

Smoothness-constrained Model for Nonparametric Item Response Theory

Toshiki Sato

Graduate School of Systems and Information Engineering, University of Tsukuba
1-1-1 Tennodai, Tsukuba-shi, Ibaraki 305-8577, JAPAN

Yuichi Takano

Faculty of Engineering, Information and Systems, University of Tsukuba
1-1-1 Tennodai, Tsukuba-shi, Ibaraki 305-8577, JAPAN

Abstract

This paper is concerned with the nonparametric item response theory (NIRT) for estimating item characteristic curves (ICCs) and latent abilities of examinees on educational and psychological tests. NIRT models can estimate various forms of ICCs under mild shape restrictions, such as the constraints of monotone homogeneity and double monotonicity. However, NIRT models frequently suffer from estimation instability because of the great flexibility of nonparametric ICCs. To improve the estimation accuracy, we propose a novel NIRT model constrained by monotone homogeneity and smoothness based on ordered latent classes. Our smoothness constraints avoid overfitting of nonparametric ICCs by keeping them close to logistic curves. We also implement a tailored expectation-maximization algorithm to calibrate our smoothness-constrained NIRT model efficiently. We conducted computational experiments to assess the effectiveness of our smoothness-constrained model in comparison with the common two-parameter logistic model and the monotone-homogeneity model. The computational results demonstrate that our model obtained more accurate estimation results than did the two-parameter logistic model when the latent abilities of examinees for some test items followed bimodal distributions. Moreover, our model outperformed the monotone-homogeneity model because of the effect of the smoothness constraints.

Keywords: item response theory, nonparametric estimation, smoothness constraint, optimization, EM algorithm, latent class

1 Introduction

Item response theory (IRT) is a family of statistical measurement methods for educational and psychological tests. In IRT models, the characteristics of each test item are examined based on the item characteristic curve (ICC), which expresses the probability of a correct answer as a function of the latent abilities of the examinees. Indeed, many testing companies use IRT models for the design, analysis, and scoring of tests.

This paper is focused on nonparametric item response theory (NIRT) models [9, 34, 36, 38]. In contrast to parametric item response theory (PIRT) models in which the ICCs are defined by parametric functions (e.g., logistic curves or normal ogives), NIRT models are capable of estimating various forms of ICCs under mild shape restrictions, such as the monotone-homogeneity constraint [19, 20] and the double monotonicity constraint [20, 21]. It has been demonstrated

that PIRT models do not always fit the data well [5, 8, 27], under which circumstances NIRT models have a clear advantage. They are also useful for evaluating the data quality [18] and the goodness of fit of PIRT models [5, 14, 15].

The existing methods for estimating nonparametric ICCs include regression splines [8, 26, 28, 30, 31], kernel smoothing [4, 7, 17, 27], isotonic regression [11], the finite mixture model [24], and monotonic polynomial regression [6, 13]. If the latent abilities of the examinees are represented as ordered latent classes, NIRT models are categorized as ordered latent class models [2, 3, 16, 39, 40]. The expectation–maximization (EM) algorithm [2, 3, 31, 40] and Markov chain Monte Carlo (MCMC) method [8, 10, 12, 16, 23, 39] have been employed to estimate both the nonparametric ICCs and the latent abilities of examinees. However, NIRT models frequently suffer from estimation instability because of the great flexibility of nonparametric ICCs, especially if there is only a small amount of item-response data [22].

Various shape restrictions have been proposed to prevent overfitting with nonparametric regression models [29, 35]. To improve the estimation accuracy of NIRT models, we make effective use of the smoothness constraints on the nonparametric ICCs. More specifically, we propose an NIRT model with monotone homogeneity and smoothness constraints based on the ordered latent classes. Our smoothness constraints keep each nonparametric ICC close to a logistic curve, and thus offer advantages to both PIRT and NIRT models, namely stability and flexibility. To the best of our knowledge, no existing study has incorporated such smoothness constraints into a monotone-homogeneity NIRT model. In addition, we implement a tailored EM algorithm to calibrate our smoothness-constrained NIRT model efficiently.

We conducted computational experiments to assess the effectiveness of our smoothness-constrained model in comparison with the common two-parameter logistic model and the monotone-homogeneity model. The computational results demonstrate that our model delivered the best estimation performance in many cases. In other words, the smoothness constraints were very effective in enhancing the estimation accuracy of the NIRT models.

The remainder of this paper is organized as follows. In Sect. 2, we present the monotone-homogeneity model for estimating the monotonically increasing ICCs and the ability classes of examinees. In Sect. 3, we formulate our smoothness-constrained model and EM algorithm. In Sect. 4, we report the computational results, and in Sect. 5 we conclude the paper with a brief summary of our work.

2 Monotone-homogeneity model

In this section, we pose our monotone-homogeneity model by following Takano et al. [37]. Let us denote by I a set of examinees and by J a set of dichotomously scored question items on a test. The test results are given as the binary item-response data

$$\mathbf{U} := (u_{ij})_{(i,j) \in I \times J} \in \{0, 1\}^{|I| \times |J|},$$

where $u_{ij} = 1$ if the i th examinee provided a correct answer to the j th item, or $u_{ij} = 0$ otherwise. We make the following assumptions throughout this paper.

Unidimensionality: the latent abilities of all examinees are evaluated unidimensionally;

Local Independence: item responses are conditionally independent of each other given an individual latent ability.

Additionally, the latent abilities of examinees are represented as ordered latent classes denoted by T .

The nonparametric ICCs of test items are defined by the decision variable

$$\mathbf{X} := (x_{jt})_{(j,t) \in J \times T},$$

where x_{jt} is the probability of the j th item being answered correctly by examinees in the t th ability class. These nonparametric ICCs are usually estimated subject to monotone-homogeneity constraints [19, 20], which require that the probability of a correct answer increases monotonically with ability class:

$$x_{jt} \leq x_{j,t+1} \quad ((j, t) \in J \times T), \quad (1)$$

$$0 \leq x_{jt} \leq 1 \quad ((j, t) \in J \times T). \quad (2)$$

The ability classes of examinees are represented by the decision variable

$$\mathbf{Y} := (y_{it})_{(i,t) \in I \times T},$$

where $y_{it} = 1$ if the i th examinee possesses the latent ability of the t th class, or $y_{it} = 0$ otherwise. The following constraints guarantee that only one ability class is assigned to each examinee:

$$\sum_{t \in T} y_{it} = 1 \quad (i \in I), \quad (3)$$

$$y_{it} \in \{0, 1\} \quad ((i, t) \in I \times T). \quad (4)$$

Given $\mathbf{x}_j := (x_{jt})_{t \in T}$ and $\mathbf{y}_i := (y_{it})_{t \in T}$, the probability of receiving response $u_{ij} \in \{0, 1\}$ is expressed as

$$\Pr(u_{ij} \mid \mathbf{x}_j, \mathbf{y}_i) := \prod_{t \in T} ((x_{jt})^{u_{ij}} (1 - x_{jt})^{1 - u_{ij}})^{y_{it}}.$$

