Arrangements and statistics, Talk No.1: Ranking and arrangements

Akimichi Takemura

University of Tokyo

August 4, 2009
Big thanks and some excuses

1. I am very much grateful, especially to Prof. Hiroaki Terao, for inviting me here and for the wonderful organization of this conference.

2. I am honored to give three talks on arrangements and statistics to this fantastic audience.

3. Actually, three talks are too much, as I already told Prof. Terao two years ago. I hoped to achieve more in the last two years...

Plan of three talks

1. Ranking and arrangements
   (with H. Kamiya)

2. Mid-hyperplane arrangement and finite-field method
   (with H. Kamiya, H. Terao and P. Orlik)

3. Other uses of hyperplane arrangement theory in statistics
   (short survey of works by other statisticians and algebraists)
References for Talk No.1


(Personally, I sent our 1997 paper to Prof. Terao in the spring of 2002. Then our collaboration started. We found out Good and Tideman’s paper later.)
1. Introduction

2. Number of rankings

3. Ranking patterns

4. Concluding remarks
Introduction
The unfolding model (Coombs (1950)).
A model for preference rankings.
Equivalent to “discriminant analysis” of many populations in statistics. The same structure is found in other fields.
- Spatial model in voting theory.
- Ideal point model in marketing science.
- Higher order Voronoi diagram.
Objects 1, 2, \ldots, m.

An individual ranks these m objects. The objects 1, 2, \ldots, m are represented by $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$.

The individual is represented by $y \in \mathbb{R}^n$ (his/her ideal point).

$\mathbb{R}^n$: the joint space.
• Preference: The nearer, the more preferred, i.e.,

\[ y \text{ prefers } i \text{ to } j \iff \|y - x_i\| < \|y - x_j\|. \]

• \( y \) has ranking \((i_1, i_2, \ldots, i_m)\) iff

\[ \|y - x_{i_1}\| < \|y - x_{i_2}\| < \cdots < \|y - x_{i_m}\|. \]
### Table: Coffee brands

<table>
<thead>
<tr>
<th></th>
<th>sourness</th>
<th>richness</th>
<th>bitterness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moca</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Columbia</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Brazil</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Kilimanjaro</td>
<td>3.5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>3.5</td>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>Blue Mountain</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Hawaii Cona</td>
<td>3.5</td>
<td>3</td>
<td>3.5</td>
</tr>
<tr>
<td>Mandheling</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Guatemala</td>
<td>4</td>
<td>2.5</td>
<td>1</td>
</tr>
</tbody>
</table>

Evaluation in 5 levels
Example
Admissible/inadmissible rankings

- In general, we can think of \( m! \) rankings among \( m \) objects.

- Not all \( m! \) rankings are generated.

\[
\begin{align*}
\exists y \in \mathbb{R}^n \text{ s.t. } \|y - x_{i_1}\| < \cdots < \|y - x_{i_m}\| & \rightarrow (i_1 i_2 \cdots i_m) : \text{admissible}; \\
\n\not\exists y \in \mathbb{R}^n \text{ s.t. } \|y - x_{i_1}\| < \cdots < \|y - x_{i_m}\| & \rightarrow (i_1 i_2 \cdots i_m) : \text{inadmissible}.
\end{align*}
\]
A very simple example: \( m = 3, \ n = 1 \)

- 3 objects (points) in \( \mathbb{R}^1 \).

- Inadmissible rankings: (132), (312).
- The object 2 in the middle cannot be the least preferred.
When the dimension is enough: $m = 3$, $n = 2$

- 3 points in $\mathbb{R}^2$.

- All rankings ($3! = 6$) appear.
Number of rankings?

- If $m - 1 \leq n$, then all rankings appear.
- Not all rankings appear because we are restricting many points on a low dimensional space. But, otherwise we suppose that the points are in “general position” in $\mathbb{R}^n$ (explained more later).

**Problem 1**

What is the number of admissible rankings?
(for $m > n + 1$)

- Common for all generic $m$ points in $\mathbb{R}^n$. 
Number of rankings
Recall

\[ y \text{ prefers } i \text{ to } j \iff \| y - x_i \| < \| y - x_j \|. \]

RHS is equivalent to \( y \) being on one side (the same side as \( x_i \)) of the perpendicular bisector of \( x_i x_j \).

