THE ORDER-PRESERVING PROPERTIES OF THE RASCH MODEL AND EXTENDED MODEL IN MARGINAL MAXIMUM LIKELIHOOD ESTIMATION

KAZUMASA MORI

Received May 13, 2014 ; revised November 22, 2014

ABSTRACT. In this study, we consider the order-preserving properties of Rasch model (Rasch, 1960) and linear logistic model(Fischer, 1994) in marginal maximum likelihood estimation (MMLE). More specially, we focus on the "manifest probability," as discussed by Cressie and Holland (1983) and derive the order-preserving statistics for the item parameters. We also derive order-preserving statistics for the ability parameters in maximum likelihood estimation under the condition that the estimates of the item parameters are already given. Both sets of statistics are derived using the characteristics of arrangement increasing functions (Hollander *et al.*, 1977, Marshall *et al.*, 2011). It is notable that the order-preserving statistics of the Rasch model in MMLE coincide with those of other estimation techniques, such as joint maximum likelihood estimation and conditional maximum likelihood estimation. However, while the marginal maximum likelihood estimates and the conditional maximum likelihood estimates are consistent, the joint maximum likelihood estimates are not. Here, we discuss the reasons for such coincidences, as well as the types of bias that occur in inconsistent estimates.

1 Introduction In this study, we consider the ordering properties of Rasch model (Rasch, 1960) and linear logistic model(Fischer, 1994) in marginal maximum likelihood estimation.

First, we introduce the Rasch model. Suppose a test comprises k items administered to n examinees. Let $X_{ij} = \{0,1\}$ be the response of the i-th examinee to the j-th item. When the i-th examinee responds with a 1 to the j-th item, the corresponding probability is

$$P_{ij}(\theta_i,\beta_j) = P(X_{ij}=1;\boldsymbol{\theta},\boldsymbol{\beta}) = \frac{\exp(\theta_i+\beta_j)}{1+\exp(\theta_i+\beta_j)}, \tag{1}$$

where θ_i is the ability parameter for the i-th examinee and β_j is the item parameter for the j-th item. In addition, $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_n)$ is an n-dimensional vector of ability parameters and $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_k)$ is a k-dimensional vector of item parameters. One of major estimation methods for the Rasch model is maximum likelihood estimation. In the Rasch model, the form of the likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta},\boldsymbol{\beta}|X) &= \prod_{i=1}^{n}\prod_{j=1}^{k}\left\{P_{ij}(\theta_i,\beta_j)^{x_{ij}}Q_{ij}(\theta_i,\beta_j)^{1-x_{ij}}\right\} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{k}\frac{\exp\left\{(\theta_i+\beta_j)\right\}^{x_{ij}}}{1+\exp(\theta_i+\beta_j)} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{k}\frac{\exp\left\{x_{ij}(\theta_i+\beta_j)\right\}}{1+\exp(\theta_i+\beta_j)}, \end{aligned} \tag{2}$$

where X represents a matrix of all responses for the test, x_{ij} is the observed response of the i-th examinee to the j-th item, and $Q_{ij}(\theta_i,\beta_j) = 1 - P_{ij}(\theta_i,\beta_j)$.

2010 *Mathematics Subject Classification.* 62P15,62B05.

Key words and phrases. Item response theory, Marginal maximum likelihood estimates, Order-preserving property, Arrangement increasing function.

Three maximum likelihood estimation techniques have been proposed, all of which use (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). The JMLE technique estimates $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ simultaneously by maximizing (2). In contrast, the CMLE and the MMLE techniques remove $\boldsymbol{\theta}$ from (2) and estimate $\boldsymbol{\beta}$ separately. Holland (1990) discussed the relationship among these estimation techniques. He compared the log-likelihood functions of the three techniques and concluded that JMLE and CMLE can both be viewed as approximations to MMLE. In other words, we can regard MMLE as being more general than JMLE and CMLE. On the other hand, Grayson (1988) and Huynh (1994) presented their basic results as the monotone likelihood ratio for the order-preserving property of the dichotomous response model. In addition, Bertoli-Barsotti (2003) derived the order-preserving property for the Rasch model using JMLE and CMLE, but not MMLE. Thus, in this study, we focus on the order-preserving property of the Rasch model based on MMLE.

