Arrangements and Ranking Patterns

Hidehiko Kamiya (Okayama Univ.)

Peter Orlik (Univ. of Wisconsin, Madison)

Akimichi Takemura (Univ. of Tokyo)

Hiroaki Terao (Hokkaido Univ.)

Items

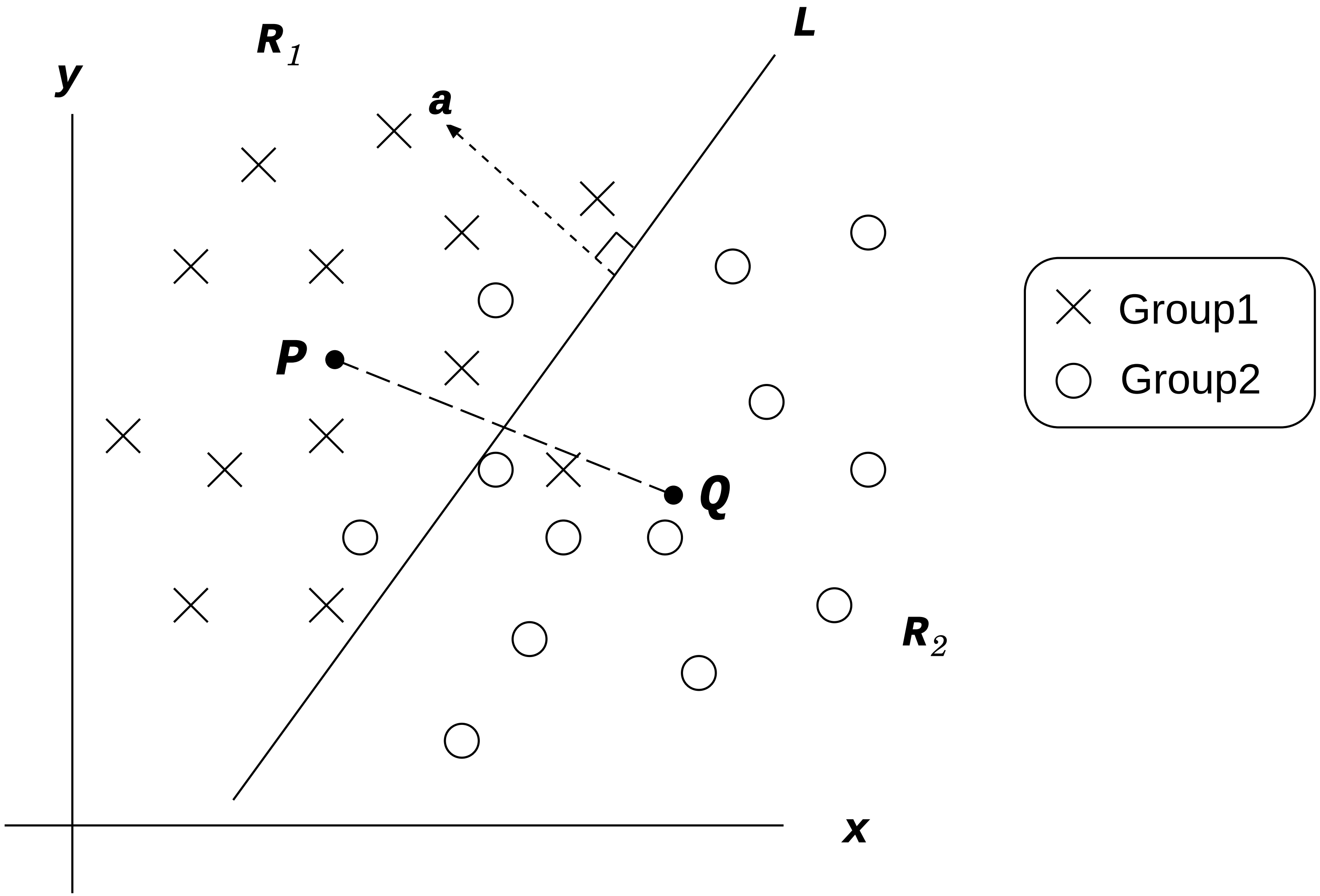
1. Discriminant analysis for two groups
2. Discriminant analysis for many groups and ranking models
3. The number of regions by the theory of hyperplane arrangements
4. Characterization of non-arising regions
5. Enumeration of ranking patterns for 1-dimensional case

Our contributions (see http://www.e.u-tokyo.ac.jp/˜takemura):

- Kamiya and Takemura (1997). On rankings generated by pairwise linear discriminant analysis of m populations. JMA.
- Kamiya and Takemura (2000). Rankings generated by spherical discriminant analysis. *Journal of the Japan Statistical Society.*
- Kamiya and Takemura (2005). Characterization of rankings generated by linear discriminant analysis. JMA.
- Kamiya, Orlik, Takemura and Terao (2006). Arrangements and Ranking Patterns. *Annals of Combinatorics.*

1 Discriminant analysis for two groups

- Divide the space into two separate groups
- Simplest discriminant method: Fisher's linear discriminant function
- When the covariance matrix $\Sigma = I$, Fisher's linear discrimination amounts to cutting the Euclidean space by "mid-hyperplane" or the "bisector" between the mean vectors of two populations.
- More complicated discrimination rules are used for practice, but we consider here the linear discrimination.

