Bayesian Estimation of Latent Rank in Neural Test Theory

SHOJIMA Kojiro

November 2007

Department of Test Analysis and Evaluation, Research Division, The National Center for University Entrance Examinations

Bayesian estimation of latent rank in neural test theory

Kojiro Shojima

Abstract

This paper describes a Bayesian method for estimating the latent ranks of examinees under neural test theory (NTT) by extending the maximum likelihood (ML) method. This method reduces to the ML method unless prior distributions for rank membership profiles are supposed. In the NTT model estimated by the ML method, the frequencies at both ends of the latent rank scale are inclined to be larger than those of the intermediate ranks. This characteristic derived from the self-organizing map can be alleviated by the Bayesian method with a trapezoidal prior distribution. This method is useful for test administrators who want to grade examinees into several groups with nearly equal frequencies.

Key words: neural test theory, graded neural test model, Bayesian estimation, maximum a posteriori rank, trapezoidal prior distribution, latent rank distribution, rank membership distribution.

ニューラルテスト理論における潜在ランクのベイズ推定

荘島宏二郎

要約

本研究では，ニューラルテスト理論における潜在ランクをベイズ推定する方法を論じた．本方法は，事前分布を仮定しないときは，最尤推定法と同一である．台形事前分布を用いると，潜在ランク分布の両端に受験者が集中するという自己組織化マップから由来する性質を和らげることができる．どの潜在ランクにもおおむね等分してランク付けしたいテスト実務家にとって有効な方法であると思われる．

キーワード: ニューラルテスト理論，段階ニューラルテストモデル，ベイズ推定，最大事後ランク，台形事前分布，潜在ランク分布，ランクメンバーシップ分布．

Department of Test Analysis and Evaluation, Research Division, The National Center for University Entrance Examinations

1 Introduction

Neural test theory (NTT; Shojima, 2007a, 2007b) is a latent rank theory (LRT; Shojima, 2007f). The LRT is a generic theory in which the assumed latent scale is rank-ordered. At present, the LRT models can be estimated by the algorithm of self-organizing map (LRT-SOM) and EM algorithm (LRT-EM). The self-organizing map (Kohonen, 1995) is a statistical learning model for clustering samples on a two-dimensional lattice. The EM (expectation maximization) algorithm (Dempster, Laird, & Rubin, 1977) is a Bayesian method for estimating structural parameters of a statistical model. The NTT models are identical to the LRT-SOM models in the framework of Shojima (2007f).

The NTT models are estimated by repeatedly selecting winner nodes and updating reference vectors in the statistical learning process. The winner node selection method is also used for estimating latent ranks of examinees. For this estimation, Shojima (2007a, 2007b) directly introduced the method of using the square of the Euclidian distance (ED^2) from the SOM procedure, and he used it in the NTT procedure as a measure for estimating the latent ranks and selecting the winner nodes. Then, Shojima (2007c) developed the maximum likelihood (ML) method for estimating them. The ML method enables NTT to provide information about the rank membership profile (RMP; Shojima, 2007c), which is useful for reviewing the probabilities of examinees belonging to respective latent ranks and the rank membership distribution (RMD; Shojima, 2007c), which is the simple sum of the RMPs of all examinees. In addition, the ML method is important for evaluating the goodness-of-fit index of the NTT model and the statistical significance of item reference profiles (Shojima, 2007d), which is the behavior of the correct answer ratios versus the latent ranks of each item. The ML method is more useful than the ED^2 method because various kinds of knowledge and techniques derived from various statistical theories become available. The NTT models became statistical rather than mathematical as a result of the implementation of the ML method.

In the RMD and the latent rank distribution (LRD), which is the frequency distribution of the latent ranks of the examinees estimated by the ML method (and also by the ED^2 method), it is often observed that the frequencies at both ends of the latent rank scale, the highest and lowest latent ranks, are larger than those of intermediate ranks (see Shojima, 2007a, 2007b, 2007c, 2007e). That is, the latent scale in NTT is not equiprobable. Although this characteristic is derived from the SOM algorithm, this fact is literally understandable if it is thought that the examinees outside the target ability of the test are obtained collectively at both ends of the latent rank scale because every test has its own target ability. However, the equiprobability scale is useful for some test administrators who want to grade the examinees into several groups with nearly equal frequencies.

In the field of the SOM, many researchers have been aware of the problem and tried to obtain the equiprobability map (e.g., Kohonen, 1995; Van Hulle, 2000). These efforts have mainly focused on not producing dead units in the map. Although similar approaches appear promising under NTT, the dead units (dead latent ranks) are rarely produced in the NTT models. This is because the possibility of the dead units in which no examinee is placed appearing increases as the number of latent ranks Q becomes larger, but Q is usually much smaller than the number of examinees. As has been repeatedly stressed concerning NTT, tests do not have high resolution for distinguishing the difference between two examinees who have nearly equal abilities, whereas weighing machines can detect the slight difference between two persons with approximately the same weights. The most that a test can do is to grade examinees into several ranks, so it has never been recommended in NTT that Q becomes larger. Consequently, another approach, different from the ones proposed in SOM, is required to solve the problem. In this study, a Bayesian method is discussed. Bayesian methods can control the shape of the LRD and RMD by assuming the prior distribution of each examinee's RMP. This method is also useful as a winner node selection method.

