Item selection criteria for Logistic Positive Exponent model-based Computerized Adaptive Testing

Thales Akira Matsumoro Ricarte*
ICMC-USP/DEs-UFSCAR, São Carlos, Brazil - thalesamr@icmc.usp.br

Mariana Cúri
ICMC-USP, São Carlos, Brazil - mcuri@icmc.usp.br

Abstract

In a Computerized Adaptive Test (CAT), the items of the evaluation are selected from a bank according to the examinee's abilities in order to avoid to present items that are too easy or too dificult to him/her. In this way, we can interpret that the item selection criteria used in a CAT would choose items with correct response probability close to 50%. Considering symmetric Item Response Theory (IRT) models and Fisher information function for item selection, this is equivalent to match difficulty levels of test items with an examinee's ability. However, for asymmetric models and other item selction criteria, this equivalence may not be true. The aim of this work is to analyse three item selection methods (Fisher Information criterium, Kullback-Leibler information criterium and Continuous Entropy Method) under an asymmetric IRT model (logistic positive exponent) in terms of the difficult parameters and probability of correct response of the selected item. The simulations show that the selected items do not have a 50% probability of correct response in either of the three methods. In this way, further studies are proposed to indicate an adequate item selection criteria for LPE model-based CAT.
Keywords: Item Response theory; Fisher Information Function; Kullback Leibler information function; Continuous Entropy Method.

1. Introduction

The aim of Computerized Adaptive Tests (CAT) is to administer a personalized test, selecting items from a bank according to the examinee's abilities. The most common item selection procedure has been based in maximizing item information. Functions based on Fisher and Kullback-Leibler information are the most commonly used for item selection when Item Response Theory (IRT) models are considered (Chang & Ying, 1996) . As well as Continuous Entropy method is often the objective function for cognitive diagnostic CAT (Xu, Chang, & Douglas, 2005; Cheng, 2009). An effective evaluation should present items neither too difficult nor too easy for examinees (Lord, 1970), i.e., individuals would have approximately 50% probability to give the correct response for the item in the cases of dichotomous models. In the theoretical foundation of IRT model-based CAT, this fact is used as a synonym of matching difficulty levels of test items with an examinee's ability. However, 50% probability of correct response implies the equality of difficulty and ability parameters only in symmetric IRT models.

Nevertheless, in some cases, symmetric models are not appropriate for the item characteristic curve. For example, when the limit of the response probability tends to 0 and 1 at different rate (Chen, Dey & Shao; 1999) and the mean ability estimates can be biased if an innapropriete model is considered (Czado & Santner, 1992). In practice, many assessments use asymmetric models, the 3 Parameter Logistic Model is implemented Test of English as a Foreign Language (TOEFL; Educational Testing Service, 2015) and the Armed Services Vocational Aptitude Battery (ASVAB; U.S. Department of Defense, 1984). Samejima (2000) proposed the asymmetric Logistic Positive Exponent (LPE) family, that generalises the Logistic IRT models and provides a better methodology for understading the complex human behaviors. CAT based on asymmetric IRT models can be criticized because maximizing Fisher and Kullback-Leibler information functions may result in administering items with too low or too high probability of success.

In this paper, our objective was to analyse selection methods based on Fisher, Kullback-Leibler Informations and the Continuous Entropy under LPE models and to verify what are the probability of correct response of the selected items (if they are close to 50%, according to Lord, 1970, or not).

2. Methodology

In this section, the LPE model and the three item selection criteria considered in ths work are presented.

2.1 Model

The Logistic Positive Exponent model with 4 item parameters is given by

$$P(X=1 \mid a,b,c,\lambda,\theta) = c + (1-c)\left[\frac{1}{1+\exp(-a(\theta-b))}\right]^{\lambda}, \quad (1)$$

where $P(X=1 \mid a,b,c,\lambda,\theta)$ is the probability of the examinee to correctly respond the item, a and b are the discrimination and dificulty item parameters, respectively, c is the guessing parameter, and λ is the acceleration parameter.
For notation simplicity, $\frac{1}{1+\exp(-a(\theta-b))}$ and $P(X=1 \mid a,b,c,\lambda,\theta)$ are going to be denoted as $p_1(\theta)$ e $p_2(\theta)$, respectively.

2.2 Fisher Information

The Fisher Information (F) function for an item based on the LPE model is given by

$$F(\theta) = \frac{[a(1-c)\lambda p_1(\theta)^{\lambda}(1-p_1(\theta))]^2}{p_2(\theta)(1-p_2(\theta))}, \quad (2)$$

The selected item maximizes the Fisher Information for the current estimated θ value.

