THE ORDER-PRESERVING PROPERTIES OF ESTIMATES IN POLYTOMOUS ITEM RESPONSE THEORY MODELS WITH APPROXIMATED LIKELIHOOD FUNCTIONS

KAZUMASA MORI

Received June 4, 2015

ABSTRACT. In this study, we consider the ordering properties of the estimates of the rating scale model(RSM) and related polytomous item response theory (IRT) models. First, we propose a kind of approximation to the likelihood functions for these IRT models. The approximated likelihood functions are derived from the inequality of arithmetic and geometric means. We then evaluate upper limits of the functions based on the mathematical result of Specht(1960). Next, we derive the order-preserving statistics for these polytomous IRT models. All sets of statistics are derived by using the characteristics of arrangement increasing functions (Hollander *et al.*, 1977, Marshall *et al.*, 2011). We also carry out simulation study and confirm that our order-preserving statistics work well in typical educational testing. Finally, it is shown that the order-preserving statistics of the RSM in three major three estimation methods coincide.

1 Introduction In this study, we consider the order-preserving properties of the estimates of the rating scale model (RSM;Rasch, 1960; Andrich,1978a, 1978b; Andersen,1996) and related polytomous item response theory (IRT) models. First, we introduce the RSM. Consider that a test comprises k items administered to n subjects and suppose that each item can take m categories. The response variable for the i-th subject and the j-th item becomes $X_{ijh} = \{0, 1\}$. When the i-th subject responds with an h to the j-th item, the corresponding probability of the RSM is

$$P_{ijh}(\theta_i, \boldsymbol{\alpha}_j) = P(X_{ijh} = 1; \theta_i, \boldsymbol{\alpha}_j) = \frac{\exp(w_h\theta_i + a_{jh})}{\sum_{h=1}^{m} \exp(w_h\theta_i + \alpha_{jh})}. \tag{1}$$

Here, θ_i is the ability parameter for the i-th subject, α_{jh} is the item parameter for the h-th category of the j-th item ($\boldsymbol{\alpha}_j = (\alpha_{j1}, \alpha_{j2}, \cdots, \alpha_{jm})$ and w_h is the weight coefficient for the h-th category. Note that w_h is assumed as given. In addition, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_n)$ is an n-dimensional vector of the ability parameters and $\boldsymbol{\alpha} = (\alpha_{11}, \cdots, \alpha_{1m}, \cdots, \alpha_{k1}, \cdots, \alpha_{km})$ is a $k \times m$-dimensional vector of the item parameters. To estimate parameters in (1), we often use the maximum likelihood principle. In the RSM, the form of the likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\alpha}|X) &= \prod_{i=1}^{n}\prod_{j=1}^{k}\prod_{h=1}^{m} P(X_{ijh} = x_{ijh}) \\ &= \frac{\exp\left(\sum_{i=1}^{n} \theta_i \sum_{j=1}^{k}\sum_{h=1}^{m} w_h x_{ijh} + \sum_{j=1}^{k}\sum_{h=1}^{m} \alpha_{jh}\sum_{i=1}^{n} x_{ijh}\right)}{\prod_{i=1}^{n}\prod_{j=1}^{k}\sum_{h=1}^{m}\exp(w_h\theta_i + a_{jh})} \\ &= \frac{\exp\left(\sum_{i=1}^{n}\theta_i t_i + \sum_{j=1}^{k}\sum_{h=1}^{m}\alpha_{jh} r_{jh}\right)}{C(\boldsymbol{\theta}, \boldsymbol{\alpha})}, \end{aligned} \tag{2}$$

2010 *Mathematics Subject Classification.* 62P15,62B05.
Key words and phrases. Polytomous item response theory, Order-preserving property, Rating scale model, Arrangement increasing, Educational testing.

where X is a response matrix that consists of all the response variables, $t_i = \sum_{j=1}^{k}\sum_{h=1}^{m} w_h x_{ijh}$ is the score for the i-th subject, $r_{jh} = \sum_{i=1}^{n} x_{ijh}$ is the number of subjects who response 1 to the h-th category of the j-th item and $C(\boldsymbol{\theta},\boldsymbol{\alpha}) = \prod_{i=1}^{n}\prod_{j=1}^{k}\sum_{h=1}^{m}\exp(w_h\theta_i + \alpha_{jh})$.

In IRT, three major estimation methods have been proposed that use a the likelihood function as in (2) : joint maximum likelihood estimation (JMLE),marginal maximum likelihood estimation (MMLE; Bock and Lieberman, 1970, Thissen, 1982), and conditional maximum likelihood estimation (CMLE; Andersen, 1972). JMLE estimates $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ simultaneously by maximizing (2) in the RSM. By contrast, CMLE and MMLE remove $\boldsymbol{\theta}$ from (2) and estimate $\boldsymbol{\alpha}$ separately. In particular, CMLE uses a conditional likelihood function in which we assume that the score t_i for the i-th subject is already given and thus remove $\boldsymbol{\theta}$ from the function. We focus on JMLE in this paper.

