95-F-23

On rankings generated by pairwise linear discriminant analysis of [illegible] populations

by

Hidehiko Kamiya
Graduate School of Economics, University of Tokyo

Akimichi Takemura
Faculty of Economics, University of Tokyo

October 1995

On rankings generated by pairwise linear discriminant analysis of m populations

Hidehiko Kamiya
Graduate School of Economics, University of Tokyo
and
Akimichi Takemura
Faculty of Economics, University of Tokyo

October 1995

1 Introduction

Consider the problem of pairwise linear discriminant analysis among m populations in R^n. For each pair of populations, we have a discriminant hyperplane. Then R^n is divided into regions by $m(m-1)/2$ such hyperplanes. Each region is indexed by an ordering of m populations, with the nearest population assigned the rank 1, the second nearest the rank 2, and so on. Therefore, we can regard pairwise multiple discriminant analysis as a generation process of rankings or orderings among m populations. This connection between multiple discriminant analysis and rankings seems to have been rarely discussed in the literature. We discuss related concepts in the existing literature in Section 1.1. For a survey of statistical analysis of ranking data, see Critchlow [4] and Fligner and Verducci [6].

Let the m populations be $N(\boldsymbol{\mu}_i, \Sigma), i = 1, \ldots, m$. For simplicity, we consider the canonical case, namely, we assume that the prior weights for the m populations are equal and that the common covariance matrix Σ is known and therefore $\Sigma = I$ (the identity matrix) without loss of generality. These assumptions are not restrictive because in general, distances to the populations are just measured by Mahalanobis distance. Thus, in the canonical case, the discriminant hyperplane between populations i and j is the bisector of the line segment connecting $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$.

There are $m!$ possible orderings among m populations. On the other hand, since each of the $m(m-1)/2$ hyperplanes cuts R^n into 2 half-spaces, the apparent maximum number of possible regions is $2^{m(m-1)/2}$. However, since we are considering partition of R^n by hyperplanes, it can be easily verified that the number of regions can not exceed $\sum_{j=0}^{n} \binom{m(m-1)/2}{j}$. Moreover, because there exist sets of three discriminant hyperplanes which necessarily share a common $(n-2)$-dimensional intersection, the maximum number of regions generated by discriminant hyperplanes is indeed $m!$. Now the question is whether all the $m!$ orderings are generated. It is easy to see that when the space is small compared with the number of populations, more precisely, if $n < m-1$, then some of the $m!$ orderings are not generated. Here arises a question: (Q–1) *How many regions arise for given n and m?* A more difficult question is: (Q–2) *How can we characterize non-arising regions?*

We review some related concepts in the literature in Section 1.1, and illustrate our problem with simple examples for $n = 2$ in Section 1.2. Basic terms and the notation are introduced in Section 1.3. Then in Section 2, we consider the question (Q–1) and give formulae for the number of regions. The number of bounded regions will be given there as well. Next, in Section 3, we take up (Q–2) and give some basic characterization of non-arising regions in the general case. For the particular case $n = m - 2$, we can completely characterize non-arising regions. Namely, *regions corresponding to the reverse orderings of bounded regions do not arise; furthermore, when $n = m - 2$, this characterizes non-arising regions.* In Section 4, we prove several results of independent interest.

1.1 Survey of various related concepts in the literature

Here we review various concepts in the literature which are closely related to our framework.

Voronoi diagram. The Voronoi diagram finds application in wide areas such as spatial interpolation, models of spatial processes, point pattern analysis, and locational optimization. It is defined as follows: Let

$P = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m\}$ be a set of points in R^n, where $2 \leq m < \infty$ and $\boldsymbol{p}_i \neq \boldsymbol{p}_j$ for $i \neq j$. Then,

$$V(\boldsymbol{p}_i) = \{\boldsymbol{x} \in R^n : \|\boldsymbol{x} - \boldsymbol{p}_i\| \leq \|\boldsymbol{x} - \boldsymbol{p}_j\| \text{ for } j \neq i\}$$

is called the n-dimensional Voronoi polyhedron associated with $\boldsymbol{p}_i$, and the set $\{V(\boldsymbol{p}_1), \ldots, V(\boldsymbol{p}_m)\}$ is called the n-dimensional Voronoi diagram generated by P.

Mathematically, the method of the Voronoi diagram is equivalent to the pairwise linear discriminant analysis. In other words, each Voronoi polyhedron is the union of the closures of the regions in this paper in which the corresponding population is given the rank 1. Furthermore, the Voronoi diagram is generalized in a variety of ways. One generalization which is closely related to our theory is the (ordered) order-k Voronoi diagram (Okabe, Boots, and Sugihara [12]). Our regions in pairwise linear discriminant analysis of m populations are the interiors of the "ordered order-m Voronoi polyhedrons."

For a comprehensive treatment of the Voronoi diagram, the reader is referred to Okabe, Boots, and Sugihara [12].

Permutahedron. The permutahedron $\Pi_{m-1} \subseteq R^m$ is defined as the convex hull of the $m!$ points in R^m whose coordinates are the orderings of $\{1, 2, \ldots, m\}$. Two vertices of Π_{m-1} are connected by an edge iff the corresponding orderings differ by an adjacent transposition. Thus, the metric version of Kendall's τ (Critchlow [4], Section II.B) is the minimum number of edges that must be traversed to get from one vertex to another.

Part of Π_{m-1} is in the dual relation to the arrangement of discriminant hyperplanes of m populations in this paper.

For more information on the permutahedron, see Thompson [19], [20] and Ziegler [24].

Ideal vector/point model. Ideal vector model and ideal point model have been studied in social choice theory, psychometry, marketing science, etc.. In these models, m objects or items $1, 2, \ldots, m$ are judged in terms of n kinds of attributes. Each attribute corresponds to a coordinate axis, and each object i is represented as a point $\boldsymbol{x}_i$ in R^n. In ideal point model, the "ideal point" $\boldsymbol{p}$ is supposed to exist, and the m objects are ranked according to the Euclidean distances to $\boldsymbol{p}$. Specifically, i is ranked better than j iff

$\|\boldsymbol{x}_i - \boldsymbol{p}\| < \|\boldsymbol{x}_j - \boldsymbol{p}\|$. In ideal vector model, on the other hand, the "ideal vector" $\boldsymbol{d}$ is supposed to exist, and the m objects are ranked according to the projections onto this direction. Specifically, i is ranked better than j iff $(\boldsymbol{d}, \boldsymbol{x}_i) > (\boldsymbol{d}, \boldsymbol{x}_j)$, where $(\ ,\ )$ denotes the inner product.

Ideal point model is related to our theory in the following way. Suppose the m objects $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in R^n$ are given. If we are given an individual's or a group of individuals' preference among the m objects in the form of an ordering σ, then the individual's or the group of individuals' ideal point must lie in the region C_σ in this paper. On the other hand, it is shown in Section 4.3 that given the m objects, the set of rankings which can occur in ideal vector model coincides with the set of rankings corresponding to unbounded regions in this paper. Because of the above connection between discriminant analysis and ideal point/vector model, we use the words "population" and "item" interchangeably from now on.

Variations of ideal point model. Various models based on ideal point model have been considered. Here, we briefly review unidimensional unfolding model, multidimensional unfolding threshold model, and ideal point discriminant analysis model.

Unidimensional unfolding model has been employed in the study of social choice problem. In this model, m options $O_1, O_2, \ldots, O_m$ are ranked by individuals. It is supposed that a "unidimensional underlying continuum," called the joint scale, exists, and that the m options are located on this continuum. Each individual I has an ideal on the joint scale, and he or she ranks the options according to the distances of the option points from this ideal, with nearer options being more preferred. Different orderings can be generated by varying the location of the ideal point. These orderings are said to be compatible with the underlying joint scale, and they are called admissible orderings. Unfolding is defined as follows: given a set of individuals' orderings, we wish to determine the joint scale on which individuals as well as options are located such that the given individuals' orderings are consistent with the orderings determined by this joint scale, although this is not always possible. Mathematically, this model can be considered a special case of our theory—pairwise multiple discriminant analysis among m populations in R^1. Admissible orderings correspond to arising regions in this paper. For unidimensional unfolding model, see Coombs [3], Luce and Raiffa [9], and van Blokland-Vogelesang [21]. This unidimensional model was extended to the multidimensional case by Bennett and Hays [1] and Hays and Bennett [7].

