PSYCHOMETRIKA—VOL. 81, NO. 1, 184–200
MARCH 2016
DOI: 10.1007/s11336-015-9443-3

EFFICIENT STANDARD ERROR FORMULAS OF ABILITY ESTIMATORS WITH DICHOTOMOUS ITEM RESPONSE MODELS

DAVID MAGIS

UNIVERSITY OF LIÈGE AND KU LEUVEN

This paper focuses on the computation of asymptotic standard errors (ASE) of ability estimators with dichotomous item response models. A general framework is considered, and ability estimators are defined from a very restricted set of assumptions and formulas. This approach encompasses most standard methods such as maximum likelihood, weighted likelihood, maximum a posteriori, and robust estimators. A general formula for the ASE is derived from the theory of M-estimation. Well-known results are found back as particular cases for the maximum and robust estimators, while new ASE proposals for the weighted likelihood and maximum a posteriori estimators are presented. These new formulas are compared to traditional ones by means of a simulation study under Rasch modeling.

Key words: item response theory, ability estimation, asymptotic standard error, maximum likelihood, weighted likelihood, Bayesian estimation, Robust estimation.

1. Introduction

One of the primary goals of item response theory (IRT) is to provide statistical methods to estimate latent ability levels of the respondents. In the framework of unidimensional IRT and dichotomously scored items, the most common ability estimators are the maximum likelihood (ML; Birnbaum 1968; Lord 1980), maximum a posteriori (MAP; Birnbaum 1969; Hambleton & Swaminathan 1985; Lord 1986; Mislevy 1986), expected a posteriori (EAP; Bock & Mislevy 1982) and weighted likelihood (WL; Warm 1989) methods. Another approach, called robust estimation (Wainer & Wright 1980; Mislevy & Bock 1982), was recently reconsidered in the field of person fit identification (Magis 2014a; Schuster & Yuan 2011). All those methods provide point estimates of the ability levels of the respondents, given their response patterns to the questionnaire and an appropriate calibration of the items according to some selected IRT model. These ability estimates can then be used to classify the respondents or select among the highest or lowest ones on the ability scale.

Although many studies focused on the statistical and psychometric properties of these various ability estimators, very few focused on the precision of these estimators, that is, the computation of their asymptotic standard errors (ASEs). Only sparse information is available from IRT textbooks and papers. The computation of ASE values remains, nevertheless, a central aspect of IRT as they measure the relative precision of the ability estimates, which is mandatory for many secondary analyses (such as classification or identification of high- or low-able respondents). The ASEs are also most useful to infer confidence intervals for ability levels or as stopping rules in computerized adaptive testing (CAT; Wainer 2000).

The most documented method is the ML approach, for which it is well known that the ASE is the inverse squared root of the expected information function (e.g., Baker & Kim 2004; Embretson & Reise 2000; Hambleton & Swaminathan 1985). The ASE of the EAP method is

Correspondence should be made to David Magis, Department of Education (B32), University of Liège, Boulevard du Rectorat 5, 4000 Liège, Belgium. Email: david.magis@ulg.ac.be

© 2015 The Psychometric Society

also well established and in line with the definition of the EAP estimator itself (Bock & Mislevy 1982; Embretson & Reise 2000). For MAP, the ASE is sometimes also referred to (and computed as) the posterior standard deviation and an explicit formula can be found in Wainer (Wainer 2000, p. 74). The ASE of some class of robust estimators was recently established (Magis 2014a).

Finally, the ASE of the WL estimator was shown to be asymptotically equivalent to that of the ML estimator (Warm 1989), but no specific formula was derived for this method so far. Making thus use of the ASE formula of the ML estimator, but plugging-in the WL estimate instead, is the strategy advised by Warm (2007) and implemented in the software ConQuest (Wu, Adams, & Wilson 1997) and the R package "PP" (Reif 2014), among others. Unfortunately, alternative formulas for the ASE of the WL estimator were proposed, among others, by Magis & Raîche (2012), Nydick (2013), and Partchev (2012), and implemented, respectively, in the R packages "catR", "catIrt," and "irtoys". This bunch of different formulas leaves some ambiguity in the correct use of an appropriate ASE value for the WL estimator.

Because each ability estimator was introduced separately and in different IRT contexts, the development of ASE formulas was also performed independently of each other approach. However, all aforementioned methods (but the EAP) belong to a common class of ability estimators, usually referred to (in statistical inference) as the class of M-estimators (Huber 1964, 1967). Noticing this common characteristic (as will be discussed more in detail in the next section) opens new possibilities to computing ASE formulas from a common framework instead of parallel derivations (per method). In the same spirit, Ogasawara (2013a) derived asymptotic cumulants of such estimators from a very general theoretical framework, including all aforementioned (but the robust) estimators.

The purpose of this paper is threefold: (a) to consider a general class of ability estimators and to derive an approximate ASE formula for the whole class; (b) to derive the particular ASE formulas for the main ability estimators (ML, MAP, WL, and robust) that belong to this class; and (c) to perform a simulation study in order to compare the traditional and the newly suggested ASE formulas, whenever the latter differed from the former (which is actually the case for the MAP and WL estimators). More precisely, the class of estimators will be defined from a general solving equation approach that directly relates to the M-estimation (Huber 1964, 1967) or estimating equation (Yuan & Jennrich 1998) frameworks.

In contrast to Ogasawara's (Ogasawara 2013a) developments, we stick to first-order approximations of Taylor expansions in the forthcoming developments. The main reason is to restrict to a simpler framework, yielding approximate yet accurate, simple, and efficient formulas for practical considerations. It is expected indeed that though relying on first-order approximations (and thus differing from the true ASE values), these simpler formulas are easy to compute, facilitating therefore their overall dissemination in scientific literature and computer software. Their accuracy, however, has to be further established by simulation studies. The first part of the paper focuses on the former objective (derivation of simple approximate ASE formulas), while the latter objective (checking the accuracy of newly suggested formulas) is investigated in the second part.

2. Framework

Consider a test of n items and let θ_0 be the true ability level, to be estimated, of the respondent of interest. Set X_i $(i = 1, \ldots, n)$ as the response of the respondent to item i, coded as zero for an incorrect response and one for a correct response. The probability of answering the item correctly is specified through an IRT model, as a function of the ability level θ:

$$P_i(\theta) = \Pr(X_i = 1|\theta, \boldsymbol{p}_i), \tag{1}$$

where the item parameters are summarized through the vector $\boldsymbol{p}_i$. In this paper, it is assumed that the item parameters are fixed to known values, obtained for instance from previous calibration of the test. Hence, only the ability level θ of the respondent must be estimated.