The RSM has many relations with other polytomous IRT models. For example, when we reparameterize $\alpha_{jh} = \sum_{p=1}^{q} v_{jhp}\eta_p$ in (2), we obtain the linear rating scale model(LRSM; Fischer and Parzer, 1991). Here, η_p is the "basic parameter," $q < m$ is a dimension of the basic parameter vector $\boldsymbol{\eta}$ and v_{jhp} is the weight coefficient which is assumed as already given. The likelihood function of the LRSM corresponding to (2) is

$$L(\boldsymbol{\theta},\boldsymbol{\eta}|X) = \frac{\exp(\sum_{i=1}^{n}\theta_i t_i + \sum_{p=1}^{q}\eta_p r'_p)}{C(\boldsymbol{\theta},\boldsymbol{\eta})}, \tag{3}$$

where $r'_p = \sum_{i=1}^{n}\sum_{j=1}^{k} v_{jhp}x_{ijh}$, $C(\boldsymbol{\theta},\boldsymbol{\eta}) = \prod_{i=1}^{n}\prod_{j=1}^{k}\sum_{h=1}^{m}\exp(w_h\theta_i + \sum_{p=1}^{q} v_{jhp}\eta_p)$ and $\boldsymbol{\eta}$ is the q-dimensional vector of the basic parameters.

Another important model related to the RSM is the partial credit model (PCM; Masters, 1982). The PCM is special case of the RSM. In other words, when we substitute $w_h = h$ for the probability function of the RSM (1), we get the that of the PCM. The likelihood function of the PCM is

$$L(\boldsymbol{\theta},\boldsymbol{\beta}|X) = \frac{\exp(\sum_{i=1}^{n}\theta_i t_i^* + \sum_{j=1}^{k}\sum_{h=1}^{m}\beta_{jh}r_{jh})}{C(\boldsymbol{\theta},\boldsymbol{\beta})}, \tag{4}$$

where β_{jh} is the item parameter for the h-th category of the j-th item, $\boldsymbol{\beta}$ is the $k \times m$-dimensional vector of the item parameters, $t_i^* = \sum_{j=1}^{k}\sum_{h=1}^{m} hx_{ijh}$, $C(\boldsymbol{\theta},\boldsymbol{\beta}) = \prod_{i=1}^{n}\prod_{j=1}^{k}\sum_{h=1}^{m}\exp(h\theta_i + \beta_{jh})$, and $\boldsymbol{\beta} = (\beta_{11},\cdots,\beta_{1m},\cdots,\beta_{11},\cdots,\beta_{k1},\cdots,\beta_{11},\cdots,\beta_{km})$. In (4), by reparameterizing $\beta_{jh} = \sum_{p=1}^{q} u_{jhp}\gamma_p$, we drive the linear partial credit model (LPCM; Glas and Verhelst,1989; Fischer and Ponocny, 1994). Here, γ_p is the basic parameter and u_{jhp} is the weight coefficient which is assumed as already given. The likelihood function of the LPCM is

$$L(\boldsymbol{\theta},\boldsymbol{\beta}|X) = \frac{\exp(\sum_{i=1}^{n}\eta_i t_i^* + \sum_{p=1}^{q}\gamma_q r_q^*)}{C(\boldsymbol{\theta},\boldsymbol{\gamma})}, \tag{5}$$

where $r_p^* = \sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{h=1}^{m} u_{jhp}x_{ijh}$, $C(\boldsymbol{\theta},\boldsymbol{\gamma})\prod_{i=1}^{n}\prod_{j=1}^{k}\sum_{h=1}^{m}\exp(h\theta_i + \sum_{p=1}^{q} u_{jhp}\gamma_p)$ and $\boldsymbol{\gamma}$ is the p-dimensional vector of the basic parameters.

Specht(1960) considered the upper limit of an inequality between the arithmetic mean and geometric mean. Following Seo (2000), let $y_1,\cdots,y_m \in [d,D]$ with $D \geq d > 0$. Then, this inequality is such that

$$S(z)\sqrt[m]{y_1 y_2 \cdots y_m} \geq \frac{y_1 + y_2 + \cdots + y_m}{m}, \tag{6}$$

where $z = D/d$ and $S(z)$ are defined as

$$S(z) = \frac{(z-1)z^{\frac{1}{z-1}}}{e\log z}(z > 1) \text{ and } S(1) = 1\ (z = 0). \tag{7}$$

Here, we call $S(z)$ Specht's ratio. Then, we consider the approximation below:

$$mAM(\boldsymbol{y}) = y_1 + y_2 + \cdots + y_m \simeq m\sqrt[m]{y_1 y_2 \cdots y_m} = mGM(\boldsymbol{y}), \tag{8}$$

where $\boldsymbol{y} = (y_1, y_2, \cdots, y_m)$. We can evaluate the difference in the above approximation by using a ratio, that is

$$DR(\boldsymbol{y}) = \frac{y_1 + y_2 + \cdots + y_m}{m\sqrt[m]{y_1 y_2 \cdots y_m}}. \tag{9}$$

Here, we call $DR(\boldsymbol{y})$ the "difference ratio" between $AM(\boldsymbol{y})$ and $GM(\boldsymbol{y})$. From (6), it holds that $S(z) \geq DR(\boldsymbol{y}) \geq 1$. This means that we can evaluate the upper limit of the least difference for (9) from $S(z)$ in (6).