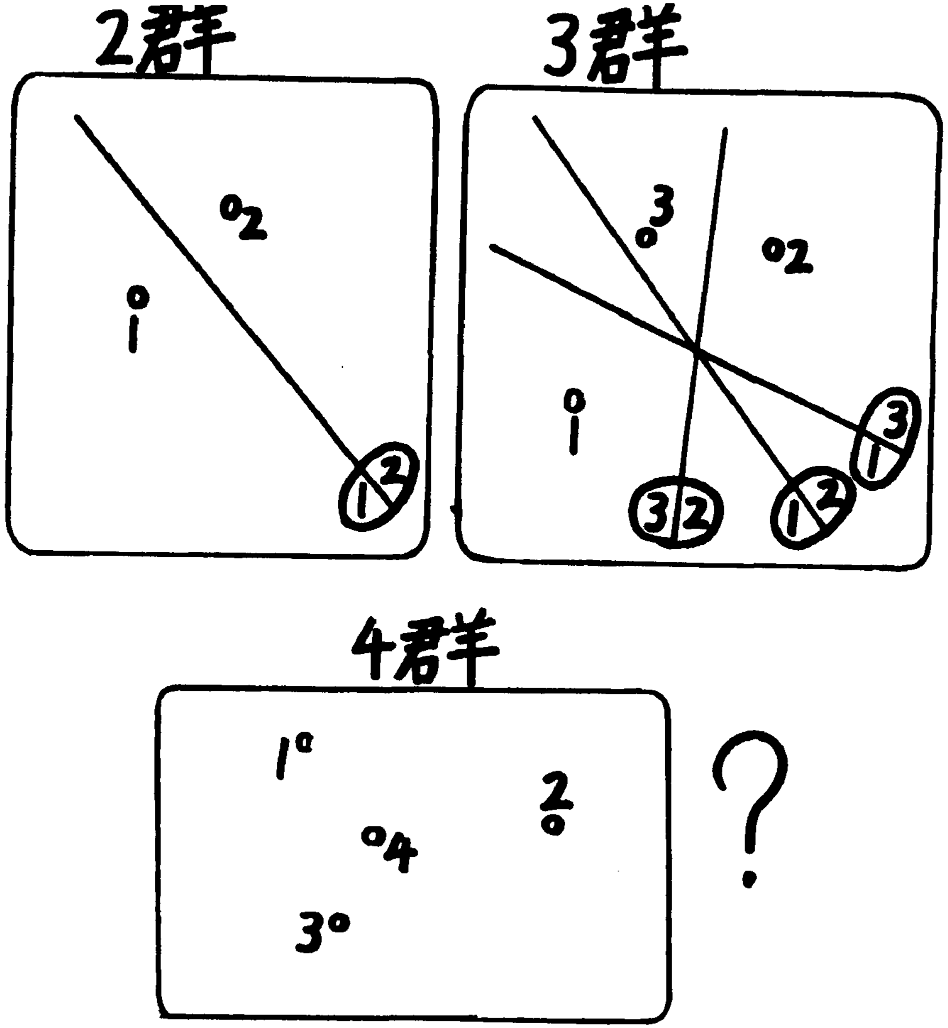


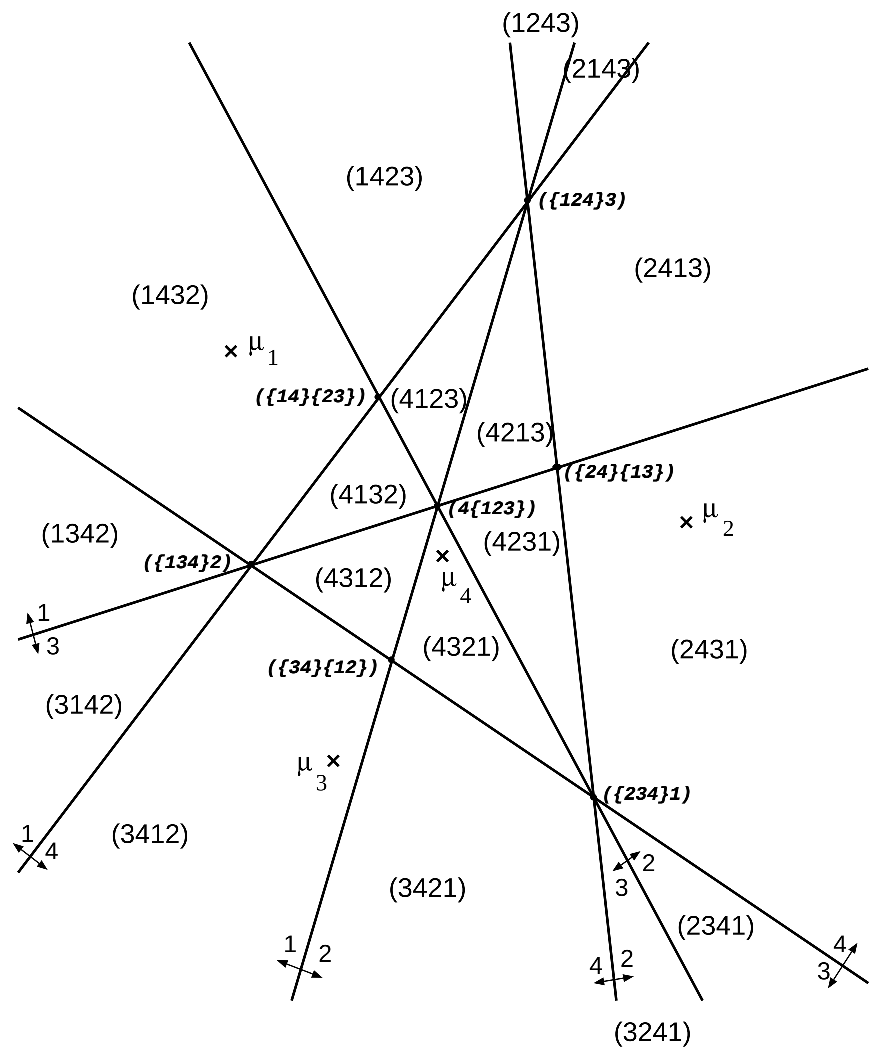
L
R1
y
a
Group1
Group2
P
Q
R2
x

2 Discriminant analysis for many groups and ranking models

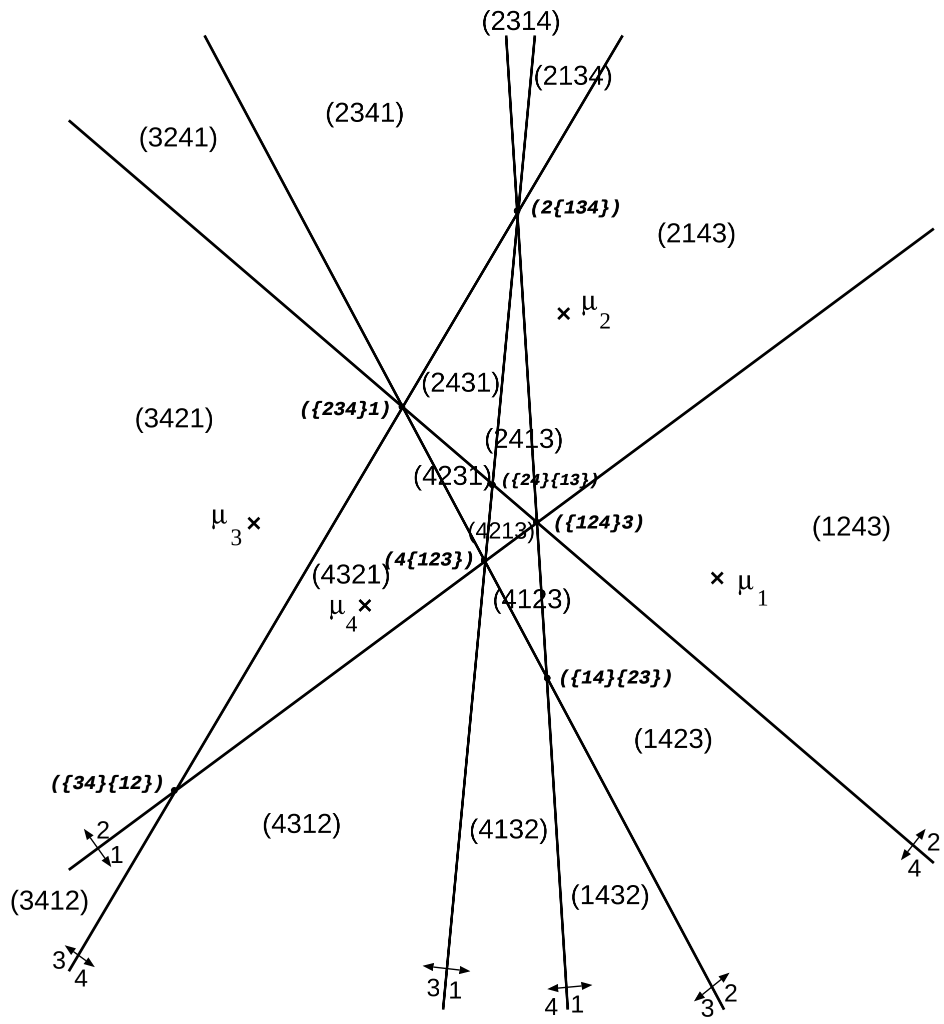
Discrimination of 3 groups

- Need to avoid the cyclic contradiction $a > b > c > a$.
- The cyclic contradiction happens if the three discriminant lines do not intersect at a single point.
- If we use the bisectors, then we do not have the contradiction.
- By the 3 bisectors, we obtain "rankings" of 3 populations, not just the most favored population.