In MMLE, we remove the ability parameter from the likelihood function (2) by integration. Cressie and Holland (1983) discussed the "manifest probability" for the Rasch model. The manifest probability can be obtained by integrating the ability parameter, θ, for each examinee. Thus, it corresponds to the marginal likelihood for each examinee. The form is

(3) $$p(\boldsymbol{x}) = \int \prod_j \left[P_j(\theta, \beta_j)^{x_j} \{1 - P_j(\theta, \beta_j)\}^{1-x_j} \right] dF(\theta),$$

where x_j is observed response for j-th item, $\boldsymbol{x} = (x_1, x_2, \cdots, x_k)$, $F(\theta)$ is the distribution function for θ and $P_j(\theta, \beta_j) = \dfrac{\exp(\theta + \beta_j)}{1 + \exp(\theta + \beta_j)}$. They also derived the log-likelihood function for the Rasch model (1). Here, the form is

(4) $$\ln L(\boldsymbol{\beta}|X) = c + n\alpha + \sum_j^k s_j \beta_j + \sum_t^k r_t \gamma(t),$$

where

$c = \log \dfrac{n!}{\prod_{\boldsymbol{x}} m(\boldsymbol{x})!}$,

$m(\boldsymbol{x})$ is the number of examinees whose item response vector is $\boldsymbol{x}$,

s_j is the number of examinees who answered 1 to the j-th item,

r_t is the number of examinees who answered t items as 1 on the test,

$\alpha = \ln p(\mathbf{0})$, where $\mathbf{0}$ is a k-dimensional vector all of whose elements are 0,

$\gamma(t) = \log \int_0^\infty u^t dG(u)$, with translation $u = \exp(\theta)$,

and $G(u)$ is a distribution function constructed from $dG(u)$ with

$$dG(u) = \frac{\exp \theta dF(\theta)}{p(\mathbf{0}) \prod_j^k \{1 + \exp(\theta + \beta_j)\}}.$$

One of extension of the Rasch model is linear logistic test model (LLTM, Fischer,1994). The LLTM is defined by adding below conditions

(5) $$\beta_j = \sum_{l=1}^p w_{jl} \delta_l$$

to (1). Here, $\delta_l, l = 1, \ldots, p$ are basic parameter of the LLTM and w_{jl} are given weights for the basic parameters δ_l.

For the MMLE of the LLTM, we substitute (5) and maximum likelihood estimate of α,

$$\hat{\alpha} = \ln(r_0/n)$$

into (4). Such modification of likelihood function was also evaluated in Tjur(1982) and Andersen(1997). Then, (4) is modified as

$$\ln L(\boldsymbol{\delta}|X) = c + n\hat{\alpha} + \sum_{l}^{p} v_l \delta_l + \sum_{t}^{k} r_t \gamma(t) = \ln L(\boldsymbol{\delta}|\boldsymbol{v}), \tag{6}$$

where $v_l = \sum_i^n \sum_j^k x_{ij} w_{jl}$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_p)$.

In this study, we use the log-likelihood function (4) and (6) to derive the order-preserving properties of the MMLE technique, as well as in related maximum likelihood estimation techniques. We use the characteristics of an arrangement increasing function (Hollander *et al.*, 1977, Marshall *et al.*, 2011) for deviations among the order-preserving properties. In our results, we assume that the maximum likelihood estimates described above exist and are unique. These assumptions are related to the form of the response matrix X and the rank of weight matrix$W = [w_{jl}]$.(for details, see Fischer (1981,1994)).

The remainder of the paper is organized as follows. The preliminaries and main theorems are presented in section 2. Finally, section 3 discusses our results and concludes the paper.

2 Preliminaries and the main results As mentioned previously, we use some characteristics of arrangement increasing (AI) functions (Hollander *et al.*, 1977) to derive the order-preserving properties of the Rasch model and the LLTM. To begin with, we introduce some definitions, as per Marshall *et al.*(2011) and Boland and Proschan(1988).

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.

Clearly, we find $(\boldsymbol{a}, \boldsymbol{b}) \stackrel{a}{=} (\boldsymbol{a}_\uparrow, \boldsymbol{b}\Pi_1) \stackrel{a}{=} (\boldsymbol{a}_\downarrow, \boldsymbol{b}\Pi_2)$, 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 the vectors with components of $\boldsymbol{a}$ arranged in ascending order and descending order, respectively.

