Maximum Likelihood Estimation of Latent Rank under Neural Test Model

SHOJIMA Kojiro

October 2007

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

Maximum likelihood estimation of latent rank under neural test model

Kojiro Shojima

Abstract

A method is described for estimating the latent ranks of examinees under neural test theory (NTT) using the maximum likelihood method under the assumption of local independence among items. It can also be used to select the winner node in the statistical learning process of NTT. The rank membership profile (RMP) based on the likelihood can be used to estimate the probability that an examinee belongs to a particular latent rank. This method was also applied under the graded neural test model.

Key words: neural test theory, grade neural test model, maximum likelihood estimation, rank membership profile.

ニューラルテスト理論における潜在ランクの最尤推定

荘島宏二郎

要約

本研究では，ニューラルテスト理論 (neural test theory, NTT; Shojima, 2007) において，潜在ランクを推定する際に，最尤推定法による潜在ランクの推定を論じた．また，この方法は，統計的学習の際の勝者ノードの決定方法としても使うことができる．ただし，そのとき，局所独立の仮定を仮定する必要がある．また，尤度に基づいて，各被験者の潜在ランクがどの程度の確からしさで求まっているのかを検討することができる．そのとき，ランクメンバシッププロフィール (rank memberhip profile) が有効である．また，本方法は，段階ニューラルテストモデルにも適用された．

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，最尤推定法，ランクメンバシッププロフィール．

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

1 Introduction

The neural test theory (NTT; Shojima, 2007a, 2007b) has a mechanism based on the self-organizing map (SOM; Kohonen, 1995). The latent scale assumed in NTT is rank-ordered, while the scales in other test theories, such as item response theory and classical test theory, are continuous. Rank-ordering is appropriate because a test is not a measurement device with high resolution like a weiging machine. Tests cannot distinguish the difference between two examinees who have nearly equal abilities whereas a weighing machine can detect a slight difference between two people who have almost the same weights. Therefore, the most that a test can do is to rank examinees into several grades (Shojima, 2007a).

In NTT, the examinees are arranged hierarchically on a latent rank scale; the latent rank of each examinee is determined by the square of the Euclidian distance between the rank reference vector (RRV) and the input data (Shojima, 2007a, 2007b). Let us assume that the sample size is N, the number of items is n, the number of latent ranks is Q, and that the RRV of each latent rank R_q $(q = 1, \cdots, Q)$ is $\boldsymbol{v}_q$ $(n \times 1)$, the latent rank of examinee i is determined by the square of the Euclidian distance:

$$R_w: \ w = \arg\min_{q \in Q} ||\boldsymbol{v}_q - \boldsymbol{u}_i||^2, \tag{1}$$

where $\boldsymbol{u}_i$ is the response vector of examinee i. This equation can also be used to select the winner node in the statistical learning process:

$$R_w^{(t)}: \ w = \arg\min_{q \in Q} ||\boldsymbol{v}_q^{(t)} - \boldsymbol{u}_h^{(t)}||^2, \tag{2}$$

where $\boldsymbol{v}_q^{(t)}$ is the RRV in the t-th period and $\boldsymbol{u}_h^{(t)}$ is the h-th input data in the t-th period.

In the graded neural test (GNT; Shojima, 2007b) model, which is the NTT model for polytomously ordered data, the latent rank is estimated using

$$R_w: \ w = \arg\min_{q \in Q} \sum_{j=1}^{n} \frac{||\boldsymbol{v}_{qj} - \boldsymbol{u}_{ij}||^2}{C_j - 1}, \tag{3}$$

where C_j is the number of categories of item j and v_{qj} $((C_j - 1) \times 1)$ is the j-th subvector of $\boldsymbol{v}_q$, i.e., the RRV of R_q. In addition, $\boldsymbol{u}_{ij}$ is the j-th subvector of examinee i's response vector, $\boldsymbol{u}_i$. The vectors v_{qj} and $\boldsymbol{u}_{ij}$ are scalars when item j is binary (true/false question), and (3) is then reduced to (1) when all items are dichotomous. In addition, the winner node in the learning process of the GNT model is determined by

$$R_w^{(t)}: \ w = \arg\min_{q \in Q} \sum_{j=1}^{n} \frac{||\boldsymbol{v}_{qj}^{(t)} - \boldsymbol{u}_{hj}^{(t)}||^2}{C_j - 1}. \tag{4}$$

In the dichotomous NTT and GNT models, the latent rank of each examinee and the winner node for each input data are selected using the equations above. That is, the numerically closest node in terms of the square of the Euclidian distance (ED^2) is selected in the latent rank estimation and winner node selection processes. However, ED^2 is not the only method that can be used as the measure of the distance between the RRV and the input data; for example, the Manhattan distance or the inner product can also be used.

In this paper, a method is described for estimating the latent rank using the maximum likelihood (ML) method. NTT is mathematical rather than statistical because the computational algorithms of the dichotomous NTT model and the GNT model are directly based on the SOM algorithm. In contrast, the ML method is statistical. However, the ML method can be implemented in the NTT computational algorithm because the RRVs obtained in the learning process are the correct answer rates for items.

2 Latent Rank Estimation by ML Method

2.1 Dichotomous neural test model

Let us assume that the finally obtained item reference matrix is $\boldsymbol{V} = \{\boldsymbol{v}_q\}$ $(q = 1, \cdots, Q)$, where each row vector $\boldsymbol{v}_q$ $(q = 1, \cdots, Q)$ is the RRV of node R_q. Also, the j-th column vector in $\boldsymbol{V}$ is the item reference profile (IRP; Shojima, 2007a) of item j $(= 1, \cdots, n)$. Let us also assume that $\boldsymbol{u}_i = \{u_{ij}\}$ is the response vector of examinee i, where u_{ij} is examinee i's response to item j, which is coded one when the response is correct and zero otherwise, and that U_{ij} is the random variable of u_{ij}.

