INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
RICE UNIVERSITY

Visual Estimation of Structure in Ranked Data

by

Keith A. Baggerly

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

David W. Scott, Chairman
Professor of Statistics

Dennis D. Cox
Professor of Statistics

Alejandro A. Schäffer
Assistant Professor of Computer Science

Georgia L. Thompson (Thesis Director)
Research Scientist
University of Texas at Dallas

Houston, Texas

May, 1995
Visual Estimation of Structure in Ranked Data

Keith A. Baggerly

Abstract

Ranked data arise when some group of judges is asked to rank order a set of $n$ items according to some preference function. A judge's ranking is denoted by a vector $x = (x_1, \ldots, x_n)$, where $x_i$ is the rank assigned to item $i$. If we treat these vectors as points in $\mathbb{R}^n$, we are led to consider the geometric structure encompassing the collection of all such vectors: the convex hull of the $n!$ points in $\mathbb{R}^n$ whose coordinates are permutations of the first $n$ integers. These structures are known as permutation polytopes.

The use of such structures for the analysis of ranked data was first proposed by Schulman [65]. Geometric constraints on the shapes of the permutation polytopes were later noted by McCullagh [56]. Thompson [77] advocated using the permutation polytopes as outlines for high-dimensional "histograms", and generalized the class of polytopes to deal with partial rankings (ties allowed).

Graphical representation of ranked data can be achieved by putting varying masses at the vertices of the generalized permutation polytopes. Each face of the permutation polytope has a specific interpretation; for example, item $i$ being ranked first. The estimation of structure in ranked data can thus be transformed into geometric (visual) problems, such as the location of faces with the highest concentrations of mass.

This thesis addresses various problems in the context of such a geometric framework: the automation of graphical displays of the permutation polytopes; illustration and estimation of parametric models; and smoothing methods using duality – where every face is replaced with a point. A new way of viewing the permutation polytopes as projections of high-dimensional hypercubes is also given. The hypercubes are built as cartesian products of the $\binom{n}{2}$ possible paired comparisons, and as such lead to methods for building rankings from collections of paired comparisons.
Acknowledgments

There are a great many people whose efforts and input have helped me with this task in one way or another. Primary among these are my advisors, Georgia Thompson and David Scott. Georgia’s willingness to help me with various intricacies of ranking structures, to argue with me and get me to refine my ideas, and to prod me along as necessary was invaluable. Dr. Scott’s interest in the visualization of high dimensional structure can be seen throughout. Further comments are due to the other members of my thesis committee, Dennis Cox and Alejandro Schäffer. My thanks to them.

In addition to the above, there are a great many others whose ideas, suggestions and/or support have helped me greatly in the past few years. These include Diane Brown, Meg Gelder Ehm, Kathy Ensor, Marek Kimmel, Monnie McGee, Mike Minnotte, Mike Pearlman, Steve Sain, Mike Swaim, James Thompson, Matt Wand, Joe Warren, and R. Allen White. Many thanks.

Finally, a note of thanks to my parents, Leo and Carole Baggerly. No more need be said.

This work has been generously supported by the NIH under Predoctoral Trainee Grant PHS GMO 8362-3, and in part by the Office of Naval Research under Contract N00014-90-J-1176, the National Science Foundation under grant DMS-9306658, and the National Security Agency under grant MOD 9086-93. Computer time and video assistance were generously provided by the Department of Geology at Southern Methodist University.
Contents

Abstract ii
Acknowledgments iii
List of Tables vi
List of Illustrations vii

1 Introduction 1

2 Previous Work: Ranked Data 7
   2.1 Metrics on Full Rankings .................................. 7
   2.2 Metrics on Partial Rankings ................................. 10
   2.3 Parametric Models ............................................. 16

3 Permutation Polytopes 19
   3.1 Basics .......................................................... 19
   3.2 Generalization to Partial Rankings ........................... 21
   3.3 A New Notation ............................................... 25
   3.4 Categorizing Faces ........................................... 28
   3.5 Marginal and Orthogonal Projections ....................... 31

4 Automating Permutation Polytopes 38
   4.1 Full Rankings .................................................. 38
   4.2 Mixed Rankings ................................................. 42
      4.2.1 Plotting .................................................... 46
      4.2.2 Fusing ..................................................... 47
   4.3 High-Dimensional Options ................................... 50
   4.4 Software Notes ............................................... 51
      4.4.1 four ......................................................... 51
      4.4.2 five ......................................................... 55
5 Duals
  5.1 Smoothing Ranked Data .................................. 58
  5.2 Duals and Multivariate Binary Data ...................... 64

6 Paired Comparisons
  6.1 Permutation Polytopes as Projections of Hypercubes .... 70
  6.2 Ties, Incomplete Pairings, and Partial Rankings .......... 80
  6.3 Linear Information and Choice Consistency ............... 82
  6.4 Implied Metrics on Mixed Rankings ....................... 86

7 Future Work
  7.1 Parallel Coordinates ..................................... 95
  7.2 Wavelets .................................................. 97
  7.3 Experimental Design ...................................... 99
  7.4 Spectral Decomposition via Group Theory ................. 99
  7.5 Spectral Decomposition via Hypercubes .................. 102

A Data Sets
  A.1 Rice Voting ............................................... 105
  A.2 German Political Preferences ............................. 107
  A.3 Attitudes Towards Science ............................... 108

Bibliography ................................................. 109
Tables

6.1 Vectors associated with inconsistent triples arising from paired comparisons of four items. 72
6.2 Vectors associated with a given item beating all others (being ranked first). 74

A.1 Rice Ballots; 300 Respondents 105
A.2 Frequencies of Preference Orderings for Four Political Goals; A = Maintaining order in the nation; B = More say in the government; C = Fighting rising prices; D = Protecting freedom of speech; 2262 Respondents. 107
A.3 Agree/Disagree Responses to Four Questions; A = The development of new ideas is the scientist’s greatest source of satisfaction; B = Scientists and engineers should be eliminated from the military draft; C = The scientist will make his maximum contribution to society when he has freedom to work on problems which interest him; D = The monetary compensation of a Nobel Prize winner in physics should be at least equal to that given popular entertainers; 2982 Respondents. 108
Illustrations

1.1 Orderings of items C,S,P ........................................ 3
1.2 Rankings of items C,S,P corresponding to the orderings in Figure 1.1. 4
1.3 Convex hull of the rankings of three items. ....................... 5

3.1 Truncated Octahedron: Orderings .............................. 22
3.2 Truncated Tetrahedron: Orderings ............................. 22
3.3 Cuboctahedron: Orderings ...................................... 30
3.4 Tetrahedron: Orderings .......................................... 30
3.5 Octahedron: Orderings .......................................... 31
3.6 Rotation of a Truncated Octahedron Suggesting a Marginal Projection. 33
3.7 Rotated Truncated Octahedron with some Overlapping Points Labelled. 34
3.8 Marginal Distribution of items A and B in rankings of 3, 4 and 5
   items (offset). .................................................. 35
3.9 Marginal Distribution of items A and B when first, second and third
   place are indicated from 11 items. ............................. 37

4.1 German Political Data: Orderings and Wireframe ............... 39
4.2 German Political Data: AC face highlighted, Wireframe hidden. . 39
4.3 Stereographic Projection of a Cube .............................. 41
4.4 Projected Polytope arising from the full rankings of five candidates.
   The highlighted hexagonal prism corresponds to candidates D and E
   being ranked in the last two positions. .......................... 43
4.5 Projected Polytope with wireframe hidden save for a highlighted face.
   The highlighted hexagonal prism corresponds to candidates D and E
   being ranked in the last two positions. .......................... 44
4.6 Projected Polytope corresponding to the full rankings of five candidates, with the projection point centered above the hexagonal prism corresponding to candidates A and B being ranked in the first two positions. The highlighted face corresponds to candidates D and E being in the last two positions. ................................................. 45

4.7 Rice Voting Data, full and partial rankings shown. Full rankings are red, first, second and third place specified are magenta, first and second place specified are aqua, and first place only is blue. .................. 48

4.8 Rice Voting Data, with positions 4 and 5 fused. The highlighted hexagon corresponds to candidates D and E being ranked in the last two positions. ......................................................... 49

5.1 Kernel Estimate of a continuous Density Function. Observed data values are indicated by vertical bars. Gaussian Kernels are indicated by dashed lines. The Density Estimate is shown as a solid curve. ... 60

5.2 Dual of a Truncated Octahedron. .................................................. 62
5.3 Rotation of the Cube to form the Dual ........................................... 63
5.4 Dual of the German Political Data .................................................. 65
5.5 Residuals after fitting Mallows' $\phi$ Model .................................. 65
5.6 Dual of Low IQ Responses. .............................................................. 67
5.7 Dual of High IQ Responses. ............................................................ 67

6.1 Cube of Paired Comparisons of Three Items, and the Associated Projection Onto the Consistent Subspace. Inconsistent Triples are Indicated by Circles. ......................................................... 71

6.2 Paired Comparison Hypercube with Inconsistent Triples Removed. ... 77

6.3 Consistent subspace projection of paired comparisons of three items, with ties allowed. The first coordinate represents AB, the second AC, and the third BC. ....................................................... 81

6.4 Two Induced Metrics on the Rankings. (a) $\tau_e$, the minimal edge path across the hypercube, and (b) $\nu$, the minimal edge path across the grid. 89

6.5 Graph Representation of the Item Ordering ($A, B, C$). ..................... 90

6.6 Graphs of Paired Comparison Triples ($AB, AC, BC$). Left, $(-1, 1, -1)$, Center, $(0, 0, 0)$, and Right, $(1, -1, 1)$. ......................... 90
6.7 Values of \((a_A, a_B, a_C)\) at corresponding grid points. Corresponding full Orderings are shown on the outer hexagon. 91

6.8 Cayley's distance. The solid dot indicates the original ranking, the dashed lines indicate axes of rotation, and the hollow dots indicate those rankings one step from the original. 93

7.1 Parallel Coordinate Representation of the Ranking \((1, 3, 4, 2)\). 96
7.2 Parallel Coordinate Representation of the German Political Data. 96
7.3 Parallel Marginal Projections. 98
7.4 Parallel Flowing from \((1, 3, 4, 2)\) to \((4, 1, 2, 3)\). 98
7.5 Decomposition of a Cube into the (a) grand mean, (b) AB main effect, (c) AC main effect, (d) BC main effect, (e) AB/AC interaction, (f) AB/BC interaction, (g) AC/BC interaction, and (h) AB/AC/BC interaction. Dark dots are positive, light dots are negative. Inconsistent points are circled in (a). 103

7.6 Induced Decomposition of Rankings into (a) grand mean, (b) AB main effect, (c) AC main effect, (d) BC main effect, (e) AB/AC interaction, (f) AB/BC interaction, (g) AC/BC interaction, and (h) AB/AC/BC interaction. Dark dots are positive, light dots are negative. Dashed lines have been added to emphasize structure. Orderings are shown in (a). 104
Chapter 1

Introduction

Ranked data arise when a group of judges is asked to rank order a set of \( n \) items according to some preference function. Examples of such functions include evaluations of a political candidate's suitability for office, the perceived order of appearance of various symptoms in the progression of a disease, or the relative importance of several factors when choosing a place to live. For example, assume a judge is given three cookies, Chocolate Chip (\( C \)), Sugar (\( S \)), and Peanut Butter (\( P \)). The ordering \( (C, S, P) \), indicating that \( C \) is most preferred, \( S \) next, and \( P \) least, would be an element of ranked data. If the same question were asked of a large group of judges, the collection of responses would comprise a set of ranked data.

Given a set of ranked data, the question arises as to whether one can infer something about the underlying "preference structure" of the population from which the sample was drawn. To do this, a more explicit definition of structure is needed. Structure in ranked data consists of some group of items being preferred to some other group of items in a predictable way which can not be reasonably ascribed to pure chance. For example, in the cookie example mentioned above, if the responses were predominantly \( (C, S, P) \) or \( (C, P, S) \), then the preference structure of the population with respect to this group of items is that \( C \) is preferred to \( P \) and \( S \). To fully define the structure, we would then need to address the issue of how often responses not exhibiting the characteristic defining the structure should occur in the population.

It can be argued that the lack of any such structure is again structure of a sort, indicating the relative indifference of the population with respect to the items being ranked. This was acknowledged by Kendall [49], who termed the uniform distribution on rankings the "null" ranking model, and presented an open invitation for others to provide "non-null" ranking models.

Ideally, defining the structure in a set of data would consist of identifying the defining characteristics and specifying a model which accurately describes the relative frequencies with which each type of response occurs in the population. These two tasks, identification of defining characteristics and specification of a model, are
in practice quite distinct. In theory, the defining characteristics should be implicit in the model, but in the absence of information of some sort (or an assumption) indicating that a given model is appropriate, the best approach is to seek the defining characteristics and attempt to find a model that fits. This thesis will primarily be concerned with the identification of the defining characteristics, but various models will also be addressed.

The ability to identify such characteristics and to recognize appropriate models is often greatly enhanced if the examiner is able to form a picture of the data. If one can construct a visual framework in which the data can be expressed in such a way as to correctly emphasize the nature and location of structure, detection of such features can be enhanced. It is thus a natural question to ask how ranked data can be graphically examined in an effective manner.

The naive method would simply consist of listing all of the possible orderings in a row and plotting bars above them, the bars being proportional in height to the number of judges picking that particular ordering. This has the familiar feel of the histogram, and indeed it allows for the discernment of certain salient features, such as which particular ordering was chosen most often. Nonetheless, the naive method poorly represents many structures and can even be misleading.

The naive method is misleading in that there is no natural linear (one-dimensional) arrangement of the orderings. There is no a priori reason why the orderings for the cookie example should be listed in the order \((C, S, P)\), \((C, P, S)\), \((S, P, C)\), \((S, C, P)\), \((P, C, S)\), \((P, S, C)\) as opposed to \((S, P, C)\), \((C, S, P)\), \((P, C, S)\), \((S, C, P)\), \((C, P, S)\), \((P, S, C)\). Indeed, if the data is plotted in this way, the appearance of a curve or shape suggesting the presence of a defining characteristic is likely an artifact due to the method of presentation (the particular ordering on the line).

A natural (albeit nonlinear) arrangement of the orderings which is properly suggestive of structure does exist. However, to find it, the issue of when one ordering can be said to be \"next to\" another must be addressed. We shall say that an ordering is next to another if one can be obtained from the other via a single transposition of adjacent elements. Thus, \((C, S, P)\) is next to \((C, P, S)\) (interchanging \(S\) and \(P\)) and to \((S, C, P)\) (interchanging \(C\) and \(S\)). Interchanging nonadjacent elements (\(C\) and \(P\)) is not allowed. If the orderings are then arranged so that each ordering is physically next to its adjacent neighbors as defined above, a hexagonal array results (see Figure 1.1; note that this is only valid for the orderings of three items). Note that this arrangement is conducive to the perception of defining characteristics; the upper
right edge corresponds to the group of people who prefer $C$ most, whereas the edge diametrically opposed corresponds to the group of people who prefer $C$ least. Each edge has an interpretation; three edges of the hexagon correspond to items being ranked first and three edges correspond to items being ranked last.

Up to this point, the data have been spoken of in terms of orderings of the item labels, but in order to generalize things more easily a formulation in terms of numbers must be introduced. To accomplish this, specify some arbitrary arrangement of the items being considered ($C, S, P$, for example). The first item in this arrangement will be said to be item 1, the next item 2, and so forth. Each ordering of the item labels can now be written as a vector of numbers, by letting the first coordinate of the vector be the rank of item 1, the second coordinate the rank of item 2, etc.

**Definition 1.1** A ranking $\pi$ is a vector of numbers,

$$\pi = (\pi(1), \pi(2), \ldots, \pi(n)),$$

such that $\pi(i)$ is the rank of item $i$.

**Definition 1.2** An ordering $\pi^{-1}$ is a vector of item labels

$$\pi^{-1} = (\pi^{-1}(1), \pi^{-1}(2), \ldots, \pi^{-1}(n))$$

such that $\pi^{-1}(i)$ is the item assigned rank $i$. 
Note the different brackets used for rankings and orderings ("(" and ")", respectively). This serves to emphasize the fact that the elements of the two vectors are of different types, in that one contains numbers and the other contains labels. The two representations contain equivalent information. Nonetheless, it seems to be easier for people to think in terms of orderings, even though the analysis is easier using rankings. The rankings corresponding to the cookie orderings (using the arbitrary arrangement \( C, S, P \)) are shown in Figure 1.2.

The ranking \((1, 2, 3)\) corresponds to \(C\) being ranked first, \(S\) second and \(P\) third, and the ranking \((2, 3, 1)\) corresponds to \(C\) being ranked second, \(S\) third, and \(P\) first. In terms of the orderings, two points are next to each other if they differ by a single transposition of adjacent elements; in terms of the rankings, two rankings are next to each other if they differ by a single transposition of consecutive integers. Thus, \((2, 3, 1)\) is next to \((1, 3, 2)\) (interchanging 1 and 2) and to \((3, 2, 1)\) (interchanging 2 and 3).

It should be noted that if the rankings are treated as points in \(\mathbb{R}^3\), then the hexagon is the convex hull of these points. This is illustrated in Figure 1.3. This figure arises as the intersection of the 6 half-spaces defined by the linear inequalities.
1 ≤ X, Y, Z and X, Y, Z ≤ 3 and the two-dimensional plane defined by the linear equality X + Y + Z = 6. The edges of the hexagon are defined by letting one of the inequalities be an equality - the edges defined by letting X = 1, Z = 3 and Y = 1 are indicated.

The geometrical frameworks arising as the convex hulls of the n! points in \( \mathbb{R}^n \) whose coordinates are permutations of n distinct numbers are known as permutation polytopes. Permutation polytopes are discussed in Yemelichev, Kovalev and Kravtsov [90]. The use of permutation polytopes in the context of analyzing ranked data (restricting the n distinct numbers to be the first n integers) was first proposed by Schulman [65].

Occasionally, a complete ordering of the items will not be given. For example, the judges in the cookie example above may have been asked to give only their first choice. Such data is said to be partially (as opposed to fully) ranked. Thompson [77] introduced a class of generalized permutation polytopes, defined as the convex hull of all points on \( \mathbb{R}^n \) whose coordinates are permutations of n not necessarily distinct
numbers, for the graphical analysis of partially ranked data. Generalized permutation polytopes will be covered in depth in Chapter 3.

The goal of this thesis is to visually assess structure in ranked data, using the geometric framework supplied by the generalized permutation polytopes. Chapter 2 provides an overview of the literature on the analysis of ranked data. Chapter 3 examines the generalized permutation polytopes and their properties, including useful projections thereof. Chapter 4 describes the development of software to automate the use of generalized permutation polytopes in analyzing ranked data. Chapter 5 describes a new way of finding structure using the duals of the generalized permutation polytopes. Connections between the duals and multivariate binary data are explored. Chapter 6 introduces a new way of viewing the permutation polytopes as projections of high-dimensional hypercubes. The hypercubes are built as cartesian products of the \( \binom{n}{2} \) paired comparisons possible among \( n \) items, and as such lead to methods for building rankings from collections of paired comparisons. The hypercube formulation induces two discrete metrics on rankings, whose properties are explored. In all chapters, attention is paid to dealing with the problems posed by the analysis of full rankings, partial rankings, and mixtures of full and partial rankings. Finally, Chapter 7 addresses some open problems and directions for future research.
Chapter 2

Previous Work: Ranked Data

Ranked data has been around for a long time, but the study of ways to analyze such data is only about a century old. Among the first papers addressing the issue of determining the degree of correlation between two rankings are two by Spearman in 1904 and 1906 ([71], [72]). One of the first probability models on rankings was set forth by Thurstone in 1927 ([80]). Both Spearman [71] and Thurstone [80] were motivated by the need to analyze psychological data. Kendall introduced another measure of correlation in 1938 [42], and various other attacks on problems connected with ranked data began to appear in the journal *Biometrika* over the next few years. The analysis of ranked data per se, however, assumed a coherent form with the first edition of Kendall’s *Rank Correlation Methods* in 1948 ([49]). In that treatise, Kendall [49] compiled and examined several extant methods for assessing degrees of correlation between ranks and specified the need for adequate models. It is not clear whether Kendall [49] was aware of Thurstone’s [80] work at the time; it is not cited.

2.1 Metrics on Full Rankings

As both Spearman and Kendall realized, the key to assessing the degree to which two rankings are in agreement lies in the definition of a measure of discrepancy. The three measures of distance between two rankings of $n$ items, $\vec{\pi}_0$ and $\vec{\pi}_1$, introduced by this stage are, in chronological order,

- Spearman’s $\rho$ (1904):

  $$\rho(\vec{\pi}_0, \vec{\pi}_1) = \left( \sum_{i=1}^{n} (\vec{\pi}_0(i) - \vec{\pi}_1(i))^2 \right)^{1/2},$$

  which can be viewed as Euclidean distance if $\vec{\pi}_0$ and $\vec{\pi}_1$ are treated as points in $\mathbb{R}^n$. 

• Spearman's Footrule (1906):

\[ F(\tilde{\pi}_0, \tilde{\pi}_1) = \sum_{i=1}^{n} |\tilde{\pi}_0(i) - \tilde{\pi}_1(i)|, \]

the sum of the absolute differences between the rankings, and

• Kendall's \(\tau\) (1938) [42]:

\[ \tau(\tilde{\pi}_0, \tilde{\pi}_1) = \sum_{i \neq j} I\{(\tilde{\pi}_0(i) - \tilde{\pi}_0(j)) \cdot (\tilde{\pi}_1(i) - \tilde{\pi}_1(j)) < 0\}. \]

Here, \(I\{S\}\) is the indicator function of the condition \(S\); \(I\{S\} = 1\) when \(S\) holds, and 0 otherwise. This corresponds to the number of pairs of items \((i, j)\) ranked in different relative orders by the two rankings. This can also be shown to be the minimal number of transpositions of adjacent elements required to get one ordering from the other (a proof of this is given in Kendall [49]).

Strictly speaking, Kendall's \(\tau\) was not discovered by Kendall; the measure in question was proposed by Fechner [27] as far back as 1897. For the involved history of various ranking measures, the interested reader is referred to Kruskal [51]. Also strictly speaking, Kendall and Spearman were both examining measures of correlation and so had normalized the measures given above to lie in the range from \(-1\) to 1. We shall only be concerned with the measures given above, not with their correlation forms.

It is worth noting that the connection structure for the permutation polytopes given in Chapter 1 (the collection of edges defining the convex hull of the \(n!\) permutations of the first \(n\) integers in \(\mathbb{R}^n\)) corresponds to connecting those vertices which are closest to one another under any one of the three above measures.

All of the measures of distance given above are actually metrics on rankings, and satisfy the standard properties of positivity, symmetry in arguments, and the triangle inequality. Furthermore, all of the above metrics are right-invariant, meaning that

\[ d(\tilde{\pi}_0, \tilde{\pi}_1) = d(\tilde{\pi}_0 S, \tilde{\pi}_1 S) \]

when \(S\) is a permutation matrix. This is simply equivalent to saying that if the elements were to be relabeled the distance between rankings of those elements would not change. For example, given two rankings of items \(A, B, C\) and \(D\), if the items were then relabeled \(B, A, C\) and \(D\), the two rankings should change to reflect the new
positions of items $A$ and $B$, but the distance from one ranking to the other should not change. Proofs of these properties may be found in Critchlow [17].

