Behaviormetrika, Vol. 23, No. 2, 153-167
1996

AN ITEM RESPONSE MODEL FOR MULTIDIMENSIONAL ANALYSIS OF MULTIPLE-CHOICE DATA*

Yoshio Takane**

An item response model, similar to that in test theory, was proposed for multiple-choice questionaire data. In this model both subjects and item categories are represented as points in a multidimensional euclidean space. The probability of a particular subject choosing a particular item category is stated as a decreasing function of the distance between the subject point and the item category point. The subject point is assumed to follow a certain distribution, and is then integrated out to derive marginal probabilities of response patterns. A marginal maximum likelihood (MML) method was developed to estimate coordinates of the item category points as well as distributional properties of the subject point. Bock and Aitkin's EM algorithm was adapted to the MML estimation of the proposed model. Examples were given to illustrate the method, which we call MAXMC.

1. Introduction and motivation

We propose a probabilistic multidimensional model for unordered categorical data. Such data arise, for example, when we ask a group of subjects in attitude surveys to endorse attitude statements expressing views close to their own, or in personality inventory to choose adjectives which adequately describe their own behavioral disposition. As a concrete example, let us look at the multiple-choice questionaire items given in the appendix. These are a sample of six questions drawn from a large scale survey on Japanese national characters conducted at the Institute of Statistical Mathematics in Tokyo (Hayashi, 1982). There are three response options for each item, from which the subjects are to choose the one that best fits their own view. Such data may be regarded as representing proximity relations between the subjects and the item categories. Systematic individual differences are common in such data, giving rise to dependencies among observations obtained from a same subject (Takane and de Leeuw, 1987). An important consideration in modeling such data is how to incorporate systematic individual differences in characterizing the item categories.

One way to allow for systematic individual differences in a model is to introduce subject parameters. In the model proposed in this paper both item categories

---

Key Words and Phrases ; proximity items, marginal maximum likelihood (MML) estimation, EM algorithm, factor analysis of discrete data

* The work reported here is based on the presentation at the Psychometric Society Meeting in Los Angeles in June, 1983 under the same title. It has been supported by research grant A6394 from Natural Sciences and Engineering Research Council of Canada. Requests for reprints should be addressed to Yoshio Takane, Department of Psychology, McGill University, 1205 Dr. Penfield Ave., Montréal, Québec H3A 1B1, Canada.

** Department of Psychology, McGill University, 1205 Docteur Penfield Avenue, Montréal, Québec H3A 1B1, Canada

and subjects are represented as points in a multidimensional euclidean space. The probability of a particular subject choosing a particular item category is stated as a decreasing function of the distance between the two points. The model is a combination of the unfolding model (Coombs, 1964) for a spatial representation of the item category and the subject points, and Luce's choice model (Luce, 1959) connecting the interpoint distances to choice probabilities.

The introduction of subject parameters, however, creates a statistically undesirable condition. The number of subject parameters increases linearly with the number of observations. Such parameters are called incidental parameters. In the presence of incidental parameters, asymptotic properties of maximum likelihood estimators (MLE) never hold. In particular, MLE may not be consistent. To avoid this difficulty, a subject point is assumed to follow a certain distribution, and is then integrated out to obtain marginal probabilities of response patterns. The marginal maximum likelihood (MML) method is then used to estimate the coordinates of item category points and parameters characterizing the distribution of the subject point. The idea is similar to that in the item response test theory (Bock & Lieberman, 1970), where essentially the same problem exists. We call our model an item response model, although the target data type for the proposed model is essentially different from that of the traditional item response models for test data. We call our method MAXMC, MAXimum likelihood IRT models for Multiple-Choice data.

In the next section we present the proposed model and the marginal maximum likelihood method for parameter estimation in some detail. We then discuss an EM algorithm for maximizing the marginal likelihood. We then introduce three examples of application for illustration. We conclude the paper with discussion.

