

A quadratic exponential model for the analysis of item response data

Un modello esponenziale quadratico per l'analisi di item di risposta

Francesco Bartolucci¹ and Fulvia Pennoni²

¹ Dipartimento di Economia, Finanza e Statistica, Università di Perugia

² Dipartimento di Statistica, Università degli studi di Milano-Bicocca
e-mail: Fulvia.Pennoni@unimib.it

Riassunto: Si propone un metodo di stima del modello logistico a due parametri (2-PL) per l'analisi di dati derivanti dalla somministrazione di un questionario con domande valutate in modo dicotomico. Il metodo è basato su un'approssimazione, formulata seguendo Cox e Wermuth (1994), del modello in questione tramite un modello esponenziale quadratico i cui parametri sono stimabili tramite una semplice procedura iterativa. Vengono considerati uno studio per simulazione ed un'applicazione su dati reali. Infine si accenna ad una possibile estensione del modello al caso multidimensionale.

Keywords: dimensionality testing, discriminant index, item banking, Rasch model.

1. Introduction

A well-known model for the analysis of dichotomously-scored item responses is the two-parameter logistic (2-PL) model of Birnbaum (1968). This model is based on the assumptions that: (i) the response variables corresponding to the items are conditional independent given a latent trait that, in educational assessment, represents the ability of the examinee; (ii) the conditional logit of the probability that an examinee responds correctly to an item is a linear function of his/her ability level with two item-specific parameters referred to as discriminant index and difficulty level. The first, in particular, measures how strongly the probability of a correct response depends on the ability. This parameter is then very useful for identifying items that, in a given item bank, must be substituted since do not discriminate adequately between more and less capable examinees.

The applicability of the 2-PL model is often limited by the difficulties that arise in the estimation of its parameters (see Hambleton and Swaminathan, 1985, Baker and Kim, 2004, and the references therein). In particular, the joint maximum likelihood method, in which the ability of any subject is represented through a fixed parameter, gives inconsistent estimates. On the other hand, the marginal maximum likelihood method, in which the ability is represented through a random parameter with a certain distribution, gives consistent estimates, but its implementation is more difficult. Note also that a conditional maximum likelihood approach, in which the ability parameters are eliminated via conditioning on suitable sufficient statistics, is not viable for the 2-PL model. These difficulties partially explain the greater diffusion of the Rasch model (Rasch, 1961) over the 2-PL model, even though the first is often considered restrictive, being based on the assumption that the discriminant index is constant across items.

In this paper we propose an approach to the estimation of the 2-PL model based on its approximation through a particular quadratic exponential (QE) model (Cox,

1972) with parameters corresponding to item difficulties and discriminant indices. This approximation is formulated on the basis of some results of Cox and Wermuth (1994). We also show that, under the approximating model, the conditional logit of the probability of responding correctly to an item, given the responses to all the other items, has an interesting interpretation within the Item Response Theory (IRT). This representation turns out to be useful in the estimation of the item parameters. The approach may be easily extended to the multidimensional case, in which the probability of responding correctly to an item depends on more than one latent trait, and used for testing unidimensionality.

The remainder of the paper is organized as follows. In Section 2 we describe more precisely the 2-PL model and the proposed approach, while in Section 3 we outline its multidimensional extension. A series of simulations and an example based on real data have been considered, the results of which are not reported here for lack of space.

2. The proposed approach

Let J denote the number of test items and let $Y_j, j = 1, \dots, J$, be a binary variable equal to 1 if a subject responds correctly to item j and to -1 otherwise. Note that, as in Cox (1972) and Cox and Wermuth (1994), we use the $(-1, 1)$ coding of the response variables, instead of the more usual $(0, 1)$ coding, since the first simplifies the algebra involved in the approach proposed here.

The 2-PL model assumes that the probability that an examinee with ability level θ responds correctly to item j is

$$p(Y_j = 1|\theta) = \frac{\exp(\alpha_j\theta - \beta_j)}{1 + \exp(\alpha_j\theta - \beta_j)}, \quad j = 1, \dots, J, \quad (1)$$

where α_j is the discriminant index of the item, which may be constrained to be non-negative in order to ensure monotonicity, and β_j is the difficulty parameter.