Let us further suppose that the probability of a correct answer for item j of the examinee with latent rank R_q, $\Pr(u_{ij} = 1|R_q)$, is independent of the probabilities of a correct answer for the other items (the assumption of local independence among items). The occurrence probability of $\boldsymbol{u}_i$, provided that examinee i's latent rank is R_q, can then be written as

$$\Pr(\boldsymbol{u}_i|R_q) = \prod_{j=1}^{n} v_{qj}^{u_{ij}}(1 - v_{qj})^{1-u_{ij}}. \tag{5}$$

Therefore, the maximum likelihood estimate (MLE) of the latent rank of examinee i is given by

$$R_w^{(ML)}: \;\; w = \arg\max_{q \in Q} \sum_{j=1}^{n} \{u_{ij} \ln v_{qj} + (1 - u_{ij}) \ln(1 - v_{qj})\}. \tag{6}$$

We do not need a particular numerical iteration method like the Newton-Raphson method to obtain the MLE because each $\ln p(\boldsymbol{u}_i|R_q)$ obtained is a real number. The node R_q that makes $\ln p(\boldsymbol{u}_i|R_q)$ maximum is the MLE of the latent rank of examinee i. The above equation can also be used in selecting the winner node in the statistical learning process. The winner node selection method should be identical to the latent rank estimation method.

2.2 Graded neural test model

The graded neural test (GNT) model is a polytomous NTT model that can deal with polytomously ordered data. First, let $\boldsymbol{V}$ be the finally obtained reference matrix. Its q-th row vector is the RRV of R_q. That is,

$$\boldsymbol{v}_q = [\boldsymbol{v}'_{q1} \cdots \boldsymbol{v}'_{qn}]' \quad (q = 1, \cdots, Q), \tag{7}$$

where $\boldsymbol{v}_{qj}$ $(j = 1, \cdots, n)$ is the j-th subvector with size $(C_j - 1)$. The $(\sum_{l=1}^{j-1} l + k)$-th column vector, $\boldsymbol{\rho}_{jk}$, in $\boldsymbol{V}$ is the boundary category reference profile (BCRP; Shojima, 2007b) for category k of item j. From the BCRP, the item category reference profile (ICRP; Shojima, 2007b) is given by

$$\boldsymbol{\pi}_{jk} = \boldsymbol{\rho}_{jk} - \boldsymbol{\rho}_{jk+1} \quad (k = 0, \cdots, C_j - 1), \tag{8}$$

where

$$\boldsymbol{\rho}_{j0} = \mathbf{1}_Q, \tag{9}$$

and

$$\boldsymbol{\rho}_{jC_j} = \mathbf{0}_Q. \tag{10}$$

The BCRP is useful for reviewing the transition of the selection ratio of each item category or higher through latent ranks, while the ICRP is convenient for examining the behavior of each item category's selection ratio through latent ranks. In addition, let us assume an expanded reference matrix of $\boldsymbol{V}$, $\boldsymbol{\Pi}$ $(Q \times \sum_j C_j)$, that contains $\boldsymbol{\pi}_{jk}$s as its column vectors,

$$\boldsymbol{\Pi} = [\boldsymbol{\pi}_{11}, \cdots, \boldsymbol{\pi}_{1C_1-1}, \cdots, \boldsymbol{\pi}_{j1}, \cdots, \boldsymbol{\pi}_{j,C_j-1}, \cdots, \boldsymbol{\pi}_{n1}, \cdots, \boldsymbol{\pi}_{nC_n-1}] \quad \Big(Q \times \sum_{j=1}^{n} C_j\Big). \tag{11}$$

Next, let us assume that the response data of examinee i is

$$\boldsymbol{x}_i = \{x_{ij} | x_{ij} \in \{0, 1, \cdots, C_j - 1\}\}, \tag{12}$$

and that higher categories are selected by examinees with higher abilities. For the statistical learning process, $\boldsymbol{x}_i$ can be coded into

$$\boldsymbol{u}_i = [\boldsymbol{u}'_{i1} \cdots \boldsymbol{u}'_{in}]' \quad \Big(\Big(\sum_{j=1}^{n} C_j - n\Big) \times 1\Big), \tag{13}$$

$$\boldsymbol{u}_{ij} = \{u_{ijk} | u_{ijk} \in \{0, 1\}\} \quad ((C_j - 1) \times 1), \tag{14}$$

$$u_{ijk} = \begin{cases} 1, \text{ If } x_{ij} \geq k \ (k = 1, \cdots, C_j - 1) \\ 0, \text{ otherwise} \end{cases}, \tag{15}$$

where $\boldsymbol{u}_{ij}$ in (14) is identical with that in (4). This coding is necessary to maintain the ordinal sequence of the latent rank scale. In addition, the following y_{ijk}s are indispensable for determining the latent rank by the ML method,

$$\boldsymbol{y}_i = [\boldsymbol{y}'_{i1} \cdots \boldsymbol{y}'_{in}]' \quad \Big(\sum_{j=1}^{n} C_j \times 1\Big), \tag{16}$$

$$\boldsymbol{y}_{ij} = \{y_{ijk} | y_{ijk} \in \{0, 1\}\} \quad (C_j \times 1), \tag{17}$$

$$y_{ijk} = \begin{cases} 1, \text{ If } x_{ij} = k \ (k = 0, \cdots, C_j - 1) \\ 0, \text{ otherwise} \end{cases}, \tag{18}$$