Multidimensional unfolding threshold model was proposed by DeSarbo and Hoffman [5] for the analysis of binary choice data in marketing research. Each of the binary data indicates whether a particular brand was chosen by the respondent or not. The model is stochastic, and the dichotomous variable y_{ij} generating the binary data is defined through the unobservable latent "disutility" variable D_{ij} :

$$D_{ij} = \|\boldsymbol{p}_i - \boldsymbol{x}_j\|^2 + \epsilon_{ij},$$

where $\boldsymbol{p}_i \in R^n$ is respondent i's ideal point, $\boldsymbol{x}_j \in R^n$ represent brand j, and ϵ_{ij} is a stochastic error component. Now, respondent i chooses brand j ($y_{ij} = 1$) if and only if respondent i's latent disutility for brand j is less than or equal to some individual threshold value d_i :

$$y_{ij} = 1 \text{ iff } D_{ij} \leq d_i.$$

Ideal point discriminant analysis was proposed by Takane, Bozdogan, and Shibayama [18]. Subjects are classified into one of m criterion groups $1, 2, \ldots, m$. It is assumed that subject i is represented as a point $\boldsymbol{y}_i$ in R^n, and that criterion group k has an ideal point $\boldsymbol{a}_k$ in the same R^n which represents the prototype of the group. Here, $\boldsymbol{y}_i \in R^n$ are supposed to be constrained as linear functions of the vectors $\boldsymbol{z}_i \in R^p$ of predictor variables:

$$\boldsymbol{y}_i = B\boldsymbol{z}_i,$$

where B is an $n \times p$ matrix of weights. Now, the probability that a particular subject i belongs to a particular criterion group k is assumed to be a decreasing function of the distance between the corresponding points $\boldsymbol{y}_i$ and $\boldsymbol{a}_k$. Specifically, the conditional probability $p_{k|i}$ that subject i belongs to criterion group k given the observation on $\boldsymbol{z}_i$ is

$$p_{k|i} = \frac{w_k \exp(-d_{ik}^2)}{\sum_{l=1}^m w_l \exp(-d_{il}^2)},$$

where w_k is a bias parameter for group k, and $d_{ik} = \|\boldsymbol{y}_i - \boldsymbol{a}_k\|^2$. This model is a special form of ideal point model combined with Luce's [8] biased choice model. The special feature of this model lies in that subject points are constrained as linear functions of the vectors of predictor variables. For extension and application of ideal point discriminant analysis, the reader is referred to Takane [15],[16], and [17].

Arrangement of hyperplanes. The problem of counting chambers, i.e., regions, in hyperplane arrangement becomes much harder when degeneracy is allowed. Zaslavsky [23] gave a formula for the number of regions in an arbitrary arrangement 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 (Lemma 2.1): The number of regions is equal to $\pi(\mathcal{A}, 1)$.

In this paper, we make extensive use of the general theory of hyperplane arrangements. For a full treatment of the theory, the reader is referred to Chapter 1 and 2 of Orlik and Terao [13].

1.2 Examples

In order to understand our problem, it is best to investigate simple examples in R^2.

If there are three populations, we have $\binom{3}{2} = 3$ discriminant lines as illustrated in Figure 1. In Figure 1, we can see all the $3! = 6$ orderings appearing. Note that three lines necessarily intersect in one point so long as the arrangement is non-degenerate or "in general position."

Now consider the case $m = 4$. We know that for any three of four populations the situation is as in the case $m = 3$. It it is not clear, however, how these $\binom{4}{3} = 4$ sub-arrangements intertwine with one another to produce the whole arrangement. If four points are placed as in Figure 2, then we have the corresponding arrangement of lines as shown in Figure 2. We see that only 18 out of $6! = 24$ regions arise in this case; the non-arising regions are

$$\begin{array}{cc} (1,2,3,4) & (1,3,2,4) \\ (2,1,3,4) & (2,3,1,4) \\ (3,1,2,4) & (3,2,1,4), \end{array}$$

where $(1,2,3,4)$, for example, is the corresponding ordering of the region where 1 is the nearest population, 2 is the second nearest population, etc..

We can explain why these regions do not occur as follows. By looking at Figure 2, we see that population 4 is "neutral," so 4 can not be the farthest

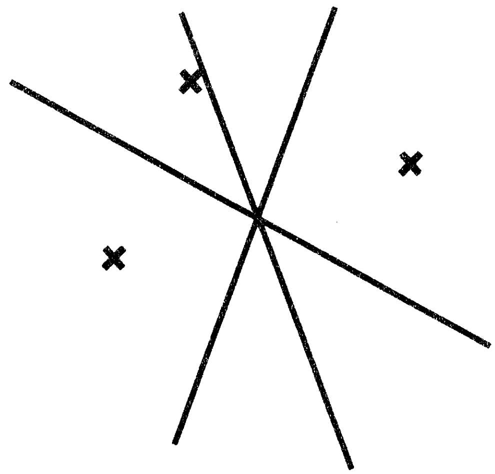

Figure 1: Discriminant analysis of three populations in R^2.

population from any point of R^2. Thus, the orderings with 4 in the last slot do not appear, and these are just the orderings listed above.

However, we can not explain non-arising regions in this way for all m and n. Even when $m = 4, n = 2$, there are cases in which this simple explanation is impossible. In fact, the four points in Figure 3 induce the arrangement of lines in Figure 3. This arrangement is of a different type from the one in Figure 2, as can be confirmed by noting that the non-arising regions are not that type of regions with a particular population assigned the last rank.

On the other hand, note that the numbers of non-arising regions do coincide in both cases, i.e., six regions do not occur in Figure 2 and Figure 3. This number seems to depend only on m and n. This can be proved using the general theory of hyperplane arrangements (Section 2).

Now, let us have a closer look at Figure 3. We may make the following observations:

1. Two neighboring regions differ by a pair of adjacent items, i and j, say; when one gets from one region to the other, the adjacent transposition (i, j) occurs. The items i and j correspond to the discriminant line containing the line segment or the half line which one has to traverse when passing between the two regions.

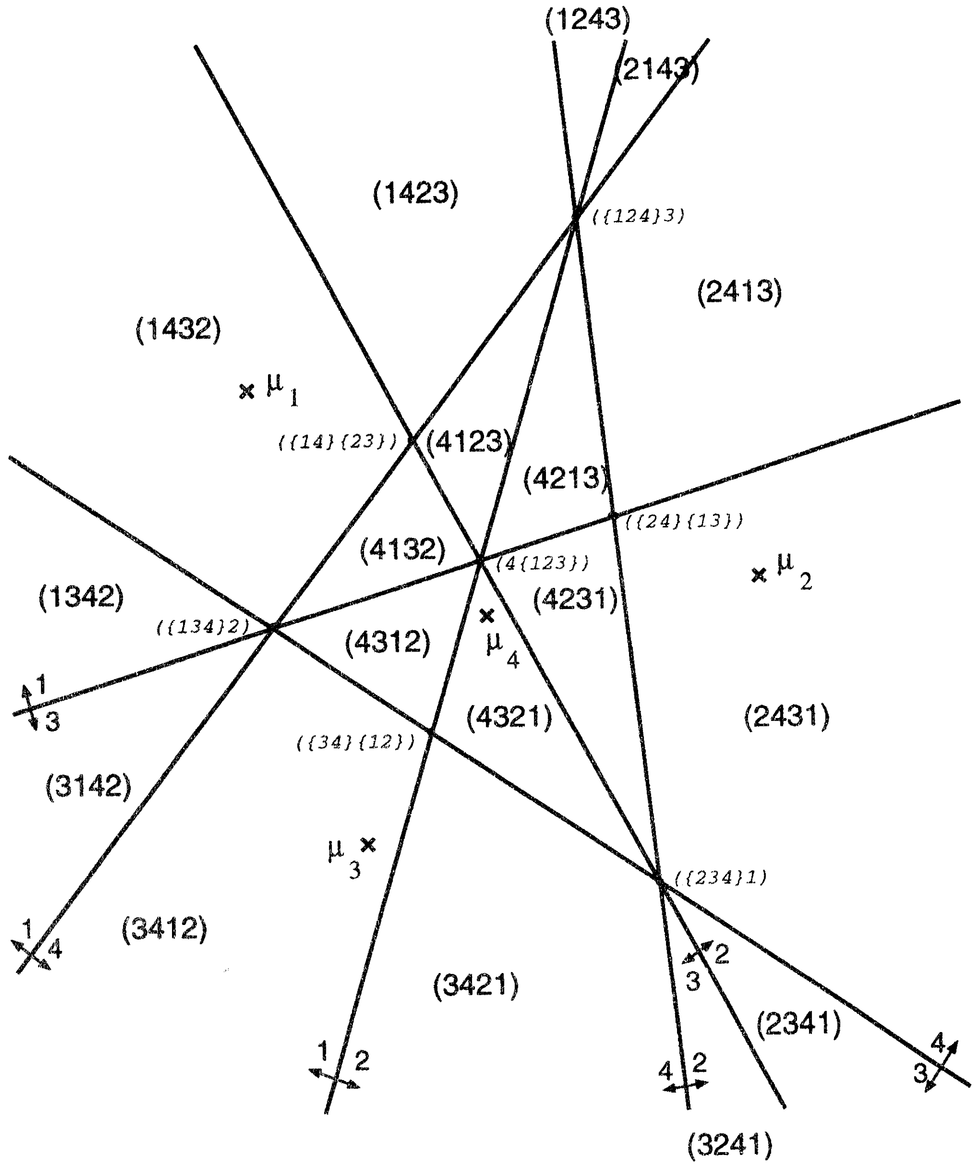

Figure 2: Discriminant analysis of four populations in R^2 : Case I. Note that commas between items are omitted for convenience of display.