The class of ability estimators to be considered is defined as follows. Set

$$G_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\theta, n) \tag{2}$$

as the arithmetic mean of suitable functions $g_i(\theta, n)$ of θ with the following generic form:

$$g_i(\theta, n) = a_i(\theta) X_i + b_i(\theta, n), \tag{3}$$

where $a_i(\theta)$ is a function of θ only and $b_i(\theta, n)$ is a function of θ and the test length n. Then, any estimator $\hat{\theta}$ of ability that satisfies the solving equation $G_n\left(\hat{\theta}\right) = 0$ belongs to the considered class. Several assumptions are required in this context:

(a) The item responses X_i are locally independent and distributed as Bernoulli variables with success probabilities given by $P_i(\theta)$.
(b) In a neighborhood of $\hat{\theta}$, the item response probabilities $P_i(\theta)$ are bounded away from zero and one.
(c) In a neighborhood of $\hat{\theta}$, the functions $a_i(\theta)$ and $b_i(\theta, n)$, as well as their first derivatives (with respect to θ) $a_i'(\theta)$ and $b_i'(\theta, n)$, are bounded away from zero (whatever the test length n).
(d) In a neighborhood of $\hat{\theta}$, the function $a_i(\theta)$ and its first derivative (with respect to θ) $a_i'(\theta)$ are bounded. That is, it exists constants c^* and c^{**} such that $|a_i(\theta)| \leq c^*$ and $\left|a_i'(\theta)\right| \leq c^{**}$ for any ability level.

These assumptions are somewhat standard for asymptotic derivations in dichotomous IRT models (Lord 1983; Magis 2014a; Warm 1989). Note that assumptions (c) and (d) imply some restrictions on the first and second derivatives of the item response probabilities $P_i(\theta)$.

2.1. *Particular Ability Estimators*

Several well-known ability estimators belong to this class, such as the maximum likelihood (ML), maximum a posteriori (MAP), the weighted likelihood (WL), and some robust estimators. Set first the likelihood $L(\theta)$ and the log-likelihood $l(\theta)$ functions as

$$L(\theta) = \prod_{i=1}^{n} P_i(\theta)^{X_i} Q_i(\theta)^{1-X_i} \tag{4}$$

and

$$l(\theta) = \sum_{i=1}^{n} \{X_i \log P_i(\theta) + (1 - X_i) \log Q_i(\theta)\}, \tag{5}$$

with $Q_i(\theta) = 1 - P_i(\theta)$. Then, the ML estimator is defined by equating the first derivative of $l(\theta)$ to zero, which is equivalent to solving the equation

$$\frac{1}{n}\sum_{i=1}^{n} [X_i - P_i(\theta)] \frac{P_{i'}(\theta)}{P_i(\theta) Q_i(\theta)} = 0, \tag{6}$$

with $P_{i'}(\theta)$ standing for the first derivative of $P_i(\theta)$ with respect to θ. In other words, the ML method belongs to the class of estimators by setting

$$a_i(\theta) = \frac{P_{i'}(\theta)}{P_i(\theta)\,Q_i(\theta)} \quad \text{and} \quad b_i(\theta, n) = -\frac{P_{i'}(\theta)}{Q_i(\theta)}. \tag{7}$$

Similarly, the MAP estimator is defined by the solving equation

$$\frac{\mathrm{d}\log f(\theta)}{\mathrm{d}\theta} + \frac{\mathrm{d}\,l(\theta)}{\mathrm{d}\theta} = 0, \tag{8}$$

where $f(\theta)$ is the prior distribution of the ability level, or equivalently

$$\frac{1}{n}\sum\nolimits_{i=1}^{n}\left\{[X_i - P_i(\theta)]\,\frac{P_{i'}(\theta)}{P_i(\theta)\,Q_i(\theta)} + \frac{1}{n}\frac{\mathrm{d}\log f(\theta)}{\mathrm{d}\theta}\right\} = 0. \tag{9}$$

Thus, it also belongs to the selected class of estimators by setting

$$a_i(\theta) = \frac{P_{i'}(\theta)}{P_i(\theta)\,Q_i(\theta)} \quad \text{and} \quad b_i(\theta, n) = -\frac{P_{i'}(\theta)}{Q_i(\theta)} + \frac{1}{n}\frac{\mathrm{d}\log f(\theta)}{\mathrm{d}\theta}. \tag{10}$$

The WL estimator is introduced as an alternative to ML estimation by weighting the likelihood function (Warm 1989). It is obtained by solving the following equation:

$$\frac{J(\theta, n)}{2\,I(\theta, n)} + \frac{\mathrm{d}\,l(\theta)}{\mathrm{d}\theta} = 0, \tag{11}$$

with $l(\theta)$ being equal to (5) and

$$I(\theta, n) = \sum\nolimits_{i=1}^{n}\frac{P_{i'}(\theta)^2}{P_i(\theta)\,Q_i(\theta)} \quad \text{and} \quad J(\theta, n) = \sum\nolimits_{i=1}^{n}\frac{P_{i'}(\theta)\,P_{i''}(\theta)}{P_i(\theta)\,Q_i(\theta)}, \tag{12}$$

and $P_{i''}(\theta)$ is the second derivative of $P_i(\theta)$ with respect to θ. Setting thus

$$a_i(\theta) = \frac{P_{i'}(\theta)}{P_i(\theta)\,Q_i(\theta)} \quad \text{and} \quad b_i(\theta, n) = -\frac{P_{i'}(\theta)}{Q_i(\theta)} + \frac{J(\theta, n)}{2\,n\,I(\theta, n)} \tag{13}$$

yields the expected belonging of the WL estimator to this class.

Finally, robust estimators of ability can also be considered in this framework, up to some appropriate constraints. By essence, robust estimators are solutions of the solving equation (Mislevy & Bock 1982):

$$\frac{1}{n}\sum\nolimits_{i=1}^{n}\omega_i(\theta)\,\frac{\mathrm{d}\,l_i(\theta)}{\mathrm{d}\theta} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\omega_i(\theta)\,[X_i - P_i(\theta)]\frac{P_{i'}(\theta)}{P_i(\theta)\,Q_i(\theta)} = 0, \tag{14}$$

where $\omega_i(\theta)$ are suitable weight functions that depend on θ but not on the item responses X_i (such weight functions are said to belong to the Mallows class; see e.g., Carroll & Pederson 1993). Possible weight functions include Tukey's biweight function (Mosteller & Tukey 1977;

see also Mislevy & Bock 1982) or Huber-type weight functions (Schuster & Yuan 2011). Whatever the choice of the weight function among the Mallows class, the robust estimator belongs to the considered class of estimators, by setting

$$a_i(\theta) = \omega_i(\theta)\frac{P_{i'}(\theta)}{P_i(\theta)\,Q_i(\theta)} \quad \text{and} \quad b_i(\theta, n) = -\omega_i(\theta)\frac{P_{i'}(\theta)}{Q_i(\theta)}. \tag{15}$$

The first two columns of Table 1 summarize the various values of $a_i(\theta)$ and $b_i(\theta, n)$ for the four aforementioned ability estimators.

It is important to mention that though being very broad, this class of ability estimators does not contain all usual methods. The EAP approach, for instance, is based on numerical integration of the posterior distribution of ability which is not in line with the current approach. In the same vein, nonparametric approaches to ability estimation (e.g., Sijisma and Molenaar 2002) also rely on different conceptual approaches that do not match the present framework. Thus, the paper restricts mostly to maximization-based estimators (in the sense that maximization of some optimization function is required to provide ability estimates).