As Critchlow [17] observes, other right-invariant metrics on rankings include

- Hamming's distance:

$$ H(\bar{\pi}_0, \bar{\pi}_1) = \sum_{i=1}^{n} I\{\bar{\pi}_0(i) \neq \bar{\pi}_1(i)\}, $$

The number of items whose rankings differ from one ranking to the other,

- Ulam's distance: $U(\bar{\pi}_0, \bar{\pi}_1) = n$ minus the maximal number of items ranked in the same order by $\bar{\pi}_0$ and $\bar{\pi}_1$, and

- Cayley's distance: the minimal number of item transpositions (not necessarily of adjacent elements) required to transform $\bar{\pi}_0$ into $\bar{\pi}_1$.

Most attention in what follows will be focused on Kendall's $\tau$, though Spearman's Footrule, Spearman's $\rho$, and Cayley's distance will be examined briefly. For further details on the various metrics the reader is referred to the monographs of Critchlow [17] and Diaconis [22].

Asymptotic properties of the various metrics under the null distribution have been examined by Kendall [42], Alvo, Cabiliio and Feigin [2], Diaconis and Graham [24], Feigin and Cohen [29], and Feigin and Alvo [28].

Diaconis [22] notes that the full rankings of $n$ items can be viewed as elements of permutation group $S_n$: the symmetric group on $n$ items. The symmetric group, in turn, can be represented as a collection of invertible matrices "in such a way that the matrix assigned to the product of two elements is the product of the matrices assigned to each element." While there are many such representations, attention here shall be focused on the one-to-one permutation representation (note that a representation can be many to one; mapping every element of the group to the constant 1 trivially preserves the multiplication structure specified above). In the permutation representation, every ranking $\bar{\pi}$ defines an $n \times n$ matrix $\pi$ of ones and zeros, the $i^{th}$ column of which is zero save for the $\bar{\pi}(i)^{th}$ entry, which is one. Diaconis then brings methods designed for the analysis of groups to bear on the analysis of ranked data. Of particular interest is his suggestion of an analog to spectral decomposition on rankings based on the Fourier decomposition of finite groups. Diaconis later expanded on this theme in his 1987 Wald lecture [23]. This aspect is further addressed in Chapter 7.
2.2 Metrics on Partial Rankings

The methods described above were all initially developed to deal with full rankings; where a clear order can be established amongst any subset of the items being ranked. However, in many cases the ranking information available is not of this form because two or more of the items being ranked are "tied" according to the judge's criterion function. In a preferential election amongst four candidates, \( A, B, C, D \), for example, a voter may return an incomplete ballot specifying \( A \) first, \( B \) second, and expressing no preference between \( C \) and \( D \). This partial ordering can be written \( (A, B, (C, D)) \), where the parentheses within the brackets serve to group items that are tied.

Several questions arise when partially ranked data are considered. First, are the data available all of the same partial ranking type (first and second specified, third and fourth tied, for example), or are mixtures of types present? Second, can the various metrics on full rankings be extended in some sense to yield measures of discrepancy for partial rankings? Third, what "rank value" should be assigned to tied items? In the example given above, two contenders are \( (1, 2, 3, 3) \), corresponding to an increment step for every group, and \( (1, 2, 3, 5, 3.5) \), the "midrank", the average of the possibly corresponding full rankings (in this case \( (1, 2, 3, 4) \) and \( (1, 2, 4, 3) \)).

The first question must be answered on a case by case basis. Addressing the third question, the midrank values are those most commonly assigned to tied ranks. Using the midrank values, Kendall \[44],[45],[49] suggests two extensions of Kendall's \( \tau \) and one extension of Spearman's \( \rho \) to mixtures of full and partial rankings. Unfortunately, the two extensions of Kendall's \( \tau \) do not agree in general, and they lack the ease of interpretation (a minimal preference shift) associated with \( \tau \) for full rankings; these will be addressed in more detail in Chapter 6. Using the midrank values, the extensions of both Spearman's \( \rho \) and Spearman's Footrule to mixtures of full and partial rankings are fairly trivial, as they correspond to the \( L_2 \) and \( L_1 \) distances between the rankings, respectively.

A major advance in treating the cases of partial rankings and mixtures of full and partial rankings was provided by Critchlow \[17\], who extended the group-theoretic approach of Diaconis \[22\] to the treatment of partially ranked data, using coset spaces of the permutation group. It should be noted that Critchlow's presentation does not make use of the matrix representation of the elements of \( S_n \). The following quote from Critchlow (\[17\], p.12-14) illustrates the concept:
The case where each judge lists his $k$ favorite items out of $n$: Let $S_n$ be the permutation group on $n$ items, and let $S_{n-k}$ be the subgroup of $S_n$ consisting of all permutations which leave the first $k$ integers fixed:

$$S_{n-k} = \{ \pi \in S_n : \pi_{i,i} = 1 \forall i = 1, \ldots, k \}.$$ 

Define an equivalence relation on $S_n$ as follows: For any two permutations on the set of $n$ items (which correspond to full rankings), $\pi, \sigma \in S_n$,

$$\pi \sim \sigma \iff \pi \sigma^{-1} \in S_{n-k} \iff \exists \tau \in S_{n-k} \text{ such that } \pi = \tau \sigma.$$ 

Note that in the above, $\sigma^{-1}$ refers to the inverse of the set of item transpositions associated with the ranking $\vec{\sigma}$ in $S_n$, not to the ordering $\vec{\sigma}^{-1}$. Thus, if $\sigma \in S_n$, $\sigma^{-1} \in S_n$.

The preceding equivalence relation partitions $S_n$ into so-called “equivalence classes”. For any $\pi \in S_n$, the equivalence class containing $\pi$ consists of all permutations equivalent to it, and is thus $\{ \tau \pi : \tau \in S_{n-k} \}$. This equivalence class is a subset of the permutation group, which we denote by $S_{n-k}\pi$, and call a right coset of $S_{n-k}$.

The importance of this lies in the observation that two full permutations $\pi, \sigma \in S_n$ belong to the same right coset of $S_{n-k}$ if and only if $\pi$ and $\sigma$ induce the same partial ranking of the first $k$ out of $n$ items.

For example, if judges were asked to indicate their first two choices out of a group of four items, then the two full rankings $(1,2,3,4)$ and $(1,2,4,3)$, corresponding to the full orderings $(A,B,C,D)$ and $(A,B,D,C)$, would both result in item $A$ being ranked first, and item $B$ second, with the difference between $C$ and $D$ going unnoticed because discrimination between third and fourth place was not requested. Thus, the two full rankings are said to be equivalent in the sense that they both induce the same partial ordering of the items, $(A,B,(C,D))$. Similarly, the full rankings $(3,2,1,4)$ and $(4,2,1,3)$ are equivalent in the sense that they both induce the partial ranking $(C,B,(A,D))$.

Consider the two rankings $(1,2,3,4)$ and $(1,2,4,3)$ in the matrix notation:

$$\pi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \sigma = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$
These are the only elements of this particular right coset of $S_4$, in that they are the only two $4 \times 4$ permutation matrices such that $\pi(1,1) = \pi(2,2) = 1$ (corresponding to item $A$ being ranked first and item $B$ being ranked second). To see that these matrices fulfill Critchlow's conditions for equivalence, note that $\pi, \sigma \in S_{4-2}$, $\sigma^{-1} = \sigma$, so $\pi \sigma^{-1} = \pi \sigma = \sigma \in S_{4-2}$, and letting $\tau = \sigma$, $\pi = \tau \sigma$. Partial rankings can thus be tied to cosets of the permutation group.

This method also allows for the measurement of distance between rankings of mixed types via the induced Hausdorff distance; this is Critchlow's "Main Theorem." Given two sets of elements, $G$ and $H$, and a metric $d(g,h)$ defined on the elements of the sets, the induced Hausdorff distance between the sets $G$ and $H$ is given by

$$D(G,H) = \max \{ \max_{g \in G} \min_{h \in H} d(g,h), \max_{h \in H} \min_{g \in G} d(g,h) \}.$$ 

By letting the sets $G$ and $H$ be distinct cosets of the permutation group, any metric on the full rankings can be extended to an induced Hausdorff metric on any type of ranking. For example, say that one judge gave the full ordering $\langle A, B, C, D \rangle$ and another gave the partial ordering $\langle B, C, (A, D) \rangle$. It is then desirable to be able to extend some metric on the full rankings (for example, Kendall's $\tau$) in such a way as to give a measure of the distance between these two orderings. The coset $C_{O_1}$ of $S_4$ corresponding to the full ordering $\langle A, B, C, D \rangle$ consists of the full ranking $a = (1,2,3,4)$, and the coset $C_{O_2}$ of $S_4$ corresponding to the partial ordering $\langle B, C, (A, D) \rangle$ consists of the full rankings $b = (3,1,2,4)$ and $c = (4,1,2,3)$. Using the standard definition of Kendall's $\tau$ for full rankings, $\tau(a,b) = 2$ and $\tau(a,c) = 3$. The Hausdorff extension of Kendall's $\tau$ (denoted $\tau^*$ by Critchlow) is thus

$$\tau^*(C_{O_1}, C_{O_2}) = \max \{ \max_{g \in C_{O_1}} \min_{h \in C_{O_2}} \tau(g,h), \min_{h \in C_{O_2}} \min_{g \in C_{O_1}} \tau(g,h) \}.$$ 

Noting that $g = a$ and that $h$ is either $b$ or $c$, the above becomes

$$\max \{ \max(\min(\tau(a,b), \tau(a,c))), \max(\min(\tau(b,a)), \min(\tau(c,a))) \} = \max \{ \max(\min(2,3), \min(\min(2,3))) \} = \max \{2,3\} = 3.$$ 

Extensions of other metrics (Spearman’s $\rho$, Spearman’s Footrule, etc) are defined and computed similarly.

Critchlow [17] also develops another type of metric extension to partial rankings; the fixed vector metric, attributed to Andrew Rukhin [64]. Unlike the Hausdorff
extension, the fixed vector method does make explicit use of matrix representations of group elements. The following proposition (Critchlow [17], p.27) illustrates this method.

**Proposition:** Let $G$ be any group, $K$ any subgroup of $G$, and let $\| \cdot \|$ denote a norm on $\mathbb{R}^n$. Let $\rho$ be a group representation of $G$; that is, $\rho$ assigns to each $\pi \in G$ a non-singular linear transformation

$$\rho(\pi) : \mathbb{R}^n \mapsto \mathbb{R}^n,$$

in such a way that

$$\rho(\pi \sigma) = \rho(\pi) \rho(\sigma) \quad \forall \pi, \sigma \in G.$$

Suppose further that $\rho$ is "norm-preserving" in the sense that

$$\| \rho(\pi)^T \vec{v} \| = \| \vec{v} \| \quad \forall \vec{v} \in \mathbb{R}^n \quad \forall \pi \in G.$$

Finally, suppose $\vec{v} \in \mathbb{R}^n$ is a vector fixed by $K$ under $\rho^T$, but not fixed by any larger subgroup. That is, suppose

$$\rho(\pi)^T \vec{v} = \vec{v} \quad \forall \pi \in K$$

and

$$\rho(\pi)^T \vec{v} \neq \vec{v} \quad \forall \pi \in GK.$$

If the preceding conditions are satisfied, then

$$d_{f_v}(K\pi, K\sigma) = \| (\rho(\pi) - \rho(\sigma))^T \vec{v} \|$$

is a right-invariant metric on the coset space $G/K$.

The most straightforward example involves letting $G = S_n$ and taking the representation $\rho$ to be the collection of $n \times n$ permutation matrices. As these matrices simply permute the order of the components of any element of $\mathbb{R}^n$, they are norm-preserving whenever the norm in use is order-indifferent. Now consider the partial ordering $\langle A, (B, C), D \rangle$. The two elements of the corresponding coset are $\langle A, B, C, D \rangle$ and $\langle A, C, B, D \rangle$, having representations

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$
respectively. For an element of $\mathbb{R}^4$ to be left unchanged by the application of one of these matrices, its middle two components must be identical, as the two are permuted, and for said element to be changed by any other permutation matrix, the first and fourth components must be distinct both from each other and from the central two. The key here is to think of the fixed vector as the vector of ranks being assigned to the items, noting that indistinguishable items are assigned the same rank. Thus, $(1,2,2,3)$ is a fixed vector, as is $(1,2.5,2.5,4)$. Of the two, the latter vector is more interesting in that it illustrates the "tied ranks" approach to partial rankings mentioned above. The new thing here is that the method suggests the use of somewhat arbitrary "pseudoranks": $(a_1,a_2,a_2,a_3)$.

To measure the fixed vector distance from $\langle A, (B, C), D \rangle$ to $\langle B, (A, D), C \rangle$, for example, assume $\bar{v}^T = (1,2.5,2.5,4)$ as the fixed vector. The two elements of the coset corresponding to $\langle B, (A, D), C \rangle$ are $\langle B, A, D, C \rangle$ and $\langle B, D, A, C \rangle$.

$$\rho(BADC)^T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \rho(BDAC)^T = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

so $\rho(BADC)^T \bar{v} = \rho(BDAC)^T \bar{v} = (2.5,1,4,2.5)^T$. If the vector norm is taken to be the $L_1$ norm, then

$$d_{fu}((A, (B, C), D), (B, (A, D), C)) = |1-2.5| + |2.5-1| + |2.5-4| + |4-2.5| = 10,$$

the Spearman’s Footrule distance between the two partial rankings if midranks are used. Similarly, if the vector norm is taken to be the $L_2$ norm, the fixed vector distance reduces to the Spearman’s $\rho$ distance between the two partial rankings.

Critchlow devotes a subsection of his monograph [17] to the comparison of the “common sense” tied ranks approach with the induced Hausdorff and fixed vector metrics. In it, he notes certain similarities; specifically, for distances between two rankings of the same type,

- the tied ranks Hamming distance is the same as the Hausdorff extension,
- the tied ranks Cayley distance is the same as the Hausdorff extension,
- the tied ranks Spearman’s Footrule is the same as the fixed vector extension, and
• the tied ranks Spearman’s $\rho$ is the same as the fixed vector extension.

He then turns to noting the limitations of the tied ranks approach, namely that they may not always be well-defined (Ulam’s metric), or that they may give either nonsensical results or may fail to exist for mixtures of different types of partial rankings. As an example of a nonsensical result, he notes that if two partial rankings of eight items are given, in the first of which the first three and the last five are tied, and in the second of which the first five and the last three are tied, then the tied ranks Hamming distance between the two is always 8, regardless of which items are placed in which group in either ranking. It is not clear that this is a nonsensical result, especially as the Hamming distance was initially designed as a measure of discrepancy between two binary sequences, so that entries either agreed exactly or failed to agree.

That the tied rank distance may fail to exist is indisputable, as is illustrated by the case of Cayley’s distance, the minimal number of item (or rank) transpositions required to take one ranking into the other. If the two rankings are $(1,2,3)$ and $(1,2.5,2.5)$, no number of transpositions will suffice. It is noted that this same problem holds for Kendall’s $\tau$ defined as the minimum number of pairwise adjacent transpositions. However, by changing the definitions of Cayley’s distance of $\tau$ slightly, this problem can be fixed. (this will be addressed further in Chapter 6).

Critchlow claims that none of these problems (being ill-defined, yielding nonsensical results, or failing to exist) arise with the Hausdorff metrics. He is largely correct. The Hausdorff metrics always exist and are always well-defined, and they can be the best measures available. Further, when they are restricted to full rankings, they reduce to the original metrics, which is decidedly a desirable property. However, the nonsensical results argument can be applied to the Hausdorff extensions as well. For example, the Hausdorff distance from a full ordering to a partial ordering must always be same as the distance from the full ordering to another full ordering. Thus, using the Hausdorff extension of Kendall’s $\tau$, the distance from $(A,B,C,D)$ to $(A,B,\overline{C},D)$ is the same as the distance from $(A,B,C,D)$ to $(A,B,D,C)$, even though the “intuitive” interpretation would suggest that an inability to discriminate between two alternatives would be less discrepant than active opposition.

Critchlow acknowledges that the tied ranks versions of Spearman’s Footrule and Spearman’s $\rho$ are “reasonable metrics”, but notes that they can be arrived at via the fixed vector approach. However, the fixed vector approach suffers from a flaw which the Hausdorff extension does not share: distances between rankings of different types
are not defined. Thus, the tied ranks approach is more general when used with either of Spearman's metrics.

In short, the argument over what type of metric extension is to be used should not rest on the basis of whether the extension works for all metrics. The occasional sacrifice of ease of interpretation when the metric is extended may be too high a price to pay. This will be discussed again briefly in Chapters 3 and 6.

2.3 Parametric Models

Non-null probability models on rankings emerged near the turn of the century and continue to appear. A good overview of many of the models currently in use is found in Critchlow, Fligner and Verducci [18]. The use of these models has thus far largely been restricted to the case of full rankings.

Perhaps the first probability models on rankings were those proposed by Thurstone [80]. Consider a random vector

\[ X = (X_1, X_2, \ldots, X_n), \]

where the \( X_i \) are random variables with known distributions. A ranking is determined by observing the relative values in a single realization of the random vector. The most commonly observed variant is the Thurstone-Mosteller model in which case the \( X_i \) are independent normal variates with equal variances and different means (Mosteller [60] provided least squares estimates of the parameters).

Luce [52] suggests a model of the following form: let \( p_i \) be positive weights adding to one (indicating, in some sense, the popularity of item \( i \)). Let \( P(\pi(1) = i) = p_i \), and proceed according to the sequential conditional process \( P(\pi(2) = j | \pi(1) = i) = \frac{p_j}{1-p_i} \), so that

\[ P(\pi) = \prod_{i=1}^{n} \frac{p_{\pi(i)}}{\sum_{j=1}^{n} p_{\pi(j)}}. \]

The Luce model can be shown to arise as a special case of the Thurstone model where the distributions of the \( X_i \) are Gumbel.

Babington Smith [5] proposed a quite general model based on the \( n(n-1)/2 \) pairwise comparison probabilities (defining \( p_{ij} \) to be the probability that item \( i \) is preferred to item \( j \) in a given comparison). Unfortunately, the number of parameters involved is quite large.

Mallows [53] begins with the paired comparison model, but makes several simplifications. First, following a suggestion of Bradley and Terry [12], he sets the pairwise
preference probability to
\[ p_{ij} = \frac{p_i}{p_i + p_j}. \]

This reduces the number of parameters to \( n - 1 \). Second, he assumes the existence of a modal ranking \( \bar{\pi}_0 \), and argues from symmetry that rankings which are equidistant from the mode in terms of Kendall's \( \tau \) and Spearman's \( \rho \) should have equal probabilities. This leads to a two parameter model,
\[ P(\bar{\pi}_1) = C(\theta, \phi) \theta^{2\tau(\bar{\pi}_1, \bar{\pi}_0)} \phi^{\tau(\bar{\pi}_1, \bar{\pi}_0)}, \]
where \( C(\theta, \phi) \) is a normalization constant, and \( \theta \) and \( \phi \) are dispersion parameters to be estimated. Two other models are generated by letting \( \theta \) and \( \phi \) be 1, respectively. In essence, these models comprise forms of exponential decay away from the modal ranking.

Diaconis [22] generalizes the \( \phi \) decay model by replacing Kendall's \( \tau \) in the exponent with other metrics on rankings. Arbitrary combinations of multimetric decays are implicit.

McCullagh [56], noting that the rankings can be viewed geometrically as points on a hypersphere, derives the rough equivalence of the Mallows-Bradley-Terry models on rankings and the Von Mises-Fisher models on spheres.

Fligner and Verducci [34] introduce a general class of multistage models, involving a decomposition of the ranking process into a series of independent stages. In what they suggestively term the forward model, a first-place item is chosen at the first stage according to some probability distribution. In the independent second stage, a second-place item is chosen from those remaining according to some different probability distribution, and so on. Backwards models, in which the last-place item is chosen at the first stage, the next-to-last-place item at the second stage and so on, or combinations of forward and backward models, are not necessarily equivalent to the forward models; these are not fully explored. By assuming a modal ranking and using a Mallows \( \phi \) model at each stage (with different dispersion parameters) they arrive at what they term the \( \phi \)-component model. This reduces to the standard Mallows \( \phi \) model if all of the dispersion parameters are the same.

More recently, Marden [54] has suggested a fundamentally different method of analyzing rankings. Marden proposes a method of orthogonal contrasts, whereby the relative rankings of one group of items are compared with that of the remaining group. These groups are then broken down in what amounts to an iterative procedure.
This has interesting implications for the location of regions of relative indifference (suggesting the imposition of a partial ranking structure). These are used to examine data roughly corresponding to a Mallows $\phi$ model in Chung and Marden [14].

As noted, in all of the above cases, the models described were being applied to fully ranked data. Extensions to partially ranked data are not always straightforward or well-defined. (It should be noted that Critchlow [17] used the Hausdorff metric induced by Spearman's Footrule to examine data constrained to the form $(A, B, (C, D), E)$ by fitting a generalized Mallows $\phi$ model.)
Chapter 3

Permutation Polytopes

Permutation polytopes provide a particularly elegant way of looking at ranked data graphically, yet they have as their basis a disarmingly simple idea: treat the rankings as points in $\mathbb{R}^n$. The permutation polytope then emerges as the convex hull of the rankings. This idea of using these structures to look at rankings was first put forth by Schulman [65], and was considered later by McCullagh [56] among others. Nonetheless, their use and development as natural histogram structures was initiated by Thompson [76],[77]. Further, Thompson [77] extended the idea to deal with partial rankings, arriving at the generalized permutation polytopes. This chapter is essentially an overview of the results in Thompson [77], albeit with a new notation that the author has developed.

3.1 Basics

As was mentioned in Chapter 2, any full ranking of $n$ items can be viewed as a permutation of the first $n$ integers. Given that realization, certain geometric restrictions are implied. Firstly, for any full ranking $\bar{\pi}_0$ (letting $\bar{\pi}_0$ denote the average ranking, in this case $n/2$),

$$\sum_{i=1}^{n} (\bar{\pi}_0(i) - \bar{\pi}_0)^2 = C_1,$$

a constant. This implies that the rankings lie imbedded in an $n$-dimensional hypersphere. Secondly,

$$\sum_{i=1}^{n} \bar{\pi}_0(i) = C_2,$$

another constant. This implies that the data lie imbedded in an $(n - 1)$-dimensional hyperplane. Taken together, these two conditions imply that the rankings of $n$ items lie naturally imbedded on the surface of an $(n - 1)$-dimensional hypersphere.
To transform the data to \((n - 1)\) coordinates, it suffices to center the rankings at zero (subtract off the mean ranking) and apply the Helmert transformation

\[
H = \begin{pmatrix}
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & \ldots & 0 \\
\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & 0 & \ldots & 0 \\
\frac{1}{\sqrt{12}} & \ldots & \ldots & -\frac{3}{\sqrt{12}} & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots \\
\frac{1}{\sqrt{n^2 - n}} & \frac{1}{\sqrt{n^2 - n}} & \ldots & \ldots & \ldots & -(n - 1)/\sqrt{n^2 - n} \\
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \ldots & \ldots & \ldots & 1/\sqrt{n}
\end{pmatrix}.
\]

For any centered ranking \(\vec{\pi}\), \(H\vec{\pi}^T\) is an \(n\)-vector with last entry zero. Furthermore, as the Helmert transform is orthonormal, no lengths or angles are distorted.