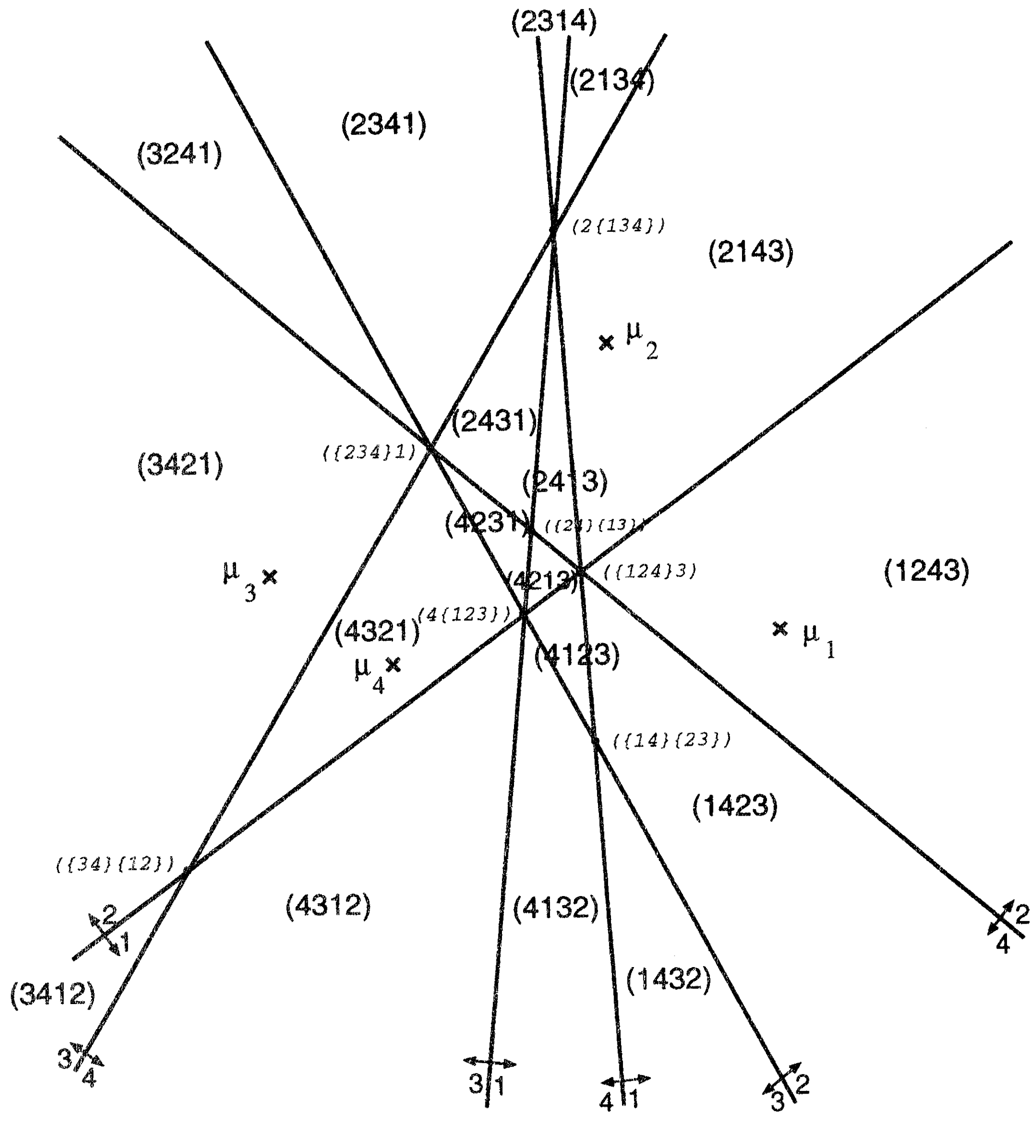

Figure 3: Discriminant analysis of four populations in R^2 : Case II.

2. A line segment connecting arbitrary points of two regions is one of the shortest paths in terms of orderings. This means that the number of crossings of borders needed to go from one point to the other, provided that one does not pass through terminal nodes (i.e., the points of intersections of discriminant lines) is equal to the minimum number of pairwise adjacent transpositions needed to transform one corresponding ordering to the other. This is just Kendall's τ between the two orderings.

3. Each terminal node is indexed by an ordering of blocks. The types are $(\{i,j\},\{k,l\}), (\{i,j,k\},l)$, and $(i,\{j,k,l\})$. Here, the order of items within a block, namely, the order of items in braces, is irrelevant, but the order of blocks is relevant.

4. Around each terminal node, there arise all regions whose corresponding orderings are obtained by giving arbitrary orders to items that are ranked together in the same block at that terminal node.

Observations 2 and 4 are verified in the general case in Sections 4.1 and 4.2, respectively.

1.3 Terminology and notation

Here, we make some basic terms precise and introduce the notation. We first explain the concepts concerning rankings.

A ranking of m items $\{1, 2, \ldots, m\}$ can be expressed as an ordering of them. The ordering $\sigma = (i_1, i_2, \ldots, i_m)$ corresponds to the ranking in which item i_1 is ranked first, item i_2 is ranked second, and so on. The rank given to item i by σ is denoted by $\sigma^{-1}(i)$. Note that σ denotes an ordering and not a "ranking," i.e., $(\sigma^{-1}(1), \ldots, \sigma^{-1}(m))$, in the terminology of Thompson [19],[20]. For an ordering $\sigma = (i_1, \ldots, i_m)$, its reverse ordering, denoted by $-\sigma$, is $(i_m, \ldots, i_1)$. 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 list some basic definitions from the theory of hyperplane arrangements. They are taken from Orlik and Terao [13].

A hyperplane arrangement $\mathcal{A}$ is a finite set of hyperplanes in $V = R^n$. $\mathcal{A}$ is called centered if $\cap_{H \in \mathcal{A}} H \neq \emptyset$. In particular, it is central if each hyperplane contains the origin. The intersection poset $L = L(\mathcal{A})$ is the set of nonempty intersections of elements of $\mathcal{A}$ endowed with the partial order defined by

$$X \leq Y \Longleftrightarrow Y \subseteq X.$$

The rank function on L is defined by $r(X) = \text{codim}(X)$. Maximal elements of $L(\mathcal{A})$ have the same rank, and the rank of $\mathcal{A}, r(\mathcal{A})$, is defined to be the rank of a maximal element of $L(\mathcal{A})$. Let $L_p = L_p(\mathcal{A}) = \{X \in L(\mathcal{A}) |\ r(X) = p\}$. The Hasse diagram of L has vertices labeled by the elements of L and arranged on levels L_p for $p \geq 0$. An edge in the Hasse diagram connects $X \in L_p$ with $Y \in L_{p+1}$ if $X < Y$. Define the Möbius function ν as follows:

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

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 a connected component of $R^n \setminus \cup_{H \in \mathcal{A}} H$. A face P is a chamber of $\mathcal{A}^X$ for some $X \in L$, where $\mathcal{A}^X$ is the restriction of $\mathcal{A}$ to X:

$$\mathcal{A}^X = \{X \cap H : H \in \mathcal{A} \backslash \mathcal{A}_X \quad \text{and} \quad X \cap H \neq \emptyset\},$$

with $\mathcal{A}_X = \{H \in \mathcal{A} : X \subseteq H\}$. However, we use the term "region" instead of "chamber" when dealing with a chamber of $\mathcal{A}^V = \mathcal{A}$ itself. The set of faces endowed with the partial order $\leq_f$:

$$P \leq_f Q \Longleftrightarrow Q \subseteq \overline{P},$$

where $\overline{P}$ denotes the closure of P, is called the face poset of $\mathcal{A}$.

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

Denote by

$$H_{ij}, \quad 1 \leq j < i \leq m,$$

the discriminant hyperplane between populations i and j, 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 I 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, i.e., each block of the latter is a union of blocks of the former. Specifically, to $X \in L$ corresponds the partition of $\{1, 2, \ldots, m\}$ into equivalence classes under the equivalence relation $\sim_X$ defined by

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

where we agree that $H_{ii} = V$ and $H_{ji} = H_{ij}$ for $i > j$. Note that $i \sim_X j$ means that $\boldsymbol{x} \in X$ is equidistant from $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$. The element of L indexed by a partition I is denoted by X_I. The Hasse diagram of "non-degenerate" (see Section 2 for definition) discriminant analysis with $m = 4, n \geq 3$ is given in Figure 4. The Hasse diagram remains the same for $n < 3$ except that vertices of rank greater than n are not present.

We rank the m populations according to the 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. The region indexed by an ordering σ is denoted by C_σ, and the ordering corresponding to a region C is denoted by σ_C. On the other hand, a face of dimension less than n is indexed by a partial ordering. The face corresponding to a partial ordering π is denoted by P_π. Elements of L of rank n, if they exist, are called terminal nodes. Each terminal node can also be considered a face, and thus it can be indexed by a partial ordering. In other words, if $X_I \in L$ is of dimension zero, the order among the blocks of the partition I is uniquely determined.