\( y \) prefers \( i \) to \( j \):

\[
\begin{array}{c}
\bullet \\
\text{RHS is equivalent to } y \text{ being on one side (the same side as } x_i ) \text{ of the perpendicular bisector of } x_i x_j .
\end{array}
\]
For each pair $i, j \ (1 \leq i < j \leq m)$, consider the perpendicular bisector of $\overline{x_i x_j}$:

$$H_{ij} := \{ y \in \mathbb{R}^n \mid \| y - x_i \| = \| y - x_j \| \}$$

$$= \{ y \in \mathbb{R}^n \mid (x_i - x_j)^T (y - \frac{x_i + x_j}{2}) = 0 \}$$

$$= \{ y \in \mathbb{R}^n \mid (x_i - x_j)^T y = \frac{1}{2} (\| x_i \|^2 - \| x_j \|^2) \}.$$
Define a hyperplane arrangement

\[ \mathcal{A}_{m,n} = \mathcal{A}_{m,n}(x_1, \ldots, x_m) := \{ H_{ij} \mid 1 \leq i < j \leq m \} \]

in \( \mathbb{R}^n \).

Call \( \mathcal{A}_{m,n} \) the **unfolding arrangement**.

\( \mathcal{A}_{m,n} = \{ H_{ij} \mid 1 \leq i < j \leq m \} \) cuts \( \mathbb{R}^n \) into **chambers** (i.e., connected components of the complement \( \mathbb{R}^n \setminus \bigcup_{H_{ij} \in \mathcal{A}_{m,n}} H_{ij} \)).
$\mathbb{R}^2$, $m = 4$

(18 chambers with 6 bounded)
Unfolding arrangement

- Each chamber is of the form

\[ C_{i_1 \ldots i_m} := \{ y \in \mathbb{R}^n | \|y - x_{i_1}\| < \|y - x_{i_2}\| < \cdots < \|y - x_{i_m}\| \} \neq \emptyset \]

for an admissible ranking \((i_1 i_2 \cdots i_m)\).
Admissible ranking $\leftrightarrow$ Chamber of $\mathcal{A}_{m,n}$

- We have
  
  $y \in \mathbb{R}^n$ gives ranking $(i_1 \cdots i_m) \iff y \in C_{i_1 \cdots i_m}$.

- One-to-one correspondence:

  \[
  \{ \text{admissible rankings} \} \leftrightarrow \{ \text{chambers of } \mathcal{A}_{m,n} \},
  \]
  
  \[
  (i_1 i_2 \cdots i_m) \leftrightarrow C_{i_1 i_2 \cdots i_m}.
  \]
Unfolding arrangement

Problem 1 ⇔ Counting chambers

Problem 1:
What is the number of admissible rankings?
→ Count the chambers of $A_{m,n}$. 
Answer to Problem 1

Theorem 1

(Kamiya and Takemura (1997, 2005)) The number of chambers of a generic $A_{m,n}$ is

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

where

$$(1 + t)(1 + 2t) \cdots (1 + (m - 1)t) = c_0 + c_1 t + \cdots + c_{m-1} t^{m-1}.$$
Answer to Problem 1

Remarks:

- (1) depends only on $m$ and $n$, and not on $x_1, \ldots, x_m$.
- When $n \geq m - 1$, (1) is equal to $m!$.
- "Generic points" in $\mathbb{R}^n$ have to satisfy several conditions. For example, no two bisectors are parallel, or no more than $n + 1$ points are on a common sphere.
- Precise conditions are given in Kamiya and Takemura (2005).
- These degeneracy conditions are non-linear for $n \geq 2$. 
Example: $m = 4$

\[(1 + t)(1 + 2t)(1 + 3t) = 1 + 6t + 11t^2 + 6t^3.\]

\[
\downarrow
\]

\[
\begin{cases}
  n = 1: & 1 + 6 = 7 = 1 + \binom{4}{2}, \\
  n = 2: & 1 + 6 + 11 = 18, \\
  n \geq 3: & 1 + 6 + 11 + 6 = 24 = 4!.
\end{cases}
\]

- Taking alternating sum gives the number of bounded regions.
The proof of Theorem 1 is based on the following proposition.

