A Bayesian Approach to Dependent Paired Comparison Rankings

Hea-Jung Kim* Dae-Hwang Kim†

Abstract

In this paper we develop a method for finding optimal ordering of K statistical models. This is based on a dependent paired comparison experimental arrangement whose results can naturally be represented by a completely oriented graph (also so called tournament graph). Introducing preference probabilities, strong transitivity conditions, and an optimal criterion to the graph, we show that a Hamiltonian path obtained from row sum ranking is the optimal ordering. Necessary theories involved in the method and computation are provided. As an application of the method, generalized variances of K multivariate normal populations are compared by a Bayesian approach.

Key Words : Oriented Graph; Hamiltonian path; paired comparison ranking; strong stochastic transitivity; generalized variances.

1 Introduction

The testing of a set of items for preference on overall suitability or on specific characteristics often requires of the respondent the ability to make very fine sensory discriminations based on complex physiological processes. To remove some of the confusion associated with simultaneously comparing several objects, the method of paired comparisons has been widely employed (See Bauer 1997 and references therein). In paired comparison experimentation, expression of preference are obtained for all possible pairs of objects in the set of interest. The data from experiments are used, among other things, to rank the objects in order of preference.

*Professor, Dongguk University, Seoul 100-715, Korea (KOSEF:R01-2001-000-00010-0)

†Graduate Student, Department of Statistics, Dongguk University, Seoul 100-715, Korea

The present paper undertakes a graphical approach to the analysis of data arising from paired comparison rankings. A graphical model is considered : A completely oriented graph (Skiena 1990, p. 175), i.e., a graph which every pair of nodes is connected by a single uniquely directed edge. The objectives are (i) to introduce preference probabilities of the model constructed by a dependent paired comparisons, and (ii) to provide a criterion for ranking of the objects in the set of interest. As an application, this paper focuses on considering how inferences might be made about a possible ordering in the generalized variance of K multivariate normal populations, $\Pi_1 \sim N_p(\theta_1, \Sigma_1), \ldots, \Pi_K \sim N_p(\theta_K, \Sigma_K)$. To this end we pursued a Bayesian approach to calculate the preference probabilities under dependent paired comparisons.

2 The Graphical Model

2.1 Paired Comparison Ranking

Consider a finite oriented graphs where ith vertex represents population (or statistical model) Π_i, $i = 1, \ldots, K$. Then the graph $G(X, R)$ consists of a set X say $X = \{1, 2, \ldots, K\}$, of vertices and a collection of arcs R where arc is of form $x \rightarrow y$ with x and y in X. A rank order of the K populations is an arrangement $(p_1, p_2, \ldots, p_K) = P$ of objects X. For a given preference graph G and an order P, let $v(P)$ denote the number of violations of the observed preference, that is the number of arcs $p_i \rightarrow p_j$ in R such that p_j precedes p_i in P.

Under the finite oriented graph, one can formulate probability model for dependent paired comparisons. Suppose that K objects are to be tested in pairs with data being obtained to the pair (i, j). Let $\Theta = \{\theta_{ij}\}$ $(i, j = 1, \ldots, K)$, where $\theta_{ij} = Pr(p_i \rightarrow p_j)$ is the probability of preference for object i over j, $\theta_{ij} + \theta_{ji} = 1$, and for convenience $\theta_{ii} = 1/2$. Then a problem arises in defining a ranking of objects based on Θ. This problem can be resolved by requiring Θ to satisfy a transitivity constraint. For a dependent paired comparisons following two conditions are satisfied:

weak stochastic transitivity (C_1) : for every triad of objects (i, j, k),

$$\theta_{ij} \geq 1/2 \text{ and } \theta_{jk} \geq 1/2 \text{ imply } \theta_{ik} \geq 1/2; \tag{1}$$

strong stochastic transitivity (C_2) : for every triad of objects (i, j, k),

$$\theta_{ij} \geq 1/2 \text{ and } \theta_{jk} \geq 1/2 \text{ imply } \theta_{ik} \geq \max(\theta_{ij}, \theta_{jk}). \tag{2}$$

An equivalent form of condition C_2 is the following: for each pair (i, j),

$$\theta_{ij} \geq 1/2 \text{ implies } \theta_{ik} \geq \theta_{jk} \quad k = 1, \ldots, K. \tag{3}$$

Condition (2) follows directly from (3), while an examination of cases establishes that (3) follows from (2).