3. Asymptotic Distributions and Standard Errors

The purpose of this section is to derive an approximate formula for the asymptotic standard error of an ability estimator $\hat{\theta}$ that belongs to the class defined by (2) and (3). The terms "approximate" and "asymptotic" deserve further explanation. First, by "asymptotic," it is meant that the test length n increases up to very large values (in theory to infinity). Second, "approximate" refers to an ASE formula derived from asymptotic Taylor expansions but limited to the first order. Although including higher-order terms in the developments lead to more complex, yet more asymptotically accurate, formulas for the ML, MAP, and WL estimators (Ogasawara 2013a), emphasis is put on deriving simple and easily applicable ASE formulas for the broad class of ability estimators (and subsequently most commonly used methods). Determining whether these approximate formulas are practically efficient is the topic of the next section.

The following approach comes from the general theory of M-estimation (Huber 1964, 1967, 1981) and follows a similar development of Stefanski & Boos (2002) and Zeileis (2006), among others. Note that unlike the classical M-estimation framework, the random variables are assumed to be independent but not identically distributed, as stated by assumption (a). Thus, weaker forms of the law of large numbers and the central limit theorem will be required. Further details about these weak forms can be found in e.g., Koralov & Sinai (2007) and Rao (1984).

The following result describes the approximate ASE formula for an ability estimator selected from the broad class of estimators defined by (2) and (3).

Result 1. *For any ability estimator $\hat{\theta}$ belonging to the class defined by* (2) *and* (3) *and satisfying the assumptions* (a)–(d)*, the approximate ASE formula $\widehat{ASE}\left(\hat{\theta}\right)$ is given by*

$$\widehat{ASE}\left(\hat{\theta}\right) = \frac{\sqrt{\sum_{i=1}^{n} a_i\left(\hat{\theta}\right)^2 P_i\left(\hat{\theta}\right) Q_i\left(\hat{\theta}\right)}}{\left|\sum_{i=1}^{n}\left\{a_i'\left(\hat{\theta}\right) P_i\left(\hat{\theta}\right) + b_i'(\hat{\theta}, n)\right\}\right|}. \tag{16}$$

Formula (16) indicates that an approximate ASE value can be computed on the basis of five functions only: the item response probability $P_i(\theta)$ and its complementary $Q_i(\theta)$, the function $a_i(\theta)$ and the derivatives $a_i'(\theta)$ and $b_i'(\theta, n)$. A sketch of the proof of this result is provided in the Appendix 1.

TABLE 1.
Summary of useful functions to compute the estimated ASE of different ability estimators.

Estimator	$a_i(\theta)$	$b_i(\theta, n)$	$a_i(\theta)^2 P_i(\theta) Q_i(\theta)$	$a_i'(\theta) P_i(\theta) + b_i'(\theta, n)$	$\widehat{ASE}(\theta)$
ML	$\frac{P_{i'}(\theta)}{P_i(\theta) Q_i(\theta)}$	$-\frac{P_{i'}(\theta)}{Q_i(\theta)}$	$I_i(\theta)$	$-I_i(\theta)$	$\frac{1}{\sqrt{I(\hat{\theta}, n)}}$
MAP	$\frac{P_{i'}(\theta)}{P_i(\theta) Q_i(\theta)}$	$-\frac{P_{i'}(\theta)}{Q_i(\theta)} + \frac{1}{n}\frac{\mathrm{d}\log f(\theta)}{\mathrm{d}\theta}$	$I_i(\theta)$	$-I_i(\theta) + \frac{1}{n}\frac{\mathrm{d}^2 \log f(\theta)}{\mathrm{d}\theta^2}$	$\frac{\sqrt{I(\hat{\theta}, n)}}{\left\lvert I(\hat{\theta}, n) - \frac{\mathrm{d}^2 \log f(\theta)}{\mathrm{d}\theta^2}\right\rvert}$
WL	$\frac{P_{i'}(\theta)}{P_i(\theta) Q_i(\theta)}$	$-\frac{P_{i'}(\theta)}{Q_i(\theta)} + \frac{J(\theta)}{2 n I(\theta)}$	$I_i(\theta)$	$-I_i(\theta) + \frac{1}{n}\ \frac{J'(\theta,n) I(\theta,n) - J(\theta,n) I'(\theta,n)}{2 I(\theta,n)^2}$	$\frac{\sqrt{I(\hat{\theta}, n)}}{\left\lvert I(\hat{\theta}) - \frac{J'(\theta,n) I(\theta,n) - J(\theta,n) I'(\theta,n)}{2 I(\theta,n)^2}\right\rvert}$
Robust	$\omega_i(\theta) \frac{P_{i'}(\theta)}{P_i(\theta) Q_i(\theta)}$	$-\omega_i(\theta) \frac{P_{i'}(\theta)}{Q_i(\theta)}$	$\omega_i(\theta)^2 I_i(\theta)$	$-\omega_i(\theta) I_i(\theta)$	$\frac{\sqrt{\sum_{i=1}^n \omega_i(\theta)^2 I_i(\theta)}}{\sum_{i=1}^n \omega_i(\theta) I_i(\theta)}$

3.1. Asymptotic Distributions of Particular Estimators

Formula (16) is general to any ability estimator that belongs to the considered class. Rewriting it to the specific case of the aforementioned ability estimators (ML, MAP, WL, and robust) yields the following formulas.

Result 2. *In the framework of Result 1, the following approximate ASE formulas hold:*

$$\widehat{ASE}\left(\hat{\theta}_{ML}\right) = \frac{1}{\sqrt{I\left(\hat{\theta}_{ML}, n\right)}} \tag{17}$$

for the ML estimator;

$$\widehat{ASE}\left(\hat{\theta}_{MAP}\right) = \frac{\sqrt{I\left(\hat{\theta}_{MAP}, n\right)}}{\left|I\left(\hat{\theta}_{MAP}, n\right) - \left[\frac{\mathrm{d}^2 \log f(\theta)}{\mathrm{d}\theta^2}\right]_{\theta=\hat{\theta}_{MAP}}\right|} \tag{18}$$

for the MAP estimator;

$$\widehat{ASE}\left(\hat{\theta}_{WL}\right) = \frac{\sqrt{I\left(\hat{\theta}_{WL}, n\right)}}{\left|I\left(\hat{\theta}_{WL}, n\right) - F_n\left(\hat{\theta}_{WL}\right)\right|} \tag{19}$$

with

$$F_n(\theta) = \frac{J'(\theta, n)\, I(\theta, n) - J(\theta, n)\, I'(\theta, n)}{2\, I(\theta, n)^2}, \tag{20}$$

for the WL estimator; and

$$\widehat{ASE}\left(\hat{\theta}_{ROB}\right) = \frac{\sqrt{\sum_{i=1}^{n} \omega_i \left(\hat{\theta}_{ROB}\right)^2 I_i\left(\hat{\theta}_{ROB}\right)}}{\sum_{i=1}^{n} \omega_i \left(\hat{\theta}_{ROB}\right) I_i\left(\hat{\theta}_{ROB}\right)} \tag{21}$$

for the robust estimator.

