Universidade Federal do Rio Grande do Sul
Instituto de Matemática

Latent Variable Models for Binary Response

Maria T. Albanese

Cadernos de Matemática e Estatística

Série A, nº 17, ABR/91
Porto Alegre, abril de 1991

[1]This paper is a summary of the Ph.D. Thesis present at University of London, september 1990, under supervision of Dr. Martin Knott.

SUMMARY

1- *Introduction*

In this paper variables will be either categorical or metrical. Categorical variables are measured in a nominal or ordinal scale while metrical variables assume values in an interval or ratio scale.

Whether categorical or metrical a variable will be either manifest (directly observable) or latent (not directly observable, and generally called a factor in factor analysis).

Bartholomew (1987) has classified latent variable models according to the type of latent and manifest variables:

	MANIFEST	
LATENT	metrical	categorical
metrical	factor analysis	factor analysis for categorical data latent trait analysis
categorical	profile analysis	latent class analysis

Bartholomew discusses these models from a new point of view, starting from a general model (5) that allows these techniques and some ones to emerge as special cases. We are concerned with metrical latent variables and categorical manifest variables.

Two different approaches are considered for the construction of the models in factor analysis for categorical data and latent trait analysis. In the first case, the ***Underlying Variable*** (UV) approach, where the categorical manifest variables are supposedly produced by underlying continuous variables, and in the second case, the ***Response Function*** (RF) approach with its origin in the theory of educational testing and developed further in Bartholomew (1980,1981). It starts

with a response function giving the probability of a positive response for an individual with variable value y.

In this paper we shall use the Response Function approach to latent trait analysis for binary response data, for which follows a review of the literature.

2- *Definition of the Model*

2.1- Notation and Assumptions

We shall consider the case when the manifest variables are binary and the latent variables are metrical. This situation happens, for example, in a survey where individuals are asked to answer questions by yes or no, agree or disagree, or in educational testing where the students may answer an item in a test right or wrong. Usually the two possible outcomes are coded as 1 (positive), otherwise 0(negative).

Thus if the test has p items and is answered by n individuals then the data matrix will be an(n x p) array of zeros and ones. We shall refer to any row of the data matrix as a **response or score pattern**, which is the set of responses of a given individual. Therefore there are 2^p different possible response patterns, which number increases quickly with p so that some patterns will probably not appear in the sample. For practical purposes in the samples we shall list only those response patterns which occur at least once.

Notation

Let $X=(X_1,X_2,\ldots,X_n)$ be a vector of p manifest variables, where X_i is equal to 1 or 0 for all i, and $Y=(Y_1,Y_2,\ldots,Y_p)'$ a vector of latent variables. Then the joint distribution of the X's is given by

$$f(x) = \int_{R_y} h(y)\ g(x|y)\ dy \qquad (1)$$

where

R_y the range space of y,

h(.) is the prior density of y,

g(.|y) is the conditional density of x given y.

Our main interest is what we can know about the latent vector Y after X has been observed. This comes from the conditional density

$$h(y|x) = h(y)\ g(x|y)\ /\ f(x),$$

which depends on our knowledge about h, g and f.

Obviously f(x) is the only density function about which inferences can be directly made, and therefore all the information we can get about g and h comes from knowledge of f. It follows that they are not uniquely determined (Bartholomew,1980). As we cannot obtain a complete specification of h(y|x), we need to make some restrictions on the class of functions to be considered.

The assumption of ***conditional independence***

$$g(x|y) = \prod_{i=1}^{p} g_i(x_i|y) \qquad (2)$$

is usually considered necessary for effective theoretical work with response functions. For it means that the association between X's (observed variables) is wholly explained by their dependence on the Y's (observed variables). Consequently, if Y is held fixed there will be no correlation between X's. This assumption cannot be tested empirically, since it is part at the definition of Y. We will come back to this point later on.

Conditional independence for g means that the set of latent variables is complete, i.e, Y is sufficient to explain the dependence between the X's.

As the X's are binary,

$$g_i(x_i|y) = [\pi_i(y)]^{x_i}\ [1-\pi_i(y)]^{1-x_i} \qquad i=1,\ldots,p \qquad (3)$$

where $\pi_i(y)=P[X_i=1|y]$ is called ***response function*** by Bartholomew (1980). In educational testing, where most of the models have been developed for a one-dimensional latent variable representing an ability of some kind, $\pi_i(y)$ is called ***item characteristic curve*** (ICC) or ***item response function*** (IRF).

Another assumption is that the response function is monotonic nondecreasing in the latent variable. This means that increasing any y, the probability of a positive response also increases or does not change.

Consequently from (2) and (3), the joint density function of the x's (1) can be written as

$$f(x) = \int_{R_y} h(y) \prod_{i=1}^{p} \left\{ [\pi_i(y)]^{x_i} [1-\pi_i(y)]^{1-x_i} \right\} dy. \qquad (4)$$

2.2- Response Function

Many suggestions about the shape of response functions and prior distribution of the latent variables have been made over the years. These have given rise to different models.

We shall present some of these models, starting from a general model (5) and deriving them as special cases.

The choice of a suitable response function was discussed by Bartholomew (1980), where he gave a set of properties that a family of response functions is desirable to possess. For instance, he says that the response function should be monotonic nondecreasing in the latent space, a property also implied by the normal ogive and logistic models, as we will see later. This implies, for example, that the probability of a correct response increases with ability (educational testing). Bartholomew also proposed a class of linear models with response functions satisfying:

$$G^{-1}(\pi_i(y)) = \alpha_{i,0} + \sum_{j=1}^{q} \alpha_{i,j} H^{-1}(y_j), \quad i=1,\ldots,p \qquad (5)$$

where

$\pi_i(y)$ is the response function,

y_j $(j=1,2,\ldots,q)$ are independently and uniformly distributed on (0,1) and functions G and H are distribution functions of random variables symmetrically distributed about zero.

In practice he limits the choice of G^{-1} and H^{-1} to the commonly used functions, the logit $\{\text{logit}(v)=\log[v/(1-v)]\}$ and the probit $[\text{probit}(v)=\Phi^{-1}(v)$, where Φ is the standard normal distribution$]$. Considering these functions, the following models can be derived:

Logit when both G^{-1} and H^{-1} are logit functions
Probit when both G^{-1} and H^{-1} are probit functions and
Logit/Probit when G^{-1} is a logit and H^{-1} is a probit function.

The logistic and normal distributions are very similar in shape, and the choice between them is without practical importance so that from one model we can obtain the approximately corresponding estimates of the parameters for the other since

$$\text{logit}(v) \cong \pi/\sqrt{3}\ \Phi^{-1}(v)$$

By transformation of $H^{-1}(y) = z$, Bartholomew (1987, Chapter 5) has proved that logit/probit model can be written in terms of normally distributed variables, as

$$\text{logit}[\pi_i(z)] = \alpha_{i,0} + \sum_{j=1}^{q} \alpha_{i,j}\ z_j \qquad i=1,\ldots,p \qquad (6)$$

Several response functions presented in the literature are special cases of the general model (5), as for example, the logistic and normal ogive models, which is shown below. For simplicity most of the results presented in this paper is done for one single latent variable.

2.2.1- Normal Ogive or Probit model

Lawley (1943) introduced a response function called the normal ogive model (Lord and Novick, 1968, p365) specified by

$$\pi_i(y) = \Phi(a_i\ (y - b_i)) \qquad i=1,\ldots,p$$

where

Φ is the cumulative distribution function of the standard normal distribution,

y, for $q=1$, is the latent ability parameter normally distributed with mean μ and variance σ^2, which characterizes the individuals and

a_i and b_i are parameters characterizing the item, called **discriminating power** and **difficulty** of item i.

Furthermore, it is assumed that $a_i \geq 0$, which means that $\pi_i(y)$ is a nondecreasing function of y.

The equivalence between the normal ogive model and the probit model, i.e,

$$\pi_i(y) = \Phi(\alpha_{i,0} + \alpha_{i,1} z),$$

may be seen by taking

$$b_i = -\alpha_{i,0}/\alpha_{i,1} \quad \text{and} \quad a_i = \alpha_{i,1}.$$

As the location and scale of the parameter estimates depend on the mean and variance of the distribution of the latent variable (or ability parameter), the equivalence between the parameter estimates is done by scaling. Bock and Aitkin (1981), for example, suggest that

$$\sum_{i=1}^{p} b_i = \sum_{i=1}^{p} \frac{-\alpha_{i,0}}{\alpha_{i,1}} = 0$$

and

$$\prod_{i=1}^{p} a_i = \prod_{i=1}^{p} \alpha_{i,1} = 1.$$

This model is the basis for numerous developments in psychometric theory, see for example, Lord (1952), Lord and Novick (1968), Bock and Lieberman (1970), Samejima (1974). Bock and Aitkin (1981) give also an extension for more than one latent variable for binary response.

2.2.2- Logistic or Logit/Probit Model

Two-parameter logistic model

Birnbaum (Lord and Novick (1968, Chapter 17) gave the two-parameter logistic model determined by assuming that the response function has the form of a logistic cumulative distribution function

$$\pi_i(y) = \frac{\exp[-da_i(y - b_i)]}{1 + \exp[-da_i(y - b_i)]}$$

where

y, is a latent ability parameter normally distributed with mean μ and variance σ^2,

a_i and b_i have the same meaning as in the normal ogive model,

d is a number that serves, at our convenience, as a unit scaling factor, with a value 1.7 corresponding to the maximum agreement between normal and logistic distributions.

The equivalence between the two-parameter logistic and the logit/probit model, i.e,

$$\pi_i(y) = \frac{\exp(\alpha_{i,0} + \alpha_{i,1} z)}{1 + \exp(\alpha_{i,0} + \alpha_{i,1} z)}$$

may be seen by taking

$$d = -1, \qquad b_i = -\alpha_{i,0}/\alpha_{i,1} \qquad \text{and} \qquad a_i = \alpha_{i,1}.$$

As for the normal ogive model, $a_i = \alpha_{i,1}$ and the possible different mean and variance for the normal distribution of Y is corrected by scaling.

Lord and Novick (1968b, Chapter 17) estimate the parameters a_i and b_i assuming that Y is N(0,1).

Rasch model

A random effect form of the model due to Rasch (1960), is a simplified form of the two-parameter logistic model with

$$\pi_i(y) = \frac{\exp(y - b_i)}{1 + \exp(y - b_i)} .$$

Here all the item discriminating powers are equal to 1, i.e, $a_i=1$, $i=1,\ldots,p$. Thus $\pi_i(y)$ depends only on the distance between the latent value y and the item difficulty b_i and as the value of b_i increases

fewer individuals will be likely to answer correctly or positively item i.

Therefore the Rasch model is equivalent to the logit/probit model when $\alpha_{i,1}=1$ and $b_i=-\alpha_{i,0}$ for all manifest variables or items i. The equivalence between the parameter estimates is obtained by scaling if a standard normal distribution is not assumed for Y.

The main advantage of this simplification is the existence of a sufficient statistic for Y, the total number of positive responses of an individual (or the total raw score of the data matrix).

There are many papers on the Rasch model and its extensions, among them Andersen (1970,1972,1973b), Gustafsson (1980a,1980b), Fischer (1981), Molenaar (1983), Thissen (1982) and many others to which we will give references later on.

Three-parameter logistic model

If in addition it is assumed that if an individual does not know the answer he will guess, and with probability c_i will guess positively then according to Lord and Novick (1968b, Chapter 17), the response function for the three-parameter logistic model is given by

$$\pi_i(y) = c_i + \frac{(1-c_i)\exp[da_i(y-b_i)]}{1+\exp[da_i(y-b_i)]}$$

Then the two-parameter logistic model is a particular case of this model, when $c_i=0$ for all i.

The three-parameter logistic model cannot be written in the general form (5), since that does not have guessing parameters.

This model has been applied by, for example, Lord (1968a,1983a), Hullin, Lissak and Drasgow (1982), Lord and Wingersky (1985), Thissen and Wainer (1982).

Since the normal ogive is equivalent to the probit model and the two-parameter logistic is equivalent to the logit/probit model, we shall use both names to refer to the same model, although we shall

generally use the notation following the general model(5) and consider Y as a latent variable.

