A Bayesian Approach to Asymmetric Multidimensional Scaling

Okada, Kensuke
Senshu University, Department of Psychology
2-1-1, Higashimita, Tama-ku, Kawasaki-shi,
Kanagawa 214-8580, Japan
E-mail: ken@psy.senshu-u.ac.jp

Introduction

Bayesian analysis of multidimensional scaling (MDS) model using Markov chain Monte Carlo (MCMC) algorithm has received considerable attention recently. There are several advantages in adopting Bayesian analysis over the standard methodology. First, the distributional properties of Bayesian analysis are exact, and do not depend on asymptotic assumptions. This implies that a small-sample inference proceeds in the same manner as a large-sample inference. On the other hand, justifications of maximum-likelihood MDS rely on the asymptotic theory, which may not be applicable in practice. Second, the results of Bayesian estimation are presented in the form of posterior probability, which is easy to interpret. For example, the result of Bayesian MDS can be used to estimate the posterior distribution of not only the distance matrix $D = \{d_{ij}\}$ but also any of its function. The posterior probability that the distance between any two objects is above some threshold c, that is, $d_{ij} > c$, can also be evaluated. In addition, the Bayesian paradigm allows the incorporation of prior information, whenever available, in the inference process.

However, existing Bayesian MDS techniques are limited to the analysis of symmetric data matrices. In contrast, asymmetric relationships among objects (variables) are frequently observed in reality. To the best of the authors' knowledge, up to now there is no Bayesian MDS method to analyze asymmetric data.

Given the above situation, we propose a Bayesian MDS model that incorporates asymmetric data structure. We consider the hill-climbing model proposed by Borg & Groenen (2005) and the jet-stream model proposed by Gower (1977). These models introduce a slope vector or a stream vector that measures to what extent it is more difficult or easier to go from one point to another in the MDS space than the Euclidean distance only. This vector reflects the asymmetric information underlying the data, and is shown within the same low-dimensional MDS space as objects. Although originally these models are built via least squares estimation technique, we consider additive errors to the observed dissimilarities and propose a Bayesian estimation method.

Bayesian Hill-climbing Model

In standard symmetric MDS, the most popular choice of the distance function is the Euclidean distance. That is, the distance between point $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined by

$$d_{ij}(\mathbf{X}) = \sqrt{\sum_{l=1}^{p} \left(x_{il} - x_{jl}\right)^2} = ||\mathbf{x}_i - \mathbf{x}_j||, \qquad (1)$$

where x_{il} is the coordinate of stimulus i $(i = 1, \ldots, n)$ on dimension l, and p is the dimensionality of the space. In contrast, the distance measure in the hill-climbing model is given by

$$d_{ij}(\mathbf{X}, \mathbf{z}) = ||\mathbf{x}_i - \mathbf{x}_j|| + \frac{\left(\mathbf{x}_i - \mathbf{x}_j\right)' \mathbf{z}}{||\mathbf{x}_i - \mathbf{x}_j||}, \qquad (2)$$

where $\mathbf{z} = (z_1, \ldots, z_p)'$ is the p-dimensional slope vector. In our probabilistic model, the observed dissimilarity $\mathbf{\Delta} = \{\delta_{ij}\}$ between object i and j is considered to be error-perturbed. We consider standard additive errors, which are given by

$$\delta_{ij} = d_{ij} + e_{ij}, \\ e_{ij} \sim N(0, \sigma^2), \quad (3)$$

independently for $i, j = 1, \dots, n$.

For Bayesian analysis of the model described in the previous section, we need to choose prior distributions for the unknown parameters. For the prior distribution of $\mathbf{X} = \{x_{il}\}$, a normal distribution is used,

$$x_{il} \sim N(0, \lambda), \quad (4)$$

independently for $i = 1, \dots, n$ objects and $l = 1, \dots, p$ dimensions. For the prior distribution of σ^2 in (3), conjugate inverse gamma distribution is used,

$$\sigma^2 \sim IG(\alpha_\sigma, \beta_\sigma). \quad (5)$$

For the hyperprior of variance λ, also an inverse gamma hyperprior is considered,

$$\lambda \sim IG(\alpha_\lambda, \beta_\lambda). \quad (6)$$

This inverse gamma is a natural conjugate distribution of the normal variance parameter. The above specifications are predominately based on Oh & Raftery (2001)'s Bayesian MDS.

