Behaviormetrika
Vol.38, No.1, 2011, 17–31

BAYESIAN NONMETRIC SUCCESSIVE CATEGORIES MULTIDIMENSIONAL SCALING

Kensuke Okada* and Shin-ichi Mayekawa**

A Bayesian nonmetric successive categories multidimensional scaling (MDS) method is proposed. The proposed method can be seen as a Bayesian alternative to the maximum likelihood multidimensional successive scaling method proposed by Takane (1981), or as a nonmetric extension of Bayesian metric MDS by Oh and Raftery (2001). The model has a graded-response type measurement model part and a latent metric MDS part. All the parameters are jointly estimated using a Markov chain Monte Carlo (MCMC) estimation technique. Moreover, WinBUGS/OpenBUGS code for the proposed methodology is also given to aid applied researchers. The proposed method is illustrated through the analysis of empirical two-mode three-way similarity data.

1. Introduction

Applications of Bayesian inference to behaviormetric models have achieved much success. The examples include exploratory factor analysis (Martin and McDonald, 1975; Press and Shigemasu, 1989), confirmatory factor analysis (Lee, 1989), item response theory (Swaminathan and Gifford, 1982), and structural equation modeling (Scheines, Hoijtink, and Boomsma, 1999; Lee and Xia, 2008).

Compared to these popular psychometric models, the introduction of Bayesian estimation is still limited for multidimensional scaling(MDS), which is also a very popular psychometric methodology. The initial study about Bayesian estimation in the MDS model was conducted by DeSarbo, Kim, Wedel, and Fong (1998). They proposed a Bayesian method for the spatial representation of pick any/J data, which areobtained from multiple response questions. DeSarbo, Kim, and Fong (1999) proposed a similar method for binary choice data. Although their work is valuable in establishing a Bayesian approach for spatial representationof such two-mode two-way type data, their method is not directly applicable to one-mode two-way or two-mode three-way data, which are historically the most popular form of input data for MDS analysis. Later, Oh and Raftery (2001) proposed the first general Bayesian MDS method for one-mode two-way data. Their method can be considered as a Bayesian alternative to the probabilistic metric MDS proposed by Ramsay (1977). Okada and Shigemasu (2009) proposed a set of R functions for Bayesian metric MDS via the use of WinBUGS. However, currently, no method (and of course, no computer program) is available for the Bayesian MDS analysis of successive categories data.

Oh and Raftery (2001) pointed out that Bayesian inference for the MDS model has many advantages. First, the distributional properties of the Bayesian analysis are ex-

---

Key Words and Phrases: Bayesian estimation, nonmetric multidimensional scaling, successive categories scaling, graded response model, Markov chain Monte Carlo, WinBUGS/OpenBUGS

∗ Senshu University
∗∗ Tokyo Institute of Technology

act for all sample sizes, and therefore need not be justified in terms of any asymptotic assumption. On the other hand, justifications of maximum likelihood MDS rely on an asymptotic theory, which may not be applicable in practice.

Second, in Bayesian inference, all the estimation results are given in the form of posterior probability, which is quite easy to interpret. For example, once the posterior distributions of the configuration matrix are derived, we can obtain the posterior distributions of any function of the configuration, such as the distance matrix ($\mathbf{D} = \{\mathrm{d_{ij}}\}$). Therefore, one can easily evaluate the posterior probability that the distance is above some criterion c, that is, $\mathrm{d_{ij}} > c$. The posterior probability is more intuitive than the p value, which is the chance of observing a value as extreme as the observed value given repeated sampling under the null hypothesis (Dunson, 2001).

Third, the Bayesian approach allows for the incorporation of meaningful prior information of the parameters, if available, into the statistical inference. This can be done by placing prior restrictions on the possible values of the parameters or by assigning a prior distribution on the basis of the summary statistics obtained from previous studies. It has been noted in many preceding studies that the incorporation of prior knowledge results in simpler estimation procedures, smaller posterior standard deviations, and improved predictive performance (e.g. Allenby, Arora, and Ginter, 1995; Gamerman and Lopes, 2006; Gelman, Carlin, Stern, and Rubin, 2002; Scheines et al., 1999).

We also note that in standard Bayesian context, hierarchical Bayesian extension of existing models is simplified by the fact that inference is based on the marginal posterior distribution, which can be determined using the Markov chain Monte Carlo (MCMC) technique. On the other hand, the high-dimensional integrals that are usually required to determine the marginal likelihood function in maximum likelihood estimation are often intractable, except for simple problems (see, e.g., Casella and George, 1992). Another advantage is that Bayesian estimation using the MCMC algorithm naturally handles missing data by treating it the same as other parameters (Levy, 2009; Little and Rubin, 2002; Tan, Tian and Ng, 2009). For further discussions of the benefit of Bayesian estimation, see e.g., Box and Tiao (1992), Carlin and Louis (2008), Gamerman and Lopes (2006), Gelman et al. (2002), and Gill (2008).

In fields such as behaviormetrics and psychometrics, similarity data are typically obtained from human judgment, and it is a common practice to obtain similarity data on successive categories, such as an M-point rating scale (M = number of possible grades). Therefore, Bayesian MDS for successive categories data would provide many benefits (such as those stated above) to this field of research.

