

Marginalized Maximum Likelihood Estimation for the IPL-AG IRT Model

Applied Psychological Measurement
2015, Vol. 39(6) 448–464
© The Author(s) 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146621615574694
apm.sagepub.com



Ryongsun Park¹, Keenan A. Pituch¹, Jiseon Kim², Barbara G. Dodd¹,
and Hyewon Chung³

Abstract

Marginal maximum likelihood estimation based on the expectation–maximization algorithm (MML/EM) is developed for the one-parameter logistic model with ability-based guessing (IPL-AG) item response theory (IRT) model. The use of the MML/EM estimator is cross-validated with estimates from NLMIXED procedure (PROC NLMIXED) in Statistical Analysis System. Numerical data are provided for comparisons of results from MML/EM and PROC NLMIXED.

Keywords

MML, EM, IPL-AG, estimator, IRT

Introduction

Psychological constructs are hypothetical concepts that cannot be observed directly, but theorized to explain human behavior. Examples include intelligence, motivation, self-esteem, mathematics proficiency, and happiness. Although these constructs cannot be observed directly, their existence can be inferred through various behaviors. For example, mathematics proficiency can be inferred by observing a person's responses to an instrument (i.e., a mathematics test). By placing an individual on this latent continuum, measuring the construct is formally established. As one of the tools for measuring constructs, item response theory (IRT) models how the trait level is related with an individual's response to an item. In applying this to education, trait levels are realized as latent variables (theta, trait continuum, or scale), which represent the individual's ability within a specific knowledge domain. Furthermore, one or more item parameters characterize an individual item. Assuming success or failure on an item follows independent Bernoulli

¹University of Texas at Austin, USA

²University of Washington, Seattle, USA

³Chungnam National University, Daejeon, South Korea

Corresponding Authors:

Ryongsun Park, University of Texas at Austin, 1 University Station, D5800, Austin, TX 78712-0383, USA.
Email: rosepark222@gmail.com

Hyewon Chung, Chungnam National University, 99 Daehakro, Yuseong-gu, Daejeon, Korea.
Email: hyewonchung7@gmail.com

distribution, the Rasch model (Lord & Novick, 1968) relates the probability of pass/fail on item with the difference between person ability and item difficulty through the logistic link function.

In terms of IRT modeling, multiple-choice (MC) items impose an interesting problem that cannot be handled with the conventional Rasch model: *guessing*. There are diverse parameterizations and interpretations used to accommodate the guessing behavior. As a result, various IRT models have been proposed to explain guessing behavior within the model: (a) a fixed value of $1/L$, with L being the number of options in MC items; (b) an average guessing parameter across items in a MC testing; (c) a guessing parameter specific to each item, as in the three-parameter logistic (3PL) model (Birnbaum, 1968); and (d) models for guessing that are dependent on person ability (Hutchinson, 1991). The fixed value of $1/L$ guessing parameter reflects the concept of random guessing and studies have shown that the real guessing probability may be larger or smaller than the $1/L$ depending on the attractiveness of incorrect options (Hambleton, Swaminathan, & Rogers, 1991). However, the 3PL was developed to explain the larger-than-zero probability of guessing on MC items for examinees of low abilities. Under 3PL, the parameter reflects item-dependent pseudo-guessing. A high value indicates a higher probability of passing an item for individuals with low ability, which may be due to the lack of attractive distractors. While the traditional 3PL explains pseudo-guessing behavior for low-ability examinees, the one-parameter logistic model with ability-based guessing (1PL-AG), has been introduced to describe guessing primarily for average-ability examinees (San Martín, del Pino, & De Boeck, 2006).

1PL-AG

One of the interpretations of the 3PL model is that item responses are composed of two processes: a p-process and a g-process (Hutchinson, 1991). The p-process is an item-solving process, whereas the g-process is a guessing process. One possible arrangement of executing the two processes is that the g-process is followed by the p-process: An examinee attempts to solve an item and resorts to guessing only if a correct response is not identified. While applying this process model, 1PL-AG was motivated from the observation that the success of guessing depends on ability. For instance, an examinee with higher ability may have a higher chance of success on the g-process, perhaps, due to his or her greater ability to recall relevant facts and connect separate knowledge to make a correct choice. According to the 1PL-AG, the probability of passing an item j given ability θ_i and item parameters is

$$P_j(\theta_i) = \frac{1}{1 + \exp(-(\theta_i - \beta_j))} + \left(1 - \frac{1}{1 + \exp(-(\theta_i - \beta_j))}\right) \frac{1}{1 + \exp(-(\alpha\theta_i - \gamma_j))}, \quad (1)$$

where $1/(1 + \exp(-(\theta_i - \beta_j)))$ is the p-process and $1/(1 + \exp(-(\alpha\theta_i - \gamma_j)))$ is the g-process. For the p-process, the probability of a correct response to the j th item is identical to the one-parameter IRT model and depends on the difference between the item difficulty, β_j and examinee's ability, θ_i . For the g-process, the probability of success for the j th item depends on ability (θ_i), an average weight on ability (α), and the guessing probability on the logistic scale (γ) for an examinee with average ability (i.e., $\theta_i = 0$).

As α is a positive value, the probability of passing an item due to the g-process increases as the ability increases. However, as the g-process is weighted by the probability of the failed p-process (i.e., $1 - \text{p-process}$), which approaches zero as ability increases, the net contribution of the g-process decreases at high ability. In addition, as the g-process approaches zero as ability decreases, the net contribution of guessing decreases at low ability. According to San Martín et al. (2006), the interpretation for the low contribution from the g-process at both extremes is that

low-ability examinees are attracted by the distractors and not able to find the correct answer, whereas high-ability examinees do not necessarily resort to guessing to choose the correct response.

Recently, researchers have raised model identification issue related with 1PL-AG (Maris & Bechger, 2009; San Martín, Rolin, & Castro, 2013). Model identifiability states that there is a one-to-one relationship between the parameters and the sampling distributions. Under the sampling theory framework, it is not possible to obtain unbiased and/or consistent estimators of parameters if identifiability does not hold (Gabrielsen, 1978; Koopmans & Reiersol, 1950). Maris and Bechger (2009) have shown that 3PL is not identifiable if all discrimination is held to a constant value (i.e., called 1PL-G). For example, it can be shown in 1PL-G that parameter values of (θ , difficulty, guessing) of both (2.0, 1.2, 0.2) and (2.0134428, 1.1694177, 0.1751562) result in same probability of passing an item, 0.7519796. Then, it cannot be determined which set of parameter should be used for the correct inference. This identification issue is relevant to the current study because 1PL-G is nested within 1PL-AG by setting α to zero.

