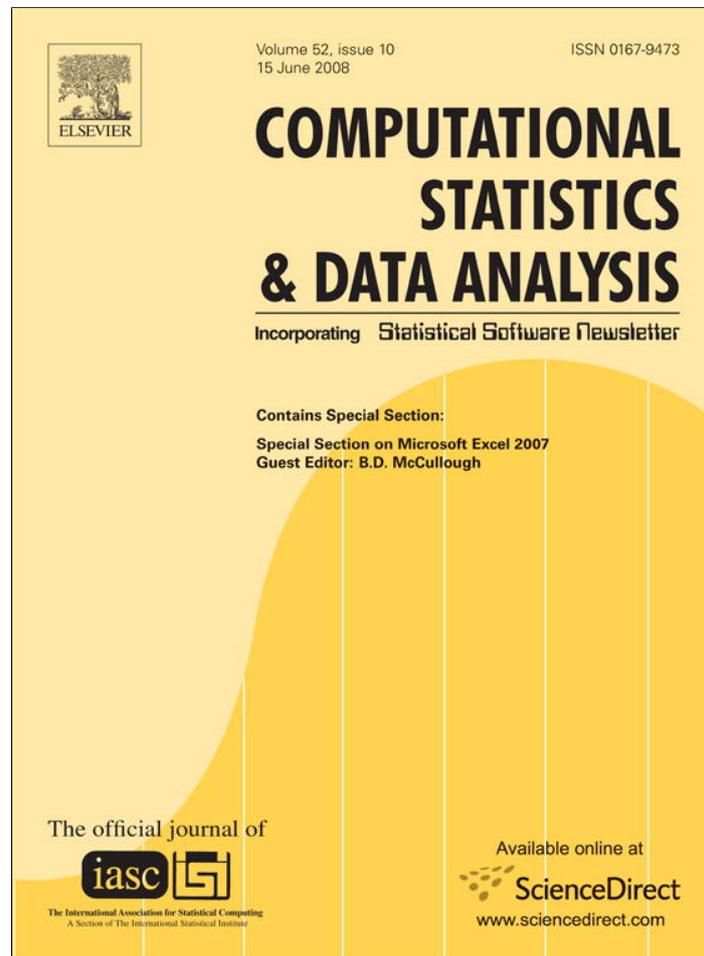


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A maximum likelihood method for an asymmetric MDS model

S. Saburi, N. Chino*

Faculty of Psychological & Physical Science, Aichi Gakuin University, 12 Araiike, Iwasaki-cho, Nisshin, Aichi, 470-0195, Japan

ARTICLE INFO

Article history:

Received 5 July 2006

Received in revised form 27 October 2007

Accepted 9 March 2008

Available online 13 March 2008

ABSTRACT

A maximum likelihood estimation method is proposed to fit an asymmetric multidimensional scaling model to a set of asymmetric data. This method is based on successive categories scaling, and enables us to analyze asymmetric proximity data measured, at least, at an ordinal scale level. It enables us to examine not only the appropriate scaling level of the data, but also the appropriate dimensionality of the model, using AIC. Prior to or in fitting the asymmetric MDS model, it is important to verify that the data are sufficiently asymmetric. Some variants of symmetry hypotheses are developed for this purpose. Since the emphasis in our paper is not on hypothesis testing, but on model diagnosis, we compare several candidate models including models with these hypotheses based on a similar model comparison idea using AIC. The method is applied to artificial data and a set of friendship data among nations in East Asia and the USA. Relations to other methods are also discussed.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Asymmetric relationships are frequently observed in animal as well as human behavior. For example, pecking order is a special asymmetric relationship among a group of hens and cocks (e.g., Rushen (1982)). Asymmetric relationships may become more complicated in humans than in animals. A typical example may be asymmetric sentiment relationships such as one-sided love and hate among members of informal groups in classrooms.

The asymmetric multidimensional scaling is a method which is specifically designed to analyze such asymmetric relationships among members and display them graphically by plotting each member in a certain dimensional space, given an asymmetric square data matrix \mathbf{S} of order n whose (i, j) th element, s_{ij} , denotes the proximity of member i to member j . If we are interested in the change in these relations over time, then we have to obtain longitudinal, asymmetric square data matrices, say $\mathbf{S}_1, \dots, \mathbf{S}_T$. In some cases s_{ij} may be dichotomous as in the pecking order data, but in other cases these may be measured at an ordinal, interval, or ratio scale level.

The asymmetric multidimensional scaling which has been developed in psychometrics since the work of Young (1975) is an extension of the method called the multidimensional scaling for symmetric data, abbreviated traditionally as MDS. Hereafter we shall abbreviate asymmetric multidimensional scaling as asymmetric MDS in order to distinguish it from the traditional symmetric MDS.

The major asymmetric MDS methods include Chino (1978, 1990), Chino and Shiraiwa (1993), Constantine and Gower (1978), Escoufier and Grouud (1980), Gower (1977), Harshman (1978), Harshman et al. (1982), Kiers and Takane (1994), Krumhansl (1978), Okada and Imaizumi (1987, 1997), Rocci and Bove (2002), Saito (1991), Saito and Takeda (1990), Sato (1988), ten Berge (1997), Tobler (1976), Trendafilov (2002), Weeks and Bentler (1982), Young (1975), and Zielman and Heiser (1996).

* Corresponding author. Tel.: +81 0561 73 1111; fax: +81 0561 73 1142.

E-mail addresses: saburi@dpc.agu.ac.jp (S. Saburi), chino@dpc.agu.ac.jp (N. Chino).

Although various asymmetric MDS methods have been proposed, these methods have remained to be descriptive. By contrast, we will propose a maximum likelihood method for asymmetric MDS, which extends MAXSCAL Takane (1981) to asymmetric relational data. It enables us, for example, to examine not only the appropriate scaling level of the data but also the appropriate dimensionality of the configuration of members using AIC (Akaike, 1973). It also enables us to perform some tests of symmetry for a given asymmetric data set. These tests enable us to verify that the data are sufficiently asymmetric.

We shall introduce two types of tests for symmetry. One type is a symmetry test to be performed prior to the application of our method to the data. It is a special test for conditional symmetry, and can be used to examine *indirectly* the validity of applying asymmetric MDS to the data. The other type of symmetry test pertains to a symmetry hypothesis to be tested in the scaling step. We shall introduce two kinds of tests of this type. These tests can be used to examine *directly* the validity of applying asymmetric MDS.

Traditionally, the symmetry hypothesis and its variants are tested using χ^2 statistics (for example, Birch (1963), Bishop et al. (1975), Bowker (1948), Caussinus (1965), Cramér (1946), Goodman (1985) and McNemar (1947)). However, these statistics have several disadvantages. For example, their use for model comparison is limited to the case in which models to be compared constitute a hierarchical structure. Ramsay (1980) shows that the asymptotic χ^2 statistic tends to be too liberal in a small sample situation. De Rooij and Heiser (2005) point out that the asymptotic χ^2 statistics tend to dismiss all models except for the saturated model when the sample size is large.

Information criteria such as AIC, on the other hand, can be used to compare various models with nonhierarchical structure. Moreover, since AIC rank-orders all competing models from the best to the worst, there is no need to worry about controlling the errors in comparing these models. In the traditional goodness-of-fit tests, on the other hand, a series of tests are typically performed to identify the best model, each comparing two nested models at a time. This renders controlling the two kinds of errors (Type 1 and Type 2) extremely difficult, if not impossible. Therefore, we shall utilize information criteria rather than the asymptotic χ^2 statistic to compare the scaling levels of the data, dimensionalities of the model, and several variants of symmetry hypotheses introduced in this paper.

