THE 1987 WALD MEMORIAL LECTURES

A GENERALIZATION OF SPECTRAL ANALYSIS WITH APPLICATION TO RANKED DATA[1]

By Persi Diaconis

Harvard University

An analog of the spectral analysis of time series is developed for data in general spaces. This is applied to data from an election in which 5738 people rank ordered five candidates. Group theoretic considerations offer an analysis of variance like decomposition which seems natural and fruitful. A variety of inferential tools are suggested. The spectral ideas are then extended to general homogeneous spaces such as the sphere.

1. Introduction. Data sometimes come in the form of ranks or preferences: A group of people may be asked to rank order five brands of chocolate chip cookies. Each person tastes the cookies and ranks all five. This results in a ranking $\pi(1)$, $\pi(2)$, $\pi(3)$, $\pi(4)$, $\pi(5)$ with $\pi(i)$ the rank given brand i. The collection of rankings makes up the data set.

Elections are sometimes based on rankings. For example, the American Psychological Association asks its members to rank order five candidates for president. Voting in Cambridge, Massachusetts and in some Australian elections is also based on rankings. A careful analysis of the data from one election is presented in the next section.

Here are some other examples of rank data sets: in testing a random number generator, people consider the relative order of k successive outputs [see Knuth (1981), page 61]. This rapidly leads to a large collection of rankings. Monte Carlo evaluation of rules for entering variables in a regression equation leads to many rankings—the order of entering variables in successive runs.

Most anyone who analyzes such data looks at simple averages such as the proportion of times each item was ranked first (or last) and the average rank for each item. These are first order statistics: They are linear combinations of the number of times item i was ranked in position j.

As will appear below, there are also natural second order statistics based on the number of times items i and i' are ranked in positions j and j'. These come in ordered and unordered modes. For example the number of times items i and i' are ranked either 12 or 21 is an unordered second order statistic. Similarly, there are third and higher order statistics of various types.

A basic tenet of data analysis is this: If you've found some structure, take it out, and look at what's left. Thus to look at second order statistics it is natural

to subtract away the observed first order structure. This leads to a natural decomposition of the original data into orthogonal pieces. The decomposition is somewhat more complicated than standard analysis of variance decompositions because of the dependence inherent in the permutation structure: If item i is ranked first, item i' has to be ranked lower.

Suppose there are n items to be ranked. Let S_n denote the symmetric group on n letters. Data can be regarded as a function f on S_n, with $f(\pi)$ being the number of rankers choosing ranking π. Group theorists have developed a natural decomposition of the space of all functions into orthogonal subspaces which is invariant under relabeling of the underlying items. This yields the decomposition

$$(1.1) \qquad\qquad\qquad f(\pi) = \sum_\rho \hat{f}_\rho(\pi),$$

where ρ indexes the various subspaces and $\hat{f}_\rho$ denotes the projection.

This can be compared with the usual spectral analysis of time series which decomposes a function f on the group of integers mod N into its projections,

$$(1.2) \qquad f(j) = \frac{1}{N}\sum \hat{f}(k)e^{-2\pi ijk/N}, \qquad \hat{f}(k) = \sum f(j)e^{2\pi ijk/N}.$$

Thus f is expressed as a linear combination of simple periodic functions. If a few of the $\hat{f}(k)$ are large and the rest are small, $f(j)$ has a simple description and approximation.

The decomposition (1.1) has a similar interpretation. One difference is that the subspaces for the integers mod N are one-dimensional while the subspaces for the permutation group have higher dimension. The choice of basis leads to interesting problems which are here resolved using a device of Mallows.

The next section presents a data analysis in some detail. The data involve both full and partially ranked data. Here, unordered pair effects are crucial to unraveling what becomes a simple, interpretable structure.

The example introduces the basic invariant subspaces in an instructional way. Section 3 gives formal descriptions of the various types of partial rankings and of the needed representation theory.

Section 4 addresses inferential issues. In the example, the respondents cannot reasonably be thought of as a sample and other considerations must be invoked to assess variability of the basic averages. Of course, sometimes multinomial (or Poisson, or normal) variability is believable and standard theory is also presented and compared.

Section 5 outlines the extension of spectral analysis to data with values in more general spaces. This includes the usual analysis of designed experiments. The extension permits all of the many considerations developed for time series and ANOVA to be brought to bear on ranked data. It also suggests some new analyses of classical designs.

Other approaches. There have been several other approaches to analyzing permutation data. Purely data-analytic methods are outlined by Cohen and Mallows (1980) and Cohen (1982). Scaling techniques are discussed by Carroll

(1980). Methods based on metrics are described by Critchlow (1985), Feigin and Alvo (1986), Fligner and Verducci (1986) and Diaconis (1988). These extend earlier work by Mallows (1957). A host of models have been suggested, the Luce model [equivalently Plackett's (1975) model] being the best known. Several models are discussed here in Section 4. Batsell and Polking (1985) survey applications in marketing. The literature on these topics is large. Diaconis (1988), Chapter 9 contains a review and pointers to further literature.

2. Data analysis.

2A. *First order analysis.* The American Psychological Association (APA) is a large professional organization of academicians, clinicians and all shades in

TABLE 1
American Psychological Association election data

Ranking	No. of votes cast of this type	Ranking	No. of votes cast of this type	Ranking	No. of votes cast of this type	Ranking	No. of votes cast of this type
54321	29	43521	91	32541	41	21543	36
54312	67	43512	84	32514	64	21534	42
54231	37	43251	30	32451	34	21453	24
54213	24	43215	35	32415	75	21435	26
54132	43	43152	38	32154	82	21354	30
54123	28	43125	35	32145	74	21345	40
53421	57	42531	58	31542	30	15432	40
53412	49	42513	66	31524	34	15423	35
53241	22	42351	24	31452	40	15342	36
53214	22	42315	51	31425	42	15324	17
53142	34	42153	52	31254	30	15243	70
53124	26	42135	40	31245	34	15234	50
52431	54	41532	50	25431	35	14532	52
52413	44	41523	45	25413	34	14523	48
52341	26	41352	31	25341	40	14352	51
52314	24	41325	23	25314	21	14325	24
52143	35	41253	22	25143	106	14253	70
52134	50	41235	16	25134	79	14235	45
51432	50	35421	71	24531	63	13542	35
51423	46	35412	61	24513	53	13524	28
51342	25	35241	41	24351	44	13452	37
51324	19	35214	27	24315	28	13425	35
51243	11	35142	45	24153	162	13254	95
51234	29	35124	36	24135	96	13245	102
45321	31	34521	107	23541	45	12543	34
45312	54	34512	133	23514	52	12534	35
45231	34	34251	62	23451	53	12453	29
45213	24	34215	28	23415	52	12435	27
45132	38	34152	87	23154	186	12354	28
45123	30	34125	35	23145	172	12345	30

TABLE 2
Percentage of voters ranking candidate i in position j

Candidate	Rank				
	1	**2**	**3**	**4**	**5**
1	18	26	23	17	15
2	14	19	25	24	18
3	28	17	14	18	23
4	20	17	19	20	23
5	20	21	20	19	20

between. The APA elects a president every year by asking each member to rank order a slate of five candidates. The elections are actively contested, not rituals to justify backroom choices. Coombs, Cohen and Chamberlain (1984) contains further background.