Consider the case of three items. These constraints suggest that the rankings must lie imbedded on a circle. This corresponds well with the regular hexagonal array illustrated in Chapter 1. Furthermore, examination of the hexagon showed that each edge had a specific interpretation. This property generalizes in such a way as to immediately lend impetus to the pursuit of this graphical technique – on the permutation polytopes, every face has a specific interpretation. Thus, visual pursuit of structure in ranked data begins with looking for clustering of data on the faces of the permutation polytopes.

Consider the case of the rankings of four items. The \(4!\) vertices lie naturally on the surface of a sphere in 3 dimensions. The resultant figure (illustrated in Figure 3.1) is known as a truncated octahedron. It has eight hexagonal faces and six square faces. It has been asserted that every face on the permutation polytope has an interpretation. On the truncated octahedron, four of the hexagons correspond to the rankings where a given item is ranked first, the other four correspond to the rankings where a given item is ranked last, and the six square faces correspond to the rankings where a specified pair of items is ranked in the first two positions (relative order within the pair is unimportant).

A basic problem of displaying the count information must now be touched on briefly - how are the relative frequencies with which each particular ranking is chosen to be indicated? The solution is to center spheres proportional in size to the corresponding frequency atop the vertices of the polytope. Of course, the phrase “proportional in size” must be made more specific.

As is noted in Cleveland [15], if circles are drawn so that the areas of the circles are directly proportional to the frequency, differences are perceptually underemphasized.
A circle corresponding to a frequency twice the magnitude of another will not appear to be "twice as large" as the other circle. Conversely, if the radii of the circles are directly proportional to the frequencies, then the differences are overemphasized. Twice the frequency will generate a circle that appears more than twice as large. An empirical correction factor is found in the work of Stevens on psychophysics – make the radii proportional to the frequency raised to the $5/7$ power (Stevens' Law). As the data from which this is derived have a high degree of variability, close approximations to $5/7$ will also work well.

### 3.2 Generalization to Partial Rankings

A thing that should be noted about the two geometric constraints given above is that the $n$ ranks $\pi_0(i)$ need not all be distinct for them to hold. Indeed, if arbitrary pseudoranks $a_i$ are assigned to the items (with the $a_i$ arranged in non-descending order), then as long as the strict inequality $a_1 < a_n$ holds, the permutations of the pseudoranks lie imbedded in an $(n-1)$-dimensional sphere. If the inequality does not hold then all of the rankings trivially degenerate to a point.

As an example of partial rankings, Thompson [77] assigns the ranking $(1,2,3,3)$ to the partial ordering $(A,B,(C,D))$. The parentheses inside the brackets serve to indicate relative indifference with respect to the items contained therein. If $\pi$ is taken to represent a partial ranking of this type (i.e. it is a vector in $\mathbb{R}^4$ whose coordinates are a permutation of the integers $1,2,3,3$), then $\pi = 9/4 = 2.25$ and

$$\sum_{i=1}^{4} (\pi(i) - \pi)^2 = 1.25^2 + .25^2 + .75^2 + .75^2 = 2.75,$$

so the vectors are imbedded in a four-dimensional sphere, and

$$\sum_{i=1}^{4} \pi(i) = 1 + 2 + 3 + 3 = 9,$$

so the vectors are imbedded in a three-dimensional hyperplane. Thus, these partial rankings are once again imbedded in the surface of an $(n-1)$-dimensional sphere. In this particular instance, the convex hull of the twelve distinct partial rankings is a truncated tetrahedron (Figure 3.2). It has four hexagonal faces and four triangular faces. Each edge of the figure is the same length. Each triangular face corresponds to a given item being ranked first; each hexagonal face corresponds to a given item being ranked last.
Figure 3.1  Truncated Octahedron: Orderings

Figure 3.2  Truncated Tetrahedron: Orderings
In this particular representation of the partial rankings the values of the rankings increase by 1 at each shift. The "tied ranks" ranking given to the partial ordering shown above is \((1, 2, 3.5, 3.5)\); the average of the two full rankings, \((1, 2, 3, 4)\) and \((1, 2, 4, 3)\), which induce this partial ordering (see Chapter 2). The objections raised to the tied ranks approach by Thompson [77] are both aesthetic and methodological. In terms of aesthetics, the generalized permutation polytope arising from the permutations of \((1, 2, 3.5, 3.5)\) is "distorted" in that some edges are longer than others. The interpretation of this is simply that the distance from \(\langle A, B, (C, D)\rangle\) to \(\langle A, C, (B, D)\rangle\) is greater than that from \(\langle A, B, (C, D)\rangle\) to \(\langle B, A, (C, D)\rangle\) (using Spearman's \(\rho\) as the metric). The use of all integer step shifts is predicated on the basic assumption that all neighboring orderings should be the same distance apart. In terms of methodology, Kendall's \(\tau\) is defined as the minimal number of pairwise adjacent transpositions required to change one ordering to another. Geometrically, each such pairwise adjacent transposition corresponds to moving along one edge of the permutation polytope. Thus, Kendall's \(\tau\) can be viewed as the length of the minimal edge path joining one ranking to another on the permutation polytope. This view is, however, problematic if the edges are of unequal lengths in that the cost of shifting along a given edge need not be one.

It should be noted that this geometric interpretation provides a quite natural extension of Kendall's \(\tau\) to partial rankings, one ranking type at a time. As Thompson [77] notes, by treating Kendall's \(\tau\) as the minimal edge path, this extension trivially satisfies the properties of a metric (positivity, symmetry in arguments, and the triangle inequality). The Hausdorff extension to Kendall's \(\tau\) proposed by Critchlow [17] yields values which correspond to the permutations of \((1, 2, 4, 4)\) for partial orderings of the form \(\langle A, B, (C, D)\rangle\). For partial orderings of the form \(\langle A, (B, C), D\rangle\), the Hausdorff metric has no readily apparent geometric interpretation.

Note that this geometric extension to partial rankings works only when the rankings are all of the same partial ranking type; combinations of \(\langle A, B, (C, D)\rangle\) and \(\langle (A, B), C, D\rangle\) (mixed ranking types) are still difficult to handle.

Thompson [77] uses linear inequality results to characterize the \(i\)-dimensional faces that can arise from a given generalized permutation polytope. Unfortunately, the notation is somewhat cumbersome. The values of the pseudoranks are given by \(0 < a_1 \leq a_2 \leq \cdots \leq a_n\). The following is taken directly from Thompson [77].

Because the permutation polytope is difficult to visualize for \(n > 4\), we will characterize all of the \(i\)-dimensional faces (\(i\)-faces) of the poly-
tope for $0 \leq i \leq n-2$, and then focus on the special characteristics of the three- and four-dimensional faces. Characterization of all of the faces requires writing a generalized permutation polytope as the solution to a system of linear inequalities. Let $N_n$ be the set $\{1, 2, \ldots, n\}$. Y[emelichev]K[ovalev]K[ravtsov] ((1984), Chapter 5, Theorem 3.1) [90] show that a permutation polytope can be defined equivalently as the intersection of the following system of linear inequalities:

$$
\sum_{i \in \omega} x_i \leq \sum_{i=1}^{\lfloor \omega \rfloor} a_{n+1-i} \quad \text{for all } \omega \subseteq N_n,
$$

(3.1)

$$
\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} a_i.
$$

(3.2)

Although YKK define permutation polytopes only for distinct values (i.e., distinct pseudoranks), their proof of this equivalence does not require the pseudoranks to be distinct.

In Theorem 3.4 of Chapter 5 YKK [90] characterize all of the faces of a permutation polytope for distinct $0 < a_1 < a_2 < \cdots < a_n$. To characterize the faces of a generalized permutation polytope, we extend their results to nondistinct pseudoranks.

**Theorem 3.1** For $0 \leq i \leq n-2$, let $\omega_1, \ldots, \omega_{n-i-1}$ be nonempty subsets of $N_n$. Then any set of solutions to

$$
\sum_{j \in \omega} x_j \leq \sum_{j=1}^{\lfloor \omega \rfloor} a_{n-j+1} \quad \text{for all } \omega \subseteq N_n,
$$

(3.3)

$$
\sum_{j \in \omega_k} x_j = \sum_{j=1}^{\lfloor \omega_k \rfloor} a_{n-j+1} \quad \text{for } k = 1, 2, \ldots, n-i
$$

(3.4)

is an $i$-face of the permutation polytope if and only if

(i) $\omega_1 \subseteq \omega_2 \subseteq \cdots \subseteq \omega_{n-i-1} \subseteq \omega_{n-i} = N_n$, and

(ii) if $|\omega_j \Delta \omega_j-1| \geq 2$, then $a_{n-|\omega_j-1|} > a_{n-|\omega_j+1|}$.

The major difference between Theorem 3.1 here and the corresponding Theorem 3.4 in Chapter 5 of YKK [90] is the inclusion of the second condition, namely, if $|\omega_j \Delta \omega_j-1| \geq 2$, then $a_{n-|\omega_j-1|} > a_{n-|\omega_j+1|}$. This condition is satisfied trivially if all the pseudoranks are distinct. If this condition is omitted with nondistinct pseudoranks, then the resulting set of solutions is a face of the polytope, but its dimension may be less than $i$. By defining $Q_k = \omega_k \setminus \omega_{k-1}$ and $j_k = |\omega_{k-1}| + 1$ for $1 \leq k \leq n-i$, Theorem 3.1 can be rephrased more usefully as follows.
Theorem 3.2  Under the assumptions of Theorem 3.1 it follows that any set of solutions to (3.3) and to
\[
\sum_{j \in Q_k} x_j = \sum_{j=j_k}^{j_{k+1}-1} a_{n-j+1} \quad \text{for } k = 1, 2, \ldots, n - i \quad (3.5)
\]
is an $i$-face of a generalized permutation polytope if and only if:

(i) $Q_1, Q_2, \ldots, Q_{n-i}$ are disjoint with $\bigcup_{j=1}^{n-i} Q_j = N_n$, and

(ii) if $|Q_j| \geq 2$, then $a_{n-|j|+1} < a_{n-|j|-1}$.

3.3 A New Notation

It is important to realize that the factors that determine the shape of the polytope produced (ignoring distortion of edges) are not the values of the pseudoranks, but rather the distribution of inequalities and equalities among them. For example, if $a_1 < a_2 < a_3 < a_4$, then the structure produced will be some (possibly distorted) version of a truncated octahedron. This can be written $(<, <, <)$; the exact values of the pseudoranks are not present, but the shape of the resultant polytope can still be inferred.

Similarly, if $a_1 < a_2 < a_3 = a_4$, then the structure produced is a (possibly distorted) truncated tetrahedron, and can be written $(<, <, =)$. This compressed inequality/equality structure shall be termed the "short form" of the polytope.

When rankings of $n$ items are considered, the number of inequalities and equalities in the short form will be $n - 1$. Also, the dimension of the polytope will be $n - 1$. Assume that one is interested in determining the $(n - 2)$-dimensional faces of the polytope. According to Thompson's [77] Theorem 3.2, this necessitates the division of the $n$ items into $n - i = n - (n - 2) = 2$ disjoint sets $Q_1$ and $Q_2$ which partition the space. Further, by (3.5),
\[
\sum_{j \in Q_1} x_j = \sum_{j=j_1}^{j_2-1} a_{n-j+1} = \sum_{j=1}^{[w_1]-1} a_{n-j+1} = \sum_{j=1}^{[Q_1]-1} a_{n-j+1},
\]
and
\[
\sum_{j \in Q_2} x_j = \sum_{j=j_2}^{j_3-1} a_{n-j+1} = \sum_{j=|Q_1|-1}^{[Q_1]-1} a_{n-j+1}
\]
so that the sum of the ranks of the items in $Q_1$ is equal to the sum of the $|Q_1|$ largest pseudoranks, and the sum of the ranks of the items in $Q_2$ is equal to the sum of the
$|Q_2|$ smallest pseudoranks. Thus, the elements of $Q_1$ are those items ranked in the last $|Q_1|$ positions, and the elements of $Q_2$ are those items ranked in the first $|Q_2|$ positions. Condition (ii) of the Theorem simply says that if there is more than one item in any subset $Q_k$, the largest and smallest ranks in that subset must be distinct.

Examine this in terms of the short form. The division of the ranking into two groups is accomplished by replacing the $|Q_2|^{th}$ entry in the short form with "". For example, assume one wished to examine the face of the truncated octahedron where one particular item was ranked first ($|Q_2| = 1$). The short form of the truncated octahedron is $(<, <, <)$, and to fix an item in first place, the first inequality is replaced with "". The short form describing that face would thus be written $(<, <, <)$. The dimensional reduction introduces a barrier between the group of items ranked "first" and the group of items ranked "last", with the geometry of each subset being defined by its own short form. The geometry of the single item being ranked first is specified by the short form "". If a judge is asked to rank a single item, he simply gives it a value by default; the geometric equivalent is a single point, as no comparison with other items is implied. The geometry of the three items ranked in the last three positions is specified by the short form $(<, <, <)$. Within this subset, a given item is ranked first, another second, and the last third. The possible arrangements trace out a hexagon. Thus, the first short form specifies a 0-dimensional space in which the first element must lie, and the second short form specifies the 2-dimensional space in which the last three elements must lie. These two spaces are orthogonal. This can be seen by considering the group of items ranked in the last three positions. A vector from one ranking, $(a_1, a_{r1,1}, a_{r1,2}, a_{r1,3})$, to another differing from the first solely by a rearrangement of items in the last three positions, $(a_1, a_{r2,1}, a_{r2,2}, a_{r2,3})$, must have the form $kI(0, a_{r1,1} - a_{r2,1}, a_{r1,2} - a_{r2,2}, a_{r1,3} - a_{r2,3})$, where $kI$ is an arbitrary constant.

A vector from one ranking, $(a_{r1}, a_2, a_3, a_4)$, to another differing from the first solely by a rearrangement of items in the first position, $(a_{r2}, a_2, a_3, a_4)$, must have the form $(a_{r1} - a_{r2}, 0, 0, 0)$ (of course, as only one item is being rearranged, $a_{r1} - a_{r2} = 0$, but the term is left in as an illustration of the general procedure). The dot product of the two vectors thus defined is zero, and hence the spaces are orthogonal. Thus, the $(n - 2)$-dimensional face specified by the conjunction $(<, <)$ is defined by the cartesian product of a point with a hexagon. As the point is a zero-dimensional object, it may be ignored when considering the shape of the resultant face; entries in the short form of $(<, <)$ may be dropped. Thus, a face of a truncated octahedron having one particular item ranked first is a hexagon.
The cartesian product formulation yields another benefit – items in the product may be rearranged without affecting the geometry. Thus, \((<, <)\) and \((<, <)(\) both describe hexagons (but with the latter corresponding to a given item ranked last). Similarly, the shape produced by \((<)(<, <)(\) is the same as that produced by \((<, <)(<)(\) or any of the other 18 distinct rearrangements of the parenthetical groups.

If a pair of items is picked to rank in the first two positions \((|Q_2| = 2)\), then the short form would be \((<)(<)\). This is the cartesian product of a segment with another segment orthogonal to the first; a rectangle (a square if \(a_2 - a_1 = a_4 - a_3\)). Thus, the short form can be used to determine all of the \((n - 2)\)-dimensional faces of the truncated octahedron. If the \((n - 3)\)-dimensional faces are desired, the allowable short form transformation is the same – replace one of the inequalities or equalities with a \)\)( pair. Note that every \(k\)-dimensional face of a permutation polytope will have \(k\) inequalities or equalities in its short form representation.

Now consider the \((n - 2)\)-dimensional faces of the truncated tetrahedron which has the short form \((<, <, =)\). Faces where a given item is ranked first are of the form \((<, =)\), a triangle. Faces where a given item is ranked last are of the form \((<, <)(\), a hexagon. Something different occurs when faces having a given pair of items ranked first are considered. These have short form \((<)(=)\). However, the short form \((=)\) is degenerate; rearrangements of the ranks within this group define no new points, and the space thus negatively defined is 0-dimensional. Thus, \((=)\) "collapses" to \(),\ a point. This is the short form equivalent of Thompson's [77] condition (ii), which simply states that no short form subset can consist solely of equalities or the dimension will drop. In this case, the implication is that faces having a pair of items ranked first have short form \((<)(=)\) which can be rewritten \((<)(\), which, as \()\)'s can be dropped, finally reduces to \((<)\), an edge (a 1-dimensional face).

Every face of a given permutation polytope can be found by starting with the short form of the whole structure, which is an \((n - 1)\)-dimensional face, and dropping one dimension at a time by replacing inequalities and equalities with \)\)( pairs. The only constraint on this operation is that no resulting parenthetical group may consist solely of equalities (or the dimensionality of the face examined will drop by more than one).

Finally, it should be observed that the resultant geometric shapes are invariant under a reversal of the order of inequalities in any parenthetical group within the short form. As \((<, <, =)\) describes a truncated tetrahedron, so does \((=, <, <)\) (albeit arising
from different partial ranking structures). This can be seen by taking a permutation polytope, centering it, reflecting it across the origin and uncentering (shifting by a constant vector until the smallest rank is 1). For example, start with $(1, 2, 3, 3)$, center it to $(-5/4, -1/4, 3/4, 3/4)$, reflect across the origin to $(5/4, 1/4, -3/4, -3/4)$, and uncenter to $(3, 2, 1, 1)$. This procedure does not affect the shape of the polytope, but it reverses the short forms being considered. In essence, the ranking scale has been reversed, so that all items ranked last are now ranked first and vice versa. That such a reversal need only apply to the entries within a given parenthetical group is due to the fact that each such group defines a space orthogonal to the others: $(<, =)(<)$ and $(=, <)(<)$ both describe the product of a triangle with a segment, a triangular prism. The resultant geometric shapes are not invariant under arbitrary permutations of the order of the inequalities and equalities within a short form. The short form $(<, =, <)$ yields a cuboctahedron, which is not geometrically equivalent to a truncated tetrahedron $(<, <, =)$: this can be seen by noting that there are square faces on the cuboctahedron $(<)(<)$ but none on the truncated tetrahedron.

### 3.4 Categorizing Faces

Thompson [77] categorizes all of the two and three dimensional faces that can occur on permutation polytopes. This is now fairly easy. The only possible two-dimensional faces are those having short forms with precisely two inequalities or equalities. These are:

- Hexagons, $(<, <)$,
- Triangles, $(<, =)$ or $(=, <)$, and
- Squares, $(<)(<)$.

Similarly, the only possible three-dimensional faces are:

- Truncated Octahedrons, $(<, <, <)$,
- Truncated Tetrahedrons, $(<, <, =)$ or $(=, <, <)$,
- Cuboctahedrons, $(<, =, <)$
- Tetrahedrons, $(<, =, =)$ or $(=, =, <)$,
- Octahedrons, $(=, <, =)$,
- Hexagonal Prisms, $(<, <)(<)$ or $(<)(<, <)$,

- Triangular Prisms, $(<, =)(<)$ or $(=, <)(<)$ or $(<)(<, =)$ or $(<)(=, <)$, and

- Cubes, $(<)(<)(<)$.

Cuboctahedrons, Tetrahedrons, and Octahedrons (with their associated orderings) are shown in Figures 3.3-3.5. There is one further observation to be made about the two- and three-dimensional faces; specifically that some of them require rankings of more than three or four items, respectively, to appear. Squares require rankings of at least four items, hexagonal and triangular prisms require at least five items, and cubes require rankings of at least six items. The minimal number of ranked items necessary to produce a given type of face can be read from the short form as the total number of equalities, inequalities, and "(" pairs, plus one.

As will be detailed in Chapter 4, it is possible to examine the four dimensional faces of a permutation polytope as well as the two and three dimensional ones. In the interests of completeness, all of the possible four dimensional faces are enumerated below (subject to arbitrary rearrangement of the parenthetical groups).

- Hyper-Truncated Octahedrons, $(<, <, <, <)$,

- Hyper-Truncated Tetrahedrons, $(<, <, <, =)$ or $(=, <, <, <)$,

- Hyper-Cuboctahedrons, $(<, <, =, <)$ or $(<, =, <, <)$,

- Hyper-Tetrahedrons, $(<, <, =, =)$ or $(=, =, <, <)$,

- Hyper-Octahedrons, $(<, =, <, =)$ or $(=, <, =, <)$,

- Long-Cuboctahedrons, $(<, =, =, <)$,

- Long-Tetrahedrons, $(<, =, =, =)$ or $(=, =, =, <)$,

- Thick-Octahedrons, $(=, <, <, =)$,

- Long-Octahedrons, $(=, <, =, =)$ or $(=, =, <, =)$,

- Truncated Octahedral Prisms, $(<, <, <)(<)$,

- Truncated Tetrahedral Prisms, $(<, <, =)(<)$ or $(=, <, <)(<)$,

- Cuboctahedral Prisms, $(<, =, <)(<)$,
Figure 3.3 Cuboctahedron: Orderings

Figure 3.4 Tetrahedron: Orderings
Figure 3.5  Octahedron: Orderings

- Tetrahedral Prisms, $(<,=,=)(<)$ or $(=,=,<)(<)$,
- Octahedral Prisms, $(=,<,=)(<)$,
- Hexahexahedrons, $(<,<)(<,<)$,
- Hexatetrahedrons, $(<,<)(<,=)$ or $(,<,<)(=,<)$,
- Tetratetrahedrons, $(<,=)(<,=)$ or $(<,=)(=,<)$ or $(=,<)(=,<)$,
- Hexagonal Hyperprisms, $(<,<)(<)(<)$,
- Triangular Hyperprisms, $(<,=)(<)(<)$ or $(=,<)(<)(<)$, and

3.5  Marginal and Orthogonal Projections

More recent work by Thompson [79] has shown that there exist certain projections of the permutation polytopes which have natural interpretations as marginal distri-
butions. An illustration of this concept is given in Figure 3.6. In these figures, a truncated octahedron is rotated so that half of the points overlap the other half.

If the last of the rotated truncated octahedrons is examined more closely, with some labels added, an interesting feature comes to light. The orderings of the points that overlap agree solely in the positioning of two items. Moreover, it is the same two items each time. This is shown in Figure 3.7, where the orderings shown agree only in the positioning of items $A$ and $B$. The resultant figure is thus said to be the marginal projection of items $A$ and $B$.

The marginal projection can also be easily arrived at from the rankings of the items. Treating the rankings as points in $\mathbb{R}^n$, so that any given ranking is of the form $(x_1, x_2, \ldots, x_n)$, item $A$ is assigned rank $x_1$ and item $B$ rank $x_2$. If the permutation polytope is then projected down onto the $x_1$-$x_2$ plane (setting all of the other coordinates to zero), the structure that remains is the marginal projection of items $A$ and $B$. Similarly, projecting the polytope onto the $x_1$ axis alone yields the marginal projection of the rankings of $A$.