2 Method

2.1 Dichotomous Neural Test Model

The dichotomous neural test (DNT; Shojima, 2007a) model is an NTT model for analyzing binary data. Let Q be the number of latent ranks and n be the number of items. Then, the reference matrix becomes $\boldsymbol{V} = \{v_{qj}\}$ $(Q \times n)$. The q-th row vector in $\boldsymbol{V}$ is the rank reference vector (RRV) of latent rank R_q $(q = 1, \cdots, Q)$, and the j-th column vector in $\boldsymbol{V}$ is the item reference profile (IRP) of item j $(= 1, \cdots, n)$.

Let us assume that the sample size is N and the response matrix of the examinees is $\boldsymbol{U} = \{u_{ij}\}$ $(N \times n)$, where u_{ij} is a dichotomous variable coded 1 if the response of examinee i to item j is correct and 0 otherwise. We also assume that $\boldsymbol{Z} = \{z_{ij}\}$ $(N \times n)$ is the missing indicator matrix (Shojima, 2007e), where z_{ij} is also a dichotomous variable that is coded 1 when u_{ij} is observed and 0 when it is missing. Let us further assume that $\boldsymbol{F} = \{f_{iq}\}$ $(N \times Q)$ is the membership indicator matrix (Shojima, 2007f), where f_{iq} is coded 1 if examinee i belongs to latent rank R_q and 0 otherwise. Then, the probability of the response vector of examinee

i, $\boldsymbol{u}_i$ $(n \times 1)$, can be expressed by

$$p(\boldsymbol{u}_i|\boldsymbol{f}_i, \boldsymbol{V}) = \prod_{q=1}^{Q} p(\boldsymbol{u}_i|f_{iq}, \boldsymbol{v}_q) = \prod_{q=1}^{Q}\prod_{j=1}^{n}\{v_{qj}^{u_{ij}}(1-v_{qj})^{1-u_{ij}}\}^{z_{ij}f_{iq}}, \tag{1}$$

where $\boldsymbol{f}_i = \{f_{iq}\}$ $(Q \times 1)$. The above equation is the likelihood of $\boldsymbol{u}_i$ with unknown parameters $\boldsymbol{f}_i$. When the latent rank to which examinee i belongs is $R_{q'}$, (1) reduces to

$$p(\boldsymbol{u}_i|f_{iq'}, \boldsymbol{v}_{q'}) = \prod_{j=1}^{n}\{v_{q'j}^{u_{ij}}(1-v_{q'j})^{1-u_{ij}}\}^{z_{ij}f_{iq'}}. \tag{2}$$

In the ML method of Shojima (2007c), the latent rank that makes the log-likelihood of (2) maximum is determined as the maximum likelihood estimate (MLE) of the latent rank. That is, the MLE of examinee i's latent rank, R_{r_i}, is obtained by

$$R_{r_i}^{(ML)}: \; r_i = \arg\max_{q \in Q} \ln p(\boldsymbol{u}_i|f_{iq}, \boldsymbol{v}_q). \tag{3}$$

However, the latent rank estimation can be considered as the estimation problem of $\boldsymbol{f}_i$ given $\boldsymbol{V}$ and $\boldsymbol{u}_i$ because each f_{iq} is estimated as a continuous value, provided that $\sum_q f_{iq} = 1$, although it is ideally a dichotomous variable, as described above. That is, each f_{iq} can be regarded as the probability that examinee i belongs to latent rank R_q, so each examinee's latent rank can be determined as the one with the maximum value of the membership probability. This procedure is the same as saying that the latent rank of each examinee is selected by the rank membership profile (Shojima, 2007c), which is also equal to the identification problem of the latent class in the context of the mixture distribution (e.g., Everitt & Hand, 1981; Titterington, Smith, & Makov, 1985; Muthén & Shedden, 1999). The latent rank estimation by the RMP can be written as

$$R_{r_i}^{(ML)}: \; r_i = \arg\max_{q \in Q} f_{iq} = \arg\max_{q \in Q} \frac{p(\boldsymbol{u}_i|f_{iq}, \boldsymbol{v}_q)}{\sum_{q'=1}^{Q} p(\boldsymbol{u}_i|f_{iq'}, \boldsymbol{v}_{q'})}. \tag{4}$$

The estimate obtained by the above equation is identical to that obtained by (3).

The RMP of examinee i, $\boldsymbol{f}_i = \{f_{iq}\}$, is a discrete probability distribution. By Bayes' theorem, when the prior probability for each f_{iq}, $p(f_{iq})$, is supposed, the posterior probability of f_{iq} then becomes proportional to the product of the likelihood and the prior probability. That is,

$$p(f_{iq}|\boldsymbol{u}_i, \boldsymbol{v}_q) \propto p(\boldsymbol{u}_i|f_{iq}, \boldsymbol{v}_q)p(f_{iq}). \tag{5}$$

Accordingly, the latent rank that gives the maximum value of the posterior membership probability is the maximum a posteriori (MAP) rank of examinee i. Thus,

$$R_{r_i}^{(MAP)}: \ r_i = \arg\max_{q \in Q} \{\ln p(\boldsymbol{u}_i | f_{iq}, \boldsymbol{v}_q) + \ln p(f_{iq})\}. \tag{6}$$