2.2.3- Properties of the Response Function

Let us consider a logit/probit model, though the same approach is valid also for the logit and probit models.

The two most important properties which response functions produce are:

(1) The choice of which the two possible outcomes is to be regarded as positive is totally arbitrary. If the positive answer has probability $\pi_i(z)$ then the negative has probability $1-\pi_i(z)$, and they are given by

$$\pi_i(z) = \frac{\exp\left(\alpha_{i,0} + \sum_{j=1}^{q} \alpha_{i,j}\, z_j\right)}{1 + \exp\left(\alpha_{i,0} + \sum_{j=1}^{q} \alpha_{i,j}\, z_j\right)}, \quad \text{i.e.,}$$

$$\pi_i(z) = \left\{ 1 + \exp\left(-\alpha_{i,0} - \sum_{j=1}^{q} \alpha_{i,j}\, z_j\right)\right\}^{-1}$$

and

$$1-\pi_i(z) = \left\{ 1 + \exp\left[\alpha_{i,0} + \sum_{j=1}^{q} \alpha_{i,j}\, z_j\right]\right\}^{-1}$$

This means that increasing any latent value z, increases the probability of a positive response and decreases, as expected, the probability of a negative response by the same amount. Thus, when $q=1$, it is possible to obtain all $\alpha_{i,1}$'s positive or zero by suitable choice of which outcome is to be considered as positive.

(2) The direction in which most latent variables are measured is arbitrary. Changing the direction of measurement involves replacing z_j by $-z_j$ in equation (6). This is equivalent to changing the sign of the corresponding $\alpha_{i,j}$ without changing the model.

2.3- Interpretation of the Parameters

The parameters of the logit/probit model may be interpreted in several ways.

The coefficient $\alpha_{i,0}$ is the value of logit $\pi_i(z)$ at z=0 and thus π_i is the probability of a positive response from a **median** individual. In the context of educational testing, $\alpha_{i,0}$ or π_i would be called the item difficulty.

The coefficients $\alpha_{i,j}$ may be interpreted in three related ways.

First, as a measure of the extent to which Z_j discriminates between individuals. For two individuals a given distance apart on the Z_j-scale, the bigger the absolute value of $\alpha_{i,j}$ the greater the difference in their probabilities of given a positive response to item i and thus easier to discriminate between them in relation to item i. Therefore $\alpha_{i,j}$ is a parameter that indicates the value of an item in the sense of the amount of information that the item provides about Z_j. In educational testing, this is the interpretation usually adopted, and $\alpha_{i,j}$ is called item discriminating power. Figure 1 illustrates the effect of differing values of $\alpha_{i,0}$ and $\alpha_{i,1}$ on the shape of the response function.

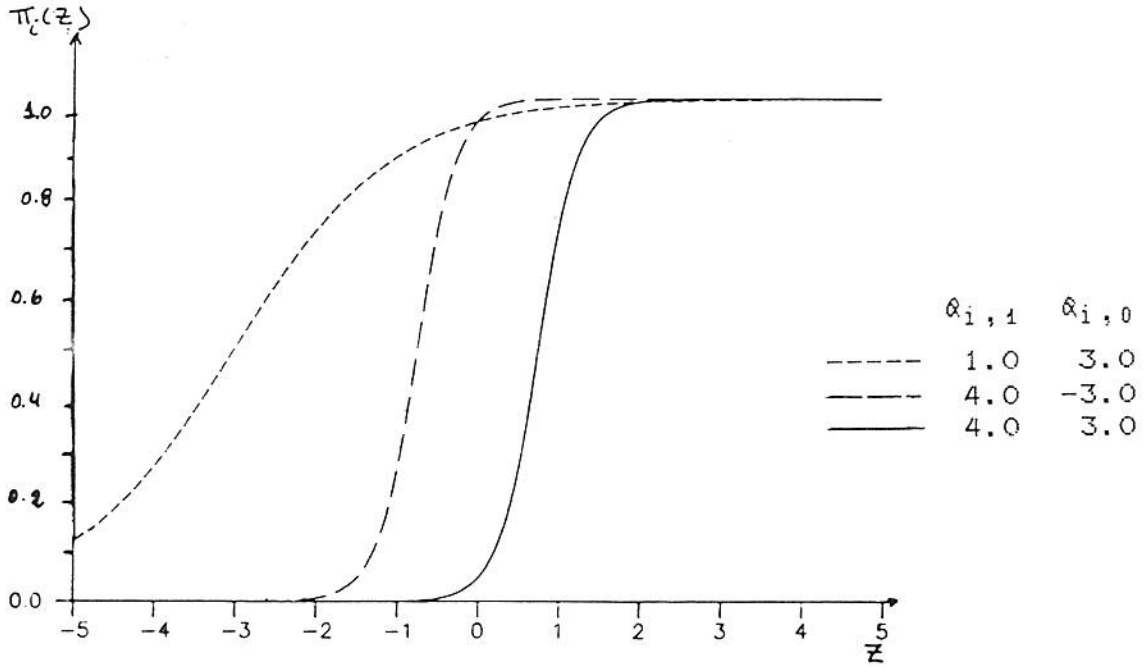

Figure 1 - The response function $\pi(z)$.

A second interpretation of the $\alpha_{i,j}$ is by analogy with linear factor analysis or principal components, where the $\alpha_{i,j}$'s are equivalent to the loadings. They are the weights of the x_i's in the determination of the component scores X_j's, i.e, $X_j = \sum_{i=1}^{p} \alpha_{i,j}\, x_i$.

Finally the $\alpha_{i,j}$ are related to correspondence analysis, where they are equivalent to the category scores. This is done by attributing the value of $\alpha_{i,j}$ to a positive response on manifest variable j and zero to a negative response. Then for each latent variable Z_j the data matrix constituted by 1 and 0 is replaced by $\alpha_{i,j}$ and 0. The individual score is thus the sum of the category scores for that latent variable Z_j.

3- *Maximum Likelihood Estimation*

3.1- **Estimation Procedures: joint, conditional and marginal likelihood functions**

In the literature we have found that the parameter of latent variable models for binary data are estimated essentially through 3 different procedures: joint maximum likelihood (ML), conditional ML and marginal ML.

As we have already pointed out when describing different shapes of response functions for the general latent model(5), Y is usually defined in the literature as a person parameter rather than a latent variable, as used in the context of this paper. However we shall refer to Y as a person or ability parameter, if necessary, when reporting research using that approach.

Joint Maximum Likelihood

A joint maximum likelihood estimation was proposed by Birnbaum (1968) for the two- and three-parameter logistic model, and for the Rasch model by Wright and Panchapakesan (1969), among others. In this approach, person abilities and item parameters (discrimination and difficulty) are estimated simultaneously so that the procedure is not conditioned on the ability parameters.

The joint ML estimation of the person and item parameters is not generally possible because the number of parameters increases with the sample size and thus standard limit theorems do not apply. Several researchers, including Wood, Wingerkly and Lord (1976) have avoided this problem by assuming that respondents who have the same score pattern, or same number of positive responses or who have been assigned provisionally to homogeneous ability groups, have the same ability. On this assumption, the number of parameters is finite and standard asymptotic theorems apply.

The assumption that abilities are fixed in size, when in fact they are not identifiable and have a distribution in the population of persons, is difficult to justify statistically. A better approach to estimation in the presence of a random nuisance parameter (person ability) is that of integrating over the parameter distribution and estimating the item parameters by maximum likelihood in the marginal distribution, which is done when using marginal ML procedure.

Baker (1988) reviews the ML estimation procedures for the one-, two- and three-logistic models.

Conditional Maximum Likelihood (CML)

Rasch (1960) showed that under his probabilistic model the 'item totals' (number of positive responses given by every person) and the 'row scores' (number of positive responses given to every item) are sufficient statistics for the person and difficulty parameters. Using Rasch results as a starting point, Andersen (1970, 1972, 1973a) developed a conditional ML procedure to estimate the difficulty parameters that did not involve the latent individual parameters. The difficulty parameter estimates are obtained from the likelihood function conditioned upon the item total scores.

Wright and Douglas (1977) have shown that the conditional ML estimation is inaccurate when a test has more than 10 or 15 items due to round-off-error. They proposed a simplified alternative procedure for conditional estimation, which is limited to 20 or 30 items due to the same precision problem, especially in the presence of extreme difficulty parameter estimates.

In order to compare the joint ML and conditional ML for tests with more than 20 items, Wright and Douglas carried out a simulation study based on 15 replications of 500 individuals each for tests of size 20 and 40. They assumed that the ability was normally distributed with mean 0, 1 and 2, and the difficulty parameters were generated from a normal distribution with mean zero. The comparison between both procedures was done in terms of MAX DIFF (maximum difference between a generated difficulty parameter and the mean over the 15 replications of its estimates), RMS (root mean square of these differences over items) and the MEAN ABS (mean of the absolute value of these differences over items). They found out that in terms of RMS and MEAN ABS both estimation procedures, conditional ML and joint ML, give approximately the same results, while the MAX DIFF's tend to increase for both algorithms, but strongly for conditional, when the mean of the sample shifts away from zero (equal 1 or 2). This later result was found to be due to the increasing discrepancy between item and sample characteristics, which made estimation difficult for the conditional ML because of accumulated round-off-error.

Marginal Maximum Likelihood (MML)

Thissen (1982) developed marginal ML procedures for the Rasch model making use of the fact that all response patterns which have the same number of positive responses have proportional likelihoods for the single latent variable. Unlike the conditional solution (CML), this estimation procedure is not conditional on the sufficient statistic for the person parameter and requires specification of the prior distribution for the person parameters.

The formulation of the model explicitly includes the item discriminating power common to all items and it is assumed that the latent ability is distributed as N(0,1).

Two algorithms have been described by Thissen (1982) for MML estimation:

(a) A gradient solution, following Bock and Lieberman (1970), where the parameters are estimated by maximum likelihood and

(b) An alternative solution, following the algorithm described by Bock and Aitkin (1981), uses Gauss-Hermite quadrature points for the N(0,1) prior distribution for latent ability (person parameters).

They also show that the MML procedure is similar to a combination of CML of the item parameters with estimation of the mean and variance of the population distribution as described by Andersen and Madsen (1977). The mean of the item difficulty parameters is essentially equivalent to Andersen and Madsen's population mean and the estimated discrimination parameter is the same as the standard deviation of the population distribution (normal) for conventionally standardized CML estimates. For this procedure the population distribution is not required to be normal, but must have finite mean and variance.

Tsutakawa (1984) derived a MML procedure employing the two-parameter logistic model. His method differs from Bock and Aitkin's method in the manner in which the prior distribution of the latent variable is handled, but it is equivalent for the special case of a discrete empirical prior. He analysed a 50-item arthritis knowledge test administered to 162 individuals, using both the joint and the marginal ML procedures. After appropriate scaling to take metric differences into account, the values of the discrimination and difficulty parameters yielded by the two methods were very similar.

Tsutakawa also used simulated data to evaluate the parameter recovery capability of the two procedures. This investigation involved two hundred simulated respondents having a unit normal distribution and a 50-item test with representative values of the item parameters. The estimated item parameters were plotted against tha underlying parameter values. The plots showed a close agreement between the two methods as well as a general 45^0 line relating the estimates and the parameters. The scatter of the item discriminating power about the line was much greater than of the item difficulty.

It follows a description of the ML procedure used in this paper to estimate the parameters of the general model(5).

3.2- **Marginal Maximum Likelihood Estimation**

For any model of the family (5) the joint probability function of $x_1,\ldots,x_p$ is

$$f(x) = \int_{Ry} \prod_{i=1}^{p} [\pi_i(y)]^{x_i} [1-\pi_i(y)]^{1-x_i} h(y)dy \qquad (7)$$

If x_s is the observed response vector for the s^{th} sample member then the loglikelihood is

$$L = \sum_{s=1}^{n} \log f(x_s) \qquad (8)$$

Bock and Lieberman (1970) maximised this function with the normal ogive model for the response function and for one latent variable distributed as N(0,1), i.e, using the probit model. The likelihood equations were solved iteratively by a Newton-Raphson method and Gauss-Hermite quadrature was employed to perform the necessary integrations. Due to the heavy numerical integration the method was considered to be limited to one latent variable and not more than 10 manifest variables.

