

ESTIMATION OF ABILITY USING PSEUDOCOUNTS IN ITEM RESPONSE THEORY

Haruhiko Ogasawara*

A method for estimation of ability using pseudocounts in dichotomous item response models is given when associated item parameters are known or estimated by a separate calibration sample of examinees with the size of an appropriate order. The pseudocount minimizing the asymptotic mean square error is algebraically obtained. Though the pseudocount depends on unknown ability, a fixed lower bound for the pseudocount is derived under the logistic model with equivalent items. The lower bound is numerically shown to be reasonable under the 3-parameter logistic model with and without model misspecification.

1. Introduction

In item response theory (IRT), estimation of the ability or proficiency level of an examinee is one of the main purposes of the associated ability test. Among various estimators of ability, the maximum likelihood estimator (MLE; Lord, 1953) has been a basic one. The maximum a posteriori or Bayes modal estimator (BME; Samejima, 1969, Chapter 2; Bock & Aitkin, 1981) is also familiar, where the standard normal prior is typically used. The so-called weighted likelihood (WL) by Warm (1989) gives the WL estimator (WLE) removing the asymptotic bias of the corresponding MLE when the IRT model is correctly specified. Warm's (1989) method can be seen as a special case of the weighted score or the penalized likelihood (see e.g., Firth, 1993).

Ogasawara (2012a) gave the asymptotic properties of the MLE under possible model misspecification (p.m.m.). Ogasawara (2013a) dealt with the MLE, BME and WLE as special cases of the estimator by the weighted score with a general weight that corresponds to the first derivative of the log prior density with respect to the ability parameter in the case of the BME. Under this unified formulation, Ogasawara (2013a) gave the asymptotic properties e.g., the asymptotic mean square error (AMSE) of the estimator with the general weight.

In the above methods, item parameters are assumed to be known or to have fixed values for estimation of ability. Ogasawara (2013b) gave the asymptotic properties of the ability estimator when the associated item parameters are estimated by a separate sample of examinees with size N for item calibration. Ogasawara (2013b) also showed that when N is of order $O(n^{5/2})$, where n is the number of items, the asymptotic cu-

Key Words and Phrases: 3-parameter logistic model; mean square error; asymptotic bias; asymptotic variance; pseudocount; Bayes modal; ability.

* Otaru University of Commerce

This work was partially supported by a Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, No.23500341.

Author's address: Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501, Japan. Email: hogasa@res.otaru-uc.jp

mulants up to the fourth order and the higher-order asymptotic variance of the ability estimator with the general weight using estimated item parameters are identical to those when the item parameters are known.

In this paper, the dichotomous response model is used, where the $2 \times n$ contingency table for correct and incorrect responses by an examinee is formulated with n items being generally distinct. In the table, there are n 1's and n 0's. The 2×1 collapsed table for the numbers of correct and incorrect responses can be obtained, where the sum of the two frequencies is n , though the number-correct score is not necessarily the sufficient statistic for the ability parameter unless special models e.g., the 1-parameter logistic model (1PLM) are used. As a special case of the 1PLM, the logistic model with equivalent items without guessing parameters (LME; Birnbaum, 1969) reduces to the binomial error model (Lord & Novick, 1968, Chapter 23).

A pseudocount (PC) is an artificial frequency to be added to the cells of the contingency table in order to have "better" estimation of associated parameters. In the case of the binomial proportion with the 2×1 contingency table, various values of the pseudocount have been used, which will be addressed later. The pseudocount method can also be seen as a weighted score method, which gives the asymptotic properties of the ability estimator with the pseudocount.

In this paper, the pseudocount with minimized AMSE is derived. Unfortunately, the optimal pseudocount depends on the population ability to be estimated. However, when the LME holds, a fixed lower bound of the pseudocount is found. The lower bound and other values of fixed pseudocounts are illustrated in the case of the familiar 3-parameter logistic model (3PLM) with and without m.m., where the item parameters are assumed to be known or estimated under the condition of $N = O(n^{5/2})$. Asymptotic and simulated estimation errors in a numerical illustration show that the lower bound obtained under the LME is reasonable even in the case of the 3PLM with and without m.m.

2. Estimation using pseudocounts

Let $U_m (m = 1, \dots, n)$ be the dichotomous variable taking values of 1 and 0 for correct and incorrect responses, respectively to the m -th item by an examinee with ability θ . In the case of the 3PLM

$$P_m \equiv \Pr(U_m = 1 | \theta, a_m, b_m, c_m) = c_m + \frac{1 - c_m}{1 + \exp\{-Da_m(\theta - b_m)\}},$$

$$Q_m \equiv 1 - P_m \quad (m = 1, \dots, n), \quad (2.1)$$

where $D = 1.7$; and a_m, b_m and c_m are the item parameters for the m -th item. The main results in this paper also hold when the item parameters are replaced by the sample counterparts under the condition $N = O(n^{5/2})$ (Ogasawara, 2013b). The notation e.g., fixed a_m rather than estimated \hat{a}_m is for simplicity. Under model misspecification (m.m.), an alternative probability P_{tm} for correct response is defined as

$$P_{Tm} \equiv E_T(U_m), \quad Q_{Tm} \equiv 1 - P_{Tm} \quad (m = 1, \dots, n), \quad (2.2)$$

where $E_T(\cdot)$ is the expectation using the true distribution of U_m at θ and $P_{Tm} \neq P_m$ for at least an item ($E(\cdot)$ will be used under correct model specification (c.m.s.)).

Define c^* as the total number of the pseudocounts for the 2×1 collapsed contingency table, where c^* is not necessarily integer-valued. That is, each cell has $0.5c^*$ for its pseudocount. In the case of the binomial proportion, $c^* = 1$ (0.5 for each cell) was derived by Haldane (1956) and Anscombe (1956) such that the asymptotic bias of order $O(n^{-1})$ for the sample logit vanishes. The value $c^* = 0.5$ (0.25 for each cell) is used by Hitchcock (1962) in the case of logistic regression. Agresti and Coull (1998) (see also Agresti & Caffo, 2000) proposed $c^* = 4$ (2 for each cell) for interval estimation of the binomial proportion. Recently, the value 0.1 of the pseudocount for each cell of the 2×2 table was used by Bonett and Price (2007).

For the 3PLM with generally distinct items, the total number of pseudocounts is to be equally distributed to n items as in the above cases. That is, in the $2 \times n$ table, each cell has $n^{-1}0.5c^*$ for its pseudocount. Let L_{PC} be the (pseudo) likelihood of θ when the pseudocount is used. Then, L_{PC} is formally written as

$$L_{PC} = \prod_{m=1}^n P_m^{U_m+n^{-1}0.5c^*} Q_m^{1-U_m+n^{-1}0.5c^*} = \left(\prod_{m=1}^n P_m^{U_m} Q_m^{1-U_m} \right) \left\{ \prod_{m=1}^n (P_m Q_m)^{0.5c^*} \right\}^{1/n} \equiv L f(\theta), \tag{2.3}$$

where $L = \prod_{m=1}^n P_m^{U_m} Q_m^{1-U_m}$ is the usual likelihood or L_{PC} with $c^* = 0$; $f(\theta) = O(1)$ is seen as a prior density or its proportional from Bayesian point of view; and $f_m(\theta) \equiv (P_m Q_m)^{0.5c^*}$ is also seen as an independent prior for the m -th item. Note that $f(\theta)$ is the geometric mean of $f_m(\theta)$ ($m = 1, \dots, n$), and becomes equal to $f_m(\theta)$ when items are equivalent e.g., under the LME.