To establish formulas (17)–(21), it is sufficient to rewrite the functions $a_i(\theta)^2 P_i(\theta) Q_i(\theta)$ and $a_i'(\theta) P_i(\theta) + b_i'(\theta, n)$ for each estimator, which is straightforward from the first columns of Table 1. For completeness, this table also displays these two functions together with the final approximate ASE formulas (17)–(21).

Several interesting conclusions can already be drawn from Result 2. First, the approximate ASE of the ML estimator corresponds to the usual formula (see e.g., Baker & Kim 2004; Birnbaum 1968; Embretson & Reise 2000; Hambleton & Swaminathan 1985; Lord 1980; Wainer 2000). Interestingly also, formula (21) for the approximate ASE of the robust estimator perfectly matches the result derived by Magis (2014a), though obtained with a different approach than that described here.

The other two estimators yield new formulas. The approximate ASE of the MAP estimator (18) differs from its traditional ASE formula, referred to as $\widehat{ASE}_{trad}\left(\hat{\theta}_{MAP}\right)$ here (e.g., Wainer 2000, p. 74):

$$\widehat{ASE}_{\text{trad}}\left(\hat{\theta}_{MAP}\right) = \frac{1}{\sqrt{I\left(\hat{\theta}_{MAP}, n\right) - \left[\frac{\mathrm{d}^2 \log f(\theta)}{\mathrm{d}\theta}\right]_{\theta=\hat{\theta}_{MAP}}}}. \tag{22}$$

In the particular case of a normal prior distribution with mean μ and variance σ^2, the formulas in (18) and (22) reduce, respectively, to

$$ASE\left(\hat{\theta}_{MAP}\right) = \frac{\sqrt{I(\hat{\theta}_{MAP}, n)}}{I\left(\hat{\theta}_{MAP}, n\right) + \frac{1}{\sigma^2}} \quad \text{and} \quad \widehat{ASE}_{\text{trad}}\left(\hat{\theta}_{MAP}\right) = \frac{1}{\sqrt{I\left(\hat{\theta}_{MAP}, n\right) + \frac{1}{\sigma^2}}}. \tag{23}$$

A direct comparison between (18) and (22) also shows the following relationship:

$$\widehat{ASE}\left(\hat{\theta}_{MAP}\right) \leq \widehat{ASE}_{\text{trad}}\left(\hat{\theta}_{MAP}\right), \tag{24}$$

the equality holding only if $\left[\frac{d^2 \log f(\theta)}{d\theta}\right]_{\theta=\hat{\theta}_{MAP}} = 0$, which means in the case of the normal prior distribution that the prior variance is infinite and consequently the prior distribution is flat. Relationship (24) indicates that the suggested approximate ASE formula will always return smaller estimated ASE values than the traditional formula. However, this is not evidence in itself towards an improvement of the ASE estimation formula for the MAP estimator. This will be investigated in the next section.

Finally, formula (19) provides a different ASE value than the traditional one. It was indeed established (Warm 1989) that the WL estimator has the same true ASE as the ML estimator, and this is a commonly considered formula in practice (we refer to it further as the *traditional ASE formula*). The present approach, however, proposes a specific ASE formula for the WL estimator and direct comparisons between both the types of ASE are therefore mandatory. Recall that other alternative formulas for the ASE of the WL estimator were proposed, among others, by Magis & Raîche (2012), Nydick (2013) and Partchev (2012). These suggestions, however, will not be considered further in this paper in order to focus on the comparison of traditional ASE values versus suggested ones in the previous result.

4. Simulation Study

A simulation study was conducted to compare both the traditional and newly suggested ASE formulas for both the MAP and WL estimators. The difficulty with such a study is that estimated ASE values should be compared to a gold standard, that is, the *exact standard error* (SE) of the estimator. Unfortunately, this exact SE cannot be computed easily under any dichotomous IRT model. One major exception, however, is the Rasch model, as in this case, the full sample distribution (and consequently the exact SE) can be computed with limited samples of response patterns and, therefore, within a reasonable computing time (see Magis 2014b, further information). Consequently, this study restricts to the Rasch model in order to compare approximate ASE values to the exact SE one.

4.1. Design

Four test lengths were considered: $n = 5$, 10, 20, and 40 items. The first values are useful to test the ASE formulas in case of small samples (e.g., at early stages of a CAT), wherein the

asymptotic conditions might not be fulfilled at all. Moreover, for each test length, thirteen ability levels were selected, from −3 to 3 by steps of 0.5. This yields 52 design settings, and for each of the 52 (item, ability) pairs of selected values, the following generation process was performed.

First, one set of item difficulty levels was randomly drawn from a standard normal distribution, and 1,000 response patterns were generated for this set of item parameters. Each item response was randomly drawn from a Bernoulli variable with success probability given by the Rasch model, the true item parameters, and the true ability level. The ability levels were estimated by the WL and MAP methods (with standard normal prior distribution), and both the traditional ASEs and the suggested ASEs were computed for each pattern and each ability estimator. In addition, the exact SE values $SE(\hat{\theta}_{MAP})$ and $SE(\hat{\theta}_{WL})$ of, respectively, the MAP and WL estimators were also computed per response pattern, using the procedure described in Magis (2014b). This led to six sets of 1,000 ASE values (three per ability estimator: traditional ASE, newly suggested ASE and exact SE). The 1,000 traditional and suggested ASE values were summarized by computing the average signed bias (ASB):

$$ASB = \frac{1}{1,000}\sum_{j=1}^{1,000}\left[\widehat{ASE}^{(j)}\left(\hat{\theta}_{MAP}\right) - SE(\hat{\theta}_{MAP})\right], \tag{25}$$

where $\widehat{ASE}^{(j)}\left(\hat{\theta}_{MAP}\right)$ is the j-th estimated ASE value of the MAP estimator, $j = 1, \ldots, 1000$, and the root mean squared error (RMSE) is

$$RMSE = \sqrt{\frac{1}{1,000}\sum_{j=1}^{1,000}\left[\widehat{ASE}^{(j)}\left(\hat{\theta}_{MAP}\right) - SE(\hat{\theta}_{MAP})\right]^2}. \tag{26}$$

Equations (25) and (26) are written for the suggested ASE (18) and corresponding values for the usual ASE (22) are derived by plugging-in $\widehat{ASE}^{(j)}_{\text{trad}}\left(\hat{\theta}_{MAP}\right)$ instead of $\widehat{ASE}^{(j)}\left(\hat{\theta}_{MAP}\right)$. Similar values are also derived for the WL estimator instead.