From this, the probability that random variable Y_{ijk} obtained is 1 when examinee i's latent rank is R_q is written as

$$\Pr(Y_{ijk} = 1 | R_q) = \pi_{qjk}, \tag{19}$$

where π_{qjk} is the q-th element of $\boldsymbol{\pi}_{jk}$. Therefore, the likelihood of the response vector of examinee i for item j is given by

$$\Pr(\boldsymbol{y}_{ij} | R_q) = \prod_{k=0}^{C_j - 1} \pi_{qjk}^{y_{ijk}}. \tag{20}$$

Finally, under the assumption of local independence among items, the MLE of the latent rank of examinee i is given by

$$R_w^{(ML)} : \ w = \arg\max_{q \in Q} \sum_{j=1}^{n} \sum_{k=0}^{C_j - 1} y_{ijk} \ln \pi_{qjk}. \tag{21}$$

As mentioned above, the winner node should be also selected by the ML method when the latent rank is estimated by the ML method. Consequently, the statistical learning process of the GNT model is as follows:

For (t=1; $t \le T$; $t = t+1$) (L1)

— Obtain $\boldsymbol{X}^{(t)}$ after Randomly sort the row vectors of $\boldsymbol{X}$. (L2)

For (h=1; $h \le N$; $h = h+1$) (L3)

— Obtain $\boldsymbol{u}_h^{(t)}$ and $\boldsymbol{y}_h^{(t)}$ from $\boldsymbol{x}_h^{(t)}$, the h-th row vector of $\boldsymbol{X}^{(t)}$, by Eq.(16). (L4)

— Obtain $\boldsymbol{\Pi}_h^{(t)}$ from $\boldsymbol{V}_h^{(t)}$ by Eq.(11). (L5)

— Select the winner node for $\boldsymbol{y}_h^{(t)}$ by Eq.(21). (L6)

— Obtain $\boldsymbol{V}_h^{(t)}$ by updating $\boldsymbol{V}_{h-1}^{(t)}$. (L7)

— $\boldsymbol{V}^{(t+1)} \Leftarrow \boldsymbol{V}_N^{(t)}$. (L8)

Line (L7) can be written as

$$\text{For } (q{=}1;\ q \le Q;\ q = q+1) \qquad \text{(L7a)}$$

$$— \boldsymbol{v}_{qh}^{(t)} = \boldsymbol{v}_{qh-1}^{(t)} + h_{qw}(t)(\boldsymbol{u}_h^{(t)} - \boldsymbol{v}_{qh-1}^{(t)}),$$

where

$$h_{qw}(t|\alpha_t, \sigma_t^2) = \alpha_t \exp\Big\{-\frac{(R_q - R_w)^2}{2\sigma_t^2}\Big\}, \tag{22}$$

$$\alpha_t = \frac{T-t+1}{T}\alpha_1, \tag{23}$$

and

$$\sigma_t = \frac{(T-t)\sigma_1 + (t-1)\sigma_0}{T-1}. \tag{24}$$

Therefore, although the winner node is selected with reference to the expanded reference matrix $\boldsymbol{\Pi}$, but the reference matrix $\boldsymbol{V}$ should be numerically updated to maintain the order of the latent rank scale.

3 Rank Membership Profile

The MLE of the latent rank of each examinee is almost always uniquely determined by (6) under the dichotomous NTT model and by (21) under the GNT model. However, simply using these formulations can result in more than two ranks being selected as MLEs. This is possible not only with selection by the ML method but also with using the square of the Euclidian distance (ED^2). However, it is highly unlikely that more than two ranks are selected when T is large. When such cases do occurr, the node with the lowest rank is determined to be the winner. The final estimates are little effected by this treatment when T is large.

With selection by the maximum likelihood method, the probability of an examinee belongings to a latent rank can be estimated. Inference based on probability theory is a strong advantage of statistical models. Using (5), we can obtain under the dichotomous NTT model the probability that examinee i belongs to latent rank R_q, p_{iq}:

$$p_{iq} = \frac{\Pr(\boldsymbol{u}_i|R_q)}{\sum_{q=1}^{Q} \Pr(\boldsymbol{u}_i|R_q)} \tag{25}$$

Similarly, under the GNT model,

$$p_{iq} = \frac{\Pr(\boldsymbol{y}_i|R_q)}{\sum_{q=1}^{Q} \Pr(\boldsymbol{y}_i|R_q)} \tag{26}$$

Apparently, $\sum_q p_{iq} = 1$ is satisfied, and

$$\boldsymbol{p}_i = [p_{i1} \cdots p_{iQ}]' \quad (Q \times 1) \tag{27}$$

is the vector of p_{iq}s. This vector is called the "rank membership profile by the maximum likelihood method (RMP-ML)". In addition, the rank membership distribution (RMD) is obtained by

$$F = \left\{ f_1, \cdots, f_Q | f_q = \sum_{i=1}^{N} p_{iq} \right\} \tag{28}$$

Furthermore, a similar profile can be obtained under selection by the square of the Euclidian distance; that is,

$$\boldsymbol{d}_i = \{d_{iq}\} \quad (Q \times 1) \tag{29}$$

where

$$d_{iq} = \frac{||\boldsymbol{v}_q - \boldsymbol{u}_i||^2}{n}, \tag{30}$$

This is called the "rank membership profile by the square of Euclidian distance per item (RMP-ED^2)".

4 Evaluation

4.1 Dichotomous neural test model

The analysis results for a world history test are presented here as an example. The sample size was 2049, and the number of items was 36. The proposed ML method was used to select the winner node in the learning process. Parameters Q, T, α_1, σ_1, and σ_0 were set to 10, 100, 0.1, 10, and 1.0, respectively, and the setup was identical to that used previously (Shojima, 2007a, section 3.1). Figure 1 shows the item reference profiles (IRPs) obtained; they were slightly different from those previously obtained (Shojima, 2007a, Figure 3).

