A maximum likelihood method for asymmetric MDS

◯ Shingo Saburi
Faculty of Letters
Graduate School of Aichi Gakuin University

Naohito Chino
Faculty of Psychological & Physical Science
Aichi Gakuin University

1 Introduction

Many models have been reported for asymmetric multidimensional scaling (MDS) since Young (1975) proposed ASYMSCAL. However, the asymmetric MDS has remained a discriptive method, whereas the symmetric MDS has some inferential methods based on the maximum likelihood (ML) method (for example Ramsay, 1977; Takane, 1981). Although Chino (1992) proposed the framework of an ML method for the asymmetric MDS which extended Takane's (1981) model to asymmetric data, the algorithm of the method has not been completed yet.

This study develops Chino's (1992) method further and attempts to complete the algorithm. Our method has three main advantages. First, it enables us to test the symmetry hypothesis for the given similarity data. Second, we can compare the goodness of fit of the data among some extant models for the asymmetric MDS using the Akaike information criterion, AIC (Akaike, 1973, 1977). Third, it allows us to employ not only the metric data but also nonmetric data.

2 The Method

2.1 Proposal of Models

According to Takane (1981), we will consider the three models, i.e., the representation model, the error model, and the response model.

representation model Any model proposed for the asymmetric MDS may be basically adopted as the representation model. For the present, however, we adopt the Hermitian Form Model (HFM) (Chino & Shiraiwa, 1993), which considers the similarity as a kind of inner product (precisely the Hermitian inner product) and embeds the objects in complex space. Following the HFM, we define the proximity from object i to object j by

$$s_{ij} = \sum_{a=1}^{A} w_a\{(r_{ia}r_{ja} + c_{ia}c_{ja}) + (c_{ia}r_{ja} - r_{ia}c_{ja})\} + \mu, \tag{1}$$

where A is the dimensionality, w_a is the weight on a-th dimension, and r and c are the real and imaginary coordinates, respectively. We put the additive constant μ since the HFM assumes the similarity matrix to be measured at the rational scale level. The configuration of the HFM has the (semi) unitary constraint, which is examined in our procedure.

error model The similarity s_{ij} in (1) is assumed to be error-perturbed by some psychological process. In our method s_{ij} can be either positive or negative theoretically, because it is a kind of inner product. Thus, we consider only the additive error model of the two error models proposed by Takane (1981), i.e.,

$$\tau_{ijk}^{(t)} = s_{ij} + e_{ijk}^{(t)}, \qquad e_{ijk}^{(t)} \sim N\Big(0, \sigma_k^2\Big), \tag{2}$$

where $\tau_{ijk}^{(t)}$ is a psychological value for subject k corresponding to the proximity from object i to object j at replication t.

response model We employ the successive categories scaling model (Torgerson, 1958), following Takane (1981). Let a rating scale be composed of M observation categories $C_1, C_2, \ldots, C_M$, and let b_{km} represent the upper boundary of the m-th category for subject k. Under the normal distribution assumption made above, we may assume

$$-\infty = b_{k0} \leq \cdots \leq b_{km} \leq \cdots \leq b_{kM} = \infty, \tag{3}$$

and define the probability of a subject's response by

$$p_{ijkm} = \Pr(o_{ijkr} \in C_m) = \Pr(b_{k(m-1)} < \tau_{ijk} \leq b_{km}) = \int_{b_{k(m-1)}}^{b_{km}} f(\tau_{ijk}) d\tau_{ijk}, \tag{4}$$

where o_{ijkr} is the proximity from object i to object j for subject k at replication r, and f is the density function of the normal distribution.

Here we make three assumptions concerning these category boundaries, accoding to Takane (1981): linear constraints without individual differences, unrestricted with individual differences, and completely unrestricted. The latter two contraints correspond to a nonmetric case.