Bock and Aitkin (1981) by a simple transformation of the Bock and Lieberman (1970) likelihood equations, found a computational solution so that the method could be applied for more than one latent variable and a large number of manifest variables. This reformulation is related to the E-M algorithm for maximum likelihood estimation as discussed by Dempster, Laird and Rubin (1977).

We shall give the main results of this method as described by Bartholomew (1987, Chapter 6).

3.2.1- An E-M Algorithm

We shall consider the logit/probit model for one latent variable expressed by Z as defined in (6), i.e,

$$\text{logit}(\pi_i(z)) = \alpha_{i,0} + \alpha_{i,1}\, z$$

Each iteration of this E-M algorithm involves two steps called the expectation step (E) and the maximization step (M) and the method starts with arbitrary values for the parameters.

E-step: Using the current values for $\{\alpha_{i0}\}$ and $\{\alpha_{i1}\}$, predict z_s for s=1,2,...,n, through

$$E(Z_s|x_s) = \int_{-\infty}^{\infty} z_s \prod_{i=1}^{p} [\pi_i(z_s)]^{x_i} [1-\pi_i(z_s)]^{1-x_i} h(z_s)\, dz_s \;/\; f(x)$$

The value of $E(Z_s|x_s)$ has to be found by numerical integration.

M-step: Treating the expected values $E(Z_s|x_s)$, s=1,2,...,n, as if they were true values z_s, estimate the parameters $\{\alpha_{i0}\}$ and $\{\alpha_{i1}\}$ by maximum likelihood, as follows:

Let the conditional loglikelihood defined by

$$L = \sum_{s=1}^{n} \sum_{i=1}^{p} \left\{ x_{is} [\log \pi_i(z_s)] + (1-x_{is}) [\log(1-\pi_i(z_s)] \right\}$$

$$= \sum_{s=1}^{n} \sum_{i=1}^{p} \left\{ x_{is}\, \text{logit}[\pi_i(z_s)] + \log(1-\pi_i(z_s)) \right\}, \qquad (9)$$

where $\text{logit}\ \pi_i(z_s) = \alpha_{i,0} + \alpha_{i,1}\, z_s$.

Then the partial derivatives with respect to the parameters $\{\alpha_{i,0}\}$ and $\{\alpha_{i,1}\}$ are

$$\frac{\partial L}{\partial \alpha_{i,0}} = \sum_{s=1}^{n} [\, x_{is} - \pi_i(z_s)\,] \qquad (10)$$

$$\frac{\partial L}{\partial \alpha_{i,1}} = \sum_{s=1}^{n} z_s \left[x_{is} - \pi_i(z_s) \right] \qquad \text{for } i=1,2,\ldots,p.$$

Thus estimating equations are obtained setting (10) equal to zero and for each variable i there is a pair of non-linear equations which can be solved for $\alpha_{i,0}$ and $\alpha_{i,1}$. Methods of solving these equations are reviewed by McFadden (1982).

Having completed the M-step, the E-step is done again, and the cycle is repeated until the estimates become stable, according to some criterion.

Bock and Aitkin (1981) reported that the convergence of the E-M algorithm is only geometric and slows up as the solution point is approached. They suggested using the acceleration technique of Ramsey (1975) to speed convergence.

The convergence properties of the E-M algorithm has been studied analytically by Wu (1983). He showed that if the likelihood function is unimodal and certain differentiability conditions are satisfied, any E-M sequence converges to the unique ML estimates of the parameters.

More than one latent variable

If there is more than one latent variable, the term $\alpha_{i,1}\, z_s$ in (9) is replaced by $\sum_{j=1}^{q} \alpha_{i,j}\, z_{j,s}$; q equations replace the second member of (10)– one for each $\{\alpha_{i,j}\}$ – and $\pi_i(z_s)$ becomes $\pi_i(z)$. In this case, the determination of $\hat{\alpha}_{i,0}$ and $\hat{\alpha}_{i,j}$, for $j=1,2,\ldots,q$, involves the solution of q+1 simultaneous non-linear equations for each i.

In order to get unique solutions, when q>1, we must impose some constraints. One possibility is to fix the values of enough α's to ensure a unique solution. For example, it is sufficient to fix $\alpha_{i,1}=0$ for some i, when q=2.

3.2.2- A Variation of the E-M Algorithm

A variation of the E-M algorithm was proposed by Bock and Aitkin (1981) also for the probit model. Bartholomew (1987, Chapter 6) discusses the same variation from a rather different perspective setting G^{-1} in (5) as the logit function. It follows the main results for one latent variable.

Even though the latent variable Z is distributed as N(0,1), it is proposed as an approximation that Z assumes values $z_1, z_2, \ldots, z_k$ with probabilities $h(z_1), h(z_2), \ldots, h(z_k)$ chosen so that the joint probability function

$$f(x_s) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(x_s|z)\, h(z)\, dz \qquad s=1,2,\ldots,n$$

can be approximated with high accuracy by Gauss-Hermite quadrature, i.e,

$$f(x_s) = \sum_{t=1}^{k} g(x_s|z_t)\, h(z_t) \qquad \text{for } s=1,2,\ldots,n \tag{11}$$

where z_t is a tabled quadrature point (node) and $h(z_t)$ is the corresponding weight (see Straud and Sechrest,1966).

The quadrature weights, $h(z_t)$, are approximately the normalized, i.e., $\sum_{t=1}^{k} h(z_t)=1$, values of the probability density of a N(0,1) random variable at the points z_t, which are choosen to best approximate the marginal probability function $f(x_s)$. This approximation becomes more accurate as the number of quadrature points increases.

From the maximization of

$$L = \sum_{s=1}^{n} \log f(x_s)$$

we obtain, for v=0,1

$$\frac{\partial L}{\partial \alpha_{i,v}} = \sum_{t=1}^{k} \frac{\partial \pi_i(z_t)}{\partial \alpha_{i,v}} \; \frac{[\, R_{i,t} - N_t\, \pi_i(z_t)\,]}{\pi_i(z_t)\,[\, 1-\pi_i(z_t)\,]} \tag{12}$$

where

$$R_{it} = \sum_{s=1}^{n} x_{is}\, h(z_t|x_s) \tag{13}$$

$$N_t = \sum_{s=1}^{n} h(z_t|x_s) \tag{14}$$

and $h(z_t|x_s)$ is the posterior probability of Z_t given x_s.

Before defining an E-M algorithm in this approach, it is useful look at the meaning of N_t and R_{it}. As the quantity $h(z_t|x)$ is the probability that an individual with response vector x is located at z_t, N is the expected number of individuals at z_t. By analogy, R_{it} is the expected number of positive responses to item i among those individuals at z_t.

Consequently if we know the allocation of each individual on the Z-scale then N_t is the number of individuals at z_t and R_{it} is the observed frequency of positive response at z_t.

The estimation of the parameters is performed by choosing any starting values for $\{\alpha_{i,0}\}$ and $\{\alpha_{i,1}\}$ followed by repeated applications of (12), (13) and (14) over the set of items, using an E-M algorithm defined as

E-step: Calculate the values of R_{it} and N_t using equations (13) and (14).

M-step: Obtain improved estimates of the $\{\alpha_{i,0}\}$ and $\{\alpha_{i,1}\}$ solving equation (12), using the values of R_{it} and N_t from the E-step.

The E-M cycles are continued until convergence is obtained. In this case the number of values that the latent variable assumes is fixed and the set of values constitues the distribution of Z.

If we use the logit/probit model for $\pi_i(z)$ then

$$\pi_i(z) = \left[\ 1 + \exp(-\alpha_{i,0} - \alpha_{i,1}\, z)\ \right]^{-1}$$

and

$$\frac{\partial\, \pi_i(z)}{\partial\, \alpha_{i,v}} = z^v\, \pi_i(z)\, [\, 1-\pi_i(z)\,] \qquad (15)$$

for v=0,1.

Substituting (12) and (15) the equations become

$$\sum_{t=1}^{k} z^{v-1}\, [\, x_{is} - \pi_i(z_t)\,]\, h(z_t|x_s) = 0 \qquad (16)$$

for v=0,1 and i=1,2,...,p, which may be compared with (10).

Even though we have presented these methods for a response function, in which G^{-1} in equation (5) was the logit function and the prior distribution of the single latent variable was approximated using Gauss-Hermite quadrature points, it may be applied for any response function and any discrete prior distribution.

If some other prior distribution of the latent variable is assumed, other points may be chosen and a normalized density point t substituted for $h(z_t)$ in (11). For example, if a rectangular prior is assumed then k points may be set at equal intervals over an appropriate range and the quadrature weight set $\{h(z_t)\}$ equal to 1/k.

Bock and Aitkin (1981) have considered besides a prior standard normal distribution, a rectangular and an empirical distribution for the single latent variable and taken k=10 (see their paper for more details). Working through the data of Section 6 and 7 for the Law School Aptitude Test (LSAT) presented in Bock and Lieberman (1970), they have obtained practically the same estimates of the parameters from these three different prior distributions. They also suggested that adequate solutions could be obtained with even smaller k, for example, k=3,5 or 7 and this would make it feasible to generalize the method to several latent variables. On the other hand, investigations made by Shea (1984) show that at least k=20 may be necessary to obtain reasonable accuracy and this puts much greater demands on computing resources.

The general model (5) as defined by Bartholomew (1980) involves an arbitrary assumption about the form of the prior distribution of the latent variable. Although the form of the analysis does not depend on this assumption, as shown by Bartholomew (1984), it does affect the estimation of the parameters. Therefore it is important to know whether the values of the estimates are sensitive to the choice of the prior distribution.

Bartholomew (1988) answered this question mainly through numerical evidence that the choice of the prior has negligible effect on the expected first- and second-marginal proportions. He concludes that the estimates are not sensitive to the choice of the prior based upon some results reported in Bartholomew (1980,1987) that the ML estimates which depend on margins of all order are usually very close to those depending only on the first-and-second order margins.

4- Behaviour of the Likelihood Function

Since the parameters of the latent variable models under investigation are usually investigated by the method of maximum likelihood (ML), it is importante to check if the behaviour of the likelihood function is suitable for the method. This means to check whether the likelihood has a smooth unimodal shape, or whether it has multiple relative maxima.

Albanese(1990, Chapter 2) investigates the behaviour of the likelihood by profiling and an approximate method. Particular attention is given to the shape of the likelihood around the maximum point in order to see whether the information matrix will a good guide to the variability of the estimates. The main results are:

(1)- The investigation of the behaviour of the likelihood function using an approximate method provides results equivalent to the profile method. Both suggest that large $\hat{\alpha}_{i,1}$ ($\geq 3/\sigma$, σ is the standard deviation of the latent distribution) probably indicates bad behaviour of the likelihood, which will be shown by the presence of a long ridge. In this case the second derivative matrix or the information matrix are not good guides to the variability of these estimates.

(2)- If $\hat{\alpha}_{i,1}$ is not large, the likelihood funtion behaves well and thus the first order asymptotic theory is appropriate.

(3)- A badly behaved likelihood function suggests either that a reparametrization is necessary, or that the model is a poor fit for the data, or that the inference is particulary difficult.

(4)- Among the several reparametrizations we tried only the one given by

$$\hat{\alpha}^{*}_{i,0} = \hat{\alpha}_{i,0} / (1 + \hat{\alpha}^{2}_{i,1})^{\frac{1}{2}}$$

provided a better behaviour of the likelihood, independent of the size of the parameter estimates.