An illustration of a two-dimensional marginal projection is shown in Figure 3.8, divided into stages indicating the structures corresponding to the full rankings of 3, 4 and 5 items. This projection indicates the rankings of items $A$ and $B$ only. If one is interested in the positions of $A$ and $B$ in the rankings of 3 items, $A, B, C$, the marginal projection would be the hexagon at the left edge of the figure. The labels here are (clockwise from upper left) $AB$ ($A$ first, $B$ second), $A_B$ ($A$ first, $B$ third), $AB$ ($A$ second, $B$ third), $BA$, $BA$, and $BA$. An underscore segment is used to indicate the presence of another item (or two or three) between $A$ and $B$ or in front of both. Underscores corresponding to items ranked after both $A$ and $B$ have been deleted. The reason for the "staging" of the figure is to demonstrate the effect of adding another item to the group being ranked. For example, the marginal projection corresponding to the positions of $A$ and $B$ in a ranking of 4 items is formed by adding the middle stage to that on the left, adding the vertices (clockwise from top) labelled $A\_\_B$ ($A$ first, $B$ fourth), $A\_B$ ($A$ second, $B$ fourth), $A\_B$, $B\_A$, and $B\_A$, and adding the edges forming a second hexagon and two squares so that the edge structure matches that shown in Figure 3.5. To get the marginal projection corresponding to the positions of $A$ and $B$ in the rankings of 5 items, one adds the rightmost stage.

If another item is added to the group being ranked, the geometric effect on the marginal projection of two items is the addition of another stage in a "honeycomb
Figure 3.6  Rotation of a Truncated Octahedron Suggesting a Marginal Projection.
cascade" progression; a hexagon is added along the central axis (the dashed line) and squares above and below so that the figure is convex. The central axis also serves as a boundary in the pairwise comparison of items A and B. At all vertices above the central axis, A is preferred to B, and at all vertices below the central axis B is preferred to A. Furthermore, distance above or below the central axis corresponds to the degree of separation between the rankings of A and B, which could be used in a weighted average to determine the degree to which item A could be said to be preferred to item B.

It should be noted that the hexagons in the 2-dimensional marginal projection structure are not regular. The vertices fall on integer coordinates in the \( x_1-x_2 \) plane (corresponding to the ranks of items A and B). The edge joining \((1,2,0,\ldots,0)\) and \((2,1,0,\ldots,0)\) is longer than the edge joining \((1,2,0,\ldots,0)\) and \((1,3,0,\ldots,0)\). The short edges correspond to shifts in the position of either A or B (not both) and some other item on the permutation polytope; the magnitude of the shift due to the other item is removed by the projection. The long edges correspond to shifts if the position of both A and B (the two items are interchanged); these edges are not shortened.
Figure 3.8 Marginal Distribution of items $A$ and $B$ in rankings of 3, 4 and 5 items (offset).
by the projection. For a long edge to occur (i.e., for the ranks of \( A \) and \( B \) to be interchanged), the ranks of \( A \) and \( B \) must be consecutive integers, so the only long edges are across the main diagonal in the \( x_1-x_2 \) plane (the central axis in the figure shown here).

The marginal projection of a single item, e.g. \( A \), ranked in a group of \( n \) items, is simply a segment divided into equal intervals with vertices representing \( A \), ranked first, \( A \) ranked second, \( A \) ranked third, and so on. The marginal projection of three items, e.g. \( A, B, C \), ranked in a group of \( n \) items, is akin to the honeycomb cascade found for the marginal projection of two items, save that the structure is now 3-dimensional and the shapes along the central axis are truncated octahedrons, and the shapes off to the sides are hexagonal prisms and cubes as opposed to squares.

Marginal projections can also be constructed if the data consist of partial rankings. For example, assume several people are asked to indicate their first, second and third choices out of a group of 11 items. The marginal projection corresponding to the relative placements of items \( A \) and \( B \) is shown in Figure 3.9.

Two things should be noticed here: that there now exists a vertex lying directly on the central axis, and that there are only three shapes (hexagon, hexagon, triangle) along the central axis (as opposed to \( n-2 \) hexagons arising from the full rankings of \( n \) items). These both arise from the fact that the rank of \( A \) can be 1 (first place), 2 (second place), 3 (third place), or 4 (not selected as one of the first three), and that both \( A \) and \( B \) can have rank 4 (neither is selected among the first three).

Actually, the marginal projection of any two items can be constructed given the shapes of the consecutive items along the central axis and the realization that the marginal projection must be convex (squares are added above and below to the figures of the central axis until the figure is convex). That the marginal projection is a convex figure follows from the fact that it is a projection of a generalized permutation polytope (which is a convex figure by definition). Alternatively, it can be noted that if \( A \) is ranked \( k-1 \) and \( B \) is ranked \( k \), this implies the existence of at least \( k \) distinct nonempty ranks, so that any pairing \((i,j), \; i,j \in \{1,\ldots,k\}, \; i \neq j\) must exist as a vertex. The only vertices not determined in this way are those corresponding to \( A \) and \( B \) having the same rank, and such vertices must lie along the central axis, as the central axis coincides with the main diagonal in the \( x_1-x_2 \) plane.

The shapes that lie on the central axis can be derived in sequence from the short form of the ranking structure. In the case mentioned above, the short form is \((<,<,<,=,=,=,=,=,=,=))\). The first figure on the central axis is defined by the first two
entries in the short form – \((<, <)\), a hexagon. Similarly, the next figure is defined by the second and third entries in the short form – again \((<, <)\), a hexagon. The third figure is determined by the third and fourth entries – \((<, =)\), a triangle pointing to the right. All further figures have short form \((=, =)\) and are degenerate. For a 2-dimensional marginal projection, only 3 distinct shapes can appear along the central axis (the degenerate point is ignored): \((<, <)\), a hexagon, \((<, =)\), a triangle pointing to the right, and \((=, <)\), a triangle pointing to the left.

The marginal projection of any three items can be constructed in a similar manner, although there are now 7 distinct shapes that can appear along the central axis: \((<, <, <)\), a truncated octahedron, \((<, <, =)\), a truncated tetrahedron pointing to the right, \((<, =, <)\), a cuboctahedron, \((<, =, =)\), a tetrahedron pointing to the right, \((=, <, <)\), a truncated tetrahedron pointing to the left, \((=, <, =)\), an octahedron, and \((=, =, <)\), a tetrahedron pointing to the left.
Chapter 4

Automating Permutation Polytopes

In an early paper outlining the usage of the permutation polytopes, Thompson [76] explicitly notes the need for the development of interactive software to make such analysis of ranked data a practical option. Such software now exists. These routines enable us to draw, rotate, and illuminate features (faces) of the generalized permutation polytopes arising from the rankings of 4 or 5 items. This chapter is devoted to the options contained therein.

4.1 Full Rankings

The most basic element, of course, is the construction of a wireframe illustrating the connection structure of the permutation polytope and the application of appropriately sized spheres to the vertices. Of only slightly less importance is the ability to rotate the structure in real time (animation). Empirically, people seem to find it much easier to detect clusters in point clouds if the clouds are in motion. It is unclear whether the wireframe outline aids or hinders in the initial location of structure; its major benefit lies in delineating the connections and suggesting interpretations once structure has been found. Hence, options are included allowing the user to hide the wireframe and/or highlight specific faces.

Animation is impossible to show clearly in text. The other basics named above (wireframe with spheres, hidden wireframe and highlighting) are shown in Figures 4.1-4.2. The data shown are from a German Political survey conducted in 1975 (see Appendix A); 2262 Germans were asked to rank four societal goals. These goals were:

- A – Maintaining order in the nation,
- B – Giving people more say in government,
- C – Fighting rising prices, and
- D – Protecting freedom of speech.
Figure 4.1  German Political Data: Orderings and Wireframe

Figure 4.2  German Political Data: AC face highlighted, Wireframe hidden.
Figure 4.1 is the starting representation: the fully ranked data and the polytope wireframe are shown, and each vertex is labelled with the appropriate ordering. Vertices can be labelled with orderings or rankings, or the labels can be hidden. There is a clear cluster in the data on the square face corresponding to items A and C being ranked in the first two positions. The sizes of the circles are proportional to the \( \frac{5}{7} \) power of the frequency of occurrence, according to Stevens' Law (see Chapter 3).

Figure 4.2 shows the representation after some options have been exercised: the wireframe and the labelings have been hidden, and a specific face \((A, C)\) first has been highlighted (edges and labels of the highlighted face being shown by default). Highlighting is accomplished by including the edges and vertices of the appropriate face in a different color.

An effective Exploratory Data Analysis (EDA) procedure seems to be: 1) Start the basic wireframe and sphere structure, with labels hidden. 2) Rotate the structure for several seconds and try to pick out clusters. Hiding the wireframe for this stage may help. 3) Having detected clusters, activate the labels to identify significant points, and highlight (one at a time) the faces including these points to identify how the structure breaks down into components.

The permutation polytopes corresponding to the rankings of 5 items can also in some sense be examined \textit{in toto}. These rankings lie imbedded in a 4-dimensional hypersphere; the structure must be “mapped down” one dimension. This can be accomplished via stereographic projection. If a regular structure (say a chandelier) is suspended in midair and a light source is placed on one side of it (a light bulb above), then a shadow is cast by the structure (onto the floor). By looking at the shadow cast by an \( n \)-dimensional object onto a “flat” \((n - 1)\)-dimensional surface, the structure has been mapped down one dimension to \((n - 2)\)-dimensions. This procedure is illustrated in Figure 4.3. A light source has been placed to the left of a wireframe of a cube, and the shadow is cast on a screen to the right.

Several features of this type of projection are apparent. First, the relative lengths of edges on the original figure are not preserved; the amount of distortion varies with the proximity of the edge to the light source. Second, the relative topology is preserved; if vertex \( A \) is connected to vertex \( B \) on the original figure, the shadow of vertex \( A \) is connected to the shadow of vertex \( B \). Third, the shape of the shadow is very dependent upon the placement of the light source. Certain viewpoints can emphasize inherent symmetries in the original figure. For the cube, the most “symmetric” shadow is cast when the light source is suspended directly above one of the
square faces. For a regular 4-dimensional polytope, the most symmetric shadows are cast when the 4-dimensional "light source" is placed directly above the center of one of the 3-dimensional faces. The structure arising from the full rankings of 5 items has short form \((<, <, <, <, <)\), and it can readily be shown that this structure possesses only two types of 3-dimensional faces: truncated octahedrons (arising as the short forms \(()(<, <, <)\) and \((<, <, <))\)) and hexagonal prisms (arising as the short forms \((<)(<, <)\) and \((<, <)(<))\). For more on the short form, the reader is referred to Chapter 3.

Examples of this structure are shown in Figures 4.4-4.6. The data shown are from an election held among the Rice University faculty (see Appendix A). In 1992, a search committee was formed to find a new president for the University. The election in question was to select a faculty member to serve on this committee. Five candidates were considered. The ballots were preferential, so that a full ballot consisted of a full ranking of the five candidates. Some respondents returned only partially completed ballots (indicating their first-place choice only, for example). A full discussion of how
to deal with mixtures of full and partial rankings is deferred to Section 2; attention at this point shall be restricted to the full rankings.

Figure 4.4 illustrates the wireframe arising as the permutation polytope associated with the full rankings of five items, with the projection point being centered above one of the truncated octahedral faces. With this structure, we are confronted with problem of “edge noise” — there are so many connecting edges that picking out specific faces (and the appropriate subsets of connections) is difficult. The distinguishing characteristic of this data set is that the data are clustered on a particular hexagonal prism (highlighted), which in this case corresponds to the three candidates $A$, $B$ and $C$ being ranked in the first three positions. In terms of the election, this says that the first three were viable candidates, and that the last two did not factor in the election. The projection point in this case was centered above the truncated octahedral face corresponding to candidate $A$ being ranked last.

Figure 4.5 illustrates the same data set, with the wireframe hidden save for the highlighted face, but with the projection point shifted (albeit still atop one of the truncated octahedral faces; in this case the projection point is centered atop the truncated octahedral face corresponding to candidate $A$ being ranked first). There is no single “best” viewing angle in an objective sense; exploration is required.

Figure 4.6 shows the same data set, but the projection point in this case has been placed over the center of a hexagonal prism. The resulting structure does not appear as spherically symmetric as that shown in Figure 4.4; rather it appears symmetric about a central axis. This “lopsided” effect does not seem to be conducive to the location of structure. A suggested EDA procedure is to flip through the projection points corresponding to truncated octahedrons only, rotating at each stage both with and without the wireframes.

### 4.2 Mixed Rankings

The Rice election referred to in Section 1 contained ballots that were not full — some people specified their first choice only, some their first and second choices only, some their first, second and third choices only, and the rest gave complete ballots (if only four choices are specified, the remaining item is assigned fifth place by default). Negative votes, in the sense of indicating someone who should not be chosen, were not allowed short of a full ballot (no one could indicate a first and fifth choice without specifying an order for the remaining three). The problem to be considered is how to
Figure 4.4  Projected Polytope arising from the full rankings of five candidates. The highlighted hexagonal prism corresponds to candidates $D$ and $E$ being ranked in the last two positions.
Figure 4.5  Projected Polytope with wireframe hidden save for a highlighted face. The highlighted hexagonal prism corresponds to candidates $D$ and $E$ being ranked in the last two positions.
Figure 4.6  Projected Polytope corresponding to the full rankings of five candidates, with the projection point centered above the hexagonal prism corresponding to candidates $A$ and $B$ being ranked in the first two positions. The highlighted face corresponds to candidates $D$ and $E$ being in the last two positions.
deal with such a mixture of full and partial rankings. This problem is quite common in practice, especially if the number of items being ranked is large.

4.2.1 Plotting

A question that must be addressed is how to plot full and partial rankings on the same figure. One possible solution is to reallocate the mass accorded to less fully specified rankings to the full rankings which might have produced them. There are several drawbacks to this approach. The first is that it is not clear what proportions should be used in the reallocation process. If the ranking comes from a preferential ballot, a partial ranking may reasonably be taken as an indication of indifference with respect to the items not ranked, so that the mass should be equally spread over all applicable full rankings. However, if the rankings are indicative of the order of appearance of symptoms in the progression of a disease, a partial ranking is more appropriately treated as a censored observation. In this case, the mass should be spread over the full rankings in accordance with the conditional distribution on the full rankings given the partial ranking, in a manner akin to that proposed by Kaplan and Meier [41].

The second drawback to the redistribution approach is that the ability to visually distinguish between full and partial rankings is lost. Indeed, at first sight it would not even be apparent that there had been any partial rankings at all.

The solution proposed, then, is to plot the partial rankings as spheres at the center of the appropriate edge, face or solid described by the fuller rankings. Analytically, this corresponds to averaging the ranks not specified in the partial ranking when determining coordinates in \( \mathbb{R}^n \). For example, the partial ranking corresponding to \( \langle A, B, (C, D) \rangle \) would be \( (1,2,3.5,3.5) \), the average of \( (1,2,3,4) \) and \( (1,2,4,3) \), the rankings corresponding to \( \langle A, B, C, D \rangle \) and \( \langle A, B, D, C \rangle \), respectively. Similarly, a partial ranking corresponding to only the first choice out of four being given would be plotted at the appropriate permutation of \( (1,3,3,3) \). This avoids the two drawbacks listed above, but can cause the display to seem overly cluttered at times. To deal with this, options have been included to hide or display mass according to the short form representation of the ranking in question. For example, one could choose to display only those rankings which were full \( (<, <, <) \) or where the ranker had specified only a first choice \( (<, =, =) \). The default is that only full rankings are shown. The short form of a given ranking can be said to determine the “type” of the ranking and the associated mass; in the sequel, “mass types” will be used to indicate masses arising
from rankings with different short forms. If a specific face is highlighted, all rankings on that face, both full and partial, are highlighted if that mass type is shown at all. The complete Rice voting data (with all ranking types present) is shown in Figure 4.7.

Rankings with different short forms can also be differentiated by assigning different colors to each, as is shown in Figure 4.7. Unfortunately, this can lead to a “Christmas-tree” effect if too many different short form types are displayed – the presence of too many colors in discrete chunks can make it difficult to pick out small clusters of dots which are important. In an initial analysis, run through the displays of all mass types, simply to get a feel for what mass types are present, and restrict comparisons to two or three different mass types at once.

Another question that arises when the partial rankings are plotted this way is how large to draw the spheres. One solution is to scale each mass type separately, so that distributions of different mass types are easy to compare. This procedure, however, disguises the relative frequencies with which different mass types are chosen. The solution used here is simply to draw all mass types to the same scale, so that the question of whether one ranking had been chosen twice as often as another can be answered, regardless of the mass type.

4.2.2 Fusing

In the subsection above, a base assumption has been that the wireframe in question should correspond to that needed to properly display the fullest rankings available. In this subsection, the effects of using a less complicated wireframe are explored.

Consider the case of the rankings of the candidates in the Rice election, as given above. Assume that the distinction between fourth and fifth place is unimportant. The data can then be plotted on a structure having short form \((\lt, \lt, \lt, \eq,\rangle\), with the orderings \((A, B, C, (D, E))\), \((A, B, C, D, E)\) and \((A, B, C, E, D)\) all being mapped to the same vertex. The abovementioned orderings have been “fused” in that the ability to distinguish between them has been removed. Thus, the process of moving from one generalized permutation polytope to another simpler one (equivalent to introducing an indifference, or changing an inequality to an equality in the short form) shall be referred to as fusing. This particular fusing is shown in Figure 4.8.

Fusing’s main disadvantage lies in the inherent indistinguishability associated with fused full and partial rankings. Conversely, this indistinguishable property reduces
Figure 4.7  Rice Voting Data, full and partial rankings shown. Full
rankings are red, first, second and third place specified are magenta, first and
second place specified are aqua, and first place only is blue.
Figure 4.8  Rice Voting Data, with positions 4 and 5 fused. The highlighted hexagon corresponds to candidates D and E being ranked in the last two positions.
high-frequency fluctuations and emphasizes gross features, so that fusing can be thought of as a smoothing operation. Another method of smoothing, using duality, will be addressed in a later chapter.

When fusing is used, the question arises as to how to display the various types of mass. In the above example, masses of type \((<, <, <, <)\) and \((<, <, <, =)\) are both plotted at the vertices of a generalized permutation polytope. If only the mass type corresponding to full rankings is displayed, then the counts for both of the full rankings corresponding to a single vertex are summed, and displayed spheres are sized appropriately. If both mass types are shown, the three counts are summed, and the spheres are resized. Dividing the spheres into pie slices via coloring (to indicate the relative proportions of each mass type present) was rejected as visually distracting. In general, if two mass types which are both shown are plotted to the same location after fusing, their respective counts are combined.

Fusing provides a quick way to visually decompose the structure in ranked data. If only first-place information is desired about the items, the full structure \((<, <, <, <)\) can be fused down to the first-place structure \((<, =, =, =)\). Last place information can be obtained by fusing to the last place structure \((=, =, =, <)\).

When rankings of five or more items are considered, fusing also has the salutary effect of reducing edge noise as it reduces the number of edges. Starting from the full rankings of \(n\) items \(((n - 1) <)\), there are \(2^{(n-1)} - 1\) possible fused structures (shifting all of the inequalities to equalities yields a degenerate structure). For \(n = 5\), this yields fifteen distinct fusings; a small enough number to be considered. It is often useful to examine at least the \((n - 1)\) fusing structures corresponding to the first \(k\) places being known (\(k\) ranging from 1 to \((n - 1))\).

### 4.3 High-Dimensional Options

If more than five items are being ranked, viewing the entire edge structure is infeasible. Projections involving only dot clouds representing the vertices are possible, but have not been pursued here.

One option is to note that the relative ranks of 3, 4 or 5 specified items can be examined. For example, if the items \(A-F\) have been assigned the ranking \((1, 5, 6, 3, 4, 2)\), then the relative ranking of the items \(A-D\) is \((1, 3, 4, 2)\). The relative rankings of these four items can then be plotted on a truncated octahedron. This does assume that the relative rankings remain unchanged by the removal of certain items from consid-
eration; this assumption is not always valid. A further drawback of this approach is that the distances between clusters can be distorted. For example, if item \( C \) is always ranked in sixth place, including it in the group of items being examined deemphasizes the degree of separation between it and the other items.

A second option relies on the fact that all of the possible 2, 3 and 4 dimensional faces of any ranking structure can be characterized. Using this characterization, computer searches can be conducted to locate the particular faces with the greatest clusterings of mass (those faces being assumed to be the ones of interest. For example, given the full ranking of six items as given above, the natural five-dimensional structure has short form \( (\), \( (\), \( (\), \( (\), \( (\), so all of the truncated octahedral faces can be characterized by: \( ()(\)\), holding one item fixed in first place, another fixed in second, and permuting the remaining four; \( ()(\)\), holding one item fixed in first place, one item fixed in last place and permuting the middle four; or \( ()()\), holding one item fixed in last place, another fixed in the next-to-last place and permuting the first four. This procedure does not bring clusters of mass closer to one another (as does the item-deletion procedure); on the other hand it does not display all of the mass (except in certain rare instances).

4.4 Software Notes

All of these routines were developed on a Silicon Graphics workstation. The routines are in C and make use of the Silicon Graphics shared graphics library.

4.4.1 four

At present, the software is comprised of a set of programs for analyzing different types of ranked data. The most general program, "four", analyzes rankings of four items. If more than four items have been ranked, the user will be prompted for a subset of four (the relative ranks of which will then be displayed).

A sample run (without pictures) is presented below.

> four
Please enter the name of the file in which your data is located.
> german.dat
There are 4 distinct items, specifically
Order
Say
Prices
Speech

Please type in the 4 numbers corresponding to the items that you wish to examine. These four will be relabelled A, B, C, D.

> 1 2 3 4

Some commands for manipulating the display:
x - rotate about the x-axis 5 degrees (pos)
y - rotate about the y-axis 5 degrees (pos)
z - rotate about the z-axis 5 degrees (pos)
c - rotate about the x-axis 5 degrees (neg)
b - rotate about the y-axis 5 degrees (neg)
a - rotate about the z-axis 5 degrees (neg)
otherwise, use the right-hand mouse button.

Rotation of the permutation polytope is controlled from the keyboard. All other features are menu-driven. When the right-hand mouse button is pressed, the main menu is summoned. The main menu options are given below.

fuse
mass
label
highlight
print
hidelattice

The first five of these options lead to their own submenus.

fuse

The "fuse" option governs the shape of the displayed permutation polytope. The default figure which is initially displayed is the truncated octahedron, which has short form (<, <, <). The other generalized permutation polytopes involving the rankings of four items can be accessed by toggling the inequalities in the short form to equalities and vice versa. The fuse submenu consists of

(1,2)
(2,3)
(3, 4)

which correspond to the first, second, and third inequalities in the short form, respectively. If the user desires to ignore the distinction between third and fourth place (treating items ranked in these positions as tied), the fuse/(3, 4) option is chosen. The resulting short form is (<, <, =), the corresponding truncated tetrahedron is drawn, and the displayed masses are redistributed to their new positions. To return to the truncated octahedron, the fuse/(3, 4) option is chosen again. The degenerate (=, =, =) short form has been disabled; the user cannot shift all of the inequalities to equalities at the same time.

mass

The "mass" option governs what data points are shown on the permutation polytope. The default is that only points corresponding to full rankings of the four items are shown. The mass submenu is given below.