This whole generation process was replicated 100 times per cell (with different item parameters per replication), leading to 100 sets of ASB and RMSE values for the ASE of both the MAP and the WL estimators. These ASB and RMSE values were eventually averaged out to return global ASB and RMSE indices for each test length and ability level. The reason for replicating this design is to remove the possible strong dependency of ASB and RMSE values to a specific set of item parameters.

Note that the ASB and the RMSE values were also computed for the MAP estimator itself, but the results are not reported since (a) they do not fit the scope of this study and (b) the behavior of the MAP estimator has been already studied in previous researches (e.g., Lord 1983, 1986), and the results from this study are in line with these previous researches.

The simulation study was conducted using an implementation in the R software (R Core Team 2014). The R code can be obtained freely from the first author.

4.2. Results

Figure 1 depicts the ASB values of the two ASE formulas, the usual one (referred to as "Usual") and the suggested one (referred to as "New"), for both the ability estimators (MAP and WL) and all true ability levels.

First, the ASB values are all positive except for the suggested ASE of the WL estimator that exhibits slightly negative ASB values around zero with very short tests. This indicates that globally the ASE formulas tend to overestimate the exact SE values. Although it tends to disappear with

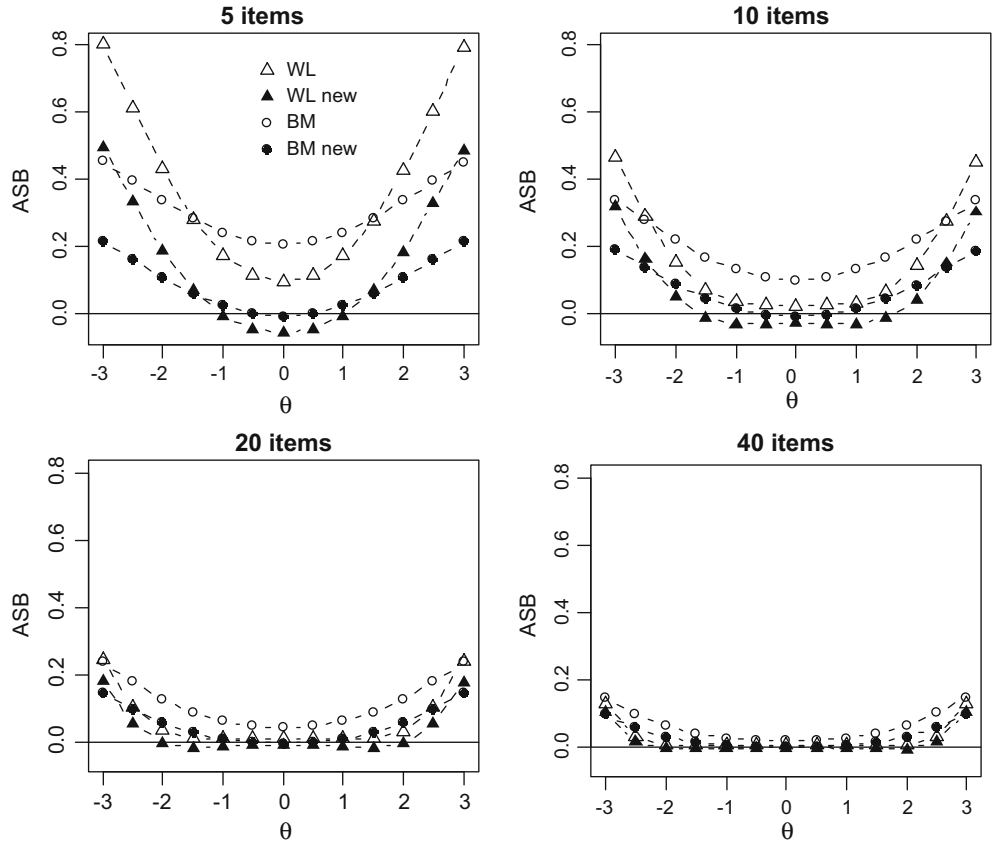

FIGURE 1.
ASB values of the traditional ASE and the suggested ("new") ASE of the MAP and WL estimators of ability (under Rasch model).

longer tests, the case of slightly negative values around the center of the ability scale deserves further investigation. One possible explanation is that in this setting, the correction term $F_n(\theta)$ (20) in the denominator of (19) leads to an important decrease of the ASE and thus to under-estimation of the true SE.

Second, the ASB is smallest for true ability levels close to zero and largest at the extremes of the ability scale, independently of the test length and the ability estimator. This is obvious as the item difficulties were drawn from standard normal distributions, leading to more informative items at the center of the difficulty scale and hence more precise estimates of ability levels. Third, the ASB curves converge toward zero as the test length increases, which is also expected since longer tests yield more information and hence more precision for ability estimation. Finally, at given ability level and test length, the ASB of the suggested ("new") ASEs are always smaller than the corresponding ASB of the usual ASEs.

The most interesting finding is that even with very short tests, the suggested ASEs tend to be almost unbiased around the center of the ability scale, while increasing the test length yields almost unbiased ASEs on the whole scale. Whatever the test length and true ability level, the usual ASEs are always more positively biased (i.e., return larger bias than the suggested ASEs). In other words, the suggested ASE of the MAP and WL estimators tend to be more accurate in this context and with short tests. Note, however, that the gap between traditional and suggested ASE formulas (in terms of ASB) decreases quickly when the test length increases.