When the prior distribution is not considered, the MAP rank given by (6) coincides with the ML rank given by (3). In addition, the posterior RMP gives

$$f_{iq}^{*} = \frac{p(\boldsymbol{u}_i | f_{iq}, \boldsymbol{v}_q) p(f_{iq})}{\sum_{q'=1}^{Q} p(\boldsymbol{u}_i | f_{iq'}, \boldsymbol{v}_{q'}) p(f_{iq'})}. \tag{7}$$

The above Bayesian method is also applicable to the method for selecting the winner nodes in the statistical learning process. That is,

$$R_w^{(MAP)}: \ w = \arg\max_{q \in Q} \Big[\sum_{j=1}^{n} z_{hj}^{(t)} \big\{u_{hj}^{(t)} \ln v_{qj}^{(t,h-1)} + (1 - u_{hj}^{(t)}) \ln(1 - v_{qj}^{(t,h-1)})\big\} + \ln p(f_{iq})\Big], \tag{8}$$

where $\boldsymbol{u}_h^{(t)} = \{u_{hj}^{(t)}\}$ $(n \times 1)$ is the h-th input data in the t-th period in the statistical learning process, and $\boldsymbol{v}_q^{(t,h-1)} = \{v_{qj}^{(t,h-1)}\}$ $(n \times 1)$ is the RRV of latent rank R_q updated after learning input data $\boldsymbol{u}_{h-1}^{(t)}$.

2.2 Graded Neural Test Model

The graded neural test (GNT; Shojima, 2007b) model is a useful NTT model for analyzing ordered polytomous data, it reduces to the dichotomous NTT model when all items are binary. Let C_j be the number of categories of item j. The reference matrix of the GNT model then becomes

$$\boldsymbol{V} = \{v_{qjk}\} = \begin{bmatrix} v_{1,1,1} & \cdots & v_{1,1,C_1-1} & v_{1,2,1} & \cdots & v_{1,n,C_n-1} \\ v_{2,1,1} & \cdots & v_{2,1,C_1-1} & v_{2,2,1} & \cdots & v_{2,n,C_n-1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ v_{Q,1,1} & \cdots & v_{Q,1,C_1-1} & v_{Q,2,1} & \cdots & v_{Q,n,C_n-1} \end{bmatrix} \quad \Big(Q \times \Big(\sum_{j=1}^{n} C_j - 1\Big)\Big), \tag{9}$$

where each column vector $\boldsymbol{v}_{jk} = \{v_{qjk}\}$ $(j = 1, \cdots, n; k = 1, \cdots, C_j - 1)$ is the boundary category reference profile (BCRP; Shojima, 2007b), and each row vector is the RRV of latent rank R_q. Each v_{qjk} stands for the probability that the examinees in latent rank R_q select category k or higher in item j. In addition, the expanded reference matrix is

$$\boldsymbol{P} = \{p_{qjk}\} = \begin{bmatrix} p_{1,1,0} & \cdots & p_{1,1,C_1-1} & p_{1,2,0} & \cdots & p_{1,n,C_n-1} \\ p_{2,1,0} & \cdots & p_{2,1,C_1-1} & p_{2,2,0} & \cdots & p_{2,n,C_n-1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p_{Q,1,0} & \cdots & p_{Q,1,C_1-1} & p_{Q,2,0} & \cdots & p_{Q,n,C_n-1} \end{bmatrix} \quad \Big(Q \times \sum_{j=1}^{n} C_j\Big), \tag{10}$$

where

$$p_{qjk} = v_{qjk} - v_{qjk+1} \quad (k = 0, 1, \cdots, C_j - 1), \tag{11}$$

provided that

$$v_{qj0} = 1, \tag{12}$$

and

$$v_{qjC_j} = 0. \tag{13}$$

Each column vector $\boldsymbol{p}_{jk} = \{p_{qjk}\}$ $(j = 1, \cdots, n; k = 0, \cdots, C_j - 1)$ in the expanded reference matrix is the item category reference profile (ICRP; Shojima, 2007b) of category k in item j. Each p_{qjk} represents the probability that the examinees located in latent rank R_q select category k of item j.

Furthermore, let $\boldsymbol{X} = \{x_{ij} | x_{ij} \in \{0, 1, \cdots, C_j - 1\}\}$ $(N \times n)$ represent the response matrix of the examinees. Then, the occurrence probability of $\boldsymbol{x}_i$ given $\boldsymbol{P}$ and $\boldsymbol{f}_i$ becomes

$$p(\boldsymbol{u}_i | \boldsymbol{f}_i, \boldsymbol{P}) = \prod_{q=1}^{Q} p(\boldsymbol{u}_i | f_{iq}, \boldsymbol{p}_q) = \prod_{q=1}^{Q} \prod_{j=1}^{n} \prod_{k=1}^{C_j-1} p_{qjk}^{u_{ijk} z_{ij} f_{iq}}, \tag{14}$$

where z_{ij} and f_{iq} are the missing indicator and membership indicator, respectively, and $\boldsymbol{u}_i = \{u_{ijk}\}$ is

$$\boldsymbol{u}_i = \{u_{ijk}\} = [u_{i,1,0}\ u_{i,1,1} \cdots u_{i,1,C_1-1}\ u_{i,2,0} \cdots u_{i,n,C_n-1}]' \quad \left(\sum_{j=1}^{n} C_j \times 1\right) \tag{15}$$

$$u_{ijk} = \begin{cases} 1 & \text{if } x_{ij} = k \\ 0 & \text{otherwise} \end{cases} \qquad (k = 0, \cdots, C_j - 1). \tag{16}$$

