PSYCHOMETRIKA—VOL. 77, NO. 1, 163–169
JANUARY 2012
DOI: 10.1007/S11336-011-9233-5

ON THE RELATIONSHIPS BETWEEN JEFFREYS MODAL AND WEIGHTED LIKELIHOOD ESTIMATION OF ABILITY UNDER LOGISTIC IRT MODELS

DAVID MAGIS

UNIVERSITY OF LIÈGE AND K.U. LEUVEN

GILLES RAÎCHE

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

This paper focuses on two estimators of ability with logistic item response theory models: the Bayesian modal (BM) estimator and the weighted likelihood (WL) estimator. For the BM estimator, Jeffreys' prior distribution is considered, and the corresponding estimator is referred to as the Jeffreys modal (JM) estimator. It is established that under the three-parameter logistic model, the JM estimator returns larger estimates than the WL estimator. Several implications of this result are outlined.

Key words: logistic model, Bayesian modal estimation, Jeffreys' prior, weighted likelihood estimation.

1. Introduction

This paper focuses on the estimation of subject ability in the framework of item response theory (IRT). Consider a test of n items and let $P_i(\theta)$ $(i = 1, \dots, n)$ be the probability of answering item i correctly. The parameter θ denotes the latent ability of the subject and has to be estimated. Set X_i as the response of the subject to item i, coded as 1 for a correct answer and 0 for an incorrect answer. The present paper is restricted to the dichotomous logistic IRT models, and in particular to the three-parameter logistic (3PL) model (Birnbaum, 1968):

$$P_i(\theta) = \Pr(X_i = 1 \mid \theta, a_i, b_i, c_i) = c_i + (1 - c_i)\frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]} \tag{1}$$

where a_i, b_i and c_i are, respectively, the discrimination, the difficulty and the pseudo-guessing parameters of item i. Fixing all pseudo-guessing parameters to zero yields the two-parameter logistic (2PL) model. The one-parameter logistic (1PL), or Rasch model (Rasch, 1960), is obtained by also fixing all discrimination parameters to one. The three item parameters are assumed to be known and not to be estimated. Subjects' abilities are estimated conditionally on these fixed item parameters. For this reason, the response probability (1) depends on the ability level θ only, which motivates the short notation $P_i(\theta)$.

The main goal of this paper is to study the particular relationships between the Bayesian modal (BM) estimator, suggested by Birnbaum (1969), and the weighted likelihood (WL) estimator introduced by Warm (1989). The BM estimator involves the selection of a suitable prior distribution for the distribution of abilities in the target population. The WL estimator was developed mainly to cancel the bias of the maximum likelihood estimator. Although conceptually different, these estimators are closely related with an accurate selection of the prior distribution. Hoijtink and Boomsma (1995, p. 57) mention that under the Rasch model, Warm's WL estimator and BM estimator are completely equivalent when the prior distribution is the Jeffreys'

Requests for reprints should be sent to David Magis, Department of Mathematics (B37), University of Liège, Grande Traverse 12, 4000 Liège, Belgium. E-mail: david.magis@ulg.ac.be non-informative prior density (Jeffreys, 1939, 1946). In the following, we call *Jeffreys modal* or JM estimator, the BM estimator with Jeffreys' prior distribution. The relationship between JM and WL estimators under the Rasch model permits bridging the gap between the (weighted) likelihood and the Bayesian estimation paradigms. Moreover, Warm (1989) noticed that when all pseudo-guessing parameters of the 3PL model are equal to zero, an appropriate choice of the weighting function is the square root of the information function (Warm, 1989, p. 431). Although it was not clearly stated by Warm, this approach corresponds to the selection of Jeffreys' prior for Bayesian estimation of ability (see also Meijer & Nering, 1999).

However, the comparison of JM and WL estimators has apparently not been studied yet under the general 3PL model. This extension is the main purpose of this paper. We start by presenting briefly the methods of ability estimation, before establishing the particular relationships between JM and WL estimators under the 3PL model.