Then, we define a partial order $\stackrel{a}{\leq}$ for vector arguments. This definition corresponds to special case denoted by Boland and Proschan(1988).

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

$$(\boldsymbol{a}, \boldsymbol{b}) \stackrel{a}{=} (\boldsymbol{a}_\uparrow, \boldsymbol{b}'). \tag{7}$$

Here, $\boldsymbol{b}' = \boldsymbol{b}\Pi_1$ and Π_1 is 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 (7) by interchanging the l-th and the m-th component ($l < m$) of $\boldsymbol{b}$ such that $b_l > b_m$. Finally, we define the partial order $\stackrel{a}{\leq}$ as

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

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

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

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

Definition 3. An AI function is a function, g, with two n-dimensional vector arguments that preserves 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

$$g(\boldsymbol{a}_\uparrow, \boldsymbol{b}_\downarrow) = g(\boldsymbol{a}_\downarrow, \boldsymbol{b}_\uparrow) \leq g(\boldsymbol{a}, \boldsymbol{b}) \leq g(\boldsymbol{a}_\uparrow, \boldsymbol{b}_\uparrow) = g(\boldsymbol{a}_\downarrow, \boldsymbol{b}_\downarrow) \tag{8}$$

for AI function g, which describes the same case as the partial order $\overset{a}{\leq}$.

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

Lemma 1. (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.

In Lemma 1, L-superadditive is the function that satisfies

$$\frac{\partial}{\partial a \partial b} \phi(a, b) \geq 0. \tag{9}$$

On the other hand, when we consider the log likelihood function in (4), we find that c, n and $\sum_t^k r_t \gamma(t)$ do not include item parameter β_j. Also, we find that

$$\alpha = \ln p(\mathbf{0}) = \ln \int \prod_j \frac{1}{1 + \exp(\theta_i + \beta_j)} dF(\theta)$$

is invariant for rearrangement within $\boldsymbol{\beta}$. Thus, for considering the order-preserving properties, we focus on a part of $\ln L$:

$$l(\boldsymbol{s}, \boldsymbol{\beta}) = \sum_j^k s_j \beta_j, \tag{10}$$

where $\boldsymbol{s}$ is a vector consisting of $s_j (j = 1, \ldots, k)$ in (4). This means that we only need to focus on $l(\boldsymbol{s}, \boldsymbol{\beta})$ in (10) to derive $\hat{\boldsymbol{\beta}}$. Here, $\hat{\boldsymbol{\beta}}$ is a vector of maximum likelihood estimates, which maximize the log-likelihood in (4).

Now, we propose the main theorem.

Theorem 1. Let $\boldsymbol{s}^*$ be a rearranged vector such that $\boldsymbol{s}^* = \boldsymbol{s}_\uparrow$ and $\tilde{\boldsymbol{\beta}}$ be the marginal maximum likelihood estimates vector that maximizes $l(\boldsymbol{s}^*, \boldsymbol{\beta})$. Then, $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_\uparrow$.

Proof. First, we find that $l(\boldsymbol{s}, \boldsymbol{\beta})$ in (10) is permutation invariant in the sense that $l(\boldsymbol{s}, \boldsymbol{\beta}) = l(\boldsymbol{s}\Pi, \boldsymbol{\beta}\Pi)$ for any permutation matrix, Π. By this permutation invariance and the uniqueness of the marginal maximum likelihood estimates, we obtain

$$l(\boldsymbol{s}, \hat{\boldsymbol{\beta}}) = l(\boldsymbol{s}^*, \hat{\boldsymbol{\beta}}\Pi^*_s) = l(\boldsymbol{s}^*, \tilde{\boldsymbol{\beta}}),$$

where Π^*_s is a permutation matrix such that $\boldsymbol{s}\Pi^*_s = \boldsymbol{s}^*$. Thus, we find that both $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ are marginal maximum likelihood estimates, and that $\tilde{\boldsymbol{\beta}}$ is a rearranged form of $\hat{\boldsymbol{\beta}}$.