Actually, $f_m(\theta)$ becomes typical priors in some cases. When the 2-parameter logistic model (2PLM) holds, $D^2 a_m^2 P_m Q_m$ is the Fisher or item information for the m -th item. Then, $(P_m Q_m)^{0.5}$ with $c^* = 1$ is proportional to the Jeffreys (1946; 1961, Section 3.10) non-informative prior for the m -th item. That is, $c^* = 1$ in (2.3) gives the geometric mean of the independent Jeffreys priors whereas the usual Jeffreys prior is given from the test information i.e., the sum of the n item informations averaged over items or the arithmetic mean of the item informations:

$$\bar{i} \equiv n^{-1} \sum_{m=1}^n P_m'^2 / (P_m Q_m), \tag{2.4}$$

where $P_m' = \partial P_m / \partial \theta$. Under the 2PLM, (2.4) becomes

$$\bar{i} = n^{-1} D^2 \sum_{m=1}^n a_m^2 P_m Q_m. \tag{2.5}$$

It is known that the Jeffreys Bayes modal estimator (JME) under canonical parametrization in the exponential family has zero asymptotic bias of order $O(n^{-1})$ when the model holds (Firth, 1993; in the current paper, the BME refers to only the

estimator using the standard normal prior). The ability estimator with fixed item parameters in the 2PLM belongs to this family. However, the corresponding estimator in the 3PLM does not belong to it. That is, Warm's WLE under the 2PLM becomes identical to the JME. Since under the LME, $f(\theta)$ with $c^* = 1$ becomes proportional to the usual Jeffreys prior, bias reduction for the ability estimator even under the 3PLM with different items is expected to some extent when $c^* = 1$ is used, which will be numerically illustrated later.

It is known that the bias of the MLE of ability is positively correlated with the population ability θ_0 (Lord, 1983; Ogasawara, 2012a). That is, when θ_0 is positive (negative), the MLE tends to be larger (smaller) than θ_0 . Equal pseudocounts for the two cells of correct and incorrect responses have the effect of centering the MLE towards 0.

When $c^* = 2$, $f_m(\theta) = P_m Q_m$ which is proportional to the informative logistic prior $D a_m P_m Q_m$, whose cumulative distribution function is P_m under the 2PLM. It is known that when $a_m = 1$, $b_m = 0$ and $c_m = 0$ with $D = 1.7$, $|P_m(\theta) - \Phi(\theta)| < 0.01$ for all θ (Lord & Novick, 1968, Inequality (17.2.2)), where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal. Thus, $f(\theta)$ with $c^* = 2$ is reduced to an approximate geometric mean of various independent normal priors. In the case of the binomial proportion, $c^* = 2$ corresponds to the flat beta prior, where the posterior mean of the binomial proportion is given by $(np + 1)/(n + 2)$ with p being the unbiased sample proportion, which is also derived by Laplace's rule of succession (see e.g., Wilson, 1927, p.210).

In similar manners, $f_m(\theta)$'s with other c^* 's are seen as priors or penalty functions for shrinkage estimators (see e.g., Gruber, 1998, Chapter 1; Lehmann & Casella, 1998, Chapter 5, Section 5), whose effect is stronger as c^* becomes larger.

Define \bar{l}_{PC} as

$$\begin{aligned}\bar{l}_{\text{PC}} &\equiv n^{-1} \log L_{\text{PC}} = n^{-1} \log L + n^{-1} \log f(\theta) \\ &\equiv \bar{l} + n^{-1} 0.5c^* \log \left(\prod_{m=1}^n P_m Q_m \right)^{1/n} \\ &\equiv \bar{l} + n^{-1} 0.5c^* \log \bar{h},\end{aligned}\tag{2.6}$$

where $\bar{h} = O(1)$ is the geometric mean of $P_m Q_m$ ($m = 1, \dots, n$) mentioned earlier (compare \bar{i} of (2.5) under the 2PLM). The estimator $\hat{\theta}_{\text{PC}}$ using the pseudocount (PCE) is given by maximizing L_{PC} or \bar{l}_{PC} , which is obtained by the solution of the following equation:

$$\begin{aligned}\frac{\partial \bar{l}_{\text{PC}}}{\partial \theta_{\text{PC}}} &= \frac{\partial \bar{l}}{\partial \theta_{\text{PC}}} + n^{-1} 0.5c^* \frac{\bar{h}'}{\bar{h}} \\ &= n^{-1} \sum_{m=1}^n \frac{1}{P_m Q_m} \{U_m - P_m + n^{-1} 0.5c^* (1 - 2P_m)\} P_m' = 0,\end{aligned}\tag{2.7}$$

where $\bar{h}' = \partial \bar{h} / \partial \theta_{\text{PC}}$. From (2.7), it is seen that the normal equation for the MLE denoted by $\hat{\theta}_{\text{ML}}$ is modified for $\hat{\theta}_{\text{PC}}$ by replacing the actual observation U_m with

$U_m + n^{-1}0.5c^*(1 - 2P_m)$ or by replacing P_m in $\{\cdot\}$ with $P_m - n^{-1}0.5c^*(1 - 2P_m)$. Under m.m. with $P_{Tm}(m = 1, \dots, n)$, the population θ_0 is defined as the solution of $E_T(\partial \bar{l} / \partial \theta_{PC}) = 0$ without $f(\theta)$.

3. Pseudocounts reducing the asymptotic mean square error

In (2.7) it is found that $0.5c^*\bar{h}'/\bar{h}$ is a special case of the general weight $g(\theta)$ for the weighted score dealt with by Ogasawara (2013a) when $\theta = \theta_{PC}$. The weights for the MLE, BME, JME, WLE and PCE are summarized as

$$\begin{aligned} g(\theta) &= 0 \quad \text{for ML, } g(\theta) = -\theta \quad \text{for BM, } g(\theta) = \bar{i}'/(2\bar{i}) \quad \text{for JM,} \\ g(\theta) &= \bar{j}/(2\bar{i}) \quad \text{for WL and } g(\theta) = c^*\bar{h}'/(2\bar{h}) \quad \text{for PC,} \end{aligned} \tag{3.1}$$

where

$$\bar{i}' = \frac{\partial \bar{i}}{\partial \theta} = n^{-1} \sum_{M=1}^n \left\{ \frac{2P_m'P_m''}{P_mQ_m} - \frac{(1 - 2P_m)P_m'^3}{(P_mQ_m)^2} \right\}, \quad \bar{j} = n^{-1} \sum_{M=1}^n \frac{P_m'P_m''}{P_mQ_m} \tag{3.2}$$

and $P_m'' = \frac{\partial^2 P_m}{\partial \theta^2}$.