Basic models, the maximum likelihood criterion, and algorithms are essentially based on MAXSCAL by Takane (1981), except that one of the constituent sub-models, that is, the representation model is replaced by that of Okada and Imaizumi (1987). For this reason, we shall call our method ASYMMAXSCAL-OI which means that it is a MAXSCAL applied to the Okada–Imaizumi model (abbreviated hereafter as the OI model) for asymmetric MDS.

The organization of this paper is as follows. In Section 2 we shall describe our method. In Section 3.1 we shall introduce a special test for symmetry to be performed prior to the scaling step, and in Section 3.2 two other tests to be performed in the scaling step. We shall show the asymptotic χ^2 statistics for these tests in these sections in order to clarify the distinction between the traditional symmetry hypotheses and ours. In Section 4 we shall discuss a model selection method. In Section 5 we shall show examples of application of our method to synthetic and actual data sets. Finally, in Section 6 we shall give some concluding remarks.

2. Method

2.1. Sub-models

In ASYMMAXSCAL as well as MAXSCAL, subjects' responses are not the usual proximities, s_{ij} 's, defined in the previous section, but rating scale categories in which subjects place certain error-perturbed proximities. For example, suppose that each subject is asked to judge the degree of friendliness of one of the governments of five nations, say, Japan to another of them, say, the USA on a 5-point rating scale with categories, "very friendly", "friendly", "neither friendly nor hostile", "hostile", "very hostile", as will be described in Section 5.2. Suppose further that a subject judges that the government of Japan is hostile to that of the USA. Then, his or her response is the fourth category, which is not the usual proximity in the context of asymmetric MDS.

Let us introduce first the population value corresponding to s_{ij} which we call the proximity model hereafter. In ASYMMAXSCAL, we assume that the proximity model is error-perturbed to give rise to a psychological value, which we call the error-perturbed proximity. We ask subjects to place it in one of the M rating scale categories, which are nothing but subjects' responses in ASYMMAXSCAL.

Subjects' responses are assumed to be determined by the following three sub-models: the representation model, the error model, and the response model, according to MAXSCAL. The representation model specifies the model of proximities (e.g., the Euclidean distance model). The error model specifies the distribution of the error-perturbed proximities (it describes how the value of the representation model is error-perturbed to generate error-perturbed proximities). In this paper we assume the normal distribution. The response model specifies the mechanism by which responses of specific form are generated. It assumes that proximities are placed in a particular category whenever error-perturbed proximities fall between the lower and the upper bounds of the category. This is similar to the law of successive categories (Torgerson, 1958).

2.1.1. The representation model

The representation model specifies the representation of the proximity. Any extant model for asymmetric MDS may be chosen as the representation model in principle. In this paper, however, we will restrict our attention to the OI model (Okada

and Imaizumi, 1987). This model represents the symmetric part and the skew-symmetric part of the square asymmetric matrix by the Euclidean distance and the difference in the radii of circles (spheres, hyperspheres) centered at the points of the objects in that space, respectively. In the OI model, the proximity model of object i to object j , g_{ij} , is defined as

$$g_{ij} = d_{ij} - r_i + r_j, \tag{1}$$

where r_i is the radius of the circle attached to object i and

$$d_{ij} = \left[\sum_{a=1}^A (x_{ia} - x_{ja})^2 \right]^{1/2},$$

where x_{ia} is the coordinate of object i on the a th dimension, and A is the dimensionality of the space. According to their model, an object with larger radius is more proximate to an object with smaller radius than vice versa. Okada and Imaizumi (1987) introduced these parameters in order to represent the skew-symmetry contained in g_{ij} , but did not consider what kinds of psychological processes underlie these parameters. However, Okada and Imaizumi (1997) interpret the term, $r_j - r_i$, as a dominance effect as discussed in Zielman and Heiser (1996).

A distinguished feature of the OI model pertains to its simple form. It is a special case of the very general model by Holman (1979)

$$p_{ij} = F \left[b_i^{(r)} + b_j^{(c)} + z_{ij} \right],$$

where p_{ij} is a real-valued function of several parameters. Here, F is an increasing function, while z_{ij} is a symmetric similarity function, and $b_i^{(r)}$ and $b_j^{(c)}$ are bias functions on the individual objects. In the OI model the row bias $b_i^{(r)}$ is equivalent to the negative of the column bias $b_i^{(c)}$ in the Holman model. However, one may consider such a model too simplistic because this model has no bias components for the symmetric part of g_{ij} (Zielman and Heiser, 1996).

The Holman model includes various asymmetric proximity models such as the quasi-symmetry models (Bishop et al., 1975; Caussinus, 1965) and the similarity choice model (Luca, 1959, 1963). These models have scored a long-standing success compared with alternative models of identification confusion (e.g., Ashby and Perrin (1988), Smith (1980), Takane and Shibayama (1986) and Townsend and Landon (1983)) as Nosofsky (1991) pointed out. The OI model has an advantage of visualizing the asymmetries contained in the estimated configuration of objects easily over various variants of the quasi-symmetric model as discussed in De Rooij and Heiser (2003).

We also introduce a “saturated” representation model abbreviated as the SR model. This model takes g_{ij} ($i, j = 1, \dots, n$) as parameters. It is used to test one of the symmetry hypotheses to be mentioned below. The estimates of g_{ij} 's are utilized to compute the initial values of the parameters for the representation model of interest.

2.1.2. The error model

As already discussed, the g_{ij} 's are assumed to be error-perturbed by some process to yield error-perturbed proximities. Takane (1981) assumes the additive error model with the normal distribution, and the multiplicative error model with the log-normal distribution. The latter implies that g_{ij} 's are nonnegative, so that it contradicts with some representation models in which g_{ij} 's can be negative as with the OI model. Thus, we assume the following additive error model only:

$$\begin{cases} \tau_{ij}^{(k)} = g_{ij} + e_{ij}^{(k)} \\ e_{ij}^{(k)} \sim N(0, \sigma^2), \end{cases} \tag{2}$$

where k indicates a subject. MAXSCAL allows for individual differences, and thus σ has subscript k . However, σ_k 's are incidental parameters, and it is often unnatural for subjects to judge the proximities among the same set of objects many times in a short period. Therefore, we assume that subjects make one and only one proximity judgment per data set. Our additive error model corresponds to the case in which subjects are viewed as replications in MAXSCAL. The index k is suppressed hereafter.