This reparametrization corresponds to the probit of the expected value of the response function of a probit model, that is,

$$\hat{\alpha}^{*}_{i,0} = \Phi^{-1}(E(\alpha_{i,0} + \alpha_{i,1} z)) =$$

$$= \Phi^{-1}(E(P(X_i=1|z)) = \Phi^{-1} (P(X_i=1))$$

5- *Sampling Variation of the Maximum Likelihood Estimators*

There has been little discussion about the magnitudes of the standard deviations of estimated parameters for commonly used latent variable models. This may be due to the fact that no simple closed formulae exist for the standard deviations as a function of the sample size and the parameters. The usual way to estimate the standard deviations for maximum likelihood estimates of the parameters is to compute the asymptotic variance-covariance matrix, using the elements of the inverse of the information matrix evaluated at the solution point. Thus if we have a set of parameters α then

$$[D(\hat{\alpha})] = E \left[\frac{-\partial^2 L}{\partial\alpha_i \, \partial\alpha_j} \right]^{-1} \Bigg|_{\alpha=\hat{\alpha}} \qquad (17)$$

where

$$\frac{\partial^2 L}{\partial\alpha_i\,\partial\alpha_j} = \left[\sum_{s=1}^{n} \frac{1}{f_s} \frac{\partial^2 f_s}{\partial\alpha_i\,\partial\alpha_j} - \frac{1}{f_s^2} \frac{\partial f_s}{\partial\alpha_i} \frac{\partial f_s}{\partial\alpha_j} \right] \Bigg|_{\alpha=\hat{\alpha}}$$

and $f_s \equiv f(x_s)$. On taking the expectation, the first term vanishes leaving

$$[D(\hat{\alpha})]^{-1} = n\,E \left[\frac{1}{f_s^2} \frac{\partial f_s}{\partial\alpha_i} \frac{\partial f_s}{\partial\alpha_j} \right] \Bigg|_{\alpha=\hat{\alpha}} \qquad (18)$$

In our case X is a response pattern taking 2^P different values and the expectation in (18) is thus

$$\sum_{s=1}^{n} \frac{1}{f^2(x_s)} \frac{\partial f(x_s)}{\partial\alpha_i} \frac{\partial f(x_s)}{\partial\alpha_j} \qquad (19)$$

If p is small it is feasible to evaluate this sum for all i and j and then to invert the resulting matrix. However if p is large some probabilities will become very small so that the computation of $1/f(x_s)$ will cause overflow on most of computers. In this case an approximation may be used, replacing the expectation of the information matrix by its observed value. This requires the computation of (17) and the inversion of the resulting matrix. Since the first term has expectation zero a further approximation may be obtained from

$$D^*(\hat{\alpha}) = \left[\sum_{s=1}^{n} \frac{1}{f^2(x_s)} \frac{\partial f(x_s)}{\partial\alpha_i} \frac{\partial f(x_s)}{\partial\alpha_j} \right]^{-1} \qquad (20)$$

The number of distinct terms in the sum of (20) will usually be less than n since more than one individual may have the same response pattern.

Louis (1982) developed a technique for computing the observed information matrix when the E-M algorithm is used to find the maximum likelihood estimates in incomplete data problems. It requires computation of the complete-data gradient and second derivative matrix which can be implemented quite simply in the E-M iterations. This procedure can be applied to obtain the asymptotic variance-covariance

matrix in latent variable models, since they involve observable (manifest) variables and not directly observable (latent) variables which corresponds to a case of incomplete data, as defined by Dempster, Laird and Rubin (1977).

Thissen and Wainer (1982) investigated the asymptotic standard errors of the item parameters for the one-, two-,and three- parameter models under the assumption that the latent value of the respondents were known and normally distributed with zero mean and unit variance. Tables of the minimum asymptotic standard errors were reported for combinations of parameter values under the three models.

An interesting set of results was given by the two-parameter and the three-parameter model with c=0 (guessing parameter). Even though the numerical values of the difficulty and the discrimination parameters would be the same, the information matrices are not. The three-parameter matrix still has a row and column corresponding to the guessing parameter. When one item was easy and had low discriminating power, the standard errors under the two-parameter model were roughly 0.09 of those reported for the three-parameter model. Clearly, the two and the three-parameter model with c=0 are not the same with respect to the standard errors of the item parameter estimates. The asymptotic standard errors for the item difficulty under the Rasch model were consistently smaller than those obtained for the other two models. In particular, the increase in standard error with departure of item difficulty from zero was much less pronounced.

Based upon the results, Thissen and Wainer (1982) suggested that when working with logistic response models we should try to fit the simplest model first, and only if it is found to be inadequate move to the more complex ones.

Lord and Wingersky (1983) have developed a method for computing the asymptotic variance-covariance matrix for the three-parameter logistic model, considering the unidimensional latent variable as a person parameter. The derivation assumes that both item and person parameters are unknown. They demonstrate that the size of the error variances are affected strongly by the restrictions introduced in order to fix the latent scale. One disadvantage of this method is that the information matrix to be inverted is very large.

On the other hand, Gruijter (1985) has shown that the method can be simplified for the Rasch model when we are only interested in the item parameters. This is done under a suitable restriction on the difficulty parameters, as for example, assuming that the

mean b is equal to zero, i.e, $b_p = -\sum_{i=1}^{p-1} b_i$, or setting $b_p=0$. Although the variance-covariance matrix for the item parameters can be obtained without difficulty, it depends on the restrictions. He points out that the first restriction seems to be preferable to the others due to its simplicity and relative accuracy of the mean.

6- *Adequacy of the Asymptotic Variance-Covariance Matrix*

When interpreting the asymptotic variance-covariance matrix of the parameter estimates it is assumed that the model is appropriate for the data. Since this assumption may be false in practice, or the sample size is not large enough for the number of parameters which have been estimated or even the standard asymptotic theory does not apply, the standard deviation and covariances obtained asymptotically will probably represent lower limits for the actual ones, and they must be analysed carefully.

Another way to look at the variance-covariance matrix that may give some idea about how the asymptotic theory is working, is through jackknife and bootstrap techniques.

9.1- Jackknife

Jackknifing is a statistical technique first proposed by Quenouille (1956), which is used for reducing bias in the estimation of parameters and for estimating the variance-covariance matrix of the estimates. Miller (1974) gives an review of the subject.

In the basic jackknife the observations are randomly divided into g groups of size h each. We shall consider the number of groups equal to the sample size n and thus each group of size one.

Let $X_1, X_2, \ldots, X_p$ be a sample of independent and identically distributed(iid) random variables and $\hat{\alpha}$ be an estimator of the parameter vector α based on the sample size n.

Let $\hat{\alpha}_{-i}$ be the corresponding estimator based on the sample of size n-1, where the i^{th} group (observation) has been deleted.

Then jackknife pseudovalues are defined by

$$\tilde{\alpha}_i = n\,\hat{\alpha} - (n-1)\,\hat{\alpha}_{-i}$$

for i=1,2,...,n.

The jackknife estimates $\tilde{\alpha}$ and its estimated variance-covariance matrix are obtained from the n pseudovalues by treating them as independently identically distributed observations from a multivariate normal distribution (Tukey,1958). These estimates are given by

$$\tilde{\alpha} = \frac{\sum \tilde{\alpha}_i}{n} \tag{21}$$

$$\tilde{\alpha} = g\,\hat{\alpha} - \frac{n-1}{n} \sum \hat{\alpha}_{-i}$$

and

$$\sum(\tilde{\alpha}) = \frac{\sum (\tilde{\alpha}_i - \tilde{\alpha})(\tilde{\alpha}_i - \tilde{\alpha})^t}{n(n-1)} \tag{22}$$

Since it often happens that $\tilde{\alpha}$ and $\hat{\alpha}$ are asymptotically equivalent, $\sum(\tilde{\alpha})$ is sometimes used to estimate the variance-covariance matrix of $\hat{\alpha}$.

The jackknife estimate of bias is the difference between the parameter estimate $\hat{\alpha}$ and $\tilde{\alpha}$ multiplied by the correction factor n/(n-1), i.e.,

$$\text{bias} = \frac{n}{n-1}(\hat{\alpha} - \tilde{\alpha}) \tag{23}$$

The jackknife technique has been applied in many areas, including factor analysis. Pennell (1972) demonstrated how the method can be used to find confidence intervals for the factor loadings, while Clarkson (1979) discussed the results of simulation studies using jackknife techniques and proposed modifications.

Clarkson's studies do not include the jackknife samples which provide Heywood cases. He argue that in these cases the jackknife estimates of the factor loadings are not representative of the 'usual' jackknife results because they are too large in absolute value.

Jorgensen (1987) gave a modification of the jackknife method for estimating the dispersion of the parameter estimates that are obtained as limits of iterative processes. He also gave examples to show how the method can be applied to the E-M algorithm and to iteratively reweighted least-squares.

Albanese(1990, Chapter 3) applies the jackknife technique to investigate the adequacy of the variance-covariance matrix for the logit/probit model for binary response data. The main results will be given in the end of this section.

9.2- **Bootstrap**

The bootstrap is a general resampling procedure introduced by Efron (1979) to estimate the distribution of statistics based on independent observations. It can be carried out non-parametrically and parametrically, depending on the distribution from which the bootstrap samples are drawn.

We shall first present the non-parametric or empirical bootstrap method.

Suppose $X_1, X_2, \ldots, X_p$ are independent and identically distributed(iid) random variables from a population with unknown distribution function F, and suppose the goal is to make inferences about the parameter vector α of the population.

Let $\hat{\alpha}(x_1, x_2, \ldots, x_p)$ be an estimator of α based on the sample size n and let $\hat{F}$ be the empirical distribution, that is, the distribution function that assign mass 1/n to each X_i.

The bootstrap approximates the sampling distribution of α under F by the sampling distribution of $\hat{\alpha}$ under $\hat{F}$. This procedure is carried out using Monte Carlo method as follow:

(1) Construct $\hat{F}$

(2) Draw a bootstrap sample, $X^*_1, X^*_2, \ldots, X^*_p$ iid with cdf $\hat{F}$ and calculate

$$\hat{\alpha}^* = \hat{\alpha}(X^*_1, X^*_2, \ldots, X^*_p)$$

(3) Independently do B times the step 2 (for some large B), obtaining

$$\hat{\alpha}^*_b \ , \ b=1,2,\ldots,B.$$

The distribution function of $\hat{\alpha}$ is approximated by

$$\hat{F}_B(y) = \#(\hat{\alpha}^*_b \leq y) \ / \ B.$$

The bootstrap estimate of α based on the B replications is the mean of the $\hat{\alpha}^*_b$ estimates, i.e,

$$\hat{\alpha}^*_\cdot = \sum \hat{\alpha}^*_b \ / \ B \qquad (24)$$

and the bootstrap variance-covariance matrix estimate of α based on the B replications is the variance-covariance matrix of the $\hat{\alpha}^*_b$ estimates, i.e,

$$\Sigma_B = (B-1)^{-1} \sum (\hat{\alpha}^*_b - \hat{\alpha}^*_\cdot) (\hat{\alpha}^*_b - \hat{\alpha}^*_\cdot)^t \qquad (25)$$

As the number of replications $B \to \infty$, $\hat{\alpha}^*_\cdot$ will approach the bootstrap estimate of α and Σ_B the corresponding bootstrap estimate of the variance-covariance matrix Σ.

The bootstrap estimate of bias is the difference between the parameter estimate $\hat{\alpha}$ and the bootstrap estimate $\hat{\alpha}^*_\cdot$, that is,

$$\text{bias} = \hat{\alpha} - \hat{\alpha}^*_\cdot \qquad (26)$$

The basic result of the bootstrap theory is that the empirical distributions of the parameter estimates obtained by this method are asymptotically the same as the sampling distribution of those parameters in sampling from the population from which the original sample was drawn.

There is nothing which says that the bootstrap must be carried out non-parametrically. If we have reason to believe that the true distribution F is Normal, for example, then we can estimate F by its parametric ML estimate $\hat{F}$. The bootstrap samples at step (1) of the algorithm could them be drawn from $\hat{F}_{normal}$ instead of $\hat{F}$ (empirical distribution) and steps (2) and (3) carried out as before.

Efron (1979) also suggests that Taylor series expansion method can be used to obtain the approximate mean and variance of the bootstrap distribution of $\hat{\alpha}^*_{\cdot}$, and he shows that it turns out to be the same as Jaeckel's infinitesimal jackknife (Miller,1974), which differ only in detail from the standard jackknife described before.

Efron and Tibshirani (1986) discuss the number of replications B necessary to give reasonable results when we are estimating the standard deviation of one parameter. They set out the following approximation

$$CV(\hat{\sigma}_B) = \{ CV(\hat{\sigma})^2 + [(E(\hat{\delta}) + 2)/4B] \}^{\frac{1}{2}}$$

where $CV(\hat{\sigma})$ is the limiting coefficient of variation of σ as $B \to \infty$, $\hat{\delta}$ is the kurtosis of the bootstrap distribution of $\hat{\alpha}^*$, given the observed data $\mathbf{x}=(x_1,x_2,\ldots,x_n)$, and $E(\hat{\delta})$ its expected value average over $\mathbf{x}$. For typical situations, $CV(\hat{\sigma})$ lies between 0.10 and 0.30.