Denote the m -th cumulant of the generic estimator $\hat{\theta}_{GW}$ using the general weight $g(\theta)$ by $\kappa_m(\hat{\theta}_{GW})$. Then, their asymptotic cumulants up to the second order are given as

$$\begin{aligned} \kappa_1(\hat{\theta}_{GW} - \theta_0) &= n^{-1}\alpha_1 + O(n^{-2}), \\ \kappa_2(\hat{\theta}_{GW}) &= n^{-1}\alpha_2 + n^{-2}\alpha_{\Delta 2} + O(n^{-3}) \quad (\alpha_2 = \alpha_{ML2}), \end{aligned} \tag{3.3}$$

where $n^{-1}\alpha_1$ is the asymptotic bias of order $O(n^{-1})$; $n^{-1}\alpha_2$ is the usual asymptotic variance of order $O(n^{-1})$, which is equal to $n^{-1}\alpha_{ML2}$ by ML; and $n^{-2}\alpha_{\Delta 2}$ is the added higher-order asymptotic variance of $\hat{\theta}_{GW}$, respectively. The actual expressions of the asymptotic cumulants under p.m.m. are given by Ogasawara (2013a, Equation (3.4), Appendix A.2)

From (3.3), the MSE of $\hat{\theta}_{GW}$ is given by

$$\begin{aligned} E_T\{(\hat{\theta}_{GW} - \theta_0)^2\} &= \{\kappa_1(\hat{\theta}_{GW} - \theta_0)\}^2 + \kappa_2(\hat{\theta}_{GW}) \\ &= n^{-1}\alpha_{ML2} + n^{-2}(\alpha_1^2 + \alpha_{\Delta 2}) + O(n^{-3}) \\ &\equiv \text{MSE}_{O(n^{-2})}(\hat{\theta}_{GW}) + O(n^{-3}), \end{aligned} \tag{3.4}$$

where $\text{MSE}_{O(n^{-2})}(\cdot)$ is the AMSE up to order $O(n^{-2})$. The actual expression of $\text{MSE}_{O(n^{-2})}(\hat{\theta}_{GW})$ under p.m.m. is given by Ogasawara (2013a, Equation (3.13)) as

$$\begin{aligned} \text{MSE}_{O(n^{-2})}(\hat{\theta}_{GW}) &= n^{-1}\alpha_{ML2} + n^{-2} \left[\alpha_{ML\Delta 2} + \alpha_{ML1}^2 \right. \\ &\quad \left. - 2\lambda^{-1}g'(\theta_0)\alpha_{ML2} - 2g(\theta_0) \left\{ \lambda^{-3}E_T \left(n \frac{\partial \bar{l}}{\partial \theta} \Big|_{\theta=\theta_0} \frac{\partial^2 \bar{l}}{\partial \theta^2} \Big|_{\theta=\theta_0} \right) \right. \right. \\ &\quad \left. \left. - \lambda^{-2} E_T \left(\frac{\partial^3 \bar{l}}{\partial \theta^3} \Big|_{\theta=\theta_0} \right) \alpha_{ML2} + \lambda^{-1}\alpha_{ML1} \right\} + \lambda^{-2}g(\theta_0)^2 \right], \end{aligned} \tag{3.5}$$

where $\alpha_{\text{ML}\Delta 2}$ and $\alpha_{\text{ML}1}$ are defined similarly to $\alpha_{\text{ML}2}$; $\lambda = E_{\text{T}}(\partial^2 \bar{l} / \partial \theta^2 |_{\theta=\theta_0})$; $g'(\theta_0) = \partial g(\theta) / \partial \theta |_{\theta=\theta_0}$.

Under c.m.s., using $\alpha_{\text{ML}2} = \bar{i}_0^{-1}$, where \bar{i}_0 is \bar{i} given by θ_0 , $\alpha_{\text{ML}1} = -\bar{j}_0 / (2\bar{i}_0^2)$ (Lord, 1983; Warm, 1989) and the Bartlett identity, (3.5) becomes

$$\begin{aligned} \text{MSE}_{O(n-2)}(\hat{\theta}_{\text{GW}}) &= n^{-1} \bar{i}_0^{-1} + n^{-2} [\alpha_{\text{ML}\Delta 2} + \alpha_{\text{ML}1}^2 + 2\bar{i}_0^{-2} g'(\theta_0) \\ &\quad - \bar{i}_0^{-3} (2\bar{i}_0' + \bar{j}_0) g(\theta_0) + \bar{i}_0^{-2} g(\theta_0)^2], \end{aligned} \quad (3.6)$$

where \bar{i}_0' and \bar{j}_0 are defined similarly to \bar{i}_0 . Note that $n^{-1} \alpha_2 + n^{-2} (\alpha_{\text{ML}\Delta 2} + \alpha_{\text{ML}1}^2)$ in (3.5) and $n^{-1} \bar{i}_0^{-1} + n^{-2} (\alpha_{\text{ML}\Delta 2} + \alpha_{\text{ML}1}^2)$ in (3.6) are $\text{MSE}_{O(n-2)}(\hat{\theta}_{\text{ML}})$'s, though $\alpha_{\text{ML}\Delta 2}$ and α_1 in (3.5) are generally different from those in (3.6), respectively.

When $g(\theta_0) = c^* \bar{h}_0' / (2\bar{h}_0)$, where \bar{h}_0' and \bar{h}_0 are population counterparts of \bar{h}' and \bar{h} , respectively, (3.6) under c.m.s. is written as

$$\begin{aligned} \text{MSE}_{O(n-2)}(\hat{\theta}_{\text{PC}}) &= \text{MSE}_{O(n-2)}(\hat{\theta}_{\text{ML}}) \\ &\quad + n^{-2} c^* \left\{ 2\bar{i}_0^{-2} \frac{\partial \bar{h}'}{\partial \theta} \frac{\bar{h}'}{2\bar{h}} \Big|_{\theta=\theta_0} - \bar{i}_0^{-3} (2\bar{i}_0' + \bar{j}_0) \frac{\bar{h}_0'}{2\bar{h}_0} \right\} + n^{-2} c^{*2} \bar{i}_0^{-2} \frac{\bar{h}_0'^2}{4\bar{h}_0^2}, \end{aligned} \quad (3.7)$$

which is minimized when c^* is

$$\begin{aligned} c_{\min}^* &\equiv -\frac{2\bar{h}_0^2}{\bar{h}_0'^2} \left(\frac{\bar{h}_0''}{\bar{h}_0} - \frac{\bar{h}_0'^2}{\bar{h}_0^2} \right) + \frac{2\bar{h}_0}{\bar{i}_0 \bar{h}_0'} \left(\bar{i}_0' + \frac{\bar{j}_0}{2} \right) \\ &= -2 \frac{\bar{h}_0 \bar{h}_0''}{\bar{h}_0'^2} + 2 + \frac{2\bar{h}_0}{\bar{i}_0 \bar{h}_0'} \left(\bar{i}_0' + \frac{\bar{j}_0}{2} \right), \end{aligned} \quad (3.8)$$

where $\bar{h}_0' \neq 0$ is assumed and $\bar{i}_0 > 0$ due to another assumption of finite θ_0 .

