

Ogasawara, H. (2009). *Computational Statistics*, 24 (2), 313-331. doi: 10.1007/s00180-008-0118-8. Self-archived with the permission of Springer Science+Business Media Inc. The original publication is available at www.springerlink.com.

Asymptotic cumulants of the parameter estimators in item response theory

May 1, 2010

(Equation (3.6) and consequently Tables 1-4 and Figures 1-2 were corrected without changing conclusions.)

Haruhiko Ogasawara
Otaru University of Commerce

This work was partially supported by Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology. Author's address: Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501 Japan. <http://www.res.otaru-uc.ac.jp/~hogasa/>

Asymptotic cumulants of the parameter estimators in item response theory

The asymptotic cumulants of the parameter estimators for the three-parameter logistic model in item response theory are derived up to the fourth order with the higher-order added asymptotic variances. The asymptotic cumulants of the corresponding Studentized estimators up to the third order are also given. The estimators are obtained by marginal maximum likelihood using the standard normal distribution for the latent variable with and without model misspecification. Numerical examples with fixed guessing parameters show advantages of the asymptotic expansions over the usual normal approximation.

Key words: item response theory, bias, skewness, marginal maximum likelihood, model misspecification, asymptotic expansion.

1. Introduction

Models in item response theory (IRT) include latent variable(s) representing e.g., abilities of subjects or examinees. When the scores of the latent variables are seen as fixed values, they are unknown parameters to be estimated with structural item parameters (see e.g., Lord & Novick, 1968, Chapter 20; Lord, 1980, Chapter 12), which gives joint maximum likelihood (JML) estimation. On the other hand, when the ability scores are seen as realized values of random variables, only the item parameters are estimated. The latter method was given by Bock and Lieberman (1970) and Bock (1972) for binary and polytomous nominal data, respectively using marginal maximum likelihood (MML) estimation, where the latent variables are integrated out in estimation. One of the problems in JML is that the number of ability parameters increases as the number of subjects (i.e., the problem of incidental parameters), from which MML is free. The method of MML was sophisticated by Bock and Aitkin (1981) using an EM algorithm, and seems to be a standard method currently used routinely in practice (Bock & Moustaki, 2007; Wirth & Edwards, 2007, Table 1).

The asymptotic distributions of the parameter estimators by MML are available using the usual normal approximation with the asymptotic standard errors (see e.g., Bock and Lieberman, 1970). The purpose of this article is to derive the asymptotic cumulants of the estimators up to the fourth order with the added higher-order asymptotic variances in the case of the three-parameter logistic model (3PLM) for binary responses with possible model misspecification. The asymptotic cumulants of the Studentized pivotal statistics up to the third order will also be derived.

By using real life data available in the literature with fixed guessing parameters, it will be shown that the asymptotic standard errors can be substantially different from the actual ones even when the sample size is as large as 1,000. It will also be shown that with the large sample size the distributions of the parameters can be biased and strongly skewed. The asymptotic standard errors currently available are based on the assumption that the model is true. In practice, models fitted to actual data are more or less misspecified especially when the unidimensional latent variable is used (note that most of IRT models used in practice are unidimensional). The correct asymptotic standard errors as well as other asymptotic cumulants under possible model misspecification will be derived. Simulations will be

performed to see the accuracy of the asymptotic results in finite samples.

2. The models and the estimators

It is assumed that the probability of success for the i^* -th examinee with ability θ_{i^*} in the j -th item is given by the logistic model:

$$\Pr(Y_{i^*j} = 1 | \theta_{i^*}, a_j, b_j, c_j) = c_j + \frac{1 - c_j}{1 + \exp\{-Da_j(\theta_{i^*} - b_j)\}} \quad (2.1)$$

$$(i^* = 1, \dots, N; j = 1, \dots, n)$$

with $\theta_{i^*} \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$ ($i^* = 1, \dots, N$), where $Y_{i^*j} = 1$ and $Y_{i^*j} = 0$ indicate the success and failure for the i^* -th examinee in the response to the j -th item, respectively. The a_j, b_j and c_j are the discrimination, difficulty and guessing parameters, respectively for the j -th item. $D=1.7$ is a constant often employed for ease of comparison to the corresponding probit model. For actual computation illustrated in a later section, c_j s are assumed to be known due to the difficulty of estimating them without strong priors or restrictions (see Thissen & Wainer, 1982; Ogasawara, 2002). When $c_j = 0$, (2.1) reduces to the two-parameter logistic model (2PLM), which was used by Bock and Lieberman (1970). With the assumption of the local independence when θ_{i^*} is given, the marginal maximum likelihood is obtained by the multinomial distribution for 2^n categories or patterns as

$$L = \frac{N!}{\prod_{k=1}^K r_k!} \prod_{k=1}^K \pi_k^{r_k},$$

$$K = 2^n,$$

$$\pi_k = \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^n \Psi_j^{X_{kj}} (1 - \Psi_j)^{1 - X_{kj}} \right\} \phi(\theta) d\theta,$$

$$\Psi_j = \Psi(\theta, a_j, b_j, c_j) = c_j + \frac{1 - c_j}{1 + \exp\{-Da_j(\theta - b_j)\}},$$

$$\phi(\theta) = \frac{1}{\sqrt{2}} \exp\left(-\frac{\theta^2}{2}\right) \quad (k = 1, \dots, K; j = 1, \dots, n), \quad (2.2)$$

where r_k is the number of examinees showing the k -th response pattern to n items ($k=1, \dots, K$). The values 1 and 0 of X_{kj} correspond to the success and failure to the j -th item in the k -th response pattern.

Note that in (2.2) the subscript i^* of θ_{i^*} is dropped since the distributions of θ_{i^*} s in the model are the same. It is difficult to obtain the integral of (2.2) in closed form. So, the result is given by the numerical method using M quadrature points Q_1, \dots, Q_M as

$$\pi_k \doteq \sum_{m=1}^M \left\{ \prod_{j=1}^n \Psi_{mj}^{X_{kj}} (1 - \Psi_{mj})^{1-X_{kj}} \right\} A(Q_m) \quad (k = 1, \dots, K), \quad (2.3)$$

where $\Psi_{mj} = \Psi(Q_m, a_j, b_j, c_j)$ ($j = 1, \dots, n$) and $A(Q_m)$ ($m = 1, \dots, M$) is the weight at the quadrature point Q_m .

Let $p_k = r_k / N$ ($k = 1, \dots, K$) be the sample proportion of the k -th response pattern. Then, when the model holds, $E(p_k) = \pi_k$. On the other hand, when the model is misspecified $E(p_k) \equiv \pi_{Tk} \neq \pi_k$ with $\pi_{Tk} - \pi_k = O(1)$ for some k , where π_{Tk} is the true probability for the k -th response pattern. In the following $\pi_{Tk} > 0$ ($k = 1, \dots, K$) are assumed. The case of the true model is given as a special case with $\pi_{Tk} = \pi_k$ ($k = 1, \dots, K$). The MML estimators are obtained by maximizing L with respect to the vector of q unknown parameters

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)' = (a_1, \dots, a_n, b_1, \dots, b_n, c_1, \dots, c_n)' \quad (2.4)$$

using (2.3) for actual computation. Note that p_k ($k = 1, \dots, K$) are the sufficient statistics for the parameters.

3. Asymptotic expansions of the distributions of the estimators.

Let α be a generic parameter representing one of the unknown parameters with α_0 and $\hat{\alpha}$ being the population value and its estimator, respectively. Then, $\hat{\alpha}$ is seen as a function of $\mathbf{p} = (p_1, \dots, p_K)'$ denoted by $\hat{\alpha} = \alpha(\mathbf{p})$ though usually the function is an

implicit one. Let $w = N^{1/2}(\hat{\alpha} - \alpha_0)$ and $\mathbf{u} = N^{1/2}(\mathbf{p} - \boldsymbol{\pi}_T)$, where $\boldsymbol{\pi}_T = (\pi_{T1}, \dots, \pi_{TK})'$. Suppose that we have the Taylor expansion of w about 0 as

$$w = \frac{\partial \alpha}{\partial \boldsymbol{\pi}_T'} \mathbf{u} + \frac{N^{-1/2}}{2} \left\{ \left(\frac{\partial}{\partial \boldsymbol{\pi}_T'} \right)^{\langle 2 \rangle} \alpha \right\} \mathbf{u}^{\langle 2 \rangle} + \frac{N^{-1}}{6} \left\{ \left(\frac{\partial}{\partial \boldsymbol{\pi}_T'} \right)^{\langle 3 \rangle} \alpha \right\} \mathbf{u}^{\langle 3 \rangle} + O_p(N^{-3/2}), \quad (3.1)$$

where $\partial \alpha / \partial \boldsymbol{\pi}_T = \partial \hat{\alpha} / \partial \mathbf{p} |_{\mathbf{p}=\boldsymbol{\pi}_T}$ with other partial derivatives defined in similar manners for simplicity of notation. $\mathbf{a}^{\langle k \rangle} = \mathbf{a} \otimes \dots \otimes \mathbf{a}$ (k times) is the k -fold Kronecker product of \mathbf{a} . Noting that \mathbf{p} is the mean of N independent observations in the multinomial distribution, we have the cumulants of w up to the fourth order as

$$\begin{aligned} \kappa_1(w) &= E(w) = N^{-1/2} \beta_1 + O(N^{-3/2}), \\ \kappa_2(w) &= E[\{w - E(w)\}^2] = \beta_2 + N^{-1} \Delta \beta_2 + O(N^{-2}), \\ \kappa_3(w) &= E[\{w - E(w)\}^3] = N^{-1/2} \beta_3 + O(N^{-3/2}), \\ \kappa_4(w) &= E[\{w - E(w)\}^4] - 3\{\kappa_2(w)\}^2 = N^{-1} \beta_4 + O(N^{-2}), \end{aligned} \quad (3.2)$$

where $N^{-1} \beta_1$, $N^{-1} \beta_2$, $N^{-2} \Delta \beta_2$, $N^{-1/2} \beta_3 / \beta_2^{3/2}$ and $N^{-1} \beta_4 / \beta_2^2$ are the asymptotic bias, variance, added higher-order variance, skewness and kurtosis of $\hat{\alpha}$, respectively.