There were about 50,000 APA members in 1980. About 15,000 members voted. Many members cast incomplete ballots, voting for their favorite q of five candidates, $1 \leq q \leq 3$. The 5738 complete ballots are tabulated in Table 1. This will be analyzed first. The incomplete ballots are analyzed at the end of this section.

The columns of Table 1 index candidates. Entries in a column show where that candidate was ranked in the given permutation. Thus 29 members ranked candidate 5 first, candidate 4 second, candidate 3 third, candidate 2 fourth and candidate 1 fifth.

It is possible to learn something by looking at these 120 numbers—some of the counts are much larger than others. For many purposes, simple averages are useful summaries. Table 2 shows the percentage of voters ranking i in position j. Thus, candidate 3 is most popular, being ranked first by 28% of the voters. Candidate 3 also had some "hate vote." Candidate 1 is strongest in the second position, has no hate vote and has a lower average rank than candidate 3. The voters seem indifferent on candidate 5.

The APA, along with many other organizations, chooses a winner by the Hare system (also known as proportional voting). This works as follows: If one of the five candidates is ranked first by more than half the voters, they win. If not, then the candidate with the fewest first place votes is eliminated, each of the remaining candidates is reranked in relative order and the method is applied inductively. Candidate 1 is the eventual winner here.

Arrow's impossibility theorem implies that there are no completely unobjectionable procedures for combining preferences into a final choice. Fishburn (1973) reviews the pitfalls and benefits of the Hare system.

Before delving further into the analysis, let's pause and ask where we are headed. The data in Table 1, along with the incomplete data analyzed below, suggest several natural questions:

1. Does the first order analysis, reported above, capture the structure of the data in Table 1 or is there further simple structure?

TABLE 3
Decomposition of the regular representation

M =	V_1	$\oplus$	V_2	$\oplus$	V_3	$\oplus$	V_4	$\oplus$	V_5	$\oplus$	V_6	$\oplus$	V_7
Dim 120	1		16		25		36		25		16		1
SS/120	2286		298		459		78		27		7		0

2. How does the Hare system work for data with the observed structure?

3. Is the partially ranked data similar to the marginal of the fully ranked data?

4. How should the partially ranked data be combined with the fully ranked data to elect a winner?

2B. *Higher order analysis.* The data vector can be regarded as the function $f(\pi)$ that assigns to π the number of people choosing ranking π. Thus $f\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 \end{pmatrix} = 29$. Let M be the space of all real valued functions on the symmetric group S_5. This is a vector space under addition of functions.

The usual inner product on M is defined by

$$(2.1) \qquad \langle f_1 | f_2 \rangle = \sum_\pi f_1(\pi) f_2(\pi).$$

The space M decomposes uniquely into the direct sum of seven subspaces. These are shown, with their dimensions, in Table 3. A more formal description of the decomposition is given in Section 3. Hopefully, the following informal description will serve for now.

The space V_1 is the set of constant functions. This has one dimension. The space V_2 will be called the space of first order functions. To explain, consider the function $\pi \to \delta_{i\pi(j)}$ which is 1 if $\pi(j) = i$ and 0 otherwise. This only depends on π through the value of one coordinate. A general first order function has the form

$$\sum_{i,j} a_{ij} \delta_{i\pi(j)}.$$

To get a direct sum decomposition, the a_{ij} must satisfy $\sum a_{ij} = 0$. The space V_2 has dimension 16.

There are two types of second order functions: unordered and ordered. A typical unordered element is $\delta_{\{1,2\},\{\pi(1),\pi(2)\}}$ which is 1 if the unordered set $\{\pi(1), \pi(2)\} = \{1, 2\}$ and 0 otherwise. The general, unordered, second order functions V_3 have the form

$$\sum_{i,i',j,j'} a_{ii',jj'} \delta_{\{i,i'\},\{\pi(j),\pi(j')\}}$$

with $a_{ii',jj'}$ chosen so that V_3 is orthogonal to $V_1 \oplus V_2$. The *ordered* second order functions V_4 are made up of elements like

$$\delta_{(i,i')(\pi(j),\pi(j'))} \quad \text{where order matters.}$$

They are defined to be orthogonal to $V_1 \oplus V_2 \oplus V_3$. Similarly there are third and higher order subspaces. The final space V_7 is spanned by the function sgn π which is ± 1 as π can be written with an even or odd number of transpositions.

The subspaces V_i are defined through invariance considerations. In this example, the values of the rankings have a natural order, but the labels assigned to candidates are arbitrary. A ranking is really a mapping π from the set of candidate names to the set of ranking values. The symmetric group S_5 naturally permutes the set of candidate names. It thus acts on functions f by σf evaluated at π being $f(\pi\sigma)$.

A subspace $V \subset M$ is *invariant* under S_5 if $f \in V$ implies $\sigma f \in V$. It seems natural to insist that basic descriptive units such as "first order functions" be invariant under irrelevant relabeling. It also seems natural that basic descriptive units be subspaces: If f and g are first order, functions like $f/10$ and $f + g$ should be first order too. This suggests invariant subspaces of M as useful objects.

The invariance discussed above has the group acting on the right, permuting candidate names. The symmetric group S_5 also acts on the left, permuting ranking labels. There is a unique decomposition—the isotypic decomposition— into subspaces invariant under both relabelings. This is the decomposition of Table 3. More refined decompositions require choosing a basis. This is discussed further below.