The model identifiability problem, however, should be considered under the context of the sampling distribution and the parameter of interest (San Martín et al., 2013). For instance, the identification problem differs significantly depending on whether the abilities are specified as fixed effects versus random effects. For fixed-effects specification, the parameters of interest during the estimation include θ , whereas the random-effects specification marginalizes the likelihood function and θ is integrated out using a known (or partially known) ability distribution. It is known that 1PL-G could be identified by fixing one guessing parameter to a known value (e.g., zero for convenience) under the random-effects specification and that the ability distribution is known up to the scale parameter (San Martín et al., 2013). Whether this issue imposes any practical limitation for 3PL or 1PL-AG model is still an open question. Furthermore, question on whether the identification problem still remains when the ability distribution is fully specified (i.e., location and scale are assumed to be known) has not answered yet.

In terms of parameter estimation, PROC NLMIXED has been the primary estimation tool because of programming ease. However, the 1PL-AG model has been vastly underused due to, we believe, the lengthy estimation time. Therefore, the current study develops and evaluates an efficient estimator based on the marginal maximum likelihood estimation with the expectation-maximization algorithm (MML/EM) for the 1PL-AG model (Bock & Aitkin, 1981). The rest of the article is organized as follows. Estimation methodologies under the IRT framework are introduced and the MML/EM for the 1PL-AG model is derived. A simulation study and its results are presented, followed by the discussion.

Method

MML/EM Estimation Algorithm for the 1PL-AG Model

The conditional probability for an examinee i to respond in pattern $Y_i = [y_{i1}, y_{i2}, \dots, y_{iJ}]$, conditional on ability θ_i is $P(Y_i|\theta_i)$. For an examinee randomly sampled from a population with ability distribution $g(\theta)$, the unconditional probability of Y_i is

$$P(Y_i) = \int P(Y_i|\theta)g(\theta)d\theta. \quad (2)$$

Under the assumption of independence of observations, the logarithmic transformation of the probability of response patterns for I randomly chosen examinees becomes

$$\log L = \sum_i^I \log P(Y_i). \quad (3)$$

The derivation of the marginal likelihood equations begins with the first derivatives with respect to a parameter ζ :

$$\begin{aligned} \frac{\partial}{\partial \zeta} \log L &= \sum_i^I \frac{1}{P(Y_i)} \int \left[\frac{\partial}{\partial \zeta} P(Y_i|\theta) \right] g(\theta) d\theta \\ &= \sum_i^I \frac{1}{P(Y_i)} \int \left[\frac{\partial}{\partial \zeta} \log P(Y_i|\theta) \right] P(Y_i|\theta) g(\theta) d\theta \\ &= \sum_i^I \int \left[\frac{\partial}{\partial \zeta} \log P(Y_i|\theta) \right] \left[\frac{P(Y_i|\theta) g(\theta)}{P(Y_i)} \right] d\theta \\ &= \sum_i^I \int \left[\frac{\partial}{\partial \zeta} \log P(Y_i|\theta) \right] P(\theta|Y_i) d\theta. \end{aligned} \quad (4)$$

Furthermore, the first derivative of the conditional likelihood function is expressed as

$$\begin{aligned} \frac{\partial}{\partial \zeta} \log P(Y_i|\theta) &= \frac{\partial}{\partial \zeta} \log \left[\prod_j^J P_j(\theta)^{y_{ij}} Q_j(\theta)^{1-y_{ij}} \right] \\ &= \sum_j^J \frac{y_{ij} - P_j(\theta)}{P_j(\theta) Q_j(\theta)} \frac{\partial P_j(\theta)}{\partial \zeta}, \end{aligned} \quad (5)$$

where y_{ij} is the dichotomous response to item j of examinee i ; $P_j(\theta)$ is the probability of being correct on item j for an examinee of ability θ ; and $Q_j(\theta)$ is $1 - P_j(\theta)$.

For each of 1PL-AG parameters, α , β_j , and γ_j , the first derivatives of the conditional likelihood functions are

$$\frac{\partial}{\partial \alpha} \log P(Y_i|\theta) = \sum_j^J \left[\frac{y_{ij} - P_j(\theta)}{P_j(\theta) Q_j(\theta)} \left[\left[\frac{\theta}{1 + \exp(\theta - \beta_j)} \right] \frac{\exp(\alpha\theta + \gamma_j)}{(1 + \exp(\alpha\theta + \gamma_j))^2} \right] \right], \quad (6)$$

$$\frac{\partial}{\partial \beta_j} \log P(Y_i|\theta) = \frac{y_{ij} - P_j(\theta)}{P_j(\theta) Q_j(\theta)} \left[\frac{\exp(\theta - \beta_j)}{(1 + \exp(\theta - \beta_j))^2} \right] \frac{-1}{1 + \exp(\alpha\theta + \gamma_j)}, \quad (7)$$

and

$$\frac{\partial}{\partial \gamma_j} \log P(Y_i|\theta) = \frac{y_{ij} - P_j(\theta)}{P_j(\theta) Q_j(\theta)} \left[\frac{1}{1 + \exp(\theta - \beta_j)} \right] \frac{\exp(\alpha\theta + \gamma_j)}{(1 + \exp(\alpha\theta + \gamma_j))^2}. \quad (8)$$

Note that except for the parameter α , which is common across items, the first derivatives of β_j and γ_j do not depend on β_h , and γ_h of $j \neq h$. Plugging Equations 6, 7, and 8 into 4, the marginal likelihood equations for α , β_j , and γ_j are

$$\frac{\partial}{\partial \alpha} \log L = \sum_i^I \int \left[\sum_j^J \left[\frac{y_{ij} - P_j(\theta)}{P_j(\theta) Q_j(\theta)} \left[\left[\frac{\theta}{1 + \exp(\theta - \beta_j)} \right] \frac{\exp(\alpha\theta + \gamma_j)}{(1 + \exp(\alpha\theta + \gamma_j))^2} \right] \right] \right] P(\theta|Y_i) d\theta, \quad (9)$$

$$\frac{\partial}{\partial \beta_j} \log L = \sum_i^I \int \left[\frac{y_{ij} - P_j(\theta)}{P_j(\theta)Q_j(\theta)} \left[\left[\frac{\exp(\theta - \beta_j)}{(1 + \exp(\theta - \beta_j))^2} \right] \frac{-1}{1 + \exp(\alpha\theta + \gamma_j)} \right] \right] P(\theta|Y_i) d\theta, \quad (10)$$

and

$$\frac{\partial}{\partial \beta_j} \log L = \sum_i^I \int \left[\frac{y_{ij} - P_j(\theta)}{P_j(\theta)Q_j(\theta)} \left[\left[\frac{1}{1 + \exp(\theta - \beta_j)} \right] \frac{\exp(\alpha\theta + \gamma_j)}{(1 + \exp(\alpha\theta + \gamma_j))^2} \right] \right] P(\theta|Y_i) d\theta. \quad (11)$$