In (3.7) and (3.8),

$$\frac{2g(\theta_0)}{c^*} = \frac{\bar{h}_0'}{\bar{h}_0} = n^{-1} \sum_{m=1}^n \frac{1 - 2P_m}{P_m Q_m} P_m' \quad (3.9)$$

(see (2.7)) and

$$\begin{aligned} \frac{2g'(\theta_0)}{c^*} &= \frac{\partial \bar{h}_0'}{\partial \theta} \frac{1}{\bar{h}_0} \Big|_{\theta=\theta_0} = \frac{\partial}{\partial \theta} n^{-1} \sum_{m=1}^n \frac{1 - 2P_m}{P_m Q_m} P_m' \Big|_{\theta=\theta_0} \\ &= n^{-1} \sum_{m=1}^n \left\{ \frac{-2P_m'^2 + (1 - 2P_m) P_m''}{P_m Q_m} - \frac{(1 - 2P_m)^2 P_m'^2}{(P_m Q_m)^2} \right\}. \end{aligned} \quad (3.10)$$

Then, (3.8) is rewritten as

$$\begin{aligned} c_{\min}^* &= -2 \left(n^{-1} \sum_{m=1}^n \frac{1 - 2P_m}{P_m Q_m} P_m' \right)^{-2} \\ &\quad \cdot n^{-1} \sum_{m=1}^n \left\{ \frac{-2P_m'^2 + (1 - 2P_m) P_m''}{P_m Q_m} - \frac{(1 - 2P_m)^2 P_m'^2}{(P_m Q_m)^2} \right\} \\ &\quad + 2\bar{i}_0^{-1} \left(n^{-1} \sum_{m=1}^n \frac{1 - 2P_m}{P_m Q_m} P_m' \right)^{-1} \left(\bar{i}_0' + \frac{\bar{j}_0}{2} \right). \end{aligned} \quad (3.11)$$

Under the 3PLM, let $\Psi_m \equiv 1/[1 + \exp\{-Da_m(\theta - b_m)\}]$. Then, in (3.11),

$$\begin{aligned} P_m' &= Da_m(1 - c_m)\Psi_m(1 - \Psi_m), \\ P_m'' &= D^2a_m^2(1 - c_m)(1 - 2\Psi_m)\Psi_m(1 - \Psi_m). \end{aligned} \quad (3.12)$$

In (3.8) or (3.11), c_{\min}^* includes unknown θ_0 and is not obtained in practice. However, under the LME, where

$$\begin{aligned} P_m = P &\equiv \frac{1}{1 + \exp\{-Da(\theta - b)\}}, & Q &\equiv 1 - P \quad (m = 1, \dots, n), \\ \bar{i}_0 = \bar{h}_0 &= P'^2/(PQ) = D^2a^2PQ, \\ \bar{i}_0' = \bar{h}_0' &= \bar{j}_0 \end{aligned} \quad (3.13)$$

with P' defined as P_m' , a fixed lower bound for c_{\min}^* is found as follows.

Theorem 1. Under the LME of (3.13) with the assumptions of $\bar{i}_0' = \bar{j}_0 \neq 0$ and finite θ_0 , a lower bound of c_{\min}^* is 3.

Proof. Using (3.13), c_{\min}^* in (3.8) becomes

$$c_{\min}^* = -2\frac{\bar{i}_0\bar{i}_0''}{\bar{i}_0'^2} + 5, \quad (3.14)$$

where $\bar{i}_0' = D^3a^3(1 - 2P)PQ$ and $\bar{i}_0'' = \partial^2\bar{i}/\partial\theta^2|_{\theta=\theta_0} = D^4a^4(1 - 6P + 6P^2)PQ$. Then, (3.14) is

$$\begin{aligned} c_{\min}^* &= -2\frac{(D^2a^2PQ)\{D^4a^4(1 - 6P + 6P^2)PQ\}}{\{D^3a^3(1 - 2P)PQ\}^2} + 5 = -2\frac{1 - 6P + 6P^2}{(1 - 2P)^2} + 5 \\ &= -2\frac{(1 - 2P)^2 - 2P(1 - P)}{(1 - 2P)^2} + 5 = 3 + \frac{4PQ}{(1 - 2P)^2} > 3. \quad \text{Q.E.D.} \end{aligned} \quad (3.15)$$

In Theorem 1, the lower bound 3 corresponding to $\theta_0 = \pm\infty$ is not attained by finite θ_0 . The term $4PQ/(1 - 2P)^2$ in the last inequality in (3.15) becomes infinitely large as θ_0 approaches 0. This corresponds to the zero bias of $\hat{\theta}_{\text{ML}}$ at $\theta_0 = 0$. It is of interest to find that the lower bound 3 is the midpoint of the pseudocounts by Laplace's rule and Agresti and Coull (1998).

In current ability tests for achievements based on IRT, the LME is seldom used. However, the familiar 3PLM used frequently in practice is similar to the LME in a crude sense. Thus, the lower bound in Theorem 1 is expected to be of practical use to some extent in the case of the 3PLM, which is numerically illustrated with and without m.m. in the next section.

4. Numerical illustration under the 3-parameter logistic model

Numerical illustration is given using artificial item parameters of the 3PLM, which are randomly generated as in Ogasawara (2012a, 2013a) and mimic actual values in practice (Lord, 1975, Figures 1–4; Kolen & Brennan, 2004, Table 6.5). The probabilities P_{Tm} ($m = 1, \dots, n$) under m.m. are given by perturbing the term $-Da_m(\theta - b_m)$

Table 1: Values of c_{\min}^* minimizing the asymptotic mean square error for the 3PLM

	θ						
n	-3	-2	-1	0	1	2	3
20	40.81	19.68	13.91	147.90	6.46	3.35	2.69
50	41.78	19.48	18.09	96.27	4.07	2.65	2.60
100	38.48	23.52	25.63	68.40	3.98	3.07	2.83

Table 2: Asymptotic and simulated errors in estimation of ability when the 3PLM holds ($n = 20$)

(Number of deleted cases)	ML	BM	WL	PC(1)	PC(2)	PC(3)	PC(4)	PC(8)	PC(c_{\min}^*)
(11)	$\theta = 1, c_{\min}^* = 6.46$								
ASE	.3779	ml	ml	ml	ml	ml	ml	ml	ml
HASE	.3989	.3399	.3796	.3789	.3578	.3353	.3113	.1863	.2423
Simulated (SD)	.4075	.3406	.3783	.3809	.3592	.3407	.3246	.2755	.2921
$\alpha_{\Delta 2}$	6.5	-10.9	.5	.3	-5.9	-12.1	-18.4	-43.2	-33.6
Simulated	9.3	-10.7	.1	.9	-5.5	-10.7	-15.0	-26.8	-23.0
α_1^2	.1	6.1	0	.1	1.1	3.2	6.3	29.3	18.5
Simulated	.2	5.4	.0	.1	1.1	2.9	5.1	17.0	12.1
HRMSE	.3993	.3618	.3796	.3793	.3617	.3471	.3358	.3285	.3238
Simulated	.4081	.3599	.3783	.3814	.3631	.3510	.3437	.3440	.3398
(5188)	$\theta = 2, c_{\min}^* = 3.35$								
ASE	.5003	ml	ml	ml	ml	ml	ml	ml	ml
HASE	.6139	.0627	.5332	.5292	.4281	.2940	-	-	.2289
Simulated (SD)	.5015	.2852	.4059	.4006	.3412	.3010	.2715	.2029	.2896
$\alpha_{\Delta 2}$	50.6	-98.6	135.6	11.9	-26.8	-65.6	-104.3	-259.2	-79.2
Simulated	.5	-67.6	-34.2	-36.0	-53.6	-63.9	-70.7	-83.7	-66.6
α_1^2	6.8	55.0	0	.0	9.3	34.4	75.5	399.2	47.1
Simulated	2.1	50.2	2.2	3.3	18.0	38.3	61.0	155.7	46.1
HRMSE	.6275	.3759	.5332	.5293	.4544	.4153	.4224	.7748	.4124
Simulated	.5067	.4547	.4128	.4107	.4018	.4316	.4756	.6560	.4462

Note. ASE = $n^{-1/2} \alpha_{\text{ML2}}^{1/2}$; HASE = $(n^{-1} \alpha_{\text{ML2}} + n^{-2} \alpha_{\Delta 2})^{1/2}$; HRMSE = $\{n^{-1} \alpha_{\text{ML2}} + n^{-2} (\alpha_{\Delta 2} + \alpha_1^2)\}^{1/2}$; ML: maximum likelihood; BM: Bayes modal; WL: weighted likelihood; PC(c^*): estimation using pseudocount c^* . The sign “ml” indicates that the value is identical to that by ML. The sign “-” indicates that the value is imaginary.

with an added independent variable following $N(0, D^2)$ and keeping $E_T(\bar{l}) = 0$. The generated item parameters are dealt with as known ones. Sample sizes $n = 20, 50$ and 100 are used.