2.1.3. The response model

Let us suppose that subjects place error-perturbed proximities in one of the M rating scale categories C_1, \dots, C_M . These categories are assumed to be represented by a set of ordered intervals with upper and lower boundaries on a psychological continuum. If we denote the upper boundary of the m th category by b_m , then

$$-\infty = b_0 \leq b_1 \leq \dots \leq b_m \leq \dots \leq b_{M-1} \leq b_M = \infty.$$

The probability that the error-perturbed proximity of object i to object j , τ_{ij} , falls in C_m is given by the

$$p_{ijm} = \text{pr} (b_{m-1} < \tau_{ij} \leq b_m). \tag{3}$$

Under the additive error model of (2) and (3) can be written as

$$p_{ijm} = \int_{b_{m-1}}^{b_m} \phi(\tau_{ij}) d\tau_{ij}, \tag{4}$$

where ϕ is the density function of the normal distribution with mean g_{ij} and variance σ^2 . By replacing $(\tau_{ij} - g_{ij})/\sigma$ with z we can obtain

$$p_{ijm} = \int_{a_{ij(m-1)}}^{a_{ijm}} f(z) dz, \tag{5}$$

where f is the density function of the standard normal distribution, and

$$\begin{cases} a_{ij(m-1)} = \frac{b_{m-1} - g_{ij}}{\sigma} \\ a_{ijm} = \frac{b_m - g_{ij}}{\sigma}. \end{cases} \tag{6}$$

If we let

$$F_{ijm} = \int_{-\infty}^{a_{ijm}} f(z) dz, \tag{7}$$

then

$$p_{ijm} = F_{ijm} - F_{ij(m-1)}. \tag{8}$$

Here, b_m 's ($m = 1, \dots, M - 1$) are not necessarily equally spaced. They are unrestricted except for the implicit ordinal restriction, that is, $b_1 \leq b_2 \leq \dots \leq b_{M-1}$. These categories constitute an ordinal scale. On the other hand, if the b_m 's are assumed to be equally spaced, then we may impose the linear constraint:

$$b_m = \alpha m + \beta, \quad (\alpha > 0; 1 \leq m \leq M - 1), \tag{9}$$

where α and β denote a scale factor and an additive constant, respectively. These categories constitute an interval scale.

2.2. Maximum likelihood criterion

Let us now introduce an indicator variable Z_{ijmk} which takes the value of one when subject k rates the proximity of object i to object j into C_m , and zero otherwise. That is,

$$Z_{ijmk} = \begin{cases} 1 & \text{when } o_{ijk} \in C_m \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

where $o_{ijk} \in C_m$ means that the proximity of object i to object j for subject k (o_{ijk}) is placed into C_m . Readers should distinguish the notation of the response o_{ijk} from the variable $\tau_{ij}^{(k)}$. The judgments for all combinations of objects are assumed to be mutually independent, so that the joint likelihood of the total observations is written as

$$L = \prod_i^n \prod_j^n \prod_m^M p_{ijm}^{Y_{ijm}}, \tag{11}$$

where $Y_{ijm} = \sum_{k=1}^n Z_{ijmk}$. We will maximize the logarithm of L in (11) over all the parameters in those three sub-models.

It is well known that all judgments are likely to be independent under the single-judgment condition, where each subject judges only one pair of objects, with randomly sampled subjects. However, under the multiple-judgment condition, where each subject judges all pairs of objects, the judgments may or may not be independent (Bock and Jones, 1968, p. 16).

2.3. Algorithm

We use Fisher's scoring algorithm as in the original MAXSCAL, to maximize the log-likelihood. We update a vector of unknown parameters $\theta = (\theta_1, \dots, \theta_{N_p})'$ by the following formula, where N_p being the total number of unknown parameters:

$$\theta^{(q+1)} = \theta^{(q)} + \varepsilon^{(q)} \mathbf{I}(\theta^{(q)})^{-1} \mathbf{u}(\theta^{(q)}), \tag{12}$$

where q is the index of iteration number, ε is a step-size parameter, and the vector $\mathbf{u}(\theta)$ and the matrix $\mathbf{I}(\theta)$ are defined by

$$\mathbf{u}(\theta) = \left(\frac{\partial \ln L}{\partial \theta} \right),$$

and

$$\mathbf{I}(\boldsymbol{\theta}) = E \left[\left(\frac{\partial \ln L}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ln L}{\partial \boldsymbol{\theta}} \right)' \right] = -E \left(\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right).$$

The information matrix $\mathbf{I}(\boldsymbol{\theta})$ is singular due to the nonuniqueness of scale and location of g_{ij} 's for the SR model. Specifically, the transformation $\tilde{g}_{ij} = \psi g_{ij} + \mu$ by arbitrary constants ψ and μ in (6) can be compensated for by the transformations on category boundaries and σ , i.e., $\tilde{b}_m = \psi b_m + \mu$ and $\tilde{\sigma} = \psi \sigma$. The solution of the OI model is also not unique with respect to the size, the origin and the rotation of the stimulus configuration, and with respect to additive constant in the radii. Thus, we replace the regular inverse in (12) with the Moore–Penrose inverse.

The next task is to evaluate $\mathbf{u}(\boldsymbol{\theta})$ and $\mathbf{I}(\boldsymbol{\theta})$ in formula (12). It is easy to see from (11) that

$$\mathbf{u}(\boldsymbol{\theta}) = \sum_i^n \sum_j^n \sum_m^M \frac{Y_{ijm}}{p_{ijm}} \left(\frac{\partial p_{ijm}}{\partial \boldsymbol{\theta}} \right),$$

and

$$\mathbf{I}(\boldsymbol{\theta}) = \sum_i^n \sum_j^n \sum_m^M \frac{n_{ij}}{p_{ijm}} \left(\frac{\partial p_{ijm}}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial p_{ijm}}{\partial \boldsymbol{\theta}} \right)',$$

where $n_{ij} = \sum_{m=1}^M Y_{ijm}$. Here, we have from (8)

$$\frac{\partial p_{ijm}}{\partial \boldsymbol{\theta}} = \frac{\partial F_{ijm}}{\partial \boldsymbol{\theta}} - \frac{\partial F_{ij(m-1)}}{\partial \boldsymbol{\theta}}, \tag{13}$$

and

$$\frac{\partial F_{ijm}}{\partial \boldsymbol{\theta}} = \frac{\partial F_{ijm}}{\partial a_{ijm}} \frac{\partial a_{ijm}}{\partial \boldsymbol{\theta}}. \tag{14}$$

We approximate the distribution function F_{ijm} of the standard normal distribution by that of the logistic distribution given by $F_{ijm} = 1 / (1 + e^{-c a_{ijm}})$, where c is a specific constant, since the density function of the normal distribution cannot be integrated in closed form, while that of the logistic function can, and consequently is easier to handle. We use $c = 1/0.569$ that is shown by Jeffres (1973) to give a better approximation than $c = \pi/3^{1/2}$ used by Takane (1981). Then the first factor on the right-hand side of (14) is given by

$$\frac{\partial F_{ijm}}{\partial a_{ijm}} = -F_{ijm}(F_{ijm} - 1)c. \tag{15}$$

Finally, we need to evaluate $\partial a_{ijm} / \partial \theta_l$ for $l = 1, \dots, N_p$.

In updating $\boldsymbol{\theta}$ in (12), it is important to choose appropriate initial values. We may use estimates of the SR model for this purpose. We may apply an appropriate method to $\mathbf{G}^* = \{g_{ij}^*\}$, where g_{ij}^* is the estimate of g_{ij} . The solution can be viewed as a set of reasonable initial values of these parameters. The b_m 's (or α and β) and σ estimated with the SR model are also viewed as their own reasonable initial values in the OI model estimation. Therefore, it might be useful to conduct the two-step procedure, that is, to deal with the SR model estimation first, and then to use the result in the subsequent OI model estimation.

3. Tests for symmetry

In this section we shall introduce a special test for conditional symmetry and two variants of symmetry tests. All of them are introduced to verify that the data are sufficiently asymmetric to apply any asymmetric MDS.