2. The model

Suppose a group of N subjects have responded to a set of I items, each having J_i ($i=1, \cdots, I$) response categories. The subjects may be classified by response patterns which are indexed by k. Define

$$g^*_{ki(j)}=\begin{cases}1, & \text{if option } j \text{ of item } i \text{ is chosen in response pattern } k\\ 0, & \text{otherwise.}\end{cases} \tag{1}$$

When there are missing data, the corresponding $g^*_{ki(j)}$ may be set equal to zero for all j. Let f_k denote the observed frequency of response pattern k. We assume that item categories are represented as points in an A dimensional euclidean space. Let $x_{i(j)a}$ denote the coordinate of category j of item i on dimension a ($j=1, \cdots, J_i$; $i=1, \cdots, I$; $a=1, \cdots, A$). We require for each i and for all a

$$\sum_j g^*_{i(j)}x_{i(j)a}=0, \tag{2}$$

to avoid the item category points drifting away from origin, where

$$g^*_{i(j)}=\sum_k f_k g^*_{ki(j)}. \tag{3}$$

We assume that a subject point is also represented in the same A dimensional euclidean space. Let $\boldsymbol{y}=(y_1, \cdots, y_A)'$ be a vector of coordinates of the subject point. We further assume that $\boldsymbol{y}$ is a random vector with its density function denoted by $h(\boldsymbol{y})$. Let $d_{i(j)}(\boldsymbol{y})$ denote the euclidean distance between an item category point (category j of item i) and a subject point, $\boldsymbol{y}$. This is defined as

$$d_{i(j)}(\boldsymbol{y})=\left\{\sum_a^A (x_{i(j)a}-y_a)^2\right\}^{1/2}. \tag{4}$$

Let $p_{i(j)}(\boldsymbol{y})$ denote the conditional probability of the subject at $\boldsymbol{y}$ choosing category j in item i. We assume that this is given by

$$p_{i(j)}(\boldsymbol{y})=\frac{\exp(-d^v_{i(j)}(\boldsymbol{y}))}{\sum_k \exp(-d^v_{i(k)}(\boldsymbol{y}))}. \tag{5}$$

The model postulates that each category has "response strength", $\exp(-d^v_{i(j)}(\boldsymbol{y}))$, which is a decreasing function of $d_{i(j)}(\boldsymbol{y})$, and that a particular category is chosen with probability proportional to its response strength relative to that of other response categories within the same item. (The denominator of (5) is just a normalization factor to make $\sum_j^{J_i} p_{i(j)}(\boldsymbol{y})=1$.) The proposed model combines Coombs' (1964) unfolding model for the representation of stimuli (item categories and subjects) and Luce's (1959) choice model for the response mechanism. The v is a prescribed power, set equal to 1.0 or 2.0. It modulates the shape of the response strength function. When $v=1.0$, the response strength function is of exponential form, and when $v=2.0$, it is of Gaussian form. The exponential form of the response strength function was initially proposed by Shepard (1957) in a stimulus generalization context, while the Gaussian form advocated by several authors (e.g., Nosofsky, 1986; Takane & Shibayama, 1986) as a model of stimulus identification. There has been a controversy as to the relative efficacy of the two forms on both theoretical and empirical grounds (Shepard, 1986; Ennis, 1988; Takane & Shibayama, 1992). We tend to prefer the Gaussian form, since the exponential form is not feasible in the unidimensional case. This is because with $v=1.0$ and $A=1$ the likelihood function is completely flat outside the range of stimuli, so that the locations of two extreme stimuli are indeterminable. In addition, the Gaussian form tends to fit the data better in the present context, albeit usually only slightly.

The conditional probability, $P_k(\boldsymbol{y})$, of response pattern k given $\boldsymbol{y}$ is stated as

$$P_k(\boldsymbol{y})=\prod_i \prod_j p_{i(j)}(\boldsymbol{y})^{g^*_{ki(j)}}. \tag{6}$$