(<, <, <)
(<, <, =)
(<, =, <)
(<, =, =)
(=, <, <)
(=, <, =)
(=, =, <)
(=, =, =)

Points corresponding to partial rankings can be added by toggling the corresponding short form. For example, to display all data points which might arise in a preferential election where negative voting is not allowed, the user would toggle (<, <, =), (<, =, =) and perhaps (=, =, =). The (<, <, <) option is toggled by default, but can be untoggled.

If the polytope has been fused, various mass types may be plotted to the same location. For example, if the fuse/(3, 4) option has been chosen, then the masses with types (<, <, <) and (<, <, =) are both plotted at the vertices of the corresponding truncated tetrahedron. If both mass types are toggled, the counts at each vertex are summed and a single larger dot is shown for the combination.
label

The "label" option labels the vertices of the permutation polytope being displayed. The submenu is given below.

rankings
orderings

Either rankings or orderings may be displayed, but not both at the same time. Labels are only shown at the vertices. Partial orderings are indicated with parentheses. As an example, if the fuse/(3,4) option has been chosen, one of the vertices would be labelled 1233 (ranking) or AB(CD) (ordering).

highlight

The "highlight" option accentuates one particular face of the permutation polytope being displayed. The default setting is that no faces are highlighted. This option has its own submenus. The basic highlight submenu is given below.

hexagon
square
triangle

Only face types actually appearing on the permutation polytope being displayed can be chosen; e.g. the triangle option is disabled for the truncated octahedron. The subsubmenus serve to identify which specific face of a given type is to be highlighted. The submenu corresponding to highlight/hexagon is given below.

A first
B first
C first
D first
A last
B last
C last
D last

As with the initial highlight submenu, only those options corresponding to faces present on the displayed polytope are enabled. When a face is highlighted, the edges
are drawn a different color than the edges of the rest of the polytope, and all displayed mass occurring on that face (at vertex, edge, or face center) is colored differently from the coloring given to regular mass.

Only one face on a polytope can be highlighted at a time. Further, the highlight option is fragile with respect to fusing, as the faces of the polytope change. If a hexagon on the truncated octahedron is highlighted, and the \texttt{fuse/(3,4)} option is then chosen, the resultant truncated tetrahedron will have no faces highlighted.

\textbf{print}

The "\texttt{print}" option dumps the displayed figure to a PostScript file. The associated submenu, \texttt{grayscale color}

is self-explanatory. This option is not yet perfected.

\textbf{hidelattice}

The "\texttt{hidelattice}" option, when toggled, erases the edge structure associated with the permutation polytope, leaving only the masses displayed. This option is overridden by the \texttt{highlight} option: if a given face is highlighted, the associated edges will be drawn and the rest of the polytope edge structure will be absent.

\textbf{4.4.2 \texttt{five}}

The programs associated with the rankings of five items are both more and less sophisticated at present. Other options are included to deal with the difficulties arising from portraying 4-dimensional figures, but the programs associated with different permutation polytopes have not been integrated, so there is no \texttt{fuse} option. Further, the various \texttt{five} programs are entirely keyboard driven.

A sample run of \texttt{five12345}, which generates the permutation polytope associated with the full rankings of five items, is provided below (without pictures).

\texttt{> five12345}

Please enter the name of the file in which your data is located. Enter a zero if no data set is involved.

(if a zero is entered, the basic edge structure will be drawn.)
> psych.dat There are 5 different variables. Please type in the 5 numbers corresponding to the positions of those that you wish to examine. For example, if the data represent ballots for candidates A, B, C, D, E, F and G respectively, an entry of 1 3 5 4 7 would correspond to choosing A, C, D, E and G.
> 1 2 3 4 5
The size of the masses is given by
(point mass/total mass)*k, where k is
a scaling factor. By using the same value
of k across several data sets, pdfs can be
compared in a straightforward way. Finding
the ideal value of k is a matter of experiment.
Please input the scaling factor k
> 1.5
Some commands for manipulating the display:
x - rotate about the x-axis 5 degrees (pos)
y - rotate about the y-axis 5 degrees (pos)
z - rotate about the z-axis 5 degrees (pos)
c - rotate about the x-axis 5 degrees (neg)
b - rotate about the y-axis 5 degrees (neg)
a - rotate about the z-axis 5 degrees (neg)
t - examine the data from a truncated octahedron (10 exist)
p - examine the data from a hexagonal prism (20)
h - examine the data from a hexagon (60)
s - examine the data from a square (90)
n - examine the data from the next face of the specified type
m - examine the data from the preceding face
d - show (hide) mass due to fully ranked data
e - show (hide) mass due to data where two entries are equal
f - show (hide) mass due to data with three equal (or 2 pairs)
g - show (hide) mass due to data with four equal (or 3 and 2)
l - show (hide) labels
o - switch between rankings (WHITE) and orderings (BLUE)
r - internal (external) view
i - highlight a truncated octahedron
j - examine from the next truncated octahedron
k - examine from the preceding truncated octahedron
esc - escape from the display

The first element which is different from the presentation of four is the inclusion of the scaling factor k. The inclusion of this factor is more important when five items are being ranked due to the distortion of edges under projection; when four items are being ranked, the sphere sizes are scaled so that the largest sphere’s radius is one-third of the length of an edge.

The rotation commands are the same as they were for four. The next several key commands (t, p, h, s, n, m) concern the location of the 4-dimensional “light source” used to project a three-dimensional shadow. As the projected edge structure exhibits more symmetry when the light source is situated on the hypersphere directly above a face, these commands indicate which face is to be used. The default is a truncated octahedral face. Other locations corresponding to the same type of face (initially other truncated octahedrons) can be cycled through by repeatedly hitting n or m.

The “show mass” key commands are cruder than in four. The default is that only fully ranked data are shown (d is toggled). Toggling e displays all masses whose corresponding short forms have one equality: (<, <, <, =), (<, <, =, <), (<, =, <, <) or (=, <, <, <). Toggling f displays masses whose short forms have two equalities, and toggling g displays masses whose short forms have three equalities.

The label activation and ordering/ranking switches (1 and o) are straightforward. The internal (external) view key, r, is different. Initially, when the structure is projected from the hypersphere, some points are mapped farther from the center than others, and can fall out of range of the screen display. This is the default (internal) view. Toggling r once the data have been projected causes the location of the viewing point to “recede” for a panoramic view - the entire structure can be seen.

The highlighting triumvirate i, j, k serves a more limited function here than in four. Here, the highlighting of various truncated octahedral faces serves to illustrate the magnitude of the distortion present. Toggling i causes a truncated octahedral face to be highlighted, and the other such faces can be highlighted in turn by repeatedly hitting j or k.

Finally, the esc option is clear.

Work to update and generalize these programs is ongoing.
Chapter 5

Duals

As the number of items being ranked increases, the high dimensionality and sheer number of vertices on the permutation polytopes makes perception of structure difficult. It is thus worthwhile to try to find related polytopes that may not suffer from these problems to the same degree. One method of constructing related polytopes which has been around for hundreds of years exploits the "duality" between faces and vertices of a given polytope. This method is most clearly illustrated in three dimensions through a consideration of the Euler equation \( v - e + f = 2 \), which states that for any convex polyhedron in three dimensions, the number of vertices \( v \) less the number of edges \( e \) plus the number of faces \( f \) is a constant (2). In the above equation, \( v \) and \( f \) appear symmetrically, so if one polyhedron satisfies the above equation, there exists another polyhedron which satisfies the equation with the number of vertices and faces reversed. Every vertex on the initial polyhedron is replaced with a face, and every face is replaced with a vertex. This method can be generalized to higher dimensions, replacing high-dimensional faces with vertices and vice-versa. Thus, employing duality, while it does not reduce the dimensionality of the problem, tends to reduce the number of vertices and gives a new way of looking for structure. It should be noted that the duals, as explained here, are just for full rankings at present.

5.1 Smoothing Ranked Data

The main idea behind using duality to examine data from permutation polytopes is quite simple. In looking for clusters on the permutation polytopes, the focus of interest lies in determining which \((n - 2)\)-dimensional faces exhibit the highest mass density. Random fluctuations, essentially equivalent to high-frequency noise, can obscure larger trends. In order to detect these larger trends a smoothing method of some type can be employed.

The use of smoothing to estimate continuous underlying distribution functions has been the subject of much research in recent years. A good recent overview of the work
done in this field is provided by Scott [67]. In the univariate case, a sample $x_1, \ldots, x_n$ of data values is assumed to come from some continuous underlying distribution. It is this underlying distribution (or, equivalently, the associated probability density function) which is the feature of interest. A first approximation to this distribution is given by the empirical distribution function. This corresponds to a discrete "density" function where probability mass of $1/n$ is assigned each of the observed values $x_i$, resulting in a collection of Dirac spikes. As a collection of spikes is not a $C_1$ function, each spike is generally replaced by a continuous scaled distribution, a kernel, thus "smoothing" the mass over a localized region. The overall value of the density at each point is then estimated by summing the contributions from each kernel at that point. This procedure is illustrated in Figure 5.1. Six data points have been observed, and a Gaussian kernel with probability mass $1/6$ has been centered over each. The resultant density estimate is shown as a solid curve.

Smoothing when the underlying distribution is known to be discrete (and the collection of possible outcomes is known) is more problematic. When the underlying distribution is continuous, the observations must be smoothed, as the probability of any exact value being replicated is zero, but the probability of another data value being observed in the same localized area is positive. This justification is not valid when the underlying distribution is discrete. Probability mass can no longer be spread across a localized continuum; rather some fraction of the probability mass is redistributed across other possible outcomes which are in some sense near the first. This "discrete kernel" method was explored by Aitchison and Aitken [1] with respect to a binary hypercube. Possible outcomes were points in $\mathbb{R}^n$ whose coordinates were entirely composed of some combination of ones and zeros: if one point was observed, some mass was also allotted to all of the $n$ other possible outcomes differing from the first in a single entry. This method was used by Titterington et. al. [85] to help classify severity of head injuries based on a collection of dichotomous and categorical predictive factors. Graphically, however, there is a drawback in that the data have not obviously been smoothed: all of the probability mass remains concentrated on the set of possible outcomes. If a smooth curve has been derived from observations from a continuous distribution, some smoothing is visibly occurring. An alternative method of smoothing that is visually obvious as smoothing is to place the averages of localized groups of possible outcomes at points which are not possible outcomes.

The method proposed here is to replace each $(n - 2)$-dimensional face with a point, and plot on that point the average mass density of the corresponding face.
Figure 5.1  Kernel Estimate of a continuous Density Function. Observed data values are indicated by vertical bars. Gaussian Kernels are indicated by dashed lines. The Density Estimate is shown as a solid curve.

This reduces the problem of finding the face with the highest mass density to the problem of finding the single largest dot. The particular connection structure to use, corresponding to which dots are connected to which, will be explored later in this section. It is now productive to examine what types of \((n - 2)\)-dimensional faces can exist on the initial permutation polytope.

From the short form representation (see Chapter 3), it follows that every \((n - 2)\)-dimensional face is of a specific structural form: \(k\) of the items are ranked "first", and the remaining \(n - k\) items are ranked "last". For example, on the truncated octahedron, the faces are hexagons with one item ranked first and three items ranked last, squares with two items ranked first and two items ranked last, and hexagons with three items ranked first and one item ranked last. The division corresponds to the location at which a "\(\langle\)" pair is substituted for an inequality or equality in the short form.

The dichotomy between first and last place items suggests a connection with a multivariate binary structure, in that items may be assigned a 1 or a 0 as they belong to the first or last groups, respectively. This in turn suggests a natural connection structure corresponding to the edges of the unit hypercube. For example, the vertex
corresponding to the square face of a truncated octahedron with items \( A \) and \( B \) ranked first would be located at \((1,1,0,0)\), and the hypercube structure would connect this vertex with \((1,0,0,0)\) \((A \text{ first, all others last})\), \((0,1,0,0)\) \((B \text{ first, all others last})\), \((1,1,1,0)\) \((D \text{ last, all others first})\), and \((1,1,0,1)\) \((C \text{ last, all others first})\). The dual of a truncated octahedron is shown in Figure 5.2 The hypercube is not quite a perfect model, however, in that the vertices of all zeros and all ones are absent (they correspond to degenerate rankings). This necessitates the following

**Definition 5.1** The connection structure of the dual of a permutation polytope corresponding to the rankings of \( n \) items is the \((n - 1)\)-dimensional structure that remains when the main diagonal is removed from the \( n \)-dimensional unit hypercube via orthogonal decomposition.

Consider the case of the rankings of three items \( A, B \) and \( C \). The edges of the permutation polytope in this case correspond to \( A \) first; \( A, B \) (unordered) first (or \( C \) last); \( B \) first; \( B, C \) first (or \( A \) last); \( C \) first; and \( A, C \) first (or \( B \) last) (going around the hexagon). These correspond to the points \((1,0,0),(1,1,0),(0,1,0),(0,1,1),(0,0,1)\) and \((1,0,1)\), respectively, shown on the cube in Figure 5.3. The resulting dual can be viewed by rotating this figure so as to sight along the main diagonal (the visual equivalent of the orthogonal decomposition) and removing all edges connected to the disallowed points (these edges are shown in gray). Hence, the dual is the projection of the highlighted figure onto the hyperplane perpendicular to the main diagonal.

As each edge of the hypercube corresponds to a unit shift along a single axis, all edges are shortened the same amount by removing the component moving in the direction of the main diagonal (a unit shift along all axes simultaneously) by symmetry. All edges of the dual permutation polytopes, then, have the same length, which yields a well-defined analog of Kendall’s \( \tau \) (when viewed as the length of the minimal edge path connecting two specified vertices).

It is obvious from the definition that the dual permutation polytope corresponding to the full rankings of \( n \) items has \( 2^n - 2 \) vertices, which for \( n > 3 \) is less than \( n! \). This saves much computational effort. In Figure 5.4, the German Political Data has been plotted on the dual of the truncated octahedron. The dot at the vertex corresponding to items \( A \) and \( C \) being ranked first is clearly the largest. Furthermore, in the German Political Data there appears to be a “decay” process at work in that the mass on a given vertex apparently decreases exponentially with the distance of that vertex from the mode (where the mode in this case is the \( AC \) vertex and the
distance in question is Kendall’s $\tau$). One of the more common parametric models on rankings is Mallows’ $\phi$ model, where the probability of a given ranking $\pi$ occurring is given by $P(\pi) = C(\phi) \exp\{-\phi \tau(\pi, \pi_0)\}$, where $\pi_0$ is the modal ranking, $\phi$ is a decay parameter, and $C(\phi)$ is a normalization constant. As the exponential decay seems as if it would be a good fit, an analog of the Mallows $\phi$ model is fitted to the data via maximum likelihood. The modal vertex is $AC$ in this case, so the only parameter remaining to be estimated is the decay constant $\phi$. The likelihood function is

$$L(\phi) = C(\phi)^n e^{-n_1\phi} e^{-2n_2\phi} e^{-3n_3\phi} e^{-4n_4\phi},$$

where $n_1$ is the number of observations lying on vertices one edge away from the mode, $n_2$ is the number of observations lying on vertices two edges away from the mode, and so on. $C(\phi)$ is

$$(c_0 + c_1 e^{-\phi} + c_2 e^{-2\phi} + c_3 e^{-3\phi} + c_4 e^{-4\phi})^{-1},$$

where $c_i$ is the number of vertices $i$ steps away from the mode, in this case 1, 4, 4, 4, 1 respectively. Maximizing the likelihood function then reduces to solving a fourth
Figure 5.3  Rotation of the Cube to form the Dual
order equation in $e^{-\phi}$; in general fitting a Mallows model requires solving a $k^{th}$ order equation where $k$ is the greatest attainable distance away from the mode. The straight residuals, observed values minus fitted values, are shown in Figure 5.5, scaled to be comparable in size with the first figure. In Figure 5.5, red indicates positive and blue negative residuals. The obvious top/bottom division in the residuals suggests the presence of low-level information hidden by the predominant characteristic, which was not apparent by other means (in this case, a slight preference for items $B$ and $C$ over items $A$ and $D$).

It should be emphasized at this point that not only has added structure in the data been revealed, but also that entirely new models on rankings are suggested by looking at the analogs of standard ranking models on duals and mapping them back to the rankings. This approach to new models is a subject for future work.

There is a problem with the duals that arises from its smoothing nature. Starting from data on permutation polytope, the process of averaging over a given face to arrive at the mass on a vertex of the dual introduces correlation between the vertex counts. Starting from a truncated octahedron, for example, the counts associated with the ordering $\langle A, B, C, D \rangle$ are divided among the three dual vertices corresponding to the hexagon defined by $A$ first, the square defined by $A, B$ first, and the hexagon defined by $A, B, C$ first. Visual determination of the correlation is impaired by the fact that it is not fully reflected in the connection structure of the dual: the vertex corresponding to $A$ first is not joined to the vertex corresponding to $A, B, C$ first, yet the two are correlated. Thus, drawing a structural inference due to a perceived pattern (as was done above in the case of the dual decay) must be approached with caution, but determination of the largest cluster is safe.

### 5.2 Duals and Multivariate Binary Data

The dual formulation addresses a somewhat different question than the permutation polytope. Whereas the permutation polytope displays the prevalence of a particular ranking of the items in a population, the dual displays the prevalence of a particular division of the items into "liked" and "disliked" groups. Whereas a minimal shift on the permutation polytope, a single edge, corresponds to an interchange in the ordering of two adjacent items, a minimal shift on the dual, again a single edge, corresponds to the addition or subtraction of a single element from the group of items "liked". This is the philosophical underpinning to the connection of the duals with multivariate
Figure 5.4  Dual of the German Political Data

Figure 5.5  Residuals after fitting Mallows' $\phi$ Model
binary data. This linkage can be used to examine the converse problem: given a collection of \( n \) items divided into liked and disliked groups by \( k \) judges, infer the relative ranking of the \( n \) items.

The following data taken from Solomon [70] serves to illustrate this. In a survey, 2982 high school seniors (divided into high and low IQ groups) were asked to agree or disagree with four statements:

- \( A \) – the development of new ideas is the scientist’s greatest source of satisfaction,
- \( B \) – scientists and engineers should be eliminated from the military draft,
- \( C \) – the scientist will make his maximum contribution to society when he has freedom to work on problems which interest him, and
- \( D \) – the monetary compensation of a Nobel Prize winner in physics should be at least equal to that given popular entertainers.

The data are listed in Appendix A, and plotted on two dual polytopes, shown in Figures 5.6-5.7.

It is immediately obvious that the discrepancy between the two groups is small, with the data clustering on the same face of the dual in both instances. The question now arises as to what a face on the dual represents. To get an \((n - 2)\) dimensional face on the dual, one fixes the “liked” states of two items and permutes the remaining \((n - 2)\). Further, of the two items fixed, one must be “liked” and the other “disliked”, for if the two agree, then permuting the remaining \((n - 2)\) will include either the entry of all ones or the entry of all zeros as a vertex. So, a face on the dual corresponds to one item being liked, and another disliked, corresponding to a paired comparison of the two items. In the IQ response data, the clustering is on the face corresponding to item \( A \) being liked and item \( B \) being disliked (everybody believes that scientists enjoy developing new ideas, but they shouldn’t be excluded from the draft), leading to the orderings \( (A, C, D, B) \) and \( (A, D, C, B) \). Discriminating between the rankings of \( C \) and \( D \) is then accomplished by contrasting the diametrically opposed faces corresponding to \( C \) liked and \( D \) disliked and \( C \) disliked and \( D \) liked. In the case of this particular data set, the distinction between the two does not seem that strong, although vertex \( AC \) has noticeably more mass than vertex \( AD \), so that \( (A, C, D, B) \) is the most likely true ordering. However, the partial ordering \( (A, (C, D), B) \) is certainly plausible. Another thing worthy of note is that the circle at the vertex corresponding
Figure 5.6  Dual of Low IQ Responses.

Figure 5.7  Dual of High IQ Responses.
to $A, C, D$ first is decidedly larger for the high IQ group, suggesting that the "distance" between items $A$ and $(C, D)$ is less for the high IQ group than it is for the low IQ group. This particular interpretation in light of information which the duals do not show – the grouping of "all items liked" is more prevalent (122 to 62) and the grouping of "all items disliked" is less prevalent (53 to 82) in the high IQ group than in the low IQ group.
Chapter 6

Paired Comparisons

A situation that often arises is that the rankings of $n$ items are not given per se; rather, the data consist of the outcomes from some subset of the $\binom{n}{2}$ possible paired comparisons, from which a ranking must be inferred. For example, if the three items $A, B, C$ were to be compared in pairs, the collection of outcomes $A > B$, $A > C$ and $B > C$ would imply the ordering $(A, B, C)$ and the associated ranking $(1, 2, 3)$ (here and in the sequel, $A > B$ means that item $A$ is preferred to item $B$). Many models on rankings arise from considering the probabilities associated with the individual paired comparisons; in particular, the Bradley-Terry models are based on estimating the pairwise probabilities $p_{ij}$ that item $i$ is preferred to item $j$, and Mallows uses the collection of paired probabilities to formulate his $\phi, \theta$ models.

Paired comparisons are of interest in their own right; early studies include those of Kendall and Babington Smith [48] and Kendall [46]. An extensive modern discussion of paired comparisons can be found in David’s monograph [21].

A common problem that arises with paired comparisons is that some collections of outcomes may be internally inconsistent, e.g. $A > B$, $B > C$, and $C > A$. These inconsistent triples were denoted circular triads by Kendall and Babington Smith [48]. How to deal with these inconsistent groupings is a matter of some dispute. In most formulations of ranking models, inconsistent collections of outcomes are simply ignored. The Bradley-Terry and Mallows models mentioned above assign probabilities to rankings based on the probabilities associated with the outcomes of the individual paired comparisons, conditional on the collection of outcomes being consistent with a ranking. However, some information can occasionally be gleaned from inconsistent collections. Given four items and the collection of outcomes $A > B$, $A > C$, $A > D$ and $B > C$, $C > D$, $D > B$, it can be asserted that item $A$ is preferred to the other three, even though the relative ordering amongst items $B$, $C$ and $D$ is not discernible.