From the assumption of local independence, the probability of the i th examinee giving the response $\mathbf{u}_i := (u_{ij})_{j \in J}$ is expressed as

$$\Pr(\mathbf{u}_i \mid \mathbf{X}, \mathbf{y}_i) := \prod_{j \in J} \Pr(u_{ij} \mid \mathbf{x}_j, \mathbf{y}_i).$$

Because the responses of different examinees are independent, the probability of receiving item response \mathbf{U} from all the examinees is given by

$$\begin{aligned} \Pr(\mathbf{U} | \mathbf{X}, \mathbf{Y}) &:= \prod_{i \in I} \Pr(\mathbf{u}_i | \mathbf{X}, \mathbf{y}_i) \\ &= \prod_{(i,j,t) \in I \times J \times T} ((x_{jt})^{u_{ij}} (1 - x_{jt})^{1-u_{ij}})^{y_{it}}. \end{aligned}$$

By treating \mathbf{X} and \mathbf{Y} as decision variables, the log-likelihood function is defined as follows:

$$\begin{aligned} \ell(\mathbf{X}, \mathbf{Y} | \mathbf{U}) &:= \log \Pr(\mathbf{U} | \mathbf{X}, \mathbf{Y}) \\ &= \sum_{(i,j,t) \in I \times J \times T} y_{it} (u_{ij} \log x_{jt} + (1 - u_{ij}) \log(1 - x_{jt})). \end{aligned} \quad (5)$$

Consequently, the monotone-homogeneity model estimates \mathbf{X} and \mathbf{Y} so that the log-likelihood function (5) is maximized subject to constraints (1)–(4):

$$\underset{\mathbf{X}, \mathbf{Y}}{\text{maximize}} \quad \sum_{(i,j,t) \in I \times J \times T} y_{it} (u_{ij} \log x_{jt} + (1 - u_{ij}) \log(1 - x_{jt})) \quad (6)$$

$$\text{subject to} \quad x_{jt} \leq x_{j,t+1} \quad ((j, t) \in J \times T), \quad (7)$$

$$0 \leq x_{jt} \leq 1 \quad ((j, t) \in J \times T), \quad (8)$$

$$\sum_{t \in T} y_{it} = 1 \quad (i \in I), \quad (9)$$

$$y_{it} \in \{0, 1\} \quad ((i, t) \in I \times T). \quad (10)$$

3 Smoothness-constrained model

In this section, we firstly formulate our smoothness-constrained model and then describe an EM algorithm for model estimation.

3.1 Smoothness constraints

To express our smoothness constraints, we use the logistic function

$$\lambda(w) := \frac{1}{1 + \exp(-w)} \quad (11)$$

and the additional decision variable

$$\mathbf{W} := (w_{jt})_{(j,t) \in J \times T}.$$

We then define the ICCs as $x_{jt} = \lambda(w_{jt})$; that is, $\lambda(w_{jt})$ denotes the probability of the j th item being answered correctly by examinees in the t th ability class. Because the logistic function

increases monotonically from zero to one, the monotone-homogeneity constraints on $\lambda(w_{jt})$ are written as

$$w_{jt} \leq w_{j,t+1} \quad ((j, t) \in J \times T). \quad (12)$$

The smoothness constraints on the nonparametric ICCs are posed as follows:

$$\sum_{t \in T} |w_{j,t+2} - 2w_{j,t+1} + w_{jt}| \leq \gamma \quad (j \in J), \quad (13)$$

where $\gamma \geq 0$ is a user-defined parameter. If γ is sufficiently large, constraints (13) are invalidated. Conversely, if $\gamma = 0$, constraints (13) are equivalent to

$$w_{j,t+2} - w_{j,t+1} = w_{j,t+1} - w_{jt} \quad ((j, t) \in J \times T), \quad (14)$$

which imply that for each $j \in J$, $\{w_{jt} \mid t \in T\}$ is a set of equally spaced points.

Figure 1 illustrates three examples of smoothness-constrained ICCs with $\gamma = 0$, where the figures on the left-hand side are graphs of the logistic function, and those on the right-hand side are the corresponding ICCs. It is clear that an S-shaped ICC can be created as shown in Fig. 1(a). In addition, although $w_{j1}, w_{j2}, \dots, w_{j5}$ must be equally spaced points, the difficulty and discrimination of a test item can be adjusted. For instance, Fig. 1(b) and (c) correspond to difficult and undiscriminating items, respectively. These examples demonstrate that the smoothness constraints (13) keep each ICC close to a logistic curve with two parameters, namely difficulty and discrimination. Therefore, our smoothness constraints provide benefits to both PIRT and NIRT models in the sense that the shapes of nonparametric ICCs are restricted by means of parametric functions.

3.2 Formulation

We begin by substituting $x_{jt} = \lambda(w_{jt})$ into the log-likelihood function (5) as follows:

$$\sum_{(i,j,t) \in I \times J \times T} y_{it} (u_{ij} \log \lambda(w_{jt}) + (1 - u_{ij}) \log(1 - \lambda(w_{jt}))). \quad (15)$$

