

CIRJE-F-15

Rankings Generated by Spherical Discriminant Analysis

Hidehiko Kamiya
The Institute of Statistical Mathematics

Akimichi Takemura
Faculty of Economics, The University of Tokyo

August 1998

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

Rankings generated by spherical discriminant analysis

Hidehiko Kamiya

The Institute of Statistical Mathematics

and

Akimichi Takemura

Faculty of Economics, University of Tokyo

August, 1998

Abstract

We consider the same problem as in Kamiya and Takemura (1997), but for discriminant analysis on $(n-1)$ -dimensional unit sphere S^{n-1} . That is, we regard pairwise discriminant analysis of m populations on S^{n-1} as a process to generate rankings among the populations, and give a formula for the number of generated rankings.

Key words: ranking, discriminant analysis, arrangement of hyperplanes, Poincaré polynomial.

1 Introduction

In a recent paper (Kamiya and Takemura(1997), abbreviated as KT throughout the paper), we investigated properties of rankings associated with linear discriminant analysis among m populations in R^n . In that paper, we regarded multiple linear discriminant analysis as a process to generate rankings among the populations, and gave formulae for the number of generated rankings, and moreover, basic characterizations of non-generated rankings.

In the present paper, we consider the same problem for discriminant analysis on $(n-1)$ -dimensional unit sphere $S^{n-1} = \{\mathbf{x} \in R^n : \|\mathbf{x}\| = 1\}$. Suppose we are given m populations on S^{n-1} . For each pair of the populations, we take an $(n-2)$ -dimensional discriminant boundary. Then S^{n-1} is divided into regions by $m(m-1)/2$ such boundaries, and each region is indexed by a ranking among the m populations according to the proximity to them. In this way, we can regard pairwise discriminant analysis on S^{n-1} as a process to generate rankings among the populations.

Let the m populations be the von Mises-Fisher distributions $M_n(\boldsymbol{\mu}_i, \kappa)$, $i = 1, \dots, m$, with common concentration parameter $\kappa > 0$:

$$\mathbf{x} = (x_1, x_2, \dots, x_n)' \sim c_n(\kappa) e^{\kappa \boldsymbol{\mu}_i' \mathbf{x}} dS_{n-1},$$

where $c_n(\kappa)$ is the normalizing constant, $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i}, \dots, \mu_{ni})' \in S^{n-1}$, $i = 1, \dots, m$, and dS_{n-1} is the volume element on S^{n-1} (Mardia, Kent, and Bibby(1979)).

Assuming equal prior probability for the m populations, we follow the following pairwise allocation rule: Assign a subject $\mathbf{x} = (x_1, x_2, \dots, x_n)' \in S^{n-1}$ to population i in preference to population j iff

$$c_n(\kappa)e^{\kappa\boldsymbol{\mu}'_i\mathbf{x}} > c_n(\kappa)e^{\kappa\boldsymbol{\mu}'_j\mathbf{x}}. \quad (1)$$

This rule is equivalent to the allocation according to the geodesic distance d on S^{n-1} : $d(\mathbf{x}, \mathbf{y}) = \cos^{-1}(\langle \mathbf{x}, \mathbf{y} \rangle)$, $\mathbf{x} = (x_1, \dots, x_n)'$, $\mathbf{y} = (y_1, \dots, y_n)' \in S^{n-1}$, $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$. Specifically, (1) is equivalent to

$$d(\mathbf{x}, \boldsymbol{\mu}_i) < d(\mathbf{x}, \boldsymbol{\mu}_j). \quad (2)$$

Thus, the discriminant boundary B_{ij} between populations i and j is the set of points on S^{n-1} equidistant from $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ in terms of d :

$$B_{ij} = \{\mathbf{x} \in S^{n-1} : d(\mathbf{x}, \boldsymbol{\mu}_i) = d(\mathbf{x}, \boldsymbol{\mu}_j)\}.$$

Now, S^{n-1} is divided into regions by $m(m-1)/2$ such boundaries B_{ij} , $1 \leq i < j \leq m$, and each region is indexed by a ranking of $\{1, 2, \dots, m\}$ in which i is ranked better than j iff (2) holds with an arbitrary point \mathbf{x} in the region.