The marginal probability, P_k, of response pattern k is then given by

$$P_k=\int P_k(\boldsymbol{y})h(\boldsymbol{y})d\boldsymbol{y}. \tag{7}$$

We assume

$$y \sim N(\mathbf{0}, \Sigma), \tag{8}$$

where Σ is assumed diagonal with the diagonal elements denoted by σ_a^2, $a=1, \cdots, A$. This can be done without loss of generality, because coordinate axes can always be set in the principal axes orientation. The assumption of zero means is a restrictive one, however, and its validity should be empirically verified. See the next section for how this can be done. The integral above may be approximated by a finite sum,

$$P_k = \sum_q P_k(\mathbf{y}_q) B(\mathbf{y}_q), \tag{9}$$

at selected points $\mathbf{y}_q$. A special table is available for the quadrature weight, $B(y_{qa})$, where $B(\mathbf{y}_q)=\prod_a B(y_{qa})$. The likelihood of the entire set of observations is now stated as

$$L = \prod_k P_k^{f_k}, \tag{10}$$

where k is taken over all possible response patterns. We determine model parameters, $\mathbf{X}=\{(x_{i(j)a}\}$ and the diagonal elements of Σ, that is, σ_a^2, $a=1, \cdots, A$, so as to maximize the likelihood.

Once (10) is maximized, we may use AIC (Akaike, 1974) or ABIC (Akaike, 1980) for goodness of fit (GOF) comparisons among competeting models. Since we are dealing with the marginal likelihood here, the latter reduces to AIC defined on the marginal likelihood. AIC is defined by

$$AIC(\pi) = -2 \ln L^*(\pi) + 2n_\pi, \tag{11}$$

where π refers to a particular model being fitted, $L^*(\pi)$ is the maximum likelihood under model π, and n_π is the effective number of parameters in model π. The model associated with the smallest value of AIC is considered the best fitting model. The dimensionality in the distance model may be determined based on the minimum AIC criterion. The effective number of parameters is calculated by

$$n_\pi = A\Big(\sum_i (J_i - 1)\Big) + A, \tag{12}$$

where A is the dimensionality of the solution space, and J_i is the number of response categories in item i. One is subtracted from J_i before multiplied by A because of the constraint, (2), and A in the second term is added because variances of $\mathbf{y}$ are estimated.

IRT models, different from the one proposed above, have previously been proposed for unordered categorical data. Bock (1972) was probably the first to propose one (also, see Hoijtink, 1990). His model (as well as Hoijtink's Parella), however, was restricted to a single dimension. Bock and Aitkin (1981) extended this model to a multidimensional case, who also used the MML estimation method and the EM algorithm. Bock and Aitkin's method was further elaborated by Bock, Gibbons, and Muraki (1988). Bartholomew (1987) also proposed a model called RF (Response Function) model similar to Bock and Aitkin's model. All the models mentioned above for multidimensional analysis of unordered categorical data use scalar products in the exponents that define the response strength functions for item categories. MAXMC, on the other hand, uses the negative (squared) euclidean distances in the exponents, following the idea of Coombs' unfolding model for preference choice data.

3. An EM algorithm

We use an EM algorithm to maximize the log of the marginal likelihood stated above. The derivation of the algorithm heavily draws on Bock and Aitkin's (1981). The algorithm alternates the following two steps until convergence is reached.

E-step. For fixed $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$, calculate

$$\bar{g}_{qi(j)} = B(\boldsymbol{y}_q) \sum_k g_{ki(j)} P_k(\boldsymbol{y}_q)/P_k, \tag{13}$$

where $g_{ki(j)} = f_k g^*_{ki(j)}$.

M-step. For each i separately, maximize

$$l_i = \sum_j \sum_q \bar{g}_{qi(j)} \ln p_{i(j)}(\boldsymbol{y}_q) \tag{14}$$

with respect to $x_{i(j)a}$ ($j=1, \cdots, J_i$; $a=1, \cdots, A$), and after all $x_{i(j)a}$'s ($i=1, \cdots, I$) are updated, maximize

$$l = \sum_i l_i \tag{15}$$

with respect to σ_a^2 ($a=1, \cdots, A$). Maximizations in the M-step may be carried out by Fisher's scoring algorithm. This algorithm is particularly attractive in the present context, because the convergence is very fast, and the number of parameters to be updated simultaneously is relatively small, since $\boldsymbol{X}_i$, i^{th} subblock of $\boldsymbol{X}$, can be independently updated for each i. The maximization of l with respect to σ_a^2 is, however, conditional on the current estimate of $\boldsymbol{X}$. The overall convergence rate of the EM algorithm can still be very slow. It may be useful to switch to the scoring algorithm to update all parameters simultaneously in the last few iterations, if the total number of parameters to be estimated is not too large (say, less than 50). The observed information matrix necessary for the scoring algorithm can be obtained by the methods proposed by Louis (1982) and by Lang (1992). This has the side benefit of yielding asymptotic variance-covariance estimates of estimated parameters. This provision has not been implemeted, however. We use a quantification method III (Q3; Hayashi, 1952) solution as an initial estimate of $\boldsymbol{X}$, and we set $\sigma_a^2=1$ for all a initially.