Figures 2(a)-(d) show the test reference profile (Shojima, 2007a), the latent rank distribution by the ML method, the rank membership distribution by the ML method, and the scatter plot of the scores and the latent MLE ranks. This marginal information is almost equivalent to that of Shojima (2007a, Figure 4). However, the latent rank distribution and rank membership distribution are very different. Further research is required to clarify the reason for the difference. In addition, Figure 2(e) is a scatter plot of the latent ranks by the ML and ED^2 methods. From the figure, it is clear that these two obtained ranks are nearly equal. Spearman's rank correlation coefficient between the two ranks was 0.959.

Figure 3(a) shows the rank membership profiles by the ML method for examinees 1–16, and 3(b) shows them by the ED^2 method for the same examinees. In general, the latent rank giving the smallest ED^2 had the largest membership probability by the ML method. That is, the location of the ravine bottom in the RMP-ED^2 results was generally that of the mountain peak in the RMP-ML results. However, the profile obtained with the ML method is more convenient than that with the ED^2 method because the peak for each RMP-ML profile is easily found, while the ravine bottom of each RMP-ED^2 profile is not.

The RMP-ML profiles directly tell us the probability of each examinee's membership in each latent rank. For example, although the latent ranks for both examinees 13 and 15 were estimated to be six, their RMP-ML profiles show that the reliability of the one for examinee 15 is higher.

4.2 Graded neural test model

Here we present an example for the graded neural test (GNT) model. The data analyzed was for an earth science test. The sample size was 1424, and the number of items was 9. Each item had four categories $(= 0, 1, 2, 3)$. Parameters Q, T, α_1, σ_1, and σ_0 were set to 10, 500, 0.1, 10, and 1.0, respectively, the same as previously (Shojima, 2007b, section 3.1).

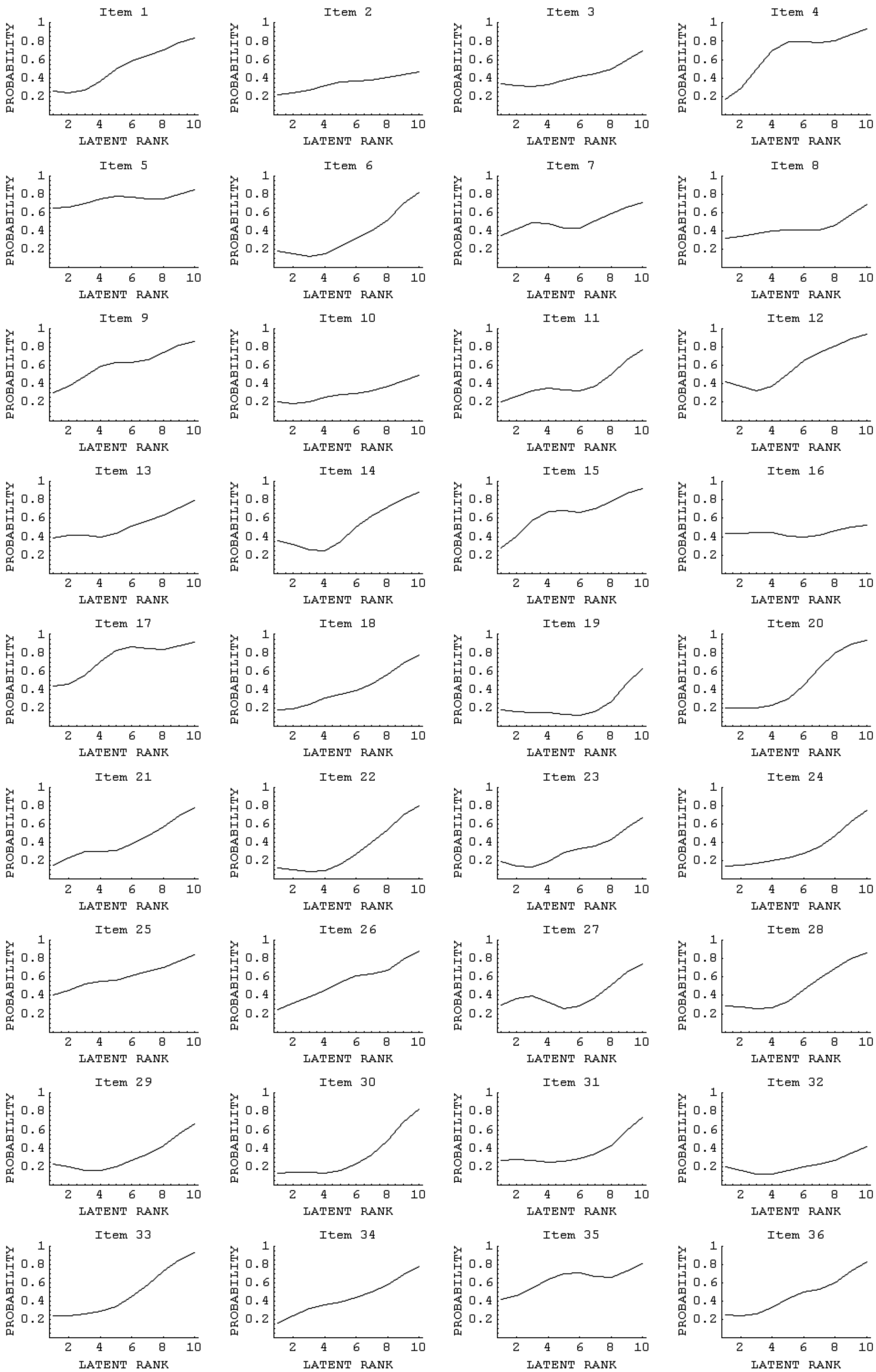
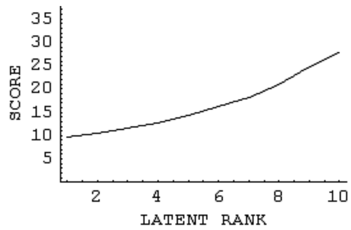
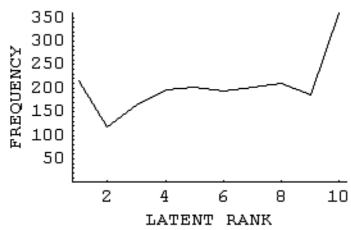