On the basis of such motivations, in this paper, we propose a Bayesian nonmetric successive categories MDS procedure. The paper is constructed as follows. In Section 2, following Takane (1981), the successive categories MDS model is introduced. Section 3 defines the prior and presents MCMC estimation and post-processing techniques. In Section 4, the proposed method is illustrated by real data analysis with and without missing observations. Finally, conclusions are presented in Section 5. The implementation of Bayesian inference has become easier with the advent of software such as WinBUGS and OpenBUGS. The BUGS code for the proposed model is given in the Appendix.

2. Successive categories MDS model

This section introduces the successive categories MDS model proposed by Takane (1981). Suppose the observed dissimilarity between objects i and j by subject k is represented by $\mathrm{o_{ijk}}$. (If the observed data are not two-mode three-way but one-mode two-way data, $\mathrm{o_{ijk}}$ can be simply replaced by $\mathrm{o_{ij}}$ and the following discussion holds the same.) In addition, in this paper, we suppose that the observed data takesuccessive categorical values, $\mathrm{o_{ijk}} = \mathrm{m}$, with $\mathrm{m} = 1, \ldots, \mathrm{M}$.

On the other hand, we suppose that every individual k shares the same configuration and has equal distances between objects behind the data. This implies that the true distance,

$$\mathrm{d_{ij}} = \sqrt{\sum_{\mathrm{l}=1}^{\mathrm{p}} (\mathrm{x_{il}} - \mathrm{x_{jl}})^2}, \tag{1}$$

does not contain the subscript k, where $\mathrm{x_{il}}$ is the coordinate of stimulus i on dimension l and p is the dimensionality of the space. Here, we suppose a standard additive error model,

$$\begin{aligned} \omega_{\mathrm{ijk}} &= \mathrm{d_{ij}} + \mathrm{e_{ijk}}, \\ \mathrm{e_{ijk}} &\sim \mathrm{N}\left(0, \sigma_{\mathrm{k}}^2\right), \end{aligned} \tag{2}$$

where $\mathrm{e_{ijk}}$ is an error term, σ_{k}^2 is its variance, and ω_{ijk} is the error-perturbed distance for subject k. In successive categories modeling, categories of the observed data are represented by a set of successive intervals, which are mutually exclusive and exhaustive (i.e., each piece of data must fall into one and only one category), of the latent variable ω_{ijk}. These intervals are demarcated by upper and lower boundaries. The upper boundary of the m-th category, which coincides with the lower boundary of the (m + 1)-th category, is denoted by $\mathrm{b_{km}}$. If the values of the boundaries $\mathbf{b} = \{\mathrm{b_{km}}\}$ are given, the category of the observed data is determined by the magnitude of ω_{ijk}. Specifically,

$$\mathrm{o_{ijk}} = \mathrm{m} \quad \text{if } \mathrm{b_{k(m-1)}} < \omega_{\mathrm{ijk}} < \mathrm{b_{km}}. \quad (\mathrm{m} = 1, \ldots, \mathrm{M}). \tag{3}$$

Without loss of generality, we set $\mathrm{b_{k0}} = -\infty$, $\mathrm{b_{kM}} = +\infty$. Then, the probability that the observed category of objects i and j for subject k is m, $\mathrm{p}\left(\mathrm{o_{ijk}} = \mathrm{m}\right)$, is given by

$$\mathrm{p}\left(\mathrm{o_{ijk}} = \mathrm{m}\right) = \int_{\mathrm{a_{ijk(m-1)}}}^{\mathrm{a_{ijkm}}} f\left(\mathrm{z}\right) \mathrm{dz}, \tag{4}$$

where $\mathrm{f}(\cdot)$ is the pdf of the standard normal distribution and

$$
\begin{aligned}
\mathrm{z} &= \frac{\omega_{\mathrm{ijk}} - \mathrm{d_{ij}}}{\sigma_{\mathrm{k}}}, \\
\mathrm{a_{ijk(m-1)}} &= \frac{\mathrm{b_{k(m-1)}} - \mathrm{d_{ij}}}{\sigma_{\mathrm{k}}}, \\
\mathrm{a_{ijkm}} &= \frac{\mathrm{b_{km}} - \mathrm{d_{ij}}}{\sigma_{\mathrm{k}}}.
\end{aligned}
\tag{5}
$$

The probability of a particular rating judgment is written as

$$
\mathrm{p_{ijk}} = \prod_{\mathrm{m}=1}^{\mathrm{M}} \mathrm{p(o_{ijk} = m)^{u_{ijkm}}}, \tag{6}
$$

where the indicator variable $\mathrm{u_{ijkm}}$ takes the value one when the observed dissimilarity between objects i and j for subject k equals category m, and zero otherwise, i.e.,

$$
\mathrm{u_{ijkm}} = \begin{cases} 1, & (\text{when } \mathrm{o_{ijk} = m}) \\ 0. & (\text{otherwise}) \end{cases} \tag{7}
$$

Assuming the independence of the additive error terms, the joint likelihood function of the model becomes

$$
\mathrm{L}\left(\boldsymbol{\theta}|\mathbf{O}\right) = \prod_{\mathrm{k}} \prod_{\mathrm{i,j}} \mathrm{p_{ijk}}, \tag{8}
$$

where $\mathbf{O}$ denotes the observed data and $\boldsymbol{\theta}$ denotes the parameters associated with the model (i.e., $\mathbf{X}$, σ^2, and $\mathbf{b}$). Note that from the psychometric viewpoint the above specification resembles Samejima's(1969) graded response model (GRM). It can be said that the Takane'snonmetricMDS model is an extension of the metric MDS model by using GRM as a measurement model.