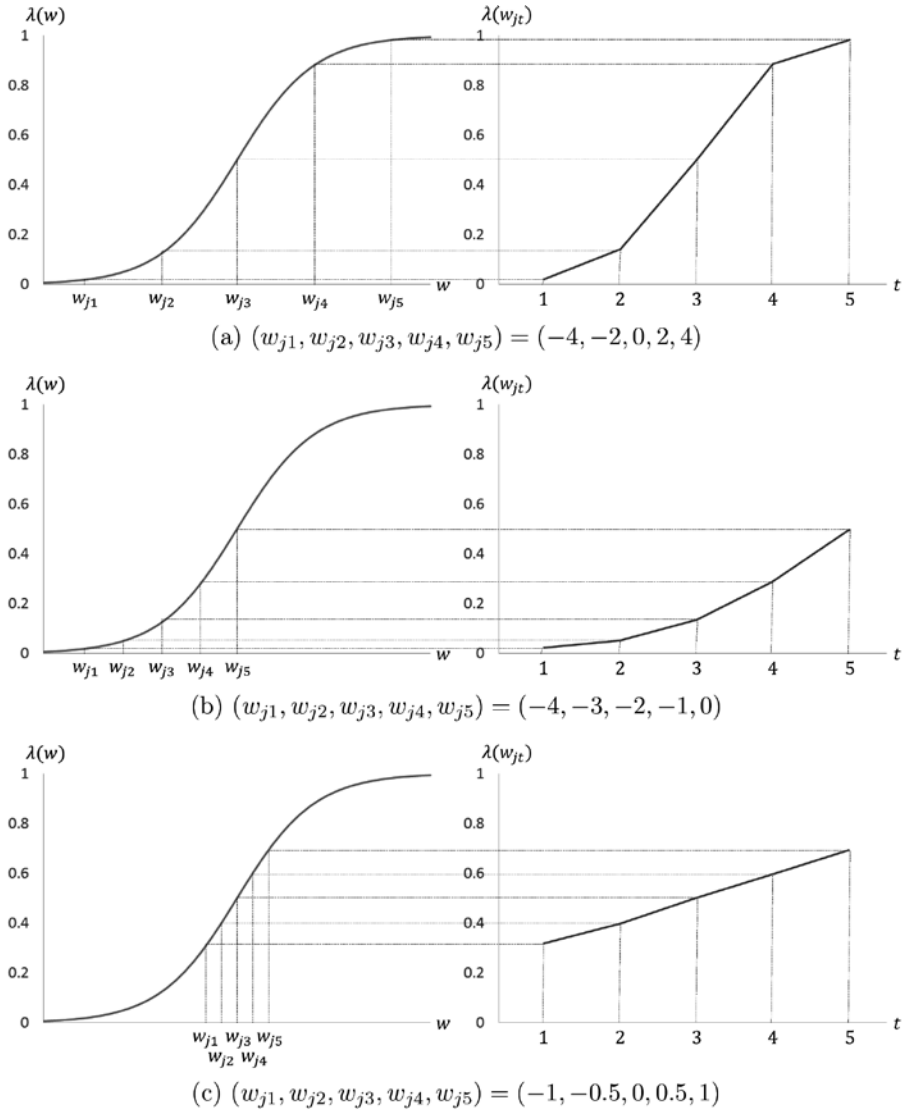
It then follows from Eq. (11) that

$$\begin{aligned} & u_{ij} \log \lambda(w_{jt}) + (1 - u_{ij}) \log(1 - \lambda(w_{jt})) \\ &= u_{ij} \log \lambda(w_{jt}) + (1 - u_{ij}) \log \lambda(-w_{jt}) \\ &= \log \lambda((2u_{ij} - 1)w_{jt}) \quad (\because u_{ij} \in \{0, 1\}) \\ &= -\log(1 + \exp((1 - 2u_{ij})w_{jt})). \end{aligned} \quad (16)$$

Therefore, maximizing the log-likelihood function (15) is equivalent to minimizing

$$\sum_{(i,j,t) \in I \times J \times T} y_{it} \log(1 + \exp((1 - 2u_{ij})w_{jt})). \quad (17)$$

We show next that the smoothness constraints (13) can be converted into a set of linear constraints.

Figure 1: Smoothness-constrained ICCs with $\gamma = 0$

Proposition 3.1. *The smoothness constraints (13) hold if and only if there exist $\mathbf{S} := (s_{jt})_{(j,t) \in J \times T}$ and $\mathbf{V} := (v_{jt})_{(j,t) \in J \times T}$ such that*

$$\sum_{t \in T} (s_{jt} + v_{jt}) \leq \gamma \quad (j \in J), \quad (18)$$

$$s_{jt} - v_{jt} = w_{j,t+2} - 2w_{j,t+1} + w_{jt} \quad ((j, t) \in J \times T), \quad (19)$$

$$s_{jt} \geq 0, \quad v_{jt} \geq 0 \quad ((j, t) \in J \times T). \quad (20)$$

Proof. Firstly, we suppose that the smoothness constraints (13) are satisfied by $\bar{\mathbf{W}} := (\bar{w}_{jt})_{(j,t) \in J \times T}$. We then set $\bar{\mathbf{S}} := (\bar{s}_{jt})_{(j,t) \in J \times T}$ and $\bar{\mathbf{V}} := (\bar{v}_{jt})_{(j,t) \in J \times T}$ as

$$\begin{cases} \bar{s}_{jt} := \bar{w}_{j,t+2} - 2\bar{w}_{j,t+1} + \bar{w}_{jt}, & \bar{v}_{jt} := 0 & \text{if } \bar{w}_{j,t+2} - 2\bar{w}_{j,t+1} + \bar{w}_{jt} \geq 0, \\ \bar{s}_{jt} := 0, & \bar{v}_{jt} := -(\bar{w}_{j,t+2} - 2\bar{w}_{j,t+1} + \bar{w}_{jt}) & \text{otherwise,} \end{cases}$$

for each $(j, t) \in J \times T$. It then follows that constraints (18)–(20) are satisfied by $\bar{\mathbf{S}}$, $\bar{\mathbf{V}}$, and $\bar{\mathbf{W}}$ because

$$\sum_{t \in T} (\bar{s}_{jt} + \bar{v}_{jt}) = \sum_{t \in T} |\bar{w}_{j,t+2} - 2\bar{w}_{j,t+1} + \bar{w}_{jt}| \leq \gamma \quad (j \in J).$$

Conversely, we suppose that constraints (18)–(20) are satisfied by $\bar{\mathbf{S}}$, $\bar{\mathbf{V}}$, and $\bar{\mathbf{W}}$. The smoothness constraints (13) are then satisfied as follows:

$$\sum_{t \in T} |\bar{w}_{j,t+2} - 2\bar{w}_{j,t+1} + \bar{w}_{jt}| = \sum_{t \in T} |\bar{s}_{jt} - \bar{v}_{jt}| \leq \sum_{t \in T} (\bar{s}_{jt} + \bar{v}_{jt}) \leq \gamma \quad (j \in J).$$

□

Consequently, our smoothness-constrained model minimizes Eq. (17) subject to constraints (3)–(4), (12), and (18)–(20):

$$\underset{\mathbf{S}, \mathbf{V}, \mathbf{W}, \mathbf{Y}}{\text{minimize}} \quad \sum_{(i,j,t) \in I \times J \times T} y_{it} \log(1 + \exp((1 - 2u_{ij})w_{jt})) \quad (21)$$

$$\text{subject to} \quad w_{jt} \leq w_{j,t+1} \quad ((j, t) \in J \times T), \quad (22)$$

$$\sum_{t \in T} (s_{jt} + v_{jt}) \leq \gamma \quad (j \in J), \quad (23)$$

$$s_{jt} - v_{jt} = w_{j,t+2} - 2w_{j,t+1} + w_{jt} \quad ((j, t) \in J \times T), \quad (24)$$

$$s_{jt} \geq 0, v_{jt} \geq 0 \quad ((j, t) \in J \times T), \quad (25)$$

$$\sum_{t \in T} y_{it} = 1 \quad (i \in I), \quad (26)$$

$$y_{it} \in \{0, 1\} \quad ((i, t) \in I \times T). \quad (27)$$

3.3 EM algorithm

We employ an EM algorithm to find a good-quality solution to the smoothness-constrained model (21)–(27) efficiently. To this end, we introduce the decision variable $\boldsymbol{\pi} := (\pi_t)_{t \in T}$ for the sizes of ability classes, such that

$$\sum_{t \in T} \pi_t = 1, \quad \pi_t > 0 \quad (t \in T). \quad (28)$$

The conditional probability of receiving response \mathbf{u}_i from the i th examinee, given that s/he belongs to the t th ability class (i.e., $y_{it} = 1$), is expressed as

$$f(\mathbf{u}_i \mid \mathbf{w}_t) := \prod_{j \in J} \lambda(w_{jt})^{u_{ij}} (1 - \lambda(w_{jt}))^{1 - u_{ij}}.$$