However, as in the case of discrimination in R^n , when n is small compared with m , not all the $m!$ rankings are generated. Here we ask the same question as in the case of R^n : How many rankings are generated from the discrimination on S^{n-1} ? Actually, our regions on S^{n-1} correspond to the unbounded regions of the linear discriminant analysis in R^n treated in our previous paper KT. Hence, as would be clear for the readers of KT, in the case of discrimination on S^n , we have that for each pair of mutually reverse rankings of $\{1, 2, \dots, m\}$, both of them or neither of them is generated. That is, a ranking is generated if and only if its reverse ranking is generated.

The organization of this paper is as follows. After explaining background concepts in Section 2, we give the main result of this paper in Section 3. In the final section, we state several results of independent interest.

2 Background concepts and definitions

In this section, we introduce background concepts from rankings and hyperplane arrangements. We first explain some concepts concerning rankings.

A ranking of m items $\{1, 2, \dots, m\}$ can be expressed as an ordering of them. The ordering $\sigma = (i_1, i_2, \dots, i_m)$ corresponds to the ranking in which item i_1 is ranked first, item i_2 is ranked second, and so on. For an ordering $\sigma = (i_1, \dots, i_m)$, its reverse ordering is defined to be $-\sigma = (i_m, \dots, i_1)$. A partial ordering π corresponds to a partial ranking, in which ties are allowed. Here we follow the convention that the order of items in braces is irrelevant; in parentheses it is relevant. So the partial ordering $\pi = (\{2, 4\}, 3, 1)$, for example, corresponds to the partial ranking in which items 2 and 4 are ranked first, item

3 is ranked third, and item 1 is ranked last. We attach the adjective “full” to ranking (ordering) when we want to emphasize the distinction from a partial ranking (ordering).

Next, we give a brief introduction to the theory of hyperplane arrangements. For a full treatment of the theory, the reader is referred to Chapters 1 and 2 of Orlik and Terao (1992).

A hyperplane arrangement \mathcal{A} is defined to be a finite set of hyperplanes in $V = R^n$. The problem of counting chambers, i.e., regions, of a hyperplane arrangement becomes much harder when degeneracy is allowed. Zaslavsky (1975) gave a formula for the number of regions in an arbitrary arrangement \mathcal{A} of hyperplanes. He introduced the method of deletion and restriction to obtain a recursion formula for chamber counting problems. By proving that the Poincaré polynomial evaluated at 1, $\pi(\mathcal{A}, 1)$, satisfies the same recursion, he obtained a beautiful result: The number of regions is equal to $\pi(\mathcal{A}, 1)$.

We list some basic definitions from the theory of hyperplane arrangements. A hyperplane arrangement \mathcal{A} is said to be centered if $\bigcap_{H \in \mathcal{A}} H \neq \emptyset$. In particular, it is said to be central if each hyperplane contains the origin. The intersection poset (partially ordered set) $L = L(\mathcal{A})$ is defined to be the set of nonempty intersections of elements of \mathcal{A} endowed with the partial order defined by

$$X \leq Y \iff Y \subseteq X.$$

The rank function on L is defined by the codimension: $r(X) = n - \dim(X)$, where $\dim(X)$ is the dimension of X . Maximal elements of $L(\mathcal{A})$ have the same rank, and the rank $r(\mathcal{A})$ of \mathcal{A} is defined to be the rank of a maximal element of $L(\mathcal{A})$. Let $L_p = L_p(\mathcal{A}) = \{X \in L(\mathcal{A}) \mid r(X) = p\}$. The Hasse diagram of L is, by definition, the diagram which has vertices labeled by the elements of L and arranged on levels L_p for $p \geq 0$, and whose edges connect $X \in L_p$ with $Y \in L_{p+1}$ iff $X < Y$. Define the Möbius function ν as follows:

$$\begin{cases} \nu(V) = 1, \\ \nu(X) = -\sum_{V \leq Y < X} \nu(Y) \quad \text{if } V < X. \end{cases}$$

The Poincaré polynomial of \mathcal{A} is defined by

$$\pi(\mathcal{A}, t) = \sum_{X \in L} \nu(X)(-t)^{r(X)},$$

where t is an indeterminate. A chamber is defined to be a connected component of $\{\mathbf{x} \in R^n : \mathbf{x} \notin \bigcup_{H \in \mathcal{A}} H\}$. However, in this paper we use the term “region” instead of “chamber.”

Now we specialize to the arrangement of discriminant hyperplanes in the pairwise linear discriminant analysis among m populations $N_n(\boldsymbol{\mu}_i, I)$, $i = 1, \dots, m$, in R^n .