Consider the data vector f as a function in M. It has a decomposition into its projections on the isotypic subspaces V_i. The squared length of each piece is shown in the third row of Table 3. As usual, the largest contribution is from the projection onto the constants. The projection onto V_2 is sizable, but not as large as the projection onto V_3. The projections onto higher order subspaces seem small by comparison.

It is customary in comparing sums of squares to divide by the dimension of the subspace. This makes sense if it is thought that the projection is reasonably spread out between a natural system of coordinate vectors so that the sum of squares is a measure of noise. If, as in the present example, the projections are likely to be quite structured, lying close to a few interpretable vectors, dividing by dimension is likely to be deceptive.

The relatively large sum of squares for V_3 also obtains for the partially ranked data (see Tables 8 and 9). This suggests a closer look at the projection of f onto V_3.

2C. *Second order analysis.* To aid interpretation, it is useful to present the first order summary of Table 2 in a form similar to the summaries of this section. This is Table 4 below which has entry i, j, the number of voters ranking candidate i in position j minus the sample size over 5, so rows (and columns) sum to zero. The entries have been rounded to integers. Table 4 is based on counts, not proportions. It is an affine function of Table 2. Thus, the same structure is apparent. The largest number, 461, indicates candidate 3 received most first place votes. The second largest number, 371, shows that candidate 1 received most second place votes, etc.

TABLE 4
First order effects

Candidate	Rank				
	1	**2**	**3**	**4**	**5**
1	−94	371	165	−145	−296
2	−372	−70	267	268	−92
3	461	−187	−354	−97	178
4	24	−175	−58	16	193
5	−18	62	−19	−41	17

The space V_3 of unordered pair effects is a 25-dimensional space of functions on S_5. There does not seem to be a natural choice of basis. Mallows has suggested the following remedy. The easily interpretable second order functions are of the form

$$\delta_{\{i, i'\}\{\pi(j), \pi(j')\}}.$$

That is, 1 if candidates j and j' are ranked in position $\{i, i'\}$ in either order and 0 otherwise. This δ is thought of as a function of π for fixed $\{i, i'\}, \{j, j'\}$. Each pair $\{i, i'\}, \{j, j'\}$ can be chosen in 10 ways, so there are 100 easily interpreted functions. Let f be the data vector and $\bar{f}$ its projection onto V_3. Compute the inner product of the easily interpretable functions with $\bar{f}$ and look at these numbers. Geometrically, the interpretable functions project to 100 points in a 25-dimensional space. The data vector projects to another point. If the data vector lies close to a few interpretable functions, we have a simple description. The inner products are shown in Table 5. For example the $\{1, 2\}, \{1, 2\}$ entry, -137, is the inner product of $\delta_{\{1,2\}, \{\pi(1), \pi(2)\}}$ (as a function of π) with $\bar{f}(\pi)$, the projection of f onto V_3. The inner product is defined by (2.1).

TABLE 5
Second order, unordered effects

Candidate	Rank									
	1, 2	**1, 3**	**1, 4**	**1, 5**	**2, 3**	**2, 4**	**2, 5**	**3, 4**	**3, 5**	**4, 5**
1, 2	−137	−20	18	140	111	22	4	6	−97	−46
1, 3	476	−88	−179	−209	−147	−169	−160	107	128	241
1, 4	−189	51	113	24	−9	98	99	−65	23	−146
1, 5	−150	57	47	45	43	49	56	−48	−53	−48
2, 3	−42	84	19	−61	30	−16	82	−76	−39	72
2, 4	157	−20	−43	−25	−93	−76	−56	8	38	112
2, 5	22	−44	7	15	−117	69	25	62	99	−138
3, 4	−265	−7	72	199	39	140	85	19	−52	−233
3, 5	−169	10	88	70	78	44	47	−51	−36	−80
4, 5	296	−24	−142	−130	−5	−163	−128	38	−9	267

The displayed second order effects have a simple structure. To understand it, consider the largest number in the table, 476, in row $\{1, 3\}$, column $\{1, 2\}$. This means there is a huge effect for ranking candidates 1 and 3 in positions 1 and 2. Similarly the last entry in row $\{1, 3\}$ shows a fair sized "hate vote."

The last row, $\{4, 5\}$, shows a big effect for the pair of candidates 4 and 5, at both ends. Now the table's structure falls into place. There are two groups of candidates, $\{1, 3\}$ and $\{4, 5\}$. The voters line up behind one group or the other. For pairs of opposite groups, like $1, 4$ or $1, 5$ or $3, 4$ or $3, 5$, there is the opposite effect; every few people like both or hate both, so the row entry begins and ends $-, -$.

This pattern makes perfect sense in the APA election. The APA divides into academicians and clinicians who are on uneasy terms. Voters seem to choose one type or the other, and then choose within, but the group effect predominates—recall that these second order effects are adjusted for individual popularity. In a similar scenario with the APA replaced by the IMS, one can imagine voters ranking the five candidates on some rough scale from statistician to probabilist and then "unfolding" a ranking about one end or the other.

Candidate 2 seems to fall in the middle, perhaps closer to 4 and 5. Further comments and findings are in the last part of this section. Table 5 has been motivated data analytically. It also has a natural group theoretic motivation suggested by James. In the present context, a natural way to investigate unordered second order structure would be to form a matrix indexed by unordered pairs, with the entry $\{i, i'\}, \{j, j'\}$ the number of people ranking items $\{j, j'\}$ in position $\{i, i'\}$. Now the symmetric group $S_n \times S_n$ acts on the rows and columns of this matrix. One can project onto the invariant subspaces. It can be shown, using the notation of Section 3, that Table 5 is precisely the projection of this matrix onto $S^{n-2, 2} \otimes S^{n-2, 2}$.

2D. *Analysis of partially ranked data.* A similar analysis is available for the voters who ranked q out of 5. These data are presented in Table 6. Again the columns index candidates, and a j under candidate i indicates that candidate i was ranked jth. Zeroes or blanks indicate unranked candidates. For example, 1022 members ranked candidate 5 first and left the others unranked.

The subspace decompositions, dimensions, sums of squares and first and second ordered projections are set out in Tables 7, 8 and 9, respectively. For example, from Table 7, there are 5141 members who only ranked $q = 1$ of the 5 candidates. The data are thus a function $f(i)$—the number of people ranking candidate i first. The space of all such functions is denoted $M^{4, 1}$. There are two invariant subspaces in the isotypic decomposition. These are the constant functions and the functions summing to zero. These are denoted S^5 and $S^{4, 1}$ to conform to later notation. In each case, the sum of squares has been divided by the dimension of the space of functions.