Accordingly, the marginal likelihood is calculated by a weighted sum of the form

$$\sum_{t \in T} \pi_t f(\mathbf{u}_i \mid \mathbf{w}_t).$$

The posterior probability of the i th examinee belonging to the t th ability class is given by Bayes' rule as follows:

$$\frac{\pi_t f(\mathbf{u}_i \mid \mathbf{w}_t)}{\sum_{\tau \in T} \pi_\tau f(\mathbf{u}_i \mid \mathbf{w}_\tau)}. \quad (29)$$

Moreover, the complete-data log-likelihood function based on \mathbf{Y} is formulated as follows:

$$\log \left(\prod_{(i,t) \in I \times T} (\pi_t f(\mathbf{u}_i \mid \mathbf{w}_t))^{y_{it}} \right) = \underbrace{\sum_{(i,t) \in I \times T} y_{it} \log \pi_t}_{(30a)} + \underbrace{\sum_{(i,t) \in I \times T} y_{it} \log f(\mathbf{u}_i \mid \mathbf{w}_t)}_{(30b)}. \quad (30)$$

Our EM algorithm starts with some initial estimate of the ability classes

$$\bar{\mathbf{Y}} := (\bar{y}_{it})_{(i,t) \in I \times T}.$$

To obtain $\bar{\mathbf{Y}}$, one may use the number of test items that each examinee answered correctly. The EM algorithm then repeats the E-step (expectation step) and M-step (maximization step) to maximize the log-likelihood function (30).

The M-step substitutes $\bar{\mathbf{Y}}$ into the log-likelihood function (30) and then maximizes it. We firstly consider maximizing Eq. (30a) subject to constraints (28). The method of Lagrange multipliers yields $\bar{\pi} := (\bar{\pi}_t)_{t \in T}$ defined by

$$\bar{\pi}_t \leftarrow \frac{\sum_{i \in I} \bar{y}_{it}}{|I|} \quad (31)$$

for each $t \in T$.

Next, we focus on maximizing Eq. (30b) after substituting Eq. (16) into it. This is equivalent to solving the smoothness-constrained model (21)–(27) with $\mathbf{Y} = \bar{\mathbf{Y}}$; that is, for each $j \in J$ we solve

$$\underset{s_j, v_j, w_j}{\text{minimize}} \quad \sum_{(i,t) \in I \times T} \bar{y}_{it} \log(1 + \exp((1 - 2u_{ij})w_{jt})) \quad (32)$$

$$\text{subject to} \quad w_{jt} \leq w_{j,t+1} \quad (t \in T), \quad (33)$$

$$\sum_{t \in T} (s_{jt} + v_{jt}) \leq \gamma, \quad (34)$$

$$s_{jt} - v_{jt} = w_{j,t+2} - 2w_{j,t+1} + w_{jt} \quad (t \in T), \quad (35)$$

$$s_{jt} \geq 0, v_{jt} \geq 0 \quad (t \in T). \quad (36)$$

This problem minimizes a convex function subject to linear constraints, so it can be solved exactly and efficiently by standard nonlinear optimization software. The estimates obtained at this step are denoted by

$$\bar{\mathbf{W}} := (\bar{w}_{jt})_{(j,t) \in J \times T}.$$

The E-step updates $\bar{\mathbf{Y}}$ with its expected value based on the current estimates $\bar{\mathbf{W}}$ and $\bar{\pi}$. This amounts to assigning the posterior probability (29) as follows:

$$\bar{y}_{it} \leftarrow \frac{\bar{\pi}_t f(\mathbf{u}_i \mid \bar{\mathbf{w}}_t)}{\sum_{\tau \in T} \bar{\pi}_\tau f(\mathbf{u}_i \mid \bar{\mathbf{w}}_\tau)} \quad (37)$$

for all $(i, t) \in I \times T$. The E-step and M-step are repeated until a termination condition is satisfied. Our EM algorithm for solving the smoothness-constrained model (21)–(27) is summarized as follows:

Step 0 (Initialization) Set $\bar{\mathbf{Y}}$ as an initial estimate, and go to Step 2.

Step 1 (E-Step) Update $\bar{\mathbf{Y}}$ according to Eq. (37) for $(i, t) \in I \times T$.

Step 2 (M-Step) Update $\bar{\pi}$ according to Eq. (31) for $t \in T$. Update $\bar{\mathbf{W}}$ by solving problem (32)–(36) for $j \in J$.

Step 3 (Termination Condition) Terminate the algorithm if a termination condition is satisfied. Otherwise, return to Step 1.

4 Computational experiments

The computational results reported in this section evaluate the effectiveness of our smoothness-constrained NIRT model.

4.1 Experimental design

We evaluated the estimation accuracy of IRT models through the simulation process illustrated in Fig. 2.

In Step 1, we randomly generate θ_i for $i \in I$ from a standard normal distribution. Next, we give a true ability class t_i^* for $i \in I$ on the basis of θ_i and Table 1. For instance, if $-1.29 \leq \theta_i < -0.81$, we set $t_i^* := 2$. The ranges of θ in Table 1 were determined so that each ability class is assigned to approximately the same number of examinees.

We used two types of function to create the true ICCs of test items. One was the two-parameter logistic (2PL) model

$$p_j^{2\text{PL}}(\theta) := \frac{1}{1 + \exp(-1.7a_j(\theta - b_j))}, \quad (38)$$