By substituting (8) with $y_h = \exp(w_h\theta_i + \alpha_{jh})$ for each i and j into (2), we get

$$\sum_{h=1}^{m} \exp(w_h\theta_i + \alpha_{jh}) \simeq m\sqrt[m]{\prod_{h=1}^{m} \exp(w_h\theta_i + \alpha_{jh})} \tag{10}$$

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\alpha}|X) &\simeq \frac{\exp(\sum_{i=1}^{n} \theta_i t_i + \sum_{j=1}^{k} \sum_{h=1}^{m} \alpha_{jh} r_{jh})}{m^{nk} \prod_{i=1}^{n} \prod_{j=1}^{k} \prod_{h=1}^{m} \exp(w_h\theta_i + \alpha_{jh})} \\ &= \frac{\exp(\sum_{i=1}^{n} \theta_i t_i + \sum_{j=1}^{k} \sum_{h=1}^{m} \alpha_{jh} r_{jh})}{\tilde{C}(\boldsymbol{\theta}, \boldsymbol{\alpha})} = \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}|X), \end{aligned} \tag{11}$$

where $\tilde{C}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = m^{nk} \prod_{i=1}^{n} \prod_{j=1}^{k} \sqrt[m]{\prod_{h=1}^{m} \exp(w_h\theta_i + \alpha_{jh})}$. These approximations of the likelihood functions can also be applied to (3), (4), and (5), which means that we can consider approximated likelihood functions to the the models related to the RSM.

In this study, we consider the ordering properties of the estimates in the RSM and the related polytomous IRT models with the approximated likelihood functions as in (11). We assume that the response matrix X is already given, all estimates derived from X exist, and each estimate is unique. Note that most conventional studies (e.g., Hemker *et al.*,1997;Van der Ark,2005, 2010) have considered the properties of other ordering: stochastic ordering (SO). In other words, they regard X as a matrix that consists of random variables and consider the ordering properties of estimators with SO.

The remainder of the paper is organized as follows. The preliminaries and main results are presented in section 2. Some performances that the approximations denoted above holds are evaluated by simulation studies in section 3. Finally, section 4 discusses our results and concludes.

2 Preliminaries and main results In this study, we use some characteristics of arrangement increasing (AI) functions (Hollander *et al.*, 1977) to consider the order preserving properties of the estimates. First, we introduce some definitions, as per Marshall *et al.* (2011), Boland and Proschan (1988) and Mori(2015).

Definition 1. Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be n-dimensional vectors. We define equality $\stackrel{a}{=}$ as

$$(\boldsymbol{a}\Pi, \boldsymbol{b}\Pi) \stackrel{a}{=} (\boldsymbol{a}, \boldsymbol{b}),$$

where Π is an arbitrary $n \times n$ permutation matrix. In this definition, we find $(\boldsymbol{a}, \boldsymbol{b}) \stackrel{a}{=} (\boldsymbol{a}\Pi_1, \boldsymbol{b}\Pi_1) \stackrel{a}{=} (\boldsymbol{a}_\uparrow, \boldsymbol{b}\Pi_1) \stackrel{a}{=} (\boldsymbol{a}\Pi_2, \boldsymbol{b}\Pi_2) \stackrel{a}{=} (\boldsymbol{a}_\downarrow, \boldsymbol{b}\Pi_2)$, where Π_1 is a matrix such that $\boldsymbol{a}\Pi_1 = \boldsymbol{a}_\uparrow$ and Π_2 is a matrix such that $\boldsymbol{a}\Pi_2 = \boldsymbol{a}_\downarrow$. Here, we use the ordered vectors $\boldsymbol{a}_\uparrow$ and $\boldsymbol{a}_\downarrow$, which are vectors with the components of $\boldsymbol{a}$ arranged in ascending order and descending order, respectively. Note it is not always hold that $\boldsymbol{b}\Pi_1 = \boldsymbol{b}_{\uparrow}$ or $\boldsymbol{b}\Pi_2 = \boldsymbol{b}_{\downarrow}$. For detail, see below Example 3.

Then, we define a partial order $\overset{a}{\leq}$ for the vector arguments.