Equations 9, 10, and 11 represent a nonlinear system of equations that need to be solved for their roots. The integral within the equations can be accomplished through numerical integration, which is the approximation of integrating a continuous curve through the summation of a rectangular area. The integrands, the functions under integration, are evaluated at a finite set of points called "integration points" (or quadrature points). The prior ability distribution, $g(\theta)$, is approximated by taking a finite number of quadrature points, X_k for $k = (1, 2, \dots, q)$ along the θ scale and its corresponding weights, $A(X_k)$. The posterior distribution of θ , Equation 4, is obtained at each X_k and its weights $A(X_k)$:

$$P(X_k|Y_i) = \frac{\prod_j^J P_j(X_k)^{y_{ij}} Q_j(X_k)^{1-y_{ij}} A(X_k)}{\sum_k^q \prod_j^J P_j(X_k)^{y_{ij}} Q_j(X_k)^{1-y_{ij}} A(X_k)}. \quad (12)$$

This equation limits the continuous distribution of the posterior ability of examinee i , θ_i , into a finite number of quadrature points. Assuming a unit normal distribution of θ , quadrature points and their weights can be obtained from Gaussian quadrature formula (Stroud & Secrest, 1966). In addition, values of $A(X_k)$ could be adjusted at each iteration by empirically estimated weights (Mislevy, 1984). As suggested from Equation 12, however, the impact of the prior ability distribution is insignificant with a practically large number of items (Bock & Aitkin, 1981). Intuitively, the ability distribution of a sample conditional on the observation of responses to items depends mainly on the item responses and their characteristics rather than on the population ability distribution, if responses are drawn from enough items. The marginal likelihood equations using Equation 12 are

$$\frac{\partial}{\partial \alpha} \log L = \sum_i^I \sum_k^q \left\{ \sum_j^J \left[\frac{y_{ij} - P_j(X_k)}{P_j(X_k)Q_j(X_k)} \left[\left[\frac{X_k}{1 + \exp(X_k - \beta_j)} \right] \frac{\exp(\alpha X_k + \gamma_j)}{(1 + \exp(\alpha X_k + \gamma_j))^2} \right] \right] P(X_k|Y_i) \right\}, \quad (13)$$

$$\frac{\partial}{\partial \beta_j} \log L = \sum_i^I \sum_k^q \left\{ \frac{y_{ij} - P_j(X_k)}{P_j(X_k)Q_j(X_k)} \left[\left[\frac{\exp(X_k - \beta_j)}{(1 + \exp(X_k - \beta_j))^2} \right] \frac{-1}{1 + \exp(\alpha X_k + \gamma_j)} \right] P(X_k|Y_i) \right\}, \quad (14)$$

and

$$\frac{\partial}{\partial \gamma_j} \log L = \sum_i^I \sum_k^q \left\{ \frac{y_{ij} - P_j(X_k)}{P_j(X_k)Q_j(X_k)} \left[\left[\frac{1}{1 + \exp(X_k - \beta_j)} \right] \frac{\exp(\alpha X_k + \gamma_j)}{(1 + \exp(\alpha X_k + \gamma_j))^2} \right] P(X_k|Y_i) \right\}. \quad (15)$$

Equations 13, 14, and 15 are marginal likelihood functions of the parameters of the *entire* set of J items through the posterior ability distribution, $P(X_k|Y_i)$. Therefore, $(2J + 1)$ marginal likelihood equations should be solved simultaneously. However, solving entire equations imposes an intractable computational burden of generating and inverting a $(2J + 1) \times (2J + 1)$ Jacobian matrix for Equations 13, 14, and 15.

Bock and Aitkin (1981) applied the EM algorithm to transform a regular MML estimation to a missing data problem. For that, two artificial quantities are introduced:

$$\bar{n}_k = \sum_i^I P(X_k|Y_i) = \sum_i^I \left[\frac{\prod_j^J P_j(X_k)^{y_{ij}} Q_j(X_k)^{1-y_{ij}} A(X_k)}{\sum_k^q \prod_j^J P_j(X_k)^{y_{ij}} Q_j(X_k)^{1-y_{ij}} A(X_k)} \right], \quad (16)$$

and

$$\bar{r}_{jk} = \sum_i^I y_{ij} P(X_k|Y_i) = \sum_i^I \left[\frac{y_{ij} \prod_j^J P_j(X_k)^{y_{ij}} Q_j(X_k)^{1-y_{ij}} A(X_k)}{\sum_k^q \prod_j^J P_j(X_k)^{y_{ij}} Q_j(X_k)^{1-y_{ij}} A(X_k)} \right], \quad (17)$$

where $\bar{n}_k$ is the expected number of examinees at ability level X_k , and $\bar{r}_{jk}$ is the expected number of correct responses at ability level X_k for item j . Using these quantities, the marginal likelihood equations become

$$\frac{\partial}{\partial \alpha} \log L = \sum_k^q \sum_j^J \left[\frac{\bar{r}_{jk} - \bar{n}_k P_j(X_k)}{P_j(X_k) Q_j(X_k)} \left[\left[\frac{X_k}{1 + \exp(X_k - \beta_j)} \right] \frac{\exp(\alpha X_k + \gamma_j)}{(1 + \exp(\alpha X_k + \gamma_j))^2} \right] \right], \quad (18)$$

$$\frac{\partial}{\partial \beta_j} \log L = \sum_k^q \left[\frac{\bar{r}_{jk} - \bar{n}_k P_j(X_k)}{P_j(X_k) Q_j(X_k)} \left[\left[\frac{\exp(X_k - \beta_j)}{(1 + \exp(X_k - \beta_j))^2} \right] \frac{-1}{1 + \exp(\alpha X_k + \gamma_j)} \right] \right], \quad (19)$$

and

$$\frac{\partial}{\partial \gamma_j} \log L = \sum_k^q \left[\frac{\bar{r}_{jk} - \bar{n}_k P_j(X_k)}{P_j(X_k) Q_j(X_k)} \left[\left[\frac{1}{1 + \exp(X_k - \beta_j)} \right] \frac{\exp(\alpha X_k + \gamma_j)}{(1 + \exp(\alpha X_k + \gamma_j))^2} \right] \right]. \quad (20)$$

To summarize, the complete steps for MML/EM algorithm are as follows:

1. *E-step*: Given the response strings, the provisional initial parameters, and the quadrature weight $A(X_k)$, compute the expected number of examinees on each X_k , $\bar{n}_k$, and compute the expected correct responses for item j on each X_k , $\bar{r}_{jk}$.
2. *M-step*: For each item, find the roots for Equations 18, 19, and 20 via a numerical method such as the Newton–Raphson method.