2群
2
1
1 2
3群
3
2
1
3 2
1 2
3 1
4群
1
2
4
3
?



18 regions, no rankings with "4" as the last element (6 of them)



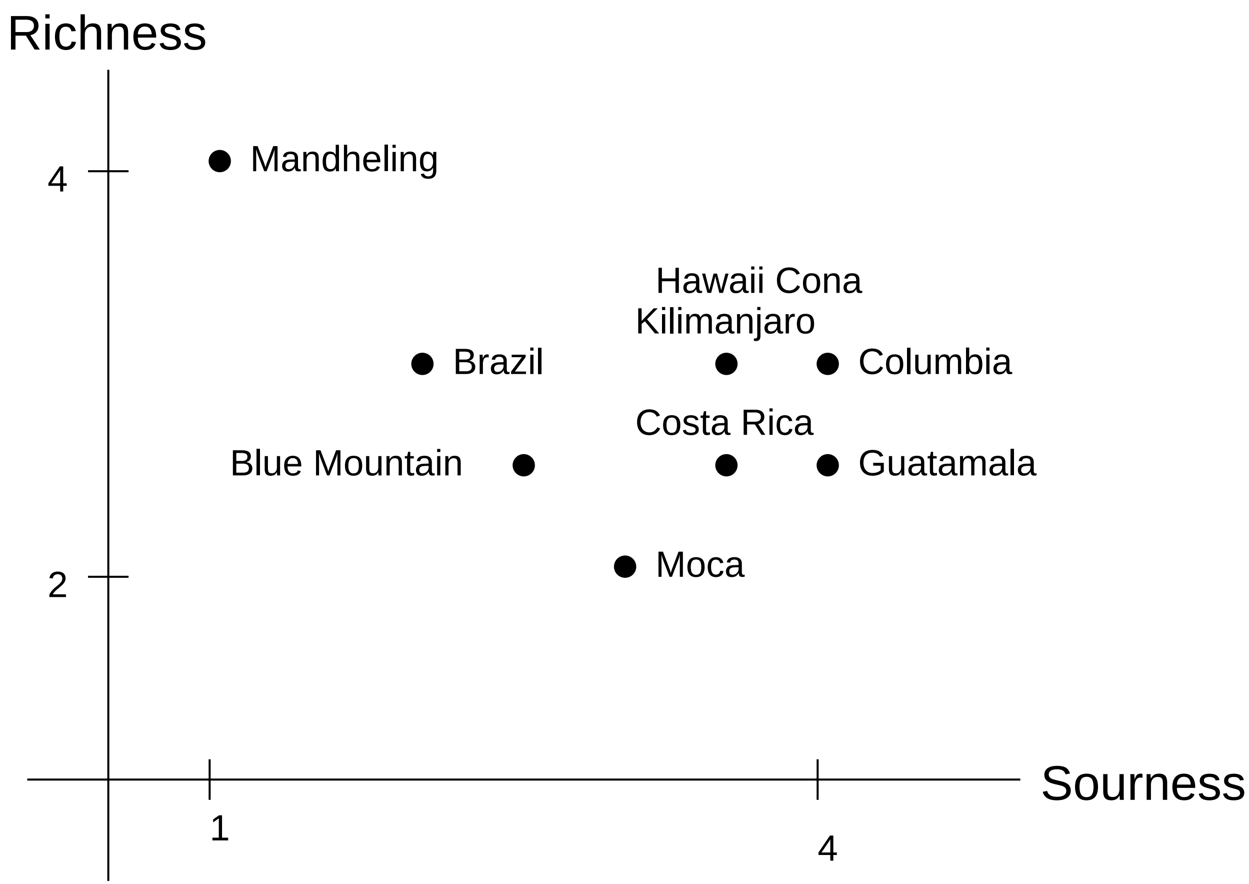
18 regions, non-arising: 1234, 1324, 1342, 3142, 3124, 3214

Ranking models

Table 1: Coffee brands

	sourness	richness	bitterness
Moca	3	2	3
Columbia	4	3	1
Brazil	2	3	3
Kilimanjaro	3.5	3	1
Costa Rica	3.5	2.5	1
Blue Mountain	2.5	2.5	2.5
Hawaii Cona	3.5	3	3.5
Mandheling	1	4	4
Guatemala	4	2.5	1

Evaluation in 5 levels



Ideal point model and Ideal vector model

3 The number of regions by the theory of hyperplane arrangements

The number of regions for different rankings can be evaluated using the theory of hyperplane arrangements (Orlik and Terao).

3.1 Preliminaries of hyperplane arrangements

- $V = R^n$
- $\mathcal{A}$: set of hyperplanes
- $L = L(\mathcal{A})$: set of non-empty intersection of hyperplanes
- Introduce a partial order into L : $X, Y \in L$

$$X \leq Y \Leftrightarrow Y \subset X$$

(V itself is the minimum element)

- $r(X) = \operatorname{codim}(X) = n - \dim X$
- $L_p = L_p(\mathcal{A}) = \{X \mid r(X) = p\}$
- Hasse diagram of the poset (partially ordered set) : draw a line between X and Y if $X \in L_p$, $Y \in L_{p+1}$, and $X < Y$.
- Möbius function ν:

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

(recursive definition)

- Poincaré polynomial

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

- Number of regions $= \pi(\mathcal{A}, 1)$
- Number of bonded regions $= (-1)^n \pi(\mathcal{A}, -1)$

3.2 The number of regions for multiple discrimination

- Poincaré polynomial for $m = 4$:

$$(1+t)(1+2t)(1+3t) = 1 + 6t + 11t^2 + 6t^3$$

- The number of arising rankings

$$1 + 6 = 7 \quad \text{in} \quad R^1$$

$$1 + 6 + 11 = 18 \quad \text{in} \quad R^2$$

$$1 + 6 + 11 + 6 = 24 = 4! \quad \text{in} \quad R^3$$

- General case with (m populations in R^n)
  - We assume that m population means are *in general position.* Let

$$
\begin{aligned}
&(1+t)(1+2t)\cdots(1+(m-1)t) \\
&\qquad = c_0 + c_1 t + c_2 t^2 + \cdots + c_{m-1} t^{m-1}
\end{aligned}
$$

  - Poincareé polynomial

$$
\pi(t) = c_0 + c_1 t + \cdots + c_n t^n \qquad (n \leq m-1)
$$

(cutting of higher order terms of “braid arrangement”)