Definition 2. Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be n-dimensional vectors. First, we permute $\boldsymbol{a}$ and $\boldsymbol{b}$ so that

$$(\boldsymbol{a}, \boldsymbol{b}) \overset{a}{=} (\boldsymbol{a}_{\uparrow}, \boldsymbol{b}'). \tag{12}$$

Here, $\boldsymbol{b}' = \boldsymbol{b}\Pi_1$ and Π_1 form the permutation matrix such that $\boldsymbol{a}\Pi_1 = \boldsymbol{a}_{\uparrow}$. Then, we generate a vector $\boldsymbol{b}^*_{l,m}$ from $\boldsymbol{b}'$ in (12) by interchanging the l-th and m-th component ($l < m$) of $\boldsymbol{b}$ such that $b_l > b_m$. Finally, we define the partial order $\overset{a}{\leq}$ as

$$(\boldsymbol{a}_{\uparrow}, \boldsymbol{b}') \overset{a}{\leq} (\boldsymbol{a}_{\uparrow}, \boldsymbol{b}^*_{l,m}).$$

Therefore, it holds that $(\boldsymbol{a}_{\uparrow}, \boldsymbol{b}_{\downarrow}) \overset{a}{=} (\boldsymbol{a}_{\downarrow}, \boldsymbol{b}_{\uparrow}) \overset{a}{\leq} (\boldsymbol{a}, \boldsymbol{b}) \overset{a}{\leq} (\boldsymbol{a}_{\uparrow}, \boldsymbol{b}_{\uparrow}) \overset{a}{=} (\boldsymbol{a}_{\downarrow}, \boldsymbol{b}_{\downarrow})$.

Here we show an example for the equality $\overset{a}{=}$ and the inequality $\overset{a}{\leq}$.

Example 3. Let $\boldsymbol{a} = (7, 5, 3, 1)$ and $\boldsymbol{b} = (6, 4, 8, 2)$. Then,

$$\begin{aligned}&(\boldsymbol{a}, \boldsymbol{b}) \overset{a}{=} ((1, 3, 5, 7), (2, 8, 4, 6)) \overset{a}{\leq} ((1, 3, 5, 7), (2, 4, 8, 6))\\ &\overset{a}{\leq} ((1, 3, 5, 7), (2, 4, 6, 8)) \overset{a}{=} ((7, 5, 3, 1), (8, 6, 4, 2)).\end{aligned}$$

Definition 4. An AI function is a function, g, with two n-dimensional vector arguments that preserve the ordering $\overset{a}{\leq}$. Thus, if g is AI, it holds that $g(\boldsymbol{a}, \boldsymbol{b}) \leq g(\boldsymbol{a}_{\uparrow}, \boldsymbol{b}^*_{l,m})$ for n-dimensional vectors $\boldsymbol{a}$,$\boldsymbol{b}$,$\boldsymbol{a}_{\uparrow}$,$\boldsymbol{b}^*_{l,m}$, such that $(\boldsymbol{a}, \boldsymbol{b}) \overset{a}{\leq} (\boldsymbol{a}_{\uparrow}, \boldsymbol{b}^*_{l,m})$.

Here, we find

$$g(\boldsymbol{a}_{\uparrow}, \boldsymbol{b}_{\downarrow}) = g(\boldsymbol{a}_{\downarrow}, \boldsymbol{b}_{\uparrow}) \leq g(\boldsymbol{a}, \boldsymbol{b}) \leq g(\boldsymbol{a}_{\uparrow}, \boldsymbol{b}_{\uparrow}) = g(\boldsymbol{a}_{\downarrow}, \boldsymbol{b}_{\downarrow}) \tag{13}$$

for AI function g.

Next, we prepare a general result as lemma (without proof) that describes the necessary and sufficient condition for AI functions containing summation forms.

Lemma 5. (Marshall *et al.*, 2011, p.233) If g has the form $g(\boldsymbol{a}, \boldsymbol{b}) = \sum_{i=1}^{n} \phi(a_i, b_i)$, then g is AI if and only if ϕ is L-superadditive.

Here, L-superadditive is the function that satisfies

$$\frac{\partial}{\partial a \partial b} \phi(a, b) \geq 0.$$

Then, we consider the log likelihoods derived from (11) for preparation:

$$\log \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{t}, \boldsymbol{r}) = \sum_{i=1}^{n} \theta_i t_i + \sum_{j=1}^{k} \sum_{h=1}^{m} \alpha_{jh} r_{jh} - \log \tilde{C}(\boldsymbol{\theta}, \boldsymbol{\alpha}). \tag{14}$$

As $\tilde{C}(\boldsymbol{\theta}, \boldsymbol{\alpha})$ is invariant for the rearrangement within $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, $\tilde{C}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \tilde{C}(\boldsymbol{\theta}\Pi_1, \boldsymbol{\alpha}\Pi_2)$ for any permutation matrices Π_1 and Π_2. Thus, we can only focus on parts of log likelihood function (14) for evaluating order-preserving properties, which are

$$\sum_{i=1}^{n} \theta_i t_i + \sum_{j=1}^{k} \sum_{h=1}^{m} \alpha_{jh} r_{jh} = \tilde{l}_1(\boldsymbol{\theta}, \boldsymbol{t}) + \tilde{l}_2(\boldsymbol{\alpha}, \boldsymbol{r}) \tag{15}$$

Here, $\boldsymbol{t}$ and $\boldsymbol{r}$ are vectors that consist of $\{t_i\}$ and $\{r_{jh}\}$, respectively.

Now, we propose our propositions.