The MML/EM algorithm is an iterative process such that the E-step and M-step are repeated until item parameters satisfy the predefined convergence criterion. A careful examination should reveal that Equations 19 and 20 for an item j do not depend on the parameters of an item h , where $j \neq h$, because $\bar{n}_k$ and $\bar{r}_{jk}$ summarize the necessary information from the item parameter estimates. Therefore, Equations 19 and 20 could be solved for individual items. However,

Equation 18 depends on the parameters of the entire set of items through summation across items, due to the parameter α being common for all items. Thissen (1982) suggested a simpler relaxation solution that could be applied for this estimation: Solve Equations 19 and 20 for each item and then Equation 18 will be solved. The application of the EM algorithm transforms multiple cycles of the $(2J+1)$ -dimensional parameter estimation into a J two-dimensional and a single univariate problem. Because each individual cycle can be solved quickly, the MML/EM algorithm is known to have a tremendous advantage in terms of estimation time. MML/EM procedure eliminates the need for the second derivative matrix because the equations of the first derivative (i.e., the marginal likelihood equations) are driven explicitly. The marginalization of the ability parameter is common for MML (Bock & Lieberman, 1970), MML/EM (Bock & Aitkin, 1981), and PROC NLMIXED. However, MML/EM is unique in that the ability parameter, or its transformed variable, is *estimated* in every iteration.

Simulation Study

The main goals of this simulation study are twofold. First, the recovery of 1PL-AG model parameters is examined using the MML/EM method presented previously. For this, the following steps are taken: (a) Implement MML/EM for the 1PL-AG in Java language; (b) generate response strings of three groups of examinees (i.e., 2,500, 3,000, and 3,500) to 20 1PL-AG items of $\alpha = (0.065, 0.265, 0.365)$ and β s and γ s from true values in Table 1. β s are difficulty parameter and γ s are guessing at $\theta = 0$ in logistic scale. Mean, standard deviation, minimum and maximum for β s and γ s are (0.641, 1.021, -1.3697, 2.097) and (-1.223, 0.644, -2.502, -0.23), respectively, and 20 values are randomly sampled within the ranges. The generated true parameter values in the current study are guided by the work from San Martín et al. (2006). (c) Perform parameter estimations for 3 (α s) $\times$ 3 (sample sizes) conditions. (d) Repeat Steps (b) and (c) 50 times and examine the performance of parameter recoveries in terms of their root mean square errors (RMSEs) and standard errors. The differences between true and estimated parameters are squared and averaged over 50 replications. Then squared root is taken for each β s, α , and γ s. Empirical standard errors for each parameter estimates were derived based on 50 replications as well.

Second, the performance comparison between MML/EM and PROC NLMIXED is made in terms of biases and estimation times. For this, 50 replications of parameter recoveries are performed with PROC NLMIXED for the same 3 $\times$ 3 conditions. The same number of Gauss-Hermite quadrature points (i.e., 10 points) and the same convergence criterion (i.e., The difference of $-2 \times \text{Log-likelihood}$ ($-2 \times l$) values between two successive EM iterations being smaller than $10E-3$) are applied for both MML/EM and PROC NLMIXED. The results from both the MML/EM and PROC NLMIXED procedures are compared based on estimation accuracy and simulation time. RMSEs are used for evaluating estimation accuracy for each items. For example, RMSE for β_j is

$$\text{RMSE}_{\beta_j} = \sqrt{\left[\frac{\sum_r^R (\beta_{jr} - \hat{\beta}_j)^2}{R} \right]}, \quad (21)$$

where R is the number of replications (i.e., 50), β_{jr} is the estimated parameter for item j in r th replication, and $\hat{\beta}_j$ is the true parameter value. Empirical standard errors were estimated across 50 replications. In addition, estimated standard errors were calculated and the ratio of average

Table 1. Recovered Estimates From MML/EM.

Parameters	True values	$\alpha = .065$			$\alpha = .265$			$\alpha = .365$		
		2,500	3,000	3,500	2,500	3,000	3,500	2,500	3,000	3,500
β_1	0.903	0.817	0.71	0.792	1.281	0.751	0.958	1.169	0.831	0.759
β_2	0.509	0.511	0.638	0.612	0.53	0.389	0.261	0.504	0.32	0.417
β_3	-1.370	-0.719	-1.202	-1.037	-1.437	-0.701	-1.576	-1.336	-1.507	-1.598
β_4	1.881	2.393	1.736	1.82	1.799	2.198	1.818	1.714	1.656	1.65
β_5	1.315	1.171	1.719	1.653	1.106	0.897	1.466	1.324	1.363	1.51
β_6	-0.351	-0.651	-0.301	-0.153	-0.234	0.044	-0.511	-0.111	-0.647	-0.193
β_7	0.418	0.307	0.221	0.329	0.133	0.321	0.155	0.49	0.393	0.871
β_8	0.738	0.425	0.745	0.691	0.958	0.596	0.917	0.763	0.723	0.955
β_9	0.329	0.446	0.483	0.418	0.538	0.501	0.366	0.354	0.62	0.143
β_{10}	-1.297	-1.68	-0.877	-0.932	-0.94	-1.411	-0.767	-1.075	-1.216	-1.003
β_{11}	1.534	1.574	1.623	1.515	1.341	1.104	1.498	2.04	1.619	1.687
β_{12}	1.046	0.819	1.107	0.977	1.488	1.172	1.209	1.379	1.032	0.989
β_{13}	-0.122	-0.607	-0.1	-0.107	-0.065	-0.148	0.165	-0.443	-0.152	0.003
β_{14}	0.399	0.535	0.431	0.412	0.037	0.478	0.493	0.767	0.085	0.395
β_{15}	-0.797	-0.704	-1.123	-1.256	-0.947	-0.807	-0.741	-0.053	-0.805	-0.569
β_{16}	1.797	1.899	2.169	2.047	1.777	1.861	1.97	1.765	2.056	1.961
β_{17}	2.097	2.202	2.005	2.004	2.302	2.529	2.102	2.307	1.939	2.163
β_{18}	0.797	0.689	0.922	0.916	0.975	0.83	0.59	0.429	0.804	0.618
β_{19}	1.197	1.146	1.472	1.485	1.245	1.425	1.368	0.769	1.773	1.51
β_{20}	1.797	1.987	1.879	1.79	1.972	1.948	1.981	1.864	1.671	2.005
α	.07	.07	.065	.065	.264	.264	.265	.367	.366	.359
γ_1	-0.494	-0.435	-0.717	-0.682	-0.182	-0.644	-0.421	-0.293	-0.597	-0.586
γ_2	-1.948	-1.806	-1.768	-1.935	-1.756	-2.145	-2.683	-2.047	-2.983	-2.27
γ_3	-1.248	-0.082	-0.739	-0.488	-1.324	0.013	-2.439	-0.896	-2.076	-2.324

(continued)

Table I. (continued)