– The number of regions

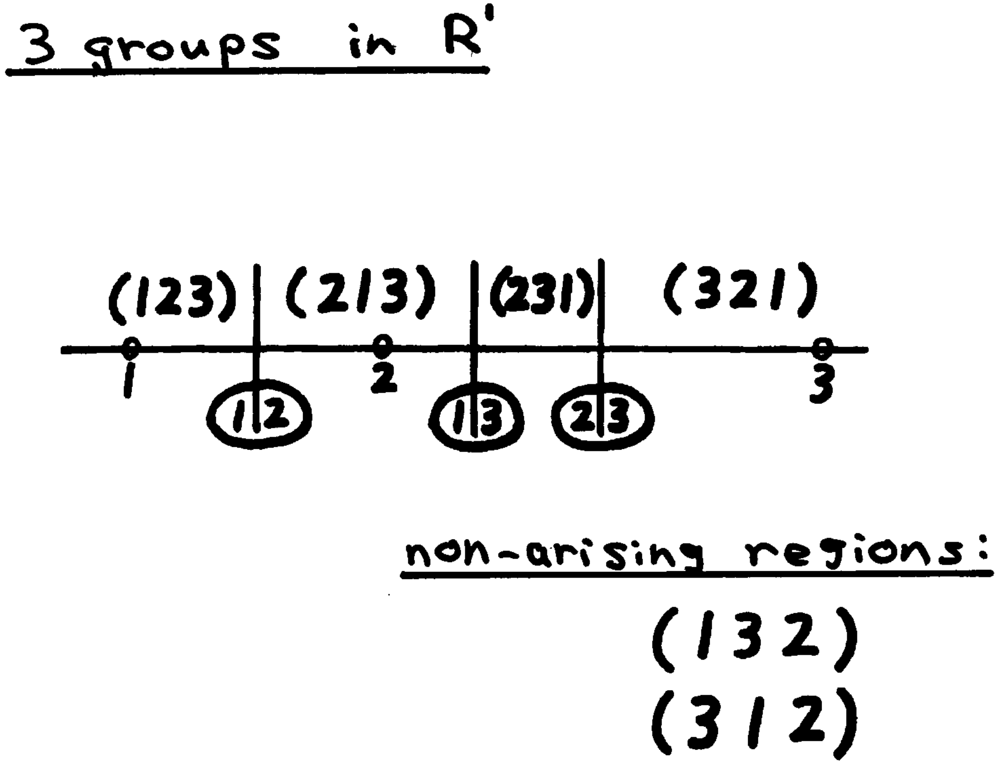
$$\pi(1) = c_0 + c_1 + \cdots + c_n$$

- The number of bounded regions:

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

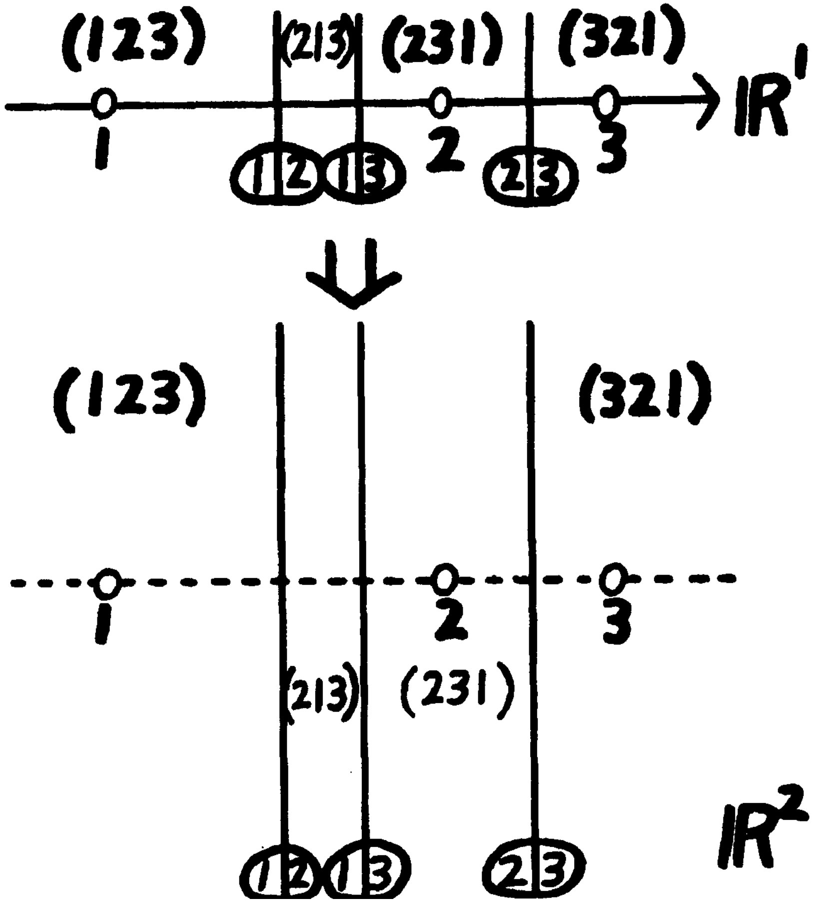
- For example, in the case of 4 populations in R^2, the number of bounded regions is $1 - 6 + 11 = 6$.
- If $n \geq m - 1$, all the regions are unbounded and all of $m!$ regions appear (i.e. braid arrangement)