The stronger condition C_2 is not necessary warranted in all situation (David, 1987). Conditions C_1 and C_2 each lead to a well defined ranking of every triad and hence to a true ranking of the set of K objects.

The so-called preference matrix representation of the graph is useful for determining $v(P)$ under the conditions of C_1 or C_2. For given P, the preference matrix contains a plus in its i, jth place if $\theta_{ij} > 1/2$ (i.e. if $p_i \rightarrow p_j$) a minus if $\theta_{ji} > 1/2$ (i.e. if $p_j \rightarrow p_i$) and a dot if $\theta_{ij} = 1/2$. Suppose, under C_2, the preference probabilities (θ_{ij}'s) yield the matrix for an ordering $P = (1, 2, \ldots, 6)$:

$$\begin{pmatrix} \cdot & + & + & - & + & + \\ - & \cdot & - & + & + & - \\ - & + & \cdot & + & + & + \\ + & - & - & \cdot & - & - \\ - & - & - & + & \cdot & + \\ - & + & - & + & - & \cdot \end{pmatrix}.$$

In this example there are five violations of the ranking $(1, 2, \ldots, 6)$ shown by the minus signs above the main diagonal. That is $v(1, 2 \ldots, 6) = 5$; other rankings yield other v's. When a dependent paired comparisons is experimented an optimal ranking P of K objects achieves

$$\min_{p} v(P) = 0$$

and we call any order P obtaining this minimum is an 'optimal adjoining ordering'.

The mathematical literature treats problem related to ours using the language of round robin tournaments, see Harary, Norman, and Cartwright (1965). For most tournaments it will be possible to order the players in a Hamiltonian path, i.e. so that each defeats his successor. In fact it is a theorem that for a complete tournament there will be at least one Hamiltonian path. Harary *et al.* see two serious difficulties with the Hamiltonian ordering of the players. First, there is no necessary relation, in general, between such a ranking of players and their scores. Secondly, a tournament may have more than one Hamiltonian path, so that several different rankings may be possible. The optimal adjoining ordering on the other hand is entirely consistent with Hamiltonian ordering and in fact provides a means for solving the two difficulties.

Theorem 1. *For a complete tournament every optimal adjoining ordering is a Hamiltonian path.*

2.2 Computational Aspects

The practical matter of determining the optimal ordering is not easy. A direct algorithm would be to enumerate all $K!$ possible orders, score v in each, and find the the ordering with zero score. Alternatively, one may obtain all the Hamiltonian pathes from the graph model with K vertices, and then score v in each path to find the optimal ordering (see, Adleman 1994; Fu, Beigel, and Zhou 2003 for the algorithms for Hamiltonian path).

The optimal ordering becomes easy when we use the strong stochastic transitivity condition C_2 in ordering K objects. Suppose that K statistical models (or objects) $\Pi_1, \ldots, \Pi_K$ are compared in all possible $K(K-1)/2$ different pairings. The simplest procedure is the Kendall-Wei method that rank according to the vector w of row-sum scores (see David 1987),

$$w = \Theta 1, \tag{4}$$

where 1 is the column vector of K ones and Θ is the matrix $\{\theta_{ij}\}$ of probabilities of preference.

3 An Application

The generalized variance can be used to rank distinct groups and populations in order of their dispersion or spread (Rencher 2002). However, due to complex sampling distribution involved in inferencing the generalized variance, the analysis of it is yet to be seen in applied settings.