Consequently, the MAP rank under the GNT model is obtained as follows:

$$R_{r_i}^{(MAP)} :\ r_i = \arg\max_{q \in Q} \{\ln p(\boldsymbol{u}_i | f_{iq}, \boldsymbol{p}_q) + \ln p(f_{iq})\} \tag{17}$$

Then, the posterior RMP is given by

$$f_{iq}^{*} = \frac{p(\boldsymbol{u}_i | f_{iq}, \boldsymbol{p}_q) p(f_{iq})}{\sum_{q'=1}^{Q} p(\boldsymbol{u}_i | f_{iq'}, \boldsymbol{p}_{q'}) p(f_{iq'})}. \tag{18}$$

Furthermore, the winner node selection method under the GNT model is

$$R_{w}^{(MAP)} :\ w = \arg\max_{q \in Q} \Big[\sum_{j=1}^{n} \sum_{k=1}^{C_j-1} z_{hj}^{(t)} u_{hjk}^{(t)} \ln p_{qjk}^{(t,h-1)} + \ln p(f_{iq}) \Big], \tag{19}$$

where $\boldsymbol{u}_h = \{u_{hjk}^{(t)}\}$ $(\sum_j C_j \times 1)$ is obtained by (16) from the h-th input data in the t-th period, $\boldsymbol{x}_h^{(t)}$. In addition, applying (11), $p_{qjk}^{(t,h-1)}$ is obtained from $v_{qjk}^{(t,h-1)}$, which is updated after learning input data $\boldsymbol{x}_{h-1}^{(t)}$.

2.3 Posterior rank distribution

The latent rank distribution (LRD; Shojima, 2007a, 2007b) is the frequency distribution of the estimated latent ranks of the examinees. In the framework of the Bayesian method, the frequency distribution of the MAP ranks can be called the posterior LRD. Similarly, the rank membership distribution (RMD; Shojima, 2007c) is the simple sum of the RMPs, and the one summing the posterior RMPs can be called the posterior RMD. It is obtained as

$$\boldsymbol{f}^* = \{f_q^*\} \quad (Q \times 1) \tag{20}$$

$$f_q^* = \sum_{i=1}^{N} f_{iq}^*. \tag{21}$$

2.4 Observation ratio profile

The observation ratio profile (ORP; Shojima, 2007e) is useful for examining the behavior of each item's response (missing) ratio through latent ranks. The unweighted ORP is obtained from the missing indicators and the ML ranks, and the weighted ORP is calculated from the missing indicators and the RMPs by the ML method. The unweighted and weighted ORPs are also available in the context of the Bayesian method. That is, the unweighted posterior ORP can be computed from the missing indicators and the MAP ranks, and the weighted ORP can be computed from the missing indicators and the posterior RMPs.

3 Analysis

3.1 Example 1

A world history test was analyzed. The sample size was 2049 and the number of items was 36. All items were binary, so the data was analyzed by the DNT model. In Example 1, the result when no prior distribution is imposed, i.e., the usual ML method, is reported as the basis for comparison. Parameters Q, T, α_1, σ_1, and σ_0 were set to 10, 500, 0.1, 10, and 1.0, respectively. Parameter α_1 is the initial value of the impact size of the winner node to learn (numerical approach) the input data. In addition, σ_1 and σ_0 are the initial and final values of the region size of the neighboring nodes in which the learning of the winner node propagates (Shojima, 2007a, 2007b, 2007e).

The IRPs estimated using the ML method are shown in Figure 1. The goodness-of-fit index of the model by the Shojima (2007d) method was $\chi^2_{(324)} = 10163.73$ ($p < 0.001$). The model was rejected because of the large size of the data. In addition, Figure 2 shows the