From (3.1) and (3.2), we have

$$\beta_1 = \frac{1}{2} \text{tr} \left(\frac{\partial^2 \alpha}{\partial \boldsymbol{\pi}_T \partial \boldsymbol{\pi}_T'} \boldsymbol{\Omega} \right), \quad \beta_2 = \frac{\partial \alpha}{\partial \boldsymbol{\pi}_T'} \boldsymbol{\Omega} \frac{\partial \alpha}{\partial \boldsymbol{\pi}_T}, \quad (3.3)$$

where

$$(\boldsymbol{\Omega})_{ij} = \omega_{ij} = E(u_i u_j) = N \text{cov}(p_i, p_j) = \delta_{ij} \pi_{Ti} - \pi_{Ti} \pi_{Tj} \quad (i, j = 1, \dots, K), \quad (3.4)$$

$(\cdot)_{ij}$ denotes the (i, j) th element of the argument matrix, u_i is the i -th element of \mathbf{u} and δ_{ij} is the Kronecker delta. The expressions of $\Delta \beta_2$, β_3 and β_4 are given from (3.1) as in the case of covariance structure analysis (see Ogasawara, 2006, 2007):

$$\Delta\beta_2 = \sum_{i,j,k=1}^K \left\{ \frac{\partial\alpha}{\partial\pi_{Ti}} \frac{\partial^2\alpha}{\partial\pi_{Tj}\partial\pi_{Tk}} J(i,j,k) + \sum_{l=1}^K \left(\frac{1}{2} \frac{\partial^2\alpha}{\partial\pi_{Ti}\partial\pi_{Tk}} \frac{\partial^2\alpha}{\partial\pi_{Tj}\partial\pi_{Tl}} + \frac{\partial\alpha}{\partial\pi_{Tl}} \frac{\partial^3\alpha}{\partial\pi_{Ti}\partial\pi_{Tj}\partial\pi_{Tk}} \right) \omega_{ij}\omega_{kl} \right\}, \quad (3.5)$$

where $\sum_{i,j,k=1}^K$ is $\sum_{i=1}^K \sum_{j=1}^K \sum_{k=1}^K$, and

$$\begin{aligned} J(i,j,k) &= N^2 \mathbf{E}\{(p_i - \pi_{Ti})(p_j - \pi_{Tj})(p_k - \pi_{Tk})\} \\ &= \delta_{ij}\delta_{ik}(\pi_{Ti} - 3\pi_{Ti}^2) - \{\delta_{ij}(1 - \delta_{ik})\pi_{Ti}\pi_{Tk} + \delta_{ik}(1 - \delta_{ij})\pi_{Ti}\pi_{Tj} \\ &\quad + \delta_{jk}(1 - \delta_{ji})\pi_{Tj}\pi_{Ti}\} + 2\pi_{Ti}\pi_{Tj}\pi_{Tk} \quad (i,j,k = 1, \dots, K) \end{aligned} \quad (3.6)$$

(for (3.6) see e.g., Stuart & Ort, 1994, Equation (7.18)),

$$\begin{aligned} \beta_3 &= \sum_{i,j,k=1}^K \frac{\partial\alpha}{\partial\pi_{Ti}} \frac{\partial\alpha}{\partial\pi_{Tj}} \frac{\partial\alpha}{\partial\pi_{Tk}} J(i,j,k) + 3 \frac{\partial\alpha}{\partial\pi_T} \mathbf{\Omega} \frac{\partial^2\alpha}{\partial\pi_T\partial\pi_T} \mathbf{\Omega} \frac{\partial\alpha}{\partial\pi_T}, \\ \beta_4 &= \sum_{i,j,k,l=1}^K \left\{ \frac{\partial\alpha}{\partial\pi_{Ti}} \frac{\partial\alpha}{\partial\pi_{Tj}} \frac{\partial\alpha}{\partial\pi_{Tk}} \frac{\partial\alpha}{\partial\pi_{Tl}} J(i,j,k,l) \right. \\ &\quad + 2 \sum_{m=1}^K \frac{\partial^2\alpha}{\partial\pi_{Ti}\partial\pi_{Tj}} \frac{\partial\alpha}{\partial\pi_{Tk}} \frac{\partial\alpha}{\partial\pi_{Tl}} \frac{\partial\alpha}{\partial\pi_{Tm}} \sum_{5C_2=10} \omega_{ij} J(k,l,m) \\ &\quad + \sum_{m,a=1}^K \left(\frac{3}{2} \frac{\partial^2\alpha}{\partial\pi_{Ti}\partial\pi_{Tj}} \frac{\partial^2\alpha}{\partial\pi_{Tk}\partial\pi_{Tl}} + \frac{2}{3} \frac{\partial^3\alpha}{\partial\pi_{Ti}\partial\pi_{Tj}\partial\pi_{Tk}} \frac{\partial\alpha}{\partial\pi_{Tl}} \right) \\ &\quad \left. \times \frac{\partial\alpha}{\partial\pi_{Tm}} \frac{\partial\alpha}{\partial\pi_{Ta}} \sum_{15} \omega_{ij}\omega_{kl}\omega_{ma} \right\} - (4\beta_1\beta_3 + 6\beta_2\Delta\beta_2 + 6\beta_2\beta_1^2), \end{aligned} \quad (3.7)$$

where $J(i,j,k,l)$ is the N^3 times the fourth-order multivariate cumulant of p_i, p_j, p_k and p_l , which is available as (see e.g., Stuart & Ort, 1994, Equation (7.18))

$$\begin{aligned}
N^3 \kappa(p_i, p_i, p_i, p_i) &= \pi_{Ti}(1 - \pi_{Ti})\{1 - 6\pi_{Ti}(1 - \pi_{Ti})\}, \\
N^3 \kappa(p_i, p_i, p_i, p_j) &= -\pi_{Ti}\pi_{Tj}\{1 - 6\pi_{Ti}(1 - \pi_{Ti})\}, \\
N^3 \kappa(p_i, p_i, p_j, p_j) &= -\pi_{Ti}\pi_{Tj}\{(1 - 2\pi_{Ti})(1 - 2\pi_{Tj}) + 2\pi_{Ti}\pi_{Tj}\}, \\
N^3 \kappa(p_i, p_i, p_j, p_k) &= 2\pi_{Ti}\pi_{Tj}\pi_{Tk}(1 - 3\pi_{Ti}), \\
N^3 \kappa(p_i, p_j, p_k, p_l) &= -6\pi_{Ti}\pi_{Tj}\pi_{Tk}\pi_{Tl} \\
&(i, j, k, l = 1, \dots, K; i \neq j, i \neq k, i \neq l, j \neq k, j \neq l, k \neq l);
\end{aligned} \tag{3.8}$$

and \sum^k in (3.7) denotes the sum of k similar terms.

When the model is true, the expression of β_2 in (3.3) becomes simplified by using the inverse of the information matrix per observation. The expression of $\Delta\beta_2$ in (3.5) has been somewhat simplified from the case of covariance structure analysis due to the exact covariance of (3.4). The expression of (3.3), (3.5) and (3.7) include partial derivatives of $\hat{\alpha}$ with respect to \mathbf{p} evaluated at $\mathbf{p} = \boldsymbol{\pi}_T$ up to the third order, which are given from the formulas of partial derivatives in implicit functions using the first-order condition of $\hat{\alpha}$. The results will be given in the appendix.

In practice, the following Studentized pivotal statistic is used since usually population asymptotic standard errors are not available:

$$t = \frac{N^{1/2}(\hat{\alpha} - \alpha_0)}{\hat{\beta}_2^{1/2}}. \tag{3.9}$$

It is assumed that the Taylor expansion for t holds as

$$\begin{aligned}
t &= \beta_2^{-1/2} \frac{\partial \alpha}{\partial \boldsymbol{\pi}_T'} \mathbf{u} + \frac{N^{-1/2}}{2} \beta_2^{-1/2} \left\{ \left(\frac{\partial}{\partial \boldsymbol{\pi}_T'} \right)^{\langle 2 \rangle} \alpha \right\} \mathbf{u}^{\langle 2 \rangle} \\
&\quad - \frac{N^{-1/2}}{2} \beta_2^{-3/2} \frac{\partial \alpha}{\partial \boldsymbol{\pi}_T'} \mathbf{u} \frac{\partial \beta_2}{\partial \boldsymbol{\pi}_T'} \mathbf{u} + O_p(N^{-1}).
\end{aligned} \tag{3.10}$$

For the asymptotic cumulants of t up to the third order, we assume that

$$\begin{aligned}
\kappa_1(t) &= N^{-1/2} \beta_1' + O(N^{-3/2}), \\
\kappa_2(t) &= 1 + O(N^{-1}), \quad (\beta_2' = 1), \\
\kappa_3(t) &= N^{-1/2} \beta_3' + O(N^{-3/2}).
\end{aligned} \tag{3.11}$$