Table 1 gives c_{\min}^* under the 3PLM for $\theta = -3, -2, \dots, 3$. When $\theta = 0$, the values are relatively large which is expected from (3.15). We find that c_{\min}^* as a function of θ is reflected N -shaped rather than unimodal. Some of the values are somewhat smaller than the lower bound 3 for the LME. Since most of the values are larger than 3 or approximately 3, the lower bound seems to be promising.

Tables 2 to 6 give asymptotic and simulated errors of estimation, when ML, BM,

Table 3: Asymptotic and simulated errors in estimation of ability when the 3PLM does not hold ($n = 20$)

(Number of deleted cases)	ML	BM	WL	PC(1)	PC(2)	PC(3)	PC(4)	PC(8)	PC(c_{\min}^*)
(2)	$\theta = 1, c_{\min}^* = 6.46, \text{cor}(P_{Tm}, P_m) = .762$								
ASE	.3057	ml	ml						
HASE	.3205	.2716	.3050	.3042	.2869	.2686	.2489	.1455	.1921
Simulated (SD)	.3259	.2756	.3045	.3059	.2892	.2748	.2622	.2233	.2365
$\alpha_{\Delta 2}$	3.7	-7.9	-.2	-.4	-4.4	-8.5	-12.6	-28.9	-22.6
Simulated	5.1	-7.0	-.3	.1	-3.9	-7.2	-9.9	-17.4	-15.0
α_1^2	.1	6.1	.0	.2	1.2	3.3	6.3	28.2	17.9
Simulated	.1	5.2	.0	.2	1.2	2.8	4.9	16.1	11.4
HRMSE	.3208	.2984	.3050	.3048	.2922	.2834	.2787	.3029	.2858
Simulated	.3263	.2984	.3046	.3067	.2942	.2873	.2848	.3001	.2908
(8185)	$\theta = 2, c_{\min}^* = 3.35, \text{cor}(P_{Tm}, P_m) = .290$								
ASE	.4878	ml	ml						
HASE	.6199	.1505	.5426	.5388	.4431	.3200	.0919	-	.2633
Simulated (SD)	.4387	.2565	.3599	.3567	.3056	.2700	.2435	.1806	.2599
$\alpha_{\Delta 2}$	58.5	-86.1	22.6	20.9	-16.7	-54.2	-91.8	-242.1	-67.5
Simulated	-18.2	-68.9	-43.4	-44.3	-57.8	-66.0	-71.5	-82.1	-68.2
α_1^2	6.3	57.5	.0	.1	10.1	36.3	78.6	410.2	49.3
Simulated	.2	56.2	5.0	6.3	23.0	44.3	67.4	162.0	52.3
HRMSE	.6325	.4078	.5426	.5390	.4707	.4394	.4528	.8112	.4389
Simulated	.4393	.4541	.3768	.3783	.3886	.4285	.4774	.6615	.4452

Note. ASE = $n^{-1/2}\alpha_{ML2}^{1/2}$; HASE = $(n^{-1}\alpha_{ML2} + n^{-2}\alpha_{\Delta 2})^{1/2}$; HRMSE = $\{n^{-1}\alpha_{ML2} + n^{-2}(\alpha_{\Delta 2} + \alpha_1^2)\}^{1/2}$; ML: maximum likelihood; BM: Bayes modal; WL: weighted likelihood; PC(c^*): estimation using pseudocount c^* . The sign "ml" indicates that the value is identical to that by ML. The sign "-" indicates that the value is imaginary.

WL, and PC (c^*) i.e., PC with c^* are used, where $c^* = 1, 2, 3, 4$ and c_{\min}^* . In addition, $c^* = 5$ or 8 is used in Tables 2, 3 and 5 to have c^* 's greater than c_{\min}^* when c_{\min}^* is not so large. It is easily seen that ML is equivalent to PC (0). Note that the results by PC (c_{\min}^*) are for comparison, where c_{\min}^* given by θ_0 is available in the experimental data though unavailable in practice. In the tables ASE = $n^{-1/2}\alpha_{ML2}^{1/2}$ is the asymptotic standard error, which is robust under m.m.; HASE = $(n^{-1}\alpha_{ML2} + n^{-2}\alpha_{\Delta 2})^{1/2}$ is the higher-order ASE; HRMSE = $\{n^{-1}\alpha_{ML2} + n^{-2}(\alpha_{\Delta 2} + \alpha_1^2)\}^{1/2} = (\text{HASE}^2 + n^{-2}\alpha_1^2)^{1/2}$ is the higher-order asymptotic root MSE up to order $O(n^{-3/2})$. Note that the asymptotic root MSE up to order $O(n^{-1/2})$ is identical to the ASE.

The simulated SE denoted by SD corresponding to ASE and HASE is the square root of the unbiased variance of 10^5 estimates for each estimator. The simulated $\alpha_{\Delta 2}$ is $n^2(\text{SD}^2 - n^{-1}\alpha_2)$. The simulated α_1^2 is n^2 times the square of the simulated bias. The simulated root MSE is given by the square root of the simulated mean of $(\hat{\theta}_{GW} - \theta_0)^2$. When $\theta \leq -2$ ($\theta \leq 0$ for $n = 20$) and $\theta = 3$, many cases of non-convergence in estimation occurred. These cases were not employed. In the tables,

Table 4: Asymptotic and simulated errors in estimation of ability when the 3PLM holds ($n = 50$)

(Number of deleted cases)	ML	BM	WL	PC(1)	PC(2)	PC(3)	PC(4)	PC(c_{\min}^*)
(591)	$\theta = -1, c_{\min}^* = 18.09$							
ASE	.3601	ml	ml	ml	ml	ml	ml	ml
HASE	.3915	.3023	.3640	.3847	.3779	.3709	.3638	.2422
Simulated (SD)	.3971	.2974	.3463	.3832	.3718	.3617	.3527	.2697
$\alpha_{\Delta 2}$	58.9	-95.7	7.1	45.8	32.7	19.6	6.6	-177.6
Simulated	70.0	-103.0	-24.4	42.9	21.3	2.9	-13.3	-142.3
α_1^2	2.1	25.3	0	.7	.0	.2	1.3	103.5
Simulated	2.6	27.2	.4	.7	.0	.4	1.7	71.3
HRMSE	.3925	.3187	.3640	.3851	.3779	.3710	.3644	3163
Simulated	.3984	.3152	.3465	.3836	.3718	.3619	.3536	.3182
(0)	$\theta = 0, c_{\min}^* = 96.27$							
ASE	.2833	ml	ml	ml	ml	ml	ml	ml
HASE	.2872	.2638	.2840	.2828	.2783	.2737	.2691	-
Simulated (SD)	.2875	.2650	.2839	.2830	.2786	.2743	.2701	.1140
$\alpha_{\Delta 2}$	5.5	-26.7	1.0	-.8	-7.1	-13.4	-19.7	-601.1
Simulated	6.0	-25.2	.7	-.5	-6.7	-12.6	-18.2	-168.2
α_1^2	.0	.0	0	.0	.2	.4	.6	304.0
Simulated	.0	.0	.0	.0	.1	.2	.4	48.0
HRMSE	.2872	.2638	.2840	.2828	.2784	.2740	.2695	-
Simulated	.2875	.2650	.2839	.2830	.2786	.2744	.2704	.1794