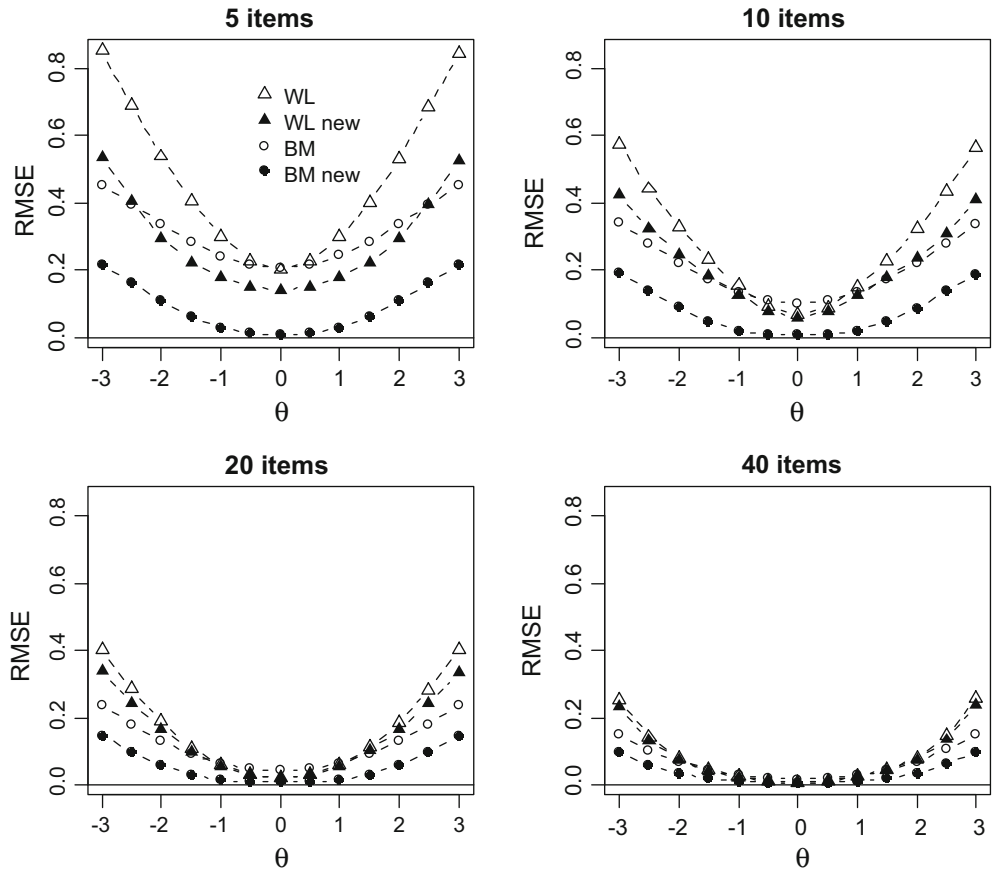

FIGURE 2.
RMSE values of the traditional ASE and the suggested ("new") ASE of the MAP and WL estimators of ability (under Rasch model).

Figure 2 shows corresponding RMSE values with the same display as in Fig. 1. Similar trends can be observed here. RMSE values decrease toward zero with increasing test length; larger RMSE values are observed at the extremes of the ability scale; and the suggested ASE formulas yield smaller RMSE values than the traditional ASE formula (for both the estimators). In other words, the bias reduction in computing the ASE with the newly suggested approach is not penalized by an important increase in variability, and therefore, the global performance (in terms of RMSE) of the suggested formulas is improved.

5. Discussion

The purpose of this paper was to consider the asymptotic distribution of a broad class of ability estimators in dichotomous IRT models. The ASE of these estimators was subsequently derived, first in its general form and then explicitly for the main four types of estimators (ML, MAP, WL, and robust) that belong to this class. Several well-known results were found back from this approach, and new ASE formulas were obtained for the MAP and WL methods.

In the limited framework of MAP and WL estimation under the Rasch model, this newly suggested ASE formula was shown to be highly competitive with the usual alternatives.

More precisely, it was shown to be less biased overall, even with small tests of no more than 5 items. Note that both the formulas yield very close results with longer tests; nevertheless, this study establishes the usefulness of the newly derived ASE formula for the MAP and WL estimators.

It is, however, important to stress once again that this simulation study was restricted in at least two ways. First, only the Rasch model was considered. Although it is not excluded that similar trends will occur with other models (such as the two- and three-parameter logistic models), this should be fully established by means of a similar study. The main issue relates to the computational burden required for obtaining the exact SE of the estimator with more general models. Indeed, for a test of n items, all 2^n possible response patterns must be considered to derive the full sample distribution of the estimator, while under the Rasch model, it requires only $n + 1$ patterns, one for each possible test score from zero to n (see e.g., Magis 2014b). In a more general design, one could derive estimates of the true standard error by drawing a large sample of response patterns and estimating the standard deviation of all ability level estimates (in line with Magis 2014a). Another option, though it has received less attention in IRT, is to use bootstrap replicates of the responses in order to generate new response patterns from a small set of existing ones.

The second restriction of the study is that the item parameters were not estimated prior to computing the MAP estimates and the ASEs, but the generated parameters were directly imputed into the ability estimation process. This permitted to focus on the accuracy of the ASE formulas of the MAP and WL estimators without any other additional nuisance than the random generation of the item responses. In practice, however, this calibration process must be realized prior to ability estimation, but very few research focused on that issue so far. For instance, Ogasawara (2013b) derived asymptotic properties of several ability estimators when item parameters are estimated, while other authors (e.g., Doebler 2012; Patton et al. 2013) considered the impact of item calibration errors in various contexts but not specifically to ASE estimation. It would then be worth considering the calibration process as an additional design factor for future studies and determine whether using true or estimated item parameters can influence the simulation results (and to what extent).

Further comparisons should also be carried out. First, Ogasawara's (Ogasawara 2013a) framework is a very general and theoretical approach to derive the distributional characteristics of the ability estimators (including asymptotic bias and SE), though restricting to Bayes and pseudo-Bayes estimators (thus excluding the robust approach). Using higher-order terms in Taylor expansions, this method returns more complex yet possibly more asymptotically accurate versions of the ASEs. Further research is therefore recommended to compare both frameworks with comparable estimators (ML, MAP, and WL). Second, as mentioned previously several alternative ASE formulas were proposed for the WL estimator, while the present study focused only on both the newly suggested and traditional versions of the ASE. It is, however, not that straightforward to determine the theoretical origin of some of these formulas, for instance from Magis & Raîche (2012) and Partchev (2012). Nevertheless, since they were formerly proposed, a careful comparison of all suggested ASE formulas is mandatory and should deserve further research. Third, this paper pointed out the equivalence between the ASE formulas of the robust estimator from this research and from Magis (2014a). Interestingly, both were derived using different approaches (the theory of estimating equations in the latter and the M-estimation theory in the former). It would then be most useful to check whether the approach of Magis (2014a) can also be adapted to this general framework and whether similar results can be derived for other IRT estimators.

Eventually, there is no theoretical restriction for not generalizing this approach to polytomous IRT models. Since all ability estimation methods discussed in this paper (but the robust method) have been extended to the polytomous framework, it would then be straightforward to study their related ASE values, for which very few information is available so far.

Acknowledgments

The author wishes to thank the associate editor and two anonymous referees for their numerous comments and suggestions. David Magis is Research Associate of the *Fonds de la Recherche Scientifique – FNRS* (Belgium). This research was funded in part by the Research Fund of KU Leuven, Belgium (GOA/15/003) and the Interuniversity Attraction Poles program financed by the Belgian government (IAP/P7/06).

Appendix: Proof of Result 1

To sketch the proof of Result 1, we recall first three important theorems and we introduce two technical lemmas.
The first result is known as *Chebyshev's theorem* (see e.g., Rao 1984, p. 57), that is, a weak form of the law of large numbers.

Chebyshev's theorem. *if* Y_i $(i = 1, \ldots, n)$ *is a sequence of independent random variables and if* $S_n = \sum_{i=1}^{n} Y_i$ *is such that*

$$\frac{V(S_n)}{n^2} \to 0 \quad as \quad n \to \infty, \tag{27}$$

(*where* $V(.)$ *stands for the mathematical variance*), *then* S_n/n *converges in probability to its expected value* $E(S_n/n)$ (*with* $E(.)$ *standing for the mathematical expectation*). *Condition* (27) *is referred to as Chebyshev's condition.*

The first technical lemma sets that Chebyshev's condition is fulfilled for a particular choice of the random variables Y_i.