From (3.10) with (3.11), we have

$$\begin{aligned}
\beta_1' &= \beta_2^{-1/2} \beta_1 - \frac{1}{2} \beta_2^{-3/2} \frac{\partial \alpha}{\partial \boldsymbol{\pi}_T'} \boldsymbol{\Omega} \frac{\partial \beta_2}{\partial \boldsymbol{\pi}_T}, \\
\beta_3' &= \beta_2^{-3/2} \beta_3 - 3 \beta_2^{-3/2} \frac{\partial \alpha}{\partial \boldsymbol{\pi}_T'} \boldsymbol{\Omega} \frac{\partial \beta_2}{\partial \boldsymbol{\pi}_T},
\end{aligned} \tag{3.12}$$

with

$$\begin{aligned}
\frac{\partial \beta_2}{\partial \pi_{Tk}} &= 2 \frac{\partial^2 \alpha}{\partial \pi_{Tk} \partial \boldsymbol{\pi}_T'} \boldsymbol{\Omega} \frac{\partial \alpha}{\partial \boldsymbol{\pi}_T} + \frac{\partial \alpha}{\partial \boldsymbol{\pi}_T'} \frac{\partial \boldsymbol{\Omega}}{\partial \pi_{Tk}} \frac{\partial \alpha}{\partial \boldsymbol{\pi}_T}, \\
\frac{\partial \boldsymbol{\Omega}}{\partial \pi_{Tk}} &= \frac{\partial \{\text{diag}(\boldsymbol{\pi}_T) - \boldsymbol{\pi}_T \boldsymbol{\pi}_T'\}}{\partial \pi_{Tk}} = \mathbf{E}_{kk} - \mathbf{e}_k \boldsymbol{\pi}_T' - \boldsymbol{\pi}_T \mathbf{e}_k' \quad (k = 1, \dots, K),
\end{aligned} \tag{3.13}$$

where $\text{diag}(\boldsymbol{\pi}_T)$ is the diagonal matrix whose diagonal elements are π_{Tk} s, \mathbf{E}_{kk} is the matrix whose (k, k) th element is 1 with others being 0, and \mathbf{e}_k is the k -th column of the $K \times K$ identity matrix.

When β_i ($i = 1, \dots, 4$) and $\Delta \beta_2$ are available, the approximation to the density by the two-term Edgeworth expansion is given by

$$\begin{aligned}
f\left(\frac{w}{\beta_2^{1/2}} = x\right) &= \left[1 + N^{-1/2} \left\{ \frac{\beta_1 x}{\beta_2^{1/2}} + \frac{\beta_3}{6\beta_2^{3/2}} (x^3 - 3x) \right\} + N^{-1} \left\{ \frac{1}{2} (\Delta \beta_2 + \beta_1^2) \frac{x^2 - 1}{\beta_2} \right. \right. \\
&\quad \left. \left. + \left(\frac{\beta_4}{24} + \frac{\beta_1 \beta_3}{6} \right) \frac{x^4 - 6x^2 + 3}{\beta_2^2} + \frac{\beta_3^2 (x^6 - 15x^4 + 45x^2 - 15)}{72\beta_2^3} \right\} \right] \phi(x) + O(N^{-3/2})
\end{aligned} \tag{3.14}$$

(see e.g., Hall, 1992a), where the approximation up to $O(N^{-1/2})$ is the single-term Edgeworth expansion. The Cornish-Fisher expansion using consistent estimators of β_1' , β_3' and β_2 is useful for constructing confidence intervals with typical asymptotic confidence coefficient $1 - \tilde{\alpha}$ (e.g., $\tilde{\alpha} = 0.05$) accurate up to $O(N^{-1/2})$:

$$\hat{\theta} + [\pm z_{\tilde{\alpha}/2} - n^{-1/2} \{ \hat{\beta}_1' + (\hat{\beta}_3' / 6)(z_{\tilde{\alpha}/2}^2 - 1) \}] n^{-1/2} \hat{\beta}_2^{1/2}, \quad (3.15)$$

where $\int_{-\infty}^{z_{\tilde{\alpha}/2}} \phi(z^*) dz^* = 1 - (\tilde{\alpha} / 2)$. Hall's (1992b) method by variable transformation removing asymptotic skewness, whose accuracy is asymptotically equivalent to the single-term Edgeworth expansion, has the property free from the anomalous phenomena of locally negative densities which can happen in (3.14) with finite N . Hall's method can also be used with $\hat{\beta}_1'$ and $\hat{\beta}_3'$ as in (3.15).

4. Numerical illustration with simulations

Two sets of numerical examples are illustrated in this section. The data sets used are binary responses by 1,000 subjects to 5 selected items each in Sections 6 and 7 of the Law School Admission Test (LSAT). The data were provided by Bock and Lieberman (1970, p.188) and reanalyzed by Bock and Aitkin (1981). While Bock and his colleagues used the two-parameter normal ogive model, 2PLM is used in this section.

The item parameters were estimated by MML with 15 quadrature points for the unidimensional ability distribution. The estimates (used as population vales in simulations) in the selected items are shown in Tables 1 and 2 for Sections 6 and 7, respectively, and were used with other estimates as the population parameters in simulations. Tables 1 and 2 show results with the assumption of true models (a misspecified case will be given later). The tables include asymptotic results for non-Studentized estimators and Studentized ones independent of N except for $HASE/ASE = \{(\beta_2 / N) + (\Delta\beta_2 / N^2)\}^{1/2} / (\beta_2 / N)^{1/2}$ for non-Studentized estimators.

In the simulations, the fitted proportions for response patterns were regarded as probabilities in the multinomial distribution. Using the distribution, random response vectors were generated. The sample sizes 1,000 and 2,000 were used. From the generated observations, the item parameters were estimated as for the original estimates. This procedure was replicated until 10,000 regular sets of estimates were obtained. The numbers of excluded samples due to nonconvergence are 40 ($N=1,000$, Table 1), 1 ($N=2,000$, Table 1), 2 ($N=1,000$, Table 2) and 0 ($N=2,000$, Table 2). The data for Section 7 seem to give more stable results than those of Section 6 though the items in Section 6 are more

homogeneous than those in Section 7 (Bock & Lieberman, 1970, p.187).

From 10,000 sets of estimates, the simulated cumulants were given from the k -statistics (unbiased estimators of cumulants) multiplied by appropriate powers of N for ease of comparison to the corresponding asymptotic values. The indexes of skewness and kurtosis of the non-Studentized estimators were given from the ratios of the associated asymptotic (simulated) cumulants raised by 1.5 or 2 for $\beta_2(\hat{\beta}_2)$. The simulated HASE/ASE is given by SD/ASE, where SD is the square root of the unbiased sample variance from 10,000 estimates for each parameter.

From the results in Table 1, we see that the speed of convergence of the simulated values to the asymptotic values for non-Studentized estimators is slow considering the original large sample size $N=1,000$. Note that $\hat{\beta}_2^{1/2}$ of \hat{b}_5 is two times of the asymptotic value with $N=1,000$, which can be partially explained by the substantial magnitude of the added higher-order variance. The ratio HASE/ASE for \hat{b}_5 is as large as 1.30 though it is still conservative (note that the corresponding simulated value is 2.00 as addressed earlier). Fortunately, the simulated cumulants of the Studentized estimators are relatively closer to the corresponding asymptotic values. In Tables 1 and 2, the bias and skewness of the a -parameters for the non-Studentized estimators are positive while those for the b -parameters are negative. It is of interest to find that the signs are reversed for the Studentized estimators

Table 3 shows the similar results in Section 7 with model misspecification. The true probabilities of the multinomial distribution were given from the original sample proportions. In the simulations, the true probabilities were estimated simply by their corresponding sample proportions. While most of the results in Table 3 are similar to those in Table 2, some of the simulated/asymptotic results are different (e.g., $\beta_1=5.4$ and 8.7 for \hat{a}_1 in Tables 2 and 3, respectively). Note that the differences in asymptotic values in the tables are of order $O(1)$ and do not vanish with $N \rightarrow \infty$.

Insert Tables 1-4 and Figures 1 and 2 about here.

Table 4 gives the root mean square errors (RMSEs) of the distribution functions of the standardized estimators $(N^{1/2}(\hat{\alpha} - \alpha_0) / \beta_2^{1/2})$ approximated by the standard normal distribution (N*), the single-term (E1) and two-term (E2) Edgeworth expansions, and Hall's (1992b) method by variable transformation (Hall). The true distribution functions were given from the simulated cumulative distributions at the 40 points -3.8 (0.2) 4.0. The RMSEs were obtained over the 40 points. The table shows that E1, E2 and Hall have smaller errors than N*, as a whole E2 has smaller errors than E1 and Hall, and Hall seems to have slightly smaller errors than E1 in these data.

Figure 1 gives the simulated (histograms) and theoretical (curved lines) densities of the selected standardized estimators in Section 7 under the true model with $N=1,000$. The simulated results were given from the simulations in Table 2. Figure 2 gives the similar results for Studentized estimators. From the figures, we find that many of the simulated distributions especially for the b -parameters are biased and substantially skewed.

5. Some conclusive remarks

From the results of the numerical examples, we find that the normal approximations to the distributions of parameter estimators are not satisfactory even with the large sample size $N=1,000$ in the data, which gives advantages for the asymptotic expansions beyond the normal approximation. This corresponds to the finding that for models of categorical data as in IRT, sample sizes larger than those for the corresponding models of continuous variables are required for stable estimation (Wirth & Edwards, 2007, p.73). This may be partially explained by loss of information by categorizing continuous latent variables. The author conjectures that the asymptotic normality for the estimators in IRT generally appears more slowly than in factor analysis for continuous observable variables.

For the distribution of θ , the standard normal was used for simplicity, while the corresponding nonparametric distribution can be estimated by histograms (Bock & Aitkin, 1981; Mislavy, 1984). In principle, the asymptotic expansions of the distributions of the estimated histograms can be obtained as well as those of the estimators of item parameters. Note that for model identification, at least two restrictions should be imposed on the histograms to remove the indeterminacy of the location and scale of the distribution of θ .