Regions fall into two types: bounded regions and unbounded ones. This distinction plays an important role in the characterization of non-arising regions. Also, as was mentioned in Section 1.1, there is a connection between unbounded regions and ideal vectors. An unbounded region is, by definition,

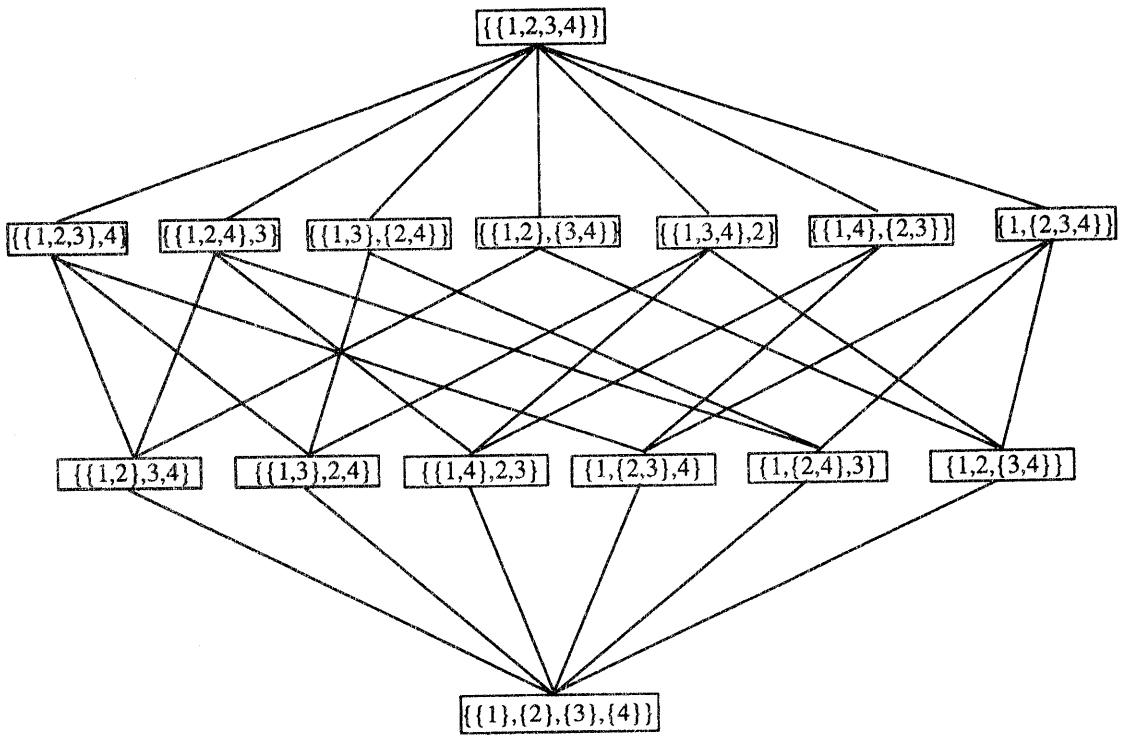

Figure 4: The Hasse diagram of non-degenerate discriminant analysis of four populations in $R^n, n \geq 3$.

a region which is not contained in any ball $B_r(\boldsymbol{x}_0) \equiv \{\boldsymbol{x} \in R^n : \|\boldsymbol{x} - \boldsymbol{x}_0\| \leq r\}, 0 < r < \infty, \boldsymbol{x}_0 \in R^n$. It is shown in Lemma 3.2 that in the non-degenerate pairwise linear discriminant analysis, an unbounded region recedes in a certain direction, i.e., there exists a direction $\boldsymbol{d} \in R^n, \|\boldsymbol{d}\| = 1$, such that the points $t\boldsymbol{d}$ are contained in the region for all sufficiently large $t > 0$.

2 The number of regions

In this section, we give expressions for the number of regions. However, before we state the theorems, we need to discuss the notion of non-degeneracy of discriminant hyperplanes.

We say that the discriminant analysis is non-degenerate if the following two assumptions hold.

(A–1) The points $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m \in R^n$ are in general position.

(A–2) The points

$$(\boldsymbol{\mu}_1', \ \|\boldsymbol{\mu}_1\|^2), \ldots, (\boldsymbol{\mu}_m', \ \|\boldsymbol{\mu}_m\|^2)$$

in R^{n+1} are in general position.

Equivalent assumptions are made in other contexts. See, e.g., Section 3 of Naiman and Wynn [11].

Remark 2.1. When $m \leq n+1$, the m points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in R^n$ are said to be in general position iff for any set of scalars $a_1, \ldots, a_m$ with $\sum_i a_i = 0$,

$$\sum_i a_i \boldsymbol{x}_i = \boldsymbol{0} \implies a_i = 0, \ \forall i.$$

This is equivalent to saying that the $m-1$ vectors $\overrightarrow{\boldsymbol{x}_1\boldsymbol{x}_2}, \ldots, \overrightarrow{\boldsymbol{x}_1\boldsymbol{x}_m}$ are linearly independent.

When $m > n+1$, the m points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are said to be in general position iff any $n+1$ of them are in general position.

Remark 2.2. When $m \leq n+1$, (A–1) implies (A–2) as follows. For any set of scalars $a_1, \ldots, a_m$ with $\sum a_i = 0$,

$$\sum a_i(\boldsymbol{x}_i', \ \|\boldsymbol{x}_i\|^2)' = \boldsymbol{0}$$

implies

$$\sum a_i \boldsymbol{x}_i = \boldsymbol{0},$$

and thus,

$$a_i = 0 \ \text{ for } \ i = 1, 2, \ldots, m.$$

So (A–1) alone suffices in this case.

Remark 2.3. Assumption (A–1) is invariant under affine transformations. On the other hand, (A–2) is invariant only under rigid motions. First, consider moving the origin to an arbitrary point $\boldsymbol{\mu}_0 \in R^n$. Then we require that

$$(\boldsymbol{\mu}_1' - \boldsymbol{\mu}_0', \ \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0\|^2), \ldots, (\boldsymbol{\mu}_m' - \boldsymbol{\mu}_0', \ \|\boldsymbol{\mu}_m - \boldsymbol{\mu}_0\|^2).$$

are in general position. By noting that

$$\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_0\|^2 = \|\boldsymbol{\mu}_i\|^2 - 2\boldsymbol{\mu}_0'\boldsymbol{\mu}_i + \|\boldsymbol{\mu}_0\|^2,$$

we have for any set of scalars $\{a_1, \ldots, a_m\}$ satisfying $\sum_i a_i = 0$,

$$\sum_i a_i(\boldsymbol{\mu}_i', \ \|\boldsymbol{\mu}_i\|^2) = 0 \iff \sum_i a_i(\boldsymbol{\mu}_i' - \boldsymbol{\mu}_0', \ \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_0\|^2) = 0.$$

Furthermore, it is easy to see that (A–2) is invariant under rotations in R^n.

Now we are ready to state the main results of this section.

Theorem 2.1. *The number of regions appearing in the non-degenerate discriminant analysis of m populations in R^n is given by*

$$c_0 + c_1 + \cdots + c_n,$$

where $c_0 = 1$ *and*

$$c_k = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq m-1} i_1 i_2 \cdots i_k, \quad k \geq 1.$$

Here,

$$(-1)^{m-k}c_{m-k} = s(m,k), \quad k = 0, 1, \ldots, m$$

are the Stirling numbers of the first kind (Macdonald [10], Section I.2, Berge [2], Section 1.5):

$$t(t-1)(t-2)\cdots(t-m+1) = \sum_{k=0}^{m} s(m,k)t^k.$$

As will be seen from the proof of Theorem 2.1, each $c_k, k = 0, \ldots, n$, is the sum of the absolute values of the Möbius functions at rank k :

$$c_k = \sum_{X:r(X)=k} (-1)^k \nu(X).$$

However, in general there exist elements of L of the same rank which are indexed by different types of partitions of $\{1, \ldots, m\}$, e.g., $\{\{1,2\},\{3,4\}\}$ and $\{1,\{2,3,4\}\}$. The following theorem identifies the contribution of each type of partitions of $\{1, \ldots, m\}$, that is, each partition of the positive integer m.

Theorem 2.2. *An alternative expression for the number of regions is*

$$\sum_{k \geq m-n} \sum_{(1^{k_1}, 2^{k_2}, \ldots, m^{k_m})} \frac{m!}{1^{k_1} 2^{k_2} \cdots m^{k_m} k_1! k_2! \cdots k_m!},$$

where the second summation is over all $(1^{k_1}, \ldots, m^{k_m})$ *such that*

$$\begin{aligned} m &= k_1 + 2k_2 + \cdots + mk_m, \\ k_i &= 0, 1, 2, \ldots, \\ k_1 + k_2 + \cdots + k_m &= k. \end{aligned}$$

Example. For $n = 2, m = 4$, c_i in Theorem 2.1 are

$$c_1 = 6, \ c_2 = 11,$$

so the number of region is

$$1 + c_1 + c_2 = 18.$$

The summation in Theorem 2.2 is over

$$(1^{k_1}, \ldots, 4^{k_4}) = (1^1, 2^0, 3^1, 4^0),\ (1^0, 2^2, 3^0, 4^0),\ (1^2, 2^1, 3^0, 4^0),\ (1^4, 2^0, 3^0, 4^0),$$

so the number of regions is again

$$8 + 3 + 6 + 1 = 18.$$

Before we prove the theorems, we state two lemmas.

In general, a face is called relatively bounded when it goes to infinity only in the directions of the relative vertices, i.e., the minimum-dimensional elements of L. For a formal definition of relative boundedness, see Zaslavsky [23], p. 25.

