Dynamic Systems and Applications 29 (2020) 101-110

PARAMETER ESTIMATION MATRIX IN THE LOGISTIC MODEL BASED ON IRT AND ITS APPLICATION

CHENG HUANGJIN AND TAN CHANGCHUN

Basic Courses, Huainan Union University, Huainan, Anhui 232038, China
Email:gold5460@163.com
School of Economics, Hefei University of Technology, Hefei, Anhui 230009, China
Email:cctan@hfut.edu.cn

ABSTRACT. After analyzing the existing methods of parameter estimate, the logistic model has been redesigned, and its matrix expression has been obtained. And then it has been applied to polytomous scoring questions of academic competitions in colleges and universities. The correlation between several sets of variables are obtained, which has provided ideas and methods for the application of the simplified models in many fields.

AMS (MOS) Subject Classification. 62F10.

1. INTRODUCTION

In recent years, various statistical indexes in the large examinations have widely investigated. And the most frequently discussed are the difficulty, discrimination and pseudo-guessing of items, in which the first two indicators have been fixed particularly on. Generally speaking, the mathematical models used to discuss those indexes derive mainly from the classical test theory (CTT) and the item response theory (IRT). The former adopts the linear model to calculate different parameters of difficulty and discrimination of items. And the disadvantage of this model lies in the assumption that all the statistical indexes depend on samples of examinees. However, the later adopts the nonlinear probability model, which associates the trait level of examinees with their reaction behaviors on the item. And the characteristic curve of the item make the estimate of related parameters independent of samples of examinees and test items. Therefore, this model is much better than the previous one [1].

The premise of application of IRT is to set up the model of the characteristic curve. Recently, its application has also been widely investigated. The model has been developed from one dimension to multi-dimensions, from 0-1 scoring to polytomous scoring, and from non-timekeeping to timekeeping. The main model is the following

Received December 20, 2019

Logistic one [2]

$$P(\theta, a, b, c) = c + \frac{1-c}{1-e^{-Da(\theta-b)}}, \tag{1.1}$$

where θ is ability level of the examinee, called the ability parameter. In (1.1), a is the discrimination of the item, b is the difficulty of the item, c is the pseudo-guessing and D is the normal adjustment constant. And a, b, c are collectively known as item parameters. The function $P(\theta, a, b, c)$ represents the probability of the correct response from the examinee upon an item.

In terms of the parameter estimate method, Birnbaum first proposed the initial joint maximum likelihood estimate method, and Bock and Lieberman got the marginal maximum likelihood estimate method and then improved it. After decades of development, many special methods for special models spring out, such as the dual two-step iterative estimate method and the SQRT/EM method based on the empirical regression[3], the Monte Carlo method based on the artificial neural network algorithm and the joint maximum likelihood estimate method developed by replacing N-R iteration with the genetic algorithm [4].

However, most scholars pay more attention to theoretical results rather than practical applications. For example, some make use of the software simulation to generate tests and datum [5], and some employ nonparametric item response models [6]. In particular, there is a lack of regression analysis between the scores of the examinees and their abilities, as well as the research that couples back the theoretical results to the actual test. One of the main reason is that the existing parameter estimates are represented by tedious algebraic expressions, lacking the unified formulation and modularity. Therefore, in practice, especially when there are a great number of items and examinees, it is difficult to use the existing softwares, which are easy to cause such problems as a large amount of manual calculation, inaccurate results and repeated calculation.

In the present article, we will adopt the piecewise function with the practical significance as the observed value of examinees score rate, designing matrixes for the parameter estimates, and then apply them to academic competitions of colleges and universities, obtaining the experimental datum by the existing calculation software. In particular, we have made the statistical analysis on the results of the parameter estimate after iterations, obtaining the regression equation between the examinees score rate and their ability parameters. We also will get the actual feedback of the experimental results. This provides the method for the application of the matrix model in the improvement of test questions, test bank building and talent selection.

2. ALGEBRAIC EXPRESSIONS OF PARAMETER ESTIMATES

It is highly unlikely for an examinee to guess the correct answer of polytomous scoring test items, especially computational problems and the application in mathematical and physical competitions. Therefore, the probability of the correct response in the three-parameter model can be regarded as the scoring rate of the examinee and the two-parameter model can be used for the research.