4 Characterization of non-arising regions

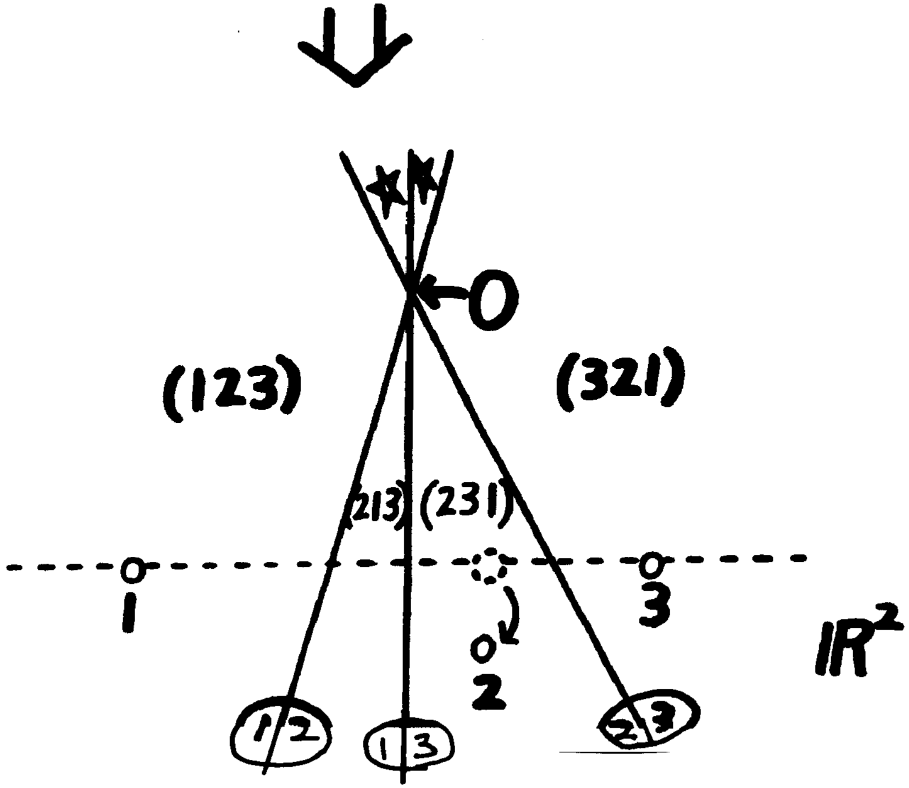


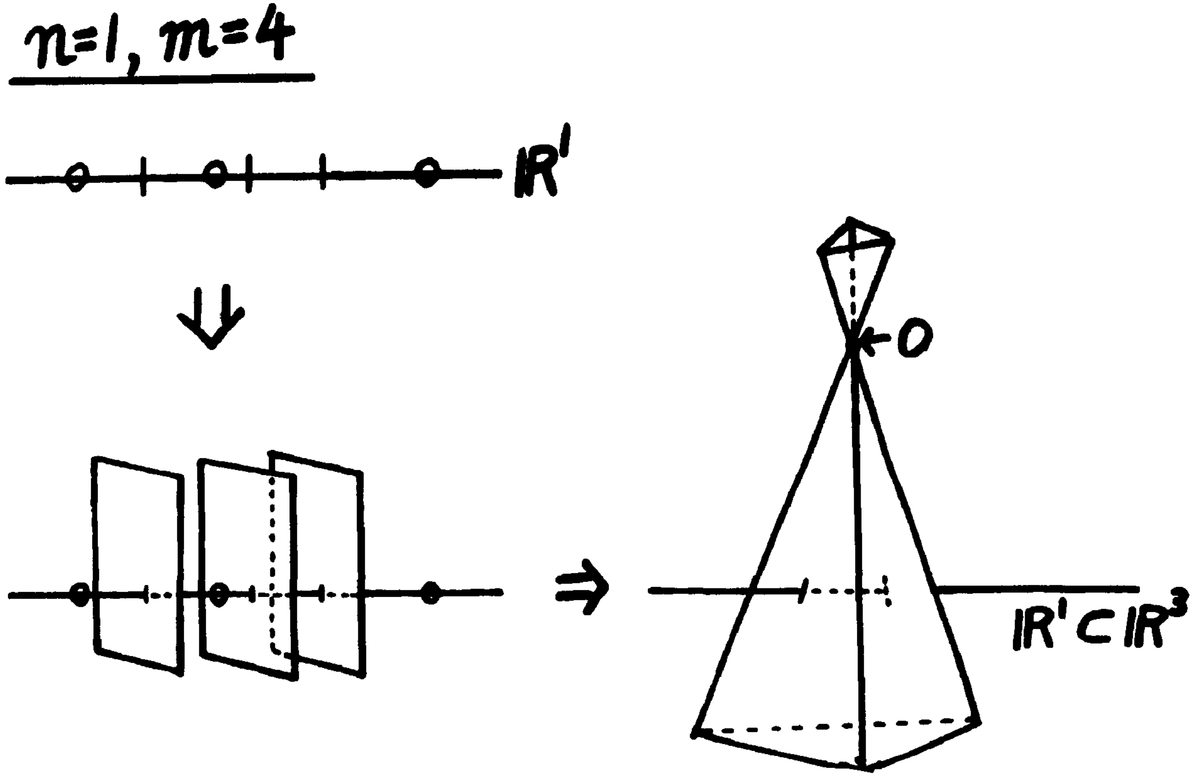
Now by "coning", regard this as

$n=1, \ m=3$



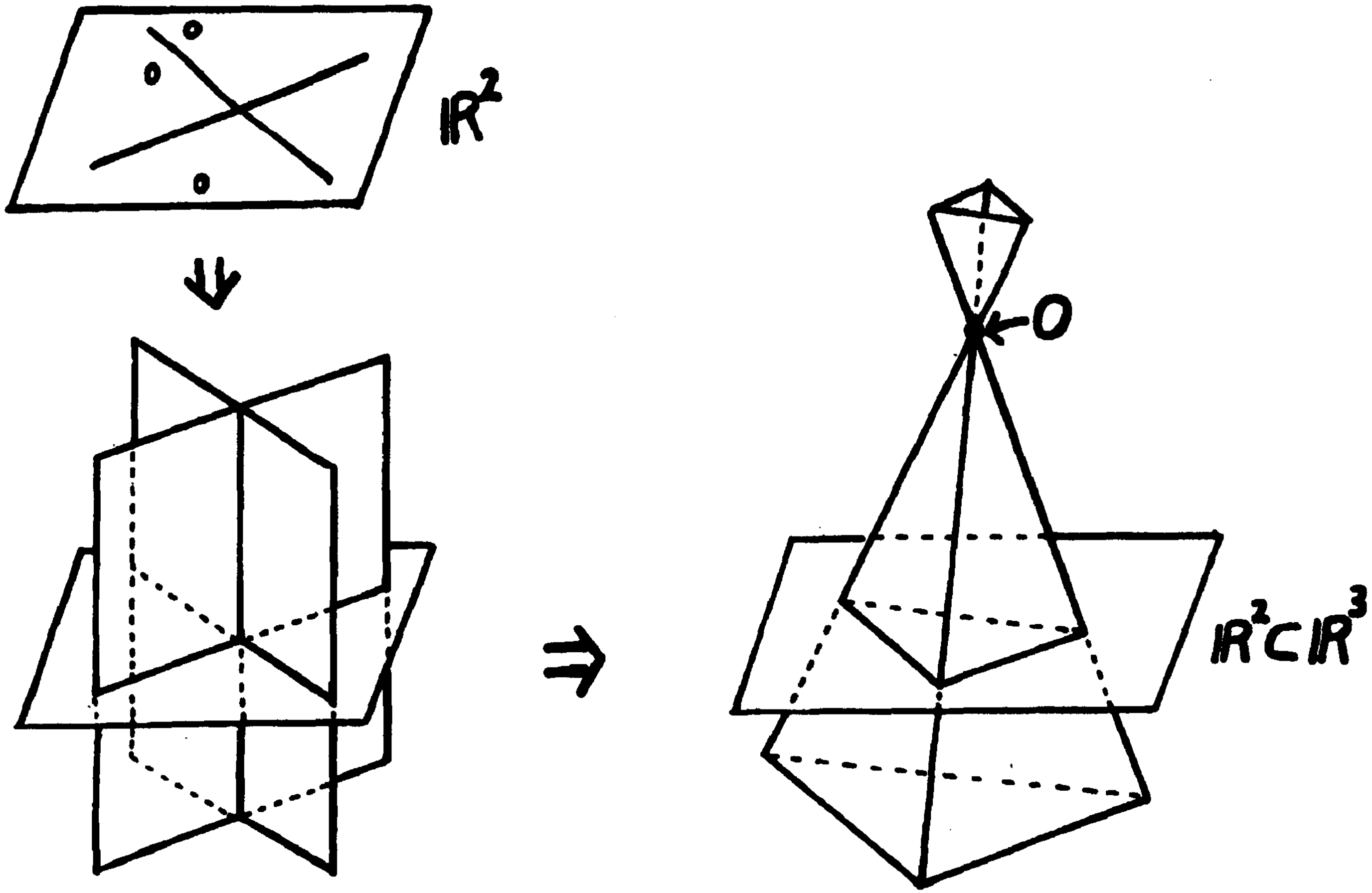
(123)
(213)
(231)
(321)
1
12
13
2
23
3
$\mathbb{R}^1$
(123)
(321)
1
2
3
(213)
(231)
$\mathbb{R}^2$
12
13
23


O
(123)
(321)
(213)
(231)
1
3
2
$\mathbb{R}^2$
12
13
23



n=1, m=4
ℝ¹
0
ℝ¹ ⊂ ℝ³

$n=2,\ m=4$



From the above consideration, the following results are obtained.