2.2 Procedure

A subject's response may be coded as

$$Z_{ijkmr} = \begin{cases} 1, & \text{when } o_{ijkr} \in C_m, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

In order to estimate the vector of all parameters $\boldsymbol{\theta}$, we may maximize the logarithm of the joint likelihood of the observations,

$$\log L = \sum_k \sum_{i,\,j} \sum_m Y_{ijkm} \log p_{ijkm}, \tag{6}$$

where $Y_{ijkm} = \sum_r Z_{ijkmr}$. Our procedure consists of two steps. In step 1, we estimate the unrestricted parameters and test the symmetry hypothesis. If it is rejected, we go to step 2, where we estimate the parameters under the unitary hypothesis and test it.

step 1 We estimate the unrestricted parameters $\boldsymbol{\theta}^*$ by Fisher's scoring method in which $\boldsymbol{\theta}$ is updated by the following formula:

$$\boldsymbol{\theta}^{(q+1)} = \boldsymbol{\theta}^{(q)} + \epsilon^{(q)} \boldsymbol{I}_{\theta^{(1)}}^{-1} \boldsymbol{u}(\boldsymbol{\theta}^{(q)}), \tag{7}$$

where (q) is the index of iteration number, ϵ is a step-size parameter, and $\boldsymbol{u}(\boldsymbol{\theta})$ and $\boldsymbol{I}_\theta$ is respectively defined by

$$\boldsymbol{u}(\boldsymbol{\theta}) = \left(\frac{\partial \log L}{\partial \boldsymbol{\theta}}\right) = \sum_k \sum_{i,\ j} \sum_m \frac{Y_{ijkm}}{p_{ijkm}} \left(\frac{\partial p_{ijkm}}{\partial \boldsymbol{\theta}}\right), \tag{8}$$

and

$$\boldsymbol{I}_\theta = -E\left(\frac{\partial^2 \log L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t}\right) = \sum_k \sum_{i,\ j} \sum_m \frac{n_{ijk}}{p_{ijkm}} \left(\frac{\partial p_{ijkm}}{\partial \boldsymbol{\theta}}\right) \left(\frac{\partial p_{ijkm}}{\partial \boldsymbol{\theta}}\right)^t, \tag{9}$$

where $n_{ijk} = \sum_{m=1}^{M} Y_{ijkm}$. When $\boldsymbol{I}_\theta$ is singular due to the nonuniqueness of the parameters, we may replace the regular inverse in (7) by the Moore-Penrose inverse, following Takane (1981). Then, we test the symmetry hypothesis $H_0^{(1)}$:

$$c_{ia} = 0, \quad i = 1, 2, \ldots, n;\ a = 1, 2, \ldots, A, \tag{10}$$

where n is the number of the objects. This hypothesis is equivalent to saying that there are no asymmetric components in the data. We use the Wald method (Aitchison & Silvey, 1960), which is based on the statistic

$$\boldsymbol{h}^t(\boldsymbol{\theta}^*)\left[\boldsymbol{H}^t_{\theta^*} \boldsymbol{I}^{-1}_{\theta^*} \boldsymbol{H}_{\theta^*}\right]^{-1} \boldsymbol{h}(\boldsymbol{\theta}^*), \tag{11}$$

where $\boldsymbol{h}(\boldsymbol{\theta})$ is the vector in which the left-hand sides of (10) are arranged in order, and $\boldsymbol{H}_\theta$ is the matrix whose (i, j)-th component is $\partial h_j(\boldsymbol{\theta})/\partial \theta_i$. Under $H_0^{(1)}$ this statistic follows the χ^2-distribution with nA degrees of freedom asymptotically.

step 2 If $H_0^{(1)}$ is rejected in step 1, we find the estimates of the restricted parameters $\boldsymbol{\theta}^\dagger$ with the unitary hypothesis $H_0^{(2)}$:

$$\begin{cases} \sum_{i=1}^{n}(r_{ia}^2 + c_{ia}^2) - 1 = 0, & a = 1, 2, \ldots, A, \\ \sum_{i=1}^{n}(r_{ia} r_{ib} + c_{ia} c_{ib}) = 0, & a < b = 2, 3, \ldots, A, \\ \sum_{i=1}^{n}(r_{ia} c_{ib} - r_{ib} c_{ia}) = 0, & a < b = 2, 3, \ldots, A, \end{cases} \tag{12}$$