From this approximation and assuming that $E(\hat{\delta})=0$, they point out that for values of $CV(\hat{\sigma}) \geqslant 0.10$, there is little improvement when B is bigger than 100. They afirm that B as small as 25 gives reasonable results. We suggest that the number of bootstrap samples be determined by the point where stability of the estimates is obtained.

Efron (1984) discusses different kinds of confidence intervals using the bootstrap and he shows that it is necessary to have at least 1000 samples to compute the BC (bias corrected percentile interval) as

defined in the same paper, and BC_α intervals while for the simplest method, percentile interval, 250 replications can give useful results.

The percentile interval is obtained by taking $\alpha \in (\hat{F}^{-1}(\alpha'), \hat{F}^{-1}(1-\alpha'))$ as an approximate $1-2\alpha'$ central interval for α.

Confidence intervals are a fundamentally more ambitious measure of statistical accuracy than standard errors, so it is not surprising that they require more computational effort.

Chatterjee (1984) gives an application of the non-parametric bootstrap method to the problem of estimating the variability of the estimates of factor loadings. The number of bootstrap samples was settled empirically; it appeared that 300 gave reasonable stability. Combining the bootstrap with graphical techniques he examines the variability of the estimator of the factor loadings. He points out that bootstrap may very well reveal when the asymptotic results are poor approximations.

Grönroos (1985) applies bootstrap methods to confirmatory factor analysis of a LISREL submodel (Jöreskog and Sörbom, 1984) to estimate factor loadings and their standard deviations.

His simulation studies involve artificial data with sample size 100, 150 and 300 and initially 300 replications. However the number of bootstrap samples become smaller, since he deletes from the analysis all those which provide the occurrence of Heywood cases.

Comparing asymptotic theory with bootstrap and Normal bootstrap results, he points out that the difference between the two bootstrap methods is very small, but it is larger, even though not essential significant, when compared with the asymptotic results.

Beran and Srivastava (1985) use bootstrap test and confidence regions for functions of the population covariance matrix, for example, eigenvalues and eigenvectors, which have the desired asymptotic levels if model restrictions, such as multiple eigenvalues, are taken into account in designing the bootstrap algorithm.

Efron and Tibshirani (1986) give a review of bootstrap methods for estimating standard errors and confidence intervals. The bootstrap is also extended to other measures of statistical accuracy such as bias

and prediction error, and to complicated data structures such as time series, censored data, and regression models.

Bootstrap confidence intervals have been discussed with new improvements by Efron (1987) and Hall(1988), and their applications to problems in a wide range of situations is given by Diciccio and Tibshirani (1987).

Albanese and Knott (1990) applies empirical and parametric bootstrap methods to find symetric percentil-t confidence intervals for $\alpha_{i,1}$ in a logit-probit model, and to test the reliability of the ranking of the respondents on the latent scale.

Albanese(1990, Chapter 3) investigates the adequacy of the asymptotic variance-covariance matrix in latent trait analysis for binary response (logit/probit model) through the jackknife and bootstrap (empirical and parametric) techniques. The investigation is carried out using e sets of real data, which represents a good range of different patterns of parameter estimates and sample sizes. Albanese also studies how close is the bootstrap distribution of the parameter estimates to a normal distribution. The main results are:

(1)- The more closely the bootstrap distribution of the parameter estimates is fitted by a normal distribution, the better is the agreement between the bootstrap and the asymptotic standard deviation.

(2)- If $\hat{\alpha}_{i,1}$ is not large, the asymptotic variance matrix can probably be trusted, since the bootstrap estimates and standard deviations are very close to the ML estimates and to the asymptotic standard deviations. Furthermore, this similarity increases as the sample size becomes larger.

(3)- Large values for $\hat{\alpha}_{i,1}$ are associated with skewed distributions or a mixture of two distributions, one normal and another with $\alpha_{i,1}$ equal to infinity. Probably the asymptotic standard deviations of the parameter estimates are smaller than the true ones.

(4)- If the sample size is small and one of the items has very large $\hat{\alpha}_{i,1}$ while the remaining ones are small, all with relative large standard deviations then it is likely that most of the estimates can not be trusted.

(5)- In summary, although the bootstrap distribution must underestimate the variation in the true sampling distribution, there is strong evidence that it gives a better guide than the usual first order normal approximation. Bootstrapping methods seem to be very useful for investigating the adequacy of the normal approximation in doubtful cases. When the discrimination parameters are small the asymptotic theory works well, but when they get large it is inadequate.

(6)- Jackknife parameter estimates and their standard deviations tend to be very similar to the original ML ones, independent of the pattern of the $\hat{\alpha}_{i,1}$'s and the sample size. Therefore, jackknife is not as good as bootstrap in warning about possible inadequacy of the asymptotic standard deviations. This undesirable result for the jackknife method may be due to the small number of different jackknife pseudovalues (16 in the case examined), and a larger number of items would provide more satisfactory results.

7- *Breakdown of the Estimation Procedure*

There are some configurations of data, analogous to those called Heywood cases in factor analysis (underlying variable model), in which the 'true' ML estimate is infinite and the iterative system proceeds in that direction indefinitely. In this situation, after some number of cycles most of the discrimination parameter estimates and the likelihood remain roughly constant while one or a few parameters increase indefinitely. The difference in goodness-of-fit with such a high discrimination parameter estimate (3.0 or bigger) is negligible. In practice in these cases the value of the estimate is a function of the stopping rule of the iterative procedure.

A Heywood case, on the other hand, is the occurence of a negative or zero estimate of the error variance Ψ for one or more variables. In the underlying model representation, $\alpha_{i,j} = -\lambda_{ij} / (\Psi_i)^{\frac{1}{2}}$, where λ_{ij} is the factor loading, so that a diverging discrimination parameter $\alpha_{i,j}$ in the response function model is equivalent to a Ψ_i (error variance) approaching zero.

In summary, according to Anderson and Gerbing (1984), Boomsma (1985) and Fachel (1986), the occurrence of Heywood cases increases as

(1)- the sample size decreases;

(2)- the number of indicators per factor and consequently the number of variables decreases, although Fachel has observed small variation between 5 and 100 variables;

(3)- the population values of the error variance are close to zero;

(4)- the factor loading are not uniform, for example, when only one factor loading increases up to 0.90 while the others remain equal to 0.5.

Van Driel (1978) identifies 3 causes for Heywood cases:

- sampling fluctuations combined with true values of the error variance close to zero;
- there does not exist any factor analysis model that fits the data;
- indefiniteness of the model (i.e.,too many true factor loadings are zero).

Bartholomew (1987) affirms, from his experience with binary estimation procedures, that the circumstances under which a slope parameter become larger and larger in the response function model are when

- the sample size is small, a few hundred or less,
- the number of variables is small and
- the discrimination parameters are very unequal,

which are equivalent to those leading to (1), (2) and (4) given above for the Heywood cases in factor analysis.

Albanese(1990, Chapter 5) investigates the stability of the discrimination parameter estimate $\hat{\alpha}_{i,1}$, for both fixed and decreasing sample size, when the number of items is reduced. The main results are:

(1)- When considering the effect of deleting items, the magnitude of the sample size must be judged in relation to the number of items.

(2)- The occurence of a large $\hat{\alpha}_{i,1}$ seems to depend more on which items are included in the test then on the sample size and test length. For the data we have analysed, large $\hat{\alpha}_{i,1}$ values were not associated to the type of question asked.

(3)- The probability of the occurence of a large $\hat{\alpha}_{i,1}$ does not increase as the number of items decreases for sample size of order 500, as it is often said to happen for Heywood cases in factor analysis.

(4)- As the number of items decreases, the largest $\hat{\alpha}_{i,1}$ tends to increase and become very large, when the test length is small.

(5)- Parameter estimates $\alpha_{i,1}$ ($\geq$0.50) and standard deviations are approximately linearly related so that larger estimates have larger standard deviations.

Albanese(1990, Chapter 6) investigates the conditions which give rise to large discrimination parameter estimates. This study leads to

(1)- A method for the generation of a (p+1)th item with any fixed $\hat{\alpha}_{p+1,1}$ and $\hat{\alpha}_{p+1,0}$. In particular, we can generate an item with large $\hat{\alpha}_{i,1}$ with patterns similar to those in real data sets. For this configuration of data, a threshold response may be the ML estimated response function, and it seems more reasonable to accept them as legitimate, and not to seek to remove them by adding or dropping items, as it has often been suggested.

These findings are also reported in Knott and Albanese (1990).

(2)- Strong evidence that the size of the discrimination parameter estimates $\hat{\alpha}_{i,1}$ may be related to the predictability of item i from the remaining ones. This can be summarise as following:

The more predictable is one item from all the remaining ones, the larger is its discrimination parameter estimate $\hat{\alpha}_{i,1}$.

An item with a large $\hat{\alpha}_{i,1}$ may not give any additional information about the latent variable in a test with 10 or more items, but for smaller length, for example 5, it may contain more information than the remaining items.

8- *Goodness-of-fit*

If the sample size (n) is large compared with 2P (number of possible response patterns) a chi-square or log-likelihood goodness-of-fit test can be carried out on the observed and expected frequencies of the response patterns. Often, there are many small expected frequencies so that pooling becomes necessary. Since the number of degrees of freedom in the unpooled case is 2P-p(q+1)-1, then situations may occur where there will be no degrees of freedom to judge the goodness of fit.

When a formal test cannot be carried out and p is not too large, the goodness of fit of the model may be judged by comparing the observed and expected frequencies of the response patterns. An additional check maybe done by comparing the observed and the fitted values of the one-and-two way marginal frequencies.

There are other checks which can be made on the data before or after fitting a model. For example, Bartholomew (1980) showed that if a one-latent variable model applies then it must be possible to label the categories so that, in the population, all the cross-product ratios exceed one. A systematic approach to the question of whether the data are consistent with an unidimensional model has been developed by Holland (1981) and extended by Rosenbaum (1984).

According to Rosenbaum, theorem 1, if a latent variable model is unidimensional for P[X=x] with nondecreasing response function then X is conditional associated, i.e, for all nondecreasing functions g(.) and f(.), all functions h(.) and all partitions and rearrangements of X into two nonoverlapping groups of items, **(S,T)**,

$$\mathrm{Cov}(g(\mathbf{S}), f(\mathbf{S}) \mid h(\mathbf{T})) \geq 0$$

where Cov(.,.|.) denotes conditional population covariance.

In particular, if we take $\mathbf{S}=(X_i,X_j)$, and $\mathbf{T}$ equal to the remaining p-2 items with $h(\mathbf{T}) = \sum_{k\neq i,j} X_k$ then a unidimensional latent variable model for $P[\mathbf{X}=\mathbf{x}]$ with nondecreasing response function implies that

$$\mathrm{Cov}((X_i, X_j) \mid \sum_{k\neq i,j} X_k = t) \geq 0$$

for all pairs of manifest variables and all values of t in the population.

Equivalently, such a model implies that there is a population cross-product ratio of at least equal 1 in every 2x2 subtable of the (p-1) layer of the 2x2x(p-1) population contingency table recording $X_i \times X_j \times \sum_{k\neq i,j} X_k$.

Tatsuoka (1984) describes the use of caution indices to identify individuals with unusual response patterns relative to a given model.

Due to the special properties of the Rasch model several goodness-of-fit test have been developed, of these, the conditional likelihood ratio test, introduced by Andersen (1973b), is perhaps the best known. The test is based on a comparison between item difficulties estimated from different subsamples formed according to the number of positive responses on the test and overall estimates obtained from the whole sample. If the Rasch model fits the data well then consistent difficulty estimates should be obtained for any subdivision of the sample into two or more groups. It was shown that, when the sample is large, the test statistic has approximately a χ^2 distributed random variable with (p-1)(g-1) degrees of freedom, where p is the number of items and g is the number of subsamples considered.