In the rest of this chapter, a geometric formulation of the collections of paired comparisons will be examined, and a direct link back to the permutation polytopes will be derived.
6.1 Permutation Polytopes as Projections of Hypercubes

When a comparison of two items is conducted, the outcome is binary – either item $A$ is preferred to item $B$, or item $B$ is preferred to item $A$. Each paired comparison can thus be thought of geometrically as defining an axis, for example the $AB$ axis, where the particular outcome determines the value along that axis: 1 if (using the example above) $A$ is preferred to $B$, and $-1 (-AB)$ if $B$ is preferred to $A$. In this manner, a space containing the overall structure arising from a collection of paired comparisons can be specified by the cartesian product of these axes, and the collections of outcomes of the $\binom{n}{2}$ paired comparisons arising from the possible pairings of 2 out of $n$ items can be viewed as vertices on an $\binom{n}{2}$-dimensional hypercube, with coordinates of either 1 or $-1$. These vertices can then be viewed as points in $\mathbb{R}^{\binom{n}{2}}$. As an example, consider the collections of paired comparisons possible among three items, $A, B, C$. If (for purposes of orientation) the axes defined are taken to correspond to $AB, AC$ and $BC$, respectively, then $(1, 1, 1)$ would indicate $A > B$, $A > C$, and $B > C$, $(-1, -1, 1)$ would indicate $A < B$, $A < C$ and $B > C$, and so on. The eight possible collections correspond to the vertices of a cube in three dimensions.

To address the problem of inconsistent sets of comparisons, consider the cube defined by the collections of paired comparisons of three items. Of the eight vertices, two correspond to triples which are linearly inconsistent: $(1, -1, 1)$ and $(-1, 1, -1)$, using the $AB-AC-BC$ coordinate system as before. The other six vertices correspond to the six possible rankings of three items. A particularly noteworthy fact about the two inconsistent triples is that they both lie along a single vector through the origin, $\overline{ABC} = (1, -1, 1)$. (in the sequel, the notation $\overline{IJK}$ will be used to indicate the vector corresponding to an inconsistent arrangement of the three arbitrary items $I, J, K$, specifically $I > J$, $J > K$ and $K > I$). This vector can be thought of as defining the “inconsistent subspace” associated with this set of paired comparisons; this in turn suggests the existence of a “consistent subspace”. According to this formulation, the linear (ranking) information present within a collection of paired comparisons can be viewed as a function of the projection of the vector associated with that particular vertex of the hypercube onto the consistent subspace. This is illustrated in Figure 6.1.

For this general procedure to be valid, it needs to be established that such a decomposition into consistent and inconsistent subspaces is feasible. As a first step, note that any triplet of items can give rise to inconsistent pairings, and as there are $\binom{n}{3}$
Figure 6.1 Cube of Paired Comparisons of Three Items, and the Associated Projection Onto the Consistent Subspace. Inconsistent Triples are Indicated by Circles.
such triplets, the vectors corresponding to inconsistencies must be linearly dependent, as \( \binom{n}{3} \) grows faster than \( \binom{n}{2} \) (the dimension of the hypercube). It is thus necessary to establish the dimension of the space spanned by the vectors corresponding to these inconsistent triplets.

A means of approaching this problem is suggested by a consideration of the vectors corresponding to the inconsistent triplets arising from the comparisons of four items. These are shown in the rows of Table 6.1. The arrangement of the axis labels across the top of the table is such that new items can be added on at the right without necessitating a revision (for example, adding item \( E \) to the group of items being considered adds \( AE, BE, CE \) and \( DE \) off to the right). An entry of 1 in the table indicates that the specified pair was preferred in the given order, a \(-1\) indicates that the pair was preferred in the reverse order, and a 0 that no direct pairing has occurred.

<table>
<thead>
<tr>
<th>Axis Labels</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>AD</th>
<th>BD</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inconsistent</td>
<td>( ABC )</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( ABC ) points to</td>
<td>( ABD )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>( A &gt; B, B &gt; C, C &gt; A )</td>
<td>( BCD )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
</tbody>
</table>

The vectors are clearly linearly dependent; this can be seen by noting that \( BCD = ABC - ABD + ACD \). However, if attention is constrained solely to the triples containing a specific item (e.g., \( A \)), those vectors are not linearly dependent.

**Lemma 6.1** The vectors corresponding to inconsistent triplets involving item \( A \) are linearly independent.

**Proof** For any items \( I \) and \( J \), the vector \( AIJ \) has a 1 in the entry corresponding to the \( IJ \) axis, and the vectors corresponding to all other inconsistent triplets containing \( A \) have a 0. Thus, \( AIJ \) cannot be formed as a linear combination of such vectors. \( \square \)
Lemma 6.2 Any vector corresponding to an inconsistent triple can be expressed as a linear combination of vectors corresponding to inconsistent triples involving item \( A \).

Proof As the lemma is trivially true if the inconsistent triple involves \( A \), it suffices to show that it holds for an arbitrary inconsistent triple \( I, J, K \) not involving \( A \). This is most easily shown using unit vectors corresponding to the various paired comparison axes; \( \overrightarrow{AB} = \overrightarrow{e}_{AB} - \overrightarrow{e}_{AC} + \overrightarrow{e}_{BC} \), for example.

\[
\overrightarrow{IJK} = \overrightarrow{e}_{IJ} - \overrightarrow{e}_{IK} + \overrightarrow{e}_{JK}
\]

\[
= \overrightarrow{e}_{IJ} - \overrightarrow{e}_{IK} + \overrightarrow{e}_{JK} + (\overrightarrow{e}_{AI} - \overrightarrow{e}_{AI}) + (\overrightarrow{e}_{AJ} - \overrightarrow{e}_{AJ}) + (\overrightarrow{e}_{AK} - \overrightarrow{e}_{AK})
\]

\[
= (\overrightarrow{e}_{AI} - \overrightarrow{e}_{AJ} + \overrightarrow{e}_{IJ}) - (\overrightarrow{e}_{AI} - \overrightarrow{e}_{AK} + \overrightarrow{e}_{IK}) + (\overrightarrow{e}_{AJ} - \overrightarrow{e}_{AK} + \overrightarrow{e}_{JK})
\]

\[
= \overrightarrow{AIJ} - \overrightarrow{AJK} + \overrightarrow{AJK}.
\]

Lemma 6.3 Any vector corresponding to an inconsistent \( k \)-tuple (for example, \( \overrightarrow{ABCD} = \overrightarrow{e}_{AB} + \overrightarrow{e}_{BC} + \overrightarrow{e}_{CD} - \overrightarrow{e}_{AD} \) corresponds to an inconsistent 4-tuple) can be written as a linear combination of vectors corresponding to inconsistent triples.

Proof The lemma holds if trivially if \( k = 3 \), and inconsistency is impossible if \( k < 3 \), so the lemma holds then as well. If \( k > 3 \), the vector corresponding to the inconsistent \( k \)-tuple can be written as the sum of a vector corresponding to an inconsistent triple and a vector corresponding to an inconsistent \( (k-1) \)-tuple as follows:

\[
\overrightarrow{IJKLN} = \overrightarrow{e}_{IJ} + \overrightarrow{e}_{JK} + \overrightarrow{e}_{KL} + \ldots + \overrightarrow{e}_{IN}
\]

\[
= (\overrightarrow{e}_{IK} - \overrightarrow{e}_{IK}) + \overrightarrow{e}_{IJ} + \overrightarrow{e}_{JK} + \overrightarrow{e}_{KL} + \ldots - \overrightarrow{e}_{IN}
\]

\[
= (\overrightarrow{e}_{II} - \overrightarrow{e}_{IK} + \overrightarrow{e}_{JK}) + \overrightarrow{e}_{IK} + \overrightarrow{e}_{KL} + \ldots - \overrightarrow{e}_{IN}
\]

\[
= \overrightarrow{IJK} + \overrightarrow{IKL} + \ldots + \overrightarrow{N}.
\]

The \( (k-1) \)-tuple can then be reduced in turn, and the lemma follows by induction. \( \square \)
Theorem 6.1  Given the space defined by the \( \binom{n}{2} \) axes associated with the paired comparisons possible among \( n \) items, the set of vectors corresponding to inconsistent triples involving item \( A \) forms a basis for the inconsistent subspace.

Proof  The theorem follows immediately from Lemmas 6.1-6.3.  \( \Box \)

Corollary 6.1  The dimension of the inconsistent subspace is \( \binom{n-1}{2} \).

Corollary 6.2  The consistent subspace exists and has dimension \( \binom{n}{2} - \binom{n-1}{2} = n - 1 \).

It is now desirable to find an orthonormal basis for the consistent subspace. First, a generic basis is found. Consider the consistent vectors corresponding to a particular item beating all others. These “first-place” vectors are shown in Table 6.2 for the case \( n = 4 \). These vectors shall be marked with a “1” suffix: the vector corresponding to item \( A \) beating all others will be denoted \( \vec{A1} \). As these vectors are consistent, they are orthogonal to the basis vectors of the inconsistent subspace. As there are \( n \) first-place vectors, they are of necessity linearly dependent. On the other hand, any \( (n-1) \) of them are not. Remove from consideration the vector corresponding to item \( A \) beating all others. The vector corresponding to an arbitrary item \( I \) beating all others has a \(-1\) in the \( AI \) column, and all other first-place vectors have 0 in that column. Thus, any \( (n-1) \) of the first-place vectors form a basis for the consistent subspace.

<table>
<thead>
<tr>
<th>Axis Labels</th>
<th>AB</th>
<th>AC</th>
<th>BC</th>
<th>AD</th>
<th>BD</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \vec{A1} )</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>First-Place</td>
<td>( \vec{B1} )</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vectors</td>
<td>( \vec{C1} )</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>( \vec{D1} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>
It remains to orthonormalize the above basis. The dot product of any first-place vector $\vec{I}$ with itself is $n - 1$, as there are $n - 1$ paired comparisons involving $I$. For any pair of first-place vectors, $\vec{I}, \vec{J}$, the dot product is $\vec{I} \cdot \vec{J} = -1$ as the only component in which $\vec{I}$ and $\vec{J}$ are jointly nonzero corresponds to the paired comparison $IJ$, which is 1 in $\vec{I}$ (as $I$ is preferred to $J$) and $-1$ in $\vec{J}$ (as $J$ is preferred to $I$). The vectors are thus symmetric with respect to their positioning in terms of one another — a good analog is the arrangement of vectors from the center to the vertices of a simplex such as a triangle or tetrahedron.

**Lemma 6.4** There exists a constant $\alpha$ such that the collection of vectors of the form $\vec{I} + \alpha \vec{A}$, where $I$ can be any item other than $A$, forms an orthogonal basis for the consistent subspace.

**Proof** As there are $(n - 1)$ such vectors, it suffices to show that there exists an $\alpha$ such that all pairs of such vectors are orthogonal and that none of the vectors are null. For any arbitrary pair of items $I, J$ other than $A$,

$$(\vec{I} + \alpha \vec{A}) \cdot (\vec{J} + \alpha \vec{A}) = \vec{I} \cdot \vec{J} + \alpha \vec{A} \cdot \vec{J} + \alpha \vec{A} \cdot \vec{I} + \alpha^2 \vec{A} \cdot \vec{A}$$

$$= (n - 1)\alpha^2 - 2\alpha - 1$$

$$= 0 \iff \alpha = \frac{1 \pm \sqrt{n}}{n - 1}.$$

Using either value of $\alpha$ found above,

$$(\vec{I} + \alpha \vec{A}) \cdot (\vec{I} + \alpha \vec{A}) = \vec{I} \cdot \vec{I} + 2\alpha \vec{A} \cdot \vec{I} + \alpha^2 \vec{A} \cdot \vec{A}$$

$$= (n - 1) - 2\alpha + (n - 1)\alpha^2$$

$$= \{(n - 1)\alpha^2 - 2\alpha - 1\} + n$$

$$= n,$$

so the vectors are not null.

\[\square\]

**Theorem 6.2** The collection of vectors $\{((\vec{I} + \alpha \vec{A})/\sqrt{n})\}$, where $I$ is any item other than $A$ and $\alpha = \frac{1 \pm \sqrt{n}}{n - 1}$, forms an orthonormal basis for the consistent subspace.
Proof The theorem follows immediately from Lemma 6.4.

Putting this into use, the projection of the 6-dimensional hypercube arising from the pairwise comparisons of 4 distinct items onto the corresponding 3-dimensional consistent subspace is shown in Figure 6.2. The completely consistent sets of paired comparisons correspond to rankings of the four items; these are indicated in red. In terms of cartesian coordinates, these vertices lie at the 24 permutations of \((0, \pm 1, \pm 2)\). Those sets of paired comparisons with only one inconsistent triple (slightly inconsistent) are shown in purple; these each have multiplicity 2 (i.e., two vertices of the initial hypercube map to each such point). In terms of cartesian coordinates, these vertices lie at the 8 permutations of \((\pm 1, \pm 1, \pm 1)\). Finally, those sets of paired comparisons with two inconsistent triples (grossly inconsistent) are shown in blue; these each have multiplicity 4. In terms of cartesian coordinates, these vertices lie at the 6 permutations of \((0, 0, \pm 1)\). It is impossible to have more than two inconsistent triples arising in the paired comparisons of 4 items (this will be shown later in this chapter). The edges defining the convex hull of these points are also shown.

Several features of this figure should be noted. First and foremost, the structure in which these points are imbedded is the permutation polytope corresponding to the rankings of four items. Second, the informational component in inconsistent paired comparisons is geometrically apparent and readily interpretable. Third, distances between different types of pairings are identical, in the sense that the distance from the vertex corresponding to the full ranking \(CBAD\) to the vertex corresponding to the full ranking \(CBDA\) is the same as the distance from the vertex at \(CBAD\) to the vertex at the center of the hexagonal face corresponding to \(C\) being ranked first, which is in turn the same as the distance from the vertex at \(CBAD\) to the vertex behind the square face corresponding to \(CB\) being ranked in the first two positions. This can be seen by considering the cartesian coordinates \((0, 1, 2)\), \((1, 1, 1)\), and \((0, 0, 1)\). These vertices are connected by the projected edges of the hypercube. Thus, there exists a well-defined analog of Kendall's \(\tau\) for determining the distance between two sets of paired comparisons.

These features all generalize nicely as the number of items being compared increases.

Lemma 6.5 The convex hull of the projection of a paired comparison hypercube onto its consistent subspace is a permutation polytope.
Figure 6.2 Paired Comparison Hypercube with Inconsistent Triples Removed.
Proof  To see this, consider the $n$-vector 

$$\vec{v}_N = (\vec{v} \cdot \vec{A}_1, \vec{v} \cdot \vec{B}_1, \ldots, \vec{v} \cdot \vec{N}_1),$$

where $\vec{v}$ is a vector pointing to a vertex on the hypercube. The vector $\vec{v}$ can be broken down into a linear combination of the unit vectors associated with the individual paired comparisons. The unit vector $\vec{e}_{IJ}$ associated with the pair of items $I, J$ will contribute 1 to the $I^{th}$ component of $\vec{v}_N$ and $-1$ to the $J^{th}$ component of $\vec{v}_N$. As each unit vector will contribute positive and negative amounts in equal proportion,

$$\sum_{i=1}^{n} \vec{v}_N(i) = 0,$$

so the data are constrained to lie in $(n - 1)$ dimensions (this was known before, but this method of presentation emphasizes the similarity to the permutation polytope formulation better). Further, if $\vec{v}$ corresponds to a full ranking, then there must be a first place item, a second place item, and so on up to an $n^{th}$ place item. As a result, there will be one component of $\vec{v}_N$ which has value $(n-1)$ (the first place item), one component which has value $(n-3)$ (the second place item), and so on up to a component which has value $-(n-1)$ (the $n^{th}$ place item). Note that the step decrement is 2, not 1, in that the value of an item is given by the number of items which it is preferred to less the number of items preferred to it. If the value were taken to be solely the number of items to which it is preferred, the step decrement would be 1. Thus, the $n!$ collections of paired comparisons which correspond to full rankings map to the $n!$ elements of $\mathbb{R}^n$ whose components are permutations of the $n$ distinct integers $-(n-1), -(n-3), \ldots, (n-3), (n-1)$. Finally, the definitions of $\vec{v}_N$ and of $\vec{v}$ lead to linear inequality constraints analogous to those associated with the permutation polytopes: none of the $n$ components of $\vec{v}_N$ can be greater than $(n-1)$ or less than $-(n-1)$, no pair of components of $\vec{v}_N$ can be greater than $2(n-2)$ or less than $-2(n-2)$ and so on. The permutations of $n$ distinct integers and the associated linear inequalities define a permutation polytope.  

\[\square\]

Lemma 6.6  The ranking associated with a given collection of paired comparisons can be inferred from the face of the permutation polytope
which is indicated by the projection of the collection onto the consistent subspace.

**Proof** Using the vectors \( \vec{u} \) and \( \vec{v}_N \) as defined in Lemma 6.5, note that if the vector \( \vec{u} \) corresponds to a full ranking, then the vector \( \vec{r} = \frac{1}{2}(n + 1)\vec{1} - \vec{v}_N \) has an entry of 1 for the first place item, 2 for the second place item, and so on; in short, \( \vec{r} \) is the ranking associated with that particular collection of paired comparisons. The vector \( \vec{r} \) still yields relative rankings for vectors \( \vec{u} \) which correspond to collections of paired comparisons which contain inconsistencies: the ranking assigned item \( I \) is determined solely by the number of items beaten directly by \( I \). As an example, consider the case of a collection of paired comparisons among four items which is "maximally inconsistent". There have been 6 comparisons made, and of the four items, two have been favored in two paired comparisons each, and the other two have been favored in just one paired comparison each (any deviation from this allows for the definition of either a clear winner or a clear loser). The vector \( \vec{v}_N \) will have components which are some permutation of \((1, 1, -1, -1)\), and the vector \( \vec{r} \) will have components which are some permutation of \((2, 2, 3, 3)\) corresponding to one pair of items being preferred to the other pair. The preferred items are those which were favored in two paired comparisons each. \( \Box \)

**Lemma 6.7** Each edge of the paired comparison hypercube is shortened equally by projection onto the consistent subspace.

**Proof** The argument is based on symmetry. Every edge of the initial hypercube corresponds to a shift along a single unit vector \( \vec{e}_{IJ} \). As the inconsistent subspace contains vectors containing inconsistent groupings where every unit vector is represented equally, and as the initial hypercube was symmetric with respect to the unit vectors (all combinations were present) each edge should be reduced in a symmetric manner. Alternatively, the magnitude of a particular unit vector onto the orthonormal basis for the consistent subspace derived earlier can be examined. Consider the unit vector \( \vec{e}_{IJ} \). If neither \( I \) nor \( J \) is item \( A \), then the projection onto the basis defined by the collection of vectors \((\vec{1} + \alpha)/\sqrt{n}\) is a vector with
two nonzero components: $1/\sqrt{n}$ and $-1/\sqrt{n}$, respectively, so that the squared magnitude of the projection is $2/n$. If the unit vector in question involves item $A$, then the projection onto the basis is a vector with $(n - 1)$ nonzero components: $(n - 2)$ components of $\alpha/\sqrt{n}$ and one component of $(\alpha - 1)/\sqrt{n}$ (up to a change of sign for all the components). The squared magnitude of this vector is

$$\frac{(n - 2)\alpha^2 + (\alpha - 1)^2}{n} = \frac{((n - 1)\alpha^2 - 2\alpha - 1) + 2}{n},$$

which, as $\alpha$ is defined to be a root of the quadratic equation given in braces above, reduces to $2/n$. Hence, all unit vectors project equally. □

The metric analogs of Kendall's $\tau$ will be discussed further in Section 6.4.

### 6.2 Ties, Incomplete Pairings, and Partial Rankings

In attempting to rank $n$ items via paired comparisons, problems can arise. One such problem is that certain paired comparisons may not have been made (leading to an incomplete collection). Alternatively, a given paired comparison may have been made, but the outcome has been a tie. The geometric solution is the same in both cases: the entry for that paired comparison is recorded as a 0 as opposed to the 1 or $-1$ recorded for a definite outcome.

If ties and incomplete pairings are allowed, the geometric structure becomes more complex. The initial hypercube is still of dimension $\binom{n}{2}$, and the vectors defining the inconsistent and consistent subspaces remain the same, but the number of allowable vertices has increased from $2^{\binom{n}{2}}$ to $3^{\binom{n}{2}}$. The projection onto the consistent subspace arising from the paired comparisons of three items is shown in Figure 6.3. The locations of the projections of each of the vertices are labelled with the coordinates of the vertex. The vertex at the top of the figure corresponds to the vertex $(1,1,1)$ on the initial cube, indicating $A > B$, $A > C$ and $B > C$. Going clockwise around the exterior of the hexagon, the vertex at the center of the edge in the upper right corresponds to the vertex $(1,1,0)$ on the initial cube, indicating $A > B$, $A > C$ and $B = C$. Vertices on the interior of the hexagon have multiplicities greater than 1; for example, the center vertex is mapped to by the three vertices $(1,-1,1)$, $(0,0,0)$ and $(-1,1,-1)$ of the initial cube.
Figure 6.3 Consistent subspace projection of paired comparisons of three items, with ties allowed. The first coordinate represents $AB$, the second $AC$, and the third $BC$. 
At this point, it should be noted that any partial ranking can be given as a set of paired comparisons if ties are allowed. For example, the partial ranking corresponding to the item \( A \) being ranked first consists of ones in the \( AI \) entries for all items \( I \), and zeros in all the others. Thus, the cube of paired comparisons with ties allowed contains all full and partial rankings. The projections of these points agree with the procedure of replacing the ranks of tied items with their average ranking before plotting them in \( \mathbb{R}^n \).

6.3 Linear Information and Choice Consistency

Kendall and Babington Smith [48] address the problem of determining the number of inconsistent triples, \( d \), that can arise given a complete set of paired comparisons with no ties allowed. They prove that

- The maximum number of inconsistent triples is

\[
(n^3 - n)/24 \quad n \text{ odd},
\]

\[
(n^3 - 4n)/24 \quad n \text{ even};
\]

and the minimum number is zero.

- These limits can always be attained by some configuration of preferences.

- For any integral number between the maximum and the minimum there exists at least one preference-configuration with that number of inconsistent triples; and in general there will be more than one.

Furthermore, they show that the number of inconsistent triples present in a given collection of paired comparisons is given by

\[
d = \frac{n^3 - n}{24} - \frac{1}{2} \sum_{i=1}^{n} (a_i - \bar{a})^2,
\]

where \( a_i \) denotes the number of items beaten by the \( i^{th} \) item, and \( \bar{a} \) denotes the average number of items beaten (namely \( (n - 1)/2 \)). The proof of this is developed through an enumeration argument involving the edges of a completely connected directed graph (this approach will be explored later in this chapter). They use this to define a coefficient of consistence, \( \zeta \), as

\[
\zeta = 1 - \frac{24d}{n^3 - n}, \quad n \text{ odd},
\]

\[
= 1 - \frac{24d}{n^3 - 4n}, \quad n \text{ even},
\]
where the second terms on the right are derived from the maximum number of inconsistent triples. This coefficient is then used to determine the capability of a given judge to accurately discriminate among a group of items. A $\zeta$ value of 1 indicates that the judge has been "perfectly consistent" - the collection of outcomes of the paired comparisons leads unambiguously to a ranking of the $n$ items. Conversely, a $\zeta$ value of 0 indicates that the judge has been "maximally inconsistent" - there exists no item whose ranking relative to any other is unambiguous. Kendall and Babington Smith [48] provide tables of the distribution of $d$ for sets of paired comparisons among $n$ items up to $n = 7$; given these and $\zeta$ one can determine whether a given set of paired comparisons is "too consistent" to have arisen by chance.