3.1 The Posterior Distribution

Suppose $X_1(i), \ldots, X_{N_i}(i)$ are independent p-variate observations from $N_p(\mu_i, \Lambda_i^{-1})$, $i = 1, \ldots, K$, where $\Lambda_i = \Sigma_i^{-1}$, the precision matrix. Let

$$\bar{X}(i) = \sum_{\ell=1}^{N_i} X_\ell(i)/N_i, \text{and } V_i = \sum_{\ell=1}^{N_i} (X_\ell(i) - \bar{X}(i))(X_\ell(i) - \bar{X}(i))'.$$

To assure very little information is contributed to the analysis by a subjective prior density, we assume diffuse prior

$$p(\cap_{k=1}^{i} \mu_i, \cap_{i=1}^{K} \Lambda_i) \propto \prod_{i=1}^{K} |\Lambda_i|^{-(p+1)/2}. \tag{5}$$

Integrating joint posterior density with respect μ_i's, we have the marginal posterior distribution of Λ_i that is independent of Λ_k, $k = 1, \ldots, K; i \neq k$:

$$\Lambda_i | \bar{X}(i), V_i \sim W_p(V_i^{-1}, N_i - 1), \quad N_i \geq p + 1, \tag{6}$$

a Wishart distribution with scale parameter V_i^{-1} and $N_i - 1$ degrees of freedom, $i = 1, \ldots, K$,.

Theorem 2. *Let* $S_i = V_i^{1/2} \Lambda_i V_i^{1/2}, i = 1, \ldots, K$, *and let* $\sum_{i=1}^{k+1} S_i = T_k' T_k$, $k = 1, \ldots, K-1$, *where* T_k *is an upper triangular* $p \times p$ *matrix. Then* $R_k = T_k'^{-1} \left(\sum_{i=1}^{k} S_i \right) T_k^{-1}$ *are independent with* $R_k \sim B_I(p; \sum_{i=1}^{k} n_i/2, n_{k+1}/2)$, $k = 1, \ldots, K-1$. *For each* k, *the joint posterior distribution of eigen values,* $\lambda_1^k, \ldots, \lambda_p^k$ *of* R_k, $p(\lambda^k | Data)$ *is*

$$c_k \prod_{\ell}^{p} \left[\left(\lambda_\ell^k \right)^{\sum_{i=1}^{k} (n_i - p - 1)/2} \left(1 - \lambda_\ell^k \right)^{(n_{k+1} - p - 1)/2} \right] \prod_{u<v}^{p} \left(\lambda_u^k - \lambda_v^k \right) I(\lambda^k), \tag{7}$$

where $\lambda^k = (\lambda_1^k, \ldots, \lambda_p^k)'$, $I(\lambda^k) = I(1 > \lambda_1^k > \cdots > \lambda_p^k > 0)$, *an indicator function,* $c_k = \pi^{p^2/2} / [\Gamma_p\{p/2\} B_p(\sum_{i=1}^{k} n_i/2, n_{k+1}/2)]$, *and* $n_k = N_k - 1$.

3.2 Posterior Probability of $|\Sigma_i| < |\Sigma_j|$

The distribution (7) enables us to obtain various integral-type posterior quantities of $|\Sigma_i|/|\Sigma_j|$'s for $i, j = 1, \ldots, K; i \neq j$. Especially the posterior probability of a set $\{|\Sigma_i|/|\Sigma_j| \in A\}$ is obtained from (7) if we set $1\{|\Sigma_i|/|\Sigma_j| \in A\}$, where $1\{\cdot\}$ denotes the indicator function. When we take $A = \{|\Sigma_i|/|\Sigma_j|; |\Sigma_i|/|\Sigma_j| < 1\}$ in the indicator function, (7) gives the posterior probability of $|\Sigma_i| < |\Sigma_j|$, i.e.,

$$\begin{aligned} \theta_{ij} &= p(|\Sigma_i| < |\Sigma_j| | Data) \\ &= \int_{I(\lambda^{K-1})} \cdots \int_{I(\lambda^1)} \eta_{ij}^*(\lambda^1, \ldots, \lambda^{K-1}) \prod_{k=1}^{K-1} p(\lambda^k | Data) \, \partial\lambda^1 \cdots \partial\lambda^{K-1}, \end{aligned} \tag{8}$$

where

$$\eta_{ij}^*(\lambda^1, \ldots, \lambda^{K-1}) = 1 \left\{ \frac{|V_i|}{|V_j|} \frac{\prod_{u=1}^{p} (1 - \lambda_u^{j-1}) \prod_{k=1}^{i-1} \left(\prod_{u=1}^{p} \lambda_u^k \right)}{\prod_{u=1}^{p} \left(1 - \lambda_u^{i-1} \right) \prod_{i=1}^{j-1} \left(\prod_{u=1}^{p} \lambda_u^k \right)} < 1 \right\}.$$

An analytic evaluation of the probability is not available because the posterior distribution $\prod_{i=1}^{K-1} p(\lambda^i | Data)$ in (7) is complicate. Fortunately, however, the weighted Monte Carlo method (cf. Chen and Shao 1999) whose variant is applicable for the computation of the probability.