We show the asymptotic χ^2 statistics for these tests in order to clarify the distinction between the traditional symmetry hypotheses and ours. However, in the actual application we recommend to use AIC as better alternatives, for reasons which we have already pointed out in Section 1. In this case we may compare values of the criteria under these hypotheses simultaneously.

3.1. A symmetry test prior to fitting the asymmetric model

In this subsection we propose a test for conditional symmetry. It enables us to diagnose the validity of applying any asymmetric model to the set of asymmetric data under study. Since we use the method of successive categories, subjects' proximity judgments from objects to objects can be compiled into a special three-way contingency table, that is, the $n \times n \times M$ table.

Fig. 1 shows two different arrangements of the original data, which correspond to two formally different sampling designs. Type A design assumes that each row $(Y_{ij1}, Y_{ij2}, \dots, Y_{ijM})$ follows a multinomial distribution with parameter n_{ij} which

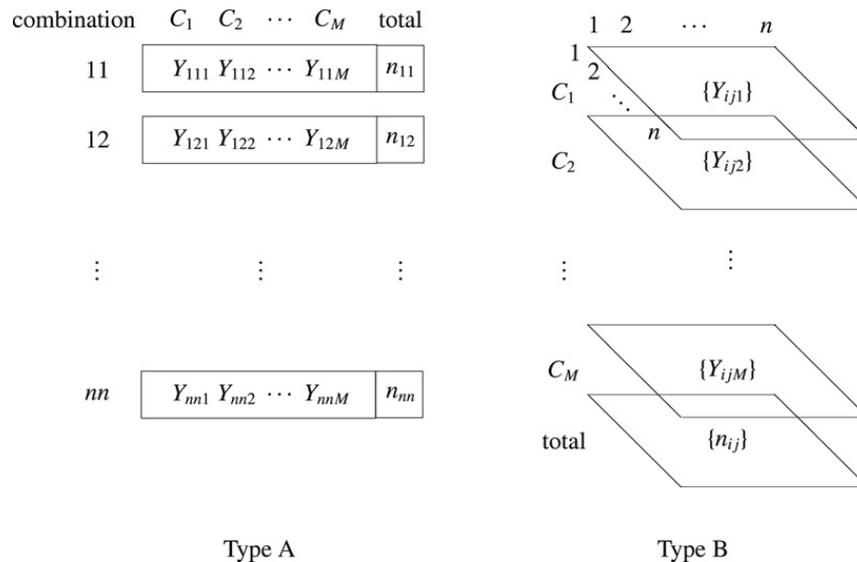


Fig. 1. Two types of sampling designs for the method of successive categories in ASYMMAXSCAL.

is fixed and certain cell probabilities $p_{ij1}, p_{ij2}, \dots, p_{ijM}$. This design is the sampling design used in our ASYMMAXSCAL as well as MAXSCAL by Takane (1981).

In Type A design it is not difficult to show that under the “special” conditional symmetry hypothesis,

$$H_0^{(cs)} : p_{ijm} - p_{jim} = 0, \quad (1 \leq i < j \leq n; 1 \leq m \leq M - 1), \quad (16)$$

the likelihood ratio test statistic is written as

$$G^2 = 2 \sum_m \sum_i \sum_j Y_{ijm} \left\{ \ln Y_{ijm} - \ln \left[\frac{n_{ij}(Y_{ijm} + Y_{jim})}{n_{ij} + n_{ji}} \right] \right\}, \quad (17)$$

which asymptotically follows the central χ^2 -distribution with $\nu_{cs} = (M-1)n(n-1)/2$ degrees of freedom under $H_0^{(cs)}$. Takane suggested this statistic (personal communication, August 26, 2005).

If the conditional symmetry hypothesis stated above is accepted, we should abandon the application of any of the extant asymmetric MDS's including ASYMMAXSCAL-OI. If it is rejected, one can proceed to the application of ASYMMAXSCAL-OI to the data.

By contrast, the Type B design data which is obtained by rearranging the Type A design data looks like the data extracted from one of the traditional designs for the special three-way contingency table, that is, the $n \times n \times M$ table (Agresti, 2002; Bishop et al., 1975; Read, 1978), but is different from any of these traditional designs. By the traditional designs we mean (1) the multiplicative Poisson model with no fixed marginals, (2) the multinomial model with the fixed total, and (3) the product (or multiplicative) multinomial model with the layer totals fixed. We call them the Type I design, the Type II design, and the Type III design for the $n \times n \times M$ table, respectively.

In order to clarify the distinction between the Type B design and the three traditional designs, let us look at the Type I design. The Type B design follows from the Type I design by conditioning on $Y_{11\bullet} = n_{11}, \dots, Y_{nn\bullet} = n_{nn}$, where the symbol ‘•’ indicates the summation over all categories. It is also the case that the joint likelihood of the data, Y_{111}, \dots, Y_{nnM} of the Type B design can be converted into that of the Type A design by putting $p_{ijm} = \lambda_{ijm}/\lambda_{ij\bullet}$, where $\lambda_{ijm} = E(Y_{ijm})$ is a Poisson parameter in the Type I design and p_{ijm} is the multinomial parameter in the Type A design. Thus, the conditional symmetry hypothesis, $H_0^{(cs)} : p_{ijm} - p_{jim} = 0$, in the Type A design is equivalent to the hypothesis, $H_0^B : \lambda_{ijm}/\lambda_{ij\bullet} - \lambda_{jim}/\lambda_{ji\bullet} = 0$, in the Type B design.

Furthermore, (17) is equivalent in form to the likelihood ratio χ^2 statistic for the traditional conditional symmetry hypothesis if and only if $n_{ij} = n_{ji}$ for all combinations of i and j . The degrees of freedom are $n(n-1)/2$ less than those for the traditional hypotheses.

The conditional symmetry test just introduced assumes none of the models for asymmetric scaling. ASYMMAXSCAL enables us to test two different symmetry hypotheses, both of which assume the error model and the response model. We shall discuss them in the next subsection.

3.2. Two symmetry tests in the scaling step

One symmetry test in the scaling step is the test for the symmetry hypothesis based on the SR model. This hypothesis is defined by

$$H_0^{(s/sr)} : g_{ij} - g_{ji} = 0, \quad (1 \leq i < j \leq n). \quad (18)$$

Table 1
The candidates for the desirable model, the corresponding hypotheses, and ν_p

No	Candidate model	Hypothesis	ν_p
1	The OI model	$g_{ij} = d_{ij} - r_i + r_j$	$n_c + nA - A(A + 1)/2 + n - 1$
2	The OI model under $H_0^{(s/oi)}$	$g_{ij} = d_{ij} - r_i + r_j$ $H_0^{(s/oi)}: r_i - r_{i+1} = 0$	$n_c + nA - A(A + 1)/2$
3	The SR model	g_{ij} with no structure	$n_c - 1 + n^2,$ (Full) $n_c - 1 + n(n - 1)$ (Truncated)
4	The SR model under $H_0^{(s/sr)}$	g_{ij} with no structure under $H_0^{(s/sr)}: g_{ij} - g_{ji} = 0$	$n_c - 1 + n(n + 1)/2,$ (Full) $n_c - 1 + n(n - 1)/2$ (Truncated)
5	The saturated model	Multinomial with fixed marginals	$(M - 1) n^2,$ (Full) $(M - 1) n(n - 1)$ (Truncated)
6	The saturated model under $H_0^{(cs)}$	Multinomial with fixed marginals under $H_0^{(cs)}: p_{jim} - p_{jim} = 0$	$(M - 1)n(n + 1)/2,$ (Full) $(M - 1)n(n - 1)/2$ (Truncated)

Note: The “truncated” means the case in which diagonals are not observed. The n_c is the number of parameters needed to specify category boundaries, i.e., $M - 1$ under the ordinal scale assumption, and 2 under the interval scale assumption.

