent regression analysis with subject-level predictors would eliminate such problems. In general, it is necessary to consider large samples in order to obtain good estimates, close to true values, with small variance. Samples of size fifty or less should be considered, in the most of the cases, too small to obtain acceptable estimates, althought a rule establishing the minimum number of items in an test does not exist.

Latent regression can be used to determine the "mean ability" of groups of students, for example to compare schools, including indicator functions as explanatory variables. Other extensions are possible, for example, explore classical and bayesian metodologies to model mean and variance parameters.

Acknowledgements

This research was supported by research grants from Facultad de Ciencias de la Universidad de los Andes.

References

Adams, R. A. & Wilson, M. (1992), A random coefficients multinomial logit: Generalizing Rasch models, Technical report, Paper presented at the annual meeting of the American Educational Research Association. San Francisco.

Anddersen, E. B. (1980), *Discrete statistical models with social science applications*, North-Holand, Publishing Company.

Andrade, D. F., Tavares, H. & Cunha, R. (2000), *Teoria da resposta ao item: conceitos e aplicacões*, ABE–Associacão Brasileira de Estatística: São Pablo.

Birnbaun, A. (1968), Some latent trait models and their use in inferring an examinee's ability, *in* 'Statistical Theories Of Mental Tests Score', Addison-Wesley, Reading, MA, pp. 395–479.

Bock, R. D. & Aitkin, M. (1981), 'Marginal maximun likelihood estimation of item parameters: An application of a EM algorithm', *Psychometrika* **46**, 179–197.

Bock, R. D. & Lieberman, M. (1970), 'Fitting a response model for n dichotomously scored items', *Psychometrika* **35**, 179–197.

Fisher, G. H. (1995), 'Some neglected problems in IRT', *Psychometrika* **60**, 449–487.

Mark, W. & Raymond, J. A. (1995), 'Rasch models for item bundles', *Psychometrika* **60**(2), 181–198.

Mislevi, R. J. (1983), 'Item response model for grouped data', *Journal of Educational Statistics* **8**(4), 271–288.

Rasch, G. (1960), *Probabilistic Models for Some Intelligence and Attainment Test*, Danish Institute for Educational Research.

Zwinderman, A. H. (1991), 'A generalized Rasch model for manifest predictors', *Psychometrika* **56**, 589–600.