The problem of fit of the Rasch model has been further discussed by Gustafsson (1980b), Van den Wollenberg (1982), Molenaar (1983), and Kelderman (1984), among others.

Gustafsson (1980b) presented a test for the hypopaper that two disjoint groups of items measure the same construct. This provides a test of unidimensionality when items are grouped a priori. Van den Wollenberg (1982) also developed test statistics for lack of equality of discrimination parameters and unidimensionality.

These type of tests are global measures of how all the items in a test fit the Rasch model. To assess goodness-of-fit to a given item response function, Gustafsson (1980b) suggested using graphical procedures. On the other hand, Molenaar (1983) has provided procedures for a more detailed analysis under the Rasch model, which also involves information about the goodness-of-fit for a given item response function.

McKinley and Mills (1985) conducted an extensive investigation of goodness-of-fit indices for a given item response fuction. They compared four such indices, those developed by Bock (1972), Yen (1981), Wright and Mead (1978), and the Likelihood Ratio (LR) statistic. The first three of these employ the standard chi-square goodness-of-fit formula and vary only with respect to the number of groups and the definition of the latent level used to compute the expected proportion of positive responses.

Nine tests with length 75 and sample sizes of 500, 1000 and 2000 individuals were used to generate the simulated data under each of the one-, two-,and three-parameter models. In addition, the normally distributed samples had means of -1, 0 and 1 on the latent scale. When the data generated by the two and three parameter models were analysed under a one-parameter model, the results indicated a consistent lack of fit. As was the case with Yen's (1981) study, analising three-parameter data using the two parameter model worked quite well.

McKinley and Mill (1985) concluded that the LR index appeared to yield the fewest erroneous rejections of the hypopaper of fit, while the Bock index yielded fewer erroneous conclusions of fit. However, the differences were slight. They also applied the four procedures to an additional 9 tests having an underlying multidimensional latent structure. In all cases, the analysis yielded a high proportion of misfits. Thus, the underlying assumption of unidimensionality appears to be critical to obtaining good fit between the ICC and the observed data.

9- *Measurement of the Latent Variable*

After the model has been fitted, our main interest is in the measurement of the latent variable. Considering Y either as a parameter or a variable has given rise to different procedures when looking for more information about Y.

In fixed effects versions of the model each individual's position on the latent scale is represented by a parameter; in the random effects versions, individuals are supposed to be sampled at random from some population so that they latent position is the value of a random variable.

In educational testing, where Y is usually treated as a parameter, some work has been done in estimating the parameters of the latent distribution function; see for example, Lord (1983b), Andersen and Madsen (1977), Samanthanan and Blumenthal (1978) and Mislevy (1984).

On the other hand, Bartholomew (1984), treating Y as a latent variable has deal with the situation by scaling the latent variable, i.e, locating the individuals in the Y-space on the basis of their observed values of **X**.

Since in this paper we are treating Y as a latent variable, we look at the scaling, instead of the estimation of the parameters of the latent distribution.

According to Bartholomew (1984) the scaling of the latent variable Y should be done via the posterior density of y given the response pattern **x**. Thus, for example, he suggests the mean $E(Y|\mathbf{x})$ (or $E(Z|\mathbf{x})$), which may not be particularly appropriate when the posterior density $h(z|\mathbf{x})$ is highly skewed.

Bartholomew also shows that $E(\Phi(y)|c_1(\mathbf{x}))$ is a nondecreasing function of the component score $c_1(\mathbf{x}) = \sum_{i=1}^{p} \alpha_{i,1} x_i$ for every nondecreasing function of $\Phi(y)$. In particular, $E(Y|c_1(\mathbf{x}))$ or $E(Z|c_1(\mathbf{x}))$ is an increasing function of $c_1(\mathbf{x})$. This means that the component score induces a stochastic ordering of the posterior distributions. Thus, for example, the rank of individuals given by the component score $c_1(\mathbf{x})$ is the same as given by the posterior means $E(Y|\mathbf{x})$ and $E(Z|\mathbf{x})$. Therefore

if we are only interested in the ranking of the individuals on the latent scale, we can use any one of these three measures, from which the component score is the easiest to be obtained.

If the $\alpha_{i,1}$'s are very similar then the ranking determined by $X = \sum_{i=1}^{p} \alpha_{i,1} x_i$ and $\sum_{i=1}^{p} x_i$ are likely to be the same whichever latent models (logit/probit, logit or probit) we are using. When this situation happens the convergence of the algorithm for estimation of the parameters (section 3.2) is obtained quicker than when at least one of the estimates differs from the other.

Furthermore, Bartholomew (1984) pointed out that for the logit model $E(Y|x)$ is an approximately linear function of the component score $c_1(x) = \sum \alpha_{i,1} x_i$, which can be justified by a Taylor expansion if all $\alpha_{i,1}$'s are small. At the same time, when all $\alpha_{i,1}$'s are equal to 1 and π_i's are equal to 0.5 then the exact value of $E(Y|x)$ is

$(1 + c_1(x)) / (2 + A)$, where $A = \sum \alpha_{i,1}$. He also found out from empirical work that the relationship between $E(Y|x)$ and $c_1(x)$ is approximately linear well outside the range of the validity of this later result. Albanese(1990, Chapter 7) shows that this is often false when at least one of the $\hat{\alpha}_{i,1}$ is large (say $\geq 3/\sigma$, where σ is the standard deviation of the latent distribution). This may be due to the fact that the component scores are strongly dependent on the values of $\hat{\alpha}_{i,1}$ while $E(Y|x)$ or $E(Z|x)$ depend on π_i, which is nearly the same for all $\hat{\alpha}_{i,1} \geq 3/\sigma$, independent of $\hat{\alpha}_{i,0}$.

Furthermore, Albanese(1990, Chapter 7) studying the relationship between component scores and posterior means found out that

(1)- Significant differences between component scores do not always reflect different positions on the latent scale, according to the $E(Y|x)$ or $E(Z|x)$. They are shown through flat sections or jumps in the curve obtained when plotting the component scores against the means $E(Y|x)$ (or $E(Z|x)$).

The occurrence of flat sections seems to depend on the number of items with large $\hat{\alpha}_{i,1}$ and test length. At the same time, we expect that the effect of 2 large $\hat{\alpha}_{i,1}$ in a test with 40 items is smaller than in a test with 20 items. Usually, they do not present a specific pattern for the items with large $\hat{\alpha}_{i,1}$.

(2)- The greater the test length, the greater the possible number of different score patterns and configuration of $\hat{\alpha}_{i,1}$'s can occur and the less likely the linearity between the posterior mean and the component score seems to be.

Very often, in practice, we are not only interested in the ranking of the individuals, which is obtained either from the component scores or from the posterior means $E(Z|\mathbf{x})$ or $E(Y|\mathbf{x})$. Thus, for example, in Educational Testing, we may be interested in comparing the lower with the higher abitity group of individuals. The criterion for the distribution (allocation) of the respondents in groups is usually based on an arbitrary percentage, for example 20%.

If we know the distribution of the individuals along the latent scale, then we can use this information to partition the sample in groups. One way to do this is to use the information given by the posterior density $h(z|\mathbf{x})$ or even the mean $E(Z|\mathbf{x})$ (or $E(Y|\mathbf{x})$).

If we intend to use the mean $E(Z|\mathbf{x})$ (or $E(Y|\mathbf{x})$) as the measure of comparison between the position of the individuals on the latent scale then we must have information about the shape of $h(z|\mathbf{x})$, at least in terms of skweness and spread.

Let us consider two individuals with different score patterns $\mathbf{x}_1$ and $\mathbf{x}_2$ and the posterior densities $h(z|\mathbf{x}_1)$ and $h(z|\mathbf{x}_2)$, which are not skew and have nearly the same dispersion. If $h(z|\mathbf{x}_1)$ and $h(z|\mathbf{x}_2)$ have roughly the same mean then $\mathbf{x}_1$ and $\mathbf{x}_2$ lead to the same beliefs about the value of Z.

In these situations the mean $E(Z|\mathbf{x})$ (or $E(Y|\mathbf{x})$ is a reliable measure to compare individuals according to their position on the latent scale.

For the logit/logit model, Bartholomew (1984) shows that when π_i and $\alpha_{i,1}$ are fixed, the posterior density $h(y|\mathbf{x})$ depends on $\mathbf{x}$ only through the component score $c_1(\mathbf{x})$. And therefore, under this conditions $c_1(\mathbf{x})$ is a Bayesian sufficient statistic of y. This property is not shared, for example, by the probit model used by Bock and Liberman (1970). However Albanese(1990) shows that this is only true if $\alpha_{i,1}$ is infinity.

Investigation of the shape of the posterior density function carried out by Albanese showed that:

(1)- As the number of items increases, the posterior distributions look more normal and less skew, though with different variances. This is even true if there are several $\alpha_{i,1}$'s estimated as large, and the relation between the posterior means and the component scores is far from linear. This result indicates that the joint analysis of the posterior maen $E(Z|\mathbf{x})$ and its standard deviation can provide a reliable measurement of the latent variable.

(2)- We do not need to determine all the $h(z|\mathbf{x})$'s to have a clear idea about the distribution of $h(z|\mathbf{x})$ along the latent scale Z. Instead, we can select a representative sample of $h(z|\mathbf{x})$, selecting the score pattern $\mathbf{x}$ so that the whole set of values assumed by $E(Y|\mathbf{x})$ (or $E(Z|\mathbf{x})$) is covered.

(3)- If we desire to make groups of individuals according to their distribution on the latent scale, we can combine the information obtained from the shape of $h(z|\mathbf{x})$'s for all $\mathbf{x}$ with the observed frequency distribution of these posterior distributions (see Albanese, 1990)).

Knott and Albanese (1990) show that for the logit model the assumption of normality for the shape of one of the posterior distributions of the factor values implies the normality of all of them, and that normal posterior distributions imply that the posterior means are linear in the component scores. The identical ordering of individuals by means of the posterior distribution and by the component scores emerges as a simple consequence of the same approach.

10- *Analysis of a real set of data using a logit/probit model*

Law School Admission Test, Section VI

The Law School Admission Test (LSAT, section VI) consists of a number of dichotomous items, which can be answered correctly (1) or incorrectly (0). The purpose of the LSAT is to produce a score for each respondent, which estimates his ability to solve the problems presented as items in the test.

Table 1 - Frequency distribution and scores obtained by fitting the logit-probit model to the Law School Admission Test Section VI data.

Response pattern	Frequency observed	Frequency expected	Total score	Posterior mean
00000	3	2.3	0	-1.90
00001	6	5.9	1	-1.48
00010	2	2.6	1	-1.46
01000	1	1.8	1	-1.43
10000	10	9.5	1	-1.37
00100	1	0.7	1	-1.32
00011	11	8.9	2	-1.03
01001	8	6.4	2	-1.01
10001	29	34.6	2	-0.94
10010	14	15.6	2	-0.92
00101	1	2.6	2	-0.90
11000	16	11.3	2	-0.90
00110	3	1.2	2	-0.88
10100	3	4.7	2	-0.79
01011	16	13.6	3	-0.55
10011	81	76.6	3	-0.48
11001	56	56.1	3	-0.46
00111	4	6.0	3	-0.44
11010	21	25.7	3	-0.44
01101	3	4.4	3	-0.42
01110	2	2.0	3	-0.40
10101	28	25.0	3	-0.35
10110	15	11.5	3	-0.33
11100	11	8.4	3	-0.30
11011	173	173.3	4	0.01
01111	15	13.9	4	0.05
10111	80	83.5	4	0.12
11101	61	62.5	4	0.15
11110	28	29.1	4	0.17
11111	298	296.7	5	0.65

Table 1 shows the frequency distribution and the results obtained by fitting a one-factor logit/probit model, using the TWOMISS program (Albanese and Knott, 1991) with 48 quadrature points and relative change of the likelihood equal to 0.00001.

The fit of the logit-probit model provided a goodness of fit measure χ^2 equal to 15.30 on 13 degrees of freedom, which indicates a very satisfactory fit. We may conclude that the items are measuring a single latent variable. The scaling given by the posterior mean is consistent with that of the total number of correct responses because the discrimination parameter estimates $\hat{\alpha}_{i,1}$ are very similar, as we can see in Table 2. In this case is reasonable to expect that the Rasch model, special case of the logit/probit model where all $\alpha_{i,1}$'s are equal, would also fit the data. We fitted the Rasch model using the RASCHMIS program (Albanese and Knott, 1991), which results confirmed our expectations and the parameter estimates are displayed in Table 2

Table 2- Parameter estimates and standard deviations obtained from fitting a logit/probit and the Rasch model to the LSAT VI data.