Parameters	True values	$\alpha = .065$			$\alpha = .265$			$\alpha = .365$		
		2,500	3,000	3,500	2,500	3,000	3,500	2,500	3,000	3,500
γ_4	-1.161	-0.917	-1.415	-1.353	-1.26	-1.038	-1.185	-1.257	-1.317	-1.288
γ_5	-0.785	-0.867	-0.556	-0.589	-0.971	-1.099	-0.683	-0.732	-0.759	-0.63
γ_6	-0.366	-0.705	-0.243	-0.084	-0.23	-0.037	-0.638	-0.066	-0.842	-0.229
γ_7	-1.517	-1.978	-1.887	-1.657	-2.072	-1.76	-1.988	-1.343	-1.657	-0.907
γ_8	-0.679	-0.993	-0.717	-0.775	-0.374	-0.881	-0.507	-0.614	-0.757	-0.552
γ_9	-2.502	-2.344	-2.13	-2.433	-1.905	-1.864	-2.433	-2.342	-1.669	-4.146
γ_{10}	-0.882	-4.183	-0.383	-0.476	-0.146	-1.405	0.076	-0.521	-0.606	-0.326
γ_{11}	-1.507	-1.47	-1.476	-1.516	-1.769	-2.252	-1.581	-1.121	-1.379	-1.384
γ_{12}	-0.946	-1.291	-0.92	-1.069	-0.613	-0.828	-0.758	-0.711	-0.946	-0.949
γ_{13}	-0.230	-0.997	-0.213	-0.229	-0.094	-0.268	0.063	-0.548	-0.267	-0.101
γ_{14}	-1.606	-1.279	-1.486	-1.563	-2.615	-1.529	-1.375	-0.954	-4.446	-1.672
γ_{15}	-0.682	-0.606	-1.324	-1.838	-0.776	-0.957	-0.657	0.116	-0.819	-0.389
γ_{16}	-0.982	-0.835	-0.876	-0.941	-0.969	-1.022	-0.795	-0.977	-0.894	-0.905
γ_{17}	-1.882	-1.74	-1.901	-1.867	-1.615	-1.661	-1.828	-1.653	-2.088	-1.712
γ_{18}	-1.882	-2.155	-1.702	-1.703	-1.488	-1.85	-2.441	-4.097	-1.858	-2.499
γ_{19}	-0.882	-0.959	-0.623	-0.636	-0.826	-0.746	-0.787	-1.387	-0.536	-0.737
γ_{20}	-2.282	-2.009	-2.247	-2.372	-2.111	-2.082	-2.121	-2.177	-2.438	-2.15

Note. MML = marginal maximum likelihood; EM = expectation-maximization algorithm.

estimated standard errors to empirical standard errors for each parameters are calculated. In addition, $-2 \times \text{Log-likelihood}$ values are calculated as fit statistics for each estimation method and simulation condition.

The simulation time for MML/EM was measured after the initial values for convergence of the Newton–Raphson method were obtained. SAS version 9.3 was used, and the simulations were performed using a computer with the following specifications: AMD Phenom II X4 CPU, 3.4 GHz, 512 KB Cache, and 6 GB of RAM.

Results

Tables 1 and 2 present parameter recoveries for individual items (in the order of β s, α , and γ s) in each condition.

Tables 3 and 4 present of RMSEs for individual items (in the order of β s, α , and γ s) in each condition.

Table 5 shows empirical standard errors. In addition, standard errors from MML/EM and PROC NLMIXED were analytically estimated, and the ratios of the average estimated standard errors to the empirical standard errors were presented within the parentheses. The values close to ones indicate that estimated standard errors are accurate to the empirical standard errors. MML/EM resulted in smaller standard errors for α s, whereas PROC NLMIXED produced smaller standard errors for β s. Standard errors for γ s were comparable for both methods. As shown in Table 5, the standard errors decreased as the sample sizes increased. Large standard error values for γ s imply large variations for estimates of guessing parameter.

Table 6 presents simulation time and a model fit statistic (i.e., $-2 \times l$) for each of the estimation methods. For PROC NLMIXED, the simulation took from 1.85 to 3.72 hr, whereas MML/EM took from 15.5 to 50 s. There is a clear pattern of increased estimation time as sample size increases. PROC NLMIXED and MML/EM methods both showed the largest simulation time for $\alpha = .265$. In terms of the $-2 \times l$, the differences between the MML/EM and PROC NLMIXED were small, demonstrating that both procedures produced similar values in terms of their likelihoods.

Discussion

The advancement of the IRT model has been intertwined with the development of parameter estimation techniques. Recent attention on the IRT model as examples of the NLMIXED model expanded the understanding and interpretation of parameters in IRT models. The relationship between two somewhat distinct models already was predicted by Bock and Lieberman (1970), when they regarded abilities as random effects. Methods of NLMIXED and MML/EM take a common approach: the maximization of the marginalized likelihood function. While widely available NLMIXED estimators perform direct optimization after marginalization (Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003), MML/EM incorporates estimations of the distributions of ability parameters. Through this augmentation of person parameter information for item parameter estimation, MML/EM achieves a faster convergence.

Most of the research on the IPL-AG model has been based on the use of PROC NLMIXED for the parameter estimation. Although it provides ease of programming, lengthy estimation time associated with PROC NLMIXED encouraged an alternative estimation procedure. The main contribution of the current study, we believe, is the development of a more efficient estimation procedure for the IPL-AG model. For the demonstration of the MML/EM method, a simulation study was conducted and compared with the current standard estimation procedure. The results indicated that MML/EM method produced estimates that were comparable with the

Table 2. Recovered Estimates From PROC NLMIXED.

Parameters	True values	$\alpha = .065$			$\alpha = .265$			$\alpha = .365$		
		2,500	3,000	3,500	2,500	3,000	3,500	2,500	3,000	3,500
β_1	0.903	0.948	0.959	0.772	0.994	0.758	0.453	1.178	0.718	0.939
β_2	0.509	0.573	0.778	0.448	0.428	0.847	0.222	0.962	0.237	0.343
β_3	-1.370	-1.18	-1.662	-1.491	-1.641	-1.682	-0.983	-1.196	-1.355	-1.359
β_4	1.881	2.072	1.81	2.008	1.758	2.079	1.739	2.311	2.027	1.695
β_5	1.315	1.087	1.214	1.429	1.516	1.248	1.083	1.86	1.308	1.511
β_6	-0.351	-0.45	-0.352	-0.544	-1.078	-0.128	-0.42	-0.359	-0.608	-0.731
β_7	0.418	0.396	0.479	0.262	0.409	0.715	0.601	1.025	0.355	0.401
β_8	0.738	0.813	0.614	0.877	0.514	1.092	0.921	0.417	0.827	0.676
β_9	0.329	0.22	0.345	0.266	0.589	0.539	0.223	0.376	0.204	0.716
β_{10}	-1.297	-1.436	-0.994	-1.151	-0.898	-1.688	-0.927	-0.799	-0.956	-1.412
β_{11}	1.534	1.599	1.613	1.339	1.826	1.611	1.651	1.428	1.79	1.711
β_{12}	1.046	1.295	1.241	0.881	0.847	1.115	0.934	1.373	1.14	1.106
β_{13}	-0.122	-0.223	0.182	-0.309	-0.462	-0.347	-0.058	0.062	0.105	-0.254
β_{14}	0.399	0.491	0.519	0.489	0.666	0.457	0.474	0.311	0.511	0.21
β_{15}	-0.797	-1.252	-1.003	-0.627	-0.489	-0.573	-1.209	-0.334	-1.169	-1.006
β_{16}	1.797	1.644	1.978	2.023	1.954	1.339	1.334	2.075	1.099	2.039
β_{17}	2.097	1.944	2.329	2.441	2.025	2.187	2.036	2.516	1.818	1.989
β_{18}	0.797	0.573	0.874	1.009	0.616	0.954	0.809	0.977	0.902	0.918
β_{19}	1.197	1.341	1.211	1.092	1.285	1.34	1.061	1.423	1.281	1.63
β_{20}	1.797	1.727	1.914	1.742	2.21	1.703	1.956	1.832	1.829	1.714
α		.063	.058	.067	.267	.266	.268	.379	.358	.367
γ_1	-0.494	-0.494	-0.539	-0.533	-0.426	-0.613	-0.804	-0.273	-0.573	-0.515
γ_2	-1.948	-1.837	-1.468	-2.011	-2.03	-1.31	-3.851	-1.191	-2.869	-2.746
γ_3	-1.248	-1.018	-2.667	-2.285	-2.402	-2.767	-3.342	-0.835	-0.923	-1.402
γ_4	-1.161	-1.019	-1.24	-1.083	-1.221	-1.134	-1.297	-1.04	-1.054	-1.291