Proposition 6. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\alpha}}$ be maximum likelihood estimates for the log likelihood function $\log \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{t}, \boldsymbol{r})$ in (14). Then, $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\alpha}}^*$ maximizes $\log \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{t}_\uparrow, \boldsymbol{r}_\uparrow)$ if and only if $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}}_\uparrow$ and $\hat{\boldsymbol{\alpha}}^* = \hat{\boldsymbol{\alpha}}_\uparrow$.

Proof. First, we assume the maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\alpha}}$ are already given. Then, we find that $\tilde{l}_1(\hat{\boldsymbol{\theta}}, \boldsymbol{t})$ and $\tilde{l}_2(\hat{\boldsymbol{\alpha}}, \boldsymbol{r})$ in (15) are permutation-invariant within each set of vectors $(\hat{\boldsymbol{\theta}}, \boldsymbol{t})$ and $(\hat{\boldsymbol{\alpha}}, \boldsymbol{r})$, which means that $\tilde{l}_1(\hat{\boldsymbol{\theta}}, \boldsymbol{t}) = \tilde{l}_1(\hat{\boldsymbol{\theta}}\Pi_1, \boldsymbol{t}\Pi_1)$ and $\tilde{l}_2(\hat{\boldsymbol{\alpha}}, \boldsymbol{r}) = \tilde{l}_2(\hat{\boldsymbol{\alpha}}\Pi_2, \boldsymbol{r}\Pi_2)$ for any permutation matrices, Π_1 and Π_2. From this permutation invariance and the uniqueness of the maximum likelihood estimates, we obtain

$$\tilde{l}_1(\hat{\boldsymbol{\theta}}, \boldsymbol{t}) + \tilde{l}_2(\hat{\boldsymbol{\alpha}}, \boldsymbol{r}) = \tilde{l}_1(\hat{\boldsymbol{\theta}}\Pi_1^*, \boldsymbol{t}_\uparrow) + \tilde{l}_2(\hat{\boldsymbol{\alpha}}\Pi_2^*, \boldsymbol{r}_\uparrow) = \tilde{l}_1(\hat{\boldsymbol{\theta}}^*, \boldsymbol{t}_\uparrow) + \tilde{l}_2(\hat{\boldsymbol{\alpha}}^*, \boldsymbol{r}_\uparrow),$$

where Π_1^* and Π_2^* are permutation matrices such that $\boldsymbol{t}\Pi_1^* = \boldsymbol{t}_\uparrow$ and $\boldsymbol{r}\Pi_2^* = \boldsymbol{r}_\uparrow$. Thus, we find that both $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\alpha}}^*$ are rearranged forms of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\alpha}}$, respectively.

On the contrary, as $\tilde{l}_1(\hat{\boldsymbol{\theta}}, \boldsymbol{t})$ is L-superadditive for variables $\hat{\theta}_i$ and t_i, it follows that $\tilde{l}_1(\hat{\boldsymbol{\theta}}, \boldsymbol{t})$ is AI according to the Lemma 5. We find that $\tilde{l}_2(\hat{\boldsymbol{\alpha}}, \boldsymbol{r})$ is AI in the same the manner. Then, from the property of the AI functions described in (13), it holds that

$$\begin{aligned} \tilde{l}_1(\hat{\boldsymbol{\theta}}_\downarrow, \boldsymbol{t}_\uparrow) &\leq \tilde{l}_1(\hat{\boldsymbol{\theta}}^*, \boldsymbol{t}_\uparrow) \leq \tilde{l}_1(\hat{\boldsymbol{\theta}}_\uparrow, \boldsymbol{t}_\uparrow), \\ \tilde{l}_2(\hat{\boldsymbol{\alpha}}_\downarrow, \boldsymbol{r}_\uparrow) &\leq \tilde{l}_2(\hat{\boldsymbol{\alpha}}^*, \boldsymbol{r}_\uparrow) \leq \tilde{l}_2(\hat{\boldsymbol{\alpha}}_\uparrow, \boldsymbol{r}_\uparrow). \end{aligned}$$

As $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\alpha}}^*$ are the estimates that maximize $\tilde{l}_1(\hat{\boldsymbol{\theta}}, \boldsymbol{t}_\uparrow)$ and $\tilde{l}_2(\hat{\boldsymbol{\alpha}}, \boldsymbol{r}_\uparrow)$ respectively, it follows that $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}_\uparrow$ and $\hat{\boldsymbol{\alpha}}^* = \hat{\boldsymbol{\alpha}}_\uparrow$.

Conversely, if we set $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}_\uparrow$ and $\hat{\boldsymbol{\alpha}}^* = \hat{\boldsymbol{\alpha}}_\uparrow$, we find that $\tilde{l}_1(\hat{\boldsymbol{\theta}}, \boldsymbol{t}_\uparrow)$ and $\tilde{l}_2(\hat{\boldsymbol{\alpha}}, \boldsymbol{r}_\uparrow)$ reach their maximum because $\tilde{l}_1$ and $\tilde{l}_2$ are AI. Then, it is shown that $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\alpha}}^*$ maximize $\log \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}|\boldsymbol{t}_\uparrow, \boldsymbol{r}_\uparrow)$. □