After the convergence is reached, a subject point can *a posteriori* be estimated for each response pattern. This is analogous to the factor score estimation in factor analysis. We apply the EAP (Bayes expected *a posteriori*) estimation method to estimate subject points (Bock and Aitkin, 1981), namely

$$E(\mathbf{y} \mid \mathbf{g}_k)=\tilde{\mathbf{y}}_k=\sum_q \mathbf{y}_q P_k(\mathbf{y}_q)B(\mathbf{y}_q)/P_k. \tag{16}$$

Variance-covariance estimates of the EAP estimators are obtained by

$$V(\mathbf{y} \mid \mathbf{g}_k)=Cov(\tilde{\mathbf{y}}_k)=\sum_q (\tilde{\mathbf{y}}_k-\mathbf{y}_q)(\tilde{\mathbf{y}}_k-\mathbf{y}_q)' P_k(\mathbf{y}_q)B(\mathbf{y}_q)/P_k. \tag{17}$$

As alluded to earlier, the distributional assumption on $\mathbf{y}$, (8), can be relaxed by estimating an empirical distribution of $\mathbf{y}$. This could make the model, (7), more in line with the observed data. It is done by re-estimating quadrature weights in each iteration according to the following formula,

$$\tilde{B}(\mathbf{y}_q)=\frac{B(\mathbf{y}_q)\sum_k f_k P_k(\mathbf{y}_q)}{\sum_{q'} B(\mathbf{y}_{q'})\sum_k f_k P_k(\mathbf{y}_{q'})}. \tag{18}$$

Estimating the quadrature weights, however, also means that a substantially larger number of parameters are to be estimated. It is thus only worthwhile when $h(\mathbf{y})$ significantly deviates from the assumed distribution of MVN.

4. Examples

In this section we discuss three examples of application. The first two have previously been analyzed by Bartholomew (1987) using his method. In reporting the results of our analyses on these data sets we will draw some comparison with his results.

4.1 Staff assessment data

The first example data set comes from Bartholomew (1987). Each of 405 managers was originally assessed on 13 aspects of his/her work using a 5-point rating scale. For illustration only three of them were used. Also, rarely used categories were combined to yield a $4\times3\times3$ contingency table. Although the categories are ordered, they were treated as if unordered as in Bartholomew (1987). The data are presented in Table 3, where response patterns are listed in order of estimated component score. Eight patterns (113, 213, 313, 411, 412, 413, 421, and 431) were unobserved out of possible 36 patterns.

Table 1 provides a summary of GOF statistics (AIC and n_π) obtained by

Table 1
Goodness of fit comparison: Staff assessment data

Saturated Model	40.4 (35)		
Number of Quadrature Points	3-point	7-point	15-point
dim=1	0.2 (8)	0.2 (8)	0.2 (8)
dim=2	8.4 (16)	8.4 (16)	

Main entries in the table are AIC values.
Effective numbers of parameters are given in parentheses.

Table 2
Parameter estimates: Staff assessment data

Item	Category	Number of Quadrature Points	
		7-point	15-point
1	1	.62	.63
	2	.24	.24
	3	−.28	−.28
	4	−.89	−.91
2	1	1.48	1.48
	2	.04	.04
	3	−.32	−.33
3	1	1.11	1.11
	2	.31	.31
	3	−.66	−.66