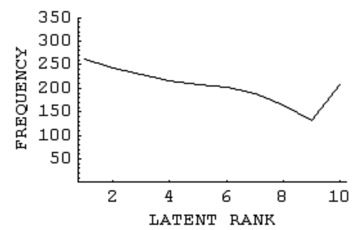
Figure 1: Item Reference Profiles



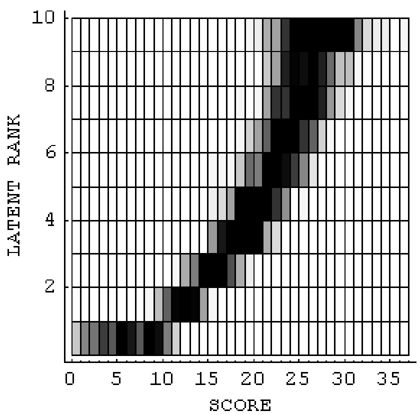
(a) Test Reference Profile



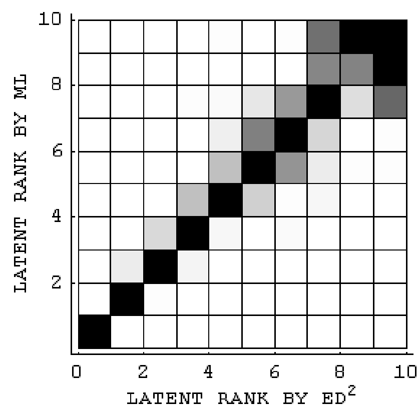
(b) Latent Rank Distribution



(c) Rank Membership Dist.



(d) Scatter Plot of Scores and Ranks



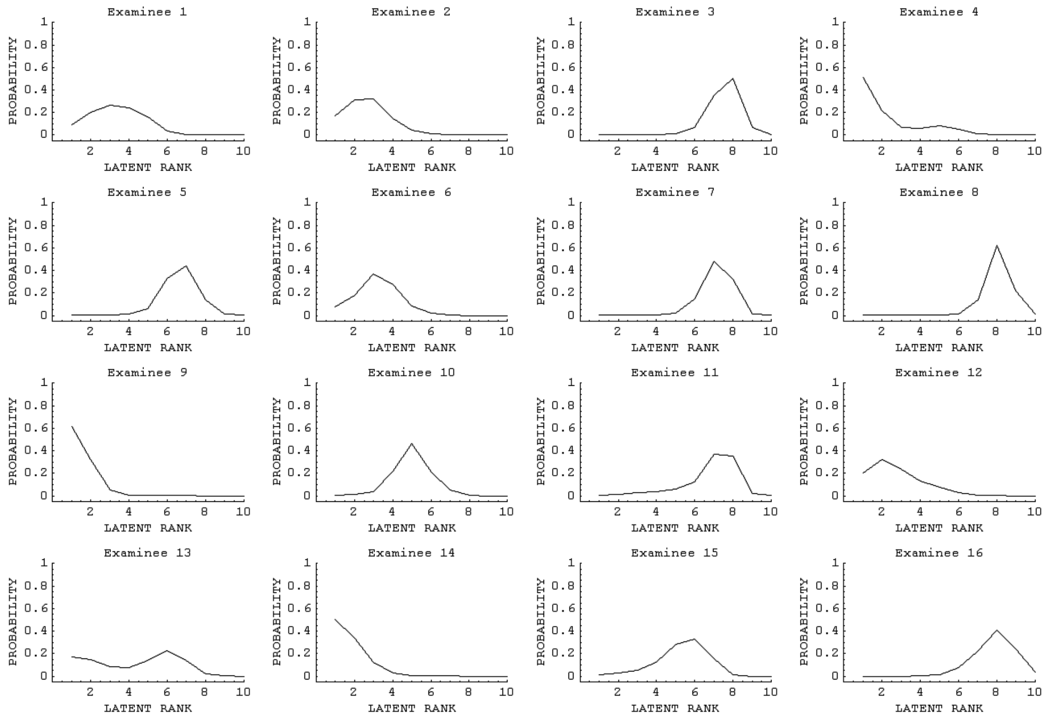
(e) Scatter Plot of Ranks of ML and ED^2