- If a ranking $(i_1, \ldots, i_m)$ appears as an unbounded region, then the reverse ranking $(i_m, \ldots, i_1)$ also appears as an unbounded region in the opposite side of R^n.
- If a ranking $(i_1, \ldots, i_m)$ appears as a bounded region, then the reverse ranking $(i_m, \ldots, i_1)$ does not appear.

(These results are from Kamiya and Takemura (1997)JMA. Recently we found that it has many overlaps with an earlier paper by I.J.Good and T.N.Tideman (1977), *Journal of Combinatorial Theory.*)

More recently, Kamiya and Takemura (2005) JMA have proved the following simple but basic facts.

- "Lumping" and "Spreading" of vectors
  - $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_k \in R^n$ are *lumped together* if $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_k$ are contained in a half-space determined by a hyperplane through the origin.
  - $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_k \in R^n$ are *spread out* if

$$c_1 \boldsymbol{\nu}_1 + \cdots + c_k \boldsymbol{\mu}_n = 0$$

for some $(c_1, \ldots, c_k) \neq (0, \ldots, 0)$.

- **Gordan's theorem**: For any given $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_M \in R^N$, one and only one of the following two alternatives holds: (i) $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_M$ are lumped together; (ii) $\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_M$ are spread out.

- Let $\tilde{\boldsymbol{\mu}}_i = (\boldsymbol{\mu}_i^T, -\|\boldsymbol{\mu}_i\|^2/2)^T,\ i = 1, \ldots, m,.$

- Assume that $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_m)$ and $(\tilde{\boldsymbol{\mu}}_1, \ldots, \tilde{\boldsymbol{\mu}}_m)$ are in general position.

- By permuting the indices if necessary, we consider whether the ranking $(1, 2, \ldots, m)$ appears.

Fist we have the following result:

- Case 1: If $(\tilde{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{\mu}}_2, \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_3, \ldots, \tilde{\boldsymbol{\mu}}_{m-1} - \tilde{\boldsymbol{\mu}}_m)$ are spread out, then neither $(1, 2, \ldots, m)$ nor $(m, m-1, \ldots, 1)$ appear.

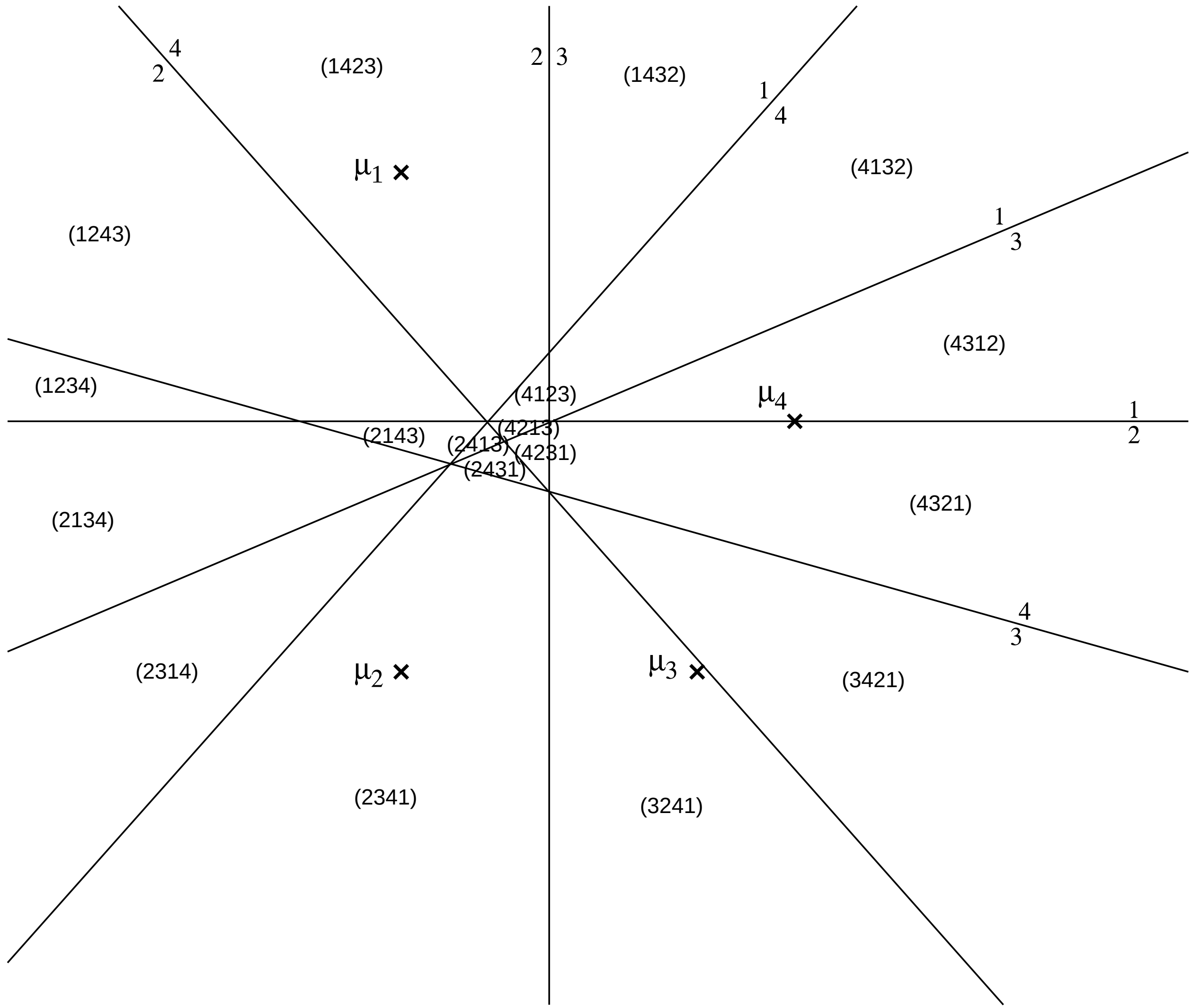
If $(\tilde{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{\mu}}_2, \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_3, \ldots, \tilde{\boldsymbol{\mu}}_{m-1} - \tilde{\boldsymbol{\mu}}_m)$ are lumped together, we need to further consider whether $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3, \ldots, \boldsymbol{\mu}_{m-1} - \boldsymbol{\mu}_m)$ are lumped together or spread out.