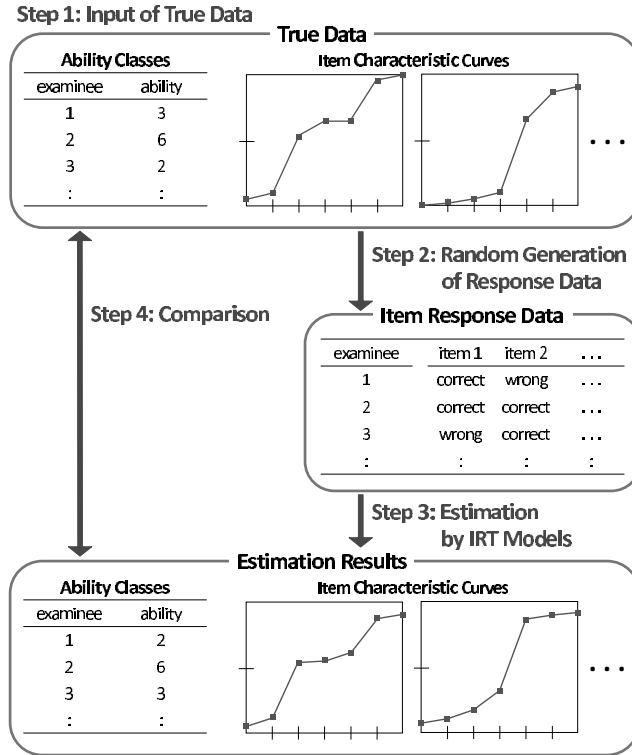


Figure 2: Simulation process of model evaluation

Table 1: Relationship between ability class $t \in T$ and normal random number θ

t	range of θ	median of θ
1	$(-\infty, -1.29)$	-1.73
2	$[-1.29, -0.81)$	-1.02
3	$[-0.81, -0.49)$	-0.64
4	$[-0.49, -0.23)$	-0.36
5	$[-0.23, 0)$	-0.12
6	$[0, 0.23)$	0.12
7	$[0.23, 0.49)$	0.36
8	$[0.49, 0.81)$	0.64
9	$[0.81, 1.29)$	1.02
10	$[1.29, \infty)$	1.73

where a_j and b_j are parameters of discrimination and difficulty, respectively. These parameters were drawn from uniform distributions for which $a_j \in [0.5, 2.0]$ and $b_j \in [-1.5, 1.5]$. Similarly to Nozawa [25], the other type of function was the extended three-parameter normal ogive (3PN) model of order two:

$$p_j^{3PN}(\theta) := \Phi(a_{j2}(\theta - b_j)^3 + \sqrt{3a_{j1}a_{j2}}(\theta - b_j)^2 + a_{j1}(\theta - b_j)), \quad (39)$$

where Φ is the normal ogive, a_{j1} and a_{j2} are shape parameters, and b_j is a difficulty parameter. These parameters were drawn from uniform distributions for which $a_{j1} \in [0.4, 0.8]$, $a_{j2} \in [0.1, 0.5]$, and $b_j \in [-0.5, 0.5]$. The 3PN model is based on the assumption that examinees' abilities follow a bimodal distribution. Accordingly, the standard two-parameter logistic IRT models have difficulty in fitting ICCs defined by the 3PN model, whereas they can accurately fit ICCs defined by the 2PL model.

When the true ICC of the j th item was defined by the 2PL model (38), it was set as $x_{j1}^* := p_j^{2PL}(-1.73)$, $x_{j2}^* := p_j^{2PL}(-1.02)$, \dots , $x_{j,10}^* := p_j^{2PL}(1.73)$ according to the median of θ in each range (see Table 1). The true ICCs defined by the 3PN model (39) were set in the same manner. The percentage of ICCs defined by the 3PN model is denoted by ρ , where $\rho \in \{0\%, 20\%, 50\%\}$ similarly to Nozawa [25]. For instance, if $|J| = 60$ and $\rho = 20\%$, true ICCs of 12 items are created by the 3PN model.

In Step 2, binary item-response data \mathbf{U} are generated randomly based on the true ability classes and ICCs specified at Step 1. Specifically, examinees in the t th ability class correctly answered the j th item with probability x_{jt}^* .

In Step 3, ability classes and ICCs are estimated from the item-response data \mathbf{U} by using the following IRT models:

2PLM: two-parameter logistic model,

MHM: monotone-homogeneity model (6)–(10),

SCM(γ): smoothness-constrained model (21)–(27) with $\gamma \in \{0, 1, 2, 4\}$.

We used the `irtoys` package in R 3.1.2 (<http://www.R-project.org>) to perform the 2PLM computations. Here, the continuous-valued ability θ estimated by 2PLM was converted into an ability class t according to Table 1 for comparison purposes. To solve MHM and SCM, we implemented the EM algorithm by using MATLAB R2015a (<https://www.mathworks.com/products/matlab.html>), in which problem (32)–(36) was solved by the `fmincon` function in the MATLAB optimization toolbox. The initial estimate $\bar{\mathbf{Y}}$ was set by dividing examinees equally into 10 ability classes according to the number of correct answers. Every time $\bar{\mathbf{Y}}$ was updated in the EM algorithm, the ability classes of examinees were determined temporarily as $\hat{t}_i := \arg \max\{\bar{y}_{it} \mid t \in T\}$ for $i \in I$. The algorithm was terminated if \hat{t}_i remained the same as the previous one for all $i \in I$.

Table 2: Root-mean-square error of item-characteristic curve estimation

$ I $	$ J $	ρ	2PLM	MHM	SCM(0)	SCM(1)	SCM(2)	SCM(4)
1000	30	0%	0.031	0.057	0.054	0.036	0.038	0.046
		20%	0.049	0.057	0.073	0.042	0.040	0.046
		50%	0.075	0.058	0.102	0.052	0.044	0.048
	60	0%	0.030	0.049	0.066	0.037	0.034	0.041
		20%	0.053	0.047	0.078	0.043	0.036	0.040
		50%	0.082	0.050	0.115	0.059	0.044	0.044
3000	30	0%	0.017	0.041	0.056	0.025	0.026	0.034
		20%	0.046	0.041	0.069	0.032	0.028	0.034
		50%	0.073	0.041	0.102	0.044	0.033	0.034
	60	0%	0.031	0.034	0.064	0.027	0.023	0.028
		20%	0.056	0.034	0.080	0.038	0.027	0.030
		50%	0.083	0.034	0.116	0.058	0.033	0.029

In Step 4, we evaluate the estimation accuracy of the IRT models by comparing the true data (Step 1) with the estimates (Step 3). Specifically, the root-mean-square error (RMSE) of the ability classes is calculated as

$$\sqrt{\frac{\sum_{i \in I} (t_i^* - \hat{t}_i)^2}{|I|}},$$

where \hat{t}_i is the estimated ability class. The RMSE of the ICCs is calculated as

$$\sqrt{\frac{\sum_{(j,t) \in J \times T} (x_{jt}^* - \hat{x}_{jt})^2}{|J||T|}},$$

where \hat{x}_{jt} is the estimated probability of a correct answer. We repeated these steps 10 times and show the average RMSEs in the following section.

4.2 Computational results

Tables 2 and 3 give the RMSEs of the ability classes and ICCs for the 12 experimental conditions. Here, the number of examinees was $|I| \in \{1000, 3000\}$, and the number of test items was $|J| \in \{30, 60\}$. Because the ordinal scale of neural test theory grades examinees into roughly 10 classes [32, 33], the number of ability classes was $|T| = 10$. Note that the minimum RMSEs for each experimental condition are given in bold face in the tables.