Figure 2: TRP, LRD, RMD, and Scatter Plots

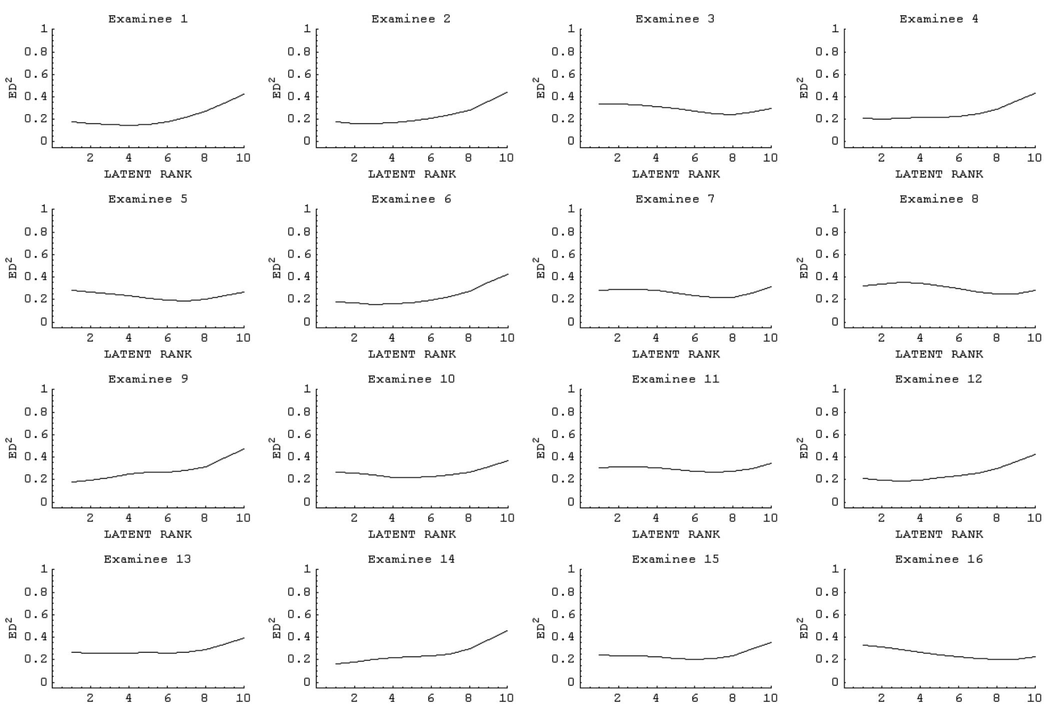
Figures 4(a) and 4(b) show the BCRPs and ICRPs, respectively. Although their general shapes were the same as for the previous ones (Shojima, 2007b, Figures 3(a) and (b)), there were some slight differences.

Figures 5(a)-(d) show the TRP, the latent rank distribution, the rank membership distribution, and the scatter plot of the scores and the MLE ranks. Figures 5(b) and 5(c) are very different, as seen in the comparison between Figures 2(b) and 2(c). Also, Figure 5(c) is smoother than Figure 5(b). In addition, Figure 5(e) is a scatter plot of the two latent ranks estimated by the ML method and the ED^2 method. The difference between the two ranks under the GNT model was larger than that under the dichotomous NTT model, and Spearman's rank correlation coefficient between the two ranks was 0.918.

Figures 6(a) and 6(b) show the rank membership profiles estimated by the ML method and the ED^2 method, respectively. As seen in the RMPs under the dichotomous NTT model, it is easier to find the mountain peaks of RMP-MLs than the ravine bottoms of RMP-ED^{2}s.