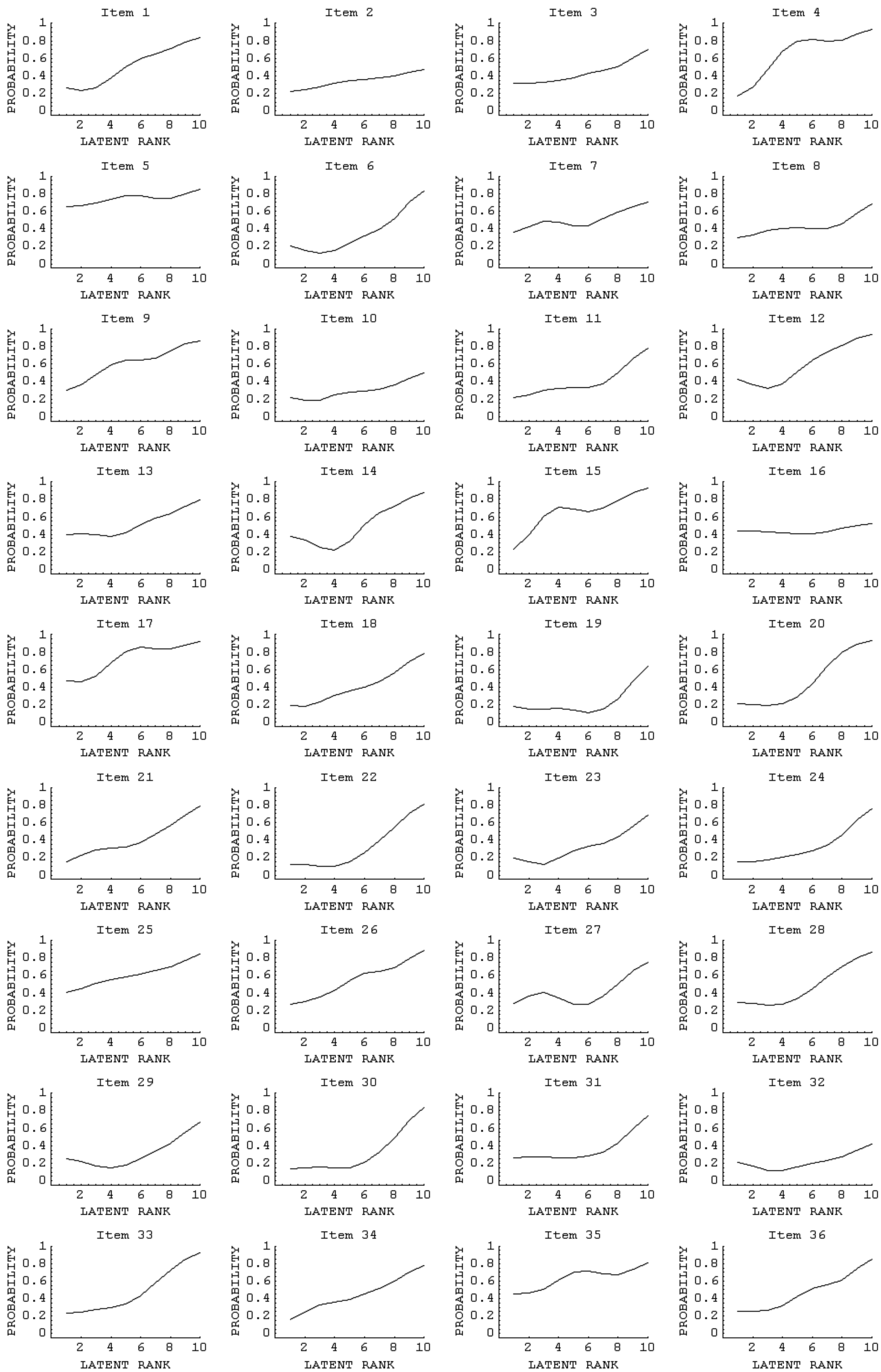
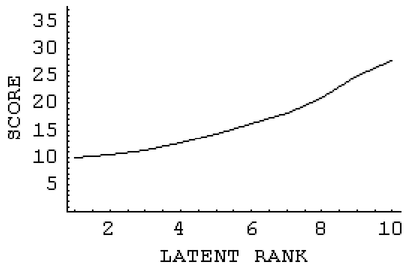
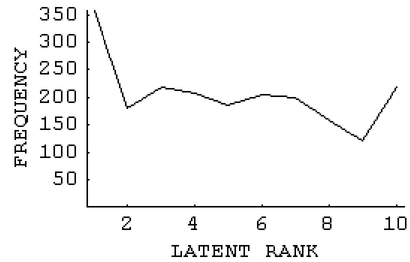


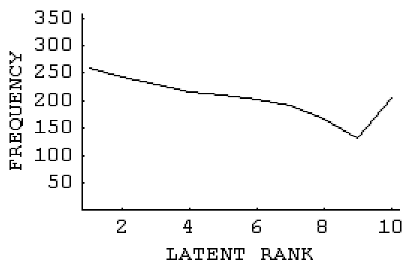
Figure 1: Item Reference Profiles (ML)



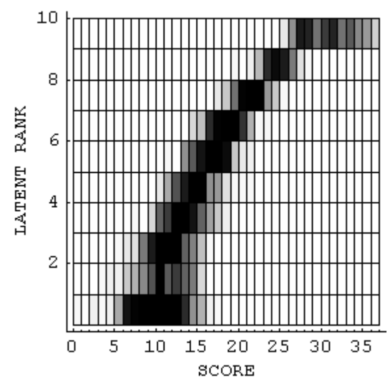
(a) Test Reference Profile (ML)