Our approach is based on the fact that if the ability of a subject is represented through a binary latent variable X , again with levels $(-1, 1)$, then the joint distribution of $(X, \mathbf{Y})$, $\mathbf{Y} = (Y_1, \dots, Y_J)$, is of QE type. From (1) we have in fact that

$$p(x, \mathbf{y}) = \pi^{(x+1)/2} (1 - \pi)^{(1-x)/2} \prod_j \frac{\exp[(y_j + 1)(\alpha_j x - \beta_j)/2]}{1 + \exp(\alpha_j x - \beta_j)},$$

where $\pi = p(X = 1)$. After some algebra, the above joint probability becomes

$$p(x, \mathbf{y}) = \exp \left(\mu + \gamma x - \frac{1}{2} \sum_j \beta_j y_j + \frac{1}{2} \sum_j \alpha_j x y_j \right), \quad (2)$$

where μ denotes the normalizing constant and γ depends on the original parameters in a suitable way. Cox and Wermuth (1994) showed that if a set of binary variables follows a QE distribution, then the marginal distribution of any subset of these variables may be accurately approximated by a distribution of the same family. Their result is based on a second-order Taylor-series expansion of the logarithm of the exact marginal probabilities around certain parameter values. By applying this rule, we have that, when $\gamma = 0$, the marginal distribution of $\mathbf{Y}$ deriving from (2) may be approximated by the QE distribution

$$p^*(\mathbf{y}) = \exp \left(\mu^* - \frac{1}{2} \sum_j \beta_j y_j + \frac{1}{4} \sum_{h < j} \alpha_h \alpha_j y_h y_j \right). \quad (3)$$

Our idea is then that of estimating the item parameters of the 2-PL model by fitting model (3). Note that, according to this model, the conditional log-odds ratio between Y_h and Y_j , $h \neq j$, given any configuration of the remaining variables, is simply equal to the product $\alpha_h \alpha_j$ between the corresponding discriminant indices. Also note that the assumption that $\gamma = 0$, made above, considerably simplifies the approximating model. This corresponds to a value of π approximately equal to $1/2$, which seems reasonable.

An interesting aspect of the approximating model (3) is that it implies that the conditional distribution of Y_j given the configuration $\mathbf{y}_{-j}$ of the remaining variables is

$$p^*(Y_j = 1 | \mathbf{y}_{-j}) = \frac{\exp(\alpha_j t_j / 2 - \beta_j)}{1 + \exp(\alpha_j t_j / 2 - \beta_j)}, \quad j = 1, \dots, J, \quad (4)$$

where $t_j = \sum_{h \neq j} \alpha_h y_h$. This expression closely resembles expression (1) on which the 2-PL model is based, the only difference being that the ability level θ is substituted by a quantity proportional to t_j . The latter, however, may be considered as a proxy of θ based on $\mathbf{y}_{-j}$ since it is the weighted sum of the elements of this vector with weights equal to the corresponding discriminant indices (Lord, 1980). This provides an intuitive explanation of model (3) and of the proposed estimation approach within IRT.

Expression (4) also suggests a very simple method for the maximum likelihood estimation of the item parameters α_j and β_j on the basis of the responses y_{ij} , $i = 1, \dots, n$, $j = 1, \dots, J$, provided by a sample of n subjects to the J test items. Starting from an initial guess for these parameters, the method is based on the following steps:

- for any j compute the weighted scores $t_{1j}, \dots, t_{nj}$, with $t_{ij} = \sum_{h \neq j} \alpha_h y_{ih}$, on the basis of the current estimates of α_h , $h \neq j$;
- for any j update the estimates of α_j and β_j by fitting a logistic model based on parametrization (4) to the data (y_{ij}, t_{ij}) , $i = 1, \dots, n$;
- iterate until convergence in the log-likelihood of the full model (3).

Note that the method at issue may be used even when there is a large number of items since it is based on a representation of the data through a $n \times J$ binary matrix whose dimension increases linearly with J . Standard errors for the maximum likelihood estimates at issue may be derived from the Fisher information matrix for (3). To take into account that the observed data come from the 2-PL model we can use a sandwich estimator instead.