Lemma 2.1. (Zaslavsky [23], Theorem A, C). *Let $\pi(\mathcal{A}, t)$ be the Poincaré polynomial for hyperplane arrangement $\mathcal{A}$. Then the number of regions is $\pi(\mathcal{A}, 1)$, and the number of relatively bounded regions is $(-1)^{r(\mathcal{A})}\pi(\mathcal{A}, -1)$. In particular, if $\mathcal{A}$ is centered, the number of relatively bounded regions is zero.*

In the non-degenerate discriminant analysis of m populations in R^n with $n \leq m - 1$, we have $r(\mathcal{A}) = n$, so the relative vertices are of dimension zero. Thus, in this case, the set of relatively bounded regions is equal to the set of bounded regions

Lemma 2.2. *Let $\mathcal{A}$ be the arrangement of discriminant hyperplanes in the non-degenerate discriminant analysis of m populations in R^n. Then, $L = L(\mathcal{A})$ is isomorphic to the poset (partially ordered set) $\mathcal{I}_{m-n}$ of partitions of $\{1, 2, \ldots, m\}$ into $k \geq m - n$ blocks. Namely, $I \in \mathcal{I}_{m-n} \mapsto X_I \in L$ is bijective and I is a refinement of J iff $X_I \leq X_J$. Furthermore,*

$$r(X) = m - k$$

for all $X \in L$.

The proof will be given in Appendix.

Proof of Theorem 2.1. First, consider the case $n \geq m-1$. In this case, because we are considering the non-degenerate case, we have by Lemma 2.2 that $L(\mathcal{A})$ is isomorphic to the intersection poset of the braid arrangement in R^m (Orlik and Terao [13], Proposition 2.9). Thus, we obtain

$$\begin{aligned} \pi(\mathcal{A}, t) &= \prod_{i=1}^{m-1}(1+it) \\ &= 1 + c_1 t + c_2 t^2 + \cdots + c_{m-1} t^{m-1}, \end{aligned} \tag{1}$$

where $c_k = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq m-1} i_1 i_2 \cdots i_k$.

In the case $n < m-1$, we have

$$r(X) \leq n, \quad X \in L.$$

Therefore, we ignore terms of order greater than n in (1) and obtain

$$\pi(\mathcal{A}, t) = 1 + c_1 t + \cdots + c_n t^n.$$

Here we used the fact that the Hasse diagram is the same for all $n \geq 1$ except that vertices of rank greater than n are not present.

Putting together both cases, we obtain

$$\pi(\mathcal{A}, t) = 1 + c_1 t + \cdots + c_n t^n$$

regardless of whether $n \geq$ or $< m-1$. Thus, by Lemma 2.1, the number of regions is

$$1 + c_1 + \cdots + c_n.$$

Q.E.D.

Proof of Theorem 2.2. Consider the partition $(1^{k_1}, \ldots, m^{k_m})$ of the integer m with $k \geq m-n$ parts:

$$\begin{aligned} m &= \underbrace{1 + \cdots + 1}_{k_1} + \cdots + \underbrace{m + \cdots + m}_{k_m} \\ &= k_1 + 2k_2 + \cdots + mk_m, \\ k &= k_1 + \cdots + k_m, \quad k_i = 0, 1, 2, \ldots, \end{aligned}$$

and let X be one of the corresponding elements in L, e.g., the element in L indexed by the partition

$$\{\{1\},\ldots,\{k_1\},\{k_1+1,k_1+2\},\ldots,\{k_1+2k_2-1,k_1+2k_2\},\ldots\}$$

of $\{1,\ldots,m\}$.

For this X, we have

$$\nu(X)=(-1)^{(k_1+\cdots+k_m)+m}((1-1)!)^{k_1}\cdots((m-1)!)^{k_m}$$

by Berge [2], Section 3.2, and

$$r(X)=m-(k_1+\cdots+k_m)$$

by Lemma 2.2. Therefore, by noting that ν and r take the same values for all $X\in L$ that correspond to the same $(1^{k_1},\ldots,m^{k_m})$, we have

$$\begin{aligned}\pi(\mathcal{A},\,t) &= \sum_{k\geq m-n}\sum_{(1^{k_1},\ldots,m^{k_m})}\frac{m!}{(1!)^{k_1}\cdots(m!)^{k_m}k_1!\cdots k_m!}\\ &\qquad\times(-1)^{(k_1+\cdots+k_m)+m}((1-1)!)^{k_1}\cdots((m-1)!)^{k_m}\\ &\qquad\times(-t)^{m-(k_1+\cdots+k_m)}\\ &= \sum_{k\geq m-n}\sum_{(1^{k_1},\ldots,m^{k_m})}\frac{m!}{1^{k_1}\cdots m^{k_m}k_1!\cdots k_m!}t^{m-(k_1+\cdots+k_m)},\end{aligned}$$

where the second summations are over all $(1^{k_1},\ldots,m^{k_m})$ such that

$$\begin{aligned}m&=k_1+2k_2+\cdots+mk_m,\\ k&=k_1+\cdots+k_m,\\ k_i&=0,1,2,\ldots.\end{aligned}$$

Thus, the number of regions is

$$\sum_{k\geq m-n}\sum_{(1^{k_1},\ldots,m^{k_m})}\frac{m!}{1^{k_1}\cdots m^{k_m}k_1!\cdots k_m!}.$$

Q.E.D.

By comparing the coefficients of the Poincaré polynomials in the proofs of Theorems 2.1 and 2.2, we obtain as a byproduct the well-known formula (Macdonald [10], Section I.2):

$$(-1)^{m-k}s(m,k) = \sum_{(1^{k_1},\ldots,m^{k_m}):k_1+\cdots+k_m=k} \frac{m!}{1^{k_1}2^{k_2}\cdots m^{k_m}k_1!k_2!\cdots k_m!}.$$

Now, as mentioned earlier, the distinction between bounded regions and unbounded ones becomes essential in the characterization of non-arising regions.

Theorem 2.3. *The number of bounded regions in the non-degenerate discriminant analysis of m populations in R^n with $n < m-1$ is given by*

$$\begin{cases} c_n - c_{n-1} + c_{n-2} - \cdots - c_1 + 1 & n\text{:even}, \\ c_n - c_{n-1} + c_{n-2} - \cdots + c_1 - 1 & n\text{:odd}. \end{cases}$$

If $n \geq m-1$, all the $m!$ regions are unbounded.

Proof. First consider the case $n < m-1$. Recalling that the Poincaré polynomial is

$$\pi(\mathcal{A},\, t) = c_n t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + 1,$$

we have by Lemma 2.1 that the number of bounded regions is

$$\begin{aligned} & (-1)^{r(\mathcal{A})}\pi(\mathcal{A},\, -1) \\ = \; & (-1)^n((-1)^n c_n + (-1)^{n-1}c_{n-1} + \cdots - c_1 + 1) \\ = \; & c_n - c_{n-1} + c_{n-2} - \cdots. \end{aligned}$$

Next, when $n \geq m-1$, the arrangement is centered, so the number of relatively bounded regions, and hence the number of bounded regions, is zero by Lemma 2.1.

Q.E.D.

3 Characterization of Non-Arising Regions

In this section, we address the second question (Q–2) in Section 1, i.e., characterization of non-arising regions. We continue to deal with the problem of non-degenerate discriminant analysis among m populations in R^n. In characterizing non-arising regions, the distinction between bounded regions and unbounded ones is important.

When $n < m - 1$, we embed our R^n in R^{m-1} and regard the m points $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m$ as those in R^{m-1}. By treating the discrimination of m populations in R^n as the degenerate one of m populations in R^{m-1}, we gain a better insight. Our idea is similar to the method of coning discussed in Orlik and Terao [13], Section 1.2.

Theorem 3.1. *In the non-degenerate discriminant analysis among m populations in R^n with $n < m - 1$, we have the following facts.*

1. *If an ordering σ arises as an unbounded region, so does the reverse ordering $-\sigma$ in the opposite direction.*
2. *If an ordering σ arises as a bounded region, then the region corresponding to $-\sigma$ does not occur.*

First, we state two lemmas.

Lemma 3.1. *In the non-degenerate discriminant analysis among m populations in R^{m-1}, the intersection of all the discriminant hyperplanes consists of exactly one point, say O,:*

$$\dim\left(\bigcap_{i>j} H_{ij}\right) = 0,$$

and all the $m!$ orderings occur as convex cones with O as their vertices. In fact, if we translate O to the origin, the closure of each region is a polyhedral convex cone.

The proof is easy and omitted. More general version of this lemma is proved as Theorem 4.2.1.