3. Bayesian analysis of the model

3.1 Prior distribution

For Bayesian analysis of the model described in the previous section, the specification of the prior distributions isrequired. For the prior distribution of $\mathbf{x}_\mathrm{i} = (\mathrm{x_{i1}}, \ldots, \mathrm{x_{ip}})'$, which corresponds to the i-th row of $\mathbf{X}$, we use a multivariate normal distribution with mean 0 and a diagonal covariance matrix $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_\mathrm{p}\}$. Specifically,

$$
\mathbf{x}_\mathrm{i} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Lambda}\right), \tag{9}
$$

independently for $\mathrm{i} = 1, \ldots, \mathrm{n}$ objects. For each element of $\boldsymbol{\Lambda}$, an inverse gamma hyperprior is considered

$$
\lambda_\mathrm{l} \sim \mathrm{IG}(\alpha_\lambda, \beta_\lambda), \tag{10}
$$

independently for $\mathrm{l} = 1, \ldots, \mathrm{p}$. This inverse gamma is a natural conjugate distribution of the normal variance parameter. The above specification is based on the Bayesian metric MDS of Oh and Raftery (2001).

For the prior distribution of $\mathbf{b}$, for technical reasons (i.e., to satisfy the order restriction on $\mathbf{b}$), the following reparameterization is introduced first:

$$\mathrm{b}^*_{\mathrm{km}} = \mathrm{b}_{\mathrm{km}} - \mathrm{b}_{\mathrm{k(m-1)}}. \tag{11}$$

Then we consider the prior distribution,

$$\mathrm{b}^*_{\mathrm{k1}} \sim \mathrm{N}(0, \sigma^2_{\mathrm{b}}), \tag{12}$$

$$\mathrm{b}^*_{\mathrm{km}} \sim \mathrm{G}\left(\alpha_{\mathrm{b}}, \beta_{\mathrm{b}}\right), \quad (\mathrm{m} \geq 2) \tag{13}$$

where G denotes a gamma distribution. This type of specification has been used in the literature of the Bayesian orderedprobit model (e.g., Lancaster, 2004).

For the prior distribution of σ^2_{k} in Eq. (2), we again use the inverse gamma distribution,

$$\sigma^2_{\mathrm{k}} \sim \mathrm{IG}(\alpha_\sigma, \beta_\sigma). \tag{14}$$

3.2 Posterior estimation

Based on the likelihood function (Eq. (8)) and prior distribution (Eqs. (9)–(13)), the posterior distribution takes the form

$$\pi\left(\boldsymbol{\theta}|\mathbf{O}\right) \propto \mathrm{L}\left(\boldsymbol{\theta}|\mathbf{O}\right)\pi\left(\boldsymbol{\theta}\right), \tag{15}$$

where $\pi\left(\boldsymbol{\theta}\right)$ denotes the product of prior distributions in Eqs. (9)–(13). Because theposterior distribution $\pi\left(\boldsymbol{\theta}|\,\mathbf{O}\right)$ is not known in closed form, an MCMC algorithm is employed to generate samples from the posterior distribution. To achieve approximately the equilibrium distribution, the MCMC process should first be run for a given number of iterations; this period is known as "burn-in." Then, after a sufficiently long burn-in run of the Markov chain, the algorithm generates random samples from the posteriordistribution.

Significant progress in facilitating the routine implementation of the MCMC algorithm has been made since the development and release of the BUGS software (Lunn, Thomas, Best, and Spiegelhalter, 2000). Once the prior and likelihood are specified, the BUGS program draws samples from the joint posterior distribution by using the MCMC algorithm. As the proposal distribution of the Metropolis method, BUGS automatically generates samples from a normal proposal distribution centered at the current point with a self-tuning variance. Some reviews of BUGS can be found in Cowles (2004), Lunn, Spiegelhalter, Thomas and Best (2009) and Ntzoufras(2009). The BUGS code of the proposed method used in the analysis in the next section is given in the Appendix.

Note that as a result of assuming Euclidean distance in Eq. (1) (which is the most popular distance assumption in MDS), posterior samples of $\mathbf{X}$ would be invariant under translation, rotation, and reflection about the origin unless strong informative priors are used. Therefore, as in Oh and Raftery (2001), the convergence of $\mathbf{D}$, rather than $\mathbf{X}$, was checked. Moreover, instead of the rather ad hoc approach of calculating the approximate mode (which is the approach taken by Oh and Raftery, 2001), post-processing was conducted after the MCMC sampling. The idea is to post-process MCMC samples using an appropriate loss function, as proposed by Celeux, Hurn, and Robert (2000) for mixture posterior distribution. In the current method, the following loss function was minimized:

$$\rho = ||\,\bar{\mathbf{X}} - \mathbf{X}^{(\mathrm{i})}\mathbf{H}^{(\mathrm{i})}||^2, \tag{16}$$

where $\mathbf{X}^{(\mathrm{i})}$ is the i-th MCMC sample of $\mathbf{X}$, $\bar{\mathbf{X}}$ is the mean of all $\mathbf{X}^{(\mathrm{i})}$'s and is the target matrix of Procrustes rotation, and $\mathbf{H}^{(\mathrm{i})}$ is a rotation matrix to be estimated corresponding to $\mathbf{X}^{(\mathrm{i})}$. Tominimize the function in Eq. (16), an alternating least squares algorithm was used. This algorithm works as follows:

(1) Rotate each $\mathbf{X}^{(\mathrm{i})}$'s by Procrustes rotation with target matrix $\bar{\mathbf{X}}$
(2) Update $\bar{\mathbf{X}}$ from the newly rotated $\mathbf{X}^{(\mathrm{i})}$'s
(3) Repeat steps (1) and (2) until convergence

4. Illustration

4.1 Real data analysis

Methods

The proposed Bayesian method is applied to the analysis of three-way similarity judgment data. The similarity data on 10 major business districts in Tokyo were collected by 21 university students. The similarity between all pairs of business districts was rated on a 7-point scale (that is, M = 7). These data weretransformed to dissimilarity by subtracting the rating from 8, and then used as inputs to the proposed method.

For the values of hyperparameters, noninformative priors are used in the analysis. All scale and shape hyperparameters of gamma and inverse gamma distributions were set as 10^{-3} (Eqs. (10), (13), and (14)). The variance hyperparameter for normal distribution was set as 10^6 (Eq. (12)). These settings seem to be adequate for noninformative Bayesian analysis (see, e.g., Albert, 2009; Lancaster, 2004). The number of dimensions p was fixed to 2. For the initial values of $\mathbf{X}$ and $\boldsymbol{\Lambda}$, we applied the suggestion of Oh and Raftery (2001). Specifically, we utilized classical MDS solutions for the mean observed dissimilarity over the individual matrix as initial values of $\mathbf{X}$. Also, $\frac{\sum_{\mathrm{i,j}}(\bar{\mathrm{o}}_{\mathrm{ij}}-\delta_{\mathrm{ij}})^2}{\mathrm{n(n-1)/2}}$, where $\bar{\mathrm{o}}_{\mathrm{ij}}$ is the mean observed dissimilarity and δ_{ij} is the corresponding distance obtained from classical MDS, is used as initial values of λ_{l}'s. For the rest of the parameters, $\mathrm{b}^*_{\mathrm{km}}$'s and σ^2_{k}'s, the vectors of 1's are utilized as initial values.

In the estimation, the first 1,000 iterations of MCMC were discarded as burn-in and the other 10,000 iterations were executed to construct posterior distributions. The convergence of the chain was monitored and checked by Heidelberger and Welch's

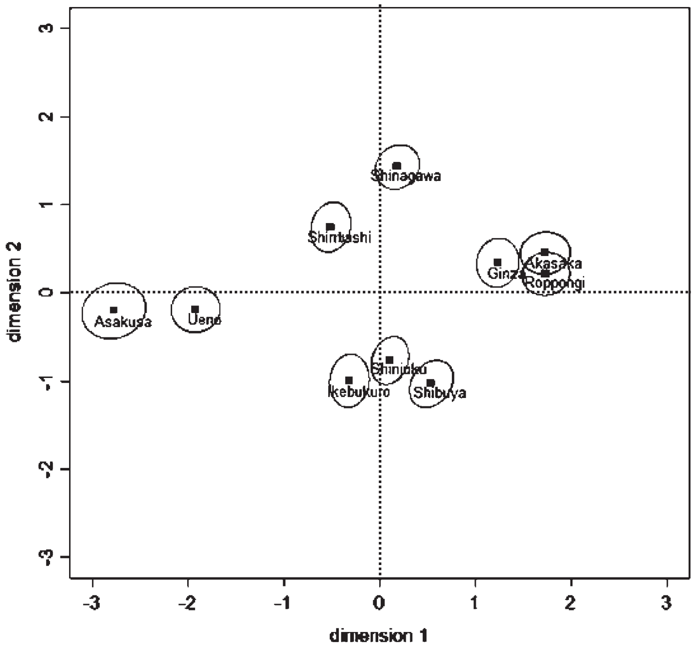

Figure 1: Estimated posterior configurations and 95% credibility regions of Tokyo business districts from the proposed method.

(1983) method, which uses the Cramer-von-Mises statistic to test the null hypothesis that the sampled values are obtained from a stationary distribution by using CODA package in R (Plummer, Best, Cowles, and Vines, 2006).The convergence of the post-processing algorithm, which minimizes Eq. (16), is rapid; in fact in our analysis, each element of $\bar{\mathbf{X}}$ did not change more than 10^{-6} after three iterations. It is confirmed that after the post-processing, each element of $\mathbf{X}$ has an approximately unimodal and equitailed posterior distribution.

Results

The resultant configuration is shown in Fig. 1. The small black squares represent the posterior mean of $\mathbf{X}$ after post-processing, which was plotted in a two-dimensional space, i.e., the posterior estimates of the business districts. The surrounding ellipses represent 95% Bayesian credibility ellipses fitted to each postprocessed MCMC sample (see, e.g., Calvetti and Somersalo, 2007). The region within each credibility ellipse is interpreted as containing the true model parameter with 95% probability. Many former studies pointed out the usefulness of such a direct probabilistic statement about parameter credibility regions obtained from Bayesian analysis; on the other hand, the frequentistconfidence interval is not a probability of parameters and, therefore, its interpretation requires the concept of repeated sampling (see, e.g., Berger, 2004).