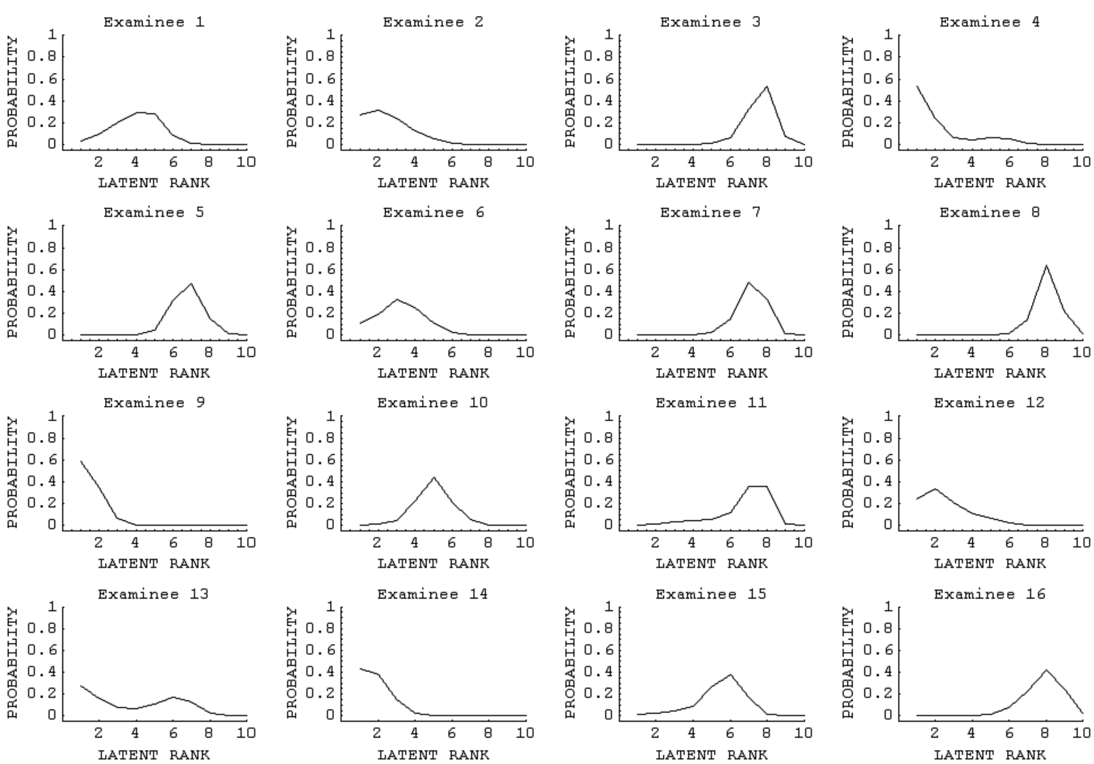
(b) Latent Rank Distribution (ML)



(c) Rank Membership Distribution (ML)



(d) Scatter Plot of Scores and Ranks (ML)



(e) Rank Membership Profiles of Examinees 1-16 (ML)

Figure 2: TRP, LRD, RMD, Scatter Plot, and RMPs Obtained by ML Method

test reference profile (TRP; Shojima, 2007a, 2007b), the latent rank distribution (LRD), the rank membership distribution (RMD), the scatter plot of the scores and the estimated ML ranks of the examinees, and the rank membership profiles of examinees 1–16.

Note that the frequencies at both ends of the latent rank scale, R_1 and R_{10}, are larger than those of intermediate ranks as a result of the characteristic derived from the SOM algorithm. Although the abilities of the examinees at both ends of the scale are outside the target ability of the test, this characteristic is probably undesirable for some test practitioners who want to grade examinees or students into ranks with nearly equal frequencies. The frequency of latent rank R_9 happened to be smaller than those of other latent ranks; this was derived not from the SOM or NTT algorithms, but from inherent characteristics attributable to the analyzed data.

3.2 Example 2

In this section, a result obtained by the Bayesian estimation method is reported. In order to apply the Bayesian method, the type of prior distribution that is selected is critical. As seen in Example 1, the NTT (and SOM) algorithm makes the frequencies at both ends of the latent ranks larger than those of intermediate ranks. Therefore, the trapezoidal prior distribution was used for alleviating the characteristic. That is,

$$p(f_{iq}) = \begin{cases} \pi & \text{if } q = 1, Q \\ (1-2\pi)/(Q-2) & \text{otherwise} \end{cases} \quad (\forall i \in N) \tag{22}$$

provided that π is smaller than or equal to $\leq 1/Q$ to make the prior probabilities of both ends of the scale become smaller than those of intermediate ranks. Here, π was set to 0.085. Accordingly, the prior probability of each intermediate rank was 0.10375.

Figure 3 shows the IRPs obtained by the Bayesian method. Their shapes are very similar to those in Figure 1. The goodness-of-fit of the model was $\chi^2_{(324)} = 10048.60$ ($p < 0.001$). Next, Figure 4 shows the TRP, posterior LRD, posterior RMP, the scatter plot of the scores and the MAP ranks, and the posterior RMP of examinees 1–16 under the Bayesian method. It is obvious that the frequencies obtained at both ends of the scale of the posterior LRD and RMD (Figures 4(b) and 4(c)) are smaller than those in the LRD and RMD obtained by the ML method (Figures 2(b) and 2(c)). This is the effect of the prior distribution.