Kendall and Babington Smith [48] also conjectured that

- $E(d) = \frac{1}{4} \binom{n}{3}$,

- $Var(d) = \frac{3}{16} \binom{n}{3}$, and that

- the distribution of $d$ tends to normality as $n$ increases.

These properties were later proved by Moran [58], who observed that if $I_{IJK}$ denotes the indicator function of items $I, J, K$ forming an inconsistent triple, then $E(I_{IJK}) = \frac{1}{4}$. This can be seen by noting that on the cube arising from the collection of possible paired comparisons amongst the three items in question, precisely two of the eight vertices are inconsistent. Similar enumerative arguments are used to prove the various properties listed above.

In terms of being able to evaluate the strength of a ranking geometrically, a more appropriate measure of the linear order inherent in the set of paired comparisons is the coefficient of linear information, $\zeta_l$.

**Definition 6.1** The coefficient of linear information, denoted $\zeta_l$, is the ratio of the magnitude of the projection of a set of paired comparisons onto the consistent subspace to the magnitude of the projection of a set of paired comparisons corresponding to a full ranking onto the consistent subspace.

For complete sets of paired comparisons (no ties and no missing values), this statistic is a monotonic transform of the coefficient of consistence $\zeta$. The particular relationship can be derived as follows.
Let \( \vec{v} \) denote a vector of the \( \binom{n}{2} \) paired comparisons among \( n \) items. Further, let \( \vec{v} \) be complete, in that no ties are allowed in the paired comparisons. Then the projection of \( \vec{v} \) onto the consistent subspace is

\[
\frac{1}{\sqrt{n}}(\vec{v} \cdot \vec{B} + \alpha \vec{A} \vec{I}), \vec{v} \cdot \vec{C} + \alpha \vec{A} \vec{I}), \ldots, \vec{v} \cdot (\vec{N} + \alpha \vec{A} \vec{I})
\]

(in terms of the orthonormal basis defined earlier). The squared magnitude of this vector is thus

\[
\frac{1}{n} \left( \sum_{I \in \{B,C, \ldots, N\}} (\vec{v} \cdot \vec{I})^2 + (n - 1)\alpha^2(\vec{v} \cdot \vec{A} \vec{I}) + 2\alpha(\vec{v} \cdot \vec{A} \vec{I}) \sum_{I \in \{B,C, \ldots, N\}} (\vec{v} \cdot \vec{I}) \right).
\]

As was observed earlier, \( \sum_{I \in \{A,B, \ldots, N\}} \vec{v} \cdot \vec{I} = 0 \) for any vector \( \vec{v} \); hence the last sum in the above expression can be replaced by \(-\vec{v} \cdot \vec{A} \vec{I}\). The sum then becomes

\[
\frac{1}{n} \left( \sum_{I \in \{B,C, \ldots, N\}} (\vec{v} \cdot \vec{I})^2 + [(n - 1)\alpha^2 - 2\alpha - 1] + 1(\vec{v} \cdot \vec{A} \vec{I})^2 \right),
\]

which, as the quadratic expression in braces disappears by the definition of \( \alpha \), reduces to

\[
\frac{1}{n} \sum_{I \in \{A,B, \ldots, N\}} (\vec{v} \cdot \vec{I})^2.
\]

Now, \( \vec{v} \cdot \vec{I} = 2a_I - (n - 1) \), where \( a_I \) is the number of items defeated by item \( I \) in direct comparison. Thus,

\[
\frac{1}{n} \sum_{I \in \{A,B, \ldots, N\}} (\vec{v} \cdot \vec{I})^2 = \frac{1}{n} \sum (2a_I - (n - 1))^2.
\]

If the items are then ordered, this can be rewritten as

\[
\frac{1}{n} \sum_{i=1}^{n} (2a_i - 2((n - 1)/2))^2,
\]

which, noting that the average number of items beaten by another is \((n-1)/2\), reduces to

\[
\frac{4}{n} \sum_{i=1}^{n} (a_i - \bar{a})^2.
\]

Recalling that Kendall and Babington Smith [48] showed that

\[
d = \frac{n^3 - n}{24} - \frac{1}{2} \sum_{i=1}^{n} (a_i - \bar{a})^2,
\]
the squared magnitude can be rewritten
\[
\frac{4}{n} \sum_{i=1}^{n} (a_i - \bar{a})^2 = \frac{n^2 - 1}{3} - \frac{8d}{n}.
\]
Thus, the ratio of the squared magnitude of an arbitrary vector \( \mathbf{v} \) to the squared magnitude of a vector corresponding to a full ranking \((d = 0)\), is
\[
\frac{\frac{1}{3}(n^2 - 1) - \frac{1}{n} 18d}{\frac{1}{3}(n^2 - 1)} = 1 - \frac{24d}{n^3 - n},
\]
the coefficient of consistence \( \zeta \). The ratio of the magnitudes, \( \zeta_l \), is simply the square root of \( \zeta \).

A restriction should be noted. The ratio of the magnitudes \( \zeta_l \) is directly related to the definition of \( \zeta \) only when \( n \) is odd. The reason for this is that Kendall and Babington Smith [48] defined the coefficient of consistence so that it would take on the value zero for a complete vector of paired rankings having the maximal number of inconsistent triples. This is reasonable for purposes of defining internal consistence, but flawed in the ranking context in that a complete vector (no ties) is required. If a complete vector is given, and \( n \) is even, the amount of ranking information present is nonzero even if the vector is inconsistent to the greatest degree possible. However, if the vector is allowed to be incomplete (in that ties are allowed), it is possible to arrive at an arrangement with no usable ranking information.

Another problem arises if ties are allowed. Consider, for example, the case of a judge faced with four items, \( A, B, C, D \). If the judge responds that \( A \) beats \( B, C \) and \( D \), but that the latter three are indistinguishable, then the response is intuitively "fully consistent", even though \( \zeta \) and \( \zeta_l \) are both less than one.

This suggests a revision of the coefficient of consistence of the following form: given that exactly \( k \) of the \( \binom{n}{2} \) paired comparisons are not ties, define \( \zeta^* \) as the ratio of the squared magnitude of the projection of the set of paired comparisons onto the consistent subspace to the maximal squared magnitude of the projection onto the consistent subspace of any set of paired comparisons exactly \( k \) of which are not ties. In the event that all \( \binom{n}{2} \) paired comparisons yield non-ties, this statistic reduces to the coefficient of consistence defined above. A problem remains, however, in that there are only certain values of \( k \) such that a \( \zeta^* \) value of 1 corresponds to a situation that most people would regard as "fully consistent" in that transitivity of the rankings is implied. Consider, for example, the case of three items, \( A, B, C \). If \( k = 1 \), then transitivity cannot hold. Assume (without loss of generality) that the
definite judgement corresponds to item $A$ beating item $B$; then $A = C$ and $B = C$ but $A \neq B$. Collections of paired comparisons which are transitive correspond to full or partial rankings of the items. If $k = \binom{n}{2}$ then a full ranking, and hence transitivity, is possible. The minimal nondegenerate value of $k$ for which transitivity can be achieved is thus $n - 1$, corresponding to a single item being ranked either first or last. This suggests that in the computation of $\xi^*$, the magnitude in the denominator should correspond to that arising from a set of $k'$ paired comparisons, where $k'$ is taken to be the smallest integer value greater than or equal to $k$ such that transitivity can be achieved. Under this definition, only the full and partial rankings achieve $\xi^*$ values of 1. Determination of the possible values of $k'$ given $n$ remains an open problem.

6.4 Implied Metrics on Mixed Rankings

The paired comparison projection formulation actually induces two distinct discrete metrics which encompass both full and partial rankings; both of these metrics are conceptually related to the "minimal edge path" formulation of Kendall's $\tau$.

The first such metric corresponds to the minimal half-edge path from one ranking to another on the hypercube, a half-edge corresponding to a shift from $\pm 1$ to 0 or from 0 to $\pm 1$ (a paired comparison response shifting from a definite preference to an indifference or vice versa). This metric has a ready interpretation in terms of minimal preference shifts, and is trivial to compute - on the order of $n^2$, corresponding to the number of possible paired comparisons. If attention is restricted to the full rankings, this metric corresponds to $2 \times \tau$. This metric shall be denoted $\tau_c$.

Two metrics (denoted $\tau_a$ and $\tau_b$) on partial rankings can be derived from suggestions of Kendall ([44], [49]) for assessing the correlation between two partial rankings.

Borrowing Kendall's notation for a moment, assume that two rankings, $X$ and $Y$, are given. Let $a_{ij}$ denote the score associated with items $i$ and $j$ in ranking $X$; $a_{ij} = 1$ if item $i$ is preferred to item $j$, $a_{ij} = -1$ if item $j$ is preferred to item $i$, and $a_{ij} = 0$ if the two items are tied. Similarly, let $b_{ij}$ denote the score associated with items $i$ and $j$ in ranking $Y$. Then

$$\tau_c(X, Y) = \sum_{i < j} |a_{ij} - b_{ij}|.$$ 

Two other extensions of Kendall's $\tau$ (denoted $\tau_a$ and $\tau_b$) to partial rankings can be derived from suggestions of Kendall ([44], [49]) for assessing the correlation between
two partial rankings. These are given by

\[ \tau_a(X, Y) = \frac{\binom{n}{2}}{2} - \frac{\sum_{i<j} a_{ij} b_{ij}}{2}, \]

and

\[ \tau_b(X, Y) = \frac{\sqrt{\left(\sum_{i<j} a_{ij}^2\right)\left(\sum_{i<j} b_{ij}^2\right)}}{2} - \frac{\sum_{i<j} a_{ij} b_{ij}}{2}. \]

(Strictly speaking, \( \tau_a \) is not a metric because \( \tau_a(X, X) \) can exceed zero.) As mentioned, both of these were initially defined as correlation coefficients,

\[ \rho_a = \frac{\sum_{i<j} a_{ij} b_{ij}}{\binom{n}{2}} \]

and

\[ \rho_b = \frac{\sum_{i<j} a_{ij} b_{ij}}{\sqrt{\left(\sum_{i<j} a_{ij}^2\right)\left(\sum_{i<j} b_{ij}^2\right)}} \]

respectively. Both of these arise quite naturally as generalizations of the standard correlation coefficient between two random variables, and both reduce to the standard Kendall's \( \tau \) when the rankings are full. Nonetheless, they both suffer from flaws: \( \tau_b \) fails to give comparable distances when several different full and partial rankings are considered, and \( \tau_a \) misattributes distances when a pair of items is tied in one ranking, and not tied in the other.

Consider the case of the two partial rankings \( X = (1, 1, 3, 3) \) and \( Y = (2, 2, 2, 4) \). The associated scores are given below:

<table>
<thead>
<tr>
<th>((i,j))</th>
<th>((1,2))</th>
<th>((1,3))</th>
<th>((2,3))</th>
<th>((2,4))</th>
<th>((3,4))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X) (a_{ij})</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(X) (b_{ij})</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(Y) (b_{ij})</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Here, \( \sum_{i<j} a_{ij} b_{ij} = 2 \), \( \sqrt{\left(\sum_{i<j} a_{ij}^2\right)\left(\sum_{i<j} b_{ij}^2\right)} = \sqrt{12} \), and \( \sum_{i<j} |a_{ij} - b_{ij}| = 3 \), so

\[ \tau_a(X, Y) = 2, \quad \tau_b(X, Y) = \sqrt{3} - 1, \quad \text{and} \quad \tau_c(X, Y) = 3. \]

Now, \( \tau_c(X, Y) \) reduces to \( 2\tau(X, Y) \) in the case of full rankings, so \( 2\tau_a(X, Y) = 4 \) and \( 2\tau_b(X, Y) = 2\sqrt{3} - 2 \) should be used for comparison. This example illustrates the problem with \( \tau_b(X, Y) \) (remembering that full rankings give rise to integer values); a range of not easily comparable values can be assumed. To illustrate the problem with \( \tau_a(X, Y) \), let \( Y = (1, 1, 3, 3) \), so that \( b_{ij} = a_{ij} \). In this case, \( 2\tau_a(X, Y) = 2 \), whereas
2τc(X, Y) = τc(X, Y) = 0. It is appealing to be able to argue that the distance between two indistinguishable rankings should be zero.

Just as Kendall’s two correlation coefficients gave rise to metrics, the proposed metric τc can be transformed into a correlation coefficient. The resulting statistic is

\[
\rho_c = 1 - \frac{\sum_{i<j} |a_{ij} - b_{ij}|}{\binom{n}{2}}.
\]

This has the added advantage of providing a clear answer in the extreme case mentioned by Kendall [44] where \(X = Y = (2.5, 2.5, 2.5, 2.5)\). In this case,

\[
\rho_a = 0/\binom{n}{2} = 0, \quad \rho_b = 0/\sqrt{0 \times 0} = ?,
\]

and

\[
\rho_c = 1 - \frac{0}{\binom{n}{2}} = 1.
\]

It can be argued that \(\rho_b = 1\) here, but the case is indeterminate. In short, τc is a viable measure of the degree of association extant between two partial rankings.

The second such metric suggested by the hypercube projection formulation actually turns out to be Spearman’s Footrule. This metric arises when nodes and edges mapping to the same nodes and edges in the consistent subspace are identified, so that if two distinct nodes \(X\) and \(Y\) map to the same node in the consistent subspace, then the distance between \(X\) and \(Y\) is said to be zero. If the edges of the paired comparison hypercube are projected down to the consistent subspace, the resulting structure corresponds to overlaying a grid atop the permutation polytope (the analogy is imperfect in that some edges map to the interior of the polytope). This new metric corresponds to the minimal number of grid edges that must be traversed get from the projected node \(X\) to the projected node \(Y\). It shall be denoted here as \(ν(X, Y)\).

The distinction between \(τ_c\) and \(ν\) is illustrated in Figure 6.4; minimal paths from one ordering of three items (\((A, B, C)\)) to another (\((C, B, A)\)) are shown. With \(τ_c\) (shown on the left), a minimal path is found on the hypercube, which is then projected down onto the consistent subspace. With \(ν\) (shown on the right), the hypercube is projected onto the consistent subspace, and a minimal path across the resultant grid is found. In the example shown, \(τ_c(X, Y) = 6\) and \(ν(X, Y) = 4\). As \(τ_c\) can be thought of as a minimal edge path across the grid subject to some constraints (two vertices connected on the grid need not be connected on the hypercube), \(ν \leq τ_c\).
Figure 6.4 Two Induced Metrics on the Rankings. (a) $\tau_e$, the minimal edge path across the hypercube, and (b) $v$, the minimal edge path across the grid.

In order to compute $v$ in terms of the minimal number of grid edges, it is useful to shift to a graph-theoretic formulation. Any set of paired comparisons among $n$ items may be illustrated on a completely connected directed graph. The special case corresponding to the ordering $\langle A, B, C \rangle$ is shown in Figure 6.5. The graph is complete in that every vertex is connected to every other vertex, and directed in that each edge has a direction assigned to it. From this, the number of items beaten by item $A$ in direct comparison, $a_A$, is easily read as the number of edges pointing away from $A$, in this case, two. Similarly, $a_B = 1$ and $a_C = 0$. If items $A$ and $B$ are tied, so that the edge $AB$ is undirected, both $A$ and $B$ shall be assigned “half a victory”. In the example given, this would result in $a_A = a_B = 1.5$ and $a_C = 0$. Referring back to Figure 6.3, three vertices of the hypercube were mapped to the center of the hexagonal grid. These vertices were $(-1,1,-1)$, $(0,0,0)$, and $(1,-1,1)$. Their respective graphs are shown in Figure 6.6. In each of the three cases shown, $a_A = a_B = a_C = 1$; this suggests that the coordinate representation to use on the grid is $(a_A,a_B,a_C)$. The values of $(a_A,a_B,a_C)$ for all points on the grid arising from the paired comparisons of three items are shown in Figure 6.7. Each edge shift corresponds to increasing one coordinate by 0.5 and decreasing another coordinate by 0.5. The minimum number of edges needed to join two points $X,Y$ on the grid is given by $v(X,Y) = \sum_{i=1}^{n} |a_i(X) - a_i(Y)|$. Of course, the $a_i$ are directly related to the rankings of $n$ items; given the representation of $X$ in the grid coordinates $a_i$, the ranking corresponding to $X$ is simply $(n - a_1,n - a_2,\ldots,n - a_n)$. Thus, as noted, $v(X,Y) = \sum_{i=1}^{n} |X(i) - Y(i)|$, otherwise known as Spearman’s Footrule.
Figure 6.5  Graph Representation of the Item Ordering \((A, B, C)\).

Figure 6.6  Graphs of Paired Comparison Triples \((AB, AC, BC)\).
Left, \((-1,1,-1)\), Center, \((0,0,0)\), and Right, \((1,-1,1)\).
Figure 6.7 Values of \((a_A, a_B, a_C)\) at corresponding grid points. Corresponding full Orderings are shown on the outer hexagon.
The hypercube projection formulation has exposed a hitherto unnoticed connection between Kendall's $\tau$ and Spearman's Footrule. Further, it provides an interpretation of $v$ as a minimal edge path, where each edge corresponds to a preference shift of a single paired comparison connecting one group of sets of paired comparisons (which are indistinguishable in terms of the amount of ranking information contained) with another group of sets of paired comparisons. Spearman's Footrule also extends easily to deal with partial rankings, using the midrank procedure to deal with ties, so that

$$v((A, B, C)^{-1}, (A, (B, C))^{-1}) = v((1, 2, 3), (1, 2.5, 2.5)) = 0 + 0.5 + 0.5 = 1.$$  

Diaconis and Graham [24] examine the properties of Spearman's Footrule (which they denote $D$) on full rankings and relate it to Kendall's $\tau$ (the minimal number of transpositions of adjacent elements required to transform one arrangement to another) and Cayley's distance (the minimal number of transpositions required to transform one arrangement to another, denoted $C$) through the combinatorial identity $\tau + C \leq D \leq 2\tau$. The second of these constraints follows trivially from the geometry, as has been noted. In the notation of this section, this can be restated as $v \leq \tau_e$. This constraint holds for mixtures of full and partial rankings as well. The geometric proof of the first constraint is not obvious, and the content is difficult to generalize for two reasons. First, Cayley's distance is not well defined for partial rankings, let alone mixtures of full and partial rankings. Second, for distances between two full rankings, $\tau_e$ is uniquely defined; this is not necessarily true for distances between two arbitrarily chosen grid points. Nonetheless, two questions are implicitly raised: first, does there exist a natural extension of Cayley's distance to full and partial rankings, and second, is there some way of uniquely defining some restriction of $\tau_e$ along the grid?

To answer the first question, it is useful to reflect on the effects of a single shift, or step, in the standard definition of Cayley's distance. Consider the ordering $(A, B, C)$ corresponding to $(a_A, a_B, a_C) = (2, 1, 0)$. Geometrically, the ordering that results from the transposition of two items, $A_1$ and $A_2$, is found by reflecting across the $a_{A_1} = a_{A_2}$ axis. This is illustrated in Figure 6.8.

It is tempting to attempt a decomposition of Cayley's distance into "half-rotations", much as $\tau_e$ decomposes $\tau$ into "half-transpositions". This attempt fails in that the center point of the hexagon is a half-rotation from all of the six vertices, so that the distance from one vertex of the hexagon to any other vertex must be the same or the triangle inequality will be violated. This distance, however, does not correspond to
Cayley's distance when restricted to the full rankings. Another approach follows from the observation that transposing two items corresponds to interchanging their labels on the appropriate graph. Thus, transposing $A$ and $C$ yields $(a_A, a_B, a_C) = (0, 1, 2)$; the values of $a_A$ and $a_C$ have been interchanged. A generalized Cayley's distance $\text{Cay}$ can be constructed by stating that a single step corresponds to adding $\delta$ to the $a_I$ and subtracting $\delta$ from $a_J$, subject to the constraint that $\delta$ takes on a value such that the ensuing $\tilde{a}$ corresponds to a full or partial ranking.

This add/subtract $\delta$ method corresponds to sliding along a single grid edge; $\text{Cay}$ is the minimal number of such edge slides required to travel from one ranking to another. Seen in the context of the geometric framework, the minimal edge slide formulation obviously satisfies symmetry, positivity and the triangle inequality; in short, it is a metric. The restriction of $\text{Cay}$ to full and partial rankings, however, excludes various points on the grid, specifically those which are to some degree linearly inconsistent. If this restriction is not employed, the extension does not reduce to Cayley's distance when restricted to full rankings. Consider the case of the two rankings $(1, 2, 3, 4)$ and $(2, 4, 1, 3)$. If no restriction is employed, then $(1, 2, 3, 4)$ to $(2, 2, 3, 3)$ to $(2, 4, 1, 3)$ is a viable path, and the distance is two; Cayley's distance applied to these full rankings yields three. Unfortunately, a polynomial-time algorithm for the computation of the distance $\text{Cay}$ for arbitrary pairs of different ranking types is not yet available.

This constraint on $\text{Cay}$ suggests a solution to the second problem mentioned above; the suitable restriction of $\tau_c$ so that it is uniquely defined on the grid points. If $\tau_c$ is constrained to act on full and partial rankings only, then the uniqueness problem...
disappears, in that the number of sets of paired comparisons corresponding to a given grid point which are fully consistent \((\zeta^* = 1)\) is either zero or one.

As was mentioned in Chapter 2, Critchlow [17] addressed limitations associated with the tied ranks approach in a general context, citing in particular the problem of dealing with mixtures of partial ranking types (Spearman’s Footrule and Spearman’s \(\rho\) excluded). In this section, that criticism has been answered with respect to Kendall’s \(\tau\) and Cayley’s Distance, though Hamming’s distance and Ulam’s distance remain problematic.
Chapter 7

Future Work

This chapter outlines some directions for further research that have become apparent as work on the thesis has progressed.

7.1 Parallel Coordinates

A problem that frequently occurs in dealing with ranked data is that the natural structure on which to display such data is high-dimensional. Various approaches to this problem have been mentioned in previous chapters (marginal projections, restricting attention to subsets, etc). This section deals with another way of looking at high-dimensional data that is not necessarily restricted to ranked data.

One way of looking at high-dimensional data (proposed by Alfred Inselberg and others [39], [40]) is to display the coordinate axes parallel to one another, rather than orthogonal. The axes are equispaced and a line orthogonal to the axes passes through all of the axis origins. A single high-dimensional point is drawn as a connected series of segments, connecting the value of the first coordinate to the value of the second coordinate, the value of the second to the value of the third, and so forth. The parallel coordinate representation of the ranking (1, 3, 4, 2) is shown in Figure 7.1.

This representation has several advantages, primary among them being the ability to portray more than three dimensions. Unfortunately, it suffers from some drawbacks as well, which become readily apparent as the number of points being examined increases. The area between coordinate axes becomes crowded with ink, and individual points become hard to distinguish. The situation becomes yet worse if the points involved are exact replicates in some coordinates (as is the case with rankings), though this problem can be somewhat alleviated by jittering the coordinate values of each point added. A graphic illustration of the problem is given in Figure 7.2, where the German Political Data are plotted in parallel coordinates.
Figure 7.1 Parallel Coordinate Representation of the Ranking $(1, 3, 4, 2)$.

