Statistics and Its Interface Volume 9 (2016) 223–232

Likelihood ratio tests in the Rasch model for item response data when the number of persons and items goes to infinity

Ting Yan*†, Zhaohai Li , Yuanzhang Li, and Hong Qin

When the number of persons and items goes to infinity simultaneously, the maximum likelihood estimator in the Rasch model for dichotomous item response data has been shown to be consistency and asymptotic normality. However, the limiting distributions of the likelihood ratio tests in the past thirty years are still unknown. In this paper, we establish the Wilks type of results for the likelihood ratio tests under some simple and composite null hypotheses. Our proof crucially depends on the approximated inverse of the Fisher information matrix with small approximation errors. Simulation studies are provided to illustrate the asymptotic results.

AMS 2000 subject classifications: Primary 62F05.
Keywords and phrases: Fisher information matrix, Likelihood ratio tests, Rasch model, Wilks type of results.

1. INTRODUCTION

Assume that there are r persons and t items engaged in dichotomous item response experiments. Rasch (1960) suggested that the probability of the correct response between person i and item j is specified by:

$$P(i \text{ correctly answers } j) = \frac{e^{\alpha_i - \nu_j}}{1 + e^{\alpha_i - \nu_j}}, \tag{1}$$

where α_i measures the ability of person i and ν_j measures the difficulty of item j.

The Rasch model plays an important role in the development of item response theory and has been extensively studied (e.g. Fischer (1974, 1978); Hambleton, et al. (1978); Rasch (1960, 1961, 1966); Andersen (1973); Lauritzen (2003)). The book by Bond and Fox (2007) contains the detailed theoretical and applied studies on this model. A more recent overview is given by Wright and Mok (2004). Fischer (1974, pages 261-263; 1981) derived the necessary and sufficient condition for the existence and uniqueness of the maximum likelihood estimate (MLE). Haberman (1977) independently discovered this condition by using the general exponential family theory developed by Berk (1972) and Barndorff-Nielsen (1978).

In the Rasch model, the total number of parameters is equal to the sum of the numbers of the persons and items. Ghosh (1995) proved that if r or t is fixed, then the MLE is not consistent. Therefore, an interesting asymptotic background is the case that r and t go to infinity simultaneously. In this asymptotic framework, Haberman (1977) proved that the MLE is uniformly consistent and asymptotically normal by assuming that $\{\alpha_i : i = 1, \ldots, r\}$ and $\{\nu_j : j = 1, \ldots, t\}$ are bounded by a constant. However, there is little known about the asymptotic behaviors of the likelihood ratio tests (LRTs) in the past thirty years.

In this paper, we will prove that the LRTs are asymptotically normal independent of nuisance parameters under some simple and composite null hypotheses when $r \to \infty$ and $t \to \infty$ in the sense that

$$\frac{\Lambda - p}{\sqrt{2p}} \to N(0,1), \quad p \to \infty,$$

where Λ is the likelihood ratio and p is the degree of freedom, which is called the Wilks type of results (a notation coined by Fan, Zhang and Zhang (2001)). For the simple null, p equals to $r + t - 1$ (i.e., the total number of free parameters). For the composite null testing the homogeneity of a set of parameters with size m, $p = m$ and the remaining $r + t - m$ parameters are nuisance. Two technical steps are important for deriving the asymptotical distribution. First, the Fisher information matrix is approximated by a simple matrix with small approximation errors. Second, the uniformly upper bound of the errors between the MLEs and their true values is established by using the Erdős-Galli graph condition. The results can be used to test whether the parameters of a set with a large dimension are equal.

The remainder of this paper is organized as follows. The main results are given in Section 2. Numerical studies are presented in Section 3. We make some discussion in Section 4. The proofs of the theorems are relegated to Section 5.

2. MAIN RESULTS

As noted by Fischer (1981), the Rasch model can be considered as the Bradley-Terry model for incomplete paired

*Corresponding author.
†Yan's research was partially supported by the National Natural Science Foundation of China (No. 11401239) and Postdoctoral Science Foundation of China (No. 2014M552064).

comparisons. In what follows, we will discuss the Rasch model under this framework. Persons and items are treated as subjects engaged in paired comparisons. Persons $1, \dots, r$ will be labeled as subjects $1, \dots, r$ and items $1, \dots, t$ as subjects $r+1, \dots, r+t$. Therefore, there are the total $r+t$ subjects. Let n_{ij} be an indictor for pair (i,j). If there is a response between i and j, denote $n_{ij}=1$; otherwise, define $n_{ij}=0$. Thus, $n_{ij}=1$ for $i=1,\dots,r$; $j=r+1,\dots,r+t$ and 0 for other i,j. Let a_{ij} be an indictor whether i wins j. If subject i correctly answers subject j, we say "subject i beats subject j" and denote $a_{ij}=1$, $a_{ji}=0$; If subject i gives the wrong response to subject j, we say "subject i loses to subject j" and denote $a_{ij}=0$, $a_{ji}=1$; If $n_{ij}=0$, denote $a_{ij}=a_{ji}=0$. The item response outcomes can be summarized by a matrix $A=(a_{ij})_{i,j=1,\dots,r+t}$.

Let $\boldsymbol{\beta}=(\beta_1,\dots,\beta_{r+t})^\top=(\alpha_1,\dots,\alpha_r,\nu_1,\dots,\nu_t)^\top$ be the vector of the merits parameters for subjects $1,\dots,r+t$ and $\mathbf{u}=(u_1,\dots,u_{r+t})^\top$, where $u_i=e^{\beta_i}$. Since the probability (1) doesn't change by adding a constant to $\boldsymbol{\beta}$, we set $\beta_1=0(u_1=1)$ for parameter identification. When it is convenient, we interchangeably use the notation β_i and u_i. The log-likelihood for the Rasch model can be represented as

(2)
$$\begin{aligned}\ell(\boldsymbol{\beta}_*) &= \sum_{1\le i<j\le r+t}[\beta_i a_{ij}+\beta_j a_{ji}-n_{ij}\log(e^{\beta_i}+e^{\beta_j})]\\ &= \sum_{i=1}^{r+t}\beta_i a_i-\sum_{1\le i<j\le r+t}n_{ij}\log(e^{\beta_i}+e^{\beta_j}).\end{aligned}$$

where $a_i=\sum_{j=1}^{r+t}a_{ij}$ and $\boldsymbol{\beta}_*=(\beta_2,\dots,\beta_{r+t})^\top$. Notice that the vector $(a_2,\dots,a_{r+t})$ is a sufficient statistic. The likelihood equations are

$$(3)\qquad a_i=\sum_{j=1}^{r+t}\frac{n_{ij}e^{\hat\beta_i}}{e^{\hat\beta_i}+e^{\hat\beta_j}}=\sum_{j=1}^{r+t}\frac{n_{ij}\hat u_i}{\hat u_i+\hat u_j},\quad i=2,\dots,r+t.$$

where $\hat{\boldsymbol{\beta}}_*=(\hat\beta_2,\dots,\hat\beta_{r+t})^\top$ is the MLE of $\boldsymbol{\beta}_*$, $\hat\beta_1=\beta_1=0$, $\hat u_i=e^{\hat\beta_i}$ and $\hat u_1=u_1=1$. Let $V=(v_{ij})_{i,j=2,\dots,r+t}$ denote the covariance matrix of $a_i, i=2,\dots,r+t$, where

$$\begin{aligned} v_{ii} &= \textstyle\sum_{j=1}^{r+t}\frac{n_{ij}u_iu_j}{(u_i+u_j)^2}, \quad i=1,\dots,r+t\\ v_{ij} &= -\frac{n_{ij}u_iu_j}{(u_i+u_j)^2}, \qquad i,j=1,\dots,r+t; i\ne j.\end{aligned}$$

Note that the dimension of V is $(r+t-1)\times(r+t-1)$. For notational convenience, we suppress the subscript $r+t-1$. V is also the Fisher information matrix of the parameters $\beta_i, i=2,\dots,r+t$. Moreover, some notations are defined in the following:

$$\begin{aligned} &M_{rt}=\max_{1\le i,j\le r+t}u_i/u_j,\quad \lambda_{rt}=\max\{r,t\},\quad \rho_{rt}=\min\{r,t\},\\ &n_i=\textstyle\sum_{j=1}^{r+t}n_{ij}, \Omega_1=\{1,\dots,r\},\quad \Omega_2=\{r+1,\dots,r+t\}.\end{aligned}$$

In order to guarantee the existence and uniqueness of the MLE for (2), the following condition is necessary and sufficient due to Haberman (1977) and Fischer (1981).