Our results in Proposition 6 hold in related models such as the LRSM, PCM and LPCM. The approximated likelihood functions corresponding to (11) in the the LRSM are

$$\tilde{L}(\boldsymbol{\theta}, \boldsymbol{\eta}|X) = \frac{\exp(\sum_{i=1}^n \theta_i t_i + \sum_{p=1}^q \eta_p r'_p)}{\tilde{C}(\boldsymbol{\theta}, \boldsymbol{\eta})}, \tag{16}$$

Here, $\tilde{C}(\boldsymbol{\theta}, \boldsymbol{\eta}) = m^{nk} \prod_{i=1}^n \prod_{j=1}^k \sqrt[m]{\prod_{h=1}^m \exp(w_h\theta_i + \sum_{p=1}^q v_{jhp}\eta_p)}$. In (16), we focus on

$$\sum_{i=1}^n \theta_i t_i + \sum_{p=1}^q \eta_q r'_q = \tilde{l}_1(\boldsymbol{\theta}, \boldsymbol{t}) + \tilde{l}_2(\boldsymbol{\eta}, \boldsymbol{r}'),$$

Then, below Proposition 6 holds.

Proposition 7. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$ be the maximum likelihood estimates for the log likelihood function $\log \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{t}, \boldsymbol{r}')$ from (2). Then, $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\eta}}^*$ maximize $\log \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{t}_\uparrow, \boldsymbol{r}'_\uparrow)$ if and only if $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}}_\uparrow$ and $\hat{\boldsymbol{\eta}}^* = \hat{\boldsymbol{\eta}}_\uparrow$.

Proof. The proof is done in the same way as in Proposition 1. We assume that$\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$ are already given. $\tilde{l}_1(\boldsymbol{\theta}, \boldsymbol{t})$ and $\tilde{l}_2(\boldsymbol{\eta}, \boldsymbol{r}')$ are permutation-invariant. Then, we find that both $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\eta}}^*$ are rearranged forms of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$, respectively. As $\tilde{l}_1(\boldsymbol{\theta}, \boldsymbol{t})$ and $\tilde{l}_2(\boldsymbol{\eta}, \boldsymbol{r}')$ are AI, l_1 and l_2 reach the maximum when $\tilde{l}_1(\boldsymbol{\theta}_\uparrow, \boldsymbol{t}_\uparrow)$ and $\tilde{l}_2(\boldsymbol{\eta}_\uparrow, \boldsymbol{r}'_\uparrow)$. Consequently, $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}}_\uparrow$ and $\hat{\boldsymbol{\eta}}^* = \hat{\boldsymbol{\eta}}_\uparrow$. Conversely, if we set $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}}_\uparrow$ and $\hat{\boldsymbol{\eta}}^* = \hat{\boldsymbol{\eta}}_\uparrow$, it holds that $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\eta}}^*$ maximize $\log \tilde{L}(\boldsymbol{\theta}, \boldsymbol{\eta}|\boldsymbol{t}_\uparrow, \boldsymbol{r}'_\uparrow)$ because l_1 and l_2 are AI. □

The approximated likelihood functions in the PCM and LPCM are

$$\tilde{L}(\boldsymbol{\theta},\boldsymbol{\beta}|\boldsymbol{t}^*,\boldsymbol{r}) = \frac{\exp(\sum_{i=1}^n \theta_i t_i^* + \sum_{j=1}^k \sum_{h=1}^m \beta_{jh} r_{jh})}{\tilde{C}(\boldsymbol{\theta},\boldsymbol{\beta})}, \tag{17}$$

$$\tilde{L}(\boldsymbol{\theta},\boldsymbol{\gamma}|\boldsymbol{t}^*,\boldsymbol{r}^*) = \frac{\exp(\sum_{i=1}^n \eta_i t_i^* + \sum_{p=1}^q \gamma_p r_p^*)}{\tilde{C}(\boldsymbol{\theta},\boldsymbol{\gamma})}. \tag{18}$$

We also find that below Corollary 8 and 9 from these likelihood functions hold.

Corollary 8. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\beta}}$ be the maximum likelihood estimates for the log likelihood function $\log \tilde{L}(\boldsymbol{\theta},\boldsymbol{\beta}|\boldsymbol{t}^*,\boldsymbol{r})$ from (17). Then, $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\beta}}^*$ maximize $\log \tilde{L}(\boldsymbol{\theta},\boldsymbol{\beta}|\boldsymbol{t}^*_{\uparrow},\boldsymbol{r}_{\uparrow})$ if and only if $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}}_{\uparrow}$ and $\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}}_{\uparrow}$.

Corollary 9. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}}$ be the maximum likelihood estimates for the log likelihood function $\log \tilde{L}(\boldsymbol{\theta},\boldsymbol{\gamma}|\boldsymbol{t}^*,\boldsymbol{r}^*)$ from (18). Then, $\hat{\boldsymbol{\theta}}^*$ and $\hat{\boldsymbol{\gamma}}^*$ maximize $\log \tilde{L}(\boldsymbol{\theta},\boldsymbol{\gamma}|\boldsymbol{t}^*_{\uparrow},\boldsymbol{r}^*_{\uparrow})$ if and only if $\hat{\boldsymbol{\theta}}^* = \hat{\boldsymbol{\theta}}_{\uparrow}$ and $\hat{\boldsymbol{\gamma}}^* = \hat{\boldsymbol{\gamma}}_{\uparrow}$.