Lemma 3.2. *If C is an unbounded region in the non-degenerate discriminant analysis among m populations in R^n, then C recedes in a certain direction, i.e., there is a direction $\boldsymbol{d} \in R^n, \|\boldsymbol{d}\| = 1$, such that*

$$t\boldsymbol{d} \in C \quad \text{for all sufficiently large} \quad t.$$

Moreover, $\boldsymbol{d}$ *can be taken so that it is not parallel to the hyperplanes.*

Proof. Express C as

$$C = \{\boldsymbol{x} \in R^n : (\boldsymbol{a}_i, \boldsymbol{x}) > b_i,\ 1 \leq i \leq m'\},\ m' = \frac{m(m-1)}{2}.$$

Then, it is obvious that C is convex and that its recession cone C' is

$$\begin{aligned} C' &\equiv \{\boldsymbol{x} \in R^n : C + \boldsymbol{x} \subseteq C\} \\ &= \{\boldsymbol{x} \in R^n : (\boldsymbol{a}_i, \boldsymbol{x}) \geq 0,\ 1 \leq i \leq m'\}, \end{aligned}$$

where $C + \boldsymbol{x}$ is the translate of C by $\boldsymbol{x} : C + \boldsymbol{x} = \{\boldsymbol{y} + \boldsymbol{x} : \boldsymbol{y} \in C\}$. For the notion of recession cone, see, for example, Rockafellar [14] or Webster [22]. Now, since C' is the same as the recession cone of $\overline{C}$, the closure of C, (Rockafellar [14], Corollary 8.3.1), and $\overline{C}$ is an unbounded, closed convex set, we have that $C' \neq \{\mathbf{o}\}$ (Rockafellar [14], Theorem 8.4). Thus, thanks to the assumption of non-degeneracy, the interior of C' is non-empty:

$$\text{int}C' = \{\boldsymbol{x} \in R^n : (\boldsymbol{a}_i, \boldsymbol{x}) > 0,\ 1 \leq i \leq m'\} \neq \emptyset,$$

and for any $\boldsymbol{d} \in \text{int}C'$, we have

$$t\boldsymbol{d} \in C \quad \text{for all sufficiently large} \quad t > 0$$

and

$$(\boldsymbol{a}_i, \boldsymbol{d}) \neq 0,\ 1 \leq i \leq m'.$$

Q.E.D.

Proof of Part 1. Embed our R^n in R^{m-1} and perturb the m points

$$\begin{matrix} (\boldsymbol{\mu}'_1, 0, \ldots, 0) \\ \vdots \\ (\boldsymbol{\mu}'_m, 0, \ldots, 0) \end{matrix} \tag{2}$$

slightly in R^{m-1} such that discriminant analysis among perturbed points are non-degenerate. For example, let the m points in (2) be perturbed as

$$\begin{gathered}(\boldsymbol{\mu}'_1, 0, \ldots, 0)\\ \vdots \\ (\boldsymbol{\mu}'_{n+1}, 0, \ldots, 0)\\ (\boldsymbol{\mu}'_{n+2}, \epsilon_1, 0, \ldots, 0)\\ (\boldsymbol{\mu}'_{n+3}, 0, \epsilon_2, 0, \ldots, 0)\\ \vdots \\ (\boldsymbol{\mu}'_m, 0, \ldots, 0, \epsilon_{m-n-1}).\end{gathered}$$

Note that if $\|\epsilon_1\|, \ldots, \|\epsilon_{m-n-1}\|$ are sufficiently small, then the face poset restricted to R^n remains the same. A similar argument is made in Remark 3.5 of Naiman and Wynn [11]. We denote the hyperplanes in R^{m-1} obtained by this perturbation by $\widetilde{H}_{ij}$.

Then, the situation is as in Lemma 3.1, and the embedded R^n is an affine subspace which does not pass through the origin. Thus, R^n can be expressed as

$$\boldsymbol{c}_0 + M$$

where $\boldsymbol{c}_0 \neq \mathbf{o},\ \boldsymbol{c}_0 \in R^{m-1}$, and $M \subseteq R^{m-1}$ is a linear subspace of dimension n.

As shown in Lemma 3.2, if an ordering σ arises as an unbounded region C_σ in R^n, then, C_σ recedes in some direction $\boldsymbol{d}, \|\boldsymbol{d}\| = 1$:

$$\exists \boldsymbol{d} \in R^n,\ \boldsymbol{c}_0 + t\boldsymbol{d} \in C_\sigma \quad \text{for all sufficiently large} \quad t > 0.$$

Here, $\boldsymbol{d}$ can be taken so that it is not parallel to the hyperplanes H_{ij}.

Now, take an arbitrary vector $\boldsymbol{d}, \|\boldsymbol{d}\| = 1$, in M which is not parallel to the hyperplanes $\widetilde{H}_{ij}$, and consider the region $\widetilde{C}_\sigma$ containing $\boldsymbol{d}$ in R^{m-1}, where σ is the corresponding ordering of this region.

Since $\boldsymbol{d} \in \widetilde{C}_\sigma$ and $\widetilde{C}_\sigma$ is open, we have for all sufficiently large $t > 0$,

$$\frac{\boldsymbol{c}_0}{t} + \boldsymbol{d} \in \widetilde{C}_\sigma$$

so that, with $\widetilde{C}_\sigma$ being a cone,

$$\boldsymbol{c}_0 + t\boldsymbol{d} \in \widetilde{C}_\sigma.$$

This implies

$$c_0 + t\boldsymbol{d} \in C_\sigma,$$

where $C_\sigma = \widetilde{C}_\sigma \cap R^n$ is the region in R^n indexed by the same ordering σ as $\widetilde{C}_\sigma$.

It follows that σ arises as an unbounded region in R^n. Also, $-\sigma$ occurs as an unbounded region in the opposite direction since $-\boldsymbol{d}$ is contained in $\widetilde{C}_{-\sigma}$.

The converse is also true, that is, if an ordering σ arises in the direction $\boldsymbol{d}$ in R^n which is not parallel to the hyperplanes H_{ij}, then, $\boldsymbol{d}$ is contained in $\widetilde{C}_\sigma$ in R^{m-1} :

$$c_0 + t\boldsymbol{d} \in C_\sigma \text{ for all sufficiently large } t \Longrightarrow \boldsymbol{d} \in \widetilde{C}_\sigma.$$

This can be shown as follows. If $\boldsymbol{d} \notin \widetilde{C}_\sigma$, then $\boldsymbol{d} \in \widetilde{C}_{\sigma'}$ for some $\sigma' \neq \sigma$, and thus

$$c_0 + t\boldsymbol{d} \in C_{\sigma'} \text{ for all sufficiently large } t$$

as shown above. This contradicts the fact that

$$c_0 + t\boldsymbol{d} \in C_\sigma \text{ for all sufficiently large } t,$$

which proves our assertion.

It follows from the assertions above that if σ occurs as an unbounded region, so does $-\sigma$ in the opposite direction.

Q.E.D.

Proof of Part 2. Suppose that $-\sigma$ arises in $R^n = c_0 + M$. Then, there exist $\boldsymbol{x}$ and $\boldsymbol{y} \in M$ such that

$$c_0 + \boldsymbol{x} \in C_\sigma \subseteq \widetilde{C}_\sigma, \quad (3)$$

$$c_0 + \boldsymbol{y} \in C_{-\sigma} \subseteq \widetilde{C}_{-\sigma}. \quad (4)$$

Here, C_σ and $\widetilde{C}_\sigma$ are the region in R^n and R^{m-1}, respectively, which correspond to σ, and the same is true for $C_{-\sigma}$ and $\widetilde{C}_{-\sigma}$.

From (4), we get

$$-c_0 - \boldsymbol{y} \in \widetilde{C}_\sigma. \quad (5)$$

Since $\widetilde{C}_\sigma$ is a convex cone, we have, by adding (3) and (5), that

$$\boldsymbol{x} - \boldsymbol{y} \in \widetilde{C}_\sigma.$$

If we note that

$$\boldsymbol{x} - \boldsymbol{y} \in M,$$

it follows from the proof of Part 1 that σ appears as an unbounded region in R^n, which is a contradiction.

Q.E.D.

Remark 3.1. Part 1 follows immediately without embedding R^n in R^{m-1} : Given an unbounded region C, take a vector $\boldsymbol{d}$ as in Lemma 3.2, and note that the line $\{t\boldsymbol{d} : t \in R\}$ intersects each hyperplane exactly once. However, the proof stated above is used also in the proof of Part 2.

Remark 3.2. When $n \geq m - 1$, Part 1 remains true by Remark 3.1. Also, Part 2 is trivially true by Theorem 2.3 in this case.

When $n = m - 2$, we have a complete characterization of non-arising regions.

Corollary 3.1. *In the case $n = m - 2$, the set of non-arising orderings is precisely the set of the reverse orderings of bounded regions.*

Proof. By virtue of Part 2 of the theorem, it is sufficient to prove that the number of non-arising regions is equal to that of bounded regions. Thanks to Theorems 2.1 and 2.3, the former is

$$m! - (c_n + c_{n-1} + \cdots + c_1 + 1),$$

and the latter is

$$c_n - c_{n-1} + c_{n-2} - \cdots.$$

Thus, the proof will be finished if we show that

$$m! = 2(c_n + c_{n-2} + \cdots).$$

Putting $t = -1$ in the identity

$$(1+t)(1+2t)\cdots(1+(m-1)t) = c_{m-1}t^{m-1} + c_{m-2}t^{m-2} + \cdots + c_1 t + 1,$$

we obtain

$$0 = -c_{m-1} + c_{m-2} - c_{m-3} + c_{m-4} - \cdots,$$

and hence,

$$\begin{aligned} m! &= c_{m-1} + c_{m-2} + \cdots + c_1 + 1 \\ &= c_{m-1} + c_{m-2} + \cdots + c_1 + 1 + (-c_{m-1} + c_{m-2} - c_{m-3} + c_{m-4} - \cdots) \\ &= 2(c_{m-2} + c_{m-4} + \cdots) \\ &= 2(c_n + c_{n-2} + \cdots). \end{aligned}$$

This completes the proof.