Condition A. All cases of outcome matrices $A=(a_{ij})$ are included except for those in which there exist sets Γ_1, Γ_2, Γ_3 and Γ_4 satisfying:

1. $\Gamma_1\bigcup\Gamma_2=\Omega_1$ and $\Gamma_3\bigcup\Gamma_4=\Omega_2$ with $\Gamma_1\bigcap\Gamma_2=\emptyset$ and $\Gamma_3\bigcap\Gamma_4=\emptyset$;
2. $\Gamma_1\ne\emptyset$ and $\Gamma_3\ne\emptyset$, or $\Gamma_2\ne\emptyset$ and $\Gamma_4\ne\emptyset$;
3. $a_{ij}=0$ for all $i\in\Gamma_1$ and $j\in\Gamma_3$;
4. $a_{ij}=1$ for all $i\in\Gamma_2$ and $j\in\Gamma_4$.

The second condition in the above Condition A includes three cases, in which the MLE does not exist: (1) $\Gamma_1=\Omega_1,\Gamma_3=\Omega_2$ ($\Gamma_2=\emptyset,\Gamma_4=\emptyset$). In this case, it says that all the persons have not given correct responses to all the items. (2) $\Gamma_2=\Omega_1,\Gamma_4=\Omega_2$ ($\Gamma_1=\emptyset,\Gamma_3=\emptyset$). In this case, it says that all the persons have given correct responses to all the items. (3) $\Gamma_1\ne\emptyset,\Gamma_2\ne\emptyset,\Gamma_3\ne\emptyset,\Gamma_4\ne\emptyset$. In this case, it says that for some partition of the items into two nonempty subsets Γ_1, Γ_2 and of the persons into two nonempty subsets Γ_3, Γ_4, all the persons in Γ_1 have not given correct responses to all the items in Γ_3 and all the persons in Γ_2 have given correct responses to all the items in Γ_4. If all the persons have large merits and all the items have relatively small merits (corresponding a large M_{rt}), then all the persons most probably correctly answer all the items such that case (1) occurs. There are similar discussions for the other two cases. On the other hand, the probability that Condition A fails depends on the size of r and t. Therefore, controlling the increasing rate of M_{rt} is necessary in order to guarantee Condition A. To establish the Wilks type of results for the Rasch model when t and r go to infinity simultaneously, we need the following three propositions.

Proposition 1. *If $M_{rt}=o(\rho_{rt}/\log\rho_{rt})$ and $\lambda_{rt}/\rho_{rt}\to c$ as r and t go to infinity, where c is a constant, then P(Condition A holds) $\to 1$.*

Proposition 2. *Let $S=(s_{ij})_{i,j=2,\dots,r+t}$ be the matrix with*

$$s_{ij}=\frac{\delta_{ij}}{v_{ii}}+\frac{1}{v_{11}}$$

where δ_{ij} is the Kroneck delta function. The upper bound of the approximation error using S to approximate the inverse of V takes:

$$(4)\qquad \|V^{-1}-S\|\le O\left(\frac{M_{rt}^4}{\rho_{rt}^2}\right)\quad as\quad \rho_{rt}\to\infty,$$

where $\|A\|=\max_{i,j}|a_{ij}|$ for a general matrix.

Proposition 3. *Let $\Delta u_i=(\hat u_i/u_i)-1$. If $M_{rt}=o(\log\rho_{rt})$ and $\lambda_{rt}/\rho_{rt}\to c$ as r and t go to infinity, then with probability approaching 1,*

(5)
$$\max_{i=1,\dots,r+t}|\Delta u_i|\le\max_{i,j=1,\dots,r+t}|\Delta u_i-\Delta u_j|\le\frac{\theta_{rt}}{1-\theta_{rt}}\to 0,$$

where

(6)

$$\theta_{rt} = \left[\frac{2(1+M_{rt})^2}{M_{rt}} + \frac{e^{c^* M_{rt}}(1+M_{rt})^4}{4M_{rt}^2}\right]\sqrt{\frac{\lambda_{rt}\log\lambda_{rt}}{\rho_{rt}^2}},$$

and c^* *is a constant.*

The proofs of Propositions 1–3 are given in the supplementary materials (http://www.intlpress.com/SII/p/2016/9-2/SII-9-2-YAN-supplement.pdf). Proposition 1 asserts that the MLE exists with probability approaching one as $r\to\infty$ and $t\to\infty$. This is a necessary condition for any desired asymptotic properties of the MLE. To derive the asymptotic distribution of the log-likelihood ratio test, one may apply the Taylor's expansion to $\ell(\hat{\boldsymbol{\beta}}_*)$ at the point of the true parameter $\boldsymbol{\beta}_*$. The first-order item is $(\mathbf{a}-E(\mathbf{a}))^\top(\hat{\boldsymbol{\beta}}_*-\boldsymbol{\beta}_*)$ and the second-order item is $\frac{1}{2}(\hat{\boldsymbol{\beta}}_*-\boldsymbol{\beta}_*)^\top V(\hat{\boldsymbol{\beta}}_*-\boldsymbol{\beta}_*)$, where $\mathbf{a}=(a_2,\ldots,a_{r+t})^\top$. It naturally requires establishing the relationship between $\hat{\boldsymbol{\beta}}_*$ and $\mathbf{a}$. Specifically, an approximate explicit expression of $\hat{\boldsymbol{\beta}}_*$ that depends on $\mathbf{a}$, is required. To this end, one needs to obtain the inverse of V. Since the inverse of V doesn't have a close form, we use a simple matrix to approximate it. This is done in Proposition 2, which provides a high accurate approximate errors. When M_{rt} is a constant, the upper bound of the errors in (4) is in the magnitude of ρ_{rt}^{-2} uniformly. Similar situations appears in Simons and Yao (1999) and Yan and Xu (2013) who used a simple matrix to approximate the inverse of the Fisher information matrix to prove the asymptotic normality of the MLE. Proposition 3 gives an upper bound of the errors between the MLE and its true value. In contrast with Haberman's (1977) result of consistency, we don't assume that M_{rt} is a constant here. A technical step in the proof of Proposition 3 uses the Erdős-Galli graph condition. The item response result can be represented by a directed bipartite graph, in which a correct answer indicates an edge from person to item and a wrong answer indicates an opposite direction. This technical skill is motivated by Chatterjee, Diaconis and Sly (2011) who proved the consistency in the β-model for undirected random graphs with the diverging number of nodes.

Now, we present the Wilks type of results for a simple null and a composite null in the following.

Theorem 1. *For a fixed* $\boldsymbol{\beta}$, *if* $M_{rt}=o(\log\rho_{rt})$, $\lambda_{rt}/\rho_{rt}\to c$ *and*

$$\sum_{i,j=1}^{r+t}\left|\frac{e^{\beta_i}-e^{\beta_j}}{e^{\beta_i}+e^{\beta_j}}\right| = o\left(\frac{\rho_{rt}^{3/2}}{(\log\rho_{rt})^{3/2}}\right), \quad (7)$$

then the log-likelihood ratio test $\ell(\hat{\boldsymbol{\beta}}_*)-\ell(\boldsymbol{\beta}_*)$ *is asymptotically normally distributed in the sense that*

$$\frac{2[\ell(\hat{\boldsymbol{\beta}}_*)-\ell(\boldsymbol{\beta}_*)]-(r+t-1)}{\sqrt{2(r+t-1)}} \xrightarrow{L} N(0,1). \quad (8)$$

Theorem 2. *Assume that* $m/\rho_{rt}\ge\tau>0$, *where* τ *is a positive constant. Under the null* $H_0:\beta_2=\cdots=\beta_m$, *if* $M_{rt}=o(\log\rho_{rt})$, $\lambda_{rt}/\rho_{rt}\to c$ *and* (7) *holds, then the log-likelihood ratio test* $\ell(\hat{\boldsymbol{\beta}}_*)-\ell(\hat{\boldsymbol{\beta}}_*^H)$ *is asymptotically normally distributed in the sense that*

$$\frac{2[\ell(\hat{\boldsymbol{\beta}}_*)-\ell(\hat{\boldsymbol{\beta}}_*^H)]-(m-1)}{\sqrt{2(m-1)}} \xrightarrow{L} N(0,1), \quad (9)$$

where $\hat{\boldsymbol{\beta}}_*^H=(\hat\beta_2^H,\ldots,\hat\beta_{r+t}^H)$ *is the MLE of* $\boldsymbol{\beta}_*$ *under the null* H_0 *and* $\hat\beta_1^H=0$.