Let

\[\begin{align*}
L(A_{m,n}) & : \text{the intersection poset of } A_{m,n}, \\
\Pi_m & : \text{the partition lattice of } \{1, \ldots, m\}.
\end{align*}\]

Then we have:

**Proposition 1**

(Kamiya and Takemura (1997))

\[L(A_{m,n}) \cong \text{the rank } n \text{ truncation of } \Pi_m \]
\[= \text{partitions of } \{1, 2, \ldots, m\} \text{ with at least } m - n \text{ blocks}.\]
$\mathbb{IR}^2$, $m = 4$

$\mathbb{IR}^2$: $\{\{1\}, \{2\}, \{3\}, \{4\}\}$

$\{\{1, 2\}, \{3\}, \{4\}\}$

$\{\{1\}, \{2, 4\}\}$

$\{\{1, 3\}, \{2, 4\}\}$

$\{\{1, 3\}, \{2\}\}$

$\{\{1, 2\}, \{3\}, \{4\}\}$

$\{\{2, 3\}, \{4\}\}$

$\{\{2, 3\}, \{4\}\}$

$\{\{1, 2\}, \{3, 4\}\}$

$\{\{1, 2, 3\}, \{4\}\}$

$\{\{1, 2\}, \{3\}, \{4\}\}$

$\{\{1, 2, 3\}, \{4\}\}$

$\{\{1, 2\}, \{3\}, \{4\}\}$

$\{\{1, 2\}, \{3\}, \{4\}\}$

$\{\{1, 2\}, \{3\}, \{4\}\}$
$L(A_4, 2)$:

![Diagram of number of rankings](image)
Now we can use Zaslavsky’s result.

\[ \#\{\text{chambers of } \mathcal{A}_{m,n}\} = \pi(\mathcal{A}_{m,n}, 1). \]

where \( \pi(\mathcal{A}_{m,n}, t) \) is the Poincaré polynomial of \( \mathcal{A}_{m,n} \).
Number of chambers of $\mathcal{A}_{m,n}$

- Write $L = L(\mathcal{A}_{m,n})$.
- Each $X \in L$ is indexed by a partition of $\{1, \ldots, m\}$:
  \[ X \in L \leftrightarrow I_X := \{1, \ldots, m\}/\sim_X, \]
  where
  \[ i \sim_X j \iff X \subseteq H_{ij} \]
  \[ (H_{ii} := \mathbb{R}^n \text{ and } H_{ij} := H_{ji} \text{ for } i > j). \]
- For $X, Y \in L$,
  \[ X \leq Y \iff I_X \text{ is a refinement of } I_Y. \]
Number of chambers of $A_{m,n}$

**Example:** $m = 4 \ (n \geq 2)$

\[
\begin{align*}
X &= H_{12} \leftrightarrow l_X = \{\{1, 2\}, \{3\}, \{4\}\}, \\
Y &= H_{12} \cap H_{23} \cap H_{13} (= H_{12} \cap H_{23}, \text{e.g.}) \\
    &\leftrightarrow l_Y = \{\{1, 2, 3\}, \{4\}\}, \\
Z &= H_{12} \cap H_{34} \\
    &\leftrightarrow l_Z = \{\{1, 2\}, \{3, 4\}\}.
\end{align*}
\]

$X < Y \leftrightarrow l_X$ is a refinement of $l_Y$, \\
$X < Z \leftrightarrow l_X$ is a refinement of $l_Z$. 
Number of chambers of $A_{m,n}$

- $\Pi_m$: the **partition lattice** of $\{1, 2, \ldots, m\}$, i.e., the lattice of partitions of $\{1, \ldots, m\}$ ordered by refinement.