The proposed approach has been studied through a series of simulations in which samples of suitable dimension are drawn from both the 2-PL models, under certain distributions of the ability, and the QE model (3). The main finding of these simulations is that, even if the samples are generated from the 2-PL model, the approximating model has a deviance usually close to that of the true model and the average estimation error for the parameters α_j and β_j is reasonable. This happens, under certain conditions, even when the assumed distribution of the ability strongly departs from a binary distribution.

3. Dealing with multidimensionality

Assume that the probability of responding correctly to an item depends on the levels of two different types of ability, θ_1 and θ_2 , as follows

$$p(Y_j = 1 | \theta_1, \theta_2) = \frac{\exp(\alpha_{1j} \theta_1 + \alpha_{2j} \theta_2 - \beta_j)}{1 + \exp(\alpha_{1j} \theta_1 + \alpha_{2j} \theta_2 - \beta_j)}, \quad (5)$$

where α_{aj} , $a = 1, 2$, $j = 1, \dots, J$, is the discriminant index for the ability of type a for item j . If we represent the two abilities through the independent binary latent variables X_1 and X_2 , the joint distribution of $(\mathbf{X}, \mathbf{Y})$, $\mathbf{X} = (X_1, X_2)$, follows the QE model

$$p(\mathbf{x}, \mathbf{y}) = \exp \left(\mu + \gamma_1 x_1 + \gamma_2 x_2 - \frac{1}{2} \sum_j \beta_j y_j + \frac{1}{2} \sum_j \alpha_{1j} x_1 y_j + \frac{1}{2} \sum_j \alpha_{2j} x_2 y_j \right).$$

This model corresponds to the latent factor model of Magidson and Vermunt (2001). Even for this model we can use the method of Cox and Wermuth (1994) for approximating the marginal distribution of $\mathbf{Y}$ when $\gamma_1 = \gamma_2 = 0$. The approximating model is in this case

$$p^*(\mathbf{y}) = \exp \left[\mu^* - \frac{1}{2} \sum_j \beta_j y_j + \frac{1}{4} \sum_{h < j} (\alpha_{1h} \alpha_{1j} + \alpha_{2h} \alpha_{2j}) y_h y_j \right]. \quad (6)$$

This generalizes model (3) by assuming a more complex structure of the second-order interaction between any pair Y_h, Y_j , $h \neq j$, of variables. Through a likelihood ratio statistic between the two models we can test the hypothesis of unidimensionality, according to which the items measure only one type of ability, against the hypothesis of bidimensionality. Also note that (6) implies a logistic model for Y_j given $\mathbf{y}_{-j}$ which closely resembles (5); in this case any ability level θ_{aj} is substituted by a quantity proportional to the proxy $t_{aj} = \sum_{h \neq j} \alpha_{ah} y_h$. This may again be used for providing an intuitive explanation of the approximating model within IRT and estimating its parameters through an iterative method similar to the one outlined at the end of Section 2.

Finally note that the approach above may be easily extended to the multidimensional case, i.e. when there are $d > 2$ types of ability. The approximating model is of QE type with two-way interaction parameter between Y_h and Y_j , $h \neq j$, equal to $\sum_{a=1}^d \alpha_{ah} \alpha_{aj} / 4$, with d that may be chosen through a suitable model selection criterion.

References

- Birnbaum A. (1968) Test scores, sufficient statistics, and the information structures of tests, in: *Statistical Theories of Mental Test Scores*, Lord F. M., Novick M. R. (Eds.), Addison-Wesley, 425–435.
- Baker F. B., Kim S. H. K. (2004) *Item Response Theory: parameters estimation techniques*, Second Edition, CRC press, New York.
- Cox D. R. (1972) The analysis of multivariate binary data, *Applied statistics*, 21, 113–120.
- Cox D. R., Wermuth N. (1994) A note on the quadratic exponential binary model, *Biometrika*, 81, 403–408.
- Hambleton R. K., Swaminathan H. (1985) *Item Response Theory: Principles and Applications*, Kluwer Nijhoff Publishing, Boston.
- Lord F. M. (1980) *Applications of item response theory to practical testing problems*, Erlbaum, Hillsdale.
- Magidson J., Vermunt J. K. (2001) Latent class factor and cluster models, bi-plots and related graphical displays, *Sociological Methodology*, 31, 223–264.
- Rasch G. (1961) On general laws and the meaning of measurement in psychology, *Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability*, 4, 321–333.