The condition $M_{rt}=o(\log\rho_{rt})$ is to control the increasing rate of M_{rt}, and it is necessary in order to guarantee the existence of the MLE with high probability. The second condition $\lambda_{rt}/\rho_{rt}\to c$ requires that the number of persons compares with that of items. Condition (7) is technical, due to the control of the remainder in the Taylor expansion of the log-likelihood function. It essentially requires that a large set of parameters do not differ too much.

Remark 1. *We only consider a special index subset, i.e.,* $i\in\{2,\ldots,m\}$ *for the complex null in Theorem 2. Since the index labels are not essential, we may change the order of index labels for persons and items. Therefore, if we change the null* $H_0:\beta_2=\cdots=\beta_m$ *to* H_0: $\beta_i, i\in S$ *are equal for a subset* $S\subset\{1,\ldots,r+t\}$ *with size* $m-1$, *Theorem 2 still holds.*

3. NUMERICAL RESULTS

In this section, we demonstrate the theoretical results via numerical studies.

3.1 Simulation studies

We conduct simulations to evaluate Theorems 1 and 2. Theorem 1 is evaluated by using the quantile-quantile (QQ) plots while Theorem 2 is done by investigating the power of the statistic (9). The parameters were set to be $u_i=[(M_{rt}-1)i/r]+1$ for $i=1,\ldots,r$; $u_i=[(M_{rt}-1)(i-r)/t]+1$ for $i=r+1,\ldots,r+t$. If the MLE doesn't exist, we define the LRTs to be zero. In practical applications, the number of persons is usually larger than that of items. Therefore, we only consider the results in the case $r>t$. We considered three cases for $M_{rt}=1,t^{1/2},t$ and a combination for $(r,t)=(50,30)$, and did 10,000 repetitions for each simulation. In the simulation results, Condition A failed with probability 0.053 when $M_{rt}=t$ and 0 for other M_{rt}. The QQ plots of the empirical distribution for the statistic (8) against the standard normal distribution are shown in Figure 1. Although the results for only a single combination of (r,t) are given here, we also tried other values and found that the phenomena are similar. Figure 1 shows that the empirical quantiles of the statistic (8) are very close to the quantiles of the standardized normal distribution when $M_{rt}=1,t^{1/2}$. On the other hand, when $M_{rt}=t$, Condition A failed with

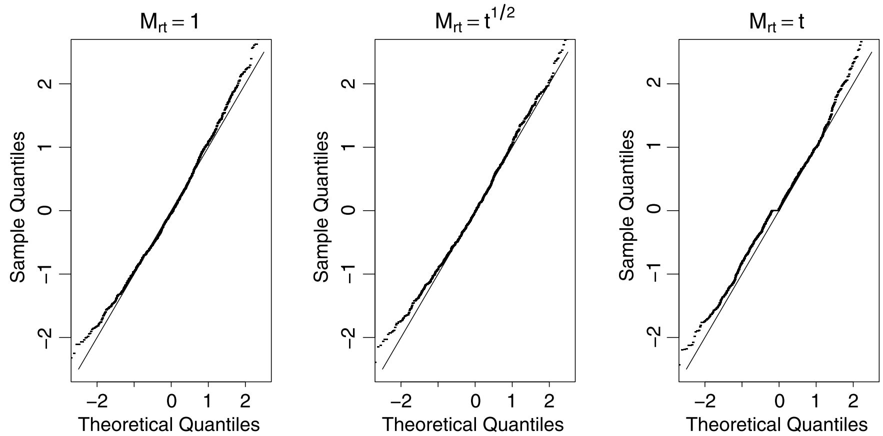

Figure 1. The QQ plots of the test (8) ($r = 50, t = 30$). The real lines are the reference line $y = x$.

Table 1. Powers of the test (9)

(r, t)	(m_1, m_2)	$\eta = 0$	$\eta = 0.3$	$\eta = 0.6$	$\eta = 0.9$	$\eta = 1.2$
(50, 30)	(15,0)	0.055	0.153	0.384	0.607	0.750
	(25,0)	0.053	0.217	0.541	0.810	0.920
	(0, 15)	0.047	0.229	0.587	0.8455	0.933
	(0, 25)	0.054	0.326	0.809	0.969	0.994
(100, 40)	(15,0)	0.054	0.185	0.539	0.754	0.869
	(25,0)	0.048	0.270	0.688	0.907	0.979
	(0, 15)	0.049	0.472	0.930	0.998	1
	(0, 25)	0.050	0.671	0.993	1	1

Table 2. The fitted parameters of 35 items for BLOT data

Item	**u**	Item	**u**	Item	**u**	Item	**u**
6	0.013	12	0.026	22	0.053	5	0.057
27	0.062	20	0.067	1	0.072	2	0.077
14	0.077	7	0.082	33	0.093	29	0.098
34	0.104	16	0.116	35	0.116	10	0.129
18	0.150	4	0.158	9	0.190	11	0.190
31	0.190	24	0.199	23	0.226	17	0.236
19	0.256	25	0.267	3	0.339	26	0.352
8	0.380	15	0.441	13	0.441	30	0.475
32	0.512	28	0.851	21	1.747		

frequency 0.053. These simulation results indicate that it is necessary to control the increasing rate of M_{rt} in order to guarantee Condition A. However, it also implies that the condition on M_{rt} in Theorem 1 is strict and could be loosened.

Next, we simulated powers of the test statistic (9). We considered the null $H_0 : u_1 = \cdots = u_{m_1} = 1, u_{r+1} = \cdots = u_{r+m_2} = 1$ and the alternative $H_1 : u_i = (i\eta M_{rt}/m_1) + 1$, $i = 1, \ldots, m_1$, $u_i = [(i-r)\eta M_{rt}/m_2]+1, i = r+1, \ldots, r+t$, where η controls the strength of the deviation from the null hypothesis. The other redundant parameters were set to be $u_i = (M_{rt} - 1)i/r$ for $i = m_1 + 1, \ldots, r$; $u_i = (M_{rt} - 1)(i - r)/t$ for $i = m_2 + 1, \ldots, r + t$. In this simulation, we let $M_{rt} = 3$. The simulated powers were put in Table 1, in which Condition A held with 100% frequencies. The simulated type I errors look very good. The powers become bigger as m_1 or m_2 increases when r and t are fixed. When $\eta \geq 0.6$, the simulated powers are very high for $m_1 \geq 25$ or $m_2 \geq 25$, which exceed 80% in many cases.

3.2 A data example

We use the data set for Bond's Logical Operations Test as an illustrated example, which is available from the web http://www.personality-project.org/r/html/blot.html. This data set was collected by Trevor G. Bond and is used as an example of Rasch modeling by Bond and Fox (2007, p. 56). It contains 150 persons and 35 items. In this data set, $r = 150$, $t = 35$ and $t/r = 0.23$. In order to guarantee Condition A, individuals 23, 27 and 46 were deleted before the analysis, who gave the correct responses for all the items. The fitted parameters of the 35 items are given in Table 2.

It may be of interest to test whether there is significant difference among 35 items. As given in Remark 1, we can use Theorem 2 to perform hypothesis testing. Under this null, the value of the test (9) is 23.25 and the corresponding p-value is 1.46×10^{-119}, indicating a very significant difference. To test the homogeneity for a subset of items, we use the null that the merits of the 26 items 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 14, 16, 17, 18, 19, 20, 22, 23, 24, 25, 27, 29, 31, 33, 34, 35, are equal as an example, in which the number of persons giving correct responses are not less than 100. Under this null, the value of the test (9) is 8.66 and the corresponding p-value is 4.86×10^{-18}, indicating a very significant difference as well. Next, we use the proposed test statistic to test whether the parameters for a subset of persons are equal. We choose the set containing the 38 persons 36, 44, 51, 52, 66, 68, 72, 75, 82, 83, 85, 86, 90, 91, 92, 94, 98, 99, 100, 101, 102, 103, 105, 106, 107, 108, 109, 111, 117, 120, 121, 124, 127, 131, 136, 142, 145, 149 as an example, who gave $20 \sim 25$ correct responses. Under this null, the value of the test (9) is -3.21 and the corresponding p-value is 1.32×10^{-3}, indicating a significant difference as well.