The fact that the plot of such a posterior probability region is possible indicates one of the advantages of Bayesian estimation. In addition, the result is easily under-

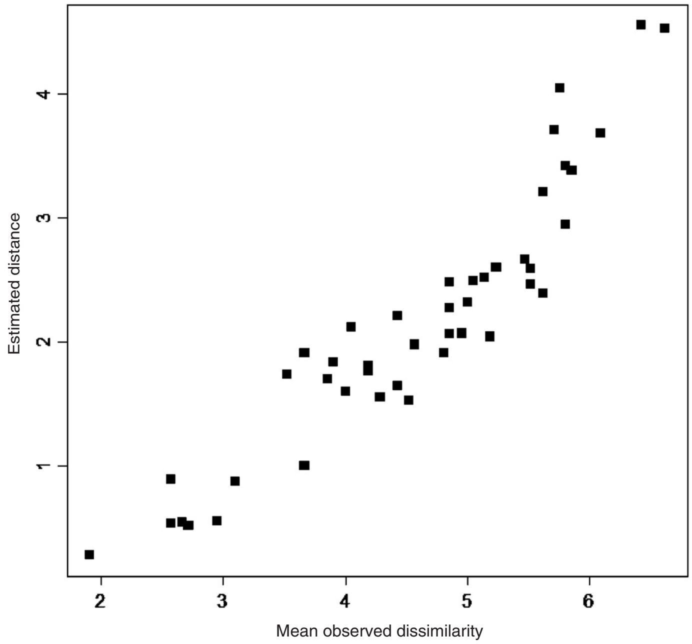

Figure 2: Scatterplot of estimated distances and mean observed dissimilarity.

standable. Dimension 1 can be interpreted as a modern-traditional dimension, with a cluster of modern business districts on one side (Ginza, Akasaka, and Roppongi) and traditional old towns on the other (Asakusa and Ueno). Similarly, dimension 2 can be interpreted as a business-leisure dimension, with business towns on one side (Shinagawa and Shimbashi) and leisure towns of youngsters on the other (Shibuya, Shinjuku, and Ikebukuro).

Fig. 2 is a plot of the mean (with respect to k individuals) of the observed dissimilarity measure ($\mathrm{o_{ijk}}$) versus the estimated distance ($\mathrm{d_{ij}}$). The points lie approximately on the diagonal line, apparently with no particular outliers. This fact can be the evidence that the cognitive locations of the downtowns are well recovered by the proposed model.

In Fig. 3, the posterior means of $\mathrm{b^*_{km}}$'s are transformed back to $\mathrm{b_{km}}$'s using the relation of Eq. (11) and are plotted against the upper boundaries of successive categories for all participants. It is easily seen that large individual differences exist in the threshold of categories, indicating the need for individual differences three-way data analysis, such as the proposed method, instead of the still common practice of averaging out the individual differences. For example, the person whose upper boundaries for categories 5 and 6 are beyond the range of this graph would be unlikely to judge the dissimilarity of two towns to be "6" and "7," although the towns are reasonably dissimilar. In addition, in this application, it is demonstrated that the estimated upper boundaries of categories 3 and 4 do not differ much for most participants. Therefore, these two categories might be merged to construct a 6-point scale, instead of 7, for

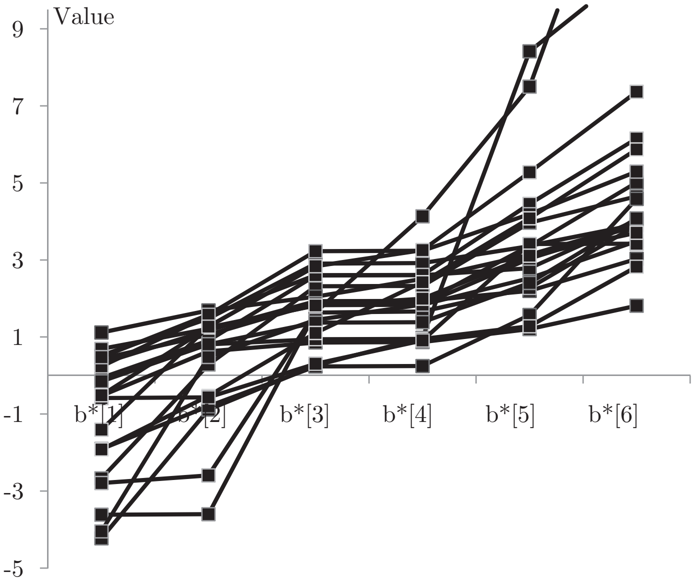

Figure 3: Values of posterior means of b_{km}'s against upper boundary of successive categories for each of the 21 individuals.

representing the same sort of cognitive difference of downtowns as observed here.

4.2 Analysis of missing data

Methods

Missing data are common in psychometrics. Although ideally one should aim to avoid missing data, this is often impossible in practice. Therefore, one needs to address the problem of missing data through statistical analysis. Bayesian methodology based on MCMC estimation is a suitable candidate for this purpose (e.g., Little and Rubin, 2002; Tan et al., 2009). In Bayesian analysis, each missing value is regarded as a hidden variable and estimated simultaneously with estimation of other parameters. One of the attractive features of MCMC estimation is that missing data are naturally handled within its scheme; at each step, a value for each missing data value is drawn on the basis of its conditional posterior distribution. Another attractive point is that BUGS will do this automatically when the data contains some missing values. Note that missing data in BUGS are treated as “missing at random,” that is, the missing-datageneration mechanism does not depend on the missing values, but perhaps on the observed values (Gelman et al., 2002).