3 Simulation studies In the next step, we evaluate the ranges within which approximation (10) holds in the RSM by using simulation studies. We set $n = 50, 100$, $k = 10, 20, 30$, and $m = 3, 5, 7, 9, 11$ and generate parameters from the settings below:

$$\begin{aligned} &\theta_i \sim N(0,1^2),\ \ \alpha_{jh} \sim N(0,1^2), w_h = \omega_h + 1, \omega_h \sim [N(0,2^2)]^+ \\ &i = 1,2,\cdots,n, j = 1,2,\cdots,k, h = 1,2,\cdots,m, \end{aligned} \tag{19}$$

where N denotes a normal distribution. Here, $[a]^+$ is a positive part of real value a, which means $[a]^+ = a$ with $a > 0$ and $[a]^+ = 0$ with $a \leq 0$. These settings are practical for educational testing. Then, we generate response matrix X, statistics t_i, and y_{jh} from (1). We also calculate a kind of "capacity factor" that is $CF = DR(\boldsymbol{y})/S(z)$ for the approximation (10). Here, $DR(\boldsymbol{y})$ and $S(z)$ are defined in (9) and (7), respectively. Finally, we evaluate Kendall's rank correlation coefficients for $(\boldsymbol{t},\boldsymbol{\theta})$ and $(\boldsymbol{y},\boldsymbol{\alpha})$ as efficiency indexes for the difference in approximation (10). We repeat the procedure above 1000 times.

Table 1, Table 2, and Table 3 show the medians of Kendall's correlations for $(\boldsymbol{t},\boldsymbol{\theta})$ and $(\boldsymbol{y},\boldsymbol{\alpha})$. First, all the correlation coefficients for $(\boldsymbol{t},\boldsymbol{\theta})$ are quite high and stable because each t_i is a sufficient statistic for θ_i, as Andersen(1996) pointed out. We also find that the correlation coefficients for $(\boldsymbol{y},\boldsymbol{\alpha})$ are high and depend on the size of category m. In other words, the correlations for $(\boldsymbol{y},\boldsymbol{\alpha})$ worsen as m increases. In section 1, we found that the least difference of the upper limit for the approximation (10) was evaluated by $S(z)$ in (7) and that $S(z)$ only depends on the maximum and minimum values of the elements $(y_1, y_2, \cdots, y_m)$. Indeed, the correlation coefficients actually decrease with the size of category m under usual conditions, although the least difference corresponding to $S(z)$ in (7) does not depend on a such criterion. This is because the differences in (10) are relatively better than the least differences, as evaluated below.

Table 4, Table 5, and Table 6 present the medians of CFs. All of the CFs are very small, which means that the DRs are very small compared with the least differences. Thus, our approximation works well in these settings. Then, we evaluate more detail of the the CFs. For each k and m, the CF increases with the size of category m. This finding means that the difference by (10) worsens as m becomes large. This result is consistent with the decreasing of the correlation coefficients denoted above.

Finally, we conclude that approximation (10) shows relatively strong performance and that this approximation and the order-preserving statistics $\boldsymbol{t}$ and $\boldsymbol{y}$ are acceptable in typical educational testing.

k	10									
n	50					100				
m	3	5	7	9	11	3	5	7	9	11
Cor $(\boldsymbol{t}, \boldsymbol{\theta})$	0.879	0.908	0.914	0.916	0.913	0.888	0.908	0.916	0.920	0.920
Cor$(\boldsymbol{y}, \boldsymbol{\alpha})$	1	0.847	0.831	0.826	0.824	1	0.872	0.851	0.846	0.851

Table 1: Medians of Kendall's correlations for $(\boldsymbol{t}, \boldsymbol{\theta})$ and $(\boldsymbol{y}, \boldsymbol{\alpha})$ $(L = 10)$

k	20									
n	50					100				
m	3	5	7	9	11	3	5	7	9	11
Cor $(\boldsymbol{t}, \boldsymbol{\theta})$	0.935	0.948	0.950	0.952	0.955	0.941	0.953	0.955	0.954	0.953
Cor$(\boldsymbol{y}, \boldsymbol{\alpha})$	1	0.872	0.831	0.828	0.822	1	0.872	0.852	0.849	0.852

Table 2: Medians of Kendall's correlations for $(\boldsymbol{t}, \boldsymbol{\theta})$ and $(\boldsymbol{y}, \boldsymbol{\alpha})$ $(L = 20)$

k	30									
n	50					100				
m	3	5	7	9	11	3	5	7	9	11
Cor $(\boldsymbol{t}, \boldsymbol{\theta})$	0.954	0.963	0.964	0.966	0.966	0.956	0.966	0.967	0.968	0.969
Cor$(\boldsymbol{y}, \boldsymbol{\alpha}$	1	0.872	0.838	0.827	0.820	1	0.872	0.850	0.846	0.852