On the other hand, as $s_j\beta_j$ is L-superadditive for variables s_j and β_j, from (9), it follows that $l(\boldsymbol{s},\boldsymbol{\beta})$ is AI by the Lemma 1. Then, by the property of AI functions described in (8), it holds that

$$l(\boldsymbol{s}^*,\tilde{\boldsymbol{\beta}}_{\downarrow}) \leq l(\boldsymbol{s}^*,\tilde{\boldsymbol{\beta}}) \leq l(\boldsymbol{s}^*,\tilde{\boldsymbol{\beta}}_{\uparrow}),$$

for given $\boldsymbol{s}^*$ and $\tilde{\boldsymbol{\beta}}$. As $\tilde{\boldsymbol{\beta}}$ is the estimate that maximizes $l(\boldsymbol{s}^*,\boldsymbol{\beta})$, it follows that $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}_{\uparrow}$. Consequently, it holds that $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\uparrow}$.□

Estimating the ability parameter $\boldsymbol{\theta}$ often occurs under the condition that estimates of $\boldsymbol{\beta}$ are already given. This estimation technique corresponds to maximizing the likelihood function with given item parameters $\hat{\boldsymbol{\beta}}$in terms of $\boldsymbol{\theta}$. The form of the likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta}|\hat{\boldsymbol{\beta}},X) &= \prod_{i=1}^{n}\prod_{j=1}^{k}\frac{\exp\left\{x_{ij}(\theta_i+\hat{\beta}_j)\right\}}{1+\exp(\theta_i+\hat{\beta}_j)} = \frac{\exp\left\{(\sum_{i=1}^{n}\theta_i t_i \sum_{j=1}^{k}\hat{\beta}_j s_j\right\}}{\prod_{i=1}^{n}\prod_{j=1}^{k}\left\{1+\exp(\theta_i+\hat{\beta}_j)\right\}} \\ &= L(\boldsymbol{\theta}|\hat{\boldsymbol{\beta}},\boldsymbol{t}), \end{aligned} \tag{11}$$

where $t_i = \sum_j^k X_{ij}$ and $\boldsymbol{t} = (t_1,t_2,\ldots,t_n)$. In other words, this maximum likelihood estimate $\hat{\boldsymbol{\theta}}_i$ maximizes $L(\boldsymbol{\theta}|\hat{\boldsymbol{\beta}},\boldsymbol{t})$ in (11). We derive the order-preserving statistics for $\hat{\boldsymbol{\beta}}$.

Theorem 2. Let $\boldsymbol{t}^*$ be a rearranged vector such that $\boldsymbol{t}^* = \boldsymbol{t}_{\uparrow}$ and let $\tilde{\boldsymbol{\theta}}$ be a vector of the maximum likelihood estimates that maximizes $L(\boldsymbol{\theta}|\hat{\boldsymbol{\beta}},\boldsymbol{t}^*)$ in (11). Then, $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\uparrow}$.

Proof. This theorem is proved in the same way as Theorem 1. First, we evaluate the log-likelihood function of (11). We write this function as

$$\ln L(\boldsymbol{\theta}|\hat{\boldsymbol{\beta}},\boldsymbol{t}) = \sum_{i=1}^{n}\theta_i t_i + \eta - h(\boldsymbol{\theta},\hat{\boldsymbol{\beta}}), \tag{12}$$

where $\eta = \sum_{j=1}^{k}\hat{\beta}_j s_j$ is a constant under the condition that $\hat{\boldsymbol{\beta}}$ is given and $h(\boldsymbol{\theta},\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n}\sum_{j=1}^{k}\log\left\{1+\exp(\theta_i+\hat{\beta}_j)\right\}$ It is clear that $h(\boldsymbol{\theta},\hat{\boldsymbol{\beta}})$ is invariant for rearrangement within $\boldsymbol{\theta}$. Thus, we focus on

$$l(\boldsymbol{t},\boldsymbol{\theta}) = \sum_{i=1}^{n}\theta_i t_i \tag{13}$$

when estimating $\boldsymbol{\theta}$. Then, we find that $l(\boldsymbol{t},\boldsymbol{\theta})$ is permutation invariant, and that $\tilde{\boldsymbol{\theta}}$ is a rearranged vector of $\hat{\boldsymbol{\theta}}$. Here, $\hat{\boldsymbol{\theta}}$ is the conditional maximum likelihood estimates for $\log L(\boldsymbol{\theta}|\hat{\boldsymbol{\beta}},\boldsymbol{t})$ in (12). As $l(\boldsymbol{t},\boldsymbol{\theta})$ is L-superadditive, $l(\boldsymbol{t},\boldsymbol{\theta})$ is AI. Then, it holds that

$$l(\boldsymbol{t}^*,\tilde{\boldsymbol{\theta}}_{\downarrow}) \leq l(\boldsymbol{t}^*,\tilde{\boldsymbol{\theta}}) \leq l(\boldsymbol{t}^*,\tilde{\boldsymbol{\theta}}_{\uparrow}).$$