To illustrate the proposed analysis in the missing data situation and to check the effect of the number of missing elements, we conducted a simulation study. The same data as in Section 4.1 were used; however, in this case, some predetermined number of elements was marked as missing for each subject. The number of missing elements *per subject* was manipulatedto from 1 to 6 (out of $\frac{10\times9}{2}$). For each condition of the number of missing element, 50 datasets with random missing elements were created. Then, they were analyzed in the same manner as in the previous analysis. Note that we can use the same code as in Section 4.1 to handle missing data. The settings of hyperparameters and initial values were also the same as in the previous section.

Results

To evaluate the effect of missing elements in the observation, the normed distances estimated from 50 datasets were plotted against the distances with no missing elements (i.e., the distances calculated from the result shown in Section 4.1); See Fig. 4. Correlation coefficients between them were also calculated and are shown in the figure. These figures and the corresponding coefficients show that the estimation works almost as well as when there exist one or two missing elements per subject (a, b). As the number of missing elements per subject grows, the probability of over- and under-estimation also grows. However, note that the correlation coefficient is still 0.91 for the condition of four missing elements per subject (d), which is large enough for practical use. The correlation coefficient drops to 0.78 for the condition of six missing elements per subject (f).

Although it may be dangerous to draw definite conclusions from a single set of simulation runs, from the above result, it would appear that at least a few missing elements per subject would not affect the result much. The proposed Bayesian method can be applied to data with a few missing elements per subject.

5. Conclusions

In this paper, we proposed a Bayesian approach to nonmetric successive categories multidimensional scaling. The model is a nonmetric extension of Oh and Raftery's (2001) Bayesian metric MDS, and can handle successive categories rating data. Our proposed model can be utilized in psychometrics and social sciences, in which successive categories rating data are common. In contrast to the standard least-squares type of MDS, in Bayesian MDS, the posterior credibility regions, which indicate the uncertainty of the parameters of each object in the resultant configuration, can be plotted by ellipses, which give additional information to the analyst. The proposed method can be easily implemented using BUGS software, which is currently a standard MCMC engine. Therefore, the applied users can easily use the proposed method using the BUGS code given in the Appendix.

The determination of the initial values in MCMC estimation is important. We used the estimates from two-mode classical MDS for parameters if possible, because classical MDS is easy to execute; for other parameters, we used rather ad hoc initial values. Note that the maximum likelihood method also requires iterations and therefore need to determine initial values. Therefore, the same prescriptions can be applied: change initial values if the convergence criterion is not satisfied (Hoshino, 2001).

To deal with the identification problem of the MCMC samples of the configuration parameter, we introduced a new post-processing approach. This approach is more in-