REMARKS. For $q = 1$, the projection onto $S^{4, 1}$ merely amounts to subtracting the number of rankers divided by 5 from the original data vector. Thus candidate 3 is most popular and the rest of the pattern is the same as for the full data (compare with the first column of Table 4).

TABLE 6
American Psychological Association election data

q = 1		q = 2		q = 3		q = 3	
Partial ranking	No. of votes cast of this type	Partial ranking	No. of votes cast of this type	Partial ranking	No. of votes cast of this type	Partial ranking	No. of votes cast of this type
1	1022	21	143	30021	75	20130	39
10	1145	12	196	30201	32	23100	83
100	1198	201	64	32001	41	20103	74
1000	881	210	48	20031	62	132	19
10000	895	102	93	20301	37	123	15
		120	56	23001	35	2103	16
		2001	70	3201	15	1302	15
		2010	114	2301	14	1032	45
		2100	89	3021	59	1320	17
		1002	80	2031	50	1203	8
		1020	87	321	20	31002	38
		1200	51	231	17	31020	45
		20001	117	30012	90	31200	32
		20010	104	30210	13	21003	17
		20100	547	32010	51	21030	31
		21000	72	20013	46	1023	55
		10002	72	20310	15	1230	9
		10020	74	23010	28	21300	31
		10200	302	3012	62	10032	35
		12000	83	3210	18	10203	49
				2310	21	10302	41
				2013	54	10320	21
				312	46	13002	31
				213	16	13020	22
				2130	17	13200	79
				3120	26	10023	44
				3102	16	10230	30
				30102	47	12003	26
				32100	57	12030	19
				30120	15	12300	27

TABLE 7
Spectral analysis for q = 1, n = 5141

$$M^{4,1} \quad = \quad S^5 \quad \oplus \quad S^{4,1}$$

Dim	5	1	4
SS/5		1,057,195	16,384

Candidate	Projection
1	−133
2	−147
3	170
4	117
5	−6

TABLE 8

Spectral analysis for q = 2, n = 2462

$$M^{3,1,1} = S^5 \oplus 2S^{4,1} \oplus S^{3,2} \oplus S^{3,1,1}$$

Dim	20	1	2·4	5	6
SS/20		15,154	4,268	7,781	652

First order analysis			Second order analysis	
Candidate	Projection 1	Projection 2	Candidate	Projection
1	38	347	12	−107
2	−202	−136	13	385
3	292	−27	14	−142
4	−30	−32	15	−136
5	−98	−51	23	−81
			24	122
			25	66
			34	−176
			35	−127
			45	197

TABLE 9

Spectral analysis for q = 3, n = 2108

$$M^{2,1,1,1} = S^5 \oplus 3S^{4,1} \oplus 3S^{3,2} \oplus 3S^{3,1,1} \oplus 2S^{2,2,1} \oplus S^{2,1,1,1}$$

Dim	60	1	3·4	3·5	3·6	2·5	4
SS/60		1,234	123	243	20	8	2

First order analysis				Second order analysis			
Candidate	1	2	3	Candidate	12	13	14
1	2	76	114	1, 2	−50	6	12
2	−78	−28	52	1, 3	150	−3	−41
3	2	−103	−116	1, 4	−71	−8	11
4	38	−7	−48	1, 5	−28	5	16
5	35	63	−1	2, 3	−2	24	28
				2, 4	57	−5	−7
				2, 5	−5	−24	−34
				3, 4	−84	−12	−4
				3, 5	−63	−8	17
				4, 5	97	26	0

For $q = 2$ there is a slight difference in first order statistics. The second order statistics correspond to the pattern found for the complete data. As a notational point, the first order subspace is denoted $2S^{4,1}$. It is an eight-dimensional space consisting of two four-dimensional subspaces, one for most preferred and one for second place.

For $q = 3$, the first order analysis seems different again, candidates 4 and 5 being clearly preferred to 1 and 3. The first column of the second order analysis matches up with the previously found pattern—a strong second order effect with $(1, 3)$ dominating $(4, 5)$.

Some differences between partially ranked ballots and fully ranked ballots are to be expected. People who only rank a few candidates probably choose favorites. People who rank everyone clearly also vote against specific candidates.

As a final remark, note that the decomposition in Table 3 (fully ranked data) appears in present notation as

$$M^{11111} = S^5 \oplus 4S^{4,1} \oplus 5S^{3,2} \oplus 6S^{3,1,1} \oplus 5S^{2,2,1} \oplus 4S^{2,1,1} \oplus S^{11111}.$$

2E. *Summary*.   The data analysis reported above shows a consistent simple structure: There is a strong effect for choosing candidates $\{1,3\}$ or $\{4,5\}$, with candidate 2 in the middle. This effect is roughly of the same size and direction (ratio of sums of squares) for $q = 2, 3$ and 4.

There is some difference between the first order effects: $q = 1$ is close to $q = 5$; $q = 2,3$ seem different. Several ad hoc significance tests not reported here reject the hypothesis that the marginal first order effect of fully ranked ballots match the first order margins for $q = 1, 2$ or 3.

The next stage of analysis is to look at other years and see if the strong two group effect is a consistent feature of APA elections. If so, this allows a solid base for investigating the Hare system and for building more detailed models. The data suggest models of the following sort: Arrange the five candidates as points on a line in some order. Regard each voter as a point on the line who orders the candidates by their distance. Carroll (1980) discusses such models. Alternatively, the data may be modeled as a mixture of two Mallows models centered at a central permutation or its reversal. See Diaconis (1988), Chapter 6A.

The preliminary sum of squares analysis simplified things considerably, eliminating a 36-dimensional space of ordered second order effects. An earlier attempt at analysis tabulated the 20×20 matrix with entry $(i, i'),(j, j')$ the number of people ranking candidates i and i' in positions (j, j'). This was hard to look at and no obvious pattern emerged.

Some further analyses of these data are in Section 4 which deals with inferential issues.

3. **Representation theory for partial rankings.**   Consider a list of n items. Let $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ be a partition of n. A *partial ranking of shape* λ is specified according to the following instructions: Choose your favorite λ_1 items from the list but don't bother to rank within. Then choose your next, λ_2, favorite items from the list but don't rank within and so on. Partial rankings are written as a collection of subsets. For example, if $\lambda = (3, 2, 1)$, a typical partial ranking is

$$\underline{1} \quad \underline{4} \quad \underline{2}$$
$$\underline{3} \quad \underline{6} \qquad ,$$
$$\underline{5}$$

so items $1, 4, 2$ are ranked first and item 5 is ranked last. Underlining denotes an unordered row.