The results of the numerical examples were given by the 2PLM. For the 3PLM with free c -parameters, the results of the partial derivatives are shown in the appendix. The theoretical results for the asymptotic expansion for the 3PLM can be obtained using the partial derivatives though the numerical results were not given in this paper due to the difficulty for the corresponding simulated results mentioned earlier. The theoretical results may be used as upper (lower) bounds for the asymptotic cumulants of the estimators with some restrictions for stable estimation used in practice.

Overall, the effect of model misspecification in the numerical examples was not large in practical sense. However, it is dangerous to generalize the conclusion to other data with e.g., gross model misspecification.

In this paper, the 3PLM with usual model specification i.e., $v_j = Da_j(\theta - b_j)$ is used, while we can use the intercept parameter as $v_j = D(a_j\theta + b_j^*)$, where $b_j^* = -a_j b_j$. The difficulty parameter b_j corresponds to the probability $c_j + \{(1 - c_j)/2\} = (1 + c_j)/2$ when $\theta = b_j$ while b_j^* is D^{-1} times the logit when $\theta = 0$. The asymptotic expansions for \hat{b}_j^* can be easily obtained by using $\partial \hat{v}_j / \partial \hat{\mathbf{h}}_j^* = D(\theta, 1)'$, $\partial^2 \hat{v}_j / \partial \hat{\mathbf{h}}_j^{*2} = \mathbf{0}$ and $\hat{\mathbf{h}}_j^* = (\hat{a}_j, \hat{b}_j^*)'$ with the remaining partial derivatives being obtained by substituting $\hat{\mathbf{h}}_j^*$ for $\hat{\mathbf{h}}_j$ in the appendix.

Appendix. The partial derivatives

A1. The partial derivatives of $\hat{\boldsymbol{\alpha}}$ with respect to \mathbf{p}

Let $l^* = N^{-1} \log L$ and \hat{l}^* be the value using the parameter estimators. Then, the first order conditions of $\hat{\boldsymbol{\alpha}}$ are given by

$$\frac{\partial \hat{l}^*}{\partial \hat{\boldsymbol{\alpha}}} = \sum_{k=1}^K \frac{p_k}{\hat{\pi}_k} \frac{\partial \hat{\pi}_k}{\partial \hat{\boldsymbol{\alpha}}} = \mathbf{0}, \quad (\text{A.1})$$

where $\hat{\pi}_k$ is π_k in (2.2) with the parameters being replaced by their estimators, and the partial derivatives denote those evaluated at the parameter estimates.

Differentiating (A.1) and noting \hat{l}^* is a linear function with respect to \mathbf{p} , we have

$$\begin{aligned}
\frac{\partial \hat{\mathbf{a}}}{\partial \mathbf{p}'} &= - \left(\frac{\partial^2 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \hat{\mathbf{a}}'} \right)^{-1} \frac{\partial^2 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \mathbf{p}'}, \\
\frac{\partial^2 \hat{\mathbf{a}}}{\partial p_i \partial p_j} &= - \left(\frac{\partial^2 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \hat{\mathbf{a}}'} \right)^{-1} \sum_{a=1}^q \left\{ \sum_{b=1}^q \frac{\partial^3 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \hat{\alpha}_a \partial \hat{\alpha}_b} \frac{\partial \hat{\alpha}_a}{\partial p_i} \frac{\partial \hat{\alpha}_b}{\partial p_j} + \sum_{(i,j)}^2 \frac{\partial^3 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \hat{\alpha}_a \partial p_i} \frac{\partial \hat{\alpha}_a}{\partial p_j} \right\}, \\
\frac{\partial^3 \hat{\mathbf{a}}}{\partial p_i \partial p_j \partial p_k} &= - \left(\frac{\partial^2 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \hat{\mathbf{a}}'} \right)^{-1} \sum_{a=1}^q \left[\sum_{b=1}^q \left\{ \sum_{c=1}^q \frac{\partial^4 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \hat{\alpha}_a \partial \hat{\alpha}_b \partial \hat{\alpha}_c} \frac{\partial \hat{\alpha}_a}{\partial p_i} \frac{\partial \hat{\alpha}_b}{\partial p_j} \frac{\partial \hat{\alpha}_c}{\partial p_k} \right. \right. \\
&\quad \left. \left. + \sum_{(i,j,k)}^3 \left(\frac{\partial^3 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \hat{\alpha}_a \partial \hat{\alpha}_b} \frac{\partial \hat{\alpha}_a}{\partial p_i} \frac{\partial^2 \hat{\alpha}_b}{\partial p_j \partial p_k} + \frac{\partial^4 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \hat{\alpha}_a \partial \hat{\alpha}_b \partial p_i} \frac{\partial \hat{\alpha}_a}{\partial p_j} \frac{\partial \hat{\alpha}_b}{\partial p_k} \right) \right\} \right. \\
&\quad \left. + \sum_{(i,j,k)}^3 \frac{\partial^3 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \hat{\alpha}_a \partial p_i} \frac{\partial^2 \hat{\alpha}_a}{\partial p_j \partial p_k} \right] \quad (i, j, k = 1, \dots, K), \tag{A.2}
\end{aligned}$$

where the existence of the inverse is assumed, $\sum_{(i,j)}^2$ is the sum of two terms with similar patterns with respect to a and b , and $\sum_{(i,j,k)}^3$ is similarly defined.

A2. The partial derivatives of \hat{l}^* with respect to $\hat{\mathbf{a}}$ and \mathbf{p}

The partial derivatives in (A.2) include those of \hat{l}^* with respect to $\hat{\mathbf{a}}$ and \mathbf{p} , which are given as follow:

$$\begin{aligned}
\frac{\partial^2 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \hat{\mathbf{a}}'} &= \sum_{i=1}^K \left(\frac{p_i}{\hat{\pi}_i} \frac{\partial^2 \hat{\pi}_i}{\partial \hat{\mathbf{a}} \partial \hat{\mathbf{a}}'} - \frac{p_i}{\hat{\pi}_i^2} \frac{\partial \hat{\pi}_i}{\partial \hat{\mathbf{a}}} \frac{\partial \hat{\pi}_i}{\partial \hat{\mathbf{a}}'} \right), \quad \frac{\partial^2 \hat{l}^*}{\partial \hat{\mathbf{a}} \partial \mathbf{p}'} = \frac{\partial \hat{\boldsymbol{\pi}}'}{\partial \hat{\mathbf{a}}} \{\text{diag}(\hat{\boldsymbol{\pi}})\}^{-1}, \\
\frac{\partial^3 \hat{l}^*}{\partial \hat{\alpha}_a \partial \hat{\alpha}_b \partial \hat{\alpha}_c} &= \sum_{i=1}^K \left(\frac{p_i}{\hat{\pi}_i} \frac{\partial^3 \hat{\pi}_i}{\partial \hat{\alpha}_a \partial \hat{\alpha}_b \partial \hat{\alpha}_c} - \sum_{(a,b,c)}^3 \frac{p_i}{\hat{\pi}_i^2} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_a} \frac{\partial^2 \hat{\pi}_i}{\partial \hat{\alpha}_b \partial \hat{\alpha}_c} \right. \\
&\quad \left. + \frac{2p_i}{\hat{\pi}_i^3} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_a} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_b} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_c} \right), \\
\frac{\partial^3 \hat{l}^*}{\partial \hat{\alpha}_a \partial \hat{\alpha}_b \partial p_k} &= \frac{1}{\hat{\pi}_k} \frac{\partial^2 \hat{\pi}_k}{\partial \hat{\alpha}_a \partial \hat{\alpha}_b} - \frac{1}{\hat{\pi}_k^2} \frac{\partial \hat{\pi}_k}{\partial \hat{\alpha}_a} \frac{\partial \hat{\pi}_k}{\partial \hat{\alpha}_b}, \\
\frac{\partial^4 \hat{l}^*}{\partial \hat{\alpha}_a \partial \hat{\alpha}_b \partial \hat{\alpha}_c \partial \hat{\alpha}_d} &= \sum_{i=1}^K \left(\frac{p_i}{\hat{\pi}_i} \frac{\partial^4 \hat{\pi}_i}{\partial \hat{\alpha}_a \partial \hat{\alpha}_b \partial \hat{\alpha}_c \partial \hat{\alpha}_d} - \sum_{(a,b,c,d)}^4 \frac{p_i}{\hat{\pi}_i^2} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_a} \frac{\partial^3 \hat{\pi}_i}{\partial \hat{\alpha}_b \partial \hat{\alpha}_c \partial \hat{\alpha}_d} \right. \\
&\quad - \sum_{(a,b,c,d)}^3 \frac{p_i}{\hat{\pi}_i^2} \frac{\partial^2 \hat{\pi}_i}{\partial \hat{\alpha}_a \partial \hat{\alpha}_b} \frac{\partial^2 \hat{\pi}_i}{\partial \hat{\alpha}_c \partial \hat{\alpha}_d} + \sum_{(a,b,c,d)}^6 \frac{2p_i}{\hat{\pi}_i^3} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_a} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_b} \frac{\partial^2 \hat{\pi}_i}{\partial \hat{\alpha}_c \partial \hat{\alpha}_d} \\
&\quad \left. - \frac{6p_i}{\hat{\pi}_i^4} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_a} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_b} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_c} \frac{\partial \hat{\pi}_i}{\partial \hat{\alpha}_d} \right), \\
\frac{\partial^4 \hat{l}^*}{\partial \hat{\alpha}_a \partial \hat{\alpha}_b \partial \hat{\alpha}_c \partial p_k} &= \frac{1}{\hat{\pi}_k} \frac{\partial^3 \hat{\pi}_k}{\partial \hat{\alpha}_a \partial \hat{\alpha}_b \partial \hat{\alpha}_c} - \sum_{(a,b,c)}^3 \frac{1}{\hat{\pi}_k^2} \frac{\partial \hat{\pi}_k}{\partial \hat{\alpha}_a} \frac{\partial^2 \hat{\pi}_k}{\partial \hat{\alpha}_b \partial \hat{\alpha}_c} \\
&\quad + \frac{2}{\hat{\pi}_k^3} \frac{\partial \hat{\pi}_k}{\partial \hat{\alpha}_a} \frac{\partial \hat{\pi}_k}{\partial \hat{\alpha}_b} \frac{\partial \hat{\pi}_k}{\partial \hat{\alpha}_c}
\end{aligned} \tag{A.3}$$