4. DISCUSSION

When the numbers of items and persons go to infinity simultaneously, we have derived Wilks type of theorems for the LRTs in the Rasch model. Theorems 1 and 2 can be used to construct the confidence interval of parameters and test the equality of a large number of parameters. For example, an approximate $1-\alpha$ confidence interval of $\bar{\boldsymbol{\beta}}_*$ is

$$\{\boldsymbol{\beta}_* : |[\ell(\boldsymbol{\beta}_*) - \ell(\bar{\boldsymbol{\beta}}_*) - (r+t-1)]/\sqrt{2(r+t-1)}| \le z_{1-\alpha/2},$$

where z_α is the α-quantile of the standard normal distribution. The condition on M_{rt} in Theorems 1 and 2 requires $M_{rt} = o(\log \rho_{rt})$. Moreover, another condition is imposed on the pairwise differences of parameters in (7). Simulation studies shed light that there are still good approximations for the likelihood ratio tests in Theorems 1 and 2 even when $M_{rt} = \rho_{rt}^{1/2}$. However, it should be noted that the asymptotic behavior of the LRTs depends not only on M_{rt}, but also on the configuration of all the parameters. It would be of interest to see if the conditions in Theorems 1 and 2 can be relaxed.

We only consider the dichotomous response data in this paper. The polytomous response data may be encountered in practice when the items are rating scales, for which successively higher integer scores are coded to indicate increasing levels of competence such as "Strongly Disagree" labelled as 0, "Disagree" as 1, "Agree" as 2 and "Strongly Agree" as 3. Andrich (1978) proposed the polytomous Rasch model to allow the multiple responses. Masters (1982) independently discovered it with a different name called "partial credit model". This generalized model also assigns a potential parameter to each subject and has more complex probability structures than the dichotomous case. However, there are still a lack of asymptotical theories for the polytomous Rasch model. Anderson, Li and Vermunt (2007) provided some simulations that shed light on the MLEs that are consistent when the number of persons and items are large enough. Establishing asymptotical properties of the MLEs as well as the Wilks type of results for the polytomous Rasch models for high dimensional situations is still an open question.

5. PROOFS OF THEOREMS

The following lemma plays an important role in the proof of Theorem 1.

Lemma 1. *(1) If $M_{rt} = o(\rho_{rt}^{1/8})$, then $\sum_{i=1}^{r+t}(a_i - E(a_i))^2/v_{ii}$ is asymptotically normal with mean $r+t$ and variance $2(r+t)$.*

(2) If $M_{rt} = o(\rho_{rt}^{1/8})$, then $(\mathbf{a} - E(\mathbf{a}))^\top V^{-1}(\mathbf{a} - E(\mathbf{a}))$ is asymptotically normal with mean $r+t$ and variance $2(r+t)$, where $\mathbf{a} = (a_2, \ldots, a_{r+t})$.

Proof. Let $\tilde{v}_{ij} = |v_{ij}|$. If $i \neq j$, then $\tilde{v}_{ij} = -v_{ij} = nu_iu_j/(u_i+u_j)^2$. For convenience, denote

$$x_{ij} = a_{ij} - E(a_{ij}),\ j \neq i, \quad x_{ii} = 0,$$

and define $q_{rt} = \min_{i,j;i\neq j} \tilde{v}_{ij} \ge M_{rt}/(1+M_{rt})^2$, $Q_{rt} = \max_{i,j;i\neq j} \tilde{v}_{ij} \le 1/4$. If $M_{rt} = o(\rho_{rt}^{1/8})$, then $Q_{rt}/q_{rt} = o(t^{1/8})$.

Since $(a_i - E(a_i))^2, i = 1, \cdots, r$ is a sequence of independent random variables and $(a_i - E(a_i))^2, i = r+1, \cdots, r+t$ is also a sequence of independent random variables, in order to prove Lemma 1 (1), it is sufficient to show:
If $Q_{rt}/q_{rt} = o(t^{1/8})$, then the following hold:

(C1) $\sum_{i=1}^{r}(a_i - E(a_i))^2/v_{ii}$ is asymptotically normally distributed with mean r and variance $2r$.
(C2) $\sum_{i=r+1}^{r+t}(a_i - E(a_i))^2/v_{ii}$ is asymptotically normally distributed with mean t and variance $2t$.
(C3) $\sum_{i=1}^{r+t}(a_i - E(a_i))^2/v_{ii}$ is asymptotically normally distributed with mean $r+t$ and variance $2(r+t)$.

The proofs of C1 and C2 are similar. We only give the proof of C1 and omit the other. Let $z_i = [(a_i - E(a_i))^2 - E(a_i - E(a_i))^2]/v_{ii}$. By direct calculation,

(10)
$$\begin{aligned} & v_{ii}^2 E(z_i^2) \\ = {} & \textstyle\sum_{j=r+1}^{r+t}[E(x_{ij}^4) - (E(x_{ij}^2))^2] + 2\sum_{j,l=r+1;j\neq l}^{r+t} \tilde{v}_{ij}\tilde{v}_{il} \\ = {} & \textstyle\sum_{j=r+1}^{r+t} \tilde{v}_{ij}(1-2p_{ij})^2 + 2\sum_{j,l=r+1;j\neq l}^{r+t} \tilde{v}_{ij}\tilde{v}_{il}. \end{aligned}$$

Since $\{z_i\}_{i=1}^r$ is a sequence of independent random variables, to prove C1, it is sufficient to show $E(z_i^2) < \infty$ and for any give $\varepsilon > 0$, the Lindeberg-Feller condition [Lindeberg (1922); Feller (1945)]

$$\frac{1}{G_r^2}\sum_{i=1}^{r} E[z_i^2 I(|z_i| > \varepsilon G_r)] \to 0, \tag{11}$$

where $G_r^2 = \sum_{i=1}^r E(z_i^2)$.

Note that $v_{ii} = \sum_{j=r+1}^{r+t}\tilde{v}_{ij}$ for $i = 1, \ldots, r$. By (10), we have that $E(z_i^2) \le 2/v_{ii} + 2$ and

$$G_r^2 \ge \sum_{i=1}^{r} \frac{2\sum_{k,l=r+1;l\neq k}^{r=t}\tilde{v}_{ik}\tilde{v}_{il}}{v_{ii}^2} \ge \frac{2r(t-1)q_{rt}^2}{tQ_{rt}^2}. \tag{12}$$

Let $\mu > 1$ and $\nu > 1$ be two constants such that $(1/\mu) + (1/\nu) = 1$. Note that for $1 \le i \le r$, $a_i - E(a_i)$ is the sum of a sequence of independent random variables $x_{ij}, j = r+1, \ldots, r+t$ with mean zero and x_{ij} is a dichotomous random variables taking values $-p_{ij}$ and $1-p_{ij}$ with success probabilities $1-p_{ij}$ and p_{ij}. It is easy to show

$$E(x_{ij}^{4\mu}) \le p_{ij}^{4\mu}(1-p_{ij}) + (1-p_{ij})^{4\mu}p_{ij} \le 2\tilde{v}_{ij}.$$

By Rosenthal's (1970) inequality, we have

$$\begin{aligned} E(a_i - E(a_i))^{4\mu} & \le c_{4\mu}[(\sum_{j=r+1}^{r+t} E(x_{ij}^2))^{2\mu} + \sum_{j=r+1}^{r+t} E(x_{ij}^{4\mu})] \\ & \le c_{4\mu}(\sum_{j=r+1}^{r+t}\tilde{v}_{ij})^{2\mu} + 2c_{4\mu}\sum_{j=r+1}^{r+t}\tilde{v}_{ij}, \end{aligned}$$

where $c_{4\mu}$ is a constant depending only on 4μ. Consequently,

(13)
$$\frac{E(a_i-Ea_i)^{4\mu}}{v_{ii}^{2\mu}} \le c_{4\mu} + \frac{2c_{4\mu}tQ_{rt}}{(tq_{rt})^{2\mu}} \le c_{4\mu} + \frac{2c_{4\mu}(Q_{rt}/q_{rt})^{2\mu}}{t^{2\mu-1}}.$$