- If $(\tilde{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{\mu}}_2, \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_3, \ldots, \tilde{\boldsymbol{\mu}}_{m-1} - \tilde{\boldsymbol{\mu}}_m)$ are lumped together
  - Case 2a: If $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3, \ldots, \boldsymbol{\mu}_{m-1} - \boldsymbol{\mu}_m)$ are lumped together, then $(1, 2, \ldots, m)$ appears as an unbounded region. Hence in this case $(m, m-1, \ldots, 1)$ also appears as an unbounded region.

– Case 2b: If $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3, \ldots \boldsymbol{\mu}_{m-1} - \boldsymbol{\mu}_m)$ are spread out, then only one of $(1, 2, \ldots, m)$ or $(m, m-1, \ldots, 1)$ appears as a bounded region. The sign of the normal vector to the hyperplane containing $\tilde{\boldsymbol{\mu}}_1 - \tilde{\boldsymbol{\mu}}_2, \tilde{\boldsymbol{\mu}}_2 - \tilde{\boldsymbol{\mu}}_3, \ldots \tilde{\boldsymbol{\mu}}_{m-1} - \tilde{\boldsymbol{\mu}}_m$ on one side determines which of the two appears.

- In general, the characterization of non-arising rankings remain difficult.

- For example, consider the case R^2 and $m = 4$. For a long time we thought that only two cases on p.8 and p.9 are possible. However recently we found another pattern!



non-arising regions: (1324),(1342), (3124),(3142), (3214),(3412)

5 Enumeration of ranking patterns for 1-dimensional case

Question 1 :

- Determine the sets of all ranking patterns, which are obtained from multiple discrimination of m populations in R^n.
- Determine whether a given set of rankings is obtained from some multiple discrimination of m populations in R^n.
- If the answer is yes, determine the set of mean vectors of m populations in R^n.

[These are very hard]

Question 2 :
Give the number of ranking patters: number of regions in “moduli space”.

- We have obtained substantial results for 1-dimensional case.
  $\rightarrow$ “Unfolding model” important in mathematical psychology.
- The same question for 2-dimension or higher remains to be very difficult.

- Consider set of rankings determined by mid-points of m points on R^1.
- Identify the m points in R^1 with a vector in R^m. Then the order relation between two midpoints corresponds to a hyperplane in R^m:

$$\frac{x_1 + x_2}{2} < \frac{x_3 + x_4}{2} \qquad \text{etc.}$$

- A set of rankings corresponds to a chamber of the the following "midhyperplane arrangement" (except for the permutation of indices):

$$\prod_{1 \le i < j \le m} (x_i - x_j) \prod_{\substack{p,q,r,s: \\ \text{distinct}}} (x_p + x_q - x_r - x_s)$$

- Let $\mathcal{A}_m$ denote the midhyperplane arrangement in R^m.
- $m = 3$
  - $\pi(\mathcal{A}_3, t) = t(1+t)(1+2t)$ (“exponents”:$\{0, 1, 2\}$)
  - $\pi(\mathcal{A}_3, 1) = 6,\qquad 6/3! = 1$
- $m = 4$
  - $\pi(\mathcal{A}_4, t) = t(1+t)(1+3t)(1+5t)$ (exponents:$\{0, 1, 3, 5\}$)
  - $\pi(\mathcal{A}_4, 1) = 48,\qquad 48/4! = 2$

- $m = 5$
  - $\pi(\mathcal{A}_5, t) = t(1+t)(1+7t)(1+8t)(1+9t)$ (exponents:$\{0, 1, 7, 8, 9\}$)
  - $\pi(\mathcal{A}_5, 1) = 1440, \qquad 1440/5! = 12$
  - So far feasible by hand and found in psychometric literature.
- $m = 6$
  - $\pi(\mathcal{A}_6, t) = t(1+t)(1+13t)(1+14t)(1+15t)(1+17t)$ (exponents:$\{0, 1, 13, 14, 15, 17\}$)
  - $\pi(\mathcal{A}_5, 1) = 120960, \qquad 120960/5! = 168$
  - Done by Stephen Szydlik and Takemura by computer.

Strategy : Consider hyperplane arrangement in the vector space over finite fields.

- Let q be a sufficiently large prime number and consider m-dimensional vector space $\mathbb{Z}_q^m$ over $\mathbb{Z}_q$. Then the poset of the midhyperplane arrangements in $\mathbb{Z}_q^m$ is isomorphic to the poset of the midhyperplane arrangements in R^m. (Terao).
- The Poincaré polynomial of the midhyperplane arrangement in $\mathbb{Z}_q^m$ is given by the number of points of $\mathbb{Z}_q^m$, which are not contained in any hyperplane of the arrangement. (Terao).
- For $m = 7$, $q \geq 72$ is sufficiently large. Therefore we can just loop through 72^7 steps of counting (feasible by computer).
- For $m = 8$, $q \geq 216$ is large enough. Looping 216^8 times requires some tweaking of computer programs.

- $m = 7$:
  - $\pi(\mathcal{A}_7, t) = t(1+t)(1+23t)(1+24t)(1+25t)(1+26t)(1+27t)$ (exponents:$\{0, 1, 23, 24, 25, 26, 27\}$)
  - $\pi(\mathcal{A}_7, 1)/7! = 4680$
- $m = 8$:

$$\begin{aligned}\pi(\mathcal{A}_8, t) &= t(1+t)(1+35t)(1+37t)(1+39t)(1+41t)\\ &\quad (1+85t+1926t^2)\end{aligned}$$

  - $\pi(\mathcal{A}_8, 1)/8! = 229368$

($m = 9$ has been done, but takes a long time on PC.)

Some qualitative differences for the $m \leq 7$ and the case $m \geq 8$.

- For $m \geq 8$, the Poincaré polynomial does not factor into linear terms (: non-existence of exponents). This can be verified by looking at the coefficient of the t^2 term:

$$2\binom{m}{3} + 15\binom{m}{4} + 120\binom{m}{5} + 375\binom{m}{6} + 630\binom{m}{7} + 315\binom{m}{8}$$