$(a, b, c, d = 1, \dots, q; k = 1, \dots, K).$

A3. The partial derivatives of $\hat{\pi}_i$ with respect to $\hat{\mathbf{a}}$

The results in (A.3) include the partial derivatives of $\hat{\pi}_i$ with respect to $\hat{\mathbf{a}}$, which are given in this subsection. Let $\mathbf{d}_j = (\mathbf{h}_j', c_j)'$ with $\mathbf{h}_j = (a_j, b_j)'$,

$$\Psi_j = \Psi(\theta, \mathbf{d}_j) = c_j + \frac{1 - c_j}{1 + \exp\{-Da_j(\theta - b_j)\}} \tag{A.4}$$

and $\hat{\Psi}_j = \Psi(\theta, \hat{\mathbf{d}}_j)$ ($j = 1, \dots, n$). Then,

$$\begin{aligned}\hat{\pi}_i &= \int_{-\infty}^{\infty} \hat{\Gamma}_i \phi(\theta) d\theta, \\ \hat{\Gamma}_i &= \prod_{j=1}^n \hat{\Psi}_j^{X_{ij}} (1 - \hat{\Psi}_j)^{1-X_{ij}} \quad (i = 1, \dots, K),\end{aligned}\tag{A.5}$$

where the integral should be replaced by summation with appropriate quadrature weights for actual computation. The partial derivatives of $\hat{\pi}_i$ with respect to $\hat{\mathbf{d}}_j$ are as follows.

$$\frac{\partial \hat{\pi}_i}{\partial \hat{\mathbf{d}}_j} = (-1)^{X_{ij}+1} \int_{-\infty}^{\infty} \frac{\partial \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j} \hat{\Gamma}_{i(j)} \phi(\theta) d\theta,\tag{A.6}$$

where

$$\begin{aligned}\frac{\partial \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j} &= \left\{ \frac{1}{1 - \hat{c}_j} (\hat{\Psi}_j - \hat{c}_j) (1 - \hat{\Psi}_j) \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j}, \frac{1 - \hat{\Psi}_j}{1 - \hat{c}_j} \right\}' \\ &= \left\{ (1 - \hat{c}_j) \hat{B}_j (1 - \hat{B}_j) \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j}, 1 - \hat{B}_j \right\}'\end{aligned}\tag{A.7}$$

with $\hat{v}_j \equiv Da_j(\theta - b_j)$, $\hat{B}_j \equiv 1/\{1 + \exp(-\hat{v}_j)\}$ and $\partial \hat{v}_j / \partial \hat{\mathbf{h}}_j = D(\theta - \hat{b}_j, -\hat{a}_j)'$;

and $\hat{\Gamma}_{i(j)} = \hat{\Gamma}_i / \{\hat{\Psi}_j^{X_{ij}} (1 - \hat{\Psi}_j)^{1-X_{ij}}\}$ ($i = 1, \dots, K$; $j = 1, \dots, n$).

The second partial derivatives are given as:

$$\frac{\partial^2 \hat{\pi}_i}{\partial \hat{\mathbf{d}}_j^{<2>}} = (-1)^{X_{ij}+1} \int_{-\infty}^{\infty} \frac{\partial^2 \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j^{<2>}} \hat{\Gamma}_{i(j)} \phi(\theta) d\theta\tag{A.9}$$

with the nonzero partial derivatives being

$$\frac{\partial^2 \hat{\Psi}_j}{\partial \hat{\mathbf{h}}_j^{<2>}} = (1 - \hat{c}_j) \hat{B}_j (1 - \hat{B}_j) \left\{ (1 - 2\hat{B}_j) \left(\frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \right)^{<2>} + \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j^{<2>}} \right\}' ,$$

where $\frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j^{<2>}} = D(0, -1, -1, 0)'$ and $\frac{\partial^2 \hat{\Psi}_j}{\partial \hat{\mathbf{h}}_j \partial \hat{c}_j} = -\hat{B}_j (1 - \hat{B}_j) \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j}$;

$$\frac{\partial^2 \hat{\pi}_i}{\partial \hat{\mathbf{d}}_j \otimes \partial \hat{\mathbf{d}}_k} = (-1)^{X_{ij}+X_{ik}} \int_{-\infty}^{\infty} \left(\frac{\partial \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j} \otimes \frac{\partial \hat{\Psi}_k}{\partial \hat{\mathbf{d}}_k} \right) \hat{\Gamma}_{i(jk)} \phi(\theta) d\theta \quad (\text{A.10})$$

with

$$\hat{\Gamma}_{i(jk)} = \hat{\Gamma}_i / \{ \hat{\Psi}_j^{X_{ij}} (1 - \hat{\Psi}_j)^{1-X_{ij}} \hat{\Psi}_k^{X_{ik}} (1 - \hat{\Psi}_k)^{1-X_{ik}} \} \quad (i=1, \dots, K; j, k=1, \dots, n, j \neq k).$$

The third partial derivatives are given below:

$$\frac{\partial^3 \hat{\pi}_i}{\partial \hat{\mathbf{d}}_j^{<3>}} = (-1)^{X_{ij}+1} \int_{-\infty}^{\infty} \frac{\partial^3 \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j^{<3>}} \hat{\Gamma}_{i(j)} \phi(\theta) d\theta \quad (\text{A.11})$$

with the nonzero partial derivatives being

$$\begin{aligned} \frac{\partial^3 \hat{\Psi}_j}{\partial \hat{\mathbf{h}}_j^{<3>}} &= (1 - \hat{c}_j) \hat{\mathbf{B}}_j (1 - \hat{\mathbf{B}}_j) \left\{ (1 - 6\hat{\mathbf{B}}_j + 6\hat{\mathbf{B}}_j^2) \left(\frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \right)^{<3>} \right. \\ &\quad \left. + (1 - 2\hat{\mathbf{B}}_j) \sum_{\langle 2 \rangle}^3 \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j^{\langle 2 \rangle}} \otimes \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \right\} \end{aligned}$$

$$\text{and } \frac{\partial^3 \hat{\Psi}_j}{\partial \hat{\mathbf{h}}_j^{<2>} \partial \hat{c}_j} = -\hat{\mathbf{B}}_j (1 - \hat{\mathbf{B}}_j) \left\{ (1 - 2\hat{\mathbf{B}}_j) \left(\frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \right)^{<2>} + \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j^{<2>}} \right\},$$

$$\text{where } \sum_{\langle 2 \rangle}^3 \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j^{\langle 2 \rangle}} \otimes \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} = \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j^{<2>}} \otimes \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} + \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \otimes \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j^{<2>}} + \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j @} \otimes \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \otimes \frac{@}{\partial \hat{\mathbf{h}}_j}$$

$$\text{with } \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j @} \otimes \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \otimes \frac{@}{\partial \hat{\mathbf{h}}_j} \equiv \left\{ \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \frac{1}{\partial \hat{\mathbf{h}}_j} \right\} \odot \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \quad \text{and } \odot$$

indicating the Hadamard product;

$$\frac{\partial^3 \hat{\pi}_i}{\partial \hat{\mathbf{d}}_j^{<2>} \otimes \partial \hat{\mathbf{d}}_k} = (-1)^{X_{ij}+X_{ik}} \int_{-\infty}^{\infty} \left(\frac{\partial^2 \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j^{<2>}} \otimes \frac{\partial \hat{\Psi}_k}{\partial \hat{\mathbf{d}}_k} \right) \hat{\Gamma}_{i(jk)} \phi(\theta) d\theta; \quad (\text{A.12})$$

$$\frac{\partial^3 \hat{\pi}_i}{\partial \hat{\mathbf{d}}_j \otimes \partial \hat{\mathbf{d}}_k \otimes \partial \hat{\mathbf{d}}_l} = (-1)^{X_{ij}+X_{ik}+X_{il}+1} \int_{-\infty}^{\infty} \left(\frac{\partial \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j} \otimes \frac{\partial \hat{\Psi}_k}{\partial \hat{\mathbf{d}}_k} \otimes \frac{\partial \hat{\Psi}_l}{\partial \hat{\mathbf{d}}_l} \right) \hat{\Gamma}_{i(jkl)} \phi(\theta) d\theta; \quad (\text{A.13})$$

where $\hat{\Gamma}_{i(jkl)}$ is defined similarly as $\hat{\Gamma}_{i(jk)}$ ($i = 1, \dots, K; j, k, l = 1, \dots, n; j \neq k, j \neq l, k \neq l$)

Finally, the fourth partial derivatives are given.