Thus, if $Q_{rt}/q_{rt} = o(\rho_{rt}^{1/8})$, then for sufficiently large ρ_{rt}, we have

(14)
$$(E(z_i^{2\mu}))^{1/\mu} \le \max\left\{(E[\frac{(a_i-Ea_i)^{4\mu}}{v_{ii}^{2\mu}}])^{1/\mu}, 1\right\} \le 1 + c_{4\mu}.$$

For any given $\varepsilon > 0$, by (12), if $Q_{rt}/q_{rt} = o(\rho_{rt}^{1/8})$, we can choose G_r such that $\varepsilon G_r > 1$ when ρ_{rt} is large enough. Note that $E(a_i - E(a_i))^2 = v_{ii}$ and

$$|z_i| > \varepsilon G_r \Leftrightarrow \frac{(a_i-E(a_i))^2}{v_{ii}} > \varepsilon G_r + 1 \text{ or } \frac{(a_i-E(a_i))^2}{v_{ii}} < 1 - \varepsilon G_r.$$

Therefore, by Hoeffding's (1963) inequality and noticing that $v_{ii} \ge tq_{rt}$ for $i = 1, \ldots, r$, we have

$$\begin{aligned} & P(|z_i| > \varepsilon G_t) \le P((a_i - Ea_i)^2 \ge \varepsilon v_{ii} G_r) \\ \le\ & 2\exp(-2\varepsilon v_{ii}G_r/t) \le 2\exp(-2\varepsilon q_{rt}G_r) \\ \le\ & 2\exp\left(-4\varepsilon r(t-1)q_{rt}^3/(tQ_{rt}^2)\right). \end{aligned}$$

Hölder's inequality gives [c.f. (12), (13)]

$$\begin{aligned} & E[z_i^2 I(|z_i| > \varepsilon G_t)] \\ \le\ & [E(z_i^{2\mu})]^{1/\mu}(P(|z_i| > \varepsilon G_t))^{1/\nu} \\ \le\ & 2(1+c_{4\mu})\exp[-4\varepsilon r(t-1)q_{rt}^2/(\nu t Q_{rt}^3)]. \qquad (15) \end{aligned}$$

By (15), we have

$$\begin{aligned} & \frac{1}{G_r^2}\sum_{i=1}^r E[z_i^2 I(|z_i| > \varepsilon G_t)] \\ \le\ & \frac{tQ_{rt}^2(1+c_{4\mu})}{(t-1)q_{rt}^2}\exp[-4\varepsilon r(t-1)q_{rt}^2/(\nu t Q_{rt}^3)]. \end{aligned}$$

Since $\mu > 1$ and $\nu > 1$ are constants, if $Q_{rt}/q_{rt} = o(t^{1/8})$, then the above expression goes to zero as $\rho_{rt} \to \infty$. This shows (11).

The variance of $\sum_{i=1}^{r+t}(a_i - E(a_i))^2/v_{ii}$ is the sum of the following two terms:

(a) $\sum_{i=1}^{r+t} Var[(a_i - E(a_i))^2/v_{ii}]$;
(b) $2\sum_{1\le i<j\le t} Cov(\frac{(a_i-E(a_i))^2}{v_{ii}}, \frac{(a_j-E(a_j))^2}{v_{jj}})$.

A direct calculation gives that

$$\begin{aligned} Var(a_i - E(a_i))^2 = & \sum_{j=1}^{r+t}[\tilde{v}_{ij}(p_{ij}^2 - p_{ij}p_{ji} + p_{ji}^2) \\ & + \tfrac{n_{ij}-1}{n_{ij}}\tilde{v}_{ij}^2 - 3\tilde{v}_{ij}^2]I(n_{ij} > 0) + 2v_{ii}^2. \end{aligned}$$

Consequently,

$$\begin{aligned} & |\frac{\sum_{i=1}^{r+t} Var(a_i - E(a_i))^2/v_{ii}^2 - 2}{r+t}| \\ \le\ & \frac{1}{r+t}\sum_{i,j=1}^{r+t}\frac{\tilde{v}_{ij} + 3\tilde{v}_{ij}^2}{v_{ii}^2} \le \frac{Q_{rt}}{q_{rt}} + \frac{3Q_{rt}^2}{q_{rt}^2}. \end{aligned}$$

Thus, if $Q_{rt}/q_{rt} = o(t^{1/2})$, we have

(16)
$$\frac{\sum_{i=1}^{r+t} Var(a_i - E(a_i))^2/v_{ii}}{r+t} = o(1) + 2.$$

Since

$$\begin{aligned} & |Cov((a_i - E(a_i))^2, (a_j - E(a_j))^2)| \\ =\ & |Cov(x_{ij}^2, x_{ji}^2)| \le 2\tilde{v}_{ij} + \tilde{v}_{ij}^2, \end{aligned}$$

we have

$$\begin{aligned} & \frac{1}{r+t}|\sum_{i\ne j} Cov(\frac{(a_i-E(a_i))^2}{v_{ii}}, \frac{(a_j-E(a_j))^2}{v_{jj}})| \\ \le\ & \frac{1}{r+t}\sum_{i\ne j}\frac{2\tilde{v}_{ij} + \tilde{v}_{ij}^2}{v_{ii}v_{jj}} \le \frac{3Q_{rt}}{q_{rt}^2 t}. \qquad (17) \end{aligned}$$

By (16) and (17), if $Q_{rt}/q_{rt} = o(t^{1/2})$, we have that

$$\frac{Var(\sum_{i=1}^{r+t}(a_i - E(a_i))^2/v_{ii})}{r+t} = o(1) + 2.$$

In view of C1 and C2, by Slutsky's theorem, we have C3. This completes the proof of Lemma 1 (1).

Note that $V^{-1} = W + S$ and

$$E[(\mathbf{a} - E(\mathbf{a}))^\top S(\mathbf{a} - E(\mathbf{a}))] = \sum_{i=1}^{r+t}\frac{(a_i - E(a_i))^2}{v_{ii}}.$$

Since $v_{11} = -\sum_{i=2}^{r+t} v_{1i}$ and for $i = 2, \ldots, r+t$,

$$\begin{aligned} (SV)_{ii} &= \sum_{j=2}^{r+t} s_{ij}v_{ji} = \sum_{j=2}^{r+t}(\frac{\delta_{ij}}{v_{ii}} + \frac{1}{v_{11}})v_{ji} \\ &= 1 + \frac{1}{v_{11}}\sum_{j=2}^{r+t} v_{ji} = 1 - \frac{v_{1i}}{v_{11}}, \end{aligned}$$

we have

$$\begin{aligned} tr(WV) &= tr((V^{-1} - S)V) \\ &= tr(I - SV) = \sum_{i=2}^{r+t}\frac{v_{1i}}{v_{11}} = -1. \end{aligned}$$

Therefore, it is sufficient to show

$$\frac{Var(\sum_{i,j=2}^{r+t}(a_i - E(a_i))w_{ij}(a_j - E(a_j)))}{2(r+t)} = o(1)$$

in order to prove

(18)
$$\frac{(\mathbf{a} - E(\mathbf{a}))^T W(\mathbf{a} - E(\mathbf{a}))}{\sqrt{r+t}} = o_p(1).$$

If this is true, then Lemma 1 (2) comes from Lemma 1 (1) immediately.

There are four cases for calculating the covariance $g_{ij\zeta\eta} = Cov((a_i - Ea_i)w_{ij}(a_j - Ea_j), (a_\zeta - Ea_\zeta)w_{\zeta\eta}(a_\eta - Ea_\eta))$.