- Rank of $I \in \Pi_m$ equals

  $$m - \text{number of blocks of } I.$$
\( \Pi_4 : \)

\[
\begin{align*}
\{1, 2, 3, 4\} & \quad \text{rank 3} \\
\{1, 2, 3\}, \{4\} & \quad \text{rank 2} \\
\{1, 2\}, \{3, 4\} & \\
\{1, 2, 4\}, \{3\} & \\
\{1, 3\}, \{2, 4\} & \\
\{1, 3, 4\}, \{2\} & \\
\{1, 4\}, \{2, 3\} & \\
\{2, 3, 4\}, \{1\} & \quad \text{rank 1} \\
\{1, 2\}, \{3\}, \{4\} & \\
\{1, 3\}, \{2\}, \{4\} & \\
\{1, 4\}, \{2\}, \{3\} & \quad \text{rank 0} \\
\{2, 3\}, \{1\}, \{4\} & \\
\{2, 4\}, \{1\}, \{3\} & \\
\{3, 4\}, \{1\}, \{2\} & \\
\end{align*}
\]
Number of chambers of $\mathcal{A}_{m,n}$

- Let
  \[ \Pi^n_m : \text{the rank } n \text{ truncation of } \Pi_m. \]

- Then we have
  \[ \Pi^n_m = \{ \text{partitions of } \{1, 2, \ldots, m\} \text{ with at least } m - n \text{ blocks} \}. \]

- Note $\Pi^n_m = \Pi_m$ when $n \geq m - 1$. 
\[ \Pi_4^2 : \]

\begin{align*}
\{\{1, 2, 3\}, \{4\}\} & \quad \{\{1, 2\}, \{3, 4\}\} & \quad \{\{1, 2, 4\}, \{3\}\} & \quad \{\{1, 3\}, \{2, 4\}\} & \quad \{\{1, 3, 4\}, \{2\}\} & \quad \{\{1, 4\}, \{2, 3\}\} & \quad \{\{2, 3, 4\}, \{1\}\} \\
2 & \quad 1 & \quad 2 & \quad 1 & \quad 2 & \quad 1 & \quad 2 \\
\{\{1, 2\}, \{3, 4\}\} & \quad \{\{1, 3\}, \{2, 4\}\} & \quad \{\{1, 4\}, \{2, 3\}\} & \quad \{\{2, 4\}, \{1, 3\}\} & \quad \{\{3, 4\}, \{1, 2\}\} & \quad \{\{1\}, \{2\}\}, \{3\}, \{4\}\} & \quad \{\{1\}, \{2\}, \{3\}, \{4\}\} \\
-1 & \quad -1 & \quad -1 & \quad -1 & \quad -1 & \quad 1 & \quad 1 \\
\{\{1\}, \{2\}, \{3\}, \{4\}\} & \quad \{\{1\}, \{2\}\}, \{3\}, \{4\}\} & \quad \{\{1\}, \{2\}\}, \{3\}, \{4\}\} & \quad \{\{1\}, \{2\}\}, \{3\}, \{4\}\} & \quad \{\{1\}, \{2\}\}, \{3\}, \{4\}\} & \quad \{\{1\}, \{2\}\}, \{3\}, \{4\}\} & \quad \{\{1\}, \{2\}\}, \{3\}, \{4\}\} \\
1 & \quad 1 & \quad 1 & \quad 1 & \quad 1 & \quad 1 & \quad 1 \\
\text{rank 2} & \quad \text{rank 1} & \quad \text{rank 1} & \quad \text{rank 0} & \quad \text{rank 0} & \quad \text{rank 0} & \quad \text{rank 0}
\end{align*}
Number of chambers of \( \mathcal{A}_{m,n} \)

- **Proposition (Kamiya and Takemura (1997))**: 
  \[ L(\mathcal{A}_{m,n}) \cong \Pi_n^m. \]

- \( L \ni X \mapsto I_X \in \Pi_n^m. \)

- \( \Pi_n^m \ni I \mapsto X_I := \bigcap H_{ij} \in L, \) where the intersection is taken over all pairs \( i, j \) such that \( i \) and \( j \) are in the same block of \( I \).