$$\frac{\partial^4 \hat{\pi}_i}{\partial \hat{\mathbf{d}}_j^{<4>}} = (-1)^{X_{ij}+1} \int_{-\infty}^{\infty} \frac{\partial^4 \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j^{<4>}} \hat{\Gamma}_{i(j)} \phi(\theta) d\theta \quad (\text{A.14})$$

with the nonzero partial derivatives being

$$\begin{aligned} \frac{\partial^4 \hat{\Psi}_j}{\partial \hat{\mathbf{h}}_j^{<4>}} &= (1 - \hat{c}_j) \hat{\mathbf{B}}_j (1 - \hat{\mathbf{B}}_j) \left\{ (1 - 14\hat{\mathbf{B}}_j + 36\hat{\mathbf{B}}_j^2 - 24\hat{\mathbf{B}}_j^3) \left(\frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \right)^{<4>} \right. \\ &\quad \left. + (1 - 6\hat{\mathbf{B}}_j + 6\hat{\mathbf{B}}_j^2) \sum \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j^{<2>}} \otimes \left(\frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \right)^{<2>} + (1 - 2\hat{\mathbf{B}}_j) \sum \left(\frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j^{<2>}} \right)^{<2>} \right\} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^4 \hat{\Psi}_j}{\partial \hat{\mathbf{h}}_j^{<3>} \partial \hat{c}_j} &= -\hat{\mathbf{B}}_j (1 - \hat{\mathbf{B}}_j) \left\{ (1 - 6\hat{\mathbf{B}}_j + 6\hat{\mathbf{B}}_j^2) \left(\frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \right)^{<3>} + (1 - 2\hat{\mathbf{B}}_j) \sum \frac{\partial^2 \hat{v}_j}{\partial \hat{\mathbf{h}}_j^{<2>}} \otimes \frac{\partial \hat{v}_j}{\partial \hat{\mathbf{h}}_j} \right\}; \\ \frac{\partial^4 \hat{\pi}_i}{\partial \hat{\mathbf{d}}_j^{<3>} \otimes \partial \hat{\mathbf{d}}_k} &= (-1)^{X_{ij}+X_{ik}} \int_{-\infty}^{\infty} \left(\frac{\partial^3 \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j^{<3>}} \otimes \frac{\partial \hat{\Psi}_k}{\partial \hat{\mathbf{d}}_k} \right) \hat{\Gamma}_{i(jk)} \phi(\theta) d\theta; \quad (\text{A.15}) \end{aligned}$$

$$\begin{aligned} &\frac{\partial^4 \hat{\pi}_i}{\partial \hat{\mathbf{d}}_j^{<2>} \otimes \partial \hat{\mathbf{d}}_k \otimes \partial \hat{\mathbf{d}}_l} \\ &= (-1)^{X_{ij}+X_{ik}+X_{il}+1} \int_{-\infty}^{\infty} \left(\frac{\partial^2 \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j^{<2>}} \otimes \frac{\partial \hat{\Psi}_k}{\partial \hat{\mathbf{d}}_k} \otimes \frac{\partial \hat{\Psi}_l}{\partial \hat{\mathbf{d}}_l} \right) \hat{\Gamma}_{i(jkl)} \phi(\theta) d\theta; \quad (\text{A.16}) \end{aligned}$$

$$\begin{aligned} &\frac{\partial^4 \hat{\pi}_i}{\partial \hat{\mathbf{d}}_j \otimes \partial \hat{\mathbf{d}}_k \otimes \partial \hat{\mathbf{d}}_l \otimes \partial \hat{\mathbf{d}}_m} \\ &= (-1)^{X_{ij}+X_{ik}+X_{il}+X_{im}} \int_{-\infty}^{\infty} \left(\frac{\partial \hat{\Psi}_j}{\partial \hat{\mathbf{d}}_j} \otimes \frac{\partial \hat{\Psi}_k}{\partial \hat{\mathbf{d}}_k} \otimes \frac{\partial \hat{\Psi}_l}{\partial \hat{\mathbf{d}}_l} \otimes \frac{\partial \hat{\Psi}_m}{\partial \hat{\mathbf{d}}_m} \right) \hat{\Gamma}_{i(jklm)} \phi(\theta) d\theta; \quad (\text{A.17}) \end{aligned}$$

with the similar definition of $\hat{\Gamma}_{i(jklm)}$

$(i = 1, \dots, K; j, k, l, m = 1, \dots, n, j \neq k, j \neq l, j \neq m, k \neq l, k \neq m, l \neq m)$.

References

- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- 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.
- Bock, R. D., & Moustaki, I. (2007). Item response theory in a general framework. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics, Vol. 26. Psychometrics* (pp.469-513). New York: Elsevier.
- Hall, P. (1992a). *The bootstrap and Edgeworth expansion*. New York: Springer. Corrected printing, 1997.
- Hall, P. (1992b). On the removal of skewness by transformation. *Journal of the Royal Statistical Society, B*, 54, 221-228.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, 49, 359-381
- Ogasawara, H. (2002). Stable response functions with unstable item parameter estimates. *Applied Psychological Measurement*, 26, 239-254
- Ogasawara, H. (2006). Asymptotic expansion of the sample correlation coefficient under nonnormality. *Computational Statistics and Data Analysis*, 50, 891-910.
- Ogasawara, H. (2007). Higher-order estimation error in structural equation modeling. *Economic Review, Otaru University of Commerce*, 57 (4), 131-160. <http://www.res.otaru-uc.ac.jp/~hogasa/>
- Stuart, A., & Ord, J. K. (1994). *Kendall's advanced theory of statistics: Distribution theory*

(6th ed., Vol.1). London: Arnold.

Thissen, D., & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika*, *47*, 397-412.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58-79.

Table 1. Simulated and asymptotic cumulants of the parameter estimators for LSAT Section 6 under a true model

	$\beta_2^{1/2}$ (dispersion)			β_1 (bias)			$\beta_3 / \beta_2^{3/2}$ (skewness)		
	Sim. (N)			Sim. (N)			Sim. (N)		
	(1000)	(2000)	Th.	(1000)	(2000)	Th.	(1000)	(2000)	Th.
a_1	5.4	5.1	4.8	12.7	8.0	7.8	23.2	15.9	11.9
a_3	5.1	4.5	4.2	25.1	22.9	18.5	48.2	32.7	25.0
a_5	4.0	3.9	3.7	4.0	3.2	4.6	14.6	11.5	9.8
b_1	52.7	35.0	27.3	-363	-308	-256	-238	-115	-54
b_3	3.4	3.2	3.1	-8	-7	-8	-27	-19	-18
b_5	52.6	34.1	26.3	-376	-305	-242	-285	-121	-52
	$(\beta_2')^{1/2}$ (dispersion)			β_1' (bias)			β_3' (skewness)		
a_1	.93	.97	1	-1.9	-2.4	-2.2	-8.2	-10.2	-11.0
a_3	.92	.96	1	-3.6	-3.2	-3.5	-17.7	-20.5	-23.0
a_5	.95	.99	1	-2.3	-2.5	-1.9	-7.4	-8.6	-9.2
b_1	1.06	1.03	1	8.9	8.5	8.5	58.1	57.5	53.2
b_3	.97	.98	1	4.2	4.1	3.6	16.2	17.6	18.2
b_5	1.04	1.03	1	7.9	8.1	8.2	51.8	56.1	52.0
HASE(SD)/ASE for non-Studentized estimators (N)									
	β_4 / β_2^2 (kurtosis)			(1000)		(2000)		Population parameters	
				Sim.	Th.	Sim.	Th.		
a_1	2010	979	557	1.11	1.07	1.05	1.04	.488	
a_3	6119	3076	1593	1.23	1.14	1.09	1.07	.521	
a_5	890	622	397	1.07	1.06	1.04	1.03	.387	
b_1	1.2×10^5	4.9×10^4	5870	1.93	1.31	1.28	1.17	-3.35	
b_3	2186	959	876	1.09	1.07	1.02	1.04	-.28	
b_5	1.7×10^5	5.3×10^4	5571	2.00	1.30	1.29	1.16	-3.12	

Note. Th.=Theoretical or asymptotic values, Sim.=Simulated values,
HASE= $\{(\beta_2 / N) + (\Delta\beta_2 / N^2)\}^{1/2}$, ASE= $(\beta_2 / N)^{1/2}$, SD=Standard deviations from simulations.

Table 2. Simulated and asymptotic cumulants of the parameter estimators for LSAT Section 7 under a true model

	$\beta_2^{1/2}$ (dispersion)			β_1 (bias)			$\beta_3 / \beta_2^{3/2}$ (skewness)		
	Sim. (N)			Sim. (N)			Sim. (N)		
	(1000)	(2000)	Th.	(1000)	(2000)	Th.	(1000)	(2000)	Th.
a_1	3.3	3.2	3.1	7.3	7.4	5.4	11.0	11.3	10.0
a_3	6.9	6.2	5.8	33.6	27.5	27.0	46.6	34.3	27.0
a_5	2.8	2.8	2.8	3.4	3.0	3.0	5.5	7.5	6.3
b_1	8.8	8.4	8.0	-33.1	-29.8	-34.8	-29.9	-27.0	-25.6
b_3	3.8	3.6	3.6	-8.8	-9.7	-8.4	-17.6	-15.3	-14.6
b_5	16.7	15.1	13.9	-92.7	-88.5	-81.8	-52.1	-38.4	-35.0
	$(\beta_2')^{1/2}$ (dispersion)			β_1' (bias)			β_3' (skewness)		
a_1	.98	.99	1	-1.0	-.9	-1.5	-8.1	-8.0	-9.0
a_3	.97	.98	1	-4.1	-4.3	-4.0	-22.6	-23.5	-24.8
a_5	.98	.99	1	-.8	-1.0	-.9	-6.1	-4.6	-5.8
b_1	1.01	1.01	1	4.7	4.9	4.1	25.7	24.4	24.9
b_3	.99	.99	1	2.6	2.2	2.5	13.0	13.6	14.6
b_5	1.02	1.01	1	5.8	5.6	5.6	31.2	32.5	34.0
HASE(SD)/ASE for non-Studentized estimators (N)									
	β_4 / β_2^2 (kurtosis)			(1000)		(2000)		Population parameters	
				Sim.	Th.	Sim.	Th.		
a_1	265	309	301	1.04	1.03	1.02	1.02	.580	
a_3	5859	3171	1675	1.19	1.11	1.06	1.06	.999	
a_5	49	35	168	1.02	1.02	1.02	1.01	.433	
b_1	1856	1325	1366	1.10	1.08	1.05	1.04	-1.88	
b_3	687	452	487	1.05	1.04	1.01	1.02	-1.06	
b_5	6359	2838	2491	1.20	1.14	1.08	1.07	-2.52	

Note. Th.=Theoretical or asymptotic values, Sim.=Simulated values, HASE= $\{(\beta_2 / N) + (\Delta\beta_2 / N^2)\}^{1/2}$, ASE= $(\beta_2 / N)^{1/2}$, SD=Standard deviations from simulations.