Note. ASE = $n^{-1/2}\alpha_{ML2}^{1/2}$; HASE = $(n^{-1}\alpha_{ML2} + n^{-2}\alpha_{\Delta 2})^{1/2}$; HRMSE = $\{n^{-1}\alpha_{ML2} + n^{-2}(\alpha_{\Delta 2} + \alpha_1^2)\}^{1/2}$; ML: maximum likelihood; BM: Bayes modal; WL: weighted likelihood; PC(c^*): estimation using pseudocount c^* . The sign "ml" indicates that the value is identical to that by ML. The sign "-" indicates that the value is imaginary.

the numbers of deleted cases until 10^5 regular observations were obtained are shown. The non-zero numbers of the deleted cases when $\theta = 1, 2$ are all due to perfect scores, where finite $\hat{\theta}_{ML}$ is unavailable and these cases are not used for other estimators.

Tables 2, 4 and 5 are given under c.m.s. while Tables 3 and 6 under m.m., where the extent of m.m. is shown by the correlation of P_{Tm} and P_m over items. Table 2 gives the results under c.m.s. with n being as small as 20. When $\theta = 1$, the asymptotic values are reasonably similar to their corresponding simulated values. The SD is closer to the corresponding HASE than the ASE. The results by WL are reasonable in that not only the asymptotic bias is zero by construction (denoted by 0 in the associated tables rather than .0 for other rounded numbers), but also $\alpha_{\Delta 2}$ is smaller than that by ML with similar simulated values. The results by PC (1) are similar to those by WL, which is expected to some extent since they are identical under the LME. It is of interest to find that the results by PC (3) using the lower bound 3 for the LME are similar to those by BM. The HRMSE by PC (4) is smaller than those by PC (3) and BM and greater than by PC (c_{\min}^*) with similar simulated tendencies.

Table 5: Asymptotic and simulated errors in estimation of ability when the 3PLM holds ($n = 50$)

(Number of deleted cases)	ML	BM	WL	PC(1)	PC(2)	PC(3)	PC(4)	PC(5)	PC(c_{\min}^*)
(0)	$\theta = 1, c_{\min}^* = 4.07$								
ASE	.2733	ml	ml						
HASE	.2778	.2545	.2743	.2717	.2655	.2591	.2525	.2458	.2521
Simulated (SD)	.2779	.2553	.2735	.2714	.2653	.2595	.2540	.2487	.2536
$\alpha_{\Delta 2}$	6.2	-24.8	1.4	-2.2	-10.6	-19.0	-27.4	-35.8	-27.9
Simulated	6.3	-23.9	.3	-2.7	-10.9	-18.5	-25.5	-32.1	-26.0
α_1^2	.3	10.0	0	.3	2.6	7.3	14.3	23.7	14.9
Simulated	.4	8.7	.0	.2	2.3	6.4	12.4	20.0	12.9
HRMSE	.2781	.2623	.2743	.2719	.2674	.2646	.2636	.2644	.2636
Simulated	.2782	.2620	.2735	.2715	.2670	.2644	.2635	.2643	.2636
(26)	$\theta = 2, c_{\min}^* = 2.65$								
ASE	.3290	ml	ml						
HASE	.3515	.2617	.3344	.3321	.3116	.2896	.2657	.2396	.2974
Simulated (SD)	.3590	.2740	.3343	.3313	.3095	.2914	.2761	.2628	.2973
$\alpha_{\Delta 2}$	38.1	-99.4	8.8	5.1	-28.0	-61.1	-94.1	-127.2	-49.6
Simulated	51.6	-83.0	8.8	3.8	-31.3	-58.4	-80.1	-98.0	-49.7
α_1^2	4.3	76.6	0	.7	14.1	44.6	92.0	156.5	32.1
Simulated	5.4	67.4	.0	.8	14.4	41.8	80.0	126.7	30.9
HRMSE	.3539	.3149	.3344	.3325	.3205	.3189	.3278	.3464	.3182
Simulated	.3620	.3194	.3343	.3318	.3186	.3188	.3290	.3461	.3174

Note. ASE = $n^{-1/2}\alpha_{ML2}^{1/2}$; HASE = $(n^{-1}\alpha_{ML2} + n^{-2}\alpha_{\Delta 2})^{1/2}$; HRMSE = $\{n^{-1}\alpha_{ML2} + n^{-2}(\alpha_{\Delta 2} + \alpha_1^2)\}^{1/2}$; ML: maximum likelihood; BM: Bayes modal; WL: weighted likelihood; PC(c^*): estimation using pseudocount c^* . The sign "ml" indicates that the value is identical to that by ML.

In the table, it is seen that $\alpha_{\Delta 2}$ and α_1^2 for PC (c^*) are monotonically decreasing and increasing, respectively in terms of c^* with similar simulated values. However, we find that HRMSE and its simulated value are not monotonic with respect to c^* .

When $\theta = 2$, some discrepancies appear especially in the results by ML, which is primarily due to the large number of deleted cases, which is as many as 5% of the regular cases. It is seen in Table 2 that when $\theta = 2$, BM gives the smallest HRMSE .3759 while PC (2) yields the smallest simulated RMSE .4018.

Table 3 gives results under a realistic condition in that the 3PLM is more or less misspecified in practice for ability tests. It is encouraging to see that the results are, in a crude sense, similar to those in Table 2. Note again that when $\theta = 2$, as many as 8185 cases corresponding to 8% of the regular cases were discarded due to perfect scores. When $\theta = 1$, the smallest HRMSE and simulated RMSE are given by PC (4) not by PC (c_{\min}^*) owing to m.m., while they are given by BM and WL, respectively when $\theta = 2$.

Tables 4 and 5 show the results with $n = 50$ under c.m.s. when $\theta = -1, 0, 1, 2$.