2. Estimation of Ability

The starting point is the maximum likelihood (ML) estimator of ability $\hat{\theta}_{\mathrm{ML}}$ (Lord, 1980). It is defined as the value of θ which maximizes the likelihood function

$$L(\theta) = \prod_{i=1}^{n} P_i(\theta)^{X_i} Q_i(\theta)^{1-X_i} \tag{2}$$

where $Q_i(\theta) = 1 - P_i(\theta)$ is the probability of an incorrect response. Equivalently, the ML estimator is obtained by maximizing the log-likelihood function

$$\log L(\theta) = \sum_{i=1}^{n} \bigl\{ X_i \log P_i(\theta) + (1 - X_i) \log Q_i(\theta) \bigr\} \tag{3}$$

or by equating the first derivative of the log-likelihood (3) to zero:

$$\frac{\partial \log L(\theta)}{\partial \theta} = 0. \tag{4}$$

The standard error of $\hat{\theta}_{\mathrm{ML}}$ is estimated by

$$se(\hat{\theta}_{\mathrm{ML}}) = \frac{1}{\sqrt{I(\hat{\theta}_{\mathrm{ML}})}} \tag{5}$$

where $I(\theta)$ is the information function:

$$I(\theta) = -E\left(\frac{\partial^2 \log L(\theta)}{\partial \theta} \right) \tag{6}$$

and E stands for the mathematical expectation. Note that for any item response model with success probability $P_i(\theta)$, the information function (6) can be expressed as follows:

$$I(\theta) = \sum_{i=1}^{n} \frac{[P_i'(\theta)]^2}{P_i(\theta) Q_i(\theta)}, \tag{7}$$

where $P_i'(\theta)$ is the first derivative of $P_i(\theta)$ with respect to θ.

The Bayes modal (or maximum a posteriori) estimator $\hat{\theta}_{\mathrm{BM}}$ is obtained by maximizing the posterior density $g(\theta)$ of θ, that is, the product of a prior density $f(\theta)$ and the likelihood function $L(\theta)$ (Birnbaum, 1969). Thus, the BM estimator is obtained by maximizing the log-posterior distribution $\log g(\theta) = \log f(\theta) + \log L(\theta)$, or equivalently, by satisfying

$$\frac{\partial \log f(\theta)}{\partial \theta} + \frac{\partial \log L(\theta)}{\partial \theta} = 0. \tag{8}$$

The prior distribution $f(\theta)$ reflects some a priori knowledge or belief about the distribution of the abilities in the target population of subjects. Standard choices for the prior distribution $f(\theta)$ are the uniform distribution (on a pre-specified range of θ values) and the normal distribution. In this paper however, we focus on Jeffreys' non-informative prior density (Jeffreys, 1939, 1946), which is proportional to the square root of the information function:

$$f(\theta) \propto \sqrt{I(\theta)}. \tag{9}$$

As announced above, the BM estimator with Jeffreys' prior distribution is referred to as the *Jeffreys modal* (JM) estimator and is denoted by $\hat{\theta}_{\mathrm{JM}}$. Inserting (9) into (8), it comes that $\hat{\theta}_{\mathrm{JM}}$ must satisfy the condition

$$\frac{I'(\theta)}{2I(\theta)} + \frac{\partial \log L(\theta)}{\partial \theta} = 0, \tag{10}$$

where $I'(\theta)$ is the first derivative of $I(\theta)$ with respect to θ. Jeffreys' prior is often called a non-informative prior distribution, in the sense that it only requires the specification of the item response model, for instance the 3PL model (1), and the item parameter values. It can therefore be seen as a "test-driven" prior, adding more prior belief to θ levels which are more informative with respect to the test.