- We have 
  \[ \text{rank of } I = m - \text{number of blocks of } I 
  = \text{codim}(X_I) = r(X_I). \]
When \( n \geq m - 1 \), we have \( \Pi_n^m = \Pi_m \) and thus
\[
\pi(A_{m,n}, t) = \sum_{I \in \Pi_m} \mu(X_I)(-t)^{r(X_I)},
\]
which is known to be
\[
(1 + t)(1 + 2t) \cdots (1 + (m - 1)t) = 1 + c_1 t + c_2 t + \cdots + c_{m-1} t^{m-1}.
\]
When $n < m - 1$, on the other hand, we have

$$
\pi(A_{m,n}, t) = \sum_{I \in \Pi^m_n} \mu(X_I)(-t)^{r(X_I)}
$$

$$
= \pi(A_{m,n}, t) \text{ for } n \geq m - 1
$$

with terms of degree $> n$ neglected

$$
= 1 + c_1 t + c_2 t + \cdots + c_n t^n.
$$
Number of chambers of $A_{m,n}$

Note that the Möbius function $\mu(X_I)$, $I \in \Pi^n_m$, takes the same values as those for the case $n \geq m - 1$, because $\mu$ is defined recursively from the bottom up.

$L(A_{4,2}) \cong \Pi_4^2$
Look at $m = 3$ in $\mathbb{R}^1$ again:

We embed it in $\mathbb{R}^2$. 
Perturbation viewpoint and inadmissible rankings

• Perturb the point 2.
Perturbation $n = 1, m = 4$

4 points on a line.

$\mathbb{R}^1$
Perturbation $n = 2, m = 4$

- 4 points on a plane.

The diagram shows four points on a plane, which are then mapped into a 3-dimensional space. The transformation from $\mathbb{R}^2$ to $\mathbb{R}^3$ is illustrated with arrows and geometric shapes, indicating the movement and arrangement of the points.
Perturbation viewpoint and inadmissible rankings

- 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 direction in \(\mathbb{R}^n\).

- 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.

- From the picture we also see that unfolding arrangement can be identified with an intersection of the braid arrangement with lower-dimensional affine spaces (in the sense of face poset).

- This corresponds to the truncation of partition lattice which is isomorphic to the intersection lattice of the braid arrangement.
However not all intersections of the braid arrangement with lower-dimensional affine spaces can be realized as an unfolding arrangement.
Ranking patterns (continued to Talk 2)
Define

\textbf{ranking pattern} := \{admissible rankings\}

for a given \(m\)-tuple \((x_1, \ldots, x_m)\).

A different \((x_1, \ldots, x_m)\),
\[\rightarrow\] a different ranking pattern.

\textbf{Problem 2}

How many ranking patterns are possible?
Difficult for general $n$.

Unidimensional case $n = 1$. (continued to Talk 2)

For this case we know that there are exactly three different patterns.

Here we take a look at $m = 4$ points in $\mathbb{R}^2$ ($n = 2$).
Three ranking patterns for $m = 4, n = 2$

- How do we know that there are three patterns?
- Look at projections of a regular tetrahedron to planes.
Three ranking patterns for $m = 4$, $n = 2$

- One vertex visible in front of three other vertices.

This is Pattern I.

- On the other hand, think of looking from the other side. Then one vertex is invisible behind three other vertices. This pattern is not realizable as an unfolding arrangement.
Three ranking patterns for $m = 4, n = 2$

- Two vertices visible in front of two other vertices.

Left: pattern II  

Right: pattern III

1 2 3 4

is degenerate, so perturb.

1 2 3 4: II

1 2 3: III
Three ranking patterns for $m = 4, n = 2$

- We cannot present a definite result at the moment, but we (Kamiya and I) think that the same reasoning can be applied to the case of $m = n + 2$. 
Concluding Remarks

- We have introduced unfolding model.
- We have formulated it as a problem in hyperplane arrangement.
- We found the number of admissible rankings.
- We have discussed ranking patterns for $m = 4$, $n = 2$. 
C. H. Coombs,
Psychological scaling without a unit of measurement, 

I. J. Good and T. N. Tideman,
Stirling numbers and a geometric structure from voting theory, 
*J. Combinatorial Theory (A)*, **23** (1977), 34–45.

H. Kamiya and A. Takemura,
On rankings generated by pairwise linear discriminant analysis of $m$ populations, 
H. Kamiya and A. Takemura,
Rankings generated by spherical discriminant analysis,

H. Kamiya and A. Takemura,
Characterization of rankings generated by linear discriminant analysis,