Variance Estimate=.750

MAXMC. All the analyses were done with $v=2.0$. The number of quadrature points was varied from 3 to 15 to examine its effects on the GOF statistic and on parameter estimates. Only a negligible effect of the number of quadrature points was found on the GOF statistic. According to the minimum AIC criterion a single dimension is sufficient to capture the variations in the three categorical variables predicting success in the job. This is consistent with Batholomew's (1987) finding. Table 2 provides estimates of parameters for the best fitting model in two different numbers of quadrature points. Hardly any differences are observed between the two sets of estimates. It seems that the number of quadrature points is not so crucial in MAXMC. Locations of the category points are consistent with their *a priori* order and (inversely) with that obtained by Bartholomew. Items 2 and 3 are slightly more predominant than item 1 in characterizing the derived dimension. Table 3 presents EAP estimates of coordinates (component scores) of response patterns along with standard errors of the estimates. The standard errors are rather large, but this is due to the small number of items (only three items) in this data set. The distribution of the subject points (component scores) is slightly positively skewed, but not to the extent that required re-estimations of quadrature weights.

4.2 Employment in small industry data

The second example also comes from Bartholomew (1987). The data set analyzed was originally collected by Leimu (1983). The study was on a sample of 469 employees from small industry in Finland, who responded to the following three questions:

Table 3
EAP estimates of coordinates of response patterns:
Staff assessment data.

Response Pattern	Observed Frequency	Coordinate	Standard Error
111	1	1.87	.53
211	7	1.66	.52
112	2	1.44	.51
311	1	1.39	.51
212	13	1.25	.51
121	3	1.11	.52
312	5	.97	.53
131	1	.91	.53
221	10	.90	.53
231	5	.69	.55
122	12	.66	.55
321	3	.59	.55
132	4	.43	.56
222	64	.42	.56
331	1	.36	.56
232	36	.18	.57
322	38	.08	.58
123	1	.04	.58
332	31	−.17	.58
133	1	−.21	.59
223	37	−.22	.59
422	4	−.35	.59
233	23	−.48	.59
323	34	−.59	.60
432	3	−.61	.60
333	41	−.86	.61
423	5	−1.05	.62
433	11	−1.35	.64
Total	405		

1. Was there any alternative choice of job when coming to your present job? (1=no, 2=don't know, 3=yes)
2. Is the job permanent? (1=very unsure or quite unsure, 2=don't know, 3=quite sure or very sure)
3. Were you unemployed in the last three years? (1=no, 2=yes)

The data are presented in Table 6. Again, response patterns are listed in order of estimated component score.

As before, all the analyses were performed with $v=2.0$. Table 4 compares the GOF of various models. Again, the effect of the number of quadrature points is minimal on the GOF statistic, and the unidimensional model has turned out to be the

Table 4
Goodness of fit comparison:
Employment in small industry data (Leimu, 1983).

Saturated Model	17.8 (17)		
Number of Quadrature Points	3-point	7-point	15-point
dim=1	5.8 (6)	6.0 (6)	6.0 (6)
dim=2	9.1 (12)	9.2 (12)	

Main entries in the table are AIC values.
Effective numbers of parameters are given in parentheses.

Table 5
Parameter estimates: Employment in
small industry data (Leimu, 1983).

Item	Category	7-Point Quadrature
1	1	.25
	2	.04
	3	−.17
2	1	.59
	2	.27
	3	−.22
3	1	−.12
	2	.61

Variance Estimate=.571
Estimates based on the 15-point quadrature are almost identical.

best fitting model according to the minimum AIC criterion. The derived dimension represents the ease with which employees can find secure employment. Table 5 presents estimates of parameters. They are given only for the 7-point quadrature, but the estimates obtained under different numbers of quadrature points are virtually indistinguishable. One interest in Bartholomew's study was to see if the "don't know" category in questions 1 and 2 indeed fall between the "yes" and the "no" categories, as often assumed. This has been confirmed in the present study as well as in Bartholomew's. Question 2 seems to be most discriminating of employment security, as indicated by the fact that response categories of this item receive the widest range of scores. Table 6 provides EAP estimates of coordinates of subject points along with their standard errors. The standard errors of the estimated subject points are rather large, but this is again due to the small number of items (only three items) in this data set.

4.3 ISM data on traditional vs modern views

The third and last example pertains to a data set collected at the Institute of

Table 6
EAP estimates of coordinates of response patterns: Employment in small industry data (Leimu, 1983).