We firstly focus on the accuracy of ICC estimation (Table 2). It is reasonable that 2PLM estimated ICCs very accurately when the percentage of 3PN ICCs was $\rho = 0\%$. However, the

Table 3: Root-mean-square error of ability-class estimation

$ I $	$ J $	ρ	2PLM	MHM	SCM(0)	SCM(1)	SCM(2)	SCM(4)	
1000	30	0%	0.814	0.884	0.814	0.785	0.801	0.845	
		20%	0.838	0.886	0.890	0.799	0.816	0.849	
		50%	0.956	0.959	1.137	0.883	0.890	0.922	
	60	0%	0.629	0.680	0.747	0.611	0.606	0.642	
		20%	0.704	0.680	0.812	0.642	0.624	0.648	
		50%	0.914	0.715	1.176	0.740	0.676	0.700	
	3000	30	0%	0.808	0.877	0.841	0.787	0.803	0.850
			20%	0.831	0.872	0.881	0.792	0.801	0.841
			50%	0.982	0.929	1.150	0.871	0.869	0.905
60		0%	0.642	0.629	0.739	0.585	0.576	0.603	
		20%	0.747	0.642	0.828	0.638	0.599	0.621	
		50%	0.997	0.694	1.190	0.761	0.666	0.675	

estimation accuracy of 2PLM was greatly reduced by increasing the percentage of 3PN ICCs. Indeed, when $\rho \geq 20\%$, the RMSEs were often smaller for MHM than for 2PLM. We also note that the estimation accuracy of MHM increased with the amount of item-response data. For instance, the RMSEs of MHM for $(|I|, |J|) = (1000, 30)$ were at least 0.057, and those for $(|I|, |J|) = (3000, 60)$ were 0.034.

The largest RMSEs of the ICCs were those for SCM(0) in almost all cases because the shapes of the ICCs were tightly restricted by the smoothness constraints with $\gamma = 0$. In contrast, SCM(1) frequently obtained higher estimation accuracy than did MHM, especially for $\rho \leq 20\%$. The smallest RMSEs were attained by SCM(2) in many cases, whereas only for $(|J|, \rho) = (60, 50\%)$ were those provided by SCM(4).

We next move on to the accuracy of ability-class estimation (Table 3). As in Table 2, the estimation accuracy of 2PLM reduced markedly with the percentage of 3PN ICCs. In contrast, MHM made relatively accurate estimates when $(|J|, \rho) = (60, 50\%)$. Very large RMSEs were still provided by SCM(0) as in Table 2. Meanwhile, the smallest RMSEs were often obtained by SCM(1) with $|J| = 30$, and those were obtained by SCM(2) with $|J| = 60$.

The results in Tables 2 and 3 confirm that smoothness constraints are very effective in improving the estimation accuracy of NIRT models. In particular, we may say that SCM(2) delivers the best estimation performance on the whole.

Table 4 gives the computation times (in seconds) required for estimating the IRT models. The computations of 2PLM were very rapid mainly because each ICC involved only two parameters. The computations of SCM(γ) became slower as γ or ρ increased. The computation time for

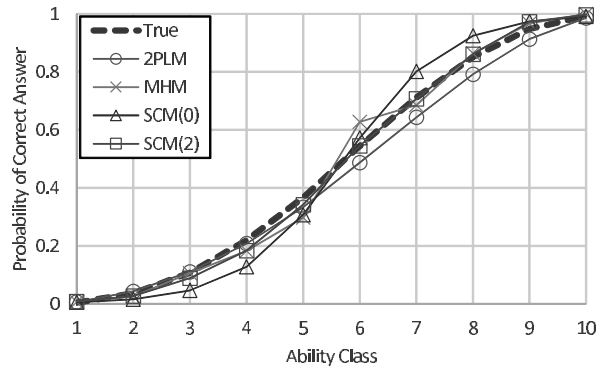
Table 4: Computation times (s)

$ I $	$ J $	ρ	2PLM	MHM	SCM(0)	SCM(1)	SCM(2)	SCM(4)
1000	30	0%	3.3	172.2	42.6	43.3	75.6	140.2
		20%	3.2	173.0	56.7	62.5	83.7	129.4
		50%	3.1	200.8	77.0	106.3	115.6	158.2
	60	0%	8.2	383.1	81.8	113.2	118.6	186.5
		20%	8.2	317.4	144.0	166.7	149.0	214.9
		50%	8.4	409.6	224.6	265.2	220.0	259.9
3000	30	0%	13.8	1140.0	88.3	128.3	310.2	597.9
		20%	14.2	1026.5	171.4	269.5	326.5	542.3
		50%	13.7	1273.7	181.6	281.2	323.0	672.6
	60	0%	36.7	1743.7	214.0	355.2	364.8	830.3
		20%	36.0	1949.8	443.8	765.2	555.2	825.2
		50%	35.8	3053.2	487.5	1008.0	898.0	1034.0

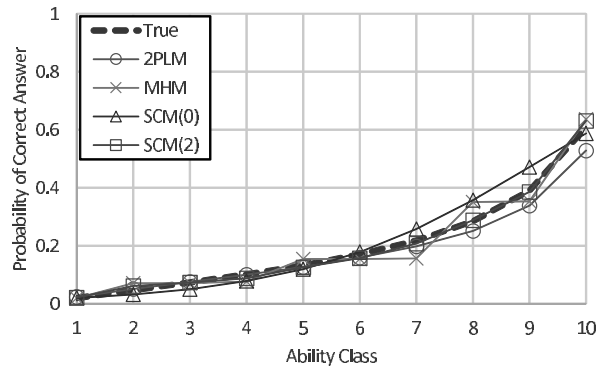
estimating MHM was always the longest among all the models. These results suggest that when many 3PN ICCs are estimated subject to loose smoothness constraints, our EM algorithm takes a relatively long time to terminate.

Figures 3 and 4 show illustrative examples of estimated ICCs together with the true ICCs. We firstly focus on Fig. 3, where $(|I|, |J|, \rho) = (3000, 60, 0\%)$, and the true ICCs were defined by the 2PL model. As expected, 2PLM fitted the true ICCs well. The MHM also fitted the true ICCs well, but it is noteworthy that the ICCs estimated by MHM moved around the true ICCs: for example, for the ability classes $t = 7, 8$, and 9 in Fig. 3(b) and $t = 3, 4$, and 6 in Fig. 3(c). The SCM(0) deviated partly from the true ICCs: for example, for the ability classes $t = 3, 4, 7$, and 8 in Fig. 3(a). In contrast, the true ICCs and those estimated by SCM(2) were almost the same because the nonparametric ICCs were made moderately less flexible by the smoothness constraints with $\gamma = 2$.