Case 1: $i = j = \zeta = \eta$. By (5),

$$
\begin{aligned}
|g_{iiii}| &\le w_{ii}^2(2v_{ii}^2 + \sum_{k=1}^{r+t}(3\tilde{v}_{ik}^2 + 2\tilde{v}_{ik})) \\
&\le w_{ii}^2(\frac{\lambda_{rt}^2}{8} + \frac{3\lambda_{rt}}{16} + \frac{\lambda_{rt}}{2}) = O(\lambda_{rt}^2 w_{ii}^2).
\end{aligned}
$$

Similarly, we have that

Case 2: only three indicates among the four indicates are the same (assume that $j = \zeta = \eta$)

$$
|g_{ijjj}| \le |w_{ij}w_{jj}|(\frac{\lambda_{rt}}{8} + \frac{3}{4}) = O(\lambda_{rt}|w_{ij}w_{jj}|);
$$

Case 3: only two indicates among the four indicates are the same (assume that $i = j$ or $j = \zeta$)

$$
\begin{aligned}
|g_{ii\eta\zeta}| &= |w_{ii}w_{\zeta\eta}(2\tilde{v}_{i\zeta}\tilde{v}_{i\eta} + v_{ii}\tilde{v}_{\zeta\eta})| \\
&\le |w_{ii}w_{\zeta\eta}|(\frac{\lambda_{rt}}{16} + \frac{\lambda_{rt}}{8}) = O(\lambda_{rt}|w_{ii}w_{\zeta\eta}|); \\
|g_{ijj\eta}| &= |w_{ii}w_{j\eta}(2\tilde{v}_{ji}\tilde{v}_{j\eta} + \tilde{v}_{ij}\tilde{v}_{j\eta})| \\
&\le \frac{3}{16}|w_{ii}w_{j\eta}| = O(|w_{ii}w_{j\eta}|).
\end{aligned}
$$

Case 4: All the four indicates are different

$$
\begin{aligned}
|g_{ij\zeta\eta}| &= |w_{ij}w_{\zeta\eta}(\tilde{v}_{i\zeta}\tilde{v}_{j\eta} + \tilde{v}_{i\eta}\tilde{v}_{j\zeta})| \\
&\le \frac{1}{8}|w_{ij}w_{\zeta\eta}| = O(|w_{ij}w_{\zeta\eta}|).
\end{aligned}
$$

Consequently, if $M_{rt} = o(\rho_{rt}^{1/8})$ and $\lambda_{rt}/\rho_{rt} \to c$, then

$$
\begin{aligned}
&\frac{Var[(\mathbf{a} - E(\mathbf{a}))^\top W(\mathbf{a} - E(\mathbf{a}))]}{2(r+t)} \\
\le\ & O(\frac{M_{rt}^{1/8}}{(r+t)\rho_{rt}^4} \times [(r+t)\lambda_{rt}^2 \\
&+ (r+t)^2\lambda_{rt} + (r+t)^3 + (r+t)^4]) \\
\le\ & O(\frac{M_{rt}^8}{\rho_{rt}}) = o(1).
\end{aligned}
$$

This completes the proof of Lemma 1 (2). □

Proof of Theorem 1. If $\max_i |(\hat{u}_i/u_i) - 1| \le 1$, then there exits a constant c_* such that $|\hat{\beta}_i - \beta_i| \le c_*|(\hat{u}_i/u_i) - 1|/2$ for all i. Let E be the event that

$$
\hat{\alpha}_{rt} := \max_{i,j} |\hat{\beta}_i - \beta_i - (\hat{\beta}_j - \beta_j)| \le \frac{c_*\theta_{rt}}{1 - \theta_{rt}}, \tag{19}
$$

where θ_{rt} is defined in (6). By Proposition 3, $P(E) \to 1$ if $M_{rt} = o(\log \rho_{rt})$ and $\lambda_{rt}/\rho_{rt} \to c$. The following calculations are based on the event E.

By Taylor's expansion, we have

$$
\ell(\hat{\boldsymbol{\beta}}_*) - \ell(\boldsymbol{\beta}_*) = (\mathbf{a} - E(\mathbf{a}))^\top(\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}_*) - \frac{1}{2}(\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}_*)^\top V(\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}_*) + z, \tag{20}
$$

where

$$
\begin{aligned}
z &= \frac{1}{6}[\sum_{i=2}^{r+t}\hat{\eta}_i(\hat{\beta}_i - \beta_i)^3 \\
&\quad + 2\sum_{i,j=2;i\ne j}^{r+t}\hat{\eta}_{ij}(\hat{\beta}_i - \beta_i)^2(\hat{\beta}_j - \beta_j)], \\
\hat{\eta}_i &= \frac{\partial^3 \ell}{\partial \beta_i^3}\Big|_{\boldsymbol{\beta}_* + \theta(\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}_*)} = \sum_{j=1}^{r+t}\frac{n_{ij}e^{\hat{\omega}_i}e^{\hat{\omega}_j}(e^{\hat{\omega}_j} - e^{\hat{\omega}_i})}{(e^{\hat{\omega}_i} + e^{\hat{\omega}_j})^3}, \\
\eta_{ij} &= \frac{\partial^3 \ell}{\partial \beta_i^2 \partial \beta_j}\Big|_{\boldsymbol{\beta}_* + \theta(\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}_*)} = \frac{n_{ij}e^{\hat{\omega}_i}e^{\hat{\omega}_j}(e^{\hat{\omega}_j} - e^{\hat{\omega}_i})}{(e^{\hat{\omega}_i} + e^{\hat{\omega}_j})^3}, \\
\hat{\omega}_i &= \beta_i + \theta(\hat{\beta}_i - \beta_i), \quad i = 2, \dots, r+t, \quad 0 \le \theta \le 1.
\end{aligned} \tag{21}
$$

By Taylor's expansion, we have

$$
\begin{aligned}
&\frac{n_{ij}e^{\hat{\beta}_i - \hat{\beta}_j}}{1 + e^{\hat{\beta}_i - \hat{\beta}_j}} - \frac{n_{ij}e^{\beta_i - \beta_j}}{1 + e^{\beta_i - \beta_j}} \\
=\ & \frac{n_{ij}e^{\beta_i - \beta_j}}{(1 + e^{\beta_i - \beta_j})^2} \times \hat{\gamma}_{ij} + \frac{n_{ij}e^{\hat{\theta}_{ij}}(1 - \hat{\theta}_{i,j})}{(1 + e^{\hat{\theta}_{ij}})^3} \times \hat{\gamma}_{ij}^2 \\
=\ & \tilde{v}_{ij}\hat{\gamma}_{ij} + \frac{n_{ij}e^{\hat{\theta}_{ij}}(1 - e^{\hat{\theta}_{ij}})}{(1 + e^{\hat{\theta}_{ij}})^3} \times \hat{\gamma}_{ij}^2,
\end{aligned}
$$

where $\hat{\gamma}_{ij} = \hat{\beta}_i - \beta_i - (\hat{\beta}_j - \beta_j)$, $\hat{\theta}_{ij} = \beta_i - \beta_j + d_{ij}(\hat{\beta}_i - \hat{\beta}_j)$ $(0 \le d_{ij} \le 1)$ and

$$
\tilde{v}_{ij} = |v_{ij}| = \frac{n_{ij}e^{\beta_i - \beta_j}}{(1 + e^{\beta_i - \beta_j})^2}, \ i \ne j; \quad \tilde{v}_{ii} = 0. \tag{22}
$$

Let

$$
h_{ij} = \frac{n_{ij}e^{\hat{\theta}_{ij}}(1 - e^{\hat{\theta}_{ij}})\hat{\gamma}_{ij}^2}{(1 + e^{\hat{\theta}_{ij}})^3}, \quad h_i = \sum_{j \ne i} h_{ij}. \tag{23}
$$

Then we have

$$
a_i - E(a_i) = \sum_{j=1}^{r+t}\tilde{v}_{ij}[(\hat{\beta}_i - \beta_i) - (\hat{\beta}_j - \beta_j)] + h_i, \quad i = 2, \dots, r+t.
$$

Write the above equations into the matrix:

$$
\mathbf{a} - E(\mathbf{a}) = V(\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}_*) + \mathbf{h}, \tag{24}
$$

where $\mathbf{h} = (h_2, \dots, h_{r+t})^\top$. Substituting $\hat{\boldsymbol{\beta}}_* - \boldsymbol{\beta}_* = V^{-1}[(\mathbf{a} - E(\mathbf{a})) - \mathbf{h}]$ into (20), it yields

$$
\ell(\hat{\boldsymbol{\beta}}_*) - \ell(\boldsymbol{\beta}_*) = \frac{1}{2}(\mathbf{a} - E(\mathbf{a}))^\top V^{-1}(\mathbf{a} - E(\mathbf{a})) - \frac{1}{2}\mathbf{h}^\top V^{-1}\mathbf{h} + z. \tag{25}
$$

In view of Lemma 1 (2), it is sufficient to prove that

$$\frac{\mathbf{h}^T V^{-1}\mathbf{h}}{\sqrt{\rho_{rt}}} = o_p(1), \quad \frac{z}{\sqrt{\rho_{rt}}} = o_p(1). \tag{26}$$

in order to prove Theorem 1.