Response Pattern	Observed Frequency	Coordinate	Standard Error
331	145	−.41	.66
231	54	−.23	.65
131	72	−.05	.64
321	33	−.01	.64
221	22	.16	.63
332	17	.22	.63
311	24	.25	.63
121	14	.33	.63
232	9	.38	.62
211	9	.41	.62
132	11	.54	.62
111	21	.57	.61
322	7	.58	.61
222	6	.74	.61
312	6	.82	.61
122	7	.89	.61
212	2	.97	.60
112	10	1.12	.60
Total	469		

Statistical Mathematics (ISM). This was part of a large scale survey on Japanese national characters conducted in Japan every five years since 1952. Questions used in the present study represent various aspects of traditional and modern views on Japanese society and culture, and are listed in the appendix. There are six questions each with three response categories. The sample size was over 3,000. Table 7 gives a summary of GOF statistics. When dim=2, the model was fitted with both v=2.0 and v=1.0. The v=2.0 fitted the data better. While this is confirmed only for dim=2, it is not likely that this tendency is reversed for other dimensionalities.

Table 7
Goodness of fit comparison: ISM data

Saturated	1,168.9 (728)		
	d^2		d
Number of Quadrature Points	3-point	5-point	3-point
dim=1	23.2 (13)	22.5 (13)	
dim=2	18.3 (26)	19.0 (26)	46.3 (26)
dim=3	28.4 (39)		

Main entries in the table are AIC values.
Effective numbers of parameters are given in parentheses.

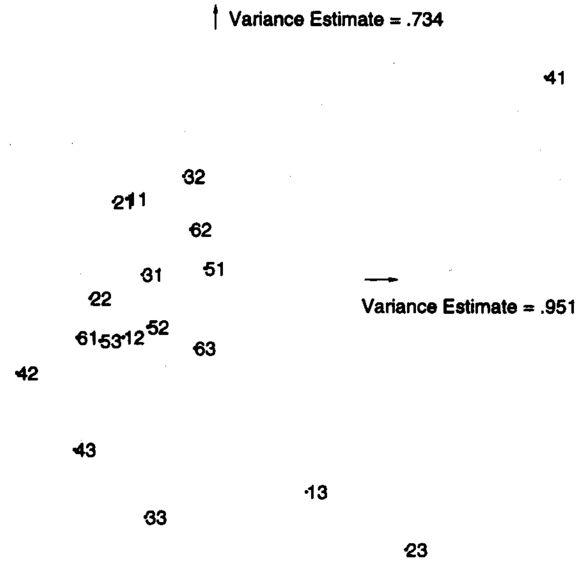

Fig. 1 The two-dimensional item category configuration for the ISM data

The two dimensional solution obtained under $v=2.0$ yields the best fitting model according to the minimum AIC criterion.

Figure 1 depicts the best fitting solution. Only the item category points are represented in this figure. The points are labelled by a pair of digits, the first one