The partition with all parts equal to 1, denoted 1^n, defines complete rankings. The partition $(q, n - q)$ defines the choice of a subset of q out of n. Such subset data arise in analysis of bettor's behavior in state lotteries—in California $n = 49$ and $q = 6$. Choice of *ordered* subset of q out of n is indexed by $n - q, 1^q$. More general partial rankings are also in use, as in the q-sort data used in psychological studies. Examples and references are in Diaconis (1988), Chapter 5. An example is given at the end of this section.

The number of partial rankings of shape λ is $n!/\lambda_1!\lambda_2! \cdots \lambda_r!$ Let X_λ denote the set of all such rankings. Let M^λ be the set of all real functions on partial rankings. This is a linear space of dimension $n!/\lambda_1! \cdots \lambda_r!$

The symmetric group S_n takes one partial ranking into another as in the example

$$
\pi \begin{pmatrix} \underline{1} & \underline{4} & \underline{2} \\ \underline{3} & \underline{6} & \\ \underline{5} & & \end{pmatrix} = \begin{matrix} \pi(1) & \pi(2) & \dot{\pi}(4) \\ \pi(3) & \pi(6) & \\ \pi(5) & & \end{matrix} .
$$

Thus, for $f \in M^\lambda$, πf is defined at x by $\pi f(x) = f(\pi^{-1}(x))$.

It is often convenient to arrange partitions in decreasing order $\lambda = (\lambda_1, \ldots, \lambda_r)$, $\lambda_1 \geq \lambda_2 \geq \cdots (\geq \lambda_r$, $\lambda_1 + \cdots + \lambda_r = n$. For example, data based on people's favorite and least favorite item on a list of n are naturally indexed by $\lambda = 1$, $n - 2, 1$. It is clearly equivalent to data indexed by $n - 2, 1, 1$. In what follows, partitions are assumed to be decreasing.

A subspace V of M^λ is *invariant* if $f \in V$ implies $\pi f \in V$. A subspace is *irreducible* if it does not contain a nontrivial invariant subspace. Two invariant subspaces V and W are *isomorphic* if there is a 1-1 linear map L from V onto W such that $\pi L f = L \pi f$ for each $f \in V$ and $\pi \in S_n$. Thus isomorphic spaces are equivalent up to change of basis.

An elementary theorem in representation theory implies that M^λ decomposes into a direct sum of invariant irreducible subspaces. Elegant, elementary treatments of representation theory are given by Serre (1977) and Ledermann (1977). Diaconis (1988) develops representation theory for statistical applications. Theorems in Diaconis (1988) are numbered to match theorems in Serre (1977). The basic decomposition is in Serre (1977), Theorem 2.

There is a well defined irreducible invariant subspace S^μ for each partition μ of n. These can be defined as suitable subspaces of M^μ. See James (1978) or Diaconis (1988), Chapter 8. As μ varies over partitions of n, the S^μ exhaust all of the possible invariant irreducible subspaces in the sense that any invariant subspace of any M^λ is isomorphic to a direct sum of S^μ's. In particular

$$
M^\lambda = \oplus k(\lambda, \mu)S^\mu.
$$

The notation means that M^λ decomposes into a direct sum of invariant irreducible subspaces, with $k(\lambda, \mu)$ of these subspaces isomorphic to S^μ.

The direct sum of all invariant subspaces isomorphic to a given S^μ is called the *isotypic subspace* belonging to the partition μ. There is no standard notation for this subspace. We will denote it V_λ^μ. Thus, $V_\lambda^\mu = k(\lambda, \mu)S^\mu$.

There is a simple method for computing the projection onto the isotypic subspace V_λ^μ. This involves the character table of the group. For each partition μ of n, there is a well defined function $\chi_\mu(\pi)$—the character associated to μ at the permutation π. Theorem 8 in Serre (1977) translates into the following.

THEOREM 1. *Let λ and μ be partitions of n. Let $f \in M^\lambda$. The orthogonal projection of f onto the isotypic subspace V_λ^μ is the function*

$$(3.1) \qquad \bar{f}_\mu(x) = \frac{\chi_\mu(\mathrm{id})}{n!} \sum_{\pi \in S_n} \chi_\mu(\pi) f\big(\pi^{-1}(x)\big)$$

with χ_μ the character corresponding to μ.

The characters of the symmetric group are tabulated for $n \le 15$. A variety of combinatorial algorithms for their computation is also available. See James and Kerber (1981), Chapter 4 and tables of Appendix 1. There is not at present a usable "formula" for $\chi_\mu(\pi)$. Observe that the sum (3.1) is over S_n. For large n, different projection formulas are available. See Diaconis (1988), Chapter 8 and Diaconis and Rockmore (1988).

EXAMPLE 1. S_5. The computations of Section 2 are all carried out using formula (3.1) and the character table of S_5 (Table 10). To explain, the character $\chi_\mu(\pi)$ is invariant under conjugation $\chi_\mu(\eta\pi\eta^{-1}) = \chi_\mu(\pi)$ and thus must only be tabulated for conjugacy classes. These only depend on the cycle type (number of fixed points, transpositions, etc.) in π. For instance, the identity has cycle type 1^5, so the first column lists $\chi_\mu(\mathrm{id})$. This equals the dimension of the irreducible representation S^μ [Serre (1977), Theorem 2]. The second column lists χ_μ at a transposition, the third at a product of two 2-cycles, the fourth at a 3-cycle, the fifth at the product of 2- and 3-cycles, the sixth at a 4-cycle and the seventh at a 5-cycle.

An interpretable function is projected onto an isotypic subspace using (3.1). Then, the inner product with the data $f(x)$ is computed to give a typical table entry.