Table 6: Asymptotic and simulated errors in estimation of ability when the 3PLM does not hold ($n = 50$)

(Number of deleted cases)	ML	BM	WL	PC(1)	PC(2)	PC(3)	PC(4)	PC(c_{\min}^*)
(83)	$\theta = -1, c_{\min}^* = 18.09, \text{cor}(P_{Tm}, P_m) = .501$							
ASE	.3608	ml	ml	ml	ml	ml	ml	ml
HASE	.3790	.2812	.3490	.3718	.3644	.3569	.3492	.2130
Simulated (SD)	.3842	.2971	.3433	.3737	.3642	.3555	.3474	.2679
$\alpha_{\Delta 2}$	33.8	-127.7	-20.9	20.2	6.6	-7.0	-20.6	-212.0
Simulated	43.7	-104.7	-30.7	23.8	6.3	-9.4	-23.7	-146.0
α_1^2	1.3	33.9	.2	.2	.1	.9	2.7	128.4
Simulated	1.7	31.2	.7	.3	.0	.8	2.6	79.7
HRMSE	.3797	.3044	.3491	.3719	.3644	.3574	.3507	.3110
Simulated	.3851	.3174	.3437	.3738	.3642	.3560	.3488	.3219
(13)	$\theta = 2, c_{\min}^* = 2.65, \text{cor}(P_{Tm}, P_m) = .470$							
ASE	.2953	ml	ml	ml	ml	ml	ml	ml
HASE	.3126	.2317	.2972	.2952	.2767	.2569	.2355	.2639
Simulated (SD)	.3181	.2459	.2974	.2950	.2764	.2608	.2476	.2659
$\alpha_{\Delta 2}$	26.3	-83.8	2.9	-.1	-26.5	-52.9	-79.4	-43.8
Simulated	35.0	-66.7	3.2	-.4	-27.0	-47.9	-64.7	-41.2
α_1^2	2.6	87.1	.2	1.8	18.4	52.3	103.7	38.6
Simulated	3.3	72.9	.2	1.7	17.6	46.8	86.6	35.4
HRMSE	.3142	.2975	.2973	.2964	.2897	.2949	.3113	.2917
Simulated	.3201	.2994	.2976	.2961	.2888	.2946	.3098	.2913

Note. ASE = $n^{-1/2}\alpha_{ML2}^{1/2}$; HASE = $(n^{-1}\alpha_{ML2} + n^{-2}\alpha_{\Delta 2})^{1/2}$; HRMSE = $\{n^{-1}\alpha_{ML2} + n^{-2}(\alpha_{\Delta 2} + \alpha_1^2)\}^{1/2}$; ML: maximum likelihood; BM: Bayes modal; WL: weighted likelihood; PC(c^*): estimation using pseudocount c^* . The sign "ml" indicates that the value is identical to that by ML.

The smallest HRMSE and simulated RMSE are given by BM or PC (c_{\min}^*). When $\theta = -1$ and 0, the value of c^* in PC (c^*) corresponding to BM in Table 4 seems to be greater than 4 reflecting the relatively large c_{\min}^* . Table 6 shows the results under m.m. when $\theta = -1$ and 2, which are found to be similar to the corresponding results in Tables 4 and 5.

5. Remarks

The results in the numerical illustration show that the lower bound 3 is reasonable especially when θ_0 is as large as 2 in data similar to those in this paper. When θ_0 is smaller than or equal to 0, $c^* = 4$ may be more reasonable. In this case, large c^* 's as c_{\min}^* may not always be recommended for use in practice though c_{\min}^* gives the smallest HRMSE under c.m.s. in the data. This is because the large c^* also gives relatively larger biases than those by e.g., BM. In the tables, BM is found to give small MSEs and in many cases comparable results by PC (c^*) with $c^* = 3$ and 4.

However, when relatively small bias is desired, WL and PC (c^*) with $c^* = 1$ and 2 may be used.

Probably, it is possible to have less biased BMEs by using more relatively vague priors in place of the standard normal. However, the prior of the BME is an informative one based on the subjective probability of θ_0 from a pure Bayesian point of view, whose reasonable one is not always easily obtained. The standard normal prior for the BME is reasonable in many cases in that item parameters are estimated typically by marginal ML assuming the standard normal for the distribution of ability. The author has tried to obtain a fixed lower bound for the BME similar to c_{\min}^* for the PCE. So far, however, the lower bound is not available. Probably, methods different from that using the MSE for $\hat{\theta}_{\text{GW}}$ e.g., empirical Bayes should be investigated.

In the tables, the squared asymptotic bias is shown as a value independent of n . The corresponding bias of order $O(n^{-1})$ is given by e.g., $(\alpha_1^2)^{1/2}/n = (25.3)^{1/2}/50 \doteq 0.1$ for BM when $\theta = -1$ and $n = 50$ in Table 4 with a similar simulated value. The value 0.1 may be accepted by many researchers especially when we consider the large reduction of the variance by $-95.7/50^2 \doteq -0.038$ (the corresponding squared asymptotic bias is $25.3/50^2 \doteq 0.010$) which reduces HASE = 0.3915 by ML to HASE = 0.3023 by BM with similar simulated values. The corresponding results by c_{\min}^* when $\theta = -1$ and $n = 50$ in Table 4 are $(\alpha_1^2)^{1/2}/n = (103.5)^{1/2}/50 \doteq 0.2$ and the reduction of the HASE = 0.3915 by ML to HASE = 0.2422 by PC (c_{\min}^*) with similar simulated values. Though the large c^* 's as c_{\min}^* were not always recommended earlier by the author, the value of the bias 0.2 may be tolerable for some researchers considering the corresponding substantially small variance.

The method of estimation using the pseudocount is almost as simple as that by ML (recall (2.7)). On the other hand, $\text{MSE}_{O(n^{-2})}(\hat{\theta}_{\text{PC}})$ (see (3.7)) is as complicated as those for other estimators with various $g(\theta)$'s (see (3.6)). It can be shown that an estimator with the same asymptotic cumulants up to the fourth order and the same added higher-order asymptotic variance as those of $\hat{\theta}_{\text{GW}}$ is given by correction of $\hat{\theta}_{\text{ML}}$ as follows:

$$\hat{\theta}_{\text{C-GW}} \equiv \hat{\theta}_{\text{ML}} + n^{-1}\hat{\alpha}_{\text{ML}2}g(\hat{\theta}_{\text{ML}}) \equiv \hat{\theta}_{\text{ML}} - n^{-1}(-\hat{\alpha}_{\text{ML}2}\hat{g}), \quad (5.1)$$

where $\hat{\alpha}_{\text{ML}2}$ is the sample version of $\alpha_{\text{ML}2}$. A typical example of $-\hat{\alpha}_{\text{ML}2}\hat{g}$ is $\hat{\alpha}_{\text{ML}1}$, which is the sample version of $\alpha_{\text{ML}1}$ by ML, yielding the asymptotically bias-corrected estimator $\hat{\theta}_{\text{ML}} - n^{-1}\hat{\alpha}_{\text{ML}1}$. Recall that $g(\theta) = \bar{j}/(2\bar{i})$ for WL, which under c.m.s. gives $-\alpha_{\text{ML}2}g(\theta) = \bar{j}/(2\bar{i}^2) = \alpha_{\text{ML}1}$.