To complete the Bayesian framework, we mention the formula for estimating the standard error of any BM estimator:

$$se(\hat{\theta}_{\mathrm{BM}}) = \frac{1}{\sqrt{-\frac{\partial^2 \log f(\theta)}{\partial \theta^2}\big|_{\hat{\theta}_{\mathrm{BM}}} + I(\hat{\theta}_{\mathrm{BM}})}}. \tag{11}$$

For instance, if $f(\theta)$ is the normal distribution with mean μ and variance σ^2, then (11) reduces to

$$se(\hat{\theta}_{\mathrm{BM}}) = \frac{1}{\sqrt{\frac{1}{\sigma^2} + I(\hat{\theta}_{\mathrm{BM}})}}, \tag{12}$$

while for JM estimator, it is equal to

$$se(\hat{\theta}_{\mathrm{JM}}) = \frac{1}{\sqrt{\frac{I''(\hat{\theta}_{\mathrm{JM}}) I(\hat{\theta}_{\mathrm{JM}}) + I'(\hat{\theta}_{\mathrm{JM}})^2}{2I(\hat{\theta}_{\mathrm{JM}})^2} + I(\hat{\theta}_{\mathrm{JM}})}}, \tag{13}$$

and $I''(\theta)$ is the second derivative of $I(\theta)$ with respect to θ.

Both ML and BM estimators are biased estimators. Lord (1983, 1984), among others, showed that their bias is proportional to the inverse of the test length n. Starting from Lord's developments, Warm (1989) suggested to maximize a weighted version of the likelihood function. Up to the selection of a convenient weighting function $f(\theta)$, the estimator of θ which maximizes $g(\theta) = f(\theta) L(\theta)$ is asymptotically unbiased. Strictly speaking, the function $f(\theta)$ is not a prior density in the Bayesian sense, but only a suitable weighing function for canceling the bias of the ML estimator (Warm, 1989). The corresponding so-called weighted likelihood (WL) estimator is the value $\hat{\theta}_{\mathrm{WL}}$ of θ which satisfies

$$\frac{J(\theta)}{2I(\theta)} + \frac{\partial \log L(\theta)}{\partial \theta} = 0, \tag{14}$$

where

$$J(\theta) = \sum_{i=1}^{n} \frac{P_i'(\theta) P_i''(\theta)}{P_i(\theta) Q_i(\theta)}, \tag{15}$$

and $P_i''(\theta)$ is the second derivative of $P_i(\theta)$ with respect to θ (Warm, 1989, pp. 430–431). Moreover, the standard error of $\hat{\theta}_{\mathrm{WL}}$ can be estimated by

$$se(\hat{\theta}_{\mathrm{WL}}) = \frac{1}{\sqrt{\frac{I'(\hat{\theta}_{\mathrm{WL}}) J(\hat{\theta}_{\mathrm{WL}}) + I(\hat{\theta}_{\mathrm{WL}}) J'(\hat{\theta}_{\mathrm{WL}})}{2I(\hat{\theta}_{\mathrm{WL}})^2} + I(\hat{\theta}_{\mathrm{WL}})}}, \tag{16}$$

and $J'(\theta)$ is the first derivative of $J(\theta)$ with respect to θ.

It is direct to notice the similarities between the conditions (10) and (14) which define, respectively, the JM estimator and the WL estimator. Both methods are nevertheless very different conceptually. The JM estimator is a Bayesian method with a prior distribution based on the test information function, while the WL estimator aims at canceling the bias of the ML estimator with an appropriate weighted likelihood function.

An important assumption for our analysis is that both the JM and the WL estimator are unique and finite over the range of ability values. In other words, (10) and (14) are fulfilled for a single θ value each, and this value is not infinite. Similarly to the ML estimator, which can hold for several values under the 3PL when the number of items is small (Lord, 1980; Magis & Raîche, 2010; Samejima, 1973), this could also occur with these two estimators. However, the situation of multiple local maxima of the posterior or weighted likelihood function is rare in practice, and for sufficiently long tests it should not occur. This assumption is nevertheless fundamental for the following comparative analysis.

3. Relationships Between JM and WL Estimators

We derive now an interesting relationship between the JM and the WL estimates of ability under the 3PL model.