Denote by H_{ij} , $1 \leq j < i \leq m$, the discriminant hyperplane between populations i and j :

$$H_{ij} = \{\mathbf{x} \in R^n : (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'(\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)) = 0\},$$

and consider the arrangement of discriminant hyperplanes

$$\mathcal{A} = \{H_{ij} : 1 \leq j < i \leq m\}.$$

Each element of $L = L(\mathcal{A})$ can be indexed by a partition J of m indices into blocks, and $X \leq Y$ for $X, Y \in L$ corresponds to the fact that the corresponding partition of X is a refinement of that of Y , so that each block of the latter is a union of some blocks of the former. Specifically, to $X \in L$ corresponds the partition of $\{1, 2, \dots, m\}$ into equivalence classes under the equivalence relation \sim_X defined by

$$i \sim_X j \iff X \subseteq H_{ij},$$

where we agree that $H_{ii} = V$ and $H_{ji} = H_{ij}$ for $i > j$. Note that $i \sim_X j$ means that $\mathbf{x} \in X$ is equidistant from $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$.

The m populations are ranked according to the Euclidean distances to $\boldsymbol{\mu}_i$. The population i with the nearest $\boldsymbol{\mu}_i$ is ranked first; j with the farthest $\boldsymbol{\mu}_j$ last. Thus, each region is indexed by a full ordering. Note that a region is open in R^n . Elements of L of rank n , if they exist, are called terminal nodes, and can be indexed by partial orderings.

Regions fall into two types: bounded regions and unbounded ones. This distinction played an important role in the characterization of non-generated rankings in KT (Section 3 of KT). Also, there is a connection between unbounded regions and ideal vectors. See Section 4.3 of KT. In general, ideal vector model is defined as follows: Objects or items $1, 2, \dots, m$ are represented as points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ in R^n for some n , and the ‘‘ideal vector’’ \mathbf{d} is supposed to exist. Then the m objects are ranked according to the projections onto this direction. Specifically, i is ranked better than j iff $\langle \mathbf{d}, \mathbf{x}_i \rangle > \langle \mathbf{d}, \mathbf{x}_j \rangle$.

3 The main result

In this section, we give a formula for the number of regions in pairwise discriminant analysis of m populations $M_n(\boldsymbol{\mu}_i, \kappa)$, $i = 1, \dots, m$, $\kappa > 0$, on S^{n-1} under the rule (1).

We begin by defining non-degeneracy of our discriminant analysis on S^{n-1} .

3.1 Definition of non-degeneracy

We say that our discriminant analysis is non-degenerate if the following condition holds:

(A) The points $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m \in S^{n-1}$ regarded as points in R^n are in general position (in the sense of Remark 3.1 below).

Remark 3.1

When $m \leq n + 1$, we say that m points $\mathbf{x}_1, \dots, \mathbf{x}_m \in R^n$ are in general position iff for any set of scalars a_1, \dots, a_m satisfying $\sum_i a_i = 0$, equation $\sum_i a_i \mathbf{x}_i = \mathbf{0}$ implies $a_i = 0$ for all $i = 1, \dots, m$.

When $m > n + 1$, we say that $\mathbf{x}_1, \dots, \mathbf{x}_m \in R^n$ are in general position iff any $n + 1$ of them are in general position.

Now we move on to the problem of finding the number of regions. We solve this problem by identifying our discrimination on S^{n-1} with linear discriminant analysis in R^n .

3.2 Identification with discrimination in R^n

Regarding $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_m \in S^{n-1}$ as points in R^n , we see that (2) is equivalent to the following rule by means of hyperplanes in R^n :

$$(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)' \mathbf{x} > 0. \quad (3)$$

Denote the hyperplanes in (3) by H_{ij} :

$$H_{ij} = \{\mathbf{x} \in R^n : (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)' \mathbf{x} = 0\}, \quad 1 \leq j < i \leq m. \quad (4)$$

Since each H_{ij} , $1 \leq j < i \leq m$, contains the origin, we see that the arrangement \mathcal{A} of these hyperplanes in R^n :

$$\mathcal{A} = \{H_{ij} : 1 \leq j < i \leq m\} \quad (5)$$

is central. Accordingly, the closure of each region of \mathcal{A} is a polyhedral cone. Hence, each of the rankings corresponding to these regions in R^n appears as a region on S^{n-1} when S^{n-1} is embedded in R^n (see Lemma 3.2 of KT). Note that (2) is equivalent to

$$\|\mathbf{x} - \boldsymbol{\mu}_i\| < \|\mathbf{x} - \boldsymbol{\mu}_j\|.$$

Conversely, it is clear that each ranking on S^{n-1} appears as a region in R^n . Therefore, we see that the sets of generated rankings agree for the discrimination in R^n and on S^{n-1} .