Table 3. Simulated and asymptotic cumulants of the parameter estimators for LSAT Section 7 with model misspecification

	$\beta_2^{1/2}$ (dispersion)			β_1 (bias)			$\beta_3 / \beta_2^{3/2}$ (skewness)		
	Sim. (N)			Sim. (N)			Sim. (N)		
	(1000)	(2000)	Th.	(1000)	(2000)	Th.	(1000)	(2000)	Th.
a_1	3.7	3.6	3.5	10.1	10.3	8.7	12.5	12.1	12.1
a_3	7.2	6.4	6.1	31.9	24.5	24.7	49.7	30.5	26.3
a_5	3.0	2.9	2.9	4.0	4.9	4.0	8.0	6.4	7.2
b_1	9.9	9.5	8.9	-39.5	-36.4	-39.1	-34.4	-31.4	-28.2
b_3	3.9	3.8	3.7	-11.7	-13.2	-11.4	-20.9	-17.9	-16.3
b_5	17.7	15.7	14.6	-98.6	-86.5	-86.9	-70.6	-42.0	-36.1
	$(\beta_2')^{1/2}$ (dispersion)			β_1' (bias)			β_3' (skewness)		
a_1	.98	1.00	1	-1.2	-1.1	-1.4	-9.6	-11.4	-11.4
a_3	.96	.97	1	-4.3	-4.6	-4.4	-21.1	-23.6	-24.4
a_5	.97	.98	1	-1.0	-.7	-1.0	-6.1	-7.2	-7.0
b_1	1.02	1.02	1	5.3	5.6	4.9	27.4	26.4	27.5
b_3	.97	.98	1	2.4	1.9	2.3	13.5	14.7	16.2
b_5	1.02	1.00	1	5.8	6.1	5.9	36.6	33.7	35.0
HASE(SD)/ASE for non-Studentized estimators (N)									
	β_4 / β_2^2 (kurtosis)			(1000)		(2000)			
				Sim.	Th.	Sim.	Th.		
a_1	237	305	403	1.04	1.04	1.03	1.02		
a_3	7231	2227	1663	1.18	1.11	1.05	1.06		
a_5	254	95	195	1.02	1.02	1.00	1.01		
b_1	2463	1812	1632	1.11	1.09	1.06	1.05		
b_3	1481	734	631	1.05	1.05	1.02	1.03		
b_5	16792	3692	2632	1.21	1.14	1.08	1.07		

Note. Th.=Theoretical or asymptotic values, Sim.=Simulated values, HASE= $\{(\beta_2 / N) + (\Delta\beta_2 / N^2)\}^{1/2}$, ASE= $(\beta_2 / N)^{1/2}$, SD=Standard deviations from simulations.

Table 4. $10^5 \times$ Root mean square errors of the asymptotic distribution functions of the standardized estimators in LSAT data

	$N=1000$				$N=2000$			
	N*	E1	E2	Hall	N*	E1	E2	Hall
Section 6, True model								
a_1	1395	782	391	748	750	412	234	381
a_3	2702	1222	555	975	1852	627	288	547
a_5	884	605	309	595	624	506	328	510
b_1	4871	2248	1406	2080	3578	1327	611	1005
b_3	1436	580	374	438	883	271	262	210
b_5	5244	2441	1315	1913	3698	1396	628	1122
Section 7, True model								
a_1	1209	494	351	505	856	334	275	333
a_3	2583	966	436	757	1614	414	264	233
a_5	652	312	181	318	492	252	188	255
b_1	2065	670	417	607	1501	507	421	519
b_3	1351	391	256	318	1032	266	254	238
b_5	3113	1033	487	943	2215	539	225	477
Section 7, Misspecified model								
a_1	1462	441	246	444	1026	356	249	353
a_3	2458	889	407	703	1544	447	300	237
a_5	711	162	157	151	583	230	243	225
b_1	2214	673	372	561	1677	509	371	509
b_3	1553	340	217	326	1246	282	220	222
b_5	3059	1076	566	766	2126	569	305	397

Note. N*=Normal approximation, E1=The single-term Edgeworth expansion, E2=The two-term Edgeworth expansion, Hall=Hall's method by variable transformation.

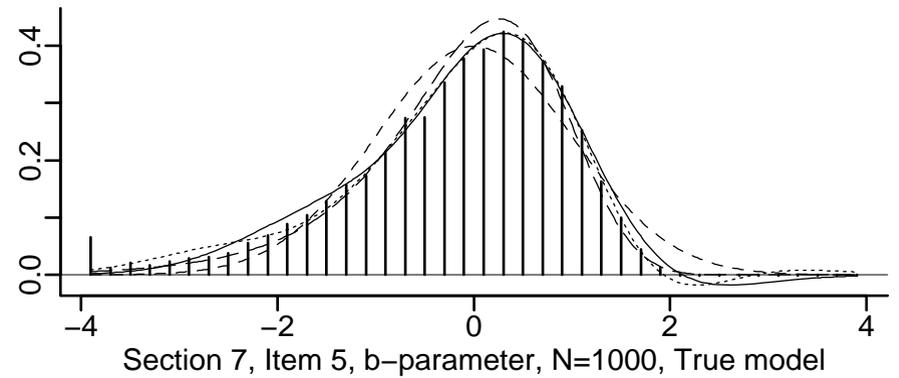
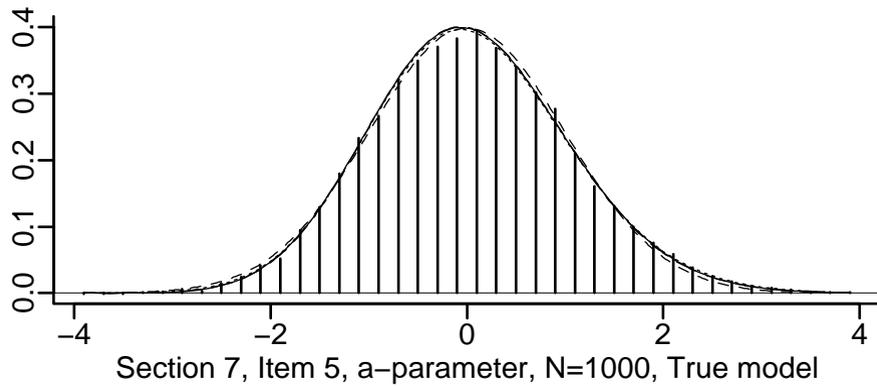
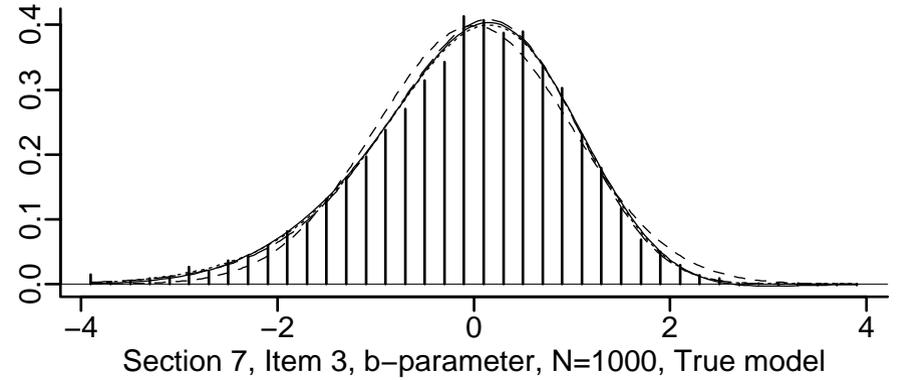
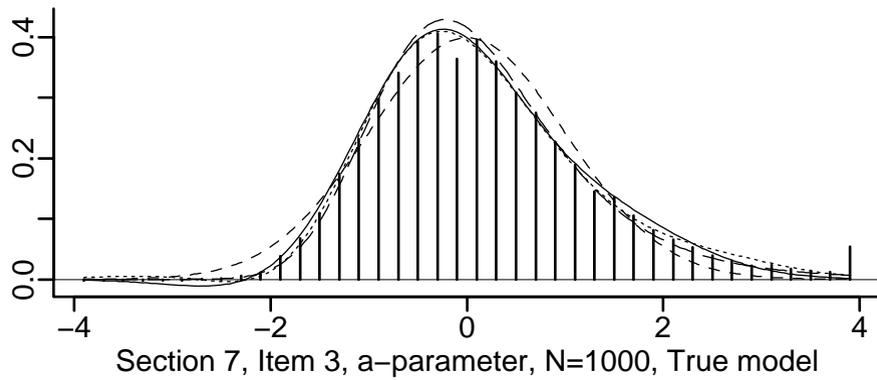
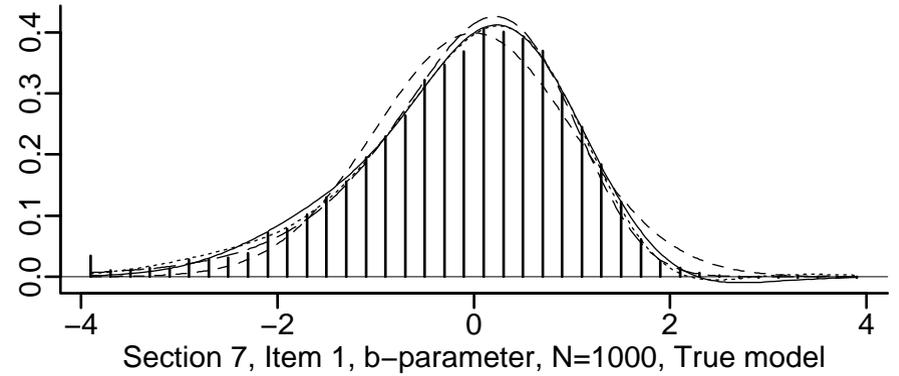
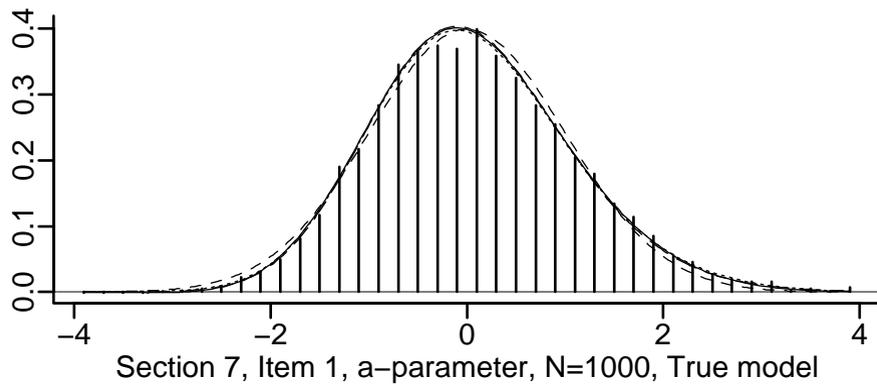