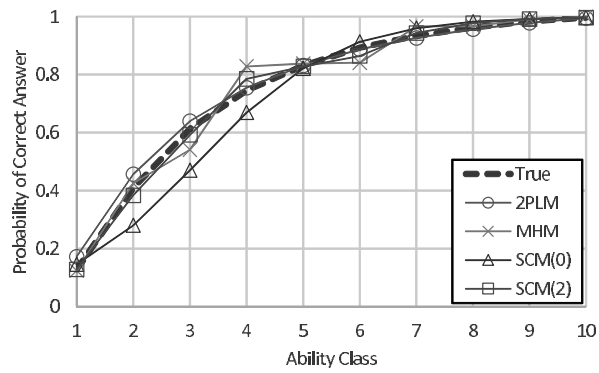
We next move on to Fig. 4, where $(|I|, |J|, \rho) = (3000, 60, 50\%)$ and the true ICCs were defined by the 3PN model. It is clear that neither 2PLM nor SCM(0) fitted the true ICC because these models create only logistic curves. In contrast, MHM and SCM(2) estimated the shapes of the true ICCs accurately, except that SCM(2) underestimated the probability of a correct answer for the ability class $t = 2$ in Fig. 4(a).



(a) test item $j = 7$

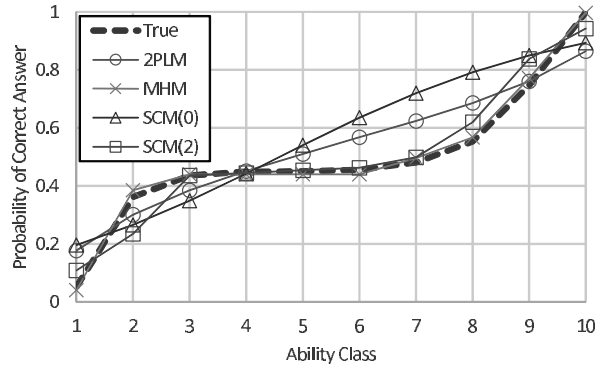


(b) test item $j = 9$

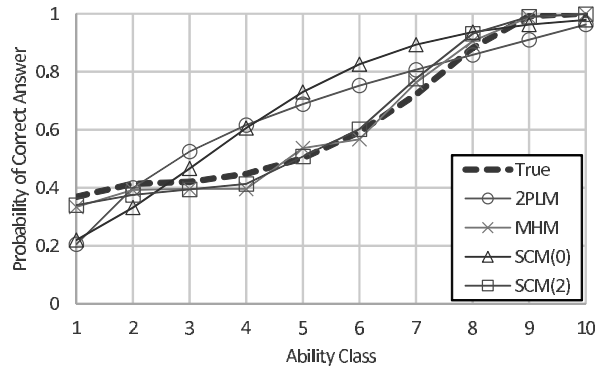


(c) test item $j = 27$

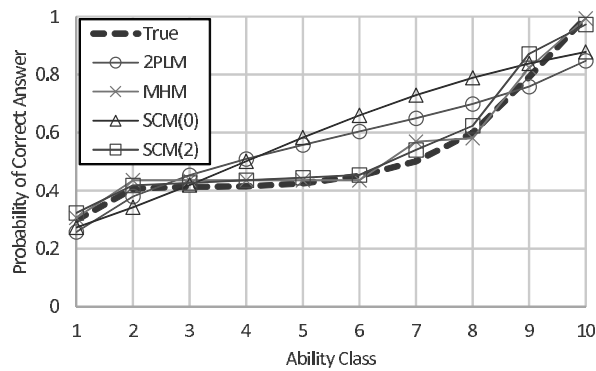
Figure 3: Estimated ICCs and true 2PL ICCs ($(|I|, |J|, \rho) = (3000, 60, 0\%)$)



(a) test item $j = 33$



(b) test item $j = 36$



(c) test item $j = 59$

Figure 4: Estimated ICCs and true 3PN ICCs ($(|I|, |J|, \rho) = (3000, 60, 50\%)$)

5 Conclusion

We devised an NIRT model with monotone homogeneity and smoothness constraints based on ordered latent classes. Our smoothness constraints avoid overfitting of nonparametric ICCs by smoothing them based on logistic curves. We also developed an EM algorithm for our smoothness-constrained NIRT model to estimate the nonparametric ICCs and the latent abilities of examinees efficiently.

The computational results demonstrated the effectiveness of our model in comparison to the two-parameter logistic model and the monotone-homogeneity model. Indeed, our model obtained more accurate estimation results than did the two-parameter logistic model when the latent abilities of examinees for some test items followed bimodal distributions. Moreover, our model outperformed the monotone-homogeneity model because of the effect of the smoothness constraints.

The contributions of this research are twofold. Firstly, we formulated the smoothness-constrained NIRT model as a mathematical optimization problem. Although we implemented the EM algorithm for the purpose of efficient computation, high-performance mixed-integer optimization algorithms could be used to compute an optimal solution to the problem [1]. Secondly, we validated the utility of shape restrictions on nonparametric ICCs in avoiding an unstable ICC estimation specific to NIRT models. A future direction of study will be to impose other shape restrictions on nonparametric ICCs and evaluate their effectiveness.

Although PIRT models are used commonly by many testing companies, it is known that they do not always fit the actual item-response data well. In contrast, NIRT models have the potential to estimate various forms of ICCs, and the estimation performance could be improved by incorporating the smoothness constraints. So, we expect that our research will extend the usefulness of NIRT models.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP26750114 and JP17K12983.

References

- [1] Bixby, R. E.: A Brief History of Linear and Mixed-integer Programming Computation. In Grötschel, M. (Ed.), *Documenta Mathematica: Optimization Stories*, Deutschen Mathematiker-Vereinigung, Berlin, pp. 107–121 (2012).
- [2] Croon, M. A.: Latent Class Analysis with Ordered Latent Classes. *British Journal of Mathematical and Statistical Psychology*, Vol. 43, No. 2, pp. 171–192 (1990).