For the prior distribution of the slope parameter, we consider the normal distribution,

$$z_l \sim IG(0, \phi^2), \quad (7)$$

independently for $l = 1, \dots, p$ dimensions. As a general practice, prior independence among parameters is assumed.

Bayesian analysis with noninformative priors is very common when little or no prior information is available. In this case, all the hyperparameters of inverse-gamma priors $(\alpha_\sigma, \beta_\sigma, \alpha_\lambda, \beta_\lambda)$ are set to be .001 and the variance hyperparameter ϕ^2 is set to be 10^6.

Based on the above-mentioned model and priors, the posterior distribution takes the form

$$\pi(\boldsymbol{\theta}|\boldsymbol{\Delta}) \propto L(\boldsymbol{\theta}|\boldsymbol{\Delta})\pi(\boldsymbol{\theta}), \quad (8)$$

where $\boldsymbol{\theta}$ denotes unknown parameters, $\pi(\boldsymbol{\theta})$ denotes the product of prior distributions in (4)–(7), and $L(\boldsymbol{\theta}|\boldsymbol{\Delta})$ denotes the likelihood function. An MCMC algorithm is employed to generate samples from the posterior distribution. The BUGS program (Lunn, Spiegelhalter, Thomas, & Best, 2009) uses built-in MCMC techniques to obtain samples from posterior distribution. To achieve the equilibrium distribution approximately, the MCMC process should be first run for a sufficient 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 posterior distribution.

Bayesian Jet-stream Model

The essential difference of the jet-stream model from the hill-climbing model is that the asymmetry factor appears in the jet-stream model in the denominator, whereas in the hill-climbing model it turns up as a separate term. We therefore describe only the differences between the two models.

In the jet-stream model, the distance measures are defined (instead of equation (2)) by

$$d_{ij}(\mathbf{X}, \mathbf{z}) = \frac{||\mathbf{x}_i - \mathbf{x}_j||}{1 + \frac{(\mathbf{x}_i - \mathbf{x}_j)'\mathbf{z}}{||\mathbf{x}_i - \mathbf{x}_j||}}, \quad (9)$$

where $\mathbf{z}$ is the p-dimensional jet-stream vector. In order to remove the scale indeterminacy associated with this model, we set different prior distribution for $\mathbf{z}$. Specifically, considering the two-dimensional case in particular, we use (instead of equation (7))

$$z_1 \sim U(0,1), \\ z_2 = \sqrt{1 - z_1^2}. \quad (10)$$

Other specifications of the model and the priors remain the same as in the case of hill-climbing model.

Simulation Study

The result of a small Monte Carlo simulation study is shown to demonstrate the parameter recovery of the proposed method when the model is correctly specified. In this numerical experiment, the true configuration matrix $\mathbf{X}$ is set as

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 1 & 0 \\ -1 & 0 \\ 1 & -1 \\ -1 & -1 \end{pmatrix}. \tag{11}$$

The slope vector or the stream vector, $\mathbf{z}$, is set as

$$\mathbf{z} = \frac{2}{3}\left(\cos\left(\frac{\pi}{3}\right), \sin\left(\frac{\pi}{3}\right)\right) \tag{10}$$

Then, the true distance matrix $\mathbf{D}$ is calculated by equations (2) and (9). Then, 100 artificial "observed" datasets are generated from the true distance matrix with additive errors for both models. The additive error is generated from a normal distribution with mean zero and standard deviation 0.3. Then, all error-perturbed datasets are analyzed with the proposed methods.

For selecting hyperparameters, we utilized standard non-informative prior settings. The number of dimensions is fixed to two. For the initial values of $\mathbf{X}$, we used the classical MDS solution of the symmetricaized data. For the initial values of other parameters, (a vector of) 1 is used. The BUGS software is used for calculating the posterior distributions. The first 5,000 MCMC samples are discarded as burn-in, and the subsequent 15,000 MCMC samples are utilized for posterior estimation. Because our Bayesian model basically consists of conjugate priors, the convergence of Bayesian MDS parameter estimation is fast, as reported in Oh & Raftery (2001). We also observed rapid convergence for both asymmetric and symmetric datasets.

Table 1 shows the true values, the means of estimated values, and 95% credibility intervals of the distance measure matrix $\mathbf{D}$ for 100 artificial datasets. From the table, it can be said that parameter recovery using the proposed Bayesian asymmetric MDS is quite accurate for both models.