We test $H_0^{(s/sr)}$ using θ^* , the estimate of θ . The Wald statistic (Wald, 1943; Aitchison and Silvey, 1960) for $H_0^{(s/sr)}$ is written as

$$Z_{s/sr} = \mathbf{h}(\theta^*)' [\mathbf{H}(\theta^*)' \mathbf{I}(\theta^*)^{-1} \mathbf{H}(\theta^*)]^{-1} \mathbf{h}(\theta^*), \tag{19}$$

which asymptotically follows the central χ^2 -distribution with $\nu_{s/sr} = n(n - 1)/2$ degrees of freedom under $H_0^{(s/sr)}$. Here, $\mathbf{h}(\theta) = (h_1, \dots, h_{\nu_{s/sr}})'$ is the vector of $g_{ij} - g_{ji}$, and $\mathbf{H}(\theta)$ is the $(N_p - 2) \times \nu_{s/sr}$ Jacobian matrix whose (i, j) th element is $\partial h_j / \partial \theta_i$. We remove σ and one of the b_m 's or β from θ^* in order to make $\mathbf{I}(\theta^*)$ nonsingular in computing (19).

The other symmetry test in the scaling step is the test for the symmetry hypothesis based on the OI representation model. This hypothesis is represented by

$$H_0^{(s/oi)}: r_i - r_{i+1} = 0, \quad (1 \leq i \leq n - 1). \tag{20}$$

If we use the Wald statistic as in (19), $Z_{s/oi}$ may be affected due to the nonuniqueness of the OI solution, and may become more complicated than that of the SR solution. An alternative method is to use the likelihood ratio statistic, $-2 \ln [L^{(s/oi)} / L^{(oi)}]$, which asymptotically follows the central χ^2 -distribution with $\nu_{s/oi} = n - 1$ degrees of freedom under $H_0^{(s/oi)}$. Here, $L^{(oi)}$ is the likelihood of the OI model, and $L^{(s/oi)}$ is that of the OI model under $H_0^{(s/oi)}$. The latter representation model is equivalent to the Euclidean distance model, $g_{ij} = d_{ij}$.

4. Model selection

In this section we shall discuss the model selection method in ASYMMAXSCAL-OI which uses an information criterion. Various alternative criteria have been proposed since AIC was introduced by Akaike (1973). Which of these criteria should we choose? Winsberg and Carroll (1989) chose AIC and BIC, but with symmetric MDS. In general, however, the two criteria do not necessarily lead to the same conclusion about the dimensionality. As Schwarz (1978) points out, BIC leans more than AIC towards lower-dimensional models when there are 8 or more observations, and the criteria differ markedly from each other. In fact, Winsberg and Carroll confronted with the difficulty of deciding the dimensionality of their model, and examined the interpretability of their solutions to overcome it.

According to the recent literature on such a conflict between AIC and BIC, it is clear that no model selection criterion with a deterministic penalty can simultaneously enjoy the properties of AIC and BIC (Yang, 2005, p. 938). Thus, Yang suggests that one should keep the specific objective of inference in mind when conducting model selection. Considering that AIC is minimax-rate optimal for estimating the regression function and BIC is consistent in selecting the true model (Yang, 2005, p. 937), we may choose AIC if the objective is to choose the best model in the sense of prediction, and may choose BIC if it is to choose the true model. We have chosen AIC since our objective in this paper is not necessarily to choose the true model, but to choose the best model. Takane (1981) uses AIC as the model selection method in MAXSCAL.

The AIC associated with a model is defined by

$$AIC = -2 \ln L + 2\nu_p,$$

where L is the maximum likelihood of the model and ν_p is the effective number of parameters in the model. The model which gives the minimum AIC value is considered the best fitting model.

Table 1 shows the various candidate models and the corresponding hypotheses as well as ν_p 's of the models. The second, fourth, and sixth models pertain to various symmetry hypotheses, while the other models do not. In ASYMMAXSCAL we shall compute AIC's of all of these models, and compare these values for choosing the most appropriate one among the models.

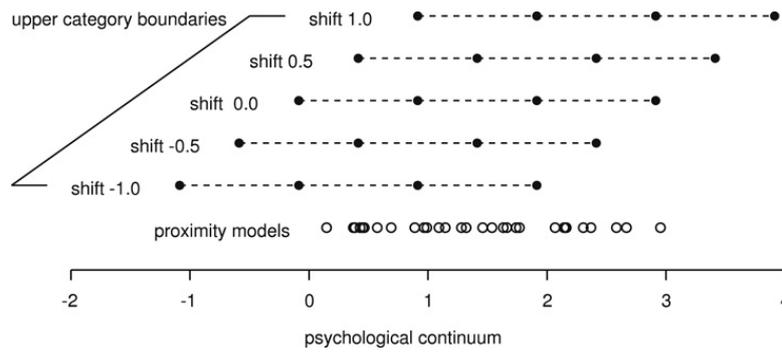


Fig. 2. Locations of the true category boundaries and the true proximity models on a psychological continuum used in the Monte Carlo study.

5. Examples

We will show the result of a small Monte Carlo study to evaluate the performance of parameter recovery. We will also show the result of an application to the friendship data among nations in East Asia and the USA.

5.1. Monte Carlo study

A small Monte Carlo study was executed in order to investigate parameter recovery with the present method for the OI model. We generated true coordinates and radii for six objects in a two-dimensional space from uniform random values, first. As in Okada and Imaizumi (1987), the dispersion of constructed g_{ij} 's which are defined by (1) is written as

$$\sum_{i \neq j} \sum_{i \neq j} (g_{ij} - \bar{g})^2 = \sum_{i \neq j} \sum_{i \neq j} (d_{ij} - \bar{d})^2 + 2n^2 \text{var}(r_i), \tag{21}$$

where \bar{g} and \bar{d} are the means of g_{ij} 's and d_{ij} 's, respectively. The first and the second terms on the right-hand side of (21) pertain to the symmetric and skew-symmetric components of g_{ij} 's, respectively. Although the ratio of the second term to the first term might affect the parameter recovery, we adjusted the true radii so that it takes 0.5, considering Okada and Imaizumi's results.

Fig. 2 illustrates the locations of the true category boundaries and the true proximity models on the psychological continuum used in the Monte Carlo study. The true category boundaries in the "shift zero" case were determined so that their mean was equivalent to \bar{g} and they were equally spaced with the unit interval in both of ordinal scale trials and interval scale trials. In other cases, the true category boundaries were shifted $-1.0, -0.5, 0.5,$ and 1.0 from the shift zero case.

We generated normal random values with mean zero and variance one (the true σ^2), and τ_{ij} 's according to (2). The number of samples per single proximity judgment was varied 10, 30, and 50. For each combination, thirty replications were run. We have assumed that the true dimensionality and the true scale level were known a priori.