We assume that n examinees take a test with the length of m. We will introduce some symbols used in the sequel. θ_i is the trait level of examinees, i.e. the ability parameter for $i = 1, 2, \cdots, n$. a_j and b_j are respectively the discrimination parameter and the difficulty parameter for $j = 1, 2, \cdots, m$. L_j means the full mark value on the j item for $j = 1, 2, \cdots, m$. x_{ij} is the score of the i examinee on the j item. P_{ij} is used to indicate the scoring rate of the i examinee on the j item. Therefore, according to IRT, we have

$$P_{ij} = \frac{1}{1 - e^{-Da_j(\theta_i - b_j)}}. \tag{2.1}$$

As mentioned above, the parameter estimate method of the previous model has gone through a long time of development and improvement. At present, the most frequently adopted is the least squares estimate method. In some literatures, the detailed parameter estimate is also presented, and the following parameter estimates are needed [1]:

$$\begin{cases} \widehat{a}_j = -\dfrac{\sum\limits_{i=1}^{n} y_{ij}\widehat{\theta}_i^{(0)}}{D\sum\limits_{i=1}^{n}\left[\widehat{\theta}_i^{(0)}\right]^2}, \\ \widehat{b}_j = -\dfrac{\sum\limits_{i=1}^{n} y_{ij}}{Dn\widehat{a}_j}, \\ \widehat{\theta}_j = -\dfrac{\sum\limits_{j=1}^{n} z_{ij}\widehat{a}_j}{D\sum\limits_{j=1}^{n}[\widehat{a}_j]^2}, \end{cases} \quad \text{for} \quad j = 1, 2, \cdots, m, \tag{2.2}$$

with

$$\begin{cases} y_{ij} = \ln\left(\dfrac{1}{\widehat{P}_{ij}} - 1\right), \\ z_{ij} = D\widehat{a}_j\widehat{b}_j, \end{cases} \tag{2.3}$$

where $\widehat{\theta}_i^{(0)}$ is the original value of the examinee's ability and $\widehat{P}_{ij}$ is the scoring rate of the examinee are obtained from original samples.

Finally, $\widehat{\theta}_i$ is implemented the transformation of the standardization. This results in the new estimate values of θ_i, which is substituted into the above-mentioned model. By the analogous way, we obtain the new estimate values of $\widehat{a}_j$ and $\widehat{b}_j$, which are just the item parameters. If we iterate in this way repeatedly for ℓ times, they will satisfy the conditions.

3. PARAMETER ESTIMATE MATRIX

In order to simplify the results and calculation, and make better use of common software such as Matlab, Mathematica, Spss and other common software, we now rewrite the algebraic forms of the above results into matrix forms.

We introduce some symbols in the form of matrix. Let

$$\begin{cases} \mathbf{A} = \left(\widehat{a}_1, \widehat{a}_2, \cdots, \widehat{a}_m\right), \\ \mathbf{B} = \operatorname{diag}\left\{\widehat{b}_1, \widehat{b}_2, \cdots, \widehat{b}_m\right\}, \\ \mathbf{\Phi} = \left(\widehat{\theta}_1, \widehat{\theta}_2, \cdots, \widehat{\theta}_m\right) \end{cases} \tag{3.1}$$

be respectively the matrix of the discrimination, the difficulty and the ability. We respectively denote the observation matrix of the scoring rate, the matrix of the examinees scoring, the transition matrix of the scoring rate and the transition matrix of the ability by

$$\mathbf{P} = (\widehat{P}_{ij})_{n\times m}, \ \ \mathbf{X} = (x_{ij})_{n\times m}, \ \ \mathbf{Y} = (y_{ij})_{n\times m}, \ \ \mathbf{Z} = (z_{ij})_{n\times m}. \tag{3.2}$$