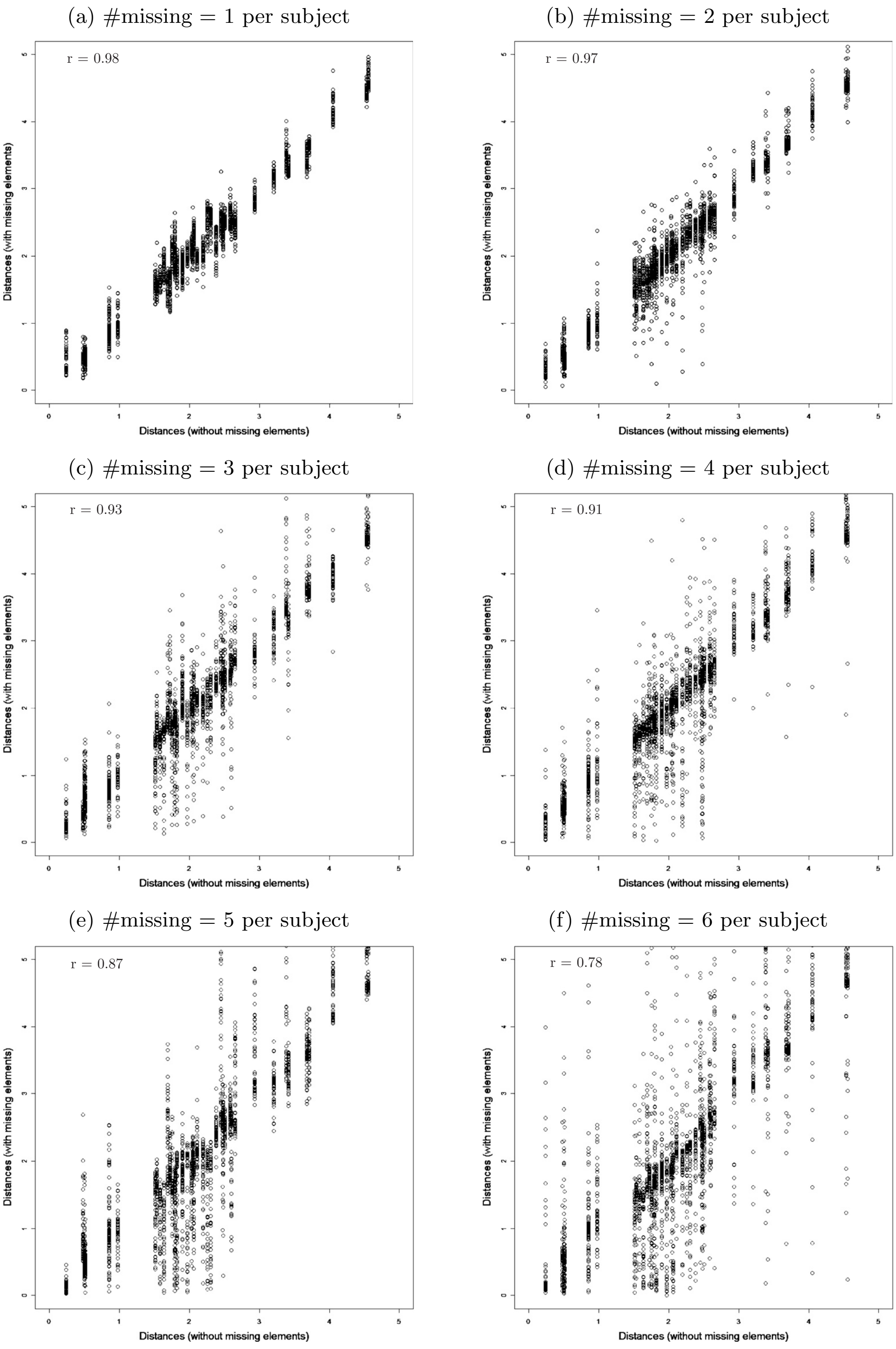

Figure 4: Plots of distances with and without missing data in 50 replications.

tuitive than the approximate mode approach used in a former study (Oh and Raftery, 2001) and is easier to implement using the BUGS estimation engine. Although further empirical study would be required, the result of the real data analysis is so far satisfactory.

Our model allows individual differences in category boundaries while keeping the "perceptual spaces" common. It would be possible, if desired, to construct the individual spaces by appropriately weighting the common space. However, as Carroll and Chang (1970) stated, "for many purposes, the individual spaces may not be necessary for an adequate comprehension of the data" (p.316).

It would be possible to expand the model presented above in numerous directions. First, individual differences of the respondents can be better represented by introducing latent class-type approaches. One of the most popular models of this type would be CLASCAL (Winsberg and De Soete, 1992), to which maximum likelihood estimation is commonly employed. As the integration of Bayesian MDS and latent class approaches have been drawing attention recently (e.g., Oh and Raftery, 2007; Park, DeSarbo, and Liechty, 2008), latent class extension of the proposed model would also be of interest. Second, when the number of items n increases, it is not easy to ask participants to makesimilarityjudgmentsfor all $n(n-1)/2$ pairs of items. In such a case, subjects may be assigned to different subgroups of items. The idea of equating in the item response theory or methods in the optimal design of experimentscould be applied to this type of missing data problem in a Bayesian context. Third, although there exist several methods for choosing the number of dimensions, such as several information criteria or Bayes factors, it is not clear which one should be preferred in this particular model. This would be an area worthy of future study.

Appendix. BUGS code

The BUGS code for estimating the parameters of the proposed model used in our illustrative example is as follows:

```
model{
  for( i in 2 : n ) {
    for( j in 1 : i-1 ) {
      for (k in 1:nind){
        obs[i,j,k] ~ dcat(p[i,j,k,1:7])
        p[i,j,k,1] <- phi((b1[k] - d[i,j])/sigma[k])
        p[i,j,k,7] <- 1- phi((b1[k]+b2[k]+b3[k]+b4[k]+b5[k]
          +b6[k]- d[i,j])/sigma[k])
        p[i,j,k,2] <- phi((b1[k]+b2[k] - d[i,j])/sigma[k]) -
          phi((b1[k] - d[i,j])/sigma[k])
        p[i,j,k,3] <- phi((b1[k]+b2[k]+b3[k] - d[i,j])
          /sigma[k]) - phi((b1[k]+b2[k] - d[i,j])/sigma[k])
        p[i,j,k,4] <- phi((b1[k]+b2[k]+b3[k]+b4[k] - d[i,j])
```

```
          /sigma[k]) - phi((b1[k]+b2[k]+b3[k] -d[i,j])
          /sigma[k])
        p[i,j,k,5] <- phi((b1[k]+b2[k]+b3[k]+b4[k]+b5[k] -
          d[i,j])/sigma[k]) - phi((b1[k]+b2[k]+b3[k]+b4[k] -
          d[i,j])/sigma[k])
        p[i,j,k,6] <-phi((b1[k]+b2[k]+b3[k]+b4[k]+b5[k]
          +b6[k] - d[i,j])/sigma[k])-phi((b1[k]+b2[k]+b3[k]
          +b4[k]+b5[k] - d[i,j])/sigma[k])
      }
      sqd[i,j] <- pow((X[i,1]-X[j,1]),2)+
        pow((X[i,2]-X[j,2]),2)
      d[i,j] <- sqrt(sqd[i,j])
    }
  }
for(l in 1 : ndim ) {
      for(k in 1 : n ) {
            X[k,l] ~ dnorm(0,invlambda[l])
      }
      invlambda[l] ~ dgamma(alpha.lam, beta.lam)
      lambda[l] <- 1/ invlambda[l]
}for (k in 1:nind){
  b1[k] ~ dnorm(0,tau.b)
  b2[k] ~ dgamma(alpha.b,beta.b)
  b3[k] ~ dgamma(alpha.b,beta.b)
  b4[k] ~ dgamma(alpha.b,beta.b)
  b5[k] ~ dgamma(alpha.b,beta.b)
  b6[k] ~ dgamma(alpha.b,beta.b)
  sigma[k] <- 1/sqrt(invsigma2[k])
  invsigma2[k] ~ dgamma(alpha.sig,beta.sig)
}
alpha.lam <- 1.0E-3
beta.lam <- 1.0E-3
tau.b <- 1.0E-6
alpha.b <- 1.0E-3
beta.b <- 1.0E-3
alpha.sig <- 1.0E-3
beta.sig <- 1.0E-3
}
```

References

Albert, J. (2009). *Bayesian Computation with R.* NY: Springer.