We computed the product moment correlation coefficients between the recovered parameters and the true ones as indices of goodness of recovery. The coefficients between the g_{ij} 's constructed from the recovered parameters and the true ones were also computed. The coefficients for category boundaries were not computed in the interval scale trials, because they take the value one for any estimate. Schönemann and Carroll's (1970) algorithm for a Procrustes problem was applied to the recovered coordinates in order to fit the true ones before computing the coefficients. The indeterminacies of the other parameters are concerned with their scales and/or locations, which do not affect the coefficients. The mean indices were obtained by converting these coefficients to z values, obtaining mean z , and converting back to the product moment correlation coefficient, as in Okada and Imaizumi (1987).

Table 2 indicates these mean indices for each combination. As regard the category boundaries in the ordinal scale trials, almost complete recoveries were attained. It might be due to their two-step estimation, that is, b_m 's (or α and β) and σ were estimated with the SR model, first, and then their estimates were used as their initial values in the subsequent OI model estimation. The indices of the other parameters increased as the sample size increased. Assigning thirty samples per single proximity judgment seems sufficient to recover the true parameters satisfactorily. However, inappropriate solutions giving some negative p_{ijm} 's were also obtained in the special cases when the true boundaries were shifted -1.0 and ten samples were assigned per single proximity judgment under both of the ordinal scale and the interval scale conditions, and in the case when the true boundaries were shifted 1.0 and ten samples were assigned per single proximity judgment under the interval scale condition. In such a situation, the τ_{ij} 's associated with the extreme values of g_{ij} 's on the psychological continuum might fall in C_1 or C_M in most cases, which may not provide enough information for estimation.

5.2. Analysis of national friendship data

We will illustrate the application of ASYMMAXSCAL-OI to the friendship data among nations in East Asia and the USA. Four hundred Japanese university students participated in the survey. The single-judgment sampling was chosen to make

Table 2
Summary of a Monte Carlo study for the OI model

	Sample size	Ordinal scale					Interval scale				
		Boundaries' shift					Boundaries' shift				
		–1.0	–0.5	0.0	0.5	1.0	–1.0	–0.5	0.0	0.5	1.0
Proximity models	10	*	0.956	0.960	0.956	*	*	0.964	0.955	0.965	0.958
	30	0.987	0.985	0.987	0.986	0.986	0.983	0.986	0.986	0.985	0.985
	50	0.992	0.992	0.993	0.992	0.991	0.992	0.992	0.992	0.992	0.991
Coordinates	10	*	0.970	0.976	0.967	*	*	0.978	0.968	0.980	0.972
	30	0.992	0.990	0.993	0.991	0.990	0.990	0.991	0.992	0.991	0.992
	50	0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.995	0.996
Radii	10	*	0.973	0.972	0.972	*	*	0.975	0.974	0.975	0.968
	30	0.991	0.989	0.988	0.992	0.991	0.982	0.991	0.990	0.992	0.989
	50	0.994	0.994	0.994	0.995	0.993	0.993	0.995	0.993	0.993	0.993
Category boundaries	10	*	0.999	0.999	0.999	*	–	–	–	–	–
	30	0.999	1.000	1.000	1.000	1.000	–	–	–	–	–
	50	1.000	1.000	1.000	1.000	1.000	–	–	–	–	–

Note: The symbol "*" indicates that inappropriate solutions were obtained.

Table 3
The results of the symmetry tests for the national friendship data

Symmetry test	Dimensionality	Ordinal scale		Interval scale	
		χ^2	d.f.	χ^2	d.f.
LR test for $H_0^{(s/oi)}$	1	34.99	4	33.55	4
	2	35.24	4	33.81	4
	3	37.10	4	34.87	4
	4	37.10	4	34.87	4
Wald test for $H_0^{(s/sr)}$	–	44.46	10	43.03	10
LR test for $H_0^{(cs)}$	–	χ^2		d.f.	
	–	77.21		40	

the independence assumption among all judgments more plausible. That is, each of them rated the extent to which he or she felt the government of nation A be friendly or hostile to that of nation B on a 5-point rating scale ranging from "very friendly" to "very hostile" for a randomly assigned combination of A and B among China, Japan, North Korea, South Korea, and the USA. Twenty students were assigned per single proximity judgment except for the so-called self-proximities.

Although the emphasis in this paper is not on the statistical tests for symmetry but on the diagnosis about candidate models including those with hypotheses about these tests, we first referred to the results of symmetry tests developed in this paper. Table 3 shows this. It is apparent that all the symmetry hypotheses are rejected at the significance level beyond the 0.01 level. These results indicate that the data are sufficiently asymmetric.

Table 4 shows the summary statistics for the national friendship data. According to AIC, the OI model, especially the three-dimensional OI model under the ordinal scale assumption, is the most realistic model.

Fig. 3 displays the estimated category boundaries with 95% asymptotic confidence intervals in the three-dimensional solution for the OI model. The categories C_2 and C_4 (labeled as "friendly" and "hostile", respectively) seem to have larger intervals than that of C_3 (labeled as "neither friendly nor hostile") under the ordinal scale assumption.

It should be noted that these confidence intervals are point-wise. Although none of the confidence intervals indicate significant departures, the joint departure (all of them combined) indicates a significant departure. That is, small piece-wise departures combine into a significant overall departure. This is confirmed by the fact that AIC indicates that the ordinal scale model is better than the interval scale model.

Fig. 4 depicted the configuration and the spheres in the three-dimensional solution for the OI model under the ordinal scale assumption. The principal-axis rotation was performed on the configuration. The sphere associated with Japan is larger than those of the other nations, showing that the friendship of the Japanese government is generally unilateral. That is, Japan is perceived as more friendly to other nations than the other way round. It should be noticed that the spheres associated with all other nations but Japan are very small. These results indicate that most of the skew-symmetric part of the data seems to relate to Japan, and the relationships among the other nations are relatively symmetric.

On the first dimension, the allies of Japan and the USA are located opposite to North Korea, while South Korea and China are located about the middle but slightly closer to these allies and North Korea, respectively. This dimension seems to show the antagonism between Japan–US allies and North Korea. The second dimension seems to represent certain active opposition between the two nations, i.e., Japan and North Korea, and the other nations. The third dimension seems to represent some antagonism of China toward South Korea.

Table 4
Summary statistics for the national friendship data

Candidate model	Dimensionality	Ordinal scale		Interval scale	
		AIC	ν_p	AIC	ν_p
The OI model	1	4.22	12	32.13	10
	2	6.21	15	34.33	13
	3	1.84	17	30.71	15
	4	3.84	18	32.71	16
The OI model under $H_0^{(s/oi)}$	1	31.21	8	57.68	6
	2	33.45	11	60.14	9
	3	30.95	13	57.58	11
	4	32.95	14	59.58	12
The SR model	–	3.83	23	32.13	21
The SR model under $H_0^{(s/sr)}$	–	30.96	13	57.58	11
		AIC		ν_p	
The saturated model	–	50.83		80	
The saturated model under $H_0^{(cs)}$	–	48.04		40	

Note: The value of one thousand is subtracted from AIC's.