Let $\mathbf{\Phi}_0 = \left(\widehat{\theta}_1^{(0)}, \widehat{\theta}_2^{(0)}, \cdots, \widehat{\theta}_m^{(0)}\right)$ be the initial matrix of the ability. And the transition matrix of the discrimination is written as

$$\mathbf{A}_1 = \begin{pmatrix} \widehat{a}_1 & \widehat{a}_1 & \cdots & \widehat{a}_1 \\ \widehat{a}_2 & \widehat{a}_2 & \cdots & \widehat{a}_2 \\ \cdots & \cdots & \cdots & \cdots \\ \widehat{a}_m & \widehat{a}_m & \cdots & \widehat{a}_m \end{pmatrix}_{m\times n}. \tag{3.3}$$

The system of equations

$$w_j = \sum_{i=1}^{n} u_{ij} v_i, \quad j = 1, 2, \cdots, m \tag{3.4}$$

can be rewritten into the matrix form

$$\mathbf{W} = \mathbf{V}\mathbf{U} \tag{3.5}$$

with

$$\mathbf{W} = (w_1, w_2, \cdots, w_m), \ \ \mathbf{V} = (v_1, v_2, \cdots, v_n), \ \ \mathbf{U} = (u_{ij})_{n\times m}. \tag{3.6}$$

Analogously, the system of equations

$$w_j = \sum_{i=1}^{n} v_i^2, \quad j = 1, 2, \cdots, m \tag{3.7}$$

can be rewritten into the matrix form

$$\mathbf{W} = \mathbf{V}\mathbf{V}^T. \tag{3.8}$$

Theorem 3.1. *If the parameters of discrimination, difficulty and examinee's ability are expressed by*

$$\begin{cases} \widehat{a}_j = -\dfrac{\sum\limits_{i=1}^{n} y_{ij}\widehat{\theta}_i^{(0)}}{D\sum\limits_{i=1}^{n}\left[\widehat{\theta}_i^{(0)}\right]^2}, & j = 1, 2, \cdots, m, \\ \widehat{b}_j = \dfrac{\sum\limits_{i=1}^{n} y_{ij}}{Dn\widehat{a}_j}, & j = 1, 2, \cdots, m, \\ \widehat{\theta}_j = -\dfrac{\sum\limits_{j=1}^{n} z_{ij}\widehat{a}_j}{D\sum\limits_{j=1}^{n}[\widehat{a}_j]^2}, & i = 1, 2, \cdots, n, \end{cases} \tag{3.9}$$

one has

$$\begin{cases} \mathbf{A} = -\dfrac{\Phi_0\mathbf{Y}}{D(\Phi_0\Phi_0^T)}, \\ \mathbf{B} = \dfrac{\mathbf{A}_1\mathbf{Y}}{Dn}, \\ \Phi = -\dfrac{\mathbf{A}\mathbf{Z}^T}{D(\mathbf{A}\mathbf{A}^T)}. \end{cases} \tag{3.10}$$

4. INSTANCE ANALYSIS

4.1. **The sample and its initial values.** Now, we use the sampled data of Higher Mathematics Competition in Huainan Union University in 2017 as a sample. The latter six questions in this test paper are polytomous scoring questions. The full marks for each of the first five questions are 10 points and the full mark for the last question is 8 points. We extract 10 test papers at random, i.e. $m = 6$ and $m = 10$.

Except for special reaction models, one has $P_{ij} \in (0,1)$. In order to gear to actual circumstances, we use the following formulas respectively

$$\widehat{P}_{ij} = \begin{cases} 0.001, & x_{ij} = 0, \\ \dfrac{x_{ij}}{L_j}, & x_{ij} \in (0, L_j), \\ 0.999, & x_{ij} = L_j \end{cases} \tag{4.1}$$

and

$$\widehat{\theta}_i^{(0)} = \ln \frac{R_i}{L - R_i}, \tag{4.2}$$

where L_i is the total scores of m items, and R_i is the total score of the i examinee on m items.