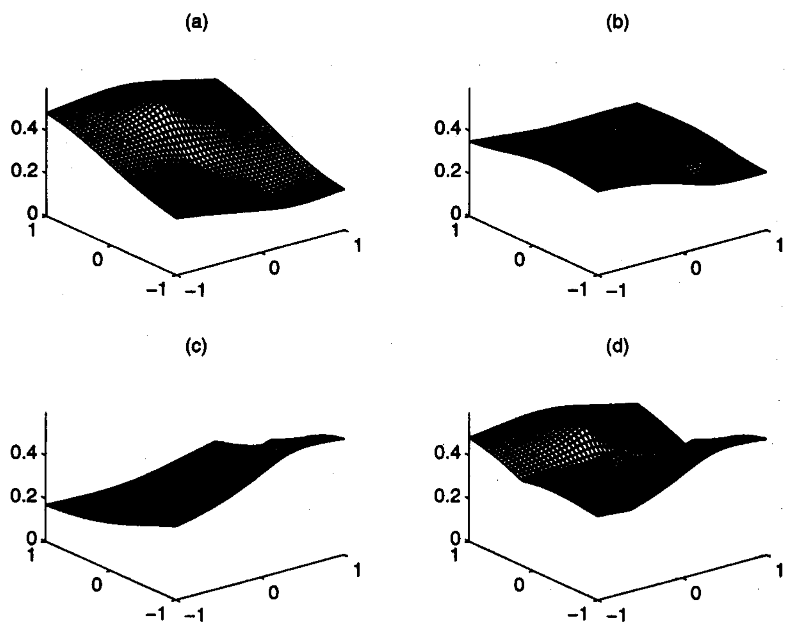

Fig. 2 Conditional probability surfaces for three response categories of item 1 in the ISM data. (a) Category 1, (b) Category 2, (c) Category 3, and (d) Maximum of the three conditional surfaces.

designating the item (question) number while the second the category number within the item, so 41, for example, indicates response category 1 of item 4. The top portion of the configuration indicates more traditional views, characterized by such item categories as 11, 21, 32, 41, etc. The middle left portion represents more modern views, featured by 12, 22, 31, 42, etc., and the bottom portion indicates indecisiveness (13, 23, and 33).

The proposed model relates the distance between an item category point and a subject point to the probability of choice via model (5). Item characteristic (conditional probability) surfaces can be drawn by evaluating $p_{i(j)}(\boldsymbol{y})$ at different values of $\boldsymbol{y}$. The item characteristic surfaces are displayed in Figures 2a, b, and c for response categories 1, 2, and 3, respectively, of item 1. In each figure there is a peak (the point where the conditional probability takes a maximum value) corresponding to the location of the category to be chosen, and two dips corresponding to the locations of the other categories of the same item. Figure 2d depicts the maximum conditional probability surface, i.e., $\max_j p_{i(j)}(\boldsymbol{y})$ as a function of $\boldsymbol{y}$. Boundaries may be seen where the most dominant category shifts from one to another that defines the maximum conditional probability surface. Similar pictures can be drawn for other items.

5. Discussion

In this paper we presented an item response model for multiple-choice questionaire data along with an MML method and the associated EM algorithm for parameter estimation. MAXMC is useful for structural analysis of unordered categorical data representing proximity relationships between subjects and item categories. MAXMC is widely applicable wherever proximity items are used. Such data arise frequently in attitude surveys, personality inventories, aptitude testing, etc.

There are, of course, other methods to analyze such data ; Q3 (Hayashi, 1952 ; also known as dual scaling (Nishisato, 1980)) and correspondence analysis (Greenacre, 1984), log-linear models (e.g., Bishop, Fienberg, and Holland, 1975), latent class analysis (LCA ; Lazarsfeld and Henry, 1968), etc. Of these, Q3 is perhaps the most widely applicable method. It requires no statistical assumptions. However, it is primarily descriptive with no built-in mechanism for statistical model evaluation. Log-linear models constitute another class of general-purpose analytic methods for categorical data. They allow statistical model evaluation under modest assumptions (large sample, independence among observations). However, they lack representations of individual differences often crucial in psychological research. LCA comes closest to the proposed model. It attempts to explain statistical dependencies among observations by postulating latent variables (latent classes) over which subjects vary. However, the representation of individual differences in LCA is discrete, which often forces discretization of intrinsically continuous processes.

This may give rise to too many latent classes which are difficult to interpret. MAXMC overcomes all of these difficulties; it allows continuous representations of individual differences, and it allows statistical model evaluation.

Takane and de Leeuw (1987) discussed the relationship between the MML estimation of IRT models and factor analysis of discrete data. Under certain conditions they are mathematically equivalent. The main difference is computational. Whereas in the IRT approach discretization of continuous latent processes precedes marginalization of subject parameters, just the opposite takes place in the factor analysis approach. MAXMC is based on the IRT approach. Presumably it can also be approached from the factor analytic perspective. However, no methods have yet been developed for unordered categorical data from the factor analytic perspective (Shigemasu, 1990).

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control.*, **19**, 716-722.

Akaike, H. (1980). Likelihood and the Bayes procedure. In J.M. Bernardo, M.H. DeGroot, & A.F.M. Smith (Eds.), *Baysian statistics* (pp. 143-203). Valencia, Spain: University Press.