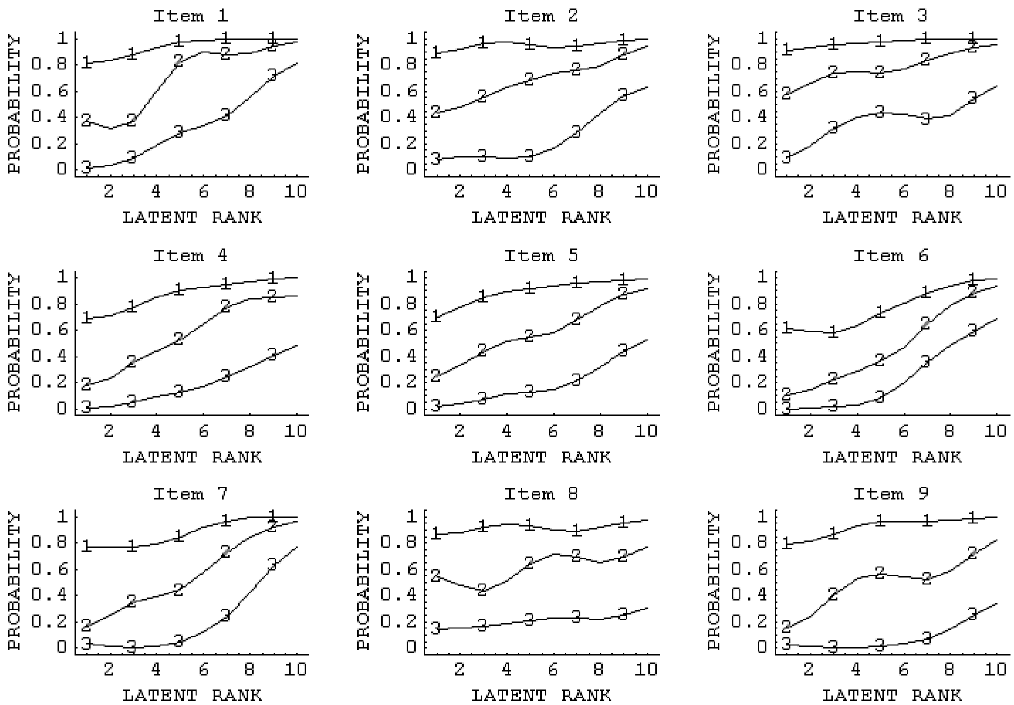


(a) RMP-ML

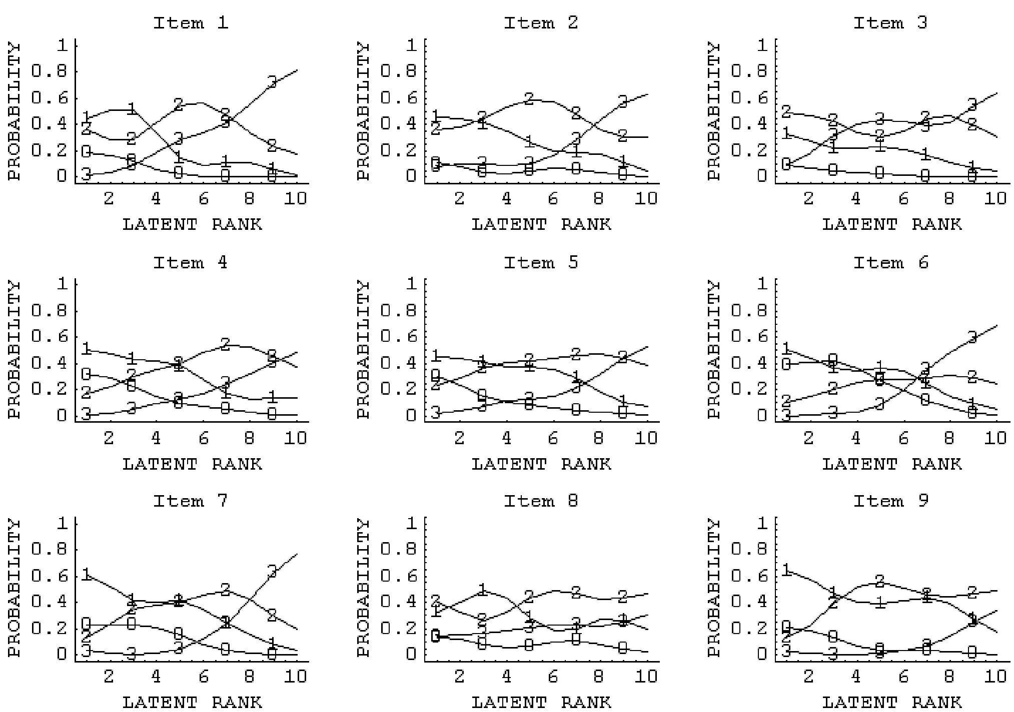


(b) RMP-ED^2

Figure 3: Rank Membership Profiles of Examinees 1-16

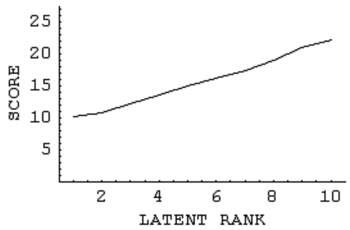


(a) Boundary Category Reference Profiles

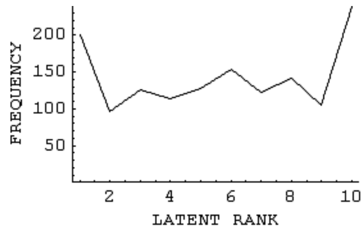


(a) Item Category Reference Profiles

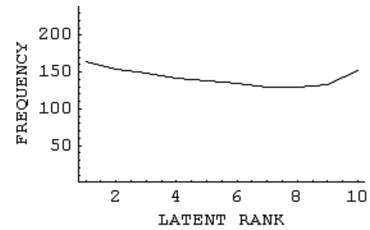
Figure 4: BCRPs and ICRPs



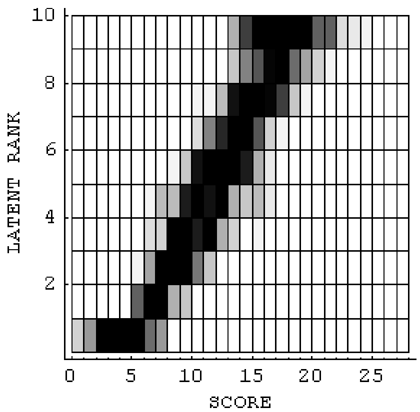
(a) Test Reference Profile



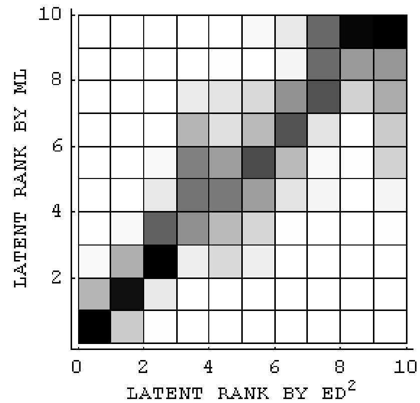
(b) Latent Rank Distribution



(c) Rank Membership Dist.