4.2. **Estimating item parameters.** We obtain the matrix of the examinees scoring $\mathbf{X}$ from the scores on the test papers, and obtain the observation matrix $\mathbf{P}$ of the scoring rates by calculation according to the following formula

(4.3)

$$\mathbf{X} = \begin{pmatrix} 2 & 5 & 0 & 7 & 9 & 6 \\ 1 & 9 & 1 & 6 & 8 & 4 \\ 2 & 8 & 1 & 9 & 6 & 6 \\ 2 & 10 & 5 & 6 & 9 & 1 \\ 2 & 9 & 7 & 7 & 5 & 4 \\ 1 & 9 & 6 & 9 & 7 & 7 \\ 2 & 8 & 6 & 10 & 6 & 7 \\ 2 & 10 & 6 & 8 & 7 & 8 \\ 9 & 9 & 4 & 8 & 8 & 8 \\ 8 & 9 & 5 & 10 & 9 & 8 \end{pmatrix}, \mathbf{P} = \begin{pmatrix} 0.1 & 0.9 & 0.1 & 0.6 & 0.8 & 0.5 \\ 0.2 & 0.5 & 0.001 & 0.7 & 0.9 & 0.75 \\ 0.2 & 0.8 & 0.1 & 0.9 & 0.6 & 0.75 \\ 0.2 & 0.9 & 0.7 & 0.7 & 0.5 & 0.5 \\ 0.2 & 0.999 & 0.5 & 0.6 & 0.999 & 0.125 \\ 0.1 & 0.9 & 0.4 & 0.9 & 0.7 & 0.875 \\ 0.2 & 0.8 & 0.6 & 0.999 & 0.6 & 0.875 \\ 0.2 & 0.999 & 0.6 & 0.8 & 0.7 & 0.999 \\ 0.9 & 0.9 & 0.4 & 0.8 & 0.8 & 0.999 \\ 0.8 & 0.9 & 0.5 & 0.999 & 0.9 & 0.999 \end{pmatrix}.$$

By (4.2), we easily gets the initial matrix of ability

(4.4)

$$\mathbf{\Phi}_0 = \begin{pmatrix} 0.000 & 0.000 & 0.208 & 0.348 & 0.348 & 0.719 & 0.719 & 0.880 & 1.344 & 1.695 \end{pmatrix}.$$

Then we have the transition matrix $\mathbf{Y}$ of the scoring rates

(4.5)

$$\begin{pmatrix} 2.197 & 1.386 & 1.386 & 1.386 & 1.386 & 2.197 & 1.386 & 1.386 & -2.197 & -1.386 \\ -2.197 & 0 & -1.386 & -2.197 & -6.907 & -2.197 & -1.386 & -6.907 & -2.197 & -2.197 \\ 2.197 & 6.907 & 2.197 & -0.847 & 0 & -0.405 & -0.405 & -0.405 & 0.405 & 0 \\ -0.405 & -0.847 & -2.197 & -0.847 & -0.405 & -2.197 & -6.907 & -1.386 & -1.386 & -6.907 \\ -1.386 & -2.197 & -0.405 & 0 & -6.907 & -0.847 & -0.405 & -0.847 & -1.386 & -2.197 \\ 0 & -1.099 & -1.099 & 0 & 1.946 & -1.946 & -1.946 & -6.907 & -6.907 & -6.907 \end{pmatrix}.$$

Finally, the matrix of item discrimination $\mathbf{A}$ and the matrix of difficulty $\mathbf{B}$ are respectively expressed by

$$\mathbf{A} = \begin{pmatrix} 0.0220 & 1.6316 & 0.0200 & 1.9305 & 0.8442 & 2.5549 \end{pmatrix}, \tag{4.6}$$

$$\mathbf{B} = \begin{pmatrix} 24.4037 & -0.9940 & 28.0839 & -0.7156 & -1.2731 & -0.5725 \end{pmatrix}, \tag{4.7}$$

where the transition matrix is

$$\mathbf{A}_1 = \begin{pmatrix} 45.455 & 45.455 & \cdots & 45.455 \\ 0.613 & 0.613 & \cdots & 0.613 \\ 0.518 & 0.518 & \cdots & 0.518 \\ 1.185 & 1.185 & \cdots & 1.185 \\ 0.391 & 0.391 & \cdots & 0.391 \end{pmatrix}_{6\times 10}. \tag{4.8}$$

4.3. **Estimate of ability parameters.** By the simple calculation, we get the transition matrix