- [3] Croon, M. A.: Investigating Mokken Scalability of Dichotomous Items by Means of Ordinal Latent Class Analysis. *British Journal of Mathematical and Statistical Psychology*, Vol. 44, No. 2, pp. 315–331 (1991).
- [4] Douglas, J.: Joint Consistency of Nonparametric Item Characteristic Curve and Ability Estimation. *Psychometrika*, Vol. 62, No. 1, pp. 7–28 (1997).
- [5] Douglas, J. and Cohen, A.: Nonparametric Item Response Function Estimation for Assessing Parametric Model Fit. *Applied Psychological Measurement*, Vol. 25, No. 3, pp. 234–243 (2001).
- [6] Falk, C. F. and Cai, L.: Semiparametric Item Response Functions in the Context of Guessing. *Journal of Educational Measurement*, Vol. 53, No. 2, pp. 229–247 (2016).
- [7] Guo, H. and Sinharay, S.: Nonparametric Item Response Curve Estimation with Correction for Measurement Error. *Journal of Educational and Behavioral Statistics*, Vol. 36, No. 6, pp. 755–778 (2011).
- [8] Johnson, M. S.: Modeling Dichotomous Item Responses with Free-knot Splines. *Computational Statistics & Data Analysis*, Vol. 51, No. 9, pp. 4178–4192 (2007).
- [9] Junker, B. and Sijtsma, K.: Nonparametric Item Response Theory in Action: An Overview of the Special Issue. *Applied Psychological Measurement*, Vol. 25, No. 3, pp. 211–220 (2001).
- [10] Karabatsos, G. and Sheu, C. F.: Order-constrained Bayes Inference for Dichotomous Models of Unidimensional Nonparametric IRT. *Applied Psychological Measurement*, Vol. 28, No. 2, pp. 110–125 (2004).
- [11] Lee, Y. S.: A Comparison of Methods for Nonparametric Estimation of Item Characteristic Curves for Binary Items. *Applied Psychological Measurement*, Vol. 31, No. 2, pp. 121–134 (2007).
- [12] Levy, R.: The Rise of Markov Chain Monte Carlo Estimation for Psychometric Modeling. *Journal of Probability and Statistics*, Article ID 537139 (2009).
- [13] Liang, L. and Browne, M. W.: A Quasi-parametric Method for Fitting Flexible Item Response Functions. *Journal of Educational and Behavioral Statistics*, Vol. 40, No. 1, pp. 5–34 (2015).
- [14] Liang, T. and Wells, C. S.: A Nonparametric Approach for Assessing Goodness-of-fit of IRT Models in a Mixed Format Test. *Applied Measurement in Education*, Vol. 28, No. 2, pp. 115–129 (2015).
- [15] Liang, T., Wells, C. S., and Hambleton, R. K.: An Assessment of the Nonparametric Approach for Evaluating the Fit of Item Response Models. *Journal of Educational Measurement*, Vol. 51, No. 1, pp. 1–17 (2014).

- [16] Ligtoet, R. and Vermunt, J. K.: Latent Class Models for Testing Monotonicity and Invariant Item Ordering for Polytomous Items. *British Journal of Mathematical and Statistical Psychology*, Vol. 65, No. 2, pp. 237–250 (2012).
- [17] Luzardo, M. and Rodríguez, P.: A Nonparametric Estimator of a Monotone Item Characteristic Curve. In van der Ark, L., Bolt, D., Wang, W. C., Douglas, J., and Chow, S. M. (Eds.), *Quantitative Psychology Research*, Springer, Cham, pp. 99–108 (2015).
- [18] Meijer, R. R., Tendeiro, J. N., and Wanders, R. B.: The Use of Nonparametric Item Response Theory to Explore Data Quality. In Reise, S. P. and Revicki, D. (Eds.), *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*, Routledge, pp. 85–110 (2015).
- [19] Meredith, W.: Some Results Based on a General Stochastic Model for Mental Tests. *Psychometrika*, Vol. 30, No. 4, pp. 419–440 (1965).
- [20] Mokken, R. J.: *A Theory and Procedure of Scale Analysis with Applications in Political Research*, De Gruyter Mouton, Berlin, Boston (2011).
- [21] Mokken, R. J. and Lewis, C.: A Nonparametric Approach to the Analysis of Dichotomous Item Responses. *Applied Psychological Measurement*, Vol. 6, No. 4, pp. 417–430 (1982).
- [22] Molenaar, I. W.: Thirty Years of Nonparametric Item Response Theory. *Applied Psychological Measurement*, Vol. 25, No. 3, pp. 295–299 (2001).
- [23] Monroe, S. and Cai, L.: Estimation of a Ramsay-curve Item Response Theory Model by the Metropolis–Hastings Robbins–Monro Algorithm. *Educational and Psychological Measurement*, Vol. 74, No. 2, pp. 343–369 (2014).
- [24] Mori, J. and Kano, Y.: Item Response Theory Using a Finite Mixture of Logistic Models with Item-specific Mixing Weights. *Journal of the Japanese Society of Computational Statistics*, Vol. 26, No. 1, pp. 17–38 (2013).
- [25] Nozawa, Y.: *Comparison of Parametric and Nonparametric IRT Equating Methods Under the Common-item Nonequivalent Groups Design*. Dissertation, University of Iowa (2008).
- [26] Ramsay, J. O.: Monotone Regression Splines in Action. *Statistical Science*, Vol. 3, No. 4, pp. 425–441 (1988).
- [27] Ramsay, J. O.: Kernel Smoothing Approaches to Nonparametric Item Characteristic Curve Estimation. *Psychometrika*, Vol. 56, No. 4, pp. 611–630 (1991).
- [28] Ramsay, J. O. and Abrahamowicz, M.: Binomial Regression with Monotone Splines: A Psychometric Application. *Journal of the American Statistical Association*, Vol. 84, No. 408, pp. 906–915 (1989).

- [29] Ramsay, J. O. and Silverman, B. W.: *Functional Data Analysis*. Springer, New York (2005).
- [30] Ramsay, J. O. and Winsberg, S.: Maximum Marginal Likelihood Estimation for Semiparametric Item Analysis. *Psychometrika*, Vol. 56, No. 3, pp. 365–379 (1991).
- [31] Rossi, N., Wang, X., and Ramsay, J. O.: Nonparametric Item Response Function Estimates with the EM Algorithm. *Journal of Educational and Behavioral Statistics*, Vol. 27, No. 3, pp. 291–317 (2002).
- [32] Shojima, K.: *Neural Test Theory*. DNC Research Note 07-02, The National Center for University Entrance Examinations (2007).
- [33] Shojima, K.: *Neural Test Theory: A Latent Rank Theory for Analyzing Test Data*. DNC Research Note 08-01, The National Center for University Entrance Examinations (2008).
- [34] Sijtsma, K. and Molenaar, I. W.: *Introduction to Nonparametric Item Response Theory*. Sage, Thousand Oaks (2002).
- [35] Simonoff, J. S.: *Smoothing Methods in Statistics*. Springer, New York (2012).
- [36] Stout, W. F.: A Nonparametric Approach for Assessing Latent Trait Unidimensionality. *Psychometrika*, Vol. 52, No. 4, pp. 589–617 (1987).
- [37] Takano, Y., Tsunoda, S., and Muraki, M.: Mathematical Optimization Models for Nonparametric Item Response Theory. *Information Science and Applied Mathematics*, Vol. 23, pp. 1–18 (2016).
- [38] van der Linden, W. J. and Hambleton, R. K. (Eds.): *Handbook of Modern Item Response Theory*. Springer, New York (2013).
- [39] van Onna, M. J. H.: Bayesian Estimation and Model Selection in Ordered Latent Class Models for Polytomous Items. *Psychometrika*, Vol. 67, No. 4, pp. 519–538 (2002).
- [40] Vermunt, J. K.: The Use of Restricted Latent Class Models for Defining and Testing Nonparametric and Parametric Item Response Theory Models. *Applied Psychological Measurement*, Vol. 25, No. 3, pp. 283–294 (2001).