2.3 Kullback-Leibler Information

The Kullback-Leibler (KL) information for an item based on the LPE model is given by

$$KL(\theta_0,\theta) = \int_{\theta_0-\sigma}^{\theta_0+\sigma}\left[p_2(\theta)\log\left(\frac{p_2(\theta)}{p_2(\theta_0)}\right) + (1-p_2(\theta)), \log\left(\frac{1-p_2(\theta)}{1-p_2(\theta_0)}\right)\right]d\theta, \quad (3)$$

where θ_0 if a fixed value of θ and σ assumes a value of a positive funtion that decreases as more itens are administered in the test. Chang and Ying (1996) recomended that $\sigma = \frac{r}{\sqrt{k}}$, where k is the number of items administered so far and r a constant, in our case set to 1.5. The item with maximum the Kullback-Leibler information is selected.

2.4 Continuous Entropy Method

Continuous Entropy (CE) method is an adaptation of Shannon Entropy (Shannon, 1948) for continuous random variables. Let $\mathbf{X}_{k-1} = \{X_1, ..., X_{k-1}\}$ be a vector of the responses $X_i, i \in \{1, ..., k-1\}$, after $k-1$ items administered in the test. The response X_i assumes value 1 if the examinee chose the correct response for the item i, otherwise it assumes value 0 . The Continuous Entropy function for LPE model after $k-1$ items administered in the test is written as

$$CE(\theta \mid \mathbf{X}_{k-1}) = -\int \pi(\theta \mid \mathbf{X}_{k-1})\log(\pi(\theta \mid \mathbf{X}_{k-1}))d\theta, \quad (4)$$

where $\pi(\theta \mid \mathbf{X}_{k-1}) \propto \prod_{i=1}^{k-1} p_2^{X_i}(1-p_2)^{1-X_i} f(\theta)$ is the posterior distribution and $f(\theta)$ is the priori of θ.
CE reaches its minimum value when the distribution $\pi(\theta \mid \mathbf{X}_{k-1})$ is concentrated on a single point i.e., $\pi(\theta=\theta_0)=1$ and $\pi(\theta \neq \theta_0)=0$ and its maximum value when θ has a uniform distribution.
Since the individual response for the kth item (X_k) is unknown, it is necessary to use the expected posterior continuous entropy (ECE) which is written as

$$ECE(\theta) = -\sum_{x=0}^{1}\int \pi(\theta \mid \mathbf{X}_{k-1}, X_k=x)\log(\pi(\theta \mid \mathbf{X}_{k-1}, X_k=x))P(X_k=x \mid \mathbf{X}_{k-1})d\theta, \quad (5)$$

where $P(X_k = x \mid \mathbf{X}_{k-1})$ is the posterior predictive distribution. The item that minimize the expected posterior continuous entropy is selected.

3. Simulation results

Figures 1, 2 and 3 present th Fisher and Kullback-Leibler Informations and Continuous Entropy Method versus the b parameter. The other parameters (a, θ, c and λ) will be assigned with fixed values in these graphics.

For each function it will be shown four graphics, each of them will present three different information curves as one of the four parameters (a, θ, c or λ) will assume three different fixed values, while the others will be attributed a single value. In the first graphic of each figure (Figures 1(a), 2(a) and 3(a)), the a item parameter will be assigned the values of 0.3, 1 and 1.7, otherwise $a = 1$. In the second graphics (Figures 1(b), 2(b) and 3(b)) θ will be assigned the values of -2, 0 and 2, otherwise $\theta = 0$. Analogously, in the third graphics (c), the c parameter will be assigned the values of 0, 0.2 and 0.5, otherwise $c = 0$. Finally, in the last graphics (d) the λ parameter will be assigned the values of 0.5, 1 and 2. otherwise $\lambda = 0$.

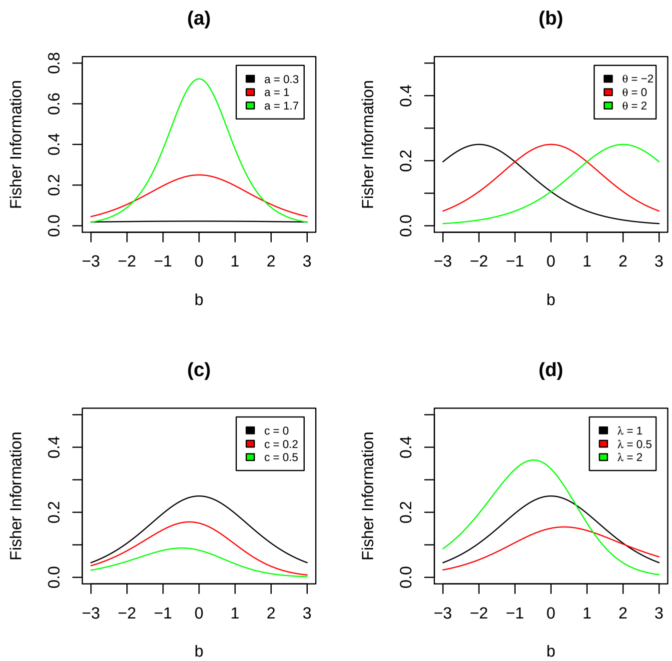

Figure 1: Fisher Information vs b.