Conclusion

In this study, we proposed a Bayesian asymmetric MDS approach by extending the hill-climbing model and the jet-stream model. The models incorporate both asymmetric and symmetric parts of the data and express the data structure without introducing many additional model parameters. The parameter recovery of the proposed methods is confirmed.

Author Note

Part of the results presented here has been submitted to a refereed journal for consideration.

REFERENCES

Borg, I., & Groenen, P. J. F. (2005). *Modern Multidimensional Scaling: Theory and Applications* (2 ed.): Springer, New York.

Gower, J. C. (1977). The analysis of asymmetry and orthogonality. In P. R. Krishnaiah & L. N. Kanal (Eds.), *Handbook of Statistics* (Vol. 2, pp. 109-123). Amsterdam: North Holland.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project: Evolution, critique and future directions (with discussion), *Statistics in Medicine*, 28, 3049-3082.

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

Table 1: True values (first row), mean of the posterior estimates (second row), and the mean of the 95% credibility intervals for each element of D for (a) Bayesian hill-climbing model and (b) Bayesian jet-stream model

(a)Bayesian hill-climbing model

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
		2.333	1.577	2.792	2.577	3.472
[1,]	-	2.346	1.575	2.783	2.556	3.464
		(2.001, 2.684)	(1.24, 1.901)	(2.477, 3.081)	(2.233, 2.87)	(3.175, 3.744)
	1.667		2.196	1.577	3.001	2.577
[2,]	1.669	-	2.164	1.581	2.974	2.568
	(1.319, 2.024)		(1.832, 2.486)	(1.226, 1.922)	(2.663, 3.271)	(2.245, 2.884)
	0.423	2.276		2.333	1.577	2.792
[3,]	0.413	2.256	-	2.314	1.592	2.802
	(0.069, 0.763)	(1.913, 2.585)		(1.989, 2.628)	(1.235, 1.934)	(2.495, 3.101)
	1.680	0.423	1.667		2.196	1.577
[4,]	1.644	0.451	1.621	-	2.166	1.602
	(1.31, 1.979)	(0.121, 0.785)	(1.287, 1.952)		(1.836, 2.487)	(1.272, 1.926)
	1.423	2.656	0.423	2.276		2.333
[5,]	1.372	2.630	0.457	2.252	-	2.352
	(1.028, 1.706)	(2.309, 2.934)	(0.129, 0.79)	(1.908, 2.583)		(2.005, 2.695)
	2.184	1.423	1.680	0.423	1.667	
[6,]	2.143	1.382	1.663	0.433	1.670	-
	(1.816, 2.46)	(1.039, 1.717)	(1.327, 1.998)	(0.093, 0.779)	(1.32, 2.028)	

(b) Bayesian jet-stream model

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
		1.500	0.634	1.437	1.268	1.721
[1,]	-	1.562	0.469	1.132	0.936	1.237
		(1.055, 2.095)	(0.211, 0.862)	(0.798, 1.548)	(0.601, 1.376)	(0.989, 1.549)
	3.000		2.329	0.634	2.412	1.268
[2,]	3.019	-	2.336	0.612	2.482	0.994
	(2.447, 3.611)		(1.746, 2.864)	(0.285, 1.041)	(1.931, 2.954)	(0.66, 1.414)
	2.366	2.150		1.500	0.634	1.437
[3,]	2.387	2.012	-	1.216	0.602	1.182
	(1.647, 3.161)	(1.424, 2.588)		(0.751, 1.734)	(0.278, 1.04)	(0.837, 1.578)
	5.040	2.366	3.000		2.329	0.634
[4,]	4.902	2.409	2.994	-	2.313	0.539
	(4.09, 5.75)	(1.669, 3.154)	(2.356, 3.608)		(1.714, 2.828)	(0.29, 0.931)
	4.732	3.418	2.366	2.150		1.500
[5,]	4.653	3.249	2.481	2.003	-	1.562
	(3.858, 5.448)	(2.574, 3.904)	(1.733, 3.237)	(1.406, 2.572)		(1.05, 2.068)
	7.944	4.732	5.040	2.366	3.000	
[6,]	7.751	4.637	4.899	2.439	3.066	-
	(6.912, 8.596)	(3.851, 5.427)	(4.1, 5.734)	(1.706, 3.204)	(2.463, 3.666)	