The AMSE of $\hat{\theta}_{\text{C-GW}}$ in (5.1) is more easily obtained than that of $\hat{\theta}_{\text{GW}}$. For simplicity, we consider the case under c.m.s. i.e., $\hat{\theta}_{\text{C-GW}} = \hat{\theta}_{\text{ML}} + n^{-1}\hat{i}^{-1}\hat{g}$, where \hat{i} is the sample version of \bar{i} . Then,

$$\begin{aligned} & \text{E}\{(\hat{\theta}_{\text{C-GW}} - \theta_0)^2\} \\ &= \text{var}(\hat{\theta}_{\text{ML}}) + n^{-2}2n \text{acov}(\hat{\theta}_{\text{ML}}, \hat{i}^{-1}\hat{g}) + \{\text{E}(\hat{\theta}_{\text{C-GW}}) - \theta_0\}^2 + O(n^{-3}) \\ &= n^{-1}\bar{i}_0^{-1} + n^{-2}[\alpha_{\text{ML}\Delta 2} + 2n \text{acov}(\hat{\theta}_{\text{ML}}, \hat{i}^{-1}\hat{g}) + \{\alpha_{\text{ML}1} + \bar{i}_0^{-1}g(\theta_0)\}^2] + O(n^{-3}), \end{aligned} \quad (5.2)$$

where $\text{acov}(\cdot, \cdot)$ is the asymptotic covariance of order $O(n^{-1})$ for the two variables. Using

$$n \text{acov}(\hat{\theta}_{\text{ML}}, \hat{i}^{-1} \hat{g}) = -\bar{i}_0^{-3} \bar{i}_0' g(\theta_0) + \bar{i}_0^{-2} g'(\theta_0), \quad (5.3)$$

(5.2) becomes

$$\begin{aligned} & n^{-1} \bar{i}_0^{-1} + n^{-2} \{ \alpha_{\text{ML}\Delta 2} + \alpha_{\text{ML}1}^2 + 2\bar{i}_0^{-2} g'(\theta_0) - \bar{i}_0^{-3} (2\bar{i}_0' + \bar{j}_0) g(\theta_0) + \bar{i}_0^{-2} g(\theta_0)^2 \} \\ & \quad + O(n^{-3}) \\ & = \text{MSE}_{O(n^{-2})}(\hat{\theta}_{\text{GW}}) + O(n^{-3}). \end{aligned} \quad (5.4)$$

The estimator $\hat{\theta}_{\text{C-GW}}$ in (5.1) was introduced to have some insight in the involved expression of (3.6). Though $\text{MSE}_{O(n^{-2})}(\cdot)$'s are the same for $\hat{\theta}_{\text{GW}}$ and $\hat{\theta}_{\text{C-GW}}$, the latter has a difficulty in the cases of zero and perfect scores, where ability estimation by ML required for $\hat{\theta}_{\text{C-GW}}$ does not give finite $\hat{\theta}_{\text{ML}}$'s. On the other hand, $\hat{\theta}_{\text{GW}}$ ' are available in these cases.

As addressed earlier, since c_{\min}^* depends on unknown θ_0 , c_{\min}^* is not available in practice. In order to overcome the difficulty an anonymous reviewer suggested a method that takes the expectation of c_{\min}^* over a distribution of θ , say $g^*(\theta)$ for its density (the author is indebted to the reviewer for the suggestion and associated results). In practice, item parameters are typically calibrated by using $\theta \sim N(0, 1)$ to obtain the model identification. So, it is quite reasonable to use the standard normal density (denoted by $\phi(\theta)$) for $g^*(\theta)$ as is used for the prior of the BME. That is, the expectation is formerly given as

$$E\{c_{\min}^*(\theta)\} = \int_{-\infty}^{\infty} c_{\min}^*(\theta) \phi(\theta) d\theta, \quad (5.5)$$

which can be approximated by using Gaussian quadrature:

$$E\{c_{\min}^*(\theta)\} \doteq \frac{1}{\sqrt{\pi}} \sum_{i=1}^M c_{\min}^*(\sqrt{2} x_i) A_i, \quad (5.6)$$

where x_i is the i -th quadrature point, A_i is the i -th quadrature weight and M is the number of the quadrature points (see Stroud & Sechrest, 1966, Table Five, pp.217–252; Ogasawara, 2012b, Part A, Subsection 3.1).

The author tried this method with $M = 10, 15$ and 20 . However, (5.6) becomes relatively large e.g., greater than 50, and tends to be unstable. This is due to the large or infinitely large value of $c_{\min}^*(\theta)$ with θ being around 0. The instability comes from the fact that when a quadrature point happens to be close to this value, (5.6) becomes large. So, some methods to remove the instability are required. A simple method is to use $c_{\min}^*(0)$ since 0 is the mean of $N(0, 1)$ though this again gives a large value (recall Table 1). Another simple method is to take the mean or minimum of $c_{\min}^*(-1)$ and $c_{\min}^*(1)$ since ± 1 are values away from the mean by a standard deviation. When the minimum is used, we have $3.98 \sim 6.46$ from Table 1. When some information about the possible values of c_{\min}^* is available, a method similar to, but different from

that suggested by the reviewer is to use the hierarchical Bayesian model. In the first stage, c_{\min}^* is seen as a power of a prior (see (2.3)). Then, using the higher-order prior representing the information available for c_{\min}^* , c_{\min}^* is integrated out over the distribution, which can avoid the instability mentioned above.

Alternatively, we are tempted to use estimates of θ_0 e.g., $\hat{\theta}_{\text{ML}}$. However, this is dangerous because the formula of $\text{MSE}_{O(n-2)}(\cdot)$ for $c_{\min}^*(\theta_0)$ no longer hold since $c_{\min}^*(\hat{\theta}_{\text{ML}}) - c_{\min}^*(\theta_0) = O_p(n^{-1/2})$.

REFERENCES

- Agresti, A., & Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, 54, 280–288.
- Agresti, A., & Coull, B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, 52, 119–126.
- Ancombe, F. J. (1956). On estimating binomial response relations. *Biometrika*, 43, 461–464.
- Birnbaum, A. (1969). Statistical theory for logistic mental test models with a prior distribution of ability. *Journal of Mathematical Psychology*, 6, 258–276.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443–459.
- Bonett, D. G., & Price, R. M. (2007). Statistical inference for generalized Yule coefficients in 2×2 contingency tables. *Sociological Methods and Research*, 35, 429–446.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27–38.
- Gruber, M. H. J. (1998). *Improving efficiency by shrinkage: The James-Stein and ridge regression estimators*. New York: Marcel Dekker.
- Haldane, J. B. S. (1956). The estimation and significance of the logarithm of a ratio of frequencies. *Annals of Human Genetics*, 20, 309–311.
- Hitchcock, S. E. (1962). A note on the estimation of the parameters of the logistic function, using the minimum logit χ^2 method. *Biometrika*, 49, 250–252.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London, Series A*, 186, 453–461.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Clarendon.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.
- Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation* (2nd ed.). New York: Springer.
- Lord, F. M. (1953). An application of confidence intervals and of maximum likelihood to the estimation of an examinee’s ability. *Psychometrika*, 18, 57–76.
- Lord, F. M. (1975). *Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters*. (Research Bulletin 75-33). Princeton, NJ: Educational Testing Service.
- Lord, F. M. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika*, 48, 233–245.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Ogasawara, H. (2012a). Asymptotic expansions for the ability estimator in item response theory. *Computational Statistics*, 27, 661–683.
- Ogasawara, H. (2012b). Supplement to the papers on “the polyserial correlation coefficients” and “discrepancy functions for general covariance structures”. *Economic Review (Otaru University of Commerce)*, 62(4), 115–164. <http://www.res.otaru-uc.ac.jp/~hogasa/>,

<http://barrel.ih.otaru-uc.ac.jp/>.

- Ogasawara, H. (2013a). Asymptotic properties of the Bayes and pseudo Bayes estimators of ability in item response theory. *Journal of Multivariate Analysis*, 114, 359–377.
- Ogasawara, H. (2013b). Asymptotic cumulants of ability estimators using fallible item parameters. *Journal of Multivariate Analysis*, 119, 144–162.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No.17.
- Stroud, A. H., & Sechrest, D. (1966). *Gaussian quadrature formulas*. Englewood Cliffs, NJ: Prentice-Hall.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427–450.
- Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22, 209–212.

(Received October 15 2013, Revised December 5 2013)