The above argument implies that, instead of our original discrimination on S^{n-1} , we may consider the discrimination in R^n by hyperplanes H_{ij} in (4). Thus we can resort to the theory of the arrangements of hyperplanes as in our previous paper KT, but the linear discriminant analysis in R^n of the present paper differs from that of KT: The linear discrimination of the present paper is “degenerate” in the sense of KT, since the arrangement \mathcal{A} in (5) is central.

3.3 The number of regions

We are now in a position to state the main result of this paper.

Theorem 3.1

The number of rankings generated by the non-degenerate discriminant analysis of m populations on S^{n-1} is given by

$$\begin{cases} 2(c_{n-1} + c_{n-3} + \dots + c_0), & n : \text{odd}, \\ 2(c_{n-1} + c_{n-3} + \dots + c_1), & n : \text{even}, \end{cases}$$

where $c_0 = 1$ and

$$c_k = \sum_{1 \leq l_1 < l_2 < \dots < l_k \leq m-1} l_1 l_2 \dots l_k, \quad k \geq 1,$$

in which l_1, l_2, \dots, l_k are natural numbers.

Note that c_k are the unsigned Stirling numbers of the first kind.

Proof. As was stated earlier, we may count the number of regions for the arrangement \mathcal{A} of hyperplanes H_{ij} in (4).

First, we treat the case $n < m - 1$.

Proceeding along the lines of the proof of Lemma 2.2 of KT, consider the partition of $\{1, \dots, m\}$ into k blocks

$$\{\{1, \dots, l_1\}, \{l_1 + 1, \dots, l_1 + l_2\}, \dots, \{m - l_k + 1, \dots, m\}\}$$

with $l_1 \geq l_2 \geq \dots \geq l_{\bar{k}} > l_{\bar{k}+1} = \dots = l_k = 1$, and the corresponding system of linear equations $A\mathbf{x} = \mathbf{0}$, where

$$A = \begin{pmatrix} \mu'_2 - \mu'_1 \\ \vdots \\ \mu'_{l_1} - \mu'_{l_1-1} \\ \mu'_{l_1+2} - \mu'_{l_1+1} \\ \vdots \\ \mu'_{l_1+l_2} - \mu'_{l_1+l_2-1} \\ \vdots \\ \mu'_{l_1+\dots+l_{\bar{k}-1}+2} - \mu'_{l_1+\dots+l_{\bar{k}-1}+1} \\ \vdots \\ \mu'_{l_1+\dots+l_{\bar{k}}} - \mu'_{l_1+\dots+l_{\bar{k}}-1} \end{pmatrix},$$

$\mathbf{x} \in R^n$, and $\mathbf{0} = (0, \dots, 0)' \in R^n$. Here, A is of size $(m-k) \times n$, and by the assumption of non-degeneracy, its rank is $\min\{m-k, n\}$. Hence, the space \mathcal{S} of solutions \mathbf{x} to $A\mathbf{x} = \mathbf{0}$ satisfies

$$\begin{cases} \dim \mathcal{S} = n - (m - k) & \text{when } k \geq m - n, \\ \mathcal{S} = \{\mathbf{0}\} & \text{when } k < m - n. \end{cases}$$

Therefore, $r(\mathcal{A}) = n$, and the Hasse diagram is the same as in the case of non-degenerate discriminant analysis of KT except that the vertices of rank n degenerate to a single vertex with index $\{\{1, \dots, m\}\}$, connected by edges with all vertices of rank $n - 1$.

Since the Poincaré polynomial of the non-degenerate case was

$$\pi'(\mathcal{A}, t) = \sum_{l=0}^n (-1)^l c_l (-t)^l$$

(KT, Proof of Theorem 2.1), the sum of the values of the Möbius function at rank l was $(-1)^l c_l$. Thus the Möbius function in the present case takes the same values at ranks less than n and $-\sum_{l \leq n-1} (-1)^l c_l$ at rank n .