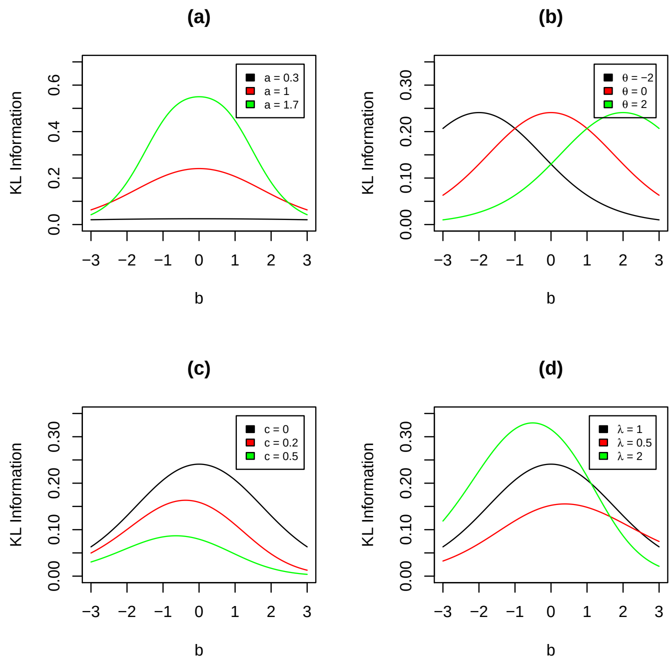

Figure 2: Kullback-Leibler Information vs b.

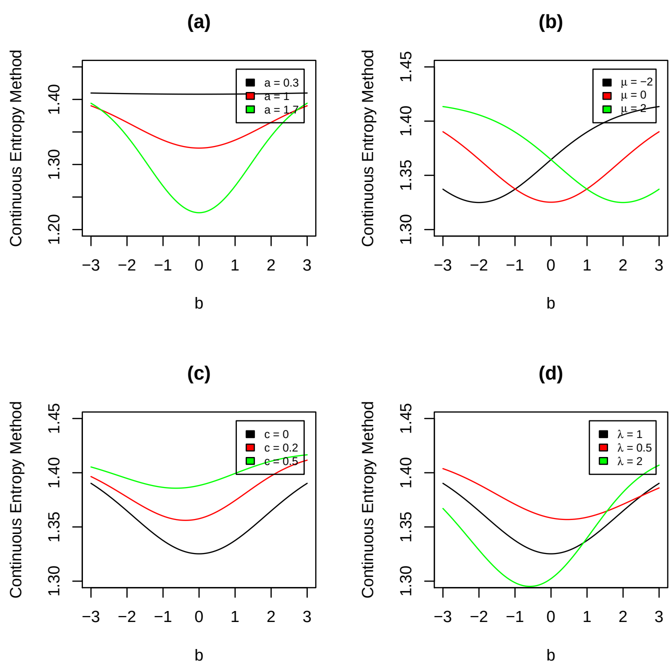

Figure 3: Continuous Entropy Method vs b.

Figures 1(a), 2(a) and 3(a) shows that if the values of b and θ ($\theta = 0$ in this case) are close, the functions increase (or decreases, for CE) as the a parameter increases. However, if b and θ are distant between each

other, this relationship disappears.
In Figures 1(b) and 2(b), as expected for the one parameter logistic model, the FI and KL information functions has higher values as the difference between b and θ are lower. As well as, in Figure 3(b), the CE function has lower value as this difference are lower.
In Figures 1(c), 2(c) and 3(c), we note that one item has less desirable to be selected if its c parameter is higher. We also observe that if all items had the same a, $c > 0$, λ and for $\theta = 0$ the item with the highest function value would have b lower than 0. The probability of the individual to select the correct response for $c = 0.2$ is 0.6537, 0.6732 and 0.6732 for Fisher, Kullback-Leibler and CEM, respectively, and for $c = 0.5$ is 0.8089, 0.8273 and 0.8273 for Fisher, Kullback-Leibler and CEM, respectively. It is worth noting that, for $\sigma = 1.5$ Kullback-Leibler and CEM selects the same item, but when σ decreases, Kullback-Leibler tends to be more likely the Fisher Information.
In Figures 1(d) and 2(d), observe that if b is lower than θ the functions increase as λ increases. In Figure 3(d), if b is lower than the mean value of θ (μ), CE function decreases as λ increases. On the other hand, if b for is greater than θ (or μ) we are not able to stablish that relationship anymore because the curves are crossing each other. Note that for $\theta = b$, the item with higher value of λ has higher information (or lower CE value). If all items had the same $a = 1$, $c = 0$, $\lambda = 0.5$ and for $\theta = 0$ the item with the highest function value (or lowest CE value) would have b higher than 0 and the probability of correctly chose the right response of the selected item being 0.6410, 0.6353 and 0.6240 for Fisher, Kullback-Leibler and CE, respectively. In the same condition as before, but with $\lambda = 2$ the item with the highest function value would have b lower than 0 and the probability of correctly chose the right response of the selected item being 0.3816, 0.3904 and 0.4139 for Fisher, Kullback-Leibler and CE, respectively.