TABLE 10
Character table of S_5

μ	1^5	$1^3, 2$	$1, 2^2$	$1^2, 3$	$2, 3$	$1, 4$	5
5	1	1	1	1	1	1	1
4, 1	4	2	0	1	-1	0	-1
3, 2	5	1	1	-1	1	-1	0
3, 1, 1	6	0	-2	0	0	0	1
2, 2, 1	5	-1	1	-1	-1	1	0
2, 1, 1, 1	4	-2	0	1	1	0	-1
1, 1, 1, 1, 1	1	-1	1	1	-1	-1	1

EXAMPLE 2. *First order projections.* The partition $\mu = (n - 1, 1)$ indexes what was called the first order projection. Here the character can be computed explicitly; see James (1978), page 26:

$$\chi_{n-1,1}(\pi) = \#(\text{fixed points in } \pi) - 1.$$

From this, an explicit form of the projection can be written down. For example, consider $\lambda = n - q, 1^q$—the partition corresponding to ordered rankings of q out of n. The projection of $f \in M^\lambda$ to $\bar{f} \in V_\lambda^{n-1,1}$ is

$$\bar{f}(x) = \frac{(n-1)}{n(n-1)\cdots(n-q)} \sum_y f(y)\{m(x, y)(n - q) - M(x, y)\}$$

with $m(x, y)$ the number of i such that $x_i = y_i$, $1 \le i \le q$, and $M(x, y)$ the number of i such that $y_i \notin \{x_1, \ldots, x_q\}$.

EXAMPLE 3. *An intuitive interpretation of the splitting.* Consider simple choice data—1 out of n. The splitting is

$$M^{n-1,1} = S^n \oplus S^{n-1,1}.$$

Here $f(i) \in M^{n-1,1}$ counts how many people choose i. The decomposition is $f = \bar{f} + (f - \bar{f})$, with $\bar{f} = (f(1) + f(2) + \cdots + f(n))/n$.

Consider next choosing an unordered pair out of n. The splitting is

$$M^{n-2,2} = S^n \oplus S^{n-1,1} \oplus S^{n-2,2}.$$

Here the projection onto S^n is the mean $\Sigma f\{i, j\}/n(n - 1)$. Use of Example 2 shows that the projection onto $S^{n-1,1}$ is equivalent to computing $\bar{f}(i) = \Sigma_j f\{i, j\}$, $1 \le i \le n - 1$. The projection onto $S^{n-2,2}$ is what's left after the mean and popularity of individual items are subtracted out.

As explained below, the splitting of ordered pair data is

$$M^{n-2,1,1} = S^n \oplus 2S^{n-1,1} \oplus S^{n-2,2} \oplus S^{n-2,1,1}.$$

Here the two copies of the space $S^{n-1,1}$ have the interpretation of the effect of item i in first and second position for $1 < i \le n - 1$. The $S^{n-2,2}$ projection is an unordered pair effect. The $S^{n-2,1,1}$ is an ordered pair effect, after the mean, first order and unordered pair effects have been removed.

YOUNG'S RULE. There is an elegant combinatorial rule for deciding which irreducible subspaces S^μ appear in the splitting of M^λ, as well as their multiplicities $k(\lambda, \mu)$. Fix a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r)$, with $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_r > 0$ and $\lambda_1 + \cdots + \lambda_r = n$. Consider λ_1 ones, λ_2 twos,..., λ_r r's. Write these symbols in left justified rows, in all ways that are weakly increasing along a row and strictly increasing down columns. Thus, if $n = 5$ and $\lambda = (3, 1, 1)$, start with $1, 1, 1, 2, 3$ and arrange as

	1,1,1,2,3	1,1,1,2	1,1,1,3	1,1,1	1,1,1
		3	2	2,3	2
					3
Partition	5	4,1	4,1	3,2	3,1,1

The partition determined by each arrangement has been written underneath. Here $M^{3,1,1}$ splits into $S^5 \oplus 2S^{4,1} \oplus S^{3,2} \oplus S^{3,1,1}$.

Young's rule says that the irreducible S^μ appears in M^λ with multiplicity $k(\lambda, \mu)$—the number of arrangements of λ_i—i's into shape μ. Young's rule is proved in James (1978), Chapter 14.

The reader may compare with the decompositions given in the examples above and in Section 2. It is an instructive exercise to compute and interpret the decomposition of $M^{n-3,1,1,1}$. A computer program for computing $k(\lambda, \mu)$ is given in Remmell and Whitney (1984).

One consequence of Young's rule which has an easy direct proof [Serre (1977), Section 2.4] is that for complete ranked data the multiplicity of S^μ equals its dimension. Thus $S^{4,1}$ is a four-dimensional representation appearing four times in the decomposition of $M^{1,1,1,1,1}$ in Table 3.

We turn next to a combinatorial rule for computing the dimension of S^λ.

THE HOOK-LENGTH FORMULA.   Consider a partition λ of n. Arrange n boxes in left justified rows with λ_i boxes in row i. Thus, for $\lambda = (3, 2, 1)$, the arrangement is
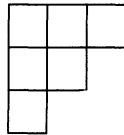
Each box determines the length of a hook centered at the box, going right and down as far as possible. Thus the second box in the first row determines the hook of length 3:
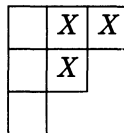
The hook lengths are thus
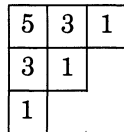
The hook-length formula says that the dimension of the irreducible representation S^λ equals $n!$ divided by the product of the hook lengths. Thus, $S^{3,2,1}$ has dimension $6!/5 \cdot 3 \cdot 3 = 16$.

The formula is originally due to Frame, Robinson and Thrall. For an elementary probabilistic proof, see Greene, Nijenhuis and Wilf (1979). For a fascinating application to determining the average length of the longest increasing subsequence in a random permutation, see Logan and Shepp (1977) and Kerov and Vershick (1985). Diaconis (1988), Chapter 8 contains further discussion.

There are many other formulas for the dimension of S^λ. For small n, the most convenient is dim $S^\lambda = \chi_\lambda(1)$ [Serre (1977), Theorem 3].

A DIFFERENT EXAMPLE. To conclude this section here is a fresh collection of examples where partially ranked data arise. In panel study data, and elsewhere, one considers many relatively short time series $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ with $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})$ and X_{ip} taking values in a finite set. For example, $\mathbf{X}_i$ might be the result of following the ith person for $p = 12$ months and noting if they were employed or not. Biostatisticians and econometricians have developed a rich collection of techniques for analyzing such models. See Hsiao (1986) for a survey.

One natural class of models has each $\mathbf{X}_i$ an independent multinomial process with parameters depending on i. This seems like a rich class of models for p small, even if n is large.

The usual approach to testing such a model compares with the known law of an ancillary conditioned on a fixed value of a sufficient statistic. Here the sufficient statistic for the ith person is T_i—a vector recording the number of X_{ij} taking each possible value. This gives a partition of p.