Figure 7.2 Parallel Coordinate Representation of the German Political Data.
The parallel coordinates procedure suffers from too much edge noise and a lack of ability to deal with overlaps. A two-stage alternative (the first of which is more restricted to ranked data) is hereby proposed.

The first stage involves noting that the parallel coordinates approach suggests a connection with the collection of one dimensional marginal ranking distributions: the marginal projection of item $A$ is plotted on the leftmost axis, the marginal projection of item $B$ on the next axis, etc. Thus, to deal with the overlapping points, remove the connecting segments and place appropriately sized circles along the coordinate axes. This is illustrated in Figure 7.3. This picture provides a quick graphical representation of the “first order” data present in a collection of rankings: the number of times item $i$ is ranked in position $j$. Unfortunately, the ability to detect individual points (and hence correlations of various degrees) has been lost.

The second stage proceeds from the opposite viewpoint: information about the collection as a whole is temporarily sacrificed in the interests of exploring individual points. The parallel coordinate representation is very powerful when only a few multivariate points are being considered. So, pick one of the points (rankings) at random, and plot it in parallel coordinates. Then, pick another point (with replacement), and “flow” from the parallel coordinate representation of the first point to the next by linearly interpolating in some small number of stages. Pick another point and repeat. This procedure is illustrated in Figure 7.4, flowing from the ranking $(1,3,4,2)$ to $(4,1,2,3)$. This procedure is “dynamically stochastic” in that it makes use of animation in a somewhat random way. This can result in several “visible bounds” on the rankings of certain items, in that if item $A$ is almost always either first or second, it will fluctuate only within that bounded range. This procedure allows for the visual perception of correlation. Finally, it should be noted that the problem of large fluctuations caused by outliers can be dealt with by replacing the single rankings with moving averages of three or four.

7.2 Wavelets

The study of wavelets, or multiresolution analysis, has suggested new attacks on the problem of isolating the informative component arising from a noisy signal. This is generally accomplished by means of division and thresholding - the observed values are divided into halves and the first half is then used to define the signal. The values associated with the first half are recorded. The second half is then compared with the
Figure 7.3 Parallel Marginal Projections.

Figure 7.4 Parallel Flowing from (1, 3, 4, 2) to (4, 1, 2, 3).
prediction arising from the first half, and the deviations of the values in the second half from the predicted values are recorded. This completes the division into signal and deviations. The values of the deviations are then examined, centered about their mean value, and the variance associated with the collection of deviations is computed. Those deviations which lie less than some threshold distance from the mean (typically two or three standard deviations) are then “zeroed” (set equal to the mean value), so that only the large deviations remain. The signal and the thresholded deviations are then recombined.

Using this approach to analyze signals on the hypercube (such as arises from either paired comparisons or the duals) suggests new attacks on the decomposition of structure in rankings.

7.3 Experimental Design

The hypercube structure arising as the tensor product of paired comparison segments is similar in nature to the full factorial design for determining the main and interaction effects arising from \( n \) distinct treatments. The identification of certain vertices to reduce the complexity of the design is also practiced, although most commonly with respect to high-order interaction terms. Whether the type of reduction equivalent to the removal of inconsistent triples yields an efficient design, or whether the identifications practiced in design yield interesting structures on which to consider rankings, should be explored.

7.4 Spectral Decomposition via Group Theory

This section expands somewhat on the use of finite Fourier decomposition over groups to examine rankings suggested by Diaconis [22].

In standard Fourier Analysis, a signal is decomposed into a sum of orthogonal sine and cosine terms, whereupon attention tends to focus on those elements which have the largest coefficients. These elements serve as a good approximation to the structure in the signal as a whole.

Each full ranking of \( n \) items may be viewed as an element of the permutation group \( S_n \). Each element of \( S_n \), in turn, may be viewed as an \( n \)-by-\( n \) matrix of zeros and ones, and these matrices define the multiplication structure of the group. Lower dimensional representations of these elements, which adhere to the same multiplication structure
(A * B = C in one representation implies A' * B' = C' in the other), can then be found and used to "decompose" the initial rankings.

Consider the case of the full rankings of three items. Conforming to standard group-theoretic notation, we let e denote the identity (1, 2, 3). Letting a denote (1, 3, 2), b (2, 1, 3), c (2, 3, 1), d (3, 1, 2) and f (3, 2, 1), with corresponding permutation matrices

\[ e = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \]

and so on, the multiplication table thus produced is

\[
\begin{array}{c|ccccccc}
* & e & a & b & c & d & f \\
\hline
e & e & a & b & c & d & f \\
\hline
a & a & e & d & f & b & c \\
b & b & c & e & a & f & d \\
c & c & b & f & d & e & a \\
d & d & f & a & e & c & b \\
f & f & d & c & b & a & e \\
\end{array}
\]

Noting that all of the permutation matrices leave the vector of all ones intact, we can redefine the unit vectors ([22], p.142) as \( w_1 = \frac{1}{\sqrt{2}}(e_1 - e_2) \), \( w_2 = \frac{1}{\sqrt{6}}(e_1 + e_2 - 2e_3) \) and \( w_3 = \frac{1}{\sqrt{3}}(e_1 + e_2 + e_3) \), yielding a set of six two by two matrices (we discount the constant last rows and columns). This we will call a two-dimensional representation. There are also two one-dimensional representations, the trivial representation which identifies each item to 1, and the alternating representation which assigns a 1 or -1 as the permutation is an even or odd number of adjacent transpositions away from
the identity, respectively. Thus a set of representations is

\[ e \quad a \quad b \]
\[ 1 \quad 1 \quad 1 \]
\[ 1 \quad -1 \quad -1 \]
\[ \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \quad \frac{1}{2} \left( \begin{array}{cc} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{array} \right) \quad \left( \begin{array}{cc} -1 & 0 \\ 0 & 1 \end{array} \right) \]

\[ c \quad d \quad f \]
\[ 1 \quad 1 \quad 1 \]
\[ 1 \quad 1 \quad -1 \]
\[ \frac{1}{2} \left( \begin{array}{cc} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{array} \right) \quad \frac{1}{2} \left( \begin{array}{cc} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{array} \right) \quad \frac{1}{2} \left( \begin{array}{cc} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{array} \right) \]

Note that these lower-dimensional representations do indeed conform to the same multiplication structure. These representations are said to be irreducible (unlike the original 3-dimensional one) in that there are no nontrivial subspaces left intact under the multiplication operation (in particular, the vector of all ones is not mapped to itself).

Using the representations, the structure on the ranks can be decomposed into orthogonal components. Let \( v \) be the vector of the number of judges arranging the items according to ranking \( i \). Let \( o_1 \) be the vector of all ones, \( o_2 \) be the vector of 1's and -1's as per the alternating representation, let \( o_3 \) be the vector arrived at by taking \( \sqrt{2} \) times the upper left entry of each of the two by two matrices, and so on (the \( \sqrt{2} \) is a scaling factor chosen so that all of the \( o_i \) will be of the same magnitude). Then

\[ v = \frac{1}{6} \sum_{i=1}^{n} (o_i \cdot v) o_i. \]

The \( o_i \)'s form an orthogonal basis for the rankings.

In this manner, orthogonal decompositions can, theoretically, be derived for the rankings of \( n \) items. Diaconis [23] also outlines how such decompositions can be applied to partial rankings. Note, though, that this decomposition is not unique; it depends upon the choice of basis vectors for the reduction of the matrices. The basis vectors used here correspond to those found in a Helmert transformation. Diaconis avoids this problem by breaking the data down into various layers of ordered effects using "a device of Mallows": \( 0^{th} \) order, the average number of votes per ranking; \( 1^{st} \)
order, the number of judges ranking candidate \( i \) in position \( j \) (minus the average), \( 2^{nd} \) order unordered, the number of judges ranking candidates \( i \) and \( j \) in positions \( k \) and \( l \) or \( l \) and \( k \), \( 2^{nd} \) order ordered, and so on. Unfortunately, the number of such contrasts grows exceedingly fast (outpacing \( n! \)), so this may quickly become impractical. Of course, if we restrict our attention to the \( 0^{th} \) through \( 2^{nd} \) order effects, the number of contrasts “only” grows as \( n^4 \).

7.5 Spectral Decomposition via Hypercubes

Another approach to the problem of spectral decomposition is suggested by the paired comparison hypercube formulation. One of the great benefits of the hypercube formulation is that it lends itself readily to orthogonal decomposition. Such decomposition is familiar to statisticians as the decomposition of a factorial experiment into main effects and the various levels of interaction effects. The orthogonal components used in the decomposition of a cube are illustrated in Figure 7.5. In the process of projecting a paired comparison hypercube down onto the consistent subspace, however, the orthogonal decomposition of the hypercube induces an orthogonal decomposition on the rankings, both full and partial. The orthogonal components induced on the full rankings by the projection of the cube onto the consistent subspace are shown in Figure 7.6.

This particular decomposition suffers from the fact that it introduces mass at “phantom locations”, and thus introduces a philosophical quandary. Nonetheless, it is worth examining. A final problem that arises is that \( 2^{(r)} \) grows very quickly, so that examining all such paired comparisons rapidly becomes infeasible. A useful reduction, then, is to consider the various order effects together - mean, combined main effects, combined second order effects, etc.
Figure 7.5  Decomposition of a Cube into the (a) grand mean, (b) AB main effect, (c) AC main effect, (d) BC main effect, (e) AB/AC interaction, 
(f) AB/BC interaction, (g) AC/BC interaction, and (h) AB/AC/BC interaction. Dark dots are positive, light dots are negative. Inconsistent points are circled in (a).
Figure 7.6  Induced Decomposition of Rankings into (a) grand mean, (b) AB main effect, (c) AC main effect, (d) BC main effect, (e) AB/AC interaction, (f) AB/BC interaction, (g) AC/BC interaction, and (h) AB/AC/BC interaction. Dark dots are positive, light dots are negative. Dashed lines have been added to emphasize structure. Orderings are shown in (a).
Appendix A

Data Sets

A.1 Rice Voting

This data set comprises the ballots cast in a preferential election to choose a faculty member to serve on the Rice Presidential Search Committee. It is published for the first time in this work.

Table A.1 Rice Ballots; 300 Respondents

<table>
<thead>
<tr>
<th>Ordering</th>
<th>Ranking</th>
<th>Freq.</th>
<th>Ordering</th>
<th>Ranking</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A(BCDE)</td>
<td>12222</td>
<td>11</td>
<td>C(ABDE)</td>
<td>22122</td>
<td>10</td>
</tr>
<tr>
<td>B(ACDE)</td>
<td>21222</td>
<td>4</td>
<td>E(ABCD)</td>
<td>22221</td>
<td>1</td>
</tr>
<tr>
<td>AB(CDE)</td>
<td>12333</td>
<td>4</td>
<td>CA(BDE)</td>
<td>23133</td>
<td>3</td>
</tr>
<tr>
<td>AC(BDE)</td>
<td>13233</td>
<td>6</td>
<td>CB(ADE)</td>
<td>32133</td>
<td>1</td>
</tr>
<tr>
<td>AE(BCD)</td>
<td>13332</td>
<td>3</td>
<td>DA(BCE)</td>
<td>23313</td>
<td>1</td>
</tr>
<tr>
<td>BA(CDE)</td>
<td>21333</td>
<td>4</td>
<td>DB(ACE)</td>
<td>32313</td>
<td>1</td>
</tr>
<tr>
<td>BC(ADE)</td>
<td>31233</td>
<td>5</td>
<td>DE(ABC)</td>
<td>33312</td>
<td>2</td>
</tr>
<tr>
<td>BD(ACE)</td>
<td>31323</td>
<td>2</td>
<td>EB(ACD)</td>
<td>32331</td>
<td>1</td>
</tr>
<tr>
<td>BE(ACD)</td>
<td>31332</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC(DE)</td>
<td>12444</td>
<td>5</td>
<td>BCE(AD)</td>
<td>41243</td>
<td>1</td>
</tr>
<tr>
<td>ABE(CD)</td>
<td>12443</td>
<td>1</td>
<td>CAB(DE)</td>
<td>23144</td>
<td>3</td>
</tr>
<tr>
<td>ACB(DE)</td>
<td>13244</td>
<td>4</td>
<td>CBD(AE)</td>
<td>42134</td>
<td>1</td>
</tr>
<tr>
<td>BAC(DE)</td>
<td>21344</td>
<td>6</td>
<td>DBA(CE)</td>
<td>32414</td>
<td>1</td>
</tr>
<tr>
<td>BAD(CE)</td>
<td>21434</td>
<td>1</td>
<td>DBC(AE)</td>
<td>42314</td>
<td>1</td>
</tr>
<tr>
<td>BAE(CD)</td>
<td>21443</td>
<td>1</td>
<td>DBE(AC)</td>
<td>42413</td>
<td>1</td>
</tr>
<tr>
<td>BCA(DE)</td>
<td>31244</td>
<td>2</td>
<td>EBA(CD)</td>
<td>32441</td>
<td>1</td>
</tr>
<tr>
<td>Ordering</td>
<td>Ranking</td>
<td>Freq.</td>
<td>Ordering</td>
<td>Ranking</td>
<td>Freq.</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>-------</td>
<td>----------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>ABCDE</td>
<td>12345</td>
<td>6</td>
<td>BEDCA</td>
<td>51432</td>
<td>2</td>
</tr>
<tr>
<td>ABCED</td>
<td>12354</td>
<td>8</td>
<td>CABDE</td>
<td>23145</td>
<td>5</td>
</tr>
<tr>
<td>ABDCE</td>
<td>12435</td>
<td>2</td>
<td>CABED</td>
<td>23154</td>
<td>18</td>
</tr>
<tr>
<td>ABDEC</td>
<td>12534</td>
<td>1</td>
<td>CADBE</td>
<td>24135</td>
<td>1</td>
</tr>
<tr>
<td>ABEDC</td>
<td>12534</td>
<td>6</td>
<td>CAEBD</td>
<td>24153</td>
<td>5</td>
</tr>
<tr>
<td>ACBDE</td>
<td>13245</td>
<td>10</td>
<td>CBADE</td>
<td>32145</td>
<td>9</td>
</tr>
<tr>
<td>ACBED</td>
<td>13254</td>
<td>20</td>
<td>CBAED</td>
<td>32154</td>
<td>4</td>
</tr>
<tr>
<td>ACDBE</td>
<td>14235</td>
<td>3</td>
<td>CBDEA</td>
<td>52134</td>
<td>1</td>
</tr>
<tr>
<td>ACDEB</td>
<td>15234</td>
<td>2</td>
<td>CBEAD</td>
<td>42153</td>
<td>1</td>
</tr>
<tr>
<td>ACEBD</td>
<td>14253</td>
<td>1</td>
<td>CDABE</td>
<td>34125</td>
<td>1</td>
</tr>
<tr>
<td>ACEDB</td>
<td>15243</td>
<td>2</td>
<td>CDABAE</td>
<td>43125</td>
<td>2</td>
</tr>
<tr>
<td>ADCBE</td>
<td>14325</td>
<td>1</td>
<td>CDBEA</td>
<td>53124</td>
<td>1</td>
</tr>
<tr>
<td>ADCEB</td>
<td>15324</td>
<td>1</td>
<td>CDEAB</td>
<td>45123</td>
<td>1</td>
</tr>
<tr>
<td>AEBCD</td>
<td>13452</td>
<td>2</td>
<td>CDEBA</td>
<td>54123</td>
<td>1</td>
</tr>
<tr>
<td>AEBDC</td>
<td>13542</td>
<td>1</td>
<td>CEADB</td>
<td>35142</td>
<td>1</td>
</tr>
<tr>
<td>AECDB</td>
<td>15342</td>
<td>1</td>
<td>CEBAD</td>
<td>43152</td>
<td>1</td>
</tr>
<tr>
<td>AEDCB</td>
<td>15432</td>
<td>1</td>
<td>CEBDA</td>
<td>53142</td>
<td>1</td>
</tr>
<tr>
<td>BACDE</td>
<td>21365</td>
<td>7</td>
<td>CEDAB</td>
<td>45132</td>
<td>1</td>
</tr>
<tr>
<td>BACED</td>
<td>21354</td>
<td>19</td>
<td>CEDBA</td>
<td>54132</td>
<td>1</td>
</tr>
<tr>
<td>BADCE</td>
<td>21435</td>
<td>1</td>
<td>DBACE</td>
<td>32415</td>
<td>3</td>
</tr>
<tr>
<td>BADEC</td>
<td>21534</td>
<td>4</td>
<td>DBAEC</td>
<td>32514</td>
<td>3</td>
</tr>
<tr>
<td>BAEDC</td>
<td>21453</td>
<td>1</td>
<td>DBCAE</td>
<td>42315</td>
<td>3</td>
</tr>
<tr>
<td>BAEDC</td>
<td>21543</td>
<td>7</td>
<td>DCBEA</td>
<td>53214</td>
<td>1</td>
</tr>
<tr>
<td>BCADE</td>
<td>31245</td>
<td>2</td>
<td>DCEAB</td>
<td>45213</td>
<td>1</td>
</tr>
<tr>
<td>BCAED</td>
<td>31254</td>
<td>11</td>
<td>DEBAC</td>
<td>43512</td>
<td>1</td>
</tr>
<tr>
<td>BCEAD</td>
<td>41253</td>
<td>4</td>
<td>EACBD</td>
<td>24351</td>
<td>1</td>
</tr>
<tr>
<td>BCEAD</td>
<td>41253</td>
<td>4</td>
<td>EACBD</td>
<td>24351</td>
<td>1</td>
</tr>
<tr>
<td>BCEAD</td>
<td>41253</td>
<td>4</td>
<td>EACBD</td>
<td>24351</td>
<td>1</td>
</tr>
<tr>
<td>BDACE</td>
<td>31425</td>
<td>1</td>
<td>EADCB</td>
<td>25431</td>
<td>1</td>
</tr>
<tr>
<td>BDCAE</td>
<td>41325</td>
<td>2</td>
<td>EBDAC</td>
<td>32541</td>
<td>1</td>
</tr>
<tr>
<td>BDEAC</td>
<td>41523</td>
<td>1</td>
<td>EBCAD</td>
<td>42351</td>
<td>1</td>
</tr>
<tr>
<td>BEADC</td>
<td>31542</td>
<td>2</td>
<td>EBDCA</td>
<td>52431</td>
<td>1</td>
</tr>
<tr>
<td>BECAD</td>
<td>41352</td>
<td>1</td>
<td>EDABC</td>
<td>34521</td>
<td>1</td>
</tr>
<tr>
<td>BECDA</td>
<td>51342</td>
<td>1</td>
<td>EDABC</td>
<td>34521</td>
<td>1</td>
</tr>
</tbody>
</table>
A.2 German Political Preferences

This data set comprises the responses of 2262 Germans to a survey about the desirability of four political goals (taken from Duncan and Brody [25]).

Table A.2 Frequencies of Preference Orderings for Four Political Goals; A = Maintaining order in the nation; B = More say in the government; C = Fighting rising prices; D = Protecting freedom of speech; 2262 Respondents.

<table>
<thead>
<tr>
<th>Ordering</th>
<th>Ranking</th>
<th>Freq.</th>
<th>Ordering</th>
<th>Ranking</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCD</td>
<td>1234</td>
<td>137</td>
<td>CABD</td>
<td>2314</td>
<td>330</td>
</tr>
<tr>
<td>ABDC</td>
<td>1243</td>
<td>29</td>
<td>CADB</td>
<td>2413</td>
<td>294</td>
</tr>
<tr>
<td>ACBD</td>
<td>1324</td>
<td>309</td>
<td>CBAD</td>
<td>3214</td>
<td>117</td>
</tr>
<tr>
<td>ACDB</td>
<td>1423</td>
<td>255</td>
<td>CBDA</td>
<td>4213</td>
<td>69</td>
</tr>
<tr>
<td>ADBC</td>
<td>1342</td>
<td>52</td>
<td>CDAB</td>
<td>3412</td>
<td>70</td>
</tr>
<tr>
<td>ADCB</td>
<td>1432</td>
<td>93</td>
<td>CDBA</td>
<td>4312</td>
<td>34</td>
</tr>
<tr>
<td>BACD</td>
<td>2134</td>
<td>48</td>
<td>DABC</td>
<td>2341</td>
<td>21</td>
</tr>
<tr>
<td>BADC</td>
<td>2143</td>
<td>23</td>
<td>DACB</td>
<td>2431</td>
<td>30</td>
</tr>
<tr>
<td>BCAD</td>
<td>3124</td>
<td>61</td>
<td>DBAC</td>
<td>3241</td>
<td>29</td>
</tr>
<tr>
<td>BCDA</td>
<td>4123</td>
<td>55</td>
<td>DBCA</td>
<td>4231</td>
<td>52</td>
</tr>
<tr>
<td>BDAC</td>
<td>3142</td>
<td>33</td>
<td>DCAB</td>
<td>3421</td>
<td>35</td>
</tr>
<tr>
<td>BDCA</td>
<td>4132</td>
<td>59</td>
<td>DCBA</td>
<td>4321</td>
<td>27</td>
</tr>
</tbody>
</table>
A.3 Attitudes Towards Science

This data set comprises the agree/disagree responses of 2982 high school seniors to four statements intended to illustrate attitudes towards scientific research. The respondents were divided equally into high and low IQ groups (taken from Solomon [70]).

Table A.3 Agree/Disagree Responses to Four Questions; A = The development of new ideas is the scientist’s greatest source of satisfaction; B = Scientists and engineers should be eliminated from the military draft; C = The scientist will make his maximum contribution to society when he has freedom to work on problems which interest him; D = The monetary compensation of a Nobel Prize winner in physics should be at least equal to that given popular entertainers; 2982 Respondents.

<table>
<thead>
<tr>
<th>(A, B, C, D)</th>
<th>Low IQ Frequency</th>
<th>High IQ Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1111</td>
<td>62</td>
<td>122</td>
</tr>
<tr>
<td>1110</td>
<td>70</td>
<td>68</td>
</tr>
<tr>
<td>1101</td>
<td>31</td>
<td>33</td>
</tr>
<tr>
<td>1100</td>
<td>41</td>
<td>25</td>
</tr>
<tr>
<td>1011</td>
<td>283</td>
<td>329</td>
</tr>
<tr>
<td>1010</td>
<td>253</td>
<td>247</td>
</tr>
<tr>
<td>1001</td>
<td>200</td>
<td>172</td>
</tr>
<tr>
<td>1000</td>
<td>305</td>
<td>217</td>
</tr>
<tr>
<td>0111</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>0110</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>0101</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>0100</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>0011</td>
<td>31</td>
<td>56</td>
</tr>
<tr>
<td>0010</td>
<td>46</td>
<td>55</td>
</tr>
<tr>
<td>0001</td>
<td>37</td>
<td>64</td>
</tr>
<tr>
<td>0000</td>
<td>82</td>
<td>53</td>
</tr>
<tr>
<td>Totals</td>
<td>1491</td>
<td>1491</td>
</tr>
</tbody>
</table>
Bibliography


[79] Thompson, G.L., Personal communication.