Lemma 1. *Using the notations of* (2) *and* (3), *set*

$$Y_i = g_i'(\theta, n) = a_i'(\theta) X_i + b_i'(\theta, n). \tag{28}$$

Then, under assumptions (a) to (d), the sequence of random variables Y_i $(i = 1, \ldots, n)$ *fulfills Chebyshev's condition. In other words,* $S_n/n = (\sum_{i=1}^{n} Y_i)/n$ *converges in probability toward*

$$c_n = \frac{1}{n}\sum_{i=1}^{n} E(g_i'(\theta, n)) = \frac{1}{n}\sum_{i=1}^{n} \left\{a_i'(\theta) P_i(\theta) + b_i'(\theta, n)\right\}. \tag{29}$$

Moreover, the limit c_n *is bounded away from zero.*

Proof. First, the Y_i variables defined by (28) are independent according to assumption (a), and thus,

$$V(S_n) = \sum_{i=1}^{n} V[g_i'(\theta, n)] = \sum_{i=1}^{n} a_i'(\theta)^2 P_i(\theta) Q_i(\theta). \tag{30}$$

By assumptions (b) and (d), one can directly notice that $V(S_n) \leq nc^{**}$ and

$$\frac{V(S_n)}{n^2} = \frac{1}{n}\sum_{i=1}^{n} \frac{a_i'(\theta)^2 P_i(\theta) Q_i(\theta)}{n} \leq \frac{c^{**}}{n}, \tag{31}$$

which is sufficient to establish that $V(S_n)/n^2$ converges toward zero with increasing nl, and thus, that Chebyshev's condition is fulfilled with this choice (28) of Y_i variables. The derivation of c_n is straightforward, and assumptions (b) and (c) ensure that it is bounded away from zero. □

The second important result is known as *Liapunov's theorem* (e.g., Rao 1984, p. 283–286), which is a weak version of the central limit theorem.

Liapunov's theorem. *Let Y_i $(i = 1, \ldots, n)$ be a sequence of independent random variables with mean μ_i and variance σ_i^2, and set $S_n = \sum_{i=1}^n Y_i$ and $s_n^2 = \sum_{i=1}^n \sigma_i^2$. If it exists $\delta > 0$ such that*

$$\frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E\left[|Y_i - \mu_i|^{2+\delta}\right] = \frac{\sum_{i=1}^n E\left[|Y_i - \mu_i|^{2+\delta}\right]}{\left[\sum_{i=1}^n \sigma_i^2\right]^{(2+\delta)/2}} \to 0 \quad as \quad n \to \infty, \tag{32}$$

then $[S_n - E(S_n)]/\sqrt{V(S_n)}$ converges in distribution to a standard normal random variable. Condition (32) is referred to as Liapunov's condition.

The second technical lemma is a particular application of Liapunov's theorem.

Lemma 2. *Using the notations of (2) and (3), set*

$$Y_i = g_i(\theta, n). \tag{33}$$

Then, under assumptions (a) to (d), the sequence of random variables Y_i $(i = 1, \ldots, n)$ fulfills Liapunov's condition with $\delta = 2$. In other words, $[S_n/\sqrt{n} - E(S_n)/\sqrt{n}]/\sqrt{V(S_n)/n}$ converges in distribution to a standard normal random variable, or equivalently the asymptotic variance of $S_n/\sqrt{n} = (\sum_{i=1}^n Y_i)/\sqrt{n}$ is equal to

$$v_n = \frac{1}{n} \sum_{i=1}^n V[g_i(\theta, n)] = \frac{1}{n} \sum_{i=1}^n a_i(\theta)^2 P_i(\theta) Q_i(\theta). \tag{34}$$

Proof. First,

$$\mu_i = E(Y_i) = a_i(\theta) P_i(\theta) + b_i(\theta, n) \tag{35}$$

and

$$s_n^2 = \sum_{i=1}^n V(Y_i) = \sum_{i=1}^n a_i(\theta)^2 P_i(\theta) Q_i(\theta). \tag{36}$$

It follows that $Y_i - \mu_i = a_i(\theta)[X_i - P_i(\theta)]$, and hence,

$$E\left[|Y_i - \mu_i|^{2+\delta}\right] = a_i(\theta)^4 E([X_i - P_i(\theta)]^4) \tag{37}$$

with $\delta = 2$. One can rewrite then, using (36) and (37),

$$\frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E\left[|Y_i - \mu_i|^{2+\delta}\right] = \frac{\sum_{i=1}^n a_i(\theta)^4 E\left([X_i - P_i(\theta)]^4\right)/n^2}{\left[\sum_{i=1}^n a_i(\theta)^2 P_i(\theta) Q_i(\theta)/n\right]^2}. \tag{38}$$

Now, by assumptions (b) to (d), the function $a_i\,(\theta)^2\,P_i\,(\theta)\,Q_i\,(\theta)$ is upper bounded and bounded away from zero for any item, so that the denominator of (38) is also upper bounded and bounded away from zero for any n. Moreover, by definition, $[X_i - P_i\,(\theta)]^4$ takes value $Q_i\,(\theta)^4$ with probability $P_i\,(\theta)$ and value $P_i\,(\theta)^4$ with probability $Q_i\,(\theta)$, whence

$$E\left([X_i - P_i\,(\theta)]^4\right) = P_i\,(\theta)\,Q_i\,(\theta)\,[P_i\,(\theta)^3 + Q_i\,(\theta)^3]. \tag{39}$$

In sum, the numerator of (38) is equal to

$$\frac{1}{n^2}\sum\nolimits_{i=1}^{n} a_i\,(\theta)^4\,P_i\,(\theta)\,Q_i\,(\theta)\,[P_i\,(\theta)^3 + Q_i\,(\theta)^3] \leq \frac{[c^*]^4}{n}, \tag{40}$$

according to assumptions (b) and (d). The inequality in (40) is sufficient to ensure that the numerator of (38) converges toward zero as n increases. Altogether, Liapunov's condition is satisfied with this specific choice (33) of Y_i variables. The asymptotic variance (34) is obtained from a straightforward calculation. □

The third important theorem to recall here is

Slutsky's theorem. *if the sequences of random variables X_i and Y_i $(i = 1, \ldots, n)$ are such that X_i converges in probability to some constant c and Y_i converges in distribution to the random variable Y, then the sequence $X_i Y_i$ converges to the random variable cY.*

We can now sketch the proof of Result 1.