To analise the parameter interactions, the objective functions for each item selection criteria were made fixing the parameter values as: a=0.3, 1 and 1.7, θ=-2, 0 and 2 (or μ=-2, 0 and 2), c=0, 0.2 and 0.5, and λ=0.5, 1 and 2. All the 81 combinations (3^4) were considered (the graphics are not shown here).
For all objective functions, $a = 0.3$ resulted in less distinguishable curves, making more dificulty to identify which item is going to be selected for each case. If $a = 1.7$, we observe the opposite effect.
For $c \in \{0, 0.2, 0.5\}$, $a \in \{0.3, 1, 1.7\}$ and $\lambda = 1$, items with higher c values are less likely to be selected. Items with $c > 0$ have the information functions maximized (or minimized, for CE method) for $b < \theta$. Additionally, the greater the a parameter is, the farther the distance between b and θ will be.
For $\lambda \in \{0.5, 1, 2\}$, $a \in \{0.3, 1, 1.7\}$ and $c = 0$, we have the same pattern as in Figures 1(d), 2(d) and 3(d). More specifically, for $c = 0$, $a \in \{0.3, 1, 1.7\}$ and $\lambda = 0.5$, items have the information functions maximized (or minimized, for CE method) for $b > \theta$. The greater the a parameter is, the farther the distance between b and θ will be. Analogouslly, for $c = 0$, $a \in \{0.3, 1, 1.7\}$ and $\lambda = 2$, items have the information functions maximized (or minimized, for CE method) for $b < \theta$, and again the greater the a parameter is, the farther the distance between b and θ will be.

4. Conclusions

The results show all the three item selection criteria favor to administer items that are not at 50% correct response for the adopted asymmetric models, in contradiction to Lord (1970) idea that too difficult or too easy items should not be presented to an individual. For the three parameter logistic model (λ=1 in equation (1)), the probability of correct response of the selected item is closer to 50% adopting Fisher Information criteria than the other two methods. As well as the LPE model, the probability of correct response of the selected item is closer to 50% adopting CE method than the others.
Future studies should be made in order to evaluate the properties of the estimates of θ in a CAT comparing F, KL and CE methods of item selection criteria and indicating the best one (or some adaptation o it) for the LPE model. The correct response probability and the exposure rate of the items should also be included in these simulations. The exposure rate is relevant because it is also expected that the items with higher λ and a and lower c values are probably be administered in a test based on the LPE model, as indicated in the results.

References

Shannon, C.E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379 −

423.

Lord, M. E (1970). Some test theory for tailored testing. In W.H. Holtzman (Ed.), Computer-assisted instruction, testing and guidance. New York: Harper and Row.

U.S. Department of Defense. (1984). Test manual for the Armed Services Vocational Aptitude Battery (DoD 1340.12AA). North Chicago, IL: U.S. Military Entrance Processing Command.

Czado, C., & Santner, T. J. (1992). The effect of link misspecification on binary regression inference. Journal of Statistical Planning and Inference, 33, 213231.

Chang, H.H., & Ying, Z.L. (1996). A global information approach to computerized adaptive testing. Applied Psychological Measurement, 20(3), $213-229$.

Chen, M. H., Dey, D. K., & Shao, Q.-M. (1999). A new skewed link model for dichotomous quantal response data. Journal of the American Statistical Association, 94, 11721186.

Samejima, F. (2000). Logistic positive exponent family of models: Virtue of asymmetric item characteristics curves. Psychometrika, 65, 319335.

Xu, X., Chang, H., & Douglas, J. (2005). Computerized adaptive testing strategies for cognitive diagnosis. Paper presented at the annual meeting of national council on measurement in education, Montreal, Canada.

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. Psychometrika, 74(4), $619-632$.

Educational Testing Service (2015), Test of English as a Foreign Language (TOEFL), Princeton, New Jersey.