Q.E.D.

For the case $n \leq m - 3$, it seems difficult to completely characterize the non-arising regions.

4 Miscellaneous results

In this section, we prove several results of independent interest.

4.1 Relation to Kendall's τ

First we consider the relation between Kendall's τ and a line segment in our discriminant arrangement.

Theorem 4.1.1. *Let $\tau(\sigma_C, \sigma_{C'})$ be Kendall's τ between the two orderings corresponding to regions C and C'. Then, $\tau(\sigma_C, \sigma_{C'})$ coincides with the number of hyperplanes meeting a line segment connecting C and C'.*

Proof.

$$\begin{aligned} \tau(\sigma_C, \sigma_{C'}) &= \#\{(i,j) : (\sigma_C^{-1}(i) - \sigma_C^{-1}(j))(\sigma_{C'}^{-1}(i) - \sigma_{C'}^{-1}(j)) < 0, \ \ i > j\} \\ &= \#\{(i,j) : C \text{ and } C' \text{ are on the opposite sides of } H_{ij}\} \\ &= \#\{H_{ij} : \overline{\boldsymbol{x}_C \boldsymbol{x}_{C'}} \cap H_{ij} \neq \emptyset\}, \end{aligned}$$

where $\boldsymbol{x}_C$ and $\boldsymbol{x}_{C'}$ are arbitrary points of C and C', respectively.

Q.E.D.

4.2 Regions around a terminal node

Next we give a characterization of regions around a terminal node.

Theorem 4.2.1. *Consider the non-degenerate discriminant analysis among m populations in R^n with $n \leq m-1$. Let $\boldsymbol{x}$ be a terminal node and let $\pi = (\pi_1, \ldots, \pi_{m-n})$ be the partial ordering corresponding to $\boldsymbol{x}$. Let $|\pi_i| = l_i, i = 1, \ldots, m-n$. Then, around $\boldsymbol{x}$, there are $\prod_{i=1}^{m-n} l_i!$ regions. These regions are obtained by giving arbitrary orders to items in $\pi_1, \ldots, \pi_{m-n}$ independently.*

Proof. Consider the terminal node X_I which is indexed, without loss of generality, by the partition

$$I = \{\{1, 2, \ldots, l_1\}, \{l_1+1, \ldots, l_1+l_2\}, \ldots, \{m - l_{m-n} + 1, \ldots, m\}\}.$$

Here, the order among the blocks is uniquely determined. Thus there corresponds a partial ordering. Assume, without loss of generality again, that the corresponding partial ordering is

$$\pi = (\{1, 2, \ldots, l_1\}, \{l_1+1, \ldots, l_1+l_2\}, \ldots, \{m - l_{m-n} + 1, \ldots, m\}).$$

First, we want to show that any region corresponding to the full ordering obtained by giving some order among the items in each block of π occurs around X_I. It is sufficient to see that the region C_σ corresponding to the ordering

$$\sigma = (1, \ldots, l_1, l_1+1, \ldots, l_1+l_2, \ldots, m - l_{m-n} + 1, \ldots, m)$$

arises around X_I.

Now, consider $X_{I'} \in L$ which corresponds to the partition

$$I' = \{\{1\}, \{2, \ldots, l_1\}, \{l_1+1, \ldots, l_1+l_2\}, \ldots, \{m - l_{m-n} + 1, \ldots, m\}\}.$$

Considering the number of blocks of I' and the refinement relation between partitions I and I', we see that in the non-degenerate case, $X_{I'}$ is of rank

$n-1$, or of dimension 1, and contains X_I. That is, $X_{I'}$ is a line passing through the terminal node X_I.

Some of the partial orderings occur which are obtained by giving particular orders among the blocks of I', as chambers of the restriction of $\mathcal{A}$ to the line $X_{I'}$, i.e., as one-dimensional faces. In other words, they arise as line segments or half lines. Moreover, the two of them whose closures contain the zero-dimensional face $P_\pi = X_I$, i.e., $<_f P_\pi$, are those two which do not contradict the order among the blocks of π, that is,

$$\pi' = (\{1\}, \{2, \ldots, l_1\}, \{l_1+1, \ldots, l_1+l_2\}, \ldots, \{m - l_{m-n}+1, \ldots, m\})$$

and

$$(\{2, \ldots, l_1\}, \{1\}, \{l_1+1, \ldots, l_1+l_2\}, \ldots, \{m - l_{m-n}+1, \ldots, m\}),$$

of which we take the former π'.

Next, consider $X_{I''} \in L$ which corresponds to the partition

$$I'' = \{\{1\}, \{2\}, \{3, \ldots, l_1\}, \{l_1+1, \ldots, l_1+l_2\}, \ldots, \{m - l_{m-n}+1, \ldots, m\}\}.$$

In a similar fashion, $X_{I''}$ is of rank $n-2$, or of dimension 2, and includes the line $X_{I'}$, that is, $X_{I''}$ is a plane including the line $X_{I'}$.

Some of the partial orderings occur which are obtained by giving particular orders among the blocks of I'', as chambers of the restriction of $\mathcal{A}$ to the plane $X_{I''}$, i.e., as two-dimensional faces. Moreover, the two of them whose closures include the one-dimensional face $P_{\pi'}$, i.e., $<_f P_{\pi'}$, are those two which do not contradict the order among the blocks of π', that is,

$$\pi'' = (\{1\}, \{2\}, \{3, \ldots, l_1\}, \{l_1+1, \ldots, l_1+l_2\}, \ldots, \{m - l_{m-n}+1, \ldots, m\})$$

and

$$(\{1\}, \{3, \ldots, l_1\}, \{2\}, \{l_1+1, \ldots, l_1+l_2\}, \ldots, \{m - l_{m-n}+1, \ldots, m\}),$$

of which we take the former π''.

Proceeding in the same way, we arrive, after $(l_1 - 1) + \cdots + (l_{m-n} - 1) = n$ steps, at the n-dimensional face $P_{\pi^{(n)}}$ with

$$\pi^{(n)} = (\{1\}, \ldots, \{l_1\}, \{l_1+1\}, \ldots, \{l_1+l_2\}, \ldots, \{m - l_{m-n}+1\}, \ldots, \{m\}),$$

which is the same as C_σ.

Next, we have to show that regions other than the ones stated in the theorem do not arise around X_I. In other words, there does not arise around X_I a region C_τ for which there exist i and j such that i is ranked better than j in π but the converse is true in τ. This can be easily verified by contradiction: Take a sequence in C_τ converging to X_I, and note that the mapping

$$\boldsymbol{x} \in R^n \mapsto \|\boldsymbol{x} - \boldsymbol{\mu}_i\| - \|\boldsymbol{x} - \boldsymbol{\mu}_j\| \in R$$

is continuous.

This completes the proof of Theorem 4.2.1

Q.E.D.

From an algorithmic viewpoint, we can enumerate all arising regions allowing repetitions by using Theorem 4.2.1—inspect all terminal nodes and list the regions around each of them. This is obvious because the closure of each region contains at least one terminal node, which, in turn, is seen by noting the following: In the non-degenerate case with $n \leq m-1$, for any set of $k < n$ discriminant hyperplanes $H_1, \ldots, H_k$ such that $r(H_1 \cap \cdots \cap H_k) = k$, there exists a hyperplane H_0 such that $r(H_0 \cap H_1 \cap \cdots \cap H_k) = k+1$.

4.3 Relation between ideal vectors and unbounded regions

As mentioned in Section 1.1, the set of orderings which can occur in ideal vector model coincides with the set of orderings corresponding to unbounded regions in our theory. More precisely, we have the following theorem.

Theorem 4.3.1. *Suppose we are given m distinct points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ in R^n. Then, for any direction $\boldsymbol{d} \in R^n, \|\boldsymbol{d}\| = 1$, such that $(\boldsymbol{d}, \overrightarrow{\boldsymbol{x}_i\boldsymbol{x}_j}) \neq 0, 1 \leq i < j \leq m$, we have that the two orderings σ_1 and σ_2 defined by*

$$\begin{aligned} \sigma_1^{-1}(i) &< \sigma_1^{-1}(j) \text{ iff } (\boldsymbol{d}, \boldsymbol{x}_i) > (\boldsymbol{d}, \boldsymbol{x}_j) \\ \sigma_2^{-1}(i) &< \sigma_2^{-1}(j) \text{ iff } \|t\boldsymbol{d} - \boldsymbol{x}_i\| < \|t\boldsymbol{d} - \boldsymbol{x}_j\| \end{aligned}$$

are equal for all sufficiently large $t > 0$.

Proof. We have

$$\|t\boldsymbol{d} - \boldsymbol{x}_i\|^2 < \|t\boldsymbol{d} - \boldsymbol{x}_j\|^2 \Longleftrightarrow -2t(\boldsymbol{d}, \boldsymbol{x}_i) + \|\boldsymbol{x}_i\|^2 < -2t(\boldsymbol{d}, \boldsymbol{x}_j) + \|\boldsymbol{x}_j\|^2.$$

Since $(\boldsymbol{d}, \boldsymbol{x}_i) \neq (\boldsymbol{d}, \boldsymbol{x}_j), i \neq j$ by assumption, for all sufficiently large $t > 0$, this is equivalent to

$$(\boldsymbol{d}, \boldsymbol{x}_i) > (\boldsymbol{d}, \boldsymbol{x}_j).$$

Q.E.D.