(continued)

Table 2. (continued)

Parameters	True values	$\alpha = .065$			$\alpha = .265$			$\alpha = .365$		
		2,500	3,000	3,500	2,500	3,000	3,500	2,500	3,000	3,500
γ_5	-0.785	-0.918	-0.82	-0.679	-0.766	-0.916	-0.883	-0.573	-0.793	-0.676
γ_6	-0.366	-0.651	-0.288	-0.603	-2.294	-0.131	-0.366	-0.338	-0.664	-1.055
γ_7	-1.517	-1.69	-1.317	-1.816	-1.708	-1.062	-1.267	-0.75	-1.602	-1.623
γ_8	-0.679	-0.685	-0.806	-0.596	-0.94	-0.46	-0.579	-1.099	-0.619	-0.821
γ_9	-2.502	-2.968	-2.458	-3.015	-1.857	-1.858	-3.749	-2.254	-2.772	-1.497
γ_{10}	-0.882	-1.43	-0.52	-0.677	-0.277	-2.366	-0.349	-0.077	-0.207	-1.016
γ_{11}	-1.507	-1.355	-1.391	-1.731	-1.367	-1.684	-1.442	-1.583	-1.25	-1.298
γ_{12}	-0.946	-0.692	-0.8	-1.116	-1.177	-0.971	-1.058	-0.735	-0.89	-0.88
γ_{13}	-0.230	-0.389	0.04	-0.425	-0.739	-0.544	-0.112	-0.035	0.027	-0.351
γ_{14}	-1.606	-1.64	-1.435	-1.43	-1.18	-1.5	-1.455	-1.909	-1.229	-2.631
γ_{15}	-0.682	-1.694	-0.976	-0.407	-0.266	-0.365	-1.434	-0.077	-1.927	-1.238
γ_{16}	-0.982	-0.977	-0.939	-0.895	-0.852	-1.235	-1.224	-0.956	-1.589	-0.915
γ_{17}	-1.882	-2.114	-1.814	-1.677	-1.897	-1.834	-1.896	-1.607	-2.343	-2.092
γ_{18}	-1.882	-2.543	-1.759	-1.567	-2.346	-1.726	-1.804	-1.58	-1.683	-1.649
γ_{19}	-0.882	-0.673	-0.939	-1.002	-0.964	-0.765	-0.905	-0.782	-0.714	-0.587
γ_{20}	-2.282	-2.389	-2.112	-2.437	-1.975	-2.583	-2.108	-2.283	-2.209	-2.407

Table 3. RMSE for the Estimates From MML/EM.

Parameters	$\alpha = .065$			$\alpha = .265$			$\alpha = .365$		
	2,500	3,000	3,500	2,500	3,000	3,500	2,500	3,000	3,500
β_1	0.296	0.283	0.261	0.359	0.331	0.405	0.604	0.503	0.424
β_2	0.164	0.156	0.137	0.263	0.214	0.199	0.396	0.296	0.278
β_3	0.345	0.286	0.350	0.444	0.451	0.367	0.625	0.633	0.608
β_4	0.300	0.256	0.278	0.526	0.548	0.430	0.519	0.521	0.393
β_5	0.274	0.265	0.292	0.433	0.453	0.370	0.577	0.516	0.433
β_6	0.321	0.258	0.264	0.392	0.399	0.410	0.424	0.398	0.417
β_7	0.205	0.192	0.206	0.251	0.233	0.234	0.409	0.294	0.291
β_8	0.203	0.208	0.216	0.366	0.377	0.378	0.547	0.500	0.399
β_9	0.153	0.134	0.142	0.293	0.263	0.224	0.358	0.307	0.277
β_{10}	0.375	0.336	0.319	0.477	0.442	0.429	0.401	0.471	0.409
β_{11}	0.209	0.206	0.231	0.257	0.258	0.249	0.358	0.312	0.330
β_{12}	0.245	0.212	0.227	0.391	0.324	0.318	0.424	0.322	0.313
β_{13}	0.311	0.281	0.321	0.406	0.414	0.375	0.480	0.367	0.365
β_{14}	0.148	0.147	0.149	0.271	0.266	0.243	0.355	0.328	0.305
β_{15}	0.304	0.294	0.274	0.418	0.407	0.392	0.605	0.493	0.455
β_{16}	0.283	0.297	0.321	0.453	0.560	0.568	0.602	0.482	0.486
β_{17}	0.205	0.201	0.204	0.342	0.377	0.306	0.449	0.437	0.381
β_{18}	0.183	0.144	0.178	0.263	0.236	0.195	0.430	0.331	0.328
β_{19}	0.223	0.251	0.229	0.366	0.346	0.433	0.680	0.466	0.386
β_{20}	0.175	0.144	0.163	0.250	0.294	0.252	0.398	0.344	0.305
α	.083	.108	.105	.101	.102	.106	.117	.092	.097
γ_1	0.212	0.249	0.204	0.266	0.258	0.271	0.352	0.251	0.304
γ_2	0.418	0.407	0.468	0.726	0.562	0.531	0.637	0.959	0.701
γ_3	1.039	1.066	0.985	1.359	1.202	1.121	1.322	1.193	1.718
γ_4	0.193	0.191	0.180	0.234	0.230	0.211	0.223	0.207	0.228
γ_5	0.191	0.214	0.219	0.272	0.261	0.246	0.336	0.279	0.276
γ_6	0.469	0.431	0.383	0.813	0.480	0.583	0.701	0.619	0.988
γ_7	0.530	0.388	0.423	0.667	0.480	0.476	0.695	0.850	0.495
γ_8	0.214	0.222	0.195	0.320	0.324	0.313	0.586	0.418	0.386
γ_9	0.671	0.538	0.665	0.826	0.922	0.821	0.976	0.853	0.786
γ_{10}	1.080	1.117	0.922	1.161	0.922	0.942	1.068	0.837	0.982
γ_{11}	0.227	0.258	0.239	0.254	0.278	0.236	0.306	0.284	0.346
γ_{12}	0.257	0.228	0.220	0.344	0.315	0.277	0.342	0.285	0.280
γ_{13}	0.338	0.349	0.344	0.479	0.457	0.413	0.659	0.446	0.461
γ_{14}	0.337	0.315	0.310	0.729	0.642	0.555	0.674	0.676	0.543
γ_{15}	0.672	0.802	0.547	0.905	0.833	0.960	0.849	1.246	0.684
γ_{16}	0.178	0.203	0.191	0.234	0.270	0.224	0.301	0.242	0.220
γ_{17}	0.224	0.226	0.210	0.268	0.242	0.253	0.325	0.324	0.328
γ_{18}	0.475	0.426	0.417	0.651	0.560	0.463	0.735	0.523	0.543
γ_{19}	0.197	0.239	0.202	0.237	0.239	0.256	0.376	0.313	0.304
γ_{20}	0.235	0.250	0.291	0.376	0.383	0.326	0.504	0.403	0.472