Figure 1. Theoretical (curved lines) and simulated (histograms) distributions of the standardized estimators in LSAT data (dashed lines=the standard normal distribution, solid lines=the single-term Edgeworth expansion, dotted lines=the two-term Edgeworth expansion, long dashed lines=Hall's variable transformation).

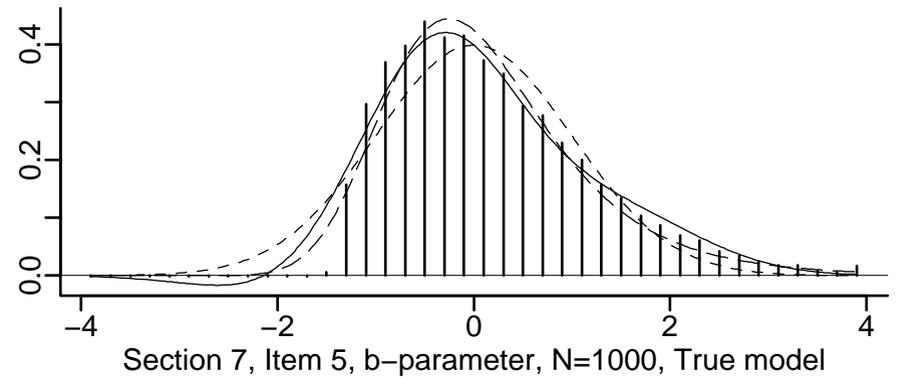
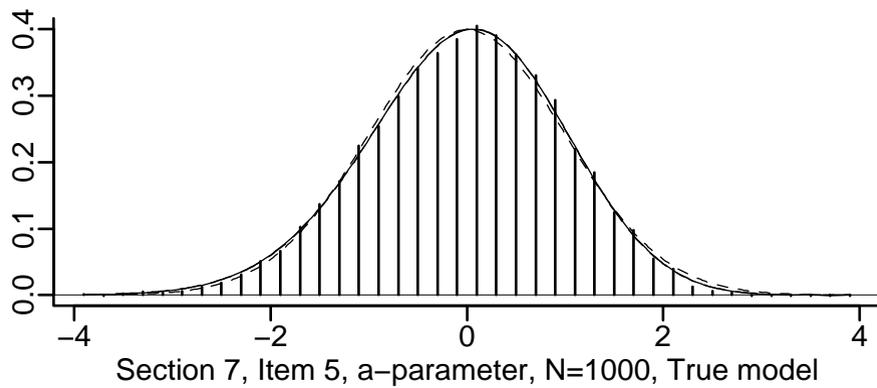
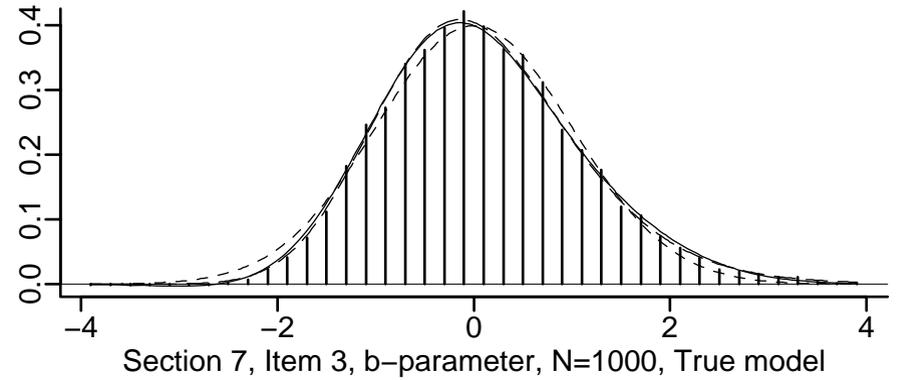
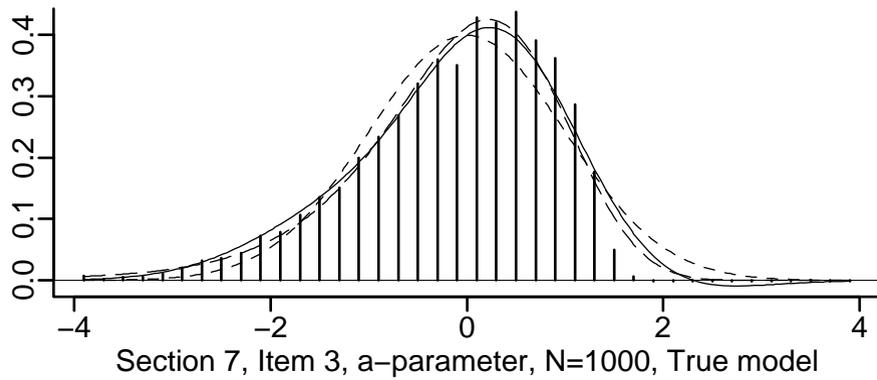
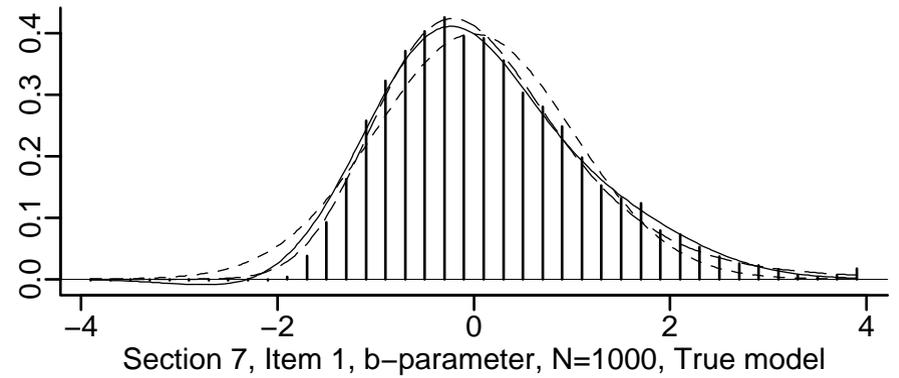
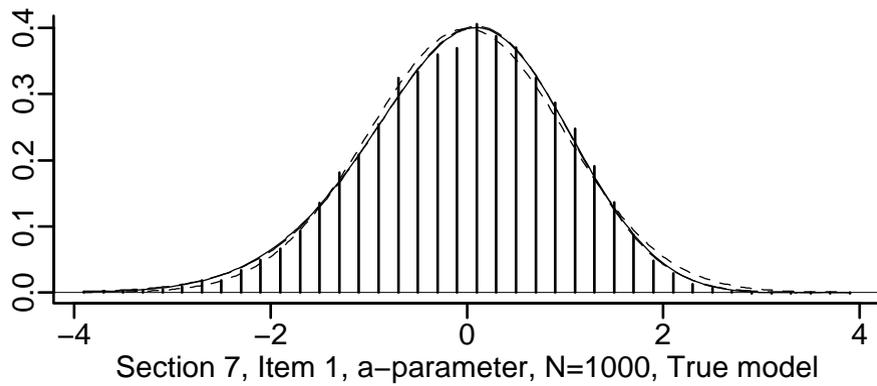


Figure 2. Theoretical (curved lines) and simulated (histograms) distributions of the Studentized estimators in LSAT data (dashed lines=the standard normal distribution, solid lines=the single-term Edgeworth expansion, long dashed lines=Hall's variable transformation).