Remark 4.3.1. Part 1 of Theorem 3.1 follows immediately from Theorem 4.3.1 as well.

5 Appendix

Here we prove Lemma 2.2.

Proof of Lemma 2.2. First, note that in general, the set of points $\boldsymbol{x} \in R^n$ equidistant from $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_l \in R^n$:

$$\{\boldsymbol{x} \in R^n : \| \boldsymbol{x} - \boldsymbol{x}_1\|^2 = \cdots = \|\boldsymbol{x} - \boldsymbol{x}_l\|^2\}$$

is equal to the set of points $\boldsymbol{x}$ satisfying

$$\begin{pmatrix} \boldsymbol{x}_2' - \boldsymbol{x}_1' \\ \boldsymbol{x}_3' - \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_l' - \boldsymbol{x}_{l-1}' \end{pmatrix} \boldsymbol{x} = \frac{1}{2} \begin{pmatrix} \|\boldsymbol{x}_2\|^2 - \|\boldsymbol{x}_1\|^2 \\ \|\boldsymbol{x}_3\|^2 - \|\boldsymbol{x}_2\|^2 \\ \vdots \\ \|\boldsymbol{x}_l\|^2 - \|\boldsymbol{x}_{l-1}\|^2 \end{pmatrix}.$$

Now, without loss of generality, consider the partition

$$\{\{1, \ldots, l_1\}, \{l_1 + 1, \ldots, l_1 + l_2\}, \ldots, \{m - l_k + 1, \ldots, m\}\}, \tag{6}$$

with $l_1 \geq l_2 \geq \cdots \geq l_{\tilde{k}} \geq 2 > l_{\tilde{k}+1} = \cdots = l_k = 1$, and the corresponding system of linear equations:

$$\begin{pmatrix} \mu'_2 - \mu'_1 \\ \vdots \\ \mu'_{l_1} - \mu'_{l_1-1} \\ \mu'_{l_1+2} - \mu'_{l_1+1} \\ \vdots \\ \mu'_{l_1+l_2} - \mu'_{l_1+l_2-1} \\ \vdots \\ \mu'_{l_1+\cdots+l_{\tilde{k}-1}+2} - \mu'_{l_1+\cdots+l_{\tilde{k}-1}+1} \\ \vdots \\ \mu'_{l_1+\cdots+l_{\tilde{k}}} - \mu'_{l_1+\cdots+l_{\tilde{k}}-1} \end{pmatrix} x = \frac{1}{2} \begin{pmatrix} \|\mu_2\|^2 - \|\mu_1\|^2 \\ \vdots \\ \|\mu_{l_1}\|^2 - \|\mu_{l_1-1}\|^2 \\ \|\mu_{l_1+2}\|^2 - \|\mu_{l_1+1}\|^2 \\ \vdots \\ \|\mu_{l_1+l_2}\|^2 - \|\mu_{l_1+l_2-1}\|^2 \\ \vdots \\ \|\mu_{l_1+\cdots+l_{\tilde{k}-1}+2}\|^2 - \|\mu_{l_1+\cdots+l_{\tilde{k}-1}+1}\|^2 \\ \vdots \\ \|\mu_{l_1+\cdots+l_{\tilde{k}}}\|^2 - \|\mu_{l_1+\cdots+l_{\tilde{k}}-1}\|^2 \end{pmatrix} \quad (7)$$

Then the following two assertions hold.

(A) If $k \geq m - n$, the set of solutions to (7) is nonempty and when considered an element $X \in L$, its rank is given by $r(X) = m - k$.

(B) If $k < m - n$, there does not exist a solution to (7).

We first prove (A). The number of rows of the matrix A on the left hand side of (7) is

$$\begin{aligned} (l_1 - 1) + (l_2 - 1) + \cdots + (l_{\tilde{k}} - 1) &= (l_1 - 1) + (l_2 - 1) + \cdots + (l_k - 1) \\ &= m - k \leq n, \end{aligned}$$

and these row vectors are linearly independent by Assumption (A–1). Therefore, a solutions to (7) exists, and the dimension of the solution space is $n - m + k$. Thus, its codimension is

$$r(X) = n - (n - m + k) = m - k,$$

and (A) is proved.

Next we verify (B). The matrix A is of size $(m-k) \times n$ with $m-k > n$, so its rank is n by Assumption (A–1). On the other hand, the $(m-k) \times (n+1)$

matrix $(A,\ \frac{1}{2}\boldsymbol{b})$ has rank $n+1$, where $\boldsymbol{b}$ is the $(m-k)\times 1$ vector on the right hand side of (7). This follows because any collection of $n+1$ row vectors in $(A,\ \boldsymbol{b})$ is linearly independent by Assumption (A–2). Thus, there does not exist a solution to (7). This establishes (B).

Now, if we denote by X_I the set of solutions to the system of linear equations corresponding to partition I, we have that (A) implies that X_I is in L for each $I \in \mathcal{I}_{m-n}$, while (B) implies that the mapping from $\mathcal{I}_{m-n}$ to L is onto. Moreover, if $X_I = X_J$ for $I, J \in \mathcal{I}_{m-n}$, then $X_I = X_J = X_{I\vee J}$, where $I \vee J$ is the finest partition of which both I and J are refinements. Now, $r(X_I) = r(X_J) = r(X_{I\vee J})$ implies $I = J$. Therefore, the mapping $I \in \mathcal{I}_{m-n} \mapsto X_I \in L$ is one-to-one. In addition, it is obvious that $X_I \leq X_J$ iff I is a refinement of J. This proves Lemma 2.2.

Q.E.D.

Acknowledgment

We thank Prof. Hajime Yamato for valuable comments.

References

[1] J.F. Bennett and W.L. Hays, *Multidimensional unfolding: Determining the dimensionality of ranked preference data,* Psychometrika, 25 (1960), pp. 27–43.

[2] C. Berge, *Principes de Combinatoire,* DUNOD, 1968.

[3] C.H. Coombs, *A Theory of Data,* Wiley, 1964.

[4] D.E. Critchlow, *Metric Methods for Analyzing Partially Ranked Data,* Springer, 1985.

[5] W.S. DeSarbo and D.L. Hoffman, *Constructing MDS joint spaces from binary choice data: A multidimensional unfolding threshold model for marketing research,* Journal of Marketing Research, XXIV (1987), pp. 40–54.

[6] M.A. Fligner and J.S. Verducci (Eds.), *Probability Models and Statistical Analyses for Ranking Data,* Springer, 1993.

[7] W.L. Hays and J.F. Bennett, *Multidimensional unfolding: Determining configuration from complete rank order preference data,* Psychometrika, 26 (1961), pp. 221–238.

[8] R.D. Luce, *Individual Choice Behavior: A Theoretical analysis,* Wiley, 1959.

[9] R.D. Luce and H. Raiffa, *Games and Decisions,* Wiley, 1957.

[10] I.G. Macdonald, *Symmetric Functions and Hall Polynomials,* Oxford University Press, 1995.

[11] D.Q. Naiman and H.P. Wynn, *Inclusion-exclusion-Bonferroni identities and inequalities for discrete tube-like problems via Euler characteristics,* The Annals of Statistics, 20 (1992), pp. 43–76.

[12] A. Okabe, B. Boots, and K. Sugihara, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams,* Wiley, 1992.

[13] P. Orlik and H. Terao, *Arrangements of Hyperplanes,* Springer, 1992.

[14] R.T. Rockafellar, *Convex Analysis,* Princeton University Press, 1970.

[15] Y. Takane, *Analysis of contingency tables by ideal point discriminant analysis,* Psychometrika, 52 (1987), pp.493–513.

[16] Y. Takane, *Ideal point discriminant analysis and ordered response categories,* Behaviormetrika, 26 (1989), pp. 31–46.

[17] Y. Takane, *Ideal point discriminant analysis: Implications for multiway data analysis,* in *Multiway Data Analysis.* R. Coppi and S. Bolasco, editors, Elsevier Science Publishers B.V., pp. 287–299, 1989.

[18] Y. Takane, H. Bozdogan, and T. Shibayama, *Ideal point discriminant analysis,* Psychometrika, 52 (1987), pp. 371–392.

[19] G.L. Thompson, *Generalized permutation polytopes and exploratory graphical methods for ranked data,* The Annals of Statistics, 21 (1993), pp. 1401–1430.

[20] G.L. Thompson, *Graphical techniques for ranked data,* in *Probability Models and Statistical Analyses for Ranking Data.* M.A. Fligner and J.S. Verducci, editors, Springer, pp. 294–298, 1993.

[21] R. van Blokland-Vogelesang, *A nonparametric distance model for unidimensional unfolding,* in *Probability Models and Statistical Analyses for Ranking Data.* M.A. Fligner and J.S. Verducci, editors, Springer, pp. 241–276, 1993.

[22] R. Webster, *Convexity,* Oxford University Press, 1994.

[23] T. Zaslavsky, *Facing up to Arrangements: Face-Count Formulas for Partitions of Space by Hyperplanes,* American Mathematical Society, 1975.

[24] G.M. Ziegler, *Lectures on Polytopes,* Springer, 1995.