Since $|e^x(1-e^x)/(1+e^x)^3| \le e^x/(1+e^x)^2 \le 1/4$, we have

$$|h_{ij}| \le n_{ij}\hat{\alpha}_{rt}^2/4, |h_i| \le \sum_{j\neq i}|h_{ij}| \le \sum_{j=1}^{r+t} n_{ij}\frac{\hat{\alpha}_{rt}^2}{4} \le \frac{\lambda_{rt}\hat{\alpha}_{rt}^2}{4}. \tag{27}$$

By (24), we have

$$\sum_{i=2}^{r+t}(a_i - E(a_i)) = \sum_{j=2}^{r+t}\tilde{v}_{j1}(\hat{\beta}_j - \beta_j) + \sum_{i=2}^{r+t} h_i.$$

Hence,

$$\begin{aligned} |\sum_{i=2}^{r+t} h_i| &= |-(a_1 - E(a_1)) - \sum_{j=2}^{r+t}\tilde{v}_{j1}(\hat{\beta}_j - \beta_j) \\ &\le |a_1 - E(a_1)| + v_{11}\hat{\alpha}_{rt}. \end{aligned} \tag{28}$$

It is easy to show that if $M_{rt} = o(\log \rho_{rt})$, then $(a_1 - E(a_1))^2/v_{11} = O_p(1)$, by noting that $a_1 = \sum_{i=r+1}^{r+t} a_{1i}$ is a sum of t independent Binomial random variables. Since

$$\frac{M_{rt}}{(1+M_{rt})^2} \le \tilde{v}_{ij} \le \frac{1}{4},\ i\neq j; \quad \frac{\rho_{rt}M_{rt}}{(1+M_{rt})^2} \le v_{ii} \le \frac{\lambda_{rt}}{4}, \tag{29}$$

by (27) and (28), we have

$$\begin{aligned} \mathbf{h}^T S\mathbf{h} &= \sum_{i=2}^{r+t}\frac{h_i^2}{v_{ii}} + \frac{(\sum_{i=2}^{r+t} h_i)^2}{v_{11}} \\ &\le \frac{\lambda_{rt}^2\hat{\alpha}_{rt}^4}{16}\times\frac{(1+M_{rt})^2}{\rho_{rt}M_{rt}} + \frac{2(a_1-E(a_1))^2}{v_{11}} + 2v_{11}\hat{\alpha}_{rt}^2 \\ &\le O(\frac{\lambda_{rt}^4 M_{rt}^9 e^{4c^*M_{rt}}(\log\lambda_{rt})^2}{\rho_{rt}^4}). \end{aligned}$$

Therefore, by Proposition 2 and the inequality (27), we have

$$\begin{aligned} & |\mathbf{h}^T V^{-1}\mathbf{h}| \\ \le\ & |\mathbf{h}^T S\mathbf{h}| + |\mathbf{h}^T W\mathbf{h}| \\ \le\ & \frac{h_i^2}{v_{ii}} + \frac{(\sum_{i=2}^{r+t} h_i)^2}{v_{11}} + \|W\|\sum_{i,j=2}^{r+t}|h_i||h_j| \\ \le\ & O\left(\frac{\lambda_{rt}^4 M_{rt}^9 e^{4c^*M_{rt}}(\log\lambda_{rt})^2}{\rho_{rt}^4}\right) \\ & + O\left(\frac{\lambda_{rt}^6 M_{rt}^{12} e^{4c^*M_{rt}}(\log\lambda_{rt})^2}{\rho_{rt}^6}\right). \end{aligned} \tag{30}$$

Assuming that $\hat{\alpha}_{rt}$ is sufficiently small, we have

$$\begin{aligned} \hat{\eta}_{ij} &\le \frac{n_{ij}}{4}\times|\frac{e^{\beta_j+\theta(\hat{\beta}_j-\beta_j)} - e^{\beta_i+\theta(\hat{\beta}_i-\beta_i)}}{e^{\beta_i+\theta(\hat{\beta}_i-\beta_i)} + e^{\beta_j+\theta(\hat{\beta}_j-\beta_j)}}| \\ &\le \frac{n_{ij}}{4}\times\left(|\frac{e^{\beta_j}-e^{\beta_i}}{e^{\beta_j}+e^{\beta_i}}| + 2\hat{\alpha}_{rt}\right). \end{aligned} \tag{31}$$

Consequently, we have

$$\begin{aligned} 6|z| &\le 3\hat{\alpha}_{rt}^3\left(\sum_{i,j=1}^{r+t}\frac{n_{ij}}{4}[|\frac{e^{\beta_i}-e^{\beta_j}}{e^{\beta_i}+e^{\beta_j}}| + 2\hat{\alpha}_{rt}]\right) \\ &\le O\left(M_{rt}^8 e^{4c^*M_{rt}}(\log\lambda_{rt})^2\frac{\lambda_{rt}^4}{\rho_{rt}^4}\right) \\ &\quad + O\left(M_{rt}^6 e^{3c^*M_{rt}}(\log\lambda_{rt})^{3/2}\sum_{i,j=1}^{r+t}|\frac{e^{\beta_i}-e^{\beta_j}}{e^{\beta_i}+e^{\beta_j}}|\frac{\lambda_{rt}^{3/2}}{\rho_{rt}^3}\right). \end{aligned} \tag{32}$$

By (30) and (32), if $M_{rt} = o(\log\rho_{rt})$ and (7) holds, then we have (26). This completes the proof. □

Let

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{12}^\top & V_{22}\end{pmatrix},$$

where V_{11} and V_{22} have the dimension $(m-1)\times(m-1)$ and $(r+t-m)\times(r+t-m)$, respectively. V_{11} and V_{22} are the covariance matrices of $\mathbf{a}^1 = (a_2,\dots,a_m)^\top$ and $\mathbf{a}^2 = (a_{m+1},\dots,a_{r+t})^\top$, respectively. Let $\mu = \sum_{i=2}^m\sum_{j\notin\{2,\dots,m\}}\tilde{v}_{ij}$ be the variance of $a_2+\cdots+a_m$ and $\boldsymbol{\omega}$ be the vector of the column sum of V_{12}, where $j\notin\{2,\dots,m\}$ denotes $j\in\{1,\dots,r+m\}\backslash\{2,\dots,m\}$. Let

$$U = \begin{pmatrix} \mu & \boldsymbol{\omega}^\top \\ \boldsymbol{\omega} & V_{22}\end{pmatrix},$$

Similar to the proof of Proposition 2, we have

Lemma 2. *Let $\bar{S} = (\bar{s}_{ij})_{i,j=m+1,\dots,r+t}$, where $\bar{s}_{ij} = \delta_{ij}/v_{ii} + 1/v_{11}$ for all $i\neq j$ except for $\bar{s}_{11} = 1/\mu + 1/v_{11}$. Then we have*

$$\|\overline{W} = U^{-1} - \bar{S}\| \le O(\frac{M_{rt}^4}{(r+t-m)^2}),$$

where the norm $\|\cdot\|$ is defined in Proposition 2.

Proof of Theorem 2. Let $\mathbf{b} = (\sum_{i=2}^m a_i, a_{m+1},\dots,a_{r+t})^\top$. The definition of z and h_i in (21) and (23) can be viewed as a function on $\hat{\boldsymbol{\beta}}_*$ and $\boldsymbol{\beta}_*$. Since $\hat{\beta}_2 = \cdots = \hat{\beta}_m$ and $\beta_2 = \cdots = \beta_m$ under H_0, similar to the proof of (20), we have

$$\ell(\hat{\boldsymbol{\beta}}_*^H) - \ell(\boldsymbol{\beta}_*) = \frac{1}{2}(\mathbf{b}-E(\mathbf{b}))^\top U^{-1}(\mathbf{b}-E(\mathbf{b})) - \frac{1}{2}\bar{\mathbf{h}}^\top U^{-1}\bar{\mathbf{h}} + \bar{z}, \tag{33}$$

where $\bar{\mathbf{h}} = (\sum_{i=2}^m\bar{h}_i, \bar{h}_{m+1},\dots,\bar{h}_{r+t})^\top$, $\bar{h}_i = h_i(\hat{\boldsymbol{\beta}}_* = \hat{\boldsymbol{\beta}}_*^H, \boldsymbol{\beta}_*)$ and $\bar{z} = z(\hat{\boldsymbol{\beta}}_* = \hat{\boldsymbol{\beta}}_*^H, \boldsymbol{\beta}_*)$.