Bartholomew, D.J. (1987). *Latent variable models and factor analysis.* London: Charles Griffin.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete multivariate analysis: theory and practice.* Cambridge, MA: The MIT Press.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, **37**, 29-51.

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Applications of an EM algorithm. *Psychometrika*, **46**, 443-459.

Bock, R.D., Gibbons, R., & Muraki, E. (1988). Full information item factor analysis. *Applied Psychological Measurement*, **12**, 261-280.

Bock, R.D. & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, **35**, 179-197.

Coombs, C.H. (1964). *A theory of data.* New York: Wiley.

Ennis, D.M. (1988). Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General*, **117**, 408-411.

Greenacre, M.J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, **3**, 69-98.

Hayashi, C. (1982). Changes in belief systems, quality of life issues and social conditions over 25 years in post-war Japan. Paper presented at the Tenth World Congress of Sociology, Mexico City.

Hoijtink, H. (1990). *Parella: Measurement of latent traits by proximity items.* Leiden, The Netherlands: DSWO Press.

Lang, J. (1992). Obtaining the observed information matrix from the Poisson log linear model with incomplete data. *Biometrika*, **79**, 405-407.

Lazarsfeld, P.F. & Henry, N. (1968). *Latent structure analysis.* New York: Houghton Mifflin.

Leimu, H. (1983). Työntekijäin työasema ja työpaikkaliikkuvuus erikokoisissa teollisuusyrityksissä. [Bluecollar work position and workplace mobility in small and large industial firms, I. Theoretical point of departure and comparison of workers in small and large industrial firms]. University of Turku, C40, Turku.

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **44**, 226-233.
Luce, R.D. (1959). *Individual choice behavior*. New York: Wiley.
Nishisato, S. (1980). *Analysis of categorical data: dual scaling and its applications*. Toronto: University of Toronto Press.
Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
Shepard, R.N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, **22**, 325-345.
Shepard, R.N. (1986). Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General*, **115**, 58-61.
Shigemasu, K. (1990). Kategorikaru deta no inshibunseki [Factor analysis of categorical data]. *Kodokeiryogaku* [*The Japanese Journal of Behaviormetrics*], **18**, 41-51.
Takane, Y. & de Leeuw, J. (1987). The relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, **52**, 393-408.
Takane, Y. & Shibayama, T. (1986). Comparison of models for stimulus identification data. In J. de Leeuw, W. Heiser, J. Meulman, & F. Critchley (Eds.), *Multidimensional data analysis* (pp. 119-148). Leiden, The Netherlands: DSWO Press.
Takane, Y. & Shibayama, T. (1992). Structures in stimulus identification data. In F.G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 335-362). Hillsdale, NJ: Earlbaum.

(Received March 1996, Revised September 1996)

Appendix

Questions used in the ISM data (Translations as given in Hayashi, 1982).

1. If you have no children, do you think it necessary to adopt a child in order to continue the family line, even if there is no blood relationship? Or do you not think this is important?

   (1) Would adopt
   (2) Would not adopt
   (3) Depends on circumstances

2. In bringing up children of primary school age, some people think that one should teach them that money is the most important thing. Do you agree with this or not?

   (1) Agree
   (2) Disagree
   (3) Undecided

3. If you think a thing is right, do you think you should go ahead and do it even if it is contrary to usual custom, or do you think you are less apt to make a mistake if you follow custom?

(1) Go ahead
(2) Follow custom
(3) Depends on circumstances

4. Some people say that if we get good political leaders, the best way to improve the country is for the people to leave everything to them, rather than for the people to discuss things among themselves. Do you agree with this, or disagree?

(1) Agree
(2) Disagree
(3) Depends on circumstances

5. Here are three opinions about man and nature. Which one of these do you think is closest to the truth?

(1) In order to be happy, man must follow nature.
(2) In order to be happy, man must make use of nature.
(3) In order to be happy, man must conquer nature.

6. Which one of the following opinions do you agree with?

(1) If individuals are made happy, then and only then will Japan as a whole improve.
(2) If Japan as a whole improves, then and only then can individuals be made happy.
(3) Improving Japan and making individuals happy are the same thing.