(d) Scatter Plot of Scores and Ranks



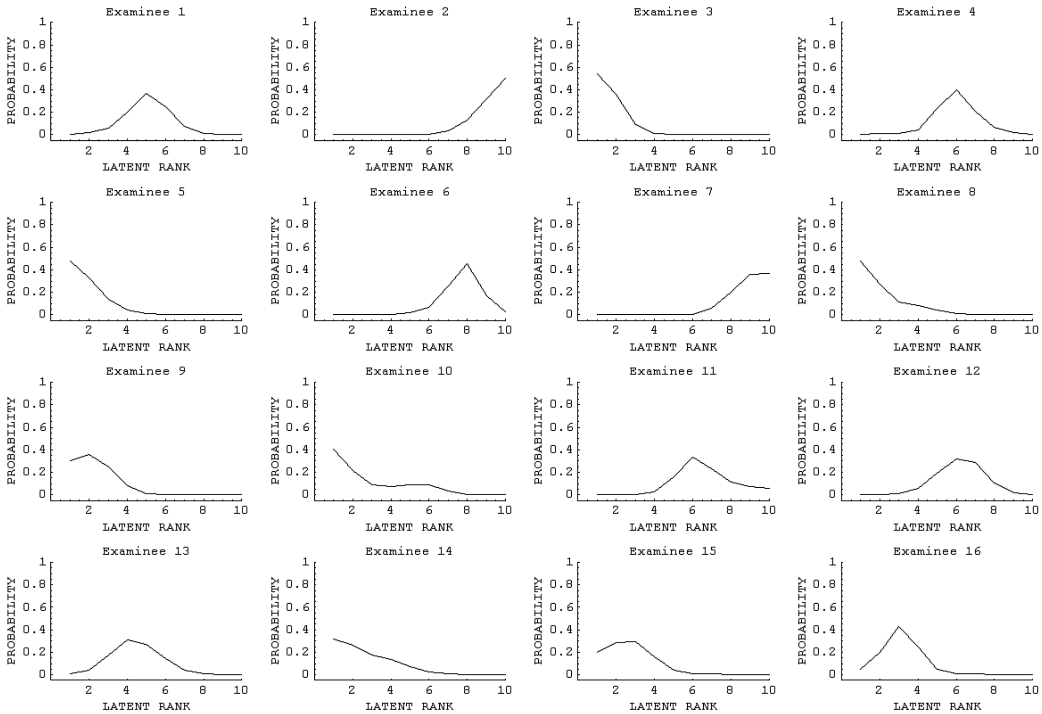
(e) Scatter Plot of Ranks of ML and ED^2

Figure 5: TRP, LRD, RMD, And Scatter Plots

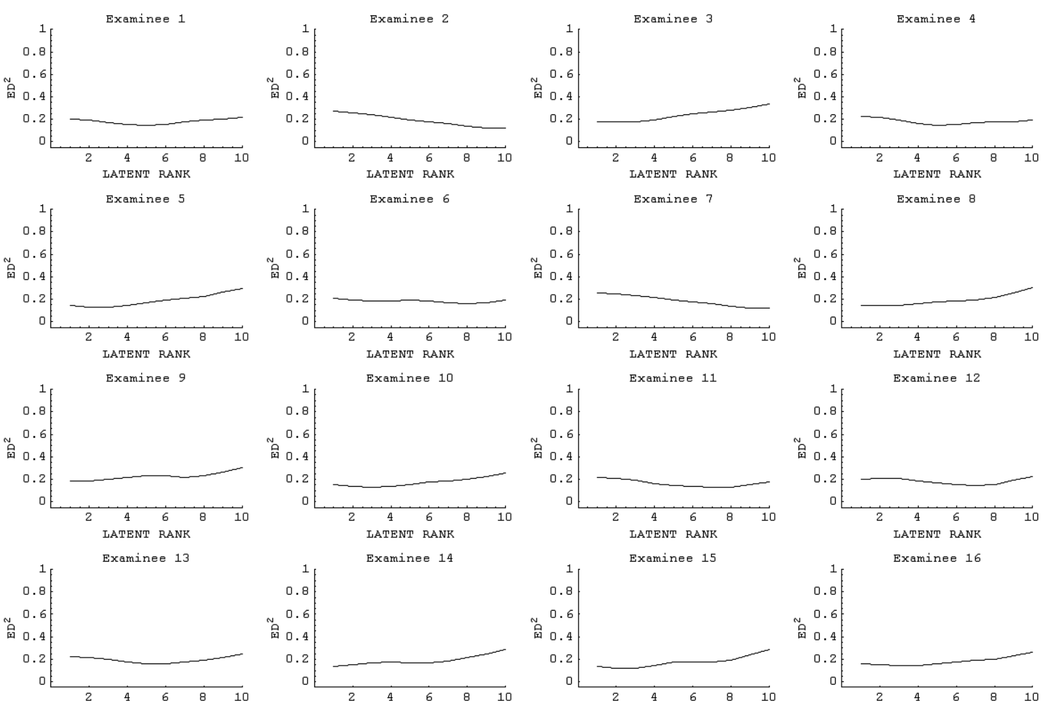
5 Discussion

Our evaluation of estimatiing the latent rank by the maximum likelihood (ML) method under the dichotomous neural test theory (NTT) model and under the graded neural test (GNT) model show that the ML method is useful for examining the accuracy of the latent rank estimated for each examinee. It also showed that the rank membership profile (RMP) obtained by the ML method is useful for reviewing the behavior of each examinee's membership probabilities through latent ranks. The ML method is compatible with the monotonically increasing constraint on the IRPs and the BCRPs (Shojima, 2007a, 2007b) in the statistical learning process.

The NTT model was developed as a mathematical model rather than a statistical model. The dichotomous NTT model (Shojima, 2007a) is a simple application of the 1-dimensional self-organizing map (SOM) for binary data, so the SOM can be said to be mathematical rather than statistical. The GNT model (Shojima, 2007b), which is a polytomous NTT model for polytomously rank-ordered data, reveals the uniqueness of the NTT because the data coded in the GNT learning process is not used in the SOM, so the boundary category



(a) RMP-ML (GNT Model)



(b) RMP-ED^2 (GNT Model)

Figure 6: Rank Membership Profiles of Examinees 1-16 (GNT Model)

reference profile (BCRP) and the item category reference profile (ICRP) developed the NTT's uniqueness. Nevertheless, the NTT models are not statistical but mathematical.

A statistical feature was thus implemented in the NTT. This means that the NTT can provide various statistical benefits such as a goodness-of-fit index for the whole model, for each item, and for each examinee. However, we must make some statistical assumptions. For example, we must assume local independence among items when calculating the likelihood, as seen in equations (5) and (20).

References

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

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

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.