Note. RMSE = root mean square error; MML = marginal maximum likelihood; EM = expectation-maximization algorithm.

PROC NLMIXED in a fraction of the time. As the performance of the PROC NLMIXED procedure has not, to our knowledge, been compared with other estimation methods, this study supports the general accuracy of this procedure as well.

The conceptualization and interpretation of guessing in educational testing settings are far from being settled. There is not a single model that fully explains underlying interactions

Table 4. RMSE for the Estimates From PROC NLMIXED.

Parameters	$\alpha = .065$			$\alpha = .265$			$\alpha = .365$		
	2,500	3,000	3,500	2,500	3,000	3,500	2,500	3,000	3,500
β_1	0.341	0.338	0.331	0.522	0.431	0.491	0.868	0.679	0.588
β_2	0.171	0.145	0.138	0.337	0.262	0.259	0.546	0.391	0.372
β_3	0.291	0.282	0.365	0.590	0.580	0.450	0.859	0.763	0.689
β_4	0.354	0.301	0.337	0.815	0.656	0.621	0.776	0.789	0.537
β_5	0.305	0.288	0.312	0.636	0.396	0.489	0.837	0.710	0.586
β_6	0.310	0.287	0.294	0.515	0.444	0.492	0.607	0.511	0.517
β_7	0.218	0.184	0.233	0.320	0.304	0.291	0.620	0.395	0.381
β_8	0.220	0.251	0.265	0.510	0.412	0.449	0.722	0.646	0.463
β_9	0.153	0.133	0.150	0.376	0.328	0.299	0.478	0.386	0.334
β_{10}	0.408	0.351	0.346	0.667	0.525	0.509	0.601	0.587	0.519
β_{11}	0.205	0.212	0.241	0.372	0.293	0.297	1.203	0.428	0.429
β_{12}	0.255	0.243	0.263	0.510	0.434	0.422	0.753	0.441	0.462
β_{13}	0.329	0.305	0.371	0.535	0.465	0.442	0.644	0.523	0.482
β_{14}	0.175	0.154	0.146	0.313	0.317	0.317	0.536	0.434	0.360
β_{15}	0.345	0.302	0.294	0.556	0.496	0.490	0.785	0.607	0.559
β_{16}	0.342	0.369	0.417	0.492	0.599	0.871	1.909	0.693	0.545
β_{17}	0.232	0.235	0.222	0.400	0.456	0.403	0.755	0.638	0.517
β_{18}	0.200	0.164	0.180	0.319	0.207	0.253	0.552	0.414	0.403
β_{19}	0.286	0.300	0.280	0.508	0.373	0.484	1.175	0.608	0.481
β_{20}	0.174	0.171	0.165	0.311	0.299	0.322	0.509	0.436	0.333
α	.119	.171	.137	.153	.121	.143	.195	.134	.120
γ_1	0.251	0.330	0.258	0.359	0.288	0.329	0.505	0.347	0.361
γ_2	0.477	0.417	0.440	0.740	0.653	0.567	1.230	1.248	0.861
γ_3	0.971	1.154	1.130	1.668	1.261	1.525	2.677	2.008	1.786
γ_4	0.231	0.230	0.219	0.306	0.262	0.276	0.318	0.285	0.268
γ_5	0.240	0.263	0.256	0.367	0.272	0.307	0.434	0.352	0.323
γ_6	0.529	0.490	0.493	0.734	0.592	0.627	1.185	1.472	1.256
γ_7	0.608	0.453	0.477	0.952	0.742	0.570	0.893	0.732	0.590
γ_8	0.270	0.288	0.238	0.425	0.364	0.371	0.715	0.513	0.431
γ_9	0.691	0.553	0.718	1.056	0.876	1.795	1.272	0.888	0.812
γ_{10}	1.177	1.701	1.161	1.346	1.013	1.030	2.735	0.968	0.892
γ_{11}	0.269	0.327	0.272	0.335	0.285	0.284	0.407	0.325	0.375
γ_{12}	0.309	0.312	0.270	0.412	0.379	0.343	0.454	0.366	0.354
γ_{13}	0.388	0.420	0.401	0.550	0.501	0.482	0.898	0.603	0.546
γ_{14}	0.420	0.383	0.352	0.923	0.646	0.601	0.811	0.635	0.609
γ_{15}	0.923	0.988	0.634	1.146	1.554	1.550	2.173	1.151	1.537
γ_{16}	0.243	0.300	0.238	0.282	0.252	0.271	0.412	0.309	0.236
γ_{17}	0.218	0.296	0.223	0.339	0.295	0.289	0.439	0.399	0.372
γ_{18}	0.602	0.558	0.462	0.581	0.538	0.425	1.149	0.555	0.635
γ_{19}	0.274	0.324	0.245	0.303	0.252	0.297	0.516	0.382	0.324
γ_{20}	0.256	0.352	0.304	0.416	0.366	0.365	0.573	0.447	0.430

Note. RMSE = root mean square error.

between ability and item characteristics occurring from guessing. Even the most popular model, the 3PL, explains the probability of guessing for low-ability examinees, whereas the 1PL-AG focuses on guessing behavior for mid-ability examinees. For the 1PL-AG, guessing is the secondary or auxiliary problem-solving process conditional on the failure of the main process used to respond to an item. Therefore, low-ability examinees lack this ability, whereas high-ability examinees do not have to depend on it.

Table 5. Average Empirical Standard Error of Estimates for MML/EM and PROC NLMIXED.