Proof of Result 1. Assume that items 1 to n are administered to a respondent with true ability level θ_0. Start by writing the difference $\hat{\theta} - \theta_0$ using Taylor series expansion of $G_n\big(\hat{\theta}\big)$, limited to the first order (as already pointed out, this limitation yields approximate ASE values):

$$G_n\big(\hat{\theta}\big) = G_n\big(\theta_0\big) + \big(\hat{\theta} - \theta_0\big)G_n'\big(\theta_0\big), \tag{41}$$

with $G_n'\big(\theta_0\big)$ being the first derivative of $G_n\big(\theta\big)$ with respect to θ and evaluated at θ_0. Equation (41) can be rewritten as

$$\sqrt{n}\big(\hat{\theta} - \theta_0\big) = -G_n'\,(\theta_0)^{-1}\,\sqrt{n}G_n\big(\theta_0\big), \tag{42}$$

provided that $G_n'\big(\theta_0\big)$ is not zero and since $G_n\big(\hat{\theta}\big) = 0$ by definition of the ability estimator. Now, by definition,

$$G_n'\,(\theta_0) = \frac{1}{n}\sum\nolimits_{i=1}^{n} g_i'\,(\theta_0, n) = \frac{1}{n}\sum\nolimits_{i=1}^{n} Y_i, \tag{43}$$

with Y_i as defined in Lemma 1. Thus, according to this Lemma, (43) converges in probability to the constant c_n set by (29). Moreover,

$$\sqrt{n}G_n\,(\theta_0) = \frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^{n} g_i\,(\theta_0, n) = \frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^{n} Y_i, \tag{44}$$

with Y_i as defined in Lemma 2. According to this Lemma, (44) converges in distribution to a normal random variable with variance v_n given by (34).
Eventually, applying Slutsky's theorem to the right-hand side of (42), using results from (43) and (44), it is established that $\sqrt{n}\left(\hat{\theta} - \theta_0\right)$ converges in distribution to a normal random variable with variance given by $c_n^{-2} v_n$, or equivalently, that the (approximate) ASE of the ability estimator is

$$\sqrt{\frac{v_n}{n\, c_n^2}} = \frac{\sqrt{\sum_{i=1}^n a_i\left(\theta_0\right)^2 P_i\left(\theta_0\right) Q_i\left(\theta_0\right)}}{\left|\sum_{i=1}^n \left\{a_i'\left(\theta_0\right) P_i\left(\theta_0\right) + b_i'(\theta_0, n)\right\}\right|}, \tag{45}$$

and can be estimated by plugging-in the ability estimator $\hat{\theta}$ instead of the true ability level θ_0:

$$\widehat{ASE}\left(\hat{\theta}\right) = \frac{\sqrt{\sum_{i=1}^n a_i\left(\hat{\theta}\right)^2 P_i\left(\hat{\theta}\right) Q_i\left(\hat{\theta}\right)}}{\left|\sum_{i=1}^n \left\{a_i'\left(\hat{\theta}\right) P_i\left(\hat{\theta}\right) + b_i'(\hat{\theta}, n)\right\}\right|}, \tag{46}$$

which corresponds to (16). □

References

Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores (chapters)* (pp. 17–20). Reading, MA: Addison-Wesley.

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., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a micro computer environment. *Applied Psychological Measurement*, *6*, 431–444.

Carroll, R. J., & Pederson, S. (1993). On robustness in the logistic regression model. *Journal of the Royal Statistical Society: Series B*, *55*, 693–706.

Doebler, A. (2012). The problem of bias in person parameter estimation in adaptive testing. *Applied Psychological Measurement*, *36*, 255–270. doi:10.1177/0146621612443304.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New York: Erlbaum.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, *35*, 73–101. doi:10.1214/aoms/1177703732.

Huber, P.J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. In *Proceeding of the 5th Berkeley Symposium*, (vol. 1, pp. 221–233).

Huber, P. J. (1981). *Robust statistics*. New York: Wiley.

Koralov, L., & Sinai, Y. G. (2007). *Theory of probability and random processes*. New York: Springer.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. M. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika*, *48*, 233–245. doi:10.1007/BF02294018.

Lord, F. M. (1986). Maximum likelihood and Bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, *23*, 157–162. doi:10.1111/j.1745-3984.1986.tb00241.x.

Magis, D. (2014a). On the asymptotic standard error of a class of robust estimators of ability in dichotomous item response models. *British Journal of Mathematical and Statistical Psychology*, *67*, 430–450. doi:10.1111/bmsp.12027.

Magis, D. (2014b). Accuracy of asymptotic standard errors of the maximum and weighted likelihood estimators of proficiency levels with short tests. *Applied Psychological Measurement*, *38*, 105–121. doi:10.1177/0146621613496890.

Magis, D., & Raîche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. *Journal of Statistical Software*, *48*, 1–31.

Mislevy, R. J. (1986). Bayes modal estimation in item response theory. *Psychometrika*, *51*, 177–195. doi:10.1007/BF02293979.

Mislevy, R. J., & Bock, R. D. (1982). Biweight estimates of latent ability. *Educational and Psychological Measurement*, *42*, 725–737. doi:10.1177/001316448204200302.

Mosteller, F., & Tukey, J. (1977). *Exploratory data analysis and regression*. Reading, MA: Addison-Wesley.

Nydick, S.W. (2013). *catIrt: An R package for simulating IRT-based computerized adaptive tests*. R package version 0.4-1.

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. doi:10.1016/j.jmva.2012.08.013.
Ogasawara, H. (2013b). Asymptotic cumulants of the ability estimators using fallible item parameters. *Journal of Multivariate Analysis*, *119*, 144–162. doi:10.1016/j.jmva.2013.04.008.
Partchev, I. (2012). *irtoys: Simple interface to the estimation and plotting of IRT models*. R package version 0.1.6.
Patton, J. M., Cheng, Y., Yuan, K.-H., & Diao, Q. (2013). The influence of item calibration error on variable-length computerized adaptive testing. *Applied Psychological Measurement*, *37*, 24–40. doi:10.1177/0146621612461727.
R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
Rao, M. M. (1984). *Probability theory with applications*. New York: Academic Press.
Reif, M. (2014). *PP: Estimation of person parameters for the 1, 2, 3, 4-PL model and the GPCM*. R package version 0.5.3.
Schuster, C., & Yuan, K.-H. (2011). Robust estimation of latent ability in item response models. *Journal of Educational and Behavioral Statistics*, *36*, 720–735. doi:10.3102/1076998610396890.
Sijisma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.
Stefanski, L. A., & Boos, D. D. (2002). The calculus of M-estimation. *The American Statistician*, *56*, 29–38. doi:10.1198/000313002753631330.
Wainer, H. (2000). *Computerized adaptive testing: A primer* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
Wainer, H., & Wright, B. D. (1980). Robust estimation of ability in the Rasch model. *Psychometrika*, *45*, 373–391. doi:10.1007/BF02293910.
Warm, T. A. (1989). Weighted likelihood estimation of ability in item response models. *Psychometrika*, *54*, 427–450. doi:10.1007/BF02294627.
Warm, T. A. (2007). Warm (Maximum) likelihood estimates of Rasch measures. *Rasch Measurement Transactions*, *21*, 1094.
Wu, M. L., Adams, R. J., & Wilson, M. R. (1997). *ConQuest: Multi-aspect test software [Computer program]*. Camberwell, Australia: Australian Council for Educational Research.
Yuan, K.-H., & Jennrich, R. I. (1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis*, *65*, 245–260. doi:10.1006/jmva.1997.1731.
Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, *16*(9), 1–16.

Manuscript Received: 4 DEC 2013
Published Online Date: 18 FEB 2015