As $\tilde{\boldsymbol{\theta}}$ is the vector of maximum likelihood estimates that maximizes $l(\boldsymbol{t},\boldsymbol{\theta})$ in (13), and $\tilde{\boldsymbol{\theta}}$ is a rearranged vector of $\hat{\boldsymbol{\beta}}$, it follows that $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\uparrow}$.□

Analogue to the MMLE of the Rasch model, the order-preserving properties holds for the MMLE of the LLTM.

Theorem 3. Let $\boldsymbol{v}^*$ be a rearranged vector such that $\boldsymbol{v}^* = \boldsymbol{v}_{\uparrow}$ and let $\hat{\boldsymbol{\delta}}$ and $\tilde{\boldsymbol{\delta}}$ be a vector of the maximum likelihood estimates that maximizes $L(\boldsymbol{\delta}|\boldsymbol{v})$ and $L(\boldsymbol{\delta}|\boldsymbol{v}^*)$ in (6), respectively. Then, $\tilde{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}_{\uparrow}$.

Proof. As with the proof of Theorem 1 and 2, we focus on

$$l(\boldsymbol{u},\boldsymbol{\delta}) = \sum_{l}^{p} v_l\delta_l \tag{14}$$

in (6). From such permutation invariant of $l(\boldsymbol{u},\boldsymbol{\delta})$ and the existence and uniqueness of the MMLE of the LLTM, we find that $\tilde{\boldsymbol{\delta}}$ is a rearranged form of $\hat{\boldsymbol{\delta}}$. On the other hand, as $l(\boldsymbol{u},\boldsymbol{\delta})$ is AI, it holds that

$$l(\boldsymbol{v}^*,\tilde{\boldsymbol{\delta}}_{\downarrow}) \leq l(\boldsymbol{v}^*,\tilde{\boldsymbol{\delta}}) \leq l(\boldsymbol{v}^*,\tilde{\boldsymbol{\delta}}_{\uparrow}).$$

Consequently, $\tilde{\boldsymbol{\delta}}$ coincides on $\hat{\boldsymbol{\delta}}_{\uparrow}$.□

Next, we consider the case when maximum likelihood estimation $\hat{\boldsymbol{\delta}}$ is already given for the the LLTM. It is clear that the same result as the Theorem 7 holds for the LLTM. We denote this result as below corollary.

Corollary 1. Define that $L(\boldsymbol{\theta}|\hat{\boldsymbol{\delta}},\boldsymbol{t})$ is likelihood function of the LLTM provided that $\hat{\boldsymbol{\delta}}$ is already given. Let $\boldsymbol{t}^*$ be a rearranged vector such that $\boldsymbol{t}^* = \boldsymbol{t}_{\uparrow}$ and let $\tilde{\boldsymbol{\theta}}$ be a vector of the maximum likelihood estimates that maximizes $L(\boldsymbol{\theta}|\hat{\boldsymbol{\delta}},\boldsymbol{t}^*)$. Then, $\tilde{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\uparrow}$.

Lastly, we consider structurally incomplete design for the LLTM. According to Fishcer(1994), we introduce following notations:

> $B = (b_{ij})$ is an $n \times k$ design matrix. If response of j-th item by i-th examinee is presented, then $b_{ij} = 1$. Otherwise, $b_{ij} = 0$.
>
> And $x_{ij} = \{0, a, 1\}$. If $b_{ij} = 1$, then $x_{ij} = \{0, 1\}$. Otherwise ($b_{ij} = 0$) $x_{ij} = a$ with $0 < a < 1$.

Then, (6) is modified as

$$\ln L(\boldsymbol{\delta}|X) = c + n\hat{\alpha} + \sum_{l}^{p} q_l\delta_l + \sum_{t}^{k} r_t\gamma(t) = \ln L(\boldsymbol{\delta}|\boldsymbol{q}), \tag{15}$$

where $q_l = \sum_i^n \sum_j^k x_{ij}b_{ij}w_{jl}$ and $\boldsymbol{q} = (q_1, q_2, \cdots, q_p)$. From (15) we get below result as a corollary of Theorem 3.