α	Number of examinees	MML/EM			PROC NLMIXED		
		$SE(\hat{\alpha})$	$SE(\hat{\beta})$	$SE(\hat{\gamma})$	$SE(\hat{\alpha})$	$SE(\hat{\beta})$	$SE(\hat{\gamma})$
.065	2,500	0.083 (1.18)	0.242 (1.16)	0.403 (1.13)	0.106 (1.33)	0.253 (1.13)	0.423 (1.22)
	3,000	0.105 (0.96)	0.225 (1.09)	0.399 (1.04)	0.146 (0.88)	0.253 (0.99)	0.449 (0.99)
	3,500	0.105 (0.92)	0.232 (1.02)	0.378 (1.02)	0.127 (0.95)	0.254 (1.01)	0.414 (1.06)
.265	2,500	0.100 (1.07)	0.343 (1.12)	0.551 (1.17)	0.140 (1.06)	0.441 (1.12)	0.652 (1.11)
	3,000	0.102 (1.13)	0.343 (1.08)	0.488 (1.11)	0.118 (1.16)	0.391 (1.07)	0.533 (1.19)
	3,500	0.105 (1.12)	0.322 (1.03)	0.471 (1.02)	0.139 (1.13)	0.394 (1.15)	0.585 (1.07)
.365	2,500	0.112 (0.77)	0.433 (0.99)	0.580 (0.82)	0.180 (0.68)	0.667 (0.86)	0.945 (0.66)
	3,000	0.089 (1.04)	0.382 (1.00)	0.545 (0.92)	0.121 (1.02)	0.478 (1.01)	0.665 (0.96)
	3,500	0.096 (1.12)	0.351 (1.06)	0.540 (1.13)	0.114 (1.14)	0.431 (1.05)	0.602 (1.18)

Note. Values within parentheses are the ratio of the estimated standard errors to the empirical standard errors. MML = marginal maximum likelihood; EM = expectation-maximization algorithm.

Table 6. Simulation Time and Fit Statistics for MML and PROC NLMIXED.

Number of examinees	α	MML/EM		PROC NLMIXED	
		Simulation time (s)	$-2 \times I$	Simulation time (hr)	$-2 \times I$
2,500	.065	15.539	61,408.91	1.853	61,425.62
	.265	32.753	60,799.13	2.372	60,795.51
	.365	31.561	60,483.32	2.058	60,470.56
3,000	.065	31.471	73,715.37	2.526	73,701.85
	.265	49.994	72,970.22	3.170	72,979.59
	.365	48.463	72,585.85	2.995	72,571.56
3,500	.065	40.761	85,978.95	2.906	85,946.54
	.265	41.535	85,154.96	3.721	85,203.74
	.365	41.973	84,672.89	3.025	84,727.60

Note. MML = marginal maximum likelihood; EM = expectation-maximization algorithm.

Modeling factors that influence examinee's responses with IRT models is an ongoing effort. Therefore, further research should follow. First, a possible instability between the discrimination and the guessing parameter in the 3PL (van der Linden & Hambleton, 1997) may be applicable to 1PL-AG model. Second, the application of EM for NLMIXED reported by Walker (1996) opens interesting comparison study with MML/EM method. Third, a more systematic approach for obtaining initial values of MML/EM will accelerate wide adoption of MML/EM estimation. As each Newton–Raphson process takes less than a minute for conditions in this study, even with an elaborate search method, MML/EM should still hold a substantial advantage in terms of computing time compared with other methods. Fourth, there has been several studies compared the performance between Markov chain Monte Carlo (MCMC) and MML estimation for binary outcomes (Albert, 1992; Baker, 1998; Kieftenbeld & Natesan, 2012; Kim, 2001). However, to our knowledge, no studies have been done comparing the performance between MCMC and NLMIXED. Therefore, future study comparing the performance between MCMC and SAS PROC NLMIXED for 1PL-AG will provide useful information for the choice of estimation method. Last, two-parameter logistic model with ability-based guessing (2PL-AG) should be investigated at as a natural extension of 1PL-AG. 2PL-AG conceptualizes the discriminating power of each item by allowing discrimination parameters to be freely estimated, while also including ability-based guessing process. However, another interesting approach is to view 2PL-AG as a 3PL model with ability-based guessing. Unlike the constant probability of guessing in 3PL, a monotonic increasing probability of success of guessing on the ability scale can be modeled in 2PL-AG. When α (i.e., the weight on ability in the guessing process) is fixed to zero, 2PL-AG is reduced to 3PL, which makes 3PL a nested model of 2PL-AG. The difference in the number of item parameters between two models is only one. It is believed that a model comparison among 2PL-AG, 3PL, and 1PL-AG would be worth further investigation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational and Behavioral Statistics, 17*, 251-269.
- Baker, F. B. (1998). An investigation of the item parameter recovery characteristics of a Gibbs sampling procedure. *Applied Psychological Measurement, 22*, 153-169.
- 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* (pp. 397-479). Reading, MA: Addison-Wesley.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443-459.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35*, 179-197.
- Gabrielsen, A. (1978). Consistency and identifiability. *Journal of Econometrics, 8*, 261-263.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

- Hutchinson, T. (1991). *Ability, partial information, guessing: Statistical modeling applied to multiple-choice tests*. Adelaide, Australia: Rumsby Scientific.
- Kieftenbeld, V., & Natesan, P. (2012). Recovery of graded response model parameters a comparison of marginal maximum likelihood and Markov chain Monte Carlo estimation. *Applied Psychological Measurement, 36*, 399-419.
- Kim, S.-H. (2001). An evaluation of a Markov chain Monte Carlo method for the Rasch model. *Applied Psychological Measurement, 25*, 163-176.
- Koopmans, T. C., & Reiersol, O. (1950). The identification of structural characteristics. *The Annals of Mathematical Statistics, 21*, 165-181.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Maris, G., & Bechger, T. (2009). On interpreting the model parameters for the three parameter logistic model. *Measurement, 7*, 75-88.
- Mislevy, R. (1984). Estimating latent distributions. *Psychometrika, 49*, 359-381.
- Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods, 8*, 185-205.
- San Martín, E., del Pino, G., & De Boeck, P. (2006). IRT models for ability-based guessing. *Applied Psychological Measurement, 30*, 183-203.
- San Martín, E., Rolin, J.-M., & Castro, L. M. (2013). Identification of the 1PL model with guessing parameter: Parametric and semi-parametric results. *Psychometrika, 78*, 341-379.
- Stroud, A., & Secrest, D. (1966). *Gaussian quadrature formulas* (Vol. 39). Englewood Cliffs, NJ: Prentice Hall.
- Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika, 47*, 175-186.
- van der Linden, W., & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In W. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 1-28). New York, NY: Springer.
- Walker, S. (1996). An EM algorithm for nonlinear random effects models. *Biometrics, 52*, 934-944.