IMPROVEMENTS IN VALIDITY WITH NEW BAYES ESTIMATION IN THE ITEM RESPONSE THEORY MODEL

KAZUMASA MORI

Bunkyo University
E-mail: morik@shonan.bunkyo.ac.jp

Abstract - In this study, we propose a simple Bayes-LINEX estimation of an item response model (IRT) under asymmetric loss and improved validity of ability estimates in educational test data. More precisely, using the Markov Chain Monte Carlo (MCMC) method, we propose a simple Bayesian estimator of the Rasch model under LINEX loss. [8] We call this estimation a simple Bayes-LINEX estimation (BLE). The BLE was applied to a total dataset for adult literacy [1], and we found that estimates using the ability parameter with the simple BLE showed better validity than the estimates using the conventional Bayesian estimation. We conclude that our simple BLE is a suitable solution for the educational test data with respect to validity.

Index Terms - Asymmetric loss, educational testing, item response theory, validity..

I. INTRODUCTION

In this study, we propose a simple Bayes estimation of the item response model(IRT) under asymmetric loss and improve the validity of ability estimates in educational test data.

As an application of the IRT model, we use a Rasch model [6]. Let $X_{ij} = \{0,1\}$ be the response of thei-th subject in j-th item. Then, the probability function of the Rasch model is

$$P(X_{ij} = 1|\theta,\delta) = \frac{\exp(\theta_i-\delta_j)}{1+\exp(\theta_i-\delta_j)}, \quad (1)$$

where θ_i is the ability parameter of thei-th subject, and δ_j is the difficulty parameter of the j-thitem.We estimate the parameter of the Rasch model using the maximum likelihood estimation or the Bayesian estimation. Here, we focus on the Bayesian estimation. In Bayesian estimation, we regard (1) as the likelihood function and set a prior distribution like:

$$\theta_i \sim N(0,1), \delta_j \sim N(0,1). \quad (2)$$

To calculate the estimator, we typically use the Markov Chain Monte Carlo (MCMC) method such as the MH-algorithm [4][2].

Although IRT models are useful in evaluating educational testing, they have some disadvantages. One of the most severe disadvantages is the problem of overestimation. [7] noted that parameters of the IRT model are overestimated if the correlations of the response matrix are high. This situation usually occurs when the educational test has a "testlet" within it. To prevent overestimation, many methods have been proposed. For example, [9] proposed a special "testlet" parameter in the IRT model. However, we can hardly explain the meaning of this parameter. In this study, we propose a new approach for the overestimation problem. In our approach, we use asymmetric loss to the IRT model. More precisely, we use LINEX loss [8]:

$$L(\hat{\vartheta} - \vartheta) = b\left[\exp\left\{a\{\hat{\vartheta} - \vartheta\}\right\} - a\{\hat{\vartheta} - \vartheta\} - 1\right], a \neq 0, b > 0, \quad (3)$$

where L is the loss function, $\vartheta,\hat{\vartheta}$ are the parameter and estimator, and a, b are setting values for the loss function. Setting a value controls the overestimation or underestimation of the estimator. If a<0, this loss function penalizes overestimation. On the other hand, if a>0. this loss function penalizes underestimation.

In this study, we propose a simple Bayes estimation of the Rasch model under LINEX loss and improve the validity of the estimates of ability in educational test data.The remainder of the paper is organized as follows: Our new estimator (Bayes-Linex estimator) is proposed in section 2. The case study of educational testing and improvement of validity is shown in section 3. The conclusion and discussion are given in section 4.