Set first

$$f_{\mathrm{JM}}(\theta) = \frac{I'(\theta)}{2I(\theta)} + \frac{\partial \log L(\theta)}{\partial \theta} \quad \text{and} \quad f_{\mathrm{WL}}(\theta) = \frac{J(\theta)}{2I(\theta)} + \frac{\partial \log L(\theta)}{\partial \theta}. \tag{17}$$

The function f_{JM} is the first derivative (with respect to θ) of the log-posterior distribution with Jeffreys prior, and setting $f_{\mathrm{JM}}(\theta) = 0$ is a simple rewriting of the condition (8). If follows that $f_{\mathrm{JM}}(\hat{\theta}_{\mathrm{JM}}) = 0$ and by the assumptions of uniqueness and finiteness of the estimator,

$$f_{\mathrm{JM}}(\theta) > 0 \quad \text{if } \theta < \hat{\theta}_{\mathrm{JM}} \quad \text{and} \quad f_{\mathrm{JM}}(\theta) < 0 \quad \text{if } \theta > \hat{\theta}_{\mathrm{JM}}. \tag{18}$$

Similarly and in the same spirit, $f_{\mathrm{WL}}(\hat{\theta}_{\mathrm{WL}}) = 0$ and

$$f_{\mathrm{WL}}(\theta) > 0 \quad \text{if } \theta < \hat{\theta}_{\mathrm{WL}} \quad \text{and} \quad f_{\mathrm{WL}}(\theta) < 0 \quad \text{if } \theta > \hat{\theta}_{\mathrm{WL}}. \tag{19}$$

Let us focus now on the difference between the two functions in (17):

$$f_{\mathrm{JM}}(\theta) - f_{\mathrm{WL}}(\theta) = \frac{I'(\theta) - J(\theta)}{2I(\theta)}. \tag{20}$$

This difference does not depend on the particular response pattern $(X_1, \ldots, X_n)$ of the examinee. Since the information function $I(\theta)$ defined by (7) is strictly positive for all ability levels, let us focus on the difference $I'(\theta) - J(\theta)$ only. The function $I'(\theta)$ can be written as

$$I'(\theta) = 2\sum_{i=1}^{n} \frac{P_i'(\theta)P_i''(\theta)}{P_i(\theta)Q_i(\theta)} - \sum_{i=1}^{n} \frac{P_i'(\theta)^3[Q_i(\theta) - P_i(\theta)]}{P_i(\theta)^2 Q_i(\theta)^2}, \tag{21}$$

by using (7). It comes then

$$I'(\theta) - J(\theta) = \sum_{i=1}^{n} \frac{P_i'(\theta)\{P_i(\theta)Q_i(\theta)P_i''(\theta) - P_i'(\theta)^2[Q_i(\theta) - P_i(\theta)]\}}{P_i(\theta)^2 Q_i(\theta)^2}. \tag{22}$$

Consider the following term in the right-hand side of (22):

$$h_i(\theta) = P_i(\theta)Q_i(\theta)P_i''(\theta) - P_i'(\theta)^2\big[Q_i(\theta) - P_i(\theta)\big], \tag{23}$$

and rewrite it under the 3PL model. To simplify the notations, we set $e_i = \exp[a_i(\theta - b_i)]$ so that (1) takes the simple form

$$P_i(\theta) = c_i + (1 - c_i)\frac{e_i}{1 + e_i} = \frac{c_i + e_i}{1 + e_i}, \tag{24}$$

and similarly,

$$Q_i(\theta) = \frac{1 - c_i}{1 + e_i}, \qquad P_i'(\theta) = \frac{a_i(1 - c_i)e_i}{(1 + e_i)^2} \quad \text{and} \quad P_i''(\theta) = \frac{a_i^2(1 - c_i)e_i(1 - e_i)}{(1 + e_i)^3}. \tag{25}$$

It follows that

$$h_i(\theta) = \frac{(1 - c_i)^2 a_i^2 c_i e_i}{(1 + e_i)^4} \tag{26}$$

which is strictly positive if $c_i > 0$ and equal to zero otherwise. It implies that $I'(\theta) > J(\theta)$ for any θ, according to (22), and hence that $f_{\mathrm{JM}}(\theta) > f_{\mathrm{WL}}(\theta)$ for any θ, according to (20).