Corollary 2. Let $\boldsymbol{q}^*$ be a rearranged vector such that $\boldsymbol{q}^* = \boldsymbol{q}_{\uparrow}$ and let $\hat{\boldsymbol{\delta}}$ and $\tilde{\boldsymbol{\delta}}$ be a vector of the maximum likelihood estimates that maximizes $L(\boldsymbol{\delta}|\boldsymbol{q})$ and $L(\boldsymbol{\delta}|\boldsymbol{q}^*)$ in (15), respectively. Then, $\tilde{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}_{\uparrow}$.

3 Discussion In this study, we examined the order-preserving property of the Rasch model and the LLTM in MMLE.

Especially, for Rasch model, our results from Theorems 1 and 2 coincide with those of Bertoli-Barsotti (2003), who focused on JMLE and CMLE. It is well known that the marginal maximum likelihood (MML) estimates and conditional maximum likelihood (CML) estimates are consistent, but that the joint maximum likelihood (JML) estimates are not (Neymann and Scott, 1948, Andersen, 1970). Nevertheless, the order-preserving statistics in the three estimation techniques coincide. This is because the biases of inconsistent estimates are positive. For example, Andersen (1980, Theorem 6.1) pointed out that the JML estimates for $\beta_1, \beta_2, \cdots, \beta_k$ have an approximate asymptotic bias of $\frac{k-1}{k}$, for infinite k, corresponding to the CML estimates. Following this result, it holds that

$$\breve{\beta}_j = \frac{k-1}{k}\hat{\hat{\beta}}_j, j = 1, 2, \cdots, k$$

for the JML estimate $\breve{\beta}_j$ and the CML estimate $\hat{\hat{\beta}}_j$. Note that the bias $\frac{k-1}{k}$ is strictly positive. Then, if it holds that $\hat{\hat{\beta}}_u \leq \hat{\hat{\beta}}_v (u \neq v)$, it also holds that $\breve{\beta}_u \leq \breve{\beta}_v$, and vice versa. Thus, the ordering of the estimates of $\boldsymbol{\beta}$ is preserved between the JML and CML estimates when k is infinite. Finally, when compared to the MML estimates, the JML estimates have positive biases.

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] 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.

[2] Andersen, E.B. (1980). *Discrete Statistical Models with Social Science Applications*. North-Holland publishing company.

[3] Andersen, E.B. (1997). The rating scale model. In Van der Linden, W.J. and Hambleton, R.K. (Eds.) *Handbook of Modern Item Response Theory*,67-83.

[4] Bertoli-Barsotti, L. (2003). An order-preserving property of maximum likelihood estimates for Rasch model. *Statistics and Probability Letters*, vol. **61**, 91-96.

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

[6] 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.

[7] Crassie, N. and Holland, P.W.(1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika*, vol. **48**, 129-141.

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

[9] Fischer, G.H. (1994). The linear logistic test model. In Fischer, G.H. and Molenaar, I.V. (Eds.) *Rasch models :foundations, recent developments and applications*. 131-155. Springer.

[10] Grayson, D.A. (1988). Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika*, vol. **53**, 383-392.

[11] Holland, P.W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, vol. **55**, 577-601.

[12] Hollander, M., Proschan, F. and Sethuraman, J. (1977). Functions decreasing in transposition and their applications in ranking problems. *Annals of Statistics*, vol. **5**, 722-733.

[13] Huynh,H. (1994). A new proof for monotone likelihood ratio for the sum of independent Bernoulli random variables. *Psychometrika*, vol. **59**, 77-79.

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

[15] Neyman, J. and Scott, E.L. (1948). Estimates based on partially consistent observations, *Econometrika*, vol. **16**, 1-32.

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

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

[18] Tjur, T.(1982). A connection between Racsh's item analysis model and multiplicative Poisson model. *Scandinavian Journal of Statistics*, vol**9**,23-30.

Communicated by *Masamori Ihara*

Kazumasa Mori
Graduate school of Arts and Sciences, The university of Tokyo,
3-8-1 Komaba, Meguro-ku, Tokyo, 153-8902 JAPAN
mailto:morikzm@nifty.com