II. THE SIMPLE BAYES-LINEX ESTIMATION

[10] showed the general form of the Bayesian estimator under LINEX loss:

$$\widehat{\vartheta_B} = -\left(\frac{1}{a}\right)\log\{E_\vartheta \exp(-a\vartheta)\}, \quad (4)$$

where E_ϑ is posterior expectation with respect to $P(\vartheta|X)$, X is general sample and prior information. Zellner is a derived special case of (4) in which posterior distribution follows normal distribution:

$$\widehat{\vartheta_B} = m - av/2, \quad (5)$$

where m is the posterior mean and v is the posterior variance. Focusing on (5), we derive a simple Bayes estimator of the Rasch model under asymmetric loss:

1. From (1) and (2), derive a conventional Bayes estimator with the MCMC method.
2. The distribution of θ_i is calculated using the MCMC method as with normal distribution.
3. Finally, derive LINEX the estimator based on the Bayes estimator:

$$\widehat{\theta_{iB}} = m_{\theta_i} - av_{\theta_i}/2, \quad (6)$$

where m_{θ_i} and v_{θ_i} denote the posterior mean and posterior variance calculated using the MCMC method. We can estimate δ_j by the same procedure. Note that we must derive m_{θ_i} and v_{θ_i} in (6) for each subject i.

We call this procedure a simple Bayes-LINEX

estimation.

III. CASE STUDIES: IMPROVEMENT IN THE VALIDITY OF EDUCATIONAL TEST DATA

In the case study, we use a total dataset for adultliteracy [1]. As this dataset includes several test datasets, we select two literacy test datasets and IQ datasets (the Wechsler Adult Intelligence Scale; WAIS). Then, weestimate the ability parameter of each examinee for the two datasets using Bayes estimation (BE) and Bayes-LINEX estimation (BLE). Finally, we evaluate the validity of these estimations and the WAIS score. For the BLE, we apply a=5, 3, 1, -1, -3, -5 in (3) and, for the BE, we use Stan package in R software.

A. The GED test dataset

First, we consider the relationship between the General Educational Development (GED) test and the normalized WISE.We show the relationship between the normalized WISE score and the BE/BLE in Fig.1 and the correlation coefficients between the normalized WISE score and the BE/BLE in Table 1. From Table 1, we find that the correlations between normalized WISE and the BLE are higher if the setting value is smaller. In fact, the correlation coefficient r=0.5532, that is between the normalized WISE and the BLE, is the best result in this analysis. This implies that conventional estimations including the BE overestimated and worsened the validity of the parameter estimation. In this case, the BLE improved the overestimation and validity of the ability parameter estimation.

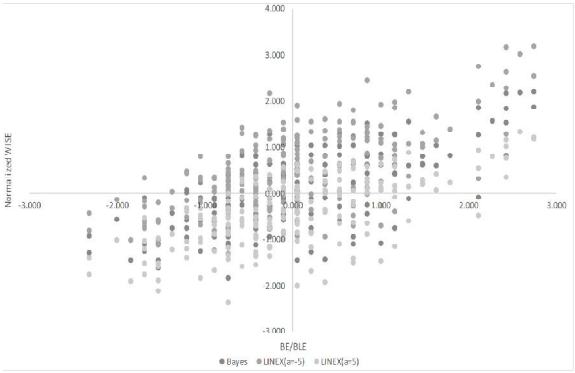

Fig1. Relationship between the normalized WISE and BE/BLE (GED dataset)

Table1. Correlation coefficients between the normalized WISE and BE/BLE (GED dataset)

	BE	BLE (a=5)	BLE (a=3)	BLE (a=1)	BLE (a=-1)	BLE (a=-3)	BLE (a=-5)
Correlation with WAIS	0.5510	0.5433	0.5470	0.5499	0.5520	0.5531	0.5532

B. The NAEP test dataset

Then, we consider the relationship between the National Assessment of Educational Progress (NAEP) test and the normalized WISE. As with the GED case, we show the relationship between the normalized WISE score and the BE/BLE in Fig. 2 and the correlation coefficients between the normalized WISE score and the BE/BLE in Table 2. From Table 2,we find that the correlations between the normalized WISE and the BLE are higher than correlation between the normalized WISE and the BE if the setting value is smaller.Therefore, a lower setting value improves the validity of the estimator of the ability parameter.