The previous inequality is strict under the 3PL model, because at least one pseudo-guessing parameter c_i is strictly positive, and thus at least one of the functions $h_i(\theta)$ in (23) takes strictly positive values. If all c_i are equal to zero, as under the 2PL model, then it comes from (22) that the functions $I'(\theta)$ and $J(\theta)$ are completely identical. This was already pointed out by Warm (1989), and this yields back the well-known equivalence between JM and WL estimators in this context.

Finally, the estimates $\hat{\theta}_{\mathrm{WL}}$ and $\hat{\theta}_{\mathrm{JM}}$ are linked together as follows. First, recall that $f_{\mathrm{JM}}(\hat{\theta}_{\mathrm{JM}}) = 0$ by definition. Second, using the previous result, one gets $f_{\mathrm{JM}}(\hat{\theta}_{\mathrm{JM}}) > f_{\mathrm{WL}}(\hat{\theta}_{\mathrm{JM}})$. This implies $f_{\mathrm{WL}}(\hat{\theta}_{\mathrm{JM}}) < 0$ and using (19), one concludes that $\hat{\theta}_{\mathrm{JM}} > \hat{\theta}_{\mathrm{WL}}$. In other words, under the 3PL model and with the assumptions of uniqueness and finiteness, the JM estimator always returns larger values than the WL estimator for the same response pattern.

This result is interesting for several reasons. First, to our knowledge, such a relationship between two distinct estimators has never been established before. Second, the inequality fixes an overall trend between the two estimators under the 3PL model. Third, it is independent of the response pattern and the test length. However, longer tests should be preferred in order to ensure the uniqueness and the finiteness of the estimators, which are central assumptions for validating the developments above.

It is important to notice that the gap between $\hat{\theta}_{\text{WL}}$ and $\hat{\theta}_{\text{JM}}$ may not be necessarily very large. The previous relationship only provides a systematic trend between the two estimates, but the magnitude of their difference is not that easy to derive. We reserve this issue for follow-up research, where the empirical bias of the two methods will be compared. However, for very large ability levels, the item response curves under the 2PL and the 3PL models are nearly identical. This means that at this extreme of the ability scale, the estimates $\hat{\theta}_{\text{WL}}$ and $\hat{\theta}_{\text{JM}}$ can be assumed to be computed under the 2PL model, and thus yield identical estimates. Thus, both estimators should return very close estimates when the true ability level is very large. This gap between $\hat{\theta}_{\text{WL}}$ and $\hat{\theta}_{\text{JM}}$ can even be characterized more precisely as follows.

First, the difference $\Delta_f(\theta) = f_{\text{JM}}(\theta) - f_{\text{WL}}(\theta)$ can be written as follows:

$$\Delta_f(\theta) = \frac{\sum_{i=1}^n y_i(\theta)}{2\sum_{i=1}^n z_i(\theta)}, \tag{27}$$

by using (17), (22), (24) and (25), and with

$$y_i(\theta) = \frac{a_i^3 c_i (1-c_i) e_i^2}{(c_i+e_i)^2(1+e_i)^2} \quad \text{and} \quad z_i(\theta) = \frac{a_i^2(1-c_i)e_i^2}{(c_i+e_i)(1+e_i)^2}. \tag{28}$$

Furthermore, $y_i(\theta) \geq 0$ and $z_i(\theta) > 0$, and the ratio $y_i(\theta)/z_i(\theta)$ equals $a_i c_i/(c_i+e_i)$ and converges towards zero as θ increases infinitely. In sum, since

$$0 \leq \Delta_f(\theta) = \frac{\sum_{i=1}^n y_i(\theta)}{2\sum_{i=1}^n z_i(\theta)} \leq \frac{1}{2}\sum_{i=1}^n \frac{y_i(\theta)}{z_i(\theta)}, \tag{29}$$

one concludes that $\Delta_f(\theta)$ decreases towards zero as θ increases. This implies that at very large ability levels, the functions $f_{\text{JM}}(\theta)$ and $f_{\text{WL}}(\theta)$ are nearly identical, and thus also the JM and WL estimates. In other words, one expects the bias of the two estimators to be similar for large positive ability levels.