(4.9)

$$\mathbf{Z} = \begin{pmatrix} 1.284 & 0.473 & 0.473 & 0.473 & 0.473 & 1.284 & 0.473 & 0.473 & -3.11 & -2.299 \\ 0.56 & 2.757 & 1.371 & 0.56 & -4.15 & 0.56 & 1.371 & -0.45 & 0.56 & 0.56 \\ 1.233 & 5.943 & 1.233 & -1.811 & -0.964 & -1.369 & -1.369 & -1.369 & -0.559 & -0.964 \\ 1.943 & 1.501 & 0.151 & 1.501 & 1.943 & 0.151 & -4.559 & 0.962 & 0.962 & -4.559 \\ 0.441 & -0.37 & 1.442 & 1.827 & -5.08 & 0.98 & 1.422 & 0.98 & 0.441 & -0.37 \\ 2.487 & 1.388 & 1.388 & 2.487 & 4.433 & 0.541 & 0.541 & -4.42 & -4.42 & -4.42 \end{pmatrix},$$

TABLE 1. Actual scoring rate and standard deviation of scores on each item

j	1	2	3	4	5	6
d_j	0.310	0.860	0.410	0.800	0.738	0.740
σ_j	2.737	1.356	2.385	1.414	1.500	2.166

TABLE 2. Total scoring rate of each sampled examinee

i	1	2	3	4	5	6	7	8	9	10
d'_i	0.500	0.500	0.552	0.569	0.586	0.672	0.672	0.707	0.793	0.845

Further, we obtain the matrix of the examinee's ability parameters

$$\boldsymbol{\Phi} = \begin{pmatrix} -0.464 & -0.464 & -0.315 & -0.173 & -0.504 & -0.147 & 0.173 & 0.664 & 0.354 & 0.884 \end{pmatrix}. \tag{4.10}$$

5. **Conclusion**

Under the conditions

$$\begin{cases} \max\left\{|\widehat{a}_j^{(t+1)} - \widehat{a}_j^{(t)}|, |\widehat{b}_j^{(t+1)} - \widehat{b}_j^{(t)}|\right\} < 0.1, \\ |\widehat{\theta}_j^{(t+1)} - \widehat{\theta}_j^{(t)}| < 0.05, \end{cases} \tag{5.1}$$

by iteration, one has

$$\mathbf{A} = \begin{pmatrix} 1.148 & 0.606 & 1.302 & 1.789 & 0.077 & 3.531 \end{pmatrix}, \tag{5.2}$$

$$\mathbf{B} = \begin{pmatrix} 0.467 & -2.677 & 0.436 & -0.772 & -9.100 & -0.414 \end{pmatrix} \tag{5.3}$$

and

$$\boldsymbol{\Phi} = \begin{pmatrix} -0.539 & -0.484 & -0.253 & -0.489 & -0.311 & 0.069 & 0.204 & 0.544 & 0.551 & 0.846 \end{pmatrix}. \tag{5.4}$$

In order to test the rationality and accuracy of the above theoretical results, we analyze the results of parameter estimate and their practical significance. For this purpose, we calculate the actual scoring rate of each item $d_j = \frac{I_j}{nL_j} \times 100\%$, the standard deviation of scores on each item σ_j and the total scoring rate of each sampled examinee $d'_i = \frac{R_i}{L} \times 100\%$, where I_j is the actual total score of each item, as shown in two tables above.

The discussion above shows that the following conclusions are valid:

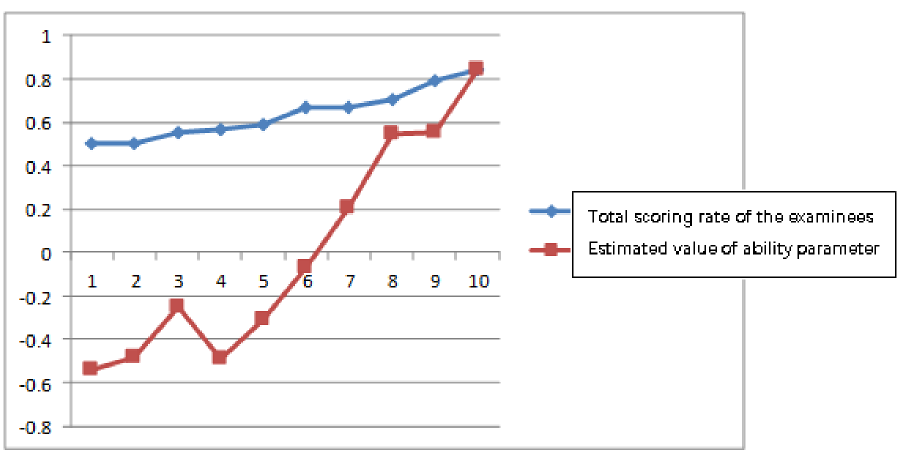
(1) Except for Item 5, when the actual scoring rate b_j of an item is low, the estimated value $\widehat{d_j}$ of the difficulty parameter of the item is positive and big. On the contrary, when the actual scoring rate b_j is high, the estimated value $\widehat{d_j}$ is negative and small. The correlation coefficient is $r \approx -0.4551$, so the degree of correlation is