Allenby, G.M., Arora, N., & Ginter, J.L. (1995). Incorporating prior knowledge into the analysis of conjoint studies. *Journal of Marketing Research*, *32*, 152–162.

Berger, J. (2004). The case for objective Bayesian analysis. *Bayesian Analysis*, 1, 1–17.

Box, G.E.P., & Tiao, G.C. (1992). *Bayesian inference in statistical analysis. (Wiley classics library edition)* New York: Wiley.

Carroll, J.D., & Chang, J-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, *35*, 283–319.

Calvetti, D. & Somersalo, E. (2007). Introduction to Bayesian Scientific Computing. New York: Springer.

Casella, G. & George, E.I. (1992). Explaining the Gibb sampler. *American Statistician*, 46, 167–174.

Carlin, B.P., & Louis, T.A. (2008). *Bayesian methods for data analysis. (3 ed)* Boca Raton, FL: Chapman & Hall/CRC.

Celeux, G., Hurn, M., & Robert, C.P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, *95*, 957–970.

Cowles, M.K. (2004). Review of WinBUGS 1.4. *American Statistician*, *58*, 330–336.

Cox, D.R. (1982). "Comments". *Journal of the Royal Statistical Society, Series A, 145*, 308–309.

DeSarbo, W.S., Kim, Y., & Fong, D. (1999). A Bayesian multidimensional scaling procedure for the spatial analysis of revealed choice data. *Journal of Econometrics*, *89*, 79–108.

DeSarbo, W.S., Kim, Y., Wedel, M., & Fong, D.K.H. (1998). A Bayesian approach to the spatial representation of market structure from consumer choice data. *European Journal of Operational Research*, *111*, 285–305.

Dunson, D.B. (2001). Commentary: Practical advantages of Bayesian analysis of epidemiologic data. *American Journal of Epidemiology*, *153*, 1222–1226.

Gamerman, D. & Lopes, H.F. (2006). Markov chain Monte Carlo: stochastic simulation for Bayesian inference. (2 ed) Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2002). *Bayesian data analysis* (2 ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gill, J. (2008). *Bayesian methods: a social and behavioral science approach. (2 ed)* London, UK: Chapman & Hall/CRC.

Heidelberger, P., & Welch, P.D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, *31*, 1109–1144.

Hoshino, T. (2001). Bayesian inference for finite mixtures in confirmatory factor analysis. *Behaviormetrika*, *28*, 37–63.

Lancaster, T. (2004). *An introduction to modern Bayesian econometrics.* Malden, MA: Blackwell.

Lee, S.-Y. (1989). A Bayesian approach to confirmatory factor analysis. *Psychometrika*, *46*, 153–160.

Lee, X.-Y. & Xia, Y.M. (2008). A robust Bayesian approach for structural equation models with missing data. *Psychometrika*, *73*, 343–364.

Levy, R. (2009). The rise of Markov chain Monte Carlo estimation for psychometric modeling. *Journal of Probability and Statistics, 537139*, 1–18.

Little, R.J.A. & Rubin, D.B. (2002). *Statistical analysis with missing data (2 ed).* Hoboken, NJ: Wiley.

Lunn, D.J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–357.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: evolution, critique and further directions. *Statistics in Medicine*, *28*, 3049–3067.

Martin, J.K., & McDonald, R.P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, *40*, 505–517.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS.* Hoboken, NJ: Wiley.

Oh, M.S., & Raftery, A.E. (2001). Bayesian multidimensional scaling and choice of dimension. *Journal of the American Statistical Association*, *96*, 1031–1044.

Oh, M.S., & Raftery, A.E. (2007). Model-based clustering with dissimilarities: A Bayesian approach. *Journal of Computational and Graphical Statistics*, *16*, 559–585.

Okada, K. & Shigemasu, K. (2009). BMDS: A collection of R functions for Bayesian multidimensional scaling. *Applied Psychological Measurement*, *33*, 570–571.

Park, J., DeSarbo, W.S., & Liechty, J. (2008). A hierarchical Bayesian multidimensional scaling methodology for accommodating both structural and preference heterogeneity. *Psychometrika*, *73*, 451–472.

Plummer, M., Best, N.G., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6, 7–11.

Press, S.J., & Shigemasu, K. (1989). Bayesian inference in factor analysis. In L. Gleser, M. Perleman, S.J. Press & A. Sampson (Eds.), *Contributions to probability and statistics* (pp.271–287). New York: Springer-Verlag.

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

Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph, No. 17.*

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, *64*, 37–52.

Swaminathan, H., & Gifford, J.A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics*, *7*, 175–191.

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

Tan, M.T., Tian, G-L, & Ng, K.W. (2009). *Bayesian missing data problems: EM, data augmentation and noniterative computation.* Boca Raton: Chapman & Hall/CRC.

Winsberg, S., & De Soete, G. (1992). A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika*, *58*, 315–330.

(Received January 15 2010, Revised November 30 2010)