Fix a partition λ of p. The ancillary information specifies the locations where the different values occur. This gives a partial ranking of shape λ. Considering all $N(\lambda)$ X_i with a fixed partition λ gives a function on these partial rankings. A test of the model can be based on this function: Under the null hypothesis, the function should look like $N(\lambda)$ balls dropped uniformly into $p!/\pi\lambda_i!$ boxes. A test statistic can be defined (e.g., chi-square or the empty cell test) and results from different tests (as λ varies) combined. Frydman and Singer (1985) carry out such a test in a binary setting and develop tests against Markov models.

With or without testing, the function can be spectrally analyzed (as suggested here) in a search for understanding model failure. Diaconis and Smith (1988) carry out and discuss such an analysis.

4. On inference.

4A. *Introduction.* It is natural to ask how sure we are about main effects. If the data are a sample from a population, sampling variability gives a notion of noise. It is straightforward to bootstrap any of the analyses suggested in Sections 2 or 3. Alternatively, we may put a prior on the "true underlying population" and use Bayesian arguments. Both routes are discussed further below.

For the voting example of Section 2, it does not seem reasonable to regard the voters as a sample. At the opposite extreme, the data may be regarded as a complete enumeration of a finite population. Then there is no inference problem.

There *is* useful work to be done for summarization. Hoaglin, Mosteller and Tukey (1985), Chapter 1 contrast the inferential and summarization positions.

Intermediate notions of variability are discussed below. Section 4B asks if we couldn't be finding patterns in noise. Section 4C refines this idea to a hierarchical, conditional version, developing a conditional Monte Carlo procedure. Section 4D shows how earlier considerations merge with conclusions drawn from exponential families through the summary statistics. This follows work of Martin-Löf quite closely.

Section 4E carries out a classical normal analysis on the square roots of the counts. The final section compares the various approaches.

4B. *Patterns in noise.* Consider the 5141 voters who ranked only one candidate:

Candidate	1	2	3	4	5
No.	895	881	1198	1145	1022

These can be summarized as "3 and 4 are most popular, 2 and 1 least popular, with 5 in between." Of course, any five numbers fall into some pattern. One natural comparison asks how different cell counts are likely to be if 5141 balls are dropped at random into five boxes. Here $\sqrt{5141 \cdot 4/25} = 29$, so the observed differences are larger than random variation.

Of course, the voting process is nothing like dropping balls into boxes. Nonetheless, ± 29 seems like a useful number: At least we aren't making the mistake of interpreting patterns in noise.

It is straightforward to compute similar estimates of variability for the kinds of linear functions and sums of squares considered in Section 2. Let X be a finite set. If N balls are dropped uniformly at random into $|X|$ boxes, let Y_x be the number in the xth box. Let $L(X)$ be the set of all real valued functions on X. Regard $\{Y_x\}$ as a function in $L(X)$. The following theorem gives the asymptotic distribution of the projection of Y_x onto a subspace $V \subset L(X)$. The elementary proof is omitted.

THEOREM 2. *With notation as above, with increasing N:*

(i) *For $V \subset L(X)$ a d-dimensional subspace of $L(X)$ orthogonal to the constants, the squared length of the projection of Y_x onto V is approximately distributed as $(N/|X|)\chi_d^2$, with χ_d^2 a chi-square variable having d degrees of freedom.*

(ii) *For V as in (i) and $g \in V$, $\langle g|Y \rangle$ is approximately distributed as a normal variable with mean zero and variance $N\|g\|^2/|X|$.*

(iii) *Projections into orthogonal subspaces are asymptotically independent.*

For example, Table 11 compares the observed and expected sums of squares for the seven isotypic subspaces of Table 3. The sums of squares have been divided by 120.

TABLE 11
Sums of squares for completely ranked data of Table 3

Subspace	$4S^{4,1}$	$5S^{3,2}$	$6S^{3,1,1}$	$5S^{2,2,1}$	$4S^{2,1,1,1}$	$S^{1,1,1,1,1}$
Dimension	16	25	36	25	16	1
SS/120	298.3	459.2	78.2	27.2	6.8	0.218
ESS/120	6.4	10.0	14.3	10.0	6.4	0.40

TABLE 12
Standard deviations for inner products in Tables 4–7

	q		
	2	3 ·	4
First order	$\pm 20(16/5)$	$\pm 18(48/5)$	$\pm 30(96/5)$
Second order	$\pm 35.1(10)$	$\pm 32.5(30)$	$\pm 53.6(60)$

REMARK. The projections onto $4S^{2,1,1,1}$ and $S^{1,1,1,1,1}$ are compatible with uniformity. The projections onto $4S^{4,1}$ and $5S^{3,2}$ show a great deal of structure. The projection onto $6S^{3,1,1}$ and perhaps $5S^{2,2,1}$ may merit further investigation.

As a second example, Table 12 gives the standard deviation of the inner product of $Y \in L(X_{5-q,1^q})$ with any "easily interpretable" function δ in the first order subspace $qS^{5-1,1}$, or in the second order subspace $\binom{q}{2}S^{5-2,2}$. The squared length of δ is shown in parentheses. Of course, these standard deviations are for the sample sizes that actually occurred in the APA data. For example, when $q = 4$ (fully ranked data), the standard deviation for the first order effects reported in Table 4 is 30. This provides one standard of comparison—the large observed differences are unlikely under uniform choice of the 5738 voters who ranked all candidates.

REMARK. For ranked data of shape $n - q, 1^q$ consider the first order function $\delta(x) = 1$ if $x_i = j$; 0 elsewhere. The first order subspace $qS^{n-1,1}$ has dimension $q(n - 1)$. It follows from Example 2 that the projection of δ onto $qS^{n-1,1}$ is $\bar{\delta}(x) = 1 - 1/n$ if $x_i = j$; $-1/n$ elsewhere. Thus $\|\bar{\delta}\|^2 = (n - 1)!(1 - 1/n)/(n - q)!$ This gives the entries in parentheses in the first row of Table 12.

REMARK. The data vector in the APA example is sufficiently close to uniform that there is only a small difference between the uniform analysis suggested here and an analysis under multinomial variability.

4C. *On conditional uniformity.* There is a more subtle use of the uniform distribution available by conditioning on observed low order statistics. Let X be a finite set and $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ a vector of points in X. Let $T(x_1, x_2, \ldots, x_N)$

be a statistic. One interpretation of the phrase "the data can be summarized by T" is that we are indifferent between two data sets with the same value of T. This gives a notion of variability for a second statistic U: Compute the conditional law of U given $T = t$ under the uniform distribution on $T^{-1}(t)$.