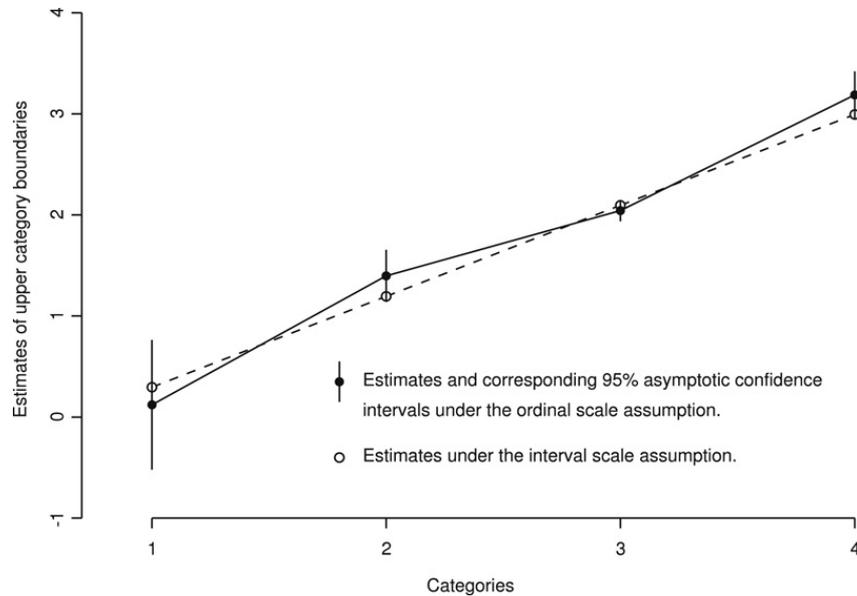


Fig. 3. Estimates of category boundaries with 95% asymptotic confidence intervals obtained from the three-dimensional solution for the OI model.

Another analysis was also done using the representation model by Saito (1991),

$$g_{ij} = d_{ij} + \gamma_i + \phi_j. \tag{22}$$

This model is more flexible than the OI model due to the biases which enables us to represent both of symmetric and skew-symmetric components of g_{ij} , but has larger number of parameters. The minimum value of AIC among models pertaining to the Saito's model was 1005.85 with the two-dimensional case under the ordinal scale assumption. It was larger than 1001.84 with the optimal model described above, i.e., the three-dimensional OI model under the ordinal scale assumption.

6. Discussion

In this paper we proposed a maximum likelihood asymmetric MDS for the OI model. Our method provides an alternative to Kruskal's (1964a, 1964b) traditional nonmetric approach taken by Okada and Imaizumi (1987, 1997), to the asymmetric ordinal relational data. Our method can potentially be more useful than their method since it provides information other than the configuration of objects, e.g., the most appropriate dimensionality of the model based on AIC, which is more sophisticated than a descriptive statistic such as stress (Kruskal, 1964b; Kruskal and Carroll, 1969). It also provides information as to the reliabilities of point locations in the form of confidence regions. It may be argued that such confidence regions can be derived in any descriptive method by the use of a bootstrap or some other resampling technique (e.g., de Leeuw and Meulman (1986)). However, developing such a procedure for MDS is not as straightforward as in other situations. Most

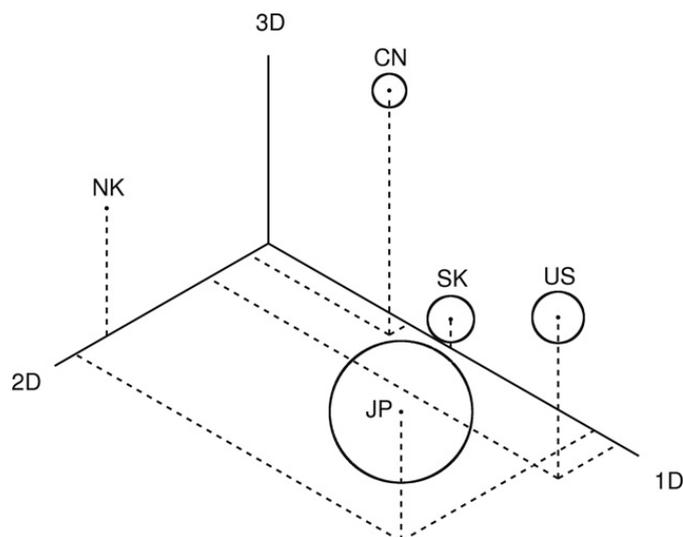


Fig. 4. Configuration and spheres obtained from the three-dimensional solution for the OI model under the ordinal scale assumption. Abbreviations indicate the following nations: CN = China, JP = Japan, NK = North Korea, SK = South Korea, and US = USA.

of the extant MDS assumes no replicated observations. Moreover, it enables us to test the validity of the assumption of asymmetry statistically using several variants of symmetry hypotheses.

The emphasis in our paper is not on hypothesis testing, but on model diagnosis using AIC. If it is of interest, however, we could also examine possible causes of asymmetry in the Type B design data, prior to fitting the asymmetric MDS model. In fact, Chino and Saburi (2006) have recently utilized various tests for symmetry and related hypotheses including the quasi-symmetry by Caussinus (1965), the marginal homogeneity by Stuart (1955), the quasi-independency by Goodman (1968), and so on, given the Type B design data. Such an approach has rarely been fully discussed in the context of asymmetric MDS.

Chino and Saburi (2006) have suggested that the conditional quasi-symmetry hypothesis plays a fundamental role in such an analysis. One of the major reasons is that the notion of quasi-symmetry is more general than that of symmetry. Another reason is that this notion is closely related to the basic feature of some of the extant asymmetric MDS models, as De Rooij and Heiser (2003) has pointed out. Although they only deal with count data, the Type A design data are not confined to count data. Therefore, Chino and Saburi's idea can be viewed as an extension of De Rooij and Heiser (2003, 2005) in this respect.

We may apply the fundamental ideas and the method proposed in this paper to any of the extant asymmetric MDS models. Some unified algorithm based on these ideas and method for extant asymmetric MDS models may be fruitful and attractive as a topic for future research.

Acknowledgements

The authors are indebted to Yoshio Takane, Yasunori Fujikoshi, and anonymous reviewers for their valuable suggestions and comments on earlier versions of this paper. The authors are also indebted to Gregory L. Rohe for proofreading an earlier version of this paper.

References

- Agresti, A., 2002. *Categorical Data Analysis*. Wiley, New York.
- Aitchison, J., Silvey, S.D., 1960. Maximum-likelihood estimation procedures and associated tests of significance. *J. Roy. Statist. Soc. Ser. B* 22, 154–171.
- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), *The Second International Symposium on Information Theory*. Akadémiai Kiadó, Budapest, pp. 267–281.
- Ashby, F.G., Perrin, N.A., 1988. Toward a unified theory of similarity and recognition. *Psychol. Rev.* 95, 124–150.
- Birch, M.W., 1963. Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. Ser. B* 25, 220–233.
- Bishop, Y.M.M., Fienberg, S.E., Holland, P.W., 1975. *Discrete Multivariate Analysis—Theory and Practice*. The MIT press, Massachusetts.
- Bock, R.D., Jones, L.V., 1968. *The Measurement and Prediction of Judgment and Choice*. Holden-Day, San Francisco.
- Bowker, A.H., 1948. A test for symmetry in contingency tables. *J. Amer. Statist. Assoc.* 43, 572–574.
- Caussinus, H., 1965. Contribution à l'analyse statistique des tableaux de corrélation [Contributions to the statistical analysis of correlation matrices]. *Annales de la Faculté des Sciences de l'Université de Toulouse* 29, 77–182.
- Chino, N., 1978. A graphical technique for representing the asymmetric relationships between N objects. *Behaviormetrika* 5, 23–40.
- Chino, N., 1990. A generalized inner product model for the analysis of asymmetry. *Behaviormetrika* 27, 25–46.
- Chino, N., Saburi, S., 2006. Tests of symmetry in asymmetric MDS. Paper Presented at the 2nd German Japanese Symposium on Classification. Berlin, March.
- Chino, N., Shiraiwa, K., 1993. Geometrical structures of some non-distance models for asymmetric MDS. *Behaviormetrika* 20, 35–47.
- Constantine, A.G., Gower, J.C., 1978. Graphical representation of asymmetric matrices. *Appl. Statist.* 27, 297–304.
- Cramér, H., 1946. *Mathematical Methods of Statistics*. Princeton University Press, Princeton.
- de Leeuw, J., Meulman, J., 1986. A special jackknife for multidimensional scaling. *J. Classif.* 3, 97–112.