	Logit-Probit				Rasch	
item	$\hat{\alpha}_{i,1}$	(SD)	$\hat{\alpha}_{i,0}$	(SD)	$\hat{\alpha}_{i,0}$	(SD)
1	0.82	(0.25)	2.77	(0.20)	2.73	(0.13)
2	0.72	(0.19)	0.99	(0.09)	0.99	(0.08)
3	0.89	(0.23)	0.25	(0.08)	0.24	(0.07)
4	0.69	(0.19)	1.28	(0.10)	1.31	(0.08)
5	0.66	(0.20)	2.05	(0.13)	2.10	(0.10)
					$\hat{\alpha}_1 = 0.76$	SD=0.07

As both logit-probit and Rasch fit the data well, the individuals may be scored using either the posterior means or the total number of correct responses. We could also have used the component scores, since no discrimination parameter is very large compared with the remaining ones. Figure 2 shows clearly that the posterior mean $E(Z|\mathbf{x})$ is a linear function of the component score $c_1(\mathbf{x}) = \sum \alpha_{i,1} x_i$.

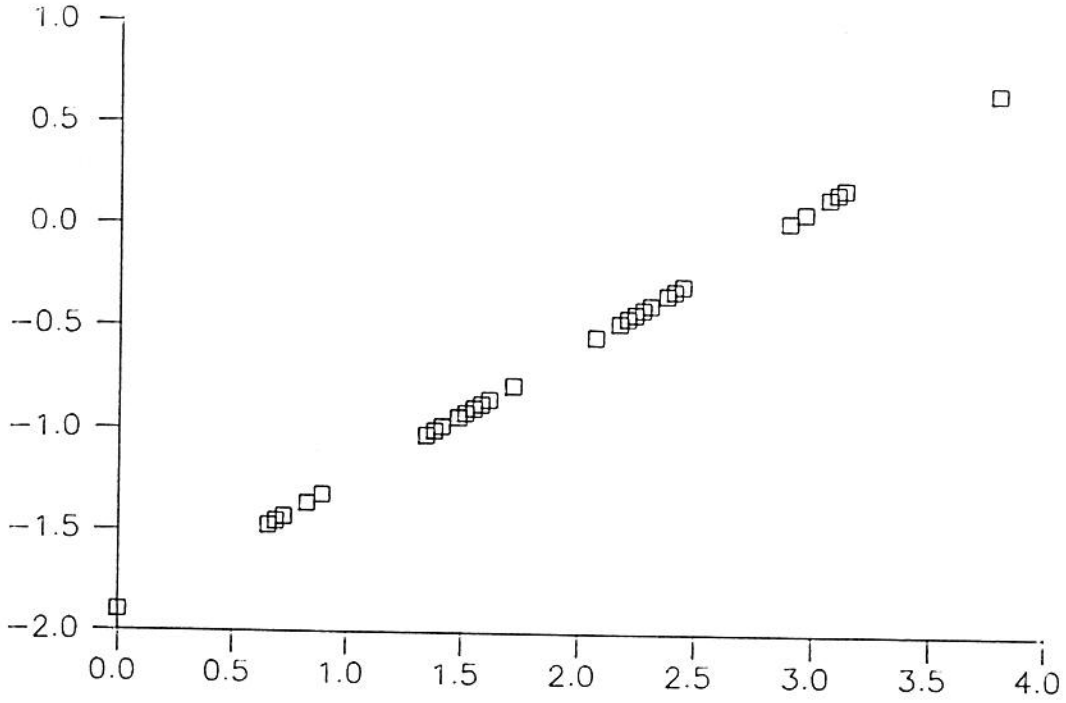

Figure 2- Relation between $E(Z|\mathbf{x})$ and $\sum \alpha_{i,1} x_i$ when fitting a logit-probit model to the LSAT VI.

From Figure 2 we can see that the response patterns are distributed into 6 groups along the line $-1.92 + 0.67\, c_1(\mathbf{x})$. Table 2 shows that they correspond to the 6 different values assumed by $\sum x_i$. As the number of positive responses increases by one unit, both posterior means, $E(Z|\mathbf{x})$ and $E(Y|\mathbf{x})$, and the component score $c_1(\mathbf{x})$ jump to higher values.

Figure 3 shows a representative collection of posterior distributions of the latent variable in a Normal scale. The posterior distributions are approximately normal with similar dispersions. The response pattern 00000 is furthest off the line in Figure 2 and has the most different variance in Figure 3.

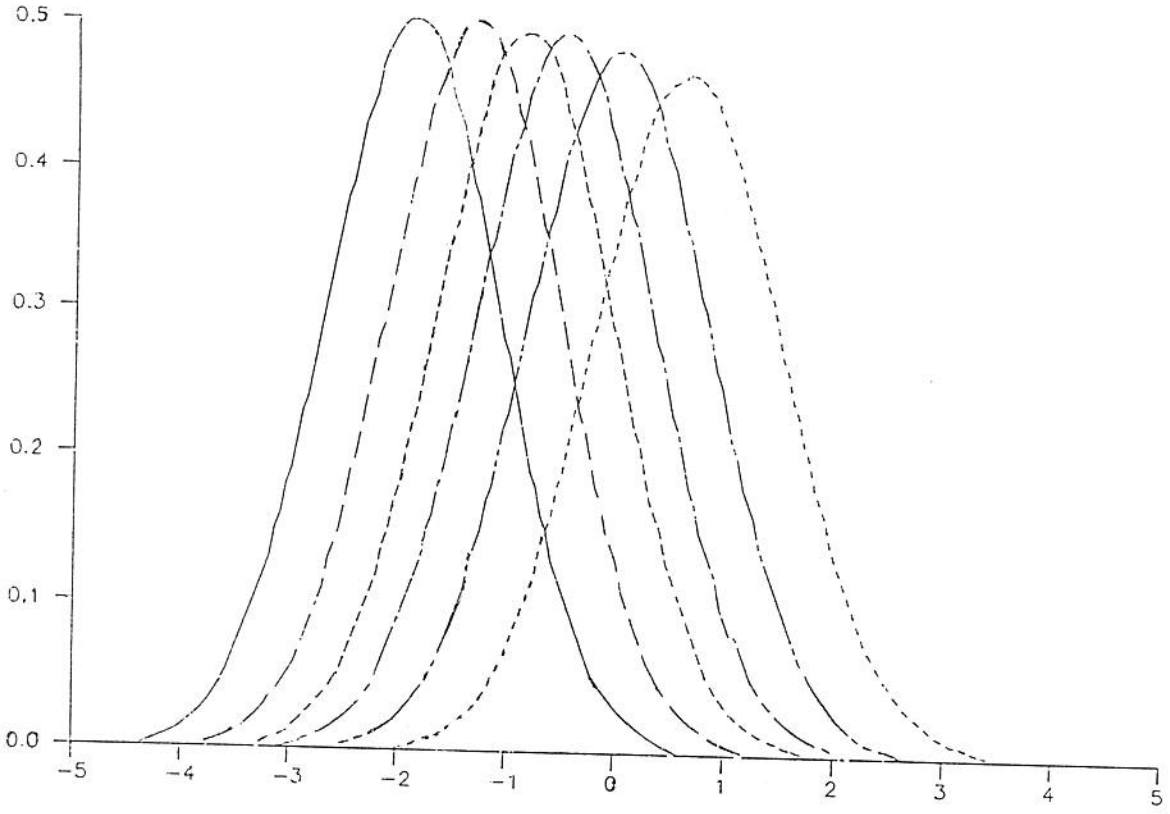

Figure 3- Posterior densities h(z|x) when fitting a logit-probit model to the LSAT VI, for the response patterns '00000', '01000', '00101', '01101', '10111' and '11111'.

It is convenient to point out that if there are one or more $\alpha_{i,1}$'s estimated as very large, the relation between the posterior means and the component scores is far from linear, but in general the posterior distributions are still very close to normal. This implies that in these cases the posterior means are more reliable measures of the position of the individuals on the latent scale than the component scores or the total number of correct responses.

11- *References*

Albanese,M.T.(1990) Latent variable models for binary response. Ph.D. thesis. University of London.

Albanese,M.T. and Knott,M.(1990) Bootstrapping latent variable models for binary response. Submitted for publication in British Journal of Mathematical and Statistical Psychology.

Knott,M. and Albanese,M.T.(1990) Posterior distribution of factor values and scaling for binary data. Submitted for publication.

Albanese,M.T. and Knott,M.(1991) TWOMISS: a computer program for fitting a two-factor logit-probit model to binary response data when observations may be missing.

Albanese,M.T. and Knott,M.(1991) RASCHMISS: a computer program for fitting the Rasch model to binary response data when observations may be missing.

Andersen,E.B.(1970) Asymptotic properties of conditional maximum likelihood estimators. J. Roy. Statist. Soc., B, 34, 283-301.

Andersen,E.B.(1972) The numerical solution of a set of conditional estimation equations. J. Roy. Statist. Soc., B, 34, 42-54.

Andersen,E.B.(1973a) Conditional inference for multiple-choice questionnaries. Br. J. Math. Statist. Psychol., 26, 31-44.

Andersen,E.B.(1973b) A goodness of fit test for the Rasch model. Psychometrika, 38, 123-140.

Andersen,E.B. and Madsen,M.(1977) Estimating parameters of the latent population distribution. Psychometrika, 42, 357-374.

Andersen,E.B.(1980) Discrete Statistical Models with Social Science Applications. Amsterdam: North-Holland Publishing Company.

Anderson,J.C. and Gerbing,D.W.(1984) The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. Psychometrika, 49, 155-173.

Baker,F.B.(1988) Methodology review: item parameter estimation under one-, two-, and three-parameter logistic models. Appl. Psychol. Measurement, 11, 111-141.

Bartholomew,D.J.(1980) Factor analysis for categorical data. J. Roy. Statist. Soc., B, 42, 293-321.

Bartholomew,D.J.(1981) Posterior analysis of the factor model. Br. J. Math. and Statist. Psychol., 34, 93-99.

Bartholomew,D.J.(1983) Latent variable models for ordered categorical data. J. Econometrics, 22, 229-243.

Bartholomew,D.J.(1984) Scaling binary data using a factor model. J. Roy. Statist. Soc., B, 46, 120-123.

Bartholomew,D.J.(1987) Latent variable models and factor analysis. London: Charles Griffin & Company Ltd.

Bartholomew,D.J.(1988) The sensitivity of latent trait analysis to choice of prior distribution. Br. J. Math. and Statist. Psychol., 41, 101-107.

Beran,R. and Srivastava,M.S.(1985) Bootstrap tests and confidence regions for functions of a covariance matrix. Ann. Statist., 13, 95-115.

Birnbaum,A.(1968) Some latent trait models and their use in infering an examinee's ability. In Lord,F.M. and Novick,M.R., Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, Chapters 17-20.

Bock,R.D.(1972) Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 37, 29-51.

Bock,R.D. and Aitkin,M.(1981) Marginal maximum likelihood estimation of item parameters: application of an E-M algorithm. Psychometrika, 46, 443-459.

Bock,R.D. and Lieberman,M.(1970) Fitting a response model for n dichotomously scored items. Psychometrika, 35, 179-197.

Boomsma, A.(1985) Nonconvergence, improper solutions and starting values in LISREL maximum likelihood estimation. Psychometrika, 50, 229-242.

Chatterjee,S.(1984) Variance estimation in factor analysis: An application of the bootstrap. Br. J. Math. Statist. Psychol., 37, 252-262.

Clarkson,D.B.(1979) Estimating the standard errors of rotated factor loadings by jackknifing. Psychometrika, 44, 297-314.

Cox,D.R.(1966) A simple example of a comparation envolving quantal data. Biometrika, 53, 215-220.

Dempster,A.P.,Laird,N.M. and Rubin,D.B.(1977) Maximum likelihood from incomplete data via the E-M algorithm. J. Roy. Statist. Soc., B, 39, 1-38.