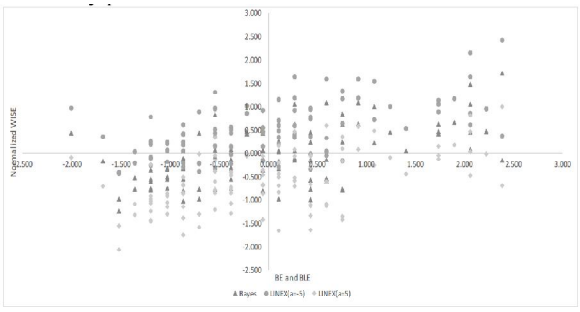

Fig2. Relationship between the normalized WISE and BE/BLE (NAEP dataset)

Table2. Correlation coefficients between the normalized WISE and BE/BLE (NAEP dataset)

	BE	BLE (a=5)	BLE (a=3)	BLE (a=1)	BLE (a=-1)	BLE (a=-3)	BLE (a=-5)
Correlation with WAIS	0.6711	0.6646	0.6693	0.6725	0.6725	0.6745	0.6754

C. Results of the data analysis

In this analysis, we improved the validity of two datasets (GED, NAEP) using the BLE. As we improved both datasets by setting the value smaller, conventional estimation such as the BE resulted in overestimation in both datasets. The reason the degree of improvement is not large is that the BE did not cause severe overestimation.

Overall, our simple BLE is a better method that improves the validity of the estimator of the ability

parameter in the Rasch model compared to conventional estimation.

CONCLUSION AND FURTHER STUDIES

In this study, we proposeda simple Bayes estimation of the Rasch model under LINEX loss and improvedthe validity of estimates of ability in educational test data. As a result, we found that the estimates of the ability parameter with the simple BLE showed better validity than the estimates using conventional Bayesian estimation. We concluded that our simple BLE is a suitable solution for the educational test data with respect to validity.

We have identified the potential for further studies First, as [7] noted, we must evaluate a "testlet" type dataset. The testlet is a type of group of related items. A testlet type dataset is composed of some datasets. If we apply the BLE to a testlet dataset, it will show excellent results. We should also extend the BLE. For example, we will apply the BLE to a polytomous IRT model such as the partial credit model [3] and the rating scale model[5].

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 16K01058.

REFERENCES

[1] Daryl, M. (2013). Improving Literacy Instruction for Adults: Reading Assessments from Adult Education Learners in Kansas and Missouri, 2005-2008. ICPSR34887-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-11-25.

[2] Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika, 57, 97-109.Masters, G.N. (1982). A Rasch Model for Partial Credit Scoring. Psychometrika, 47, 149-174.

[3] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller. E. (1953). Equations of State Calculations by Fast Computing Machines. Journal of Chemical Physics, 21, 1087-1092.

[4] Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Danish Institute for Educational Research.

[5] Rasch, G. (1961). On General Laws and the Meaning of Measurement in Psychology. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV. Berkeley, California: University of California Press. 321–334.

[6] Thissen, D., Steinberg, L. and Mooney, J. A. (1989) ，Trace Lines for Testlets - A Use of Multiple-categorical-response models. Journal of educational measurement, 26, 247-260.

[7] Varian, H.R. (1975). A Bayesian Approach to Real Estate Assessment. In Studies in Bayesian Economics and Statistics in Honor of Leonard J. Savage, Fienberg, S.E. and Zellner, A. (eds.) Amsterdam: North Holland. 195-208.

[8] Wainer, H., Bladlow, E., and Wang. X. (2007). Testlet Response Theory and its Applications. Cambridge University Press.

[9] Zellner A. (1986). Bayesian Estimation and Prediction Using Asymmetric Loss Functions. Journal of American Statistical Association, 81, 446-451.

★★★