- De Rooij, M., Heiser, W.J., 2003. A distance representation of the quasi-symmetry model and related distance models. In: Yanai, H., Okada, A., Shigemasa, K., Kano, Y., Meulman, J.J. (Eds.), *New Developments in Psychometrics*. Springer, Tokyo, pp. 487–494.
- De Rooij, M., Heiser, W.J., 2005. Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data. *Psychometrika* 70, 99–122.
- Escoufier, Y., Grorud, A., 1980. Analyse factorielle des matrices carrees non symetriques [Factor analysis of square asymmetric matrices]. In: Diday, E., et al. (Eds.), *Data Analysis and Informatics*. North-Holland, Amsterdam, pp. 263–276.
- Goodman, L.A., 1968. The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries. *J. Amer. Statist. Assoc.* 63, 1091–1131.
- Goodman, L.A., 1985. The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.* 13, 10–69.
- Gower, J.C., 1977. The analysis of asymmetry and orthogonality. In: Barra, J.R., Brodeau, F., Romer, G., van Cutsem, B. (Eds.), *Recent Developments in Statistics*. North-Holland, Amsterdam, pp. 109–123.
- Harshman, R.A., 1978. Models for analysis of asymmetrical relationships among N objects or stimuli. Paper presented at the First Joint Meeting of the Psychometric Society and the Society for Mathematical Psychology. Hamilton, Canada.
- Harshman, R.A., Green, P.E., Wind, Y., Lundy, M.E., 1982. A model for the analysis of asymmetric data in marketing research. *Market. Sci.* 1, 205–242.
- Holman, E.W., 1979. Monotonic models for asymmetric proximities. *J. Math. Psychol.* 20, 1–15.
- Jeffres, L.A., 1973. The logistic distribution as an approximation to the normal curve. *J. Acoust. Soc. Amer.* 53, 1296.
- Kiers, H.A.L., Takane, Y., 1994. A generalization of GIPSCAL for the analysis of nonsymmetric data. *J. Classif.* 11, 79–99.
- Krumhansl, C.L., 1978. Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychol. Rev.* 85, 445–463.
- Kruskal, J.B., 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1–27.
- Kruskal, J.B., 1964b. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29, 115–129.
- Kruskal, J.B., Carroll, J.D., 1969. Geometrical models and badness-of-fit functions. In: Krishnaiah, P.K. (Ed.), *Multivariate Analysis II*. Academic Press, New York, pp. 639–671.
- Luce, R.D., 1959. *Individual Choice Behavior*. Wiley, New York.
- Luce, R.D., 1963. Detection and recognition. In: Luce, R.D., Bush, R.R., Galanter, E. (Eds.), *Handbook of Mathematical Psychology*, vol. 1. Wiley, New York, pp. 103–189.
- McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12, 153–157.
- Nosofsky, R.M., 1991. Stimulus bias, asymmetric similarity, and classification. *Cognit. Psychol.* 23, 94–140.
- Okada, A., Imaizumi, T., 1987. Nonmetric multidimensional scaling of asymmetric proximities. *Behaviormetrika* 21, 81–96.
- Okada, A., Imaizumi, T., 1997. Asymmetric multidimensional scaling of two-mode three-way proximities. *J. Classif.* 14, 195–224.
- Ramsay, J.O., 1980. Some small sample results for maximum likelihood estimation in multidimensional scaling. *Psychometrika* 45, 139–144.
- Read, C.B., 1978. Tests of symmetry in three-way contingency tables. *Psychometrika* 43, 409–420.
- Rocci, R., Bove, G., 2002. Rotation techniques in asymmetric multidimensional scaling. *J. Comput. Graph. Statist.* 11, 405–419.
- Rushen, J., 1982. The peck orders of chickens: How do they develop and why are they linear? *Anim. Behav.* 30, 1129–1137.
- Saito, T., 1991. Analysis of asymmetric proximity matrix by a model of distance and additive terms. *Behaviormetrika* 29, 45–60.
- Saito, T., Takeda, S., 1990. Multidimensional scaling of asymmetric proximity: Model and method. *Behaviormetrika* 28, 49–80.
- Sato, Y., 1988. An analysis of sociometric data by MDS in Minkowski space. In: Matsushita, K. (Ed.), *Statistical Theory and Data Analysis II*. North-Holland, Amsterdam, pp. 385–396.
- Schönemann, P.H., Carroll, R.M., 1970. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika* 35, 245–255.
- Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Statist.* 6, 461–464.
- Smith, J.E.K., 1980. Models of identification. In: Nickerson, R. (Ed.), *Attention and Performance VIII*. Erlbaum, Hillsdale, NJ.
- Stuart, A., 1955. A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika* 42, 412–416.
- Takane, Y., 1981. Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika* 46, 9–28.
- Takane, Y., Shibayama, T., 1986. Comparison of models for stimulus recognition data. In: de Leeuw, J., Heiser, W., Meulman, J., Critchley, F. (Eds.), *Multidimensional Data Analysis*. DSWO Press, Leiden, pp. 119–138.
- ten Berge, J.M.F., 1997. Reduction of asymmetry by rank-one matrices. *Comput. Statist. Data Anal.* 24, 357–366.
- Tobler, W., 1976–77. Spatial interaction patterns. *J. Environ. Syst.* 6, 271–301.
- Torgerson, W.S., 1958. *Theory and Methods of Scaling*. Wiley, New York.
- Townsend, J.T., Landon, D.E., 1983. Mathematical models of recognition and confusion in psychology. *Math. Soc. Sci.* 4, 25–71.
- Trendafilov, N.T., 2002. GIPSCAL revisited. A projected gradient approach. *Statist. Comput.* 12, 135–145.
- Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* 54, 426–482.
- Weeks, D.G., Bentler, P.M., 1982. Restricted multidimensional scaling models for asymmetric proximities. *Psychometrika* 47, 201–208.
- Winsberg, S., Carroll, J.D., 1989. A quasi-nonmetric method for multidimensional scaling via an extended Euclidean model. *Psychometrika* 54, 217–229.
- Yang, Y., 2005. Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika* 92, 937–950.
- Young, F.W., 1975. An asymmetric Euclidean model for multi-process asymmetric data. Paper presented at US-Japan Seminar on MDS, San Diego, USA. pp. 79–88.
- Zielman, B., Heiser, W.J., 1996. Models for asymmetric proximities. *Br. J. Math. Statist. Psychol.* 49, 127–146.