Diciccio,T. and Tibshirani(1987) Bootstrap confidence intervals and bootstrap approximations. J. Amer. Statist. Assoc., 82, 163-170.

Dinero,T. and Haertel,E.(1977) Applicability of the Rasch model with varying item discriminations. Appl. Psychol. Measurement, 1, 581-592.

Efron,B.(1979) Bootstrap methods: another look at the jackknife. Ann. Statist., 7, 1-26.

Efron,B.(1984) Better bootstrap confidence intervals. Tech. Rep. Stanford University Dept. Statist.

Efron,B. and Tibshirani,R.(1986) Bootstrap method for standard errors, confidence intervals and other measures of statistical accuracy. Statist. Science, 1, 54-75.

Efron,B.(1987) Better bootstrap confidence intervals. J. Amer. Statist. Assoc., 82, 171-185.

Fachel,J.M.G.(1986) The C-type Distribution as an Underlying Model for Categorical Data and its use in Factor Analysis. Ph.D. Thesis, University of London.

Fisher,G.H.(1981) On the existence and uniqueness of maximum likelihood estimates in the Rasch model. Psychometrika, 46, 59-77.

Goldstein,H.(1980) Dimensionality, bias, independence and measurement scale problems in latent trait test score models. Br. J. Math. Statist. Psychol., 33, 234-246.

Grönroos,M.(1985) Boothstrapping in Factor Analysis. 45th Biennal Session of the International Statistical Institute, 12-22.

Gruijter,D.N.M.(1985) A note on the assymptotic variance-covariance matrix of item parameter estimates in the Rasch model. Psychometrika, 50, 247-249.

Gustafsson,J.E.(1980a) A solution of the conditional estimation problem for long tests in the Rasch model for dichotomous. Educ. and Psychol. Measurement, 40, 377-385.

Gustafsson,J.E.(1980b) Testing and obtaining fit of data to the Rasch model. Br. J. Math. Statist. Psychol., 33, 205-233.

Haberman,S.J.(1977) Maximum likelihood estimates in exponencial response models. Ann. Statist., 5, 815-841.

Hambleton,R.K. and Swaminathan,H.(1985) Item Response Theory: Principles and Applications. Boston: Kluwer-Nijhoff.

Holland,P.W.(1981) When are item response models consistent with observed data? Psychometrika, 46, 79-92.

Hulin,C., Lissak,R. and Drasgow,F.(1982) Recovery of two and three-parameter logistic item characteristic curves: a Monte Carlo study. Appl. Psychol. Measurement, 6, 249-260.

Jorgensen,M.A.(1987) Jackknifing fixed points of iterations. Biometrika, 74, 207-211.

Jöreskog,K.G. and Sörbom,D.(1984) LISREL VI user's guide. Mooresville, Indiana: Scientific Software, Inc.

Kelderman,H.(1984) Loglinear Rasch model tests. Psychometrika, 49, 223-245.

Kendall,M.G. and Stuart,A.(1979) The Advanced Theory of Statistics, vol.2, New York: Hafner.

Knott,M. and Albanese,M.T.(1990) Large values for maximum likelihood estimates of parameters in latent variable models. Submitted for publication in Psychometrika.

Lord,F.M.(1952) A theory of test scores. Psychometrika Monograph, number 7, 17.

Lord,F.M.(1968a) An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. Educ. and Psychol. Measurement
28, 989,1020.

Lord,F.M. and Novick,M.R.(1968b) Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley.

Lord,F.M.(1983a) Statistical bias in maximum likelihood estimators of item parameters. Psychometrika, 48, 425-435.

Lord,F.M.(1983b) Maximum likelihood estimation of item response parameters when some responses are omitted. Psychometrika, 48, 477-482.

Lord,F.M. and Wingersky,M.S.(1983) Sampling variances and covariances of parameter estimates in item response theory. In D.J.Weiss(Ed.), Proceedings of the 1982 Item Response Theory and Computadorized Adaptative Testing Conference. Minneapolis: University of Minnesota, Department of Psychology, Computerized Adaptative Testing Laboratory. (p.69-88).

Louis,T.A.(1982) Finding the observed information matrix when using the E-M algorithm. J. Roy. Statist. Soc., B, 44, 226-233.

Mantel,N. and Haenszel,W.(1959) Statistical aspects of the analysis of data from retrospective studies of disease. J. Nat. Cancer Inst., 22, 719-748.

McFadden,D.(1982) Qualitative response models. Advances in Econometrics, W. Hildenbrand, Cambridge: Univ. Press, 1-37.

McKinley,R.L. and Mills,C.N.(1985) A comparison of several of goodness-of-fit statistics. Appl. Psychol. Measurement, 9, 49-57.

Miller,R.G.(1974) The jackknife _ a review. Biometrika, 61, 1-15.

Mislevy,R.J.(1984) Estimating latent distributions. Psychometrika, 49, 359-381.

Mislevy,R.J.(1985) Estimation of latent group effects. J. Am. Statist. Assoc., 80, 993-997.

Molenaar,I.W.(1983) Some improved diagnostics for failure of the Rasch model. Psychometrika, 48, 49-72.

Pennell,R.(1972) Routinely computable confidence intervals for factor loadings using the 'jackkife'. Br. J. Math. Statist. Psychol., 25, 107-114.

Quenouille,M.H.(1956) Notes on bias estimation. Biometrika, 43, 353-360.

Ramsey,J.(1975) Solving implicit equationsin psychometric data analysis. Psychometrika, 40, 337-360.

Rasch,G.(1960) Probabilistic models for some intelligence and attainment tests. Copenhagen: Paedagogiske Institut.

Rosenbaum,P.R.(1984) Testing the conditional independence and monotonicity assumptions of the item response. Psychometrika, 49, 425-435.

Sanathanan,L. and Blumenthal,S.(1978) The logistic model and estimation of latent structure. J. Amer. Statist. Assoc., 73, 794-799.

Samejima,F.(1974) Normal ogive model on the continuos response level in the multi-dimensional latent space. Psychometrika, 39, 111-121.

Shea,B.L.(1984) FACONE: A computer program for fitting the logit latent model by maximum likelihood. Department of Statistics, London School of Economics.

Stouffer,S. and Toby,J.(1951) Role conflit and personality. J. Sociology, 56, 395-406.

Stouffer,S., Guttman,L., Suchman,E., Lazarsfeld,P., Star,S. and Clausen,J.(1950) Measurement and prediction, Volume 4 of Studies in Social Psychology during World War. Princeton, NJ: Princeton University Press. Reprinted, 1966, New York: John Wiley.

Stroud,A.H. and Secrest,D.(1966) Gaussian Quadrature Formulas. Prentice-Hall.

Tatsuoka,K.K.(1984) Caution indices based on item response theory. Psychometrika, 49, 95-110.

Thissen,D.(1982) Marginal maximum likelihood estimation for the one-parameter logistic model. Psychometrika, 47, 175-186.

Thissen,D. and Wainer,H.(1982) Some standard errors in item response theory. Psychometrika, 47, 141-147.

Tsutakawa,R.K.(1984) Estimation of two-parameter logistic item response curves. J. Educ. Statist., 9, 263-276.

Tukey,J.W.(1958) Bias and confidence in not quite large samples. Ann. Math. Statist., 29,614 [δ1]

van de Driel,O.P.(1978) On various causes of improper solutions in maximum likelihood factor analysis. Psychometrika, 43, 225-243.

van de Vijner,F.J.(1986) The robustness of Rasch estimates. Appl. Psychol. Measurement, 10, 45-57.

van den Wollenberg,A.L.(1982) Two new test statistics for the Rasch model. Psychometrika, 47, 123-140.

van den Wollenberg,A.L., Wierda,F.W. and Jansen,P.G.W. (1988) Consistency of Rasch model parameter estimation: a simulation study. Appl. Psychol. Measurement, 12, 307-313.

Wood,R.C.;Wingersky,M.S. and Lord, F.M.(1976) Logistic- a computer program for estimating examinee ability and item characteristic curve parameters (Research Memorandum 76-6) Princeton NJ: Educational Testing Service.

Wright,B. and Douglas,G.A.(1977) Conditional versus unconditional procedures for sample-free item analysis. Educ. and Psychol. Measurement, 37, 573-586.

Wright,B. and Panchapakesan,W.(1969) A procedure for sample-free item analysis. Educ. and Psychol. Measurement, 29, 23-48.

Wright,B.D. and Mead,R.J.(1978) Bical: calibrating items and scales with the Rasch model (research Memorandum Nº 23A). Chicago: University of Chicago, Statistical Laboratory.

Wu,C.F.J.(1983) On the convergence properties of the EM algorithm. Ann. Statist., 3, 95-103.

Yen,W.M.(1981) Using simulation results to choose a latent trait model. Appl. Psychol. Measurement, 5, 245-262.

Yen,W.M.(1985) Increasing item complexity; a possible cause of scale shrinkage for unidimensional item response theory. Psychometrika, 50, 399-410.

Publicações do Instituto de Matemática da UFRGS
Cadernos de Matemática e Estatística

Série A: Trabalho de Pesquisa

1. Marcos Sebastiani - Transformation des Singularités - MAR/89.
2. Jaime Bruck Ripoll - On a Theoremof R. Langevin About Curvature and Complex Singularities - MAR/89.
3. Eduardo Cisneros, Miguel Ferrero e Maria Inés Gonzales - Prime Ideals of Skew Polynomial Rings and Skew Laurent Polynomial Rings - ABR/89.
4. Oclide José Dotto - ϵ - Dilations - JUN/89.
5. Jaime Bruck Ripoll - A Characterization of Helicoids - JUN/89.
6. Mark Thompson, V. B. Moscatelli - Asymptotic Distribution of Liusternik-Schnirelman Eigenvalues for Elliptic Nonlinear Operators -JUL/89.
7. Mark Thompson - The Formula of Weyl for Regions with a Self- Similar Fractal Boundary - JUL/89.
8. Jaime Bruck Ripoll - A Note on Compact Surfaces with Non Zero Constant Mean Curvature - OUT/89.
9. Jaime Bruck Ripoll - Compact ϵ - Convex Hypersurfaces - NOV/89.
10. Jandyra Maria G. Fachel - Coeficientes de Correlação Tipo -Contingência - JAN/90.
11. Jandyra Maria G. Fachel - The Probability of Ocurrence of Heywood Cases - JAN/90.
12. Jandyra Maria G. Fachel - Heywood Cases in Unrestricted Factor Analysis - JAN/90.
13. Julio Cesar R. Claeyssen, Tereza Tsukazan - Dynamical Solutions of Linear Matrix Differential Equations - JUL/90.
14. Maria T. Albanese - Behaviour of the Likelihood Function in Latent Trait Analysis of Binary Data - ABR/91.
15. Maria T. Albanese - Measurement of the Latente Variable in Latent Trait Analysis of Binary Data - ABR/91.
16. Maria T. Albanese - Adequacy of the Asymptotic Variance-Covariance Matrix using Bootstrap and Jackknife Techniques in Latent Trait Analysis of Binary Data - ABR/91.
17. Maria T. Albanese - Latent Variable Models for Binary Response - ABR/91.

Universidade Federal do Rio Grande Sul
Reitor: Professor Tuiskon Dick

Instituto de Matemática
Diretor: Professor Aron Taitelbaum
Núcleo de Atividades Extra Curriculares
Coordenador: Professora Jandyra G. Fachel
Secretária: Rosaura Monteiro Pinheiro

Os Cadernos de Matemática e Estatística publicam as seguintes séries:
Série A: Trabalho de Pesquisa
Série B: Trabalho de Apoio Didático
Série C: Colóquio de Matemática SBM/UFRGS
Série D: Trabalho de Graduação
Série E: Dissertações de Mestrado
Série F: Trabalho de Divulgação
Série G: Textos para Discussão

Toda correspondência com solicitação de números publicados e demais informações deverá ser enviada para:

NAEC - Núcleo de Atividades Extra Curriculares
Instituto de Matemática - UFRGS
Av. Bento Gonçalves, 9500
91.500 - Agronomia - POA/RS
Telefone: 36.11.59 ou 36.17.85 Ramal: 252