Table 3: Medians of Kendall's correlations for $(\boldsymbol{t}, \boldsymbol{\theta})$ and $(\boldsymbol{y}, \boldsymbol{\alpha})$ $(L = 30)$

k	10									
n	50					100				
m	3	5	7	9	11	3	5	7	9	11
CF	2.87×10^{-6}	3.70×10^{-6}	4.99×10^{-6}	5.95×10^{-6}	9.29×10^{-6}	4.15×10^{-7}	5.59×10^{-7}	7.71×10^{-7}	9.14×10^{-7}	1.14×10^{-7}

Table 4: Medians of the CFs for approximation (10) $(L = 10)$

k	20									
n	50					100				
m	3	5	7	9	11	3	5	7	9	11
CF	1.95×10^{-6}	4.18×10^{-6}	3.69×10^{-6}	7.75×10^{-6}	7.92×10^{-6}	2.98×10^{-7}	6.92×10^{-7}	7.32×10^{-7}	1.08×10^{-6}	1.45×10^{-7}

Table 5: Medians of the CFs for approximation (10) $(L = 20)$

k	30									
n	50					100				
m	3	5	7	9	11	3	5	7	9	11
CF	2.99×10^{-6}	3.46×10^{-6}	5.64×10^{-6}	6.16×10^{-6}	8.92×10^{-6}	1.94×10^{-7}	4.95×10^{-7}	6.48×10^{-7}	7.72×10^{-7}	1.03×10^{-6}

Table 6: Medians of the CFs for approximation (10) $(L = 30)$

4 Conclusion and discussion In this study, we considered the ordering properties of the RSM and related polytoumous IRT models in JMLE with approximation (10). We also evaluated the difference in such an approximation by using simulation study and concluded that this approximation and the order-preserving statistics proposed in this study are acceptable in typical educational testing.

For the other estimation methods in RSM, namely CMLE and MMLE, the order-preserving statistics concur with those in JMLE when (10) holds. First we consider the relations between the estimates in JMLE and CMLE. The estimates in JMLE are biased comparing with those in CMLE (e.g. Andersen,1980, Theorem 6.1) and the bias is positive. Thus, the ordering of estimates in JMLE and CMLE concur, although the estimates in JMLE are biased. Consequently, the order-preserving statistics in JMLE agree with those in CMLE.

Then, we consider the relations between the estimates in CMLE and MMLE. Andersen(1996) found that CMLE and MMLE agree when n is large, that means that estimates in the CMLE and the MMLE concur. Thus, it is clear that the ordering of estimates and the order-preserving statistics in these estimations concur. Finally, we find that the order-preserving statistics in all three estimations agree.

Acknowledgments The author thanks the referee for careful reading of the paper and constructive comments. This work was supported by JSPS Grant-in-Aid for Young Scientists (B) Grant Number 26750075.

REFERENCES

[1] Andrich, D. (1978a). A rating formulation for ordered response categories. *Psychometrika*, **43**, 561-573.

[2] Andrich, D. (1978b). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied psychological measurement*, **2**, 581-594.

[3] Andersen, E.B. (1972). The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society: Series B*, vol. **34**, 42-54.

[4] Andersen, E.B. (1980). *Discrete statistical models with social science applications*. North-Holland publishing company.

[5] Andersen, E.B. (1996). The rating scale model. In van der Linden, W.J. and Hambleton, R.K. (Eds.) *Handbook of modern item response theory*. 67-84. Springer.

[6] Bock, R.D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, vol. **45**, 179-197.

[7] Boland,P.J. and Proschan, F.(1988). Multivariate arrangement increasing functions with applications in probability and statistics. *Journal of Multivariate Analysis*, vol. **25**, 286-298.

[8] Fischer, G.H. and Parzer, P. (1991). An Extension of the Rating Scale Model with an Application to the Measurement of Change. *Psychometrika*, **56**, 637-651.

[9] Fischer G.H. and Ponocny, I. (1994). An Extension of the Partial Credit Model with an Application to the Measurement of Change. *Psychometrika*, **59**, 177-192.

[10] Glas, C.A.W. and Verhelst, N. (1989). Extensions of the Partial Credit Model. *Psychometrika*, vol. **54**, 635-659.

[11] Marshall A.W., Olkin, O. and Arnold, B.C. (2011). *Inequalities:Theory of majorization and its applications 2nd edition*. Springer.

[12] Masters,G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, vol. **47**, 149-174.

[13] Mori, K. (2015). The order-preserving properties of the Rasch model and extended model in marginal maximum likelihood estimation. *Scientiae Mathematicae Japonicae online*,**e-2015**, 41-47.

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

[15] Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, vol. **47**, 175-185.

[16] Seo, Y.(2000). Specht's ratio and complementary inequalities to Golden-Thompson type inequalities on the Hadamard product. *Linear Algebra and its Applications*, **320**, 15-22.

[17] Specht, W. (1960). Zur theorie der elementaren Mittel. *Math. Z.*, **74**,91-98.

Communicated by *Koyu Uematsu*