Therefore, the Poincaré polynomial of the present case is

$$\pi(\mathcal{A}, t) = \sum_{l \leq n-1} c_l t^l + \left(-\sum_{l \leq n-1} (-1)^l c_l\right) (-t)^n,$$

and the number of regions is

$$\begin{aligned}\pi(\mathcal{A}, 1) &= \sum_{l \leq n-1} (1 + (-1)^{l+n+1})c_l \\ &= 2(c_{n-1} + c_{n-3} + c_{n-5} + \cdots).\end{aligned}$$

The proof for the case $n < m - 1$ is complete.

The formula is still valid for the case $n \geq m - 1$ because of

$$\begin{aligned}2(c_{m-2} + c_{m-4} + \cdots) &= m!, \\ c_{m-1} + c_{m-3} + \cdots &= c_{m-2} + c_{m-4} + \cdots\end{aligned}$$

(KT, p.21) and $c_i = 0$ for $i \geq m$. ■

Remark 3.2

Perturb the m points $\mu_1, \dots, \mu_m \in S^{n-1}$ slightly in R^n in such a way that

$$(\mu'_1, \|\mu_1\|^2), \dots, (\mu'_m, \|\mu_m\|^2)$$

are in general position in R^{n+1} as well as μ_1, \dots, μ_m in R^n (KT, p.12). Then, comparing the two arrangements before and after the perturbation, we see that the arrangement before perturbation is obtained by degenerating all the terminal nodes of the arrangement after perturbation to a single point. Accordingly, regions of the linear discrimination in R^n of the present paper are expected to correspond to regions of unbounded regions of non-degenerate discriminant analysis in R^n of KT. Indeed, the number of unbounded regions in KT is

$$(c_n + c_{n-1} + \cdots + c_0) - (c_n - c_{n-1} + c_{n-2} - \cdots) = 2(c_{n-1} + c_{n-3} + \cdots)$$

(KT, Theorems 2.1 and 2.3), which agrees with the result in Theorem 3.1.

Remark 3.3

As is seen from Theorem 3.1, the number of rankings generated by discriminant analysis on S^{n-1} is always even. This can be understood by considering for each $\mathbf{x} \in S^{n-1}$ its antipodal point $-\mathbf{x} \in S^{n-1}$. That is, for each pair of mutually reverse rankings, both of them or neither of them is generated.

4 Miscellaneous results

Other results in KT for linear discrimination in R^n remain to hold with minor modifications in our spherical discrimination.

First, the shortest path in terms of the usual metric on S^{n-1} is also the shortest path in terms of a particular distance for orderings. Specifically, the numerator of Kendall's τ between two generated orderings—that is, the minimum number of pairwise adjacent transpositions needed to transform one ordering to the other—is equal to the number of

discriminant boundaries B_{ij} meeting the geodesic path on S^{n-1} between points of the two corresponding regions. The proof is almost the same as that of Theorem 4.1.1 of KT.

Secondly, we can characterize the regions around a terminal node in the following way. First note that, in view of the identification with discrimination in R^n in Section 3.2, a terminal node in our spherical discrimination—that is, a point on S^{n-1} given as the intersection of some of the discriminant boundaries B_{ij} —can be indexed by a partial ordering with $m - n + 1$ blocks

$$\pi = (\pi_1, \dots, \pi_{m-n+1}). \quad (6)$$

Now, when $n \leq m$, around a terminal node $x \in S^{n-1}$ corresponding to partial ordering (6), there arise those, and only those, regions which are obtained by giving arbitrary orders to items in $\pi_1, \dots, \pi_{m-n+1}$ independently. The proof is similar to that of Theorem 4.2.1 of KT.

Thirdly, the equivalence of our ranking rule and the rule by ideal vector model in R^n is obvious from the equivalence of (2) and $\langle \mathbf{x}, \boldsymbol{\mu}_i \rangle > \langle \mathbf{x}, \boldsymbol{\mu}_j \rangle$, $\mathbf{x} \in S^{n-1}$.

Acknowledgment

We thank Prof. Shinto Eguchi for valuable comments.

References

- [1] Kamiya, H. and Takemura, A. (1997). On rankings generated by pairwise linear discriminant analysis of m populations. *J. Multivariate Anal.*, **61**, 1–28.
- [2] Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press, San Diego.
- [3] Orlik, P. and Terao, H. (1992). *Arrangements of Hyperplanes*. Springer-Verlag, Berlin.
- [4] Zaslavsky, T. (1975). *Facing up to Arrangements: Face-Count Formulas for Partitions of Space by Hyperplanes*. American Mathematical Society, Providence, RI.