3.3 A Weighted Monte Carlo Method

Suppose that $\{\lambda_{(t)}^k, t = 1, \ldots, m; k = 1, \ldots K-1\}$ is a simulated sample from an appropriate importance function $g(\lambda) = \prod_{k=1}^{K-1} g_k(\lambda^k)$. Write the importance sampling weight as

$$w_{(t)} = \prod_{k=1}^{K-1} \{p(\lambda_{(t)}^k | Data)/g(\lambda_{(t)}^k)\}. \tag{9}$$

Then, the weighted Monte Carlo(MC) estimator of the posterior probability (9) is given by

$$\hat{\theta}_{ij} = \sum_{t=1}^{m} w_{(t)} \eta_{ij}^*(\lambda_{(t)}^1, \ldots, \lambda_{(t)}^{K-1}) / \sum_{\ell=1}^{m} w_{(\ell)}, \tag{10}$$

where $\hat{\theta}_{ij} = \hat{p}(|\Sigma_i| < |\Sigma_j||Data)$.

Notice that (10) gives a Monte Carlo estimate of the posterior probability using iid samples from the importance distribution in such a way that $\prod_{k=1}^{K-1} p(\lambda^k|Data)$ and $\prod_{k=1}^{K-1} g_k(\lambda^k)$ need only be known up to a constant of proportionality. Geweke (1989) showed that

$$\hat{\theta}_{ij} \overset{a.s.}{\to} \theta_{ij}, \quad i, j = 1, \ldots, K, \tag{11}$$

as $m \to \infty$. Further the law of large numbers implies that

$$m^{1/2}(\hat{\theta}_{ij} - \theta_{ij}) \to N(0, \sigma^2),$$

where $\sigma^2 = \sigma_V^2/S^2$ and it can be estimated by $\hat{\sigma}^2 = \hat{\sigma}_V^2/\hat{S}^2$ with $\hat{S} = 1/m \sum_{t=1}^{m} w_{(t)}$ and $\hat{\sigma}_V^2 = 1/m \sum_{t=1}^{m} (w_{(t)} \eta_{ij}^*(\lambda_{(t)}^1, \ldots, \lambda_{(t)}^{K-1}) - w_{(t)}\hat{\theta}_{ij})^2$.

As one would expect, (10) will converge faster and generally be better behaved the closer the importance function is to the joint posterior distribution $\prod_{k=1}^{K-1} p(\lambda^k|Data)$. For this reason, choosing a good candidate is the main issue in applications of the importance sampling. Three choices of important distributions are considered.

References

Adelman, L. M. (1994). Molecular computation of solutions. *Science* **266**, 1021-1024.

Bauer, P.(1997). A note on multiple testing procedure in dose finding. *Biometric* **53**, 1125-1128.

Chen, M.-H, and Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of the Computational and Graphical Statistics* **8**, 69-92.

David, H. A. (1987). Ranking from unbalanced paired-comparison data. *Biometrika* **74**, 432-436.

Fu, B., Beigel, R., and Zhou, F. (2003). An $O(2^n)$ volume molecular algorithm for Hamiltonian path. www.cis.temple.edu/$\sim$ beigel/papers.

Geweke, J.(1989). Bayesian inference in econometrics models using Monte Carlo integration. *Econometrica* **57**, 1371-1340.

Harary, F., Norman, R. Z., and Cartwright, D. (1965). *Structural Models: An Introduction to the Theory of Directed Graphs.* New York: John Wiley & Sons.

Rencher, A. C. (2002).*Methods of Multivariate Analysis.* New York: Wiley.

Skiena, S. (1990). *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica.* Reading, MA: Addison-Wesley, 1990.