It is not easy to derive some similar trend for small abilities. Warm (1989) noticed that in this case, the bias of the WL estimator is positive, that is, the estimate is larger than the true level on average. Because of the systematic trend between JM and WL estimates, one can also predict that for very small ability levels the bias of the JM estimator will be positive and larger than that of the WL estimator.

4. Conclusions

This paper proposed the comparative study of two estimators of ability: the WL estimator and the JM estimator. The latter is the usual BM estimator with Jeffreys' prior distribution. Both methods are defined by closed form (10) and (14), and they are completely equivalent under the 2PL model, as stated previously in the literature. Under the 3PL model however, the JM estimator always returns larger values than the WL estimator, with the same test and response pattern. At very large positive ability levels the two estimators perform similarly, while at lower ability levels the JM estimator tends to be more positively biased.

Not only the precision of the estimators, but also their variability should be compared.

However, it is very difficult to obtain meaningful information by comparing directly the standard errors of the JM and WL estimators, i.e. (13) and (16) respectively. This topic should be further investigated.

Nevertheless, it is worth mentioning that a small simulation study was conducted in this regard. The design of the study was nearly the same as that used by Warm (1989) to generate the so-called conventional tests. It turned out that: (a) the WL and the JM estimators are globally equivalent in terms of bias and standard error for large, positive ability levels, as expected from the previous developments; (b) at small negative ability levels, the JM estimator is more positively biased than the WL estimator, but surprisingly, it is also less variable; and (c) with extremely small ability levels, the WL estimator tends to perform best.

The WL estimator was specifically developed to reduce, and even cancel, the bias of the ML estimator. It is therefore logical to observe that the JM estimator does not outperform the WL estimator in terms of bias, and the main benefit of Jeffreys' prior distribution consists probably in a decrease of the standard error. These differences in estimator performances tend to vanish with longer tests. The JM estimator, however, seems to be a convenient estimator for small tests when ability levels are not extremely low.

Acknowledgements

The authors wish to thank the associate editor and two anonymous reviewers for insightful advice, and Rémi Lambert (University of Liège, Belgium) for his fruitful suggestions. This research was supported by a grant "Chargé de recherches" of the National Funds for Scientific research (FNRS, Belgium), the Research Funds of the K.U. Leuven, and the Social Science and Humanities Research Council of Canada (SSHRC).

References

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*. Reading: Addison-Wesley (Chaps. 17–20).

Birnbaum, A. (1969). Statistical theory for logistic mental test models with a prior distribution of ability. *Journal of Mathematical Psychology*, *6*, 258–276.

Hoijtink, H., & Boomsma, A. (1995). On person parameter estimation in the dichotomous Rasch model. In G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models. Foundations, recent developments, and applications* (pp. 53–68). New York: Springer.

Jeffreys, H. (1939). *Theory of probability*. Oxford: Oxford University Press.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *186*, 453–461.

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

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. (1984). *Maximum likelihood and Bayesian parameter estimation in item response theory* (Research Report No. RR-84-30-ONR). Princeton, NJ: Educational Testing Service.

Magis, D., & Raîche, G. (2010). An iterative maximum a posteriori estimation of proficiency level to detect multiple local likelihood maxima. *Applied Psychological Measurement*, *34*, 75–90.

Meijer, R.R., & Nering, M.L. (1999). Computerized adaptive testing: Overview and introduction. *Applied Psychological Measurement*, *23*, 187–194.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.

Samejima, F. (1973). A comment on Birnbaum's three-parameter logistic model in the latent trait theory. *Psychometrika*, *38*, 221–223.

Warm, T.A. (1989). Weighted likelihood estimation of ability in item response models. *Psychometrika*, *54*, 427–450.

Manuscript Received: 12 AUG 2010
Final Version Received: 24 APR 2011
Published Online Date: 1 NOV 2011