low. If we remove $\widehat{b}_5$, then $r \approx -0.8927$. And hence the correlation is significantly improved. Evidently, among the six items in this test paper, the difficulty of Item 5 is unreasonable, so it is not the appropriate question to be used in the academic competitions of colleges and universities.

(2) Except for Item 6, when the estimated value $\widehat{a}_j$ of the discrimination parameter of an item is so big that the examinees' abilities are around the corresponding $\widehat{b}_j$ value dispersedly, the standard deviation σ_j of score is also big. On the contrary, when the $\widehat{a}_j$ value is so small that the examinees' abilities are around the corresponding $\widehat{b}_j$ value intensely, the standard deviation of score is small. The correlation coefficient between the discrimination parameter of each item $\widehat{a}_j$ and the standard deviation of score σ_j is $r \approx 0.3101$, so the degree of correlation is low. If we delete the data $\widehat{a}_6$, then $r \approx 0.8323$. Evidently, compared with the other items, the discrimination of Item 6 is big.

(3) Except for Item 4, there is a significant overall synchronous increased relationship between the estimated value $\widehat{\theta}_i$ of each examinees ability parameter and the total scoring rate d'_i of each examinee, and the correlation coefficient between them is $r \approx 0.9572$ while the standard deviation is $\sigma \approx 0.1548$. Evidently, there is a significant correlation between them (as shown in the following figure). Meanwhile, the regression equation between them is represented as

$$d'_i = 4.0632\widehat{\theta}_i - 2.5988.$$



In conclusion, the matrix design of parameter estimate method in this article is more conducive to the practical application of the model. The unity, feasibility and accuracy of the model have been verified in practical applications. The analysis of the irrationality in the difficulty or discrimination in the experiment will be conducive to the improvement of the quality of subsequent question setting and the overall level of the question bank. Meanwhile, we can estimate the overall scoring in the next examination of the same kind according to the regression equation between the ability and the total scoring rate of examinees. This provides the way for the rationality of talent selection.

Acknowledgements The authors acknowledge the Natural Science Foundation of Anhui Province in China (Grant: KJ2017A583).

REFERENCES

[1] Cheng Huangjin, Wang Yongguo, IRT-based Parameter Estimation in a Polytomous Scoring Model, *College Mathematics*, **26**(5): 156–160, 2010.

[2] Njål Foldnes, Steffen Grønneberg, On Identification and Non-normal Simulation in Ordinal Covariance and Item Response Models, *Psychometrika*, **84** (4):1000–1017, 2019.

[3] Merkle Edgar C, Furr Daniel, Rabe-Hesketh Sophia, Bayesian Comparison of Latent Variable Models: Conditional Versus Marginal Likelihoods, *Psychometrika*, **84** (3): 802–829, 2019).

[4] Joo Seang-Hwane, Chun Seokjoon, Stark Stephen, Chernyshenko Olexander S, Item Parameter Estimation With the General Hyperbolic Cosine Ideal Point IRT Model, *Applied psychological measurement*, **43** (1):18–33, 2019.

[5] Huang Huacai, Ding Shuliang, Luo Fen, SQRT/EM Method of Estimating Parameters in IRT and Its Appilications, *Journal of Jiangxi Normal University*, **29**(3):231–234, 2005.

[6] Insu Paek, Sun-Geun Baek, Mark Wilson, An IRT modeling of change over time for repeated measures item response data using a random weights linear logistic test model approach, *Asia Pacific Education Review*, **13**(3):487–494, 2012.