so that our model represents the HFM precisely. We use the Lagrange multiplier method (Aitchison & Silvey, 1960), which updates $\boldsymbol{\theta}$ by the following formula:

$$\begin{pmatrix} \boldsymbol{\theta}^{(q+1)} \\ \boldsymbol{\lambda}^{(q+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta}^{(q)} \\ \boldsymbol{\lambda}^{(q)} \end{pmatrix} + \begin{pmatrix} \frac{1}{N_T} \boldsymbol{I}_{\theta^{(1)}} & -\boldsymbol{H}_{\theta^{(1)}} \\ -\boldsymbol{H}^t_{\theta^{(1)}} & \boldsymbol{O} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{N_T} \boldsymbol{u}_{\theta^{(q)}} + \boldsymbol{H}_{\theta^{(q)}} \boldsymbol{\lambda}^{(q)} \\ \boldsymbol{h}(\boldsymbol{\theta}^{(q)}) \end{pmatrix}, \tag{13}$$

where $\boldsymbol{\lambda}$ is the vector of the Lagrange multipliers, and N_T is the total number of the similarity judgements. As in step 1, $\boldsymbol{h}(\boldsymbol{\theta})$ is the vector in which the left-hand sides of (12) were arranged, and $\boldsymbol{H}_\theta$ is the matrix whose (i, j)-th component is $\partial h_j(\boldsymbol{\theta})/\partial \theta_i$. Then, we test $H_0^{(2)}$ by the statistic

$$\boldsymbol{u}^t(\boldsymbol{\theta}^\dagger) \boldsymbol{I}^{-1}_{\theta^\dagger} \boldsymbol{u}(\boldsymbol{\theta}^\dagger), \tag{14}$$

which follows the χ^2-distribution with A^2 degrees of freedom asymptotically under $H_0^{(2)}$.

2.3 Model Comparison

In this study we adopted the HFM, one of the non-distance models, as the representation model. Of course, any other model may be adopted, for exapmle, Young's (1975) ASYMSCAL in which the proximity from object i to object j is represented as the square of the weighted Euclidian distance:

$$d_{ij}^2 = \sum_{a=1}^{A} w_{ia}(x_{ia} - x_{ja})^2, \quad w_{ia} \geq 0, \tag{15}$$

where w_{ia} is the stimulus weight of object i on a-th dimension. Then, the symmetry hypothesis may be given by

$$w_{ia} - w_{i+1,a} = 0, \quad i = 1, 2, \ldots, n-1; \; a = 1, 2, \ldots, A. \tag{16}$$

For the given similarity data, we can compare the goodness of fit among some representation models, using AIC defined by

$$\text{AIC} = -2 \log L + 2\text{d.f.}, \tag{17}$$

where d.f. is the effective number of parameters in a model (Akaike, 1973, 1977).

3 The Result

The result of applications to empirical data may be presented at the meeting.

References

Aitchison, J. & Silvey, S. D. (1960). Maximum-likelihood estimation procedures and associated tests of significance. *Journal of the Royal Statistical Society*, Series B (Methodological), **22**, 154-171.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *The second international symposium on information theory*. Budapest: Akadémiai Kiado.

Akaike, H. (1977). On entropy maximization principle. P. R. Krishnaiah (Ed.) *Applications of Statistics*. Holland: North-Holland Publishing Co.

Chino, N. (1992). Metric and nonmetric Hermitian canonical models for asymmetric MDS. *Proceedings of the 20th annual meeting of the Behaviormetric Society of Japan* (pp.246-249). Tokyo, Japan.

Chino, N., & Shiraiwa, K. (1993). Geometrical structures of some non-distance models for asymmetric MDS. *Behaviormetrika*, **20**, 35-47.

Ramsay, J. O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika*, **42**, 241-266.

Takane, Y. (1981). Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika*, **46**, 9-28.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Young, F. W. (1975). An asymmetirc Euclidean model for multi-process asymmetric data. *Paper presented at U.S.-Japan Seminar on MDS*, San Diego, U.S.A.