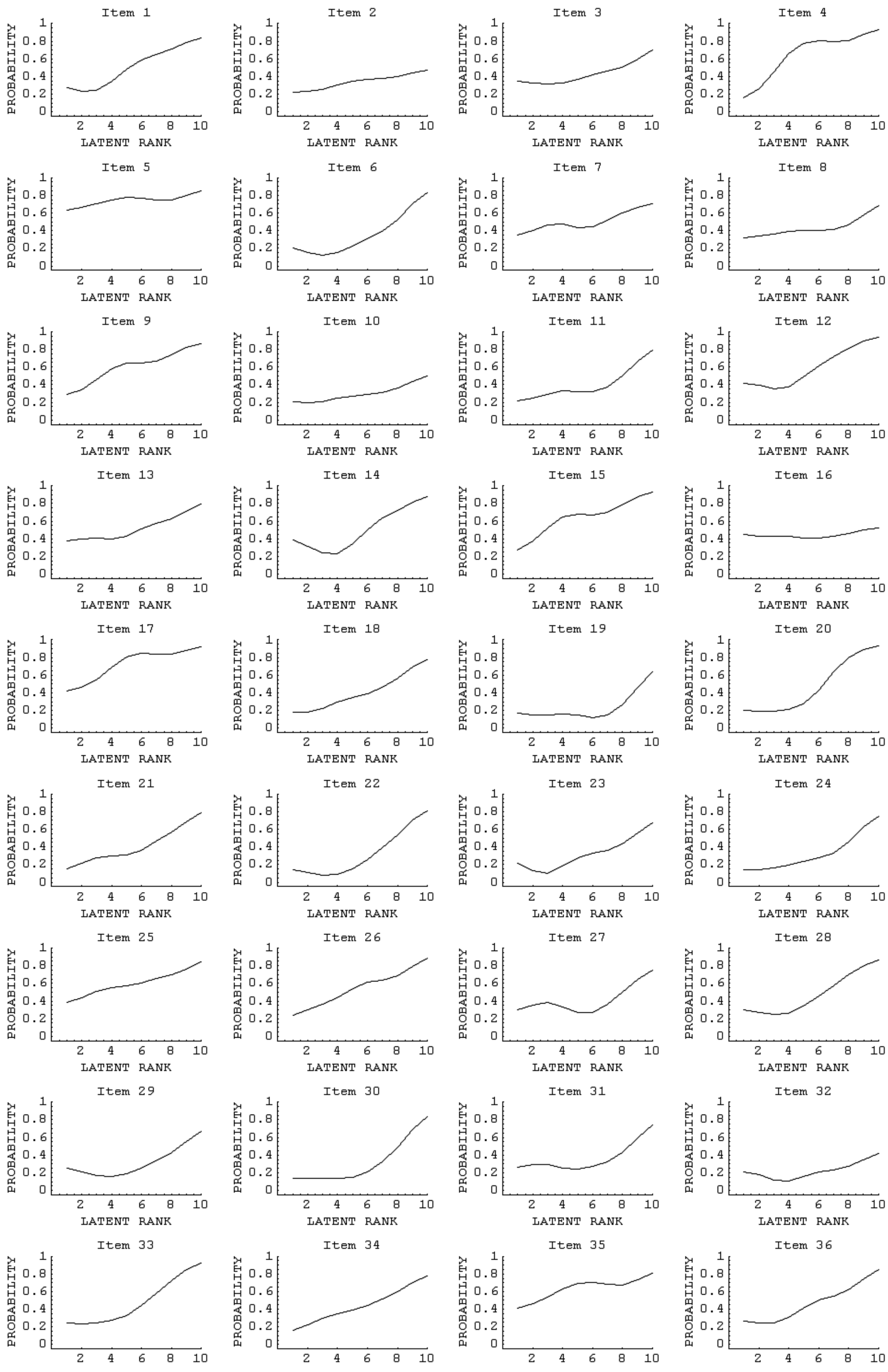
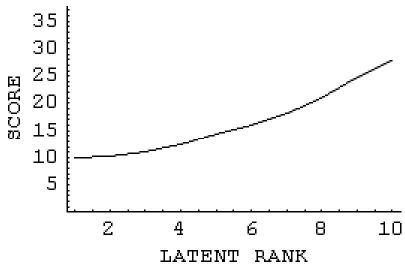
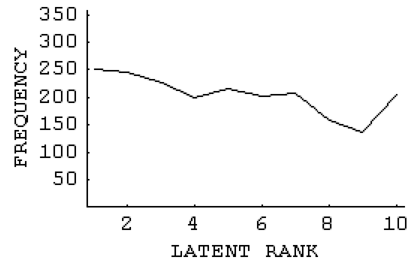


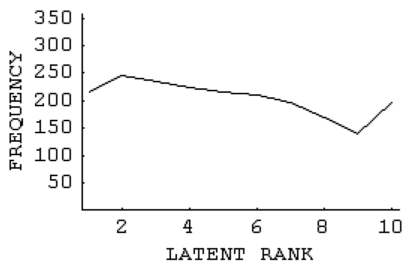
Figure 3: Item Reference Profiles (Trapezoidal Prior Distribution)



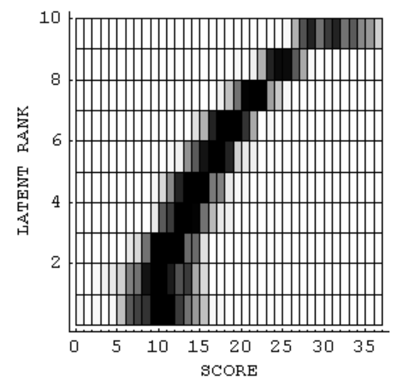
(a) TRP (Trapezoidal Prior Dist.)