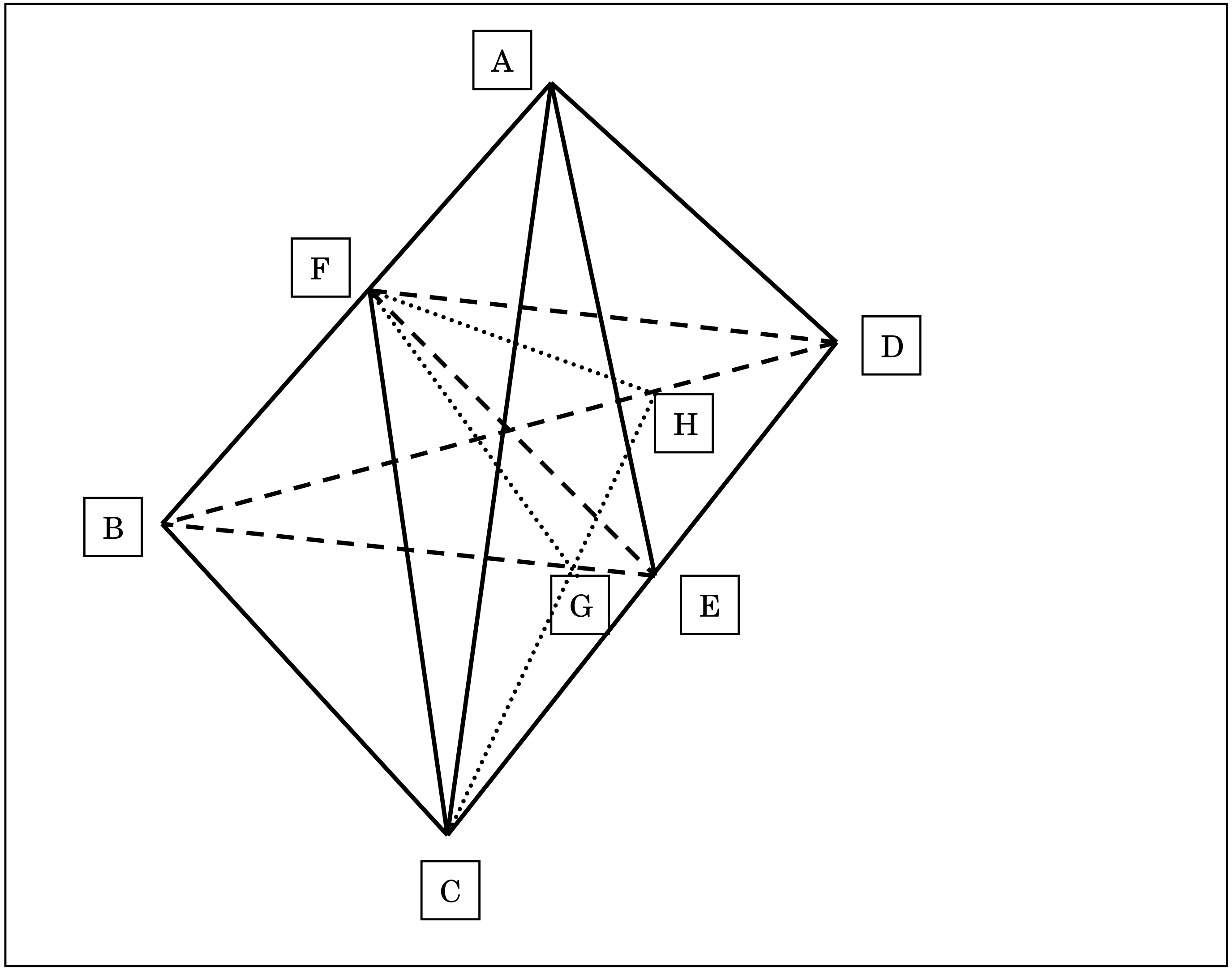
- Let $a_n = n(n^{n-1} - 1)((n-2)!)/(n-1)$. Then for some unknown reason

$$\pi(\mathcal{A}_m, 1)/m! = a_{m-2}$$

holds for $m \leq 7$. But this equality does not hold for $m = 8$.

How do the regions look like for $m = 5$?

- Let $\mathbf{x} = (x_1, \ldots, x_5) \in \mathbb{R}^5$ and consider the midhyperplane arrangement in R^5. $\mathbb{R}^5$ is divided into $\pi(\mathcal{A}_5, 1) = 1440$ cones.
- Each cone contains the line $x_1 = \cdots = x_5$ and we can restrict out attention to the orthogonal complement of $(1, \ldots, 1)$.
- Furthermore consider the intersection with the unit sphere. Then the 3-dimensional unit sphere $\mathbb{S}^3$ is divided into 1440 regions.
- There are only 6 different shapes of the regions.
- Note that not all of the regions are simplicial. One is a quadrangular cone.


A
F
D
H
B
G
E
C

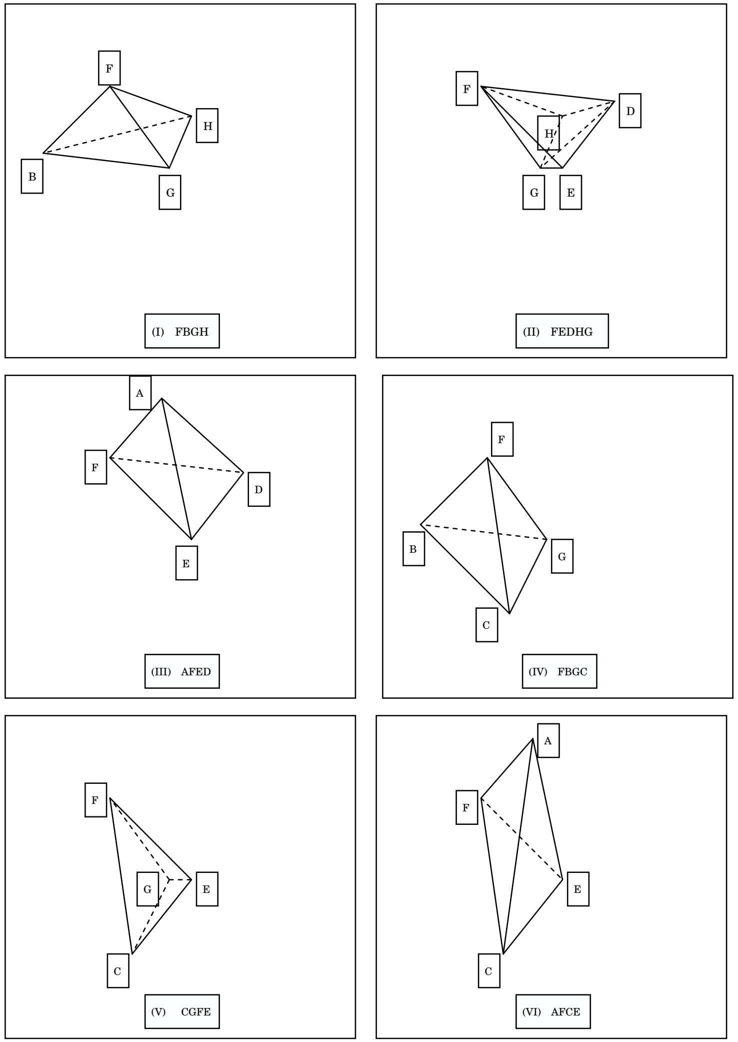

F
H
B
G
(I) FBGH
F
D
H
G
E
(II) FEDHG
A
F
D
E
(III) AFED
F
B
G
C
(IV) FBGC
F
G
E
C
(V) CGFE
A
F
E
C
(VI) AFCE

Figure 2

Summary and concluding remarks

- Ranking models arise when we combine pairwise discriminant analyses.
- When the dimension is small compared to number of points, there are restrictions on the number of rankings.
- For Fisher's linear discriminant functions, these questions can be handled by the theory of hyperplane arrangements.
- Enumeration of Ranking patterns (when the mean vectors move) are hard.
- Non-linear discriminant functions are hard.