If $M_{rt} = o(\log\rho_{rt})$, $\lambda_{rt}/\rho_{rt}\to c$ and (7) holds, similar to the proof of (26), we have

$$\frac{|\bar{\mathbf{h}}^\top U^{-1}\bar{\mathbf{h}}|}{\sqrt{r+t-m}} = o_p(1), \quad \frac{|\bar{z}|}{\sqrt{r+t-m}} = o_p(1).$$

Since $m/\rho_{rt} \geq \tau > 0$ and τ is a constant, we have

$$\frac{\ell(\hat{\boldsymbol{\beta}}_*^H) - \ell(\boldsymbol{\beta}_*)}{\sqrt{2(m-1)}} = \frac{\frac{1}{2}(\mathbf{b} - E(\mathbf{b}))^\top U^{-1}(\mathbf{b} - E(\mathbf{b}))}{\sqrt{2(m-1)}} + o_p(1).$$

Similar to the proof of (18), we have

$$\frac{(\mathbf{b} - E(\mathbf{b}))^\top \overline{W}(\mathbf{b} - E(\mathbf{b}))}{r+t-m} = o_p(1).$$

Consequently,

$$(34)\quad \frac{\ell(\hat{\boldsymbol{\beta}}_*^H) - \ell(\boldsymbol{\beta}_*)}{\sqrt{2(m-1)}} = \frac{\frac{1}{2}(\mathbf{b} - E(\mathbf{b}))^T \bar{S}(\mathbf{b} - E(\mathbf{b}))}{\sqrt{2(m-1)}} + o_p(1).$$

Note that

$$(35)\qquad (\mathbf{a} - E(\mathbf{a}))^\top S(\mathbf{a} - E(\mathbf{a})) = \sum_{i=1}^{r+t} \frac{(a_i - E(a_i))^2}{v_{ii}},$$

$$(36)\ (\mathbf{b} - E(\mathbf{b}))^T \bar{S}(\mathbf{b} - E(\mathbf{b})) = \sum_{i=m+1}^{r+t} \frac{(a_i - E(a_i))^2}{v_{ii}} + \frac{[\sum_{i=2}^m (a_i - E(a_i))]^2}{\mu} + \frac{(a_1 - E(a_1))^2}{v_{11}}.$$

Moreover, it is easy to show $[\sum_{i=2}^m (a_i - E(a_i))]^2/\mu = O_p(1)$ by referring to the central limit theorem for the bounded case (Loève (1977), page 289) if $\mu = \sum_{i=2}^m v_{ii}$ diverges. In view of that $m/\rho_{rt} \geq \tau > 0$, by (20), (26) and (18), we have

$$(37)\quad \frac{\ell(\hat{\boldsymbol{\beta}}_*) - \ell(\boldsymbol{\beta}_*)}{\sqrt{2(m-1)}} = \frac{\frac{1}{2}(\mathbf{a} - E(\mathbf{a}))^T S(\mathbf{a} - E(\mathbf{a}))}{\sqrt{2(m-1)}} + o_p(1).$$

Combining (34), (35), (36) and (37), it yields

$$\frac{2(\ell(\hat{\boldsymbol{\beta}}_*) - \ell(\hat{\boldsymbol{\beta}}_*^H)) - (m-1)}{\sqrt{2(m-1)}} = \frac{\sum_{i=2}^m (a_i - E(a_i))^2/v_{ii} - (m-1)}{\sqrt{2(m-1)}} + o_p(1).$$

Similar to the proof of Lemma 1 (2), the main item of the right expression in the above equation is asymptotically normal if $M_{rt} = o(\log \rho_{rt})$ holds. This completes the proof. □

ACKNOWLEDGEMENT

We are very grateful to two anonymous referees, an associate editor, and the Editor for their valuable comments that have led to an improvement of the manuscript.

Received 15 August 2014

REFERENCES

Andersen, E. B. (1973b). A goodness of fit test for the Rasch model. *Psychometrika* **38**, 123–140. MR0311064

Anderson C., Li Z. and Vermunt J. (2007). Estimation of models in the Rasch family for polytomous items and multiple latent variables. *Journal of Statistical Software* **20**(6), 1–36.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika* **43**, 561–73.

Barndorff-Nielsen, O. (1978). Information and Exponential Families in Statistical Theory. New York: Wiley. MR0489333

Berk, R. H. (1972). Consistency and asymptotic normality of the MLE's for exponential models. *Ann. Math. Statist.* **43**, 193–204. MR0298810

Bond T. G. and Fox C. M. (2007). Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 2nd ed. Mahwah, NJ, US: Lawrence Erlbaum.

Bradley, R. A. and Terry, M. E. (1952). The rank analysis of incomplete block designs I. The method of paired comparisons. *Biometrika* **39**, 324–345. MR0070925

Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence, *Ann. Appl. Probab.* **21**, 1400–1435. MR2857452

Ghosh M. (1995). Inconsistent maximum likelihood estimators for the Rasch model. *Statistics & Probability Letters* **23**, 165–170. MR1341360

Fan J., Zhang C., and Zhang J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon, *Ann. Statist.* **29**, 153–193. MR1833962

Feller W. (1945). The fundamental limit theorems in probability. *Bull. Amer. Math. Soc.* **51**, 800–832. MR0013252

Fischer, G. H. (1974). Einfuhrung in die Theorie Psychologischer Tests [Introduction to the Theory of Psychological Tests]. Bern: Huber. (In German).

Fischer, G. H. (1978). Probabilistic test models and their applications. *The German Journal of Psychology* **2**, 298–319.

Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika* **46**, 59–77. MR0655008

Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics* **5**, 815–841. MR0501540

Hambleton, R. K., Swaminathan, H., Cook, L. L., Eignor, D. R. and Gifford, J. A. (1978). Developments in latent trait theory: Models, technical issues, and applications. *Review of Educational Research* **48**, 467–510.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30. MR0144363

Lauritzen, S. L. (2003). Rasch models with exchangeable rows and columns. In Bernardo J. M., Bayarri M. J., Berger J. O., Dawid A. P., Heckerman D., Smith A. F. M., and West M. (Eds.), Bayesian Statistics 7, Oxford University Press, 215–232. MR2003175

Lindeberg J. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15**, 211–225. MR1544569

Loève, M. (1977). Probability Theory I. 4th ed. New York: Springer-Verlag. MR0651017

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika* **47**, 149–174.

Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Paedagogiske Institut.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In Neyman J. (Ed.), Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 4). Berkeley: University of California Press, 321–333. MR0135196

Rasch, G. (1966). An individualistic approach to item analysis. In Lazarsfeld P. F. and Henry N. W. (Eds.), Readings in Mathematical Social Science. Cambridge, Mass.: M.I.T. Press, 89–107.

Rosenthal H. P. (1970). On the subspaces of $L_p(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math.* **8**, 273–303. MR0271721

Simons, G. and Yao, Y.-C. (1999). Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons. *Ann. Statist.* **27**, 1041–1060. MR1724040

Wright B. D. and Mok M. C. (2004). An overview of the family of Rasch measurement models. In Smith E., Smith R. (Eds.), Introduction to Rasch Measurement. Chicago, Jampress.

Yan T. and Xu J. (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika* **100**, 519–524. MR3068452

Ting Yan
Department of Statistics
Central China Normal University
Wuhan, 430079
China
E-mail address: tingyanty@mail.ccnu.edu.cn

Zhaohai Li
Department of Statistics
The George Washington University
2140 Pennsylvania Ave., N.W.
Washington, DC 20052
USA
E-mail address: zli@gwu.edu

Yuanzhang Li
Walter Reed Army Institute of Research
503 Robert Grant Ave, Silver Spring
Maryland, 20910
USA
E-mail address: Liy.Li@us.army.mil

Hong Qin
Department of Statistics
Central China Normal University
Wuhan, 430079
China
E-mail address: qinhong@mail.ccnu.edu.cn