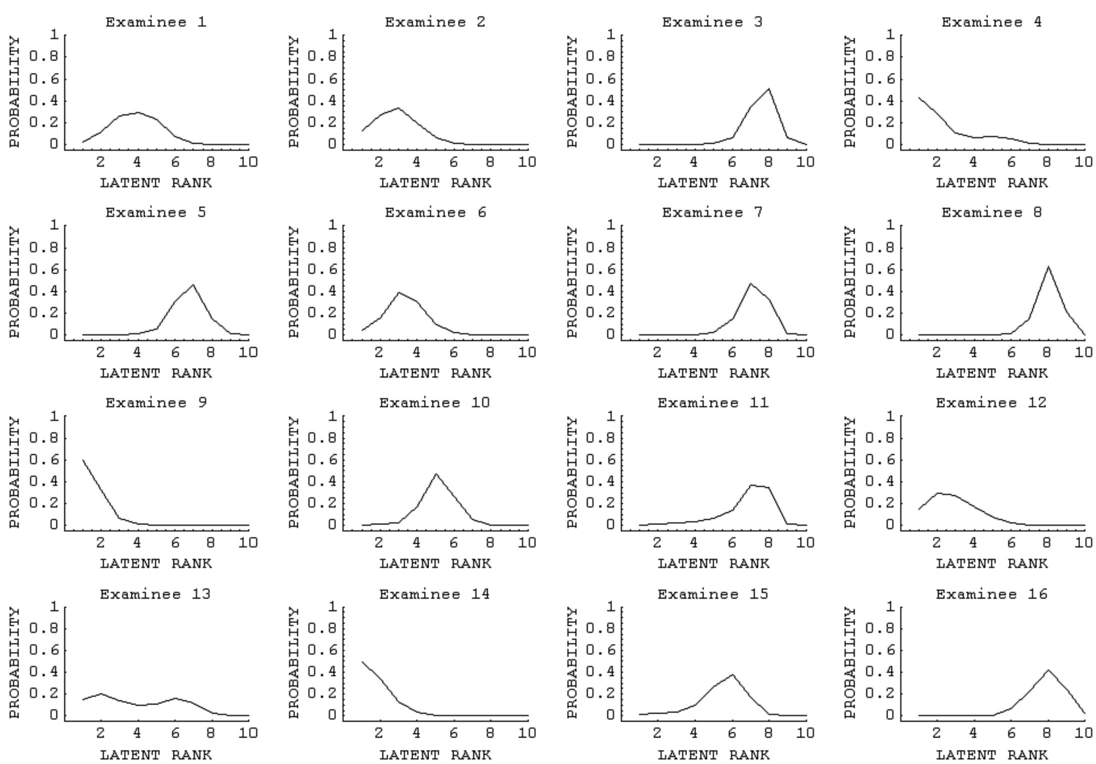
(b) Posterior LRD (Trapezoidal Prior Dist.)



(c) Posterior RMD (Trapezoidal Prior Dist.)



(d) Scatter Plot (Trapezoidal Prior Dist.)



(e) Posterior Rank Membership Profiles of Examinees 1-16 (Trapezoidal Prior Dist.)

Figure 4: Figures by Bayesian Method with Trapezoidal Prior Distribution

4 Discussion

In this study, a Bayesian method of estimating the latent ranks and selecting the winner nodes was described. For test administrators or practitioners who want to divide examinees into several latent ranks with nearly equal frequencies, this method is a useful solution that can overcome the characteristic derived from the SOM that the frequencies at both ends of the latent rank scale become larger.

Generally speaking, the influence of prior distribution becomes smaller as the number of items becomes larger. Conversely, the selected prior distribution has a large effect when the number of items is small. In this case, it is necessary to make the prior distribution flatter because it is undesirable for the prior distribution to interfere unduly with the relationship between the model and the data. Further research is required to control the effect of the prior distribution according to the number of items. Furthermore, while the trapezoidal distribution was examined in the simulation study, other discrete distributions, such as the triangular or binomial distribution, may be useful candidates.

References

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

Kohonen, T. (1995) *Self-organizing maps*. Springer.

Lord, F. M. (1980) *Applications of item response theory to practical testing problems*. Lawrence Erlbaum Associates.

Muthén, B. & Shedden, K. (1999) Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, **55**, 463-469.

Shojima, K. (2007a) Neural test theory. *DNC Research Note*, 07-02.

Shojima, K. (2007b) The graded neural test model: A neural test model for ordered polytomous data. *DNC Research Note*, 07-03.

Shojima, K. (2007c) Maximum likelihood estimation of latent rank under the neural test model. *DNC Research Note*, 07-04.

Shojima, K. (2007d) Chi-square goodness-of-fit test under the neural test model. *DNC Research Note*, 07-05.

Shojima, K. (2007e) Estimation for neural test models with missing data. *DNC Research Note*, 07-09.

Shojima, K. (2007f) Latent rank theory: Estimation of item reference profile by marginal maximum likelihood method with EM algorithm. *DNC Research Note*, 07-12.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985) *Statistical analysis of finite mixture distributions*. Wiley.

Van Hulle, M. M. (2000) *Faithful representations and topographic maps*. John Wiley & Sons, Inc.