This idea is very close in spirit to proposals suggested by Martin-Löf (1970, 1974, 1975). It is also close in spirit to the modern Bayesian approach to de Finetti's theorem on exchangeability. Lauritzen (1984) and Diaconis and Freedman (1984) contain extensive reviews of this work and its interconnections.

In the present setting $X = X_\lambda$—the partial rankings of n items of shape $\lambda = (\lambda_1, \ldots, \lambda_r)$, T is the first order summary statistic—the projection of the data onto the $r - 1$ copies of $S^{n-1,1}$ in some fixed coordinate system. U is the second order summary.

One way to investigate the variability of the second order effects is to generate a Monte Carlo sample from the uniform distribution conditional on the observed first order effects. A method for doing this is given by Diaconis and Gangolli (1987). A second method for approximate generation is described in Section 4D below.

A conditionally uniform sample provides a test of higher order structure. In the example, the variability is essentially the same as the unconditional variability reported in Section 4B. Besides this, the second order structure seems sufficiently clear without a test.

VOLUME TESTING. If voters rank uniformly, the resulting *function* has a multinomial distribution. There is another notion of uniformity which has a direct appeal. This regards $f(\pi)$ as being chosen at random from among all functions with a given sum. Thus, f has a Bose–Einstein distribution. This distribution arises as the marginal distribution if a uniform distribution is put on the underlying multinomial probabilities. Diaconis and Efron (1985) developed this point of view for contingency table testing. It does not seem so easy to simulate this kind of variability at present.

4D. *Related exponential families.* The first and second order statistics introduced above can be thought of as estimates of parameters in a model. This permits sampling theory or Bayesian methods to give confidence intervals and tests for the need of higher order terms. Moreover, Martin-Löf has pointed out that the conditional uniform approach of Sections 4B and 4C, which seems useful from the summarization view, translates into calculations within the model.

As in Section 4C, let X be a finite set and $\mathbf{x} = (x_1, \ldots, x_N)$ a vector of points in X. Let $T_N(\mathbf{x}) = (1/N)\sum_{i=1}^N T(x_i)$ be a statistic. Here is a fundamental mathematical fact borrowed from statistical mechanics. The conditional uniform distribution over all N-tuples $y_1, \ldots, y_N$ with $T_N(\mathbf{y}) = t$ is approximately the same as the law of $Y_1, Y_2, \ldots, Y_N$ where Y_i are independent and identically distributed with law

$$(4.1) \qquad P_{\hat{\theta}}(y) = C(\hat{\theta})e^{\theta \cdot T(y)}, \qquad C(\theta) \text{ a normalizing constant,}$$

and $\hat{\theta}$ contained in a Euclidean space of the dimension of t chosen to satisfy

(4.2) $E\{T(Y)\} = t.$

This result is called the equivalence of ensembles—the microcanonical ensemble (conditionally uniform) is approximately the same as the macrocanonical ensemble (i.i.d. through the matching exponential family).

Regularity conditions and rigorous statement have been suppressed above. The result has some delicate aspects. For example, if the statistic T can take irrational values, the data can be recovered from $T_N(x)$ (the statistics we deal with have values in the integers). The sense of the approximation must also be specified. Rigorous versions can be found in Martin-Löf (1970), Lauritzen (1984), Lanford (1971) and Diaconis and Freedman (1980, 1984, 1987, 1988). These last authors give references to related papers. It must be said that much remains to be done in rigorizing things and the present discussion should be regarded speculatively.

In the examples here, $X = X_\lambda$—the partial rankings of n items of shape $\lambda = (\lambda_1, \ldots, \lambda_r)$, T is the first order summary statistic—the projection onto the $(r-1)$ copies of $S^{n-1,1}$ in some fixed coordinate system. The parameter space is all of $\mathbb{R}^{(r-1)(n-1)}$ and the family is well parametrized [see Diaconis (1988), Lemma 1, Chapter 9]. The $\hat{\theta}$ satisfying (4.2) is the maximum likelihood estimate in this family. As Fisher (1922) pointed out, the projection T_N is an efficient estimate for the mean value parameter in this family.

Models of the form (4.1) for ranked data have been suggested by Holland and Silverberg [see Silverberg (1980)]. Verducci (1982, 1987, 1989) gives an imaginative investigation of such models and their submodels, as well as efficient algorithms for calculating $C(\theta)$ and $\hat{\theta}$. Diaconis (1988), Chapter 9 contains an extensive discussion and more careful pointers to the literature.

Classical theory is available to investigate $\hat{\theta}$. The main interest here is on notions of variability for the observed values of T_N and higher order statistics.

To fix ideas, consider the problem of determining the law of the second order statistics $U_N(\mathbf{x})$ reported in Table 5 under the uniform distribution conditional on the first order statistics of Table 4.

One route through the computations estimates $\hat{\theta}$ and then the distribution of the second order statistics under $P_{\hat{\theta}}$. For samples of the size considered here, the distribution will be well approximated by a normal law with mean $E_{\hat{\theta}}(U(X))$ and covariance matrix $\text{Var}_{\hat{\theta}}(U(X))$. For small n, these can be obtained by directly summing over X_λ. For larger n, the Metropolis algorithm is recommended [see Hammersley and Hanscomb (1964), Chapter 9].

In the example, a first order model was fitted using the algorithm in Verducci (1989). The variability for the second order analysis is essentially the same as in Section 4B above.

The above discussion does not do justice to the program outlined by Martin-Löf. He develops an analog of Fisher's exact test and an asymptotically equivalent parametric version with a reasonably tractable chi-square approximation. Testing if the second order effects are real did not seem warranted in the example (they are clearly huge). It would be of interest to test if ordered second order and higher order effects have been foolishly neglected.

Martin-Löf offers a coding theory explanation of his basic test statistic as the decrease in the number of bits needed to specify the data when the regularities detected by the test are taken into account.

4E. *A normal theory analysis.* For data which are a sample from a larger population, classical approaches to estimating main effects and variability are available. One straightforward route uses the Poisson approximation and then square roots to stabilize the variance. For large mean parameters, the Poisson is approximately normal so all the tools of analysis of variance are applicable.

In more detail, let S_n be the permutation group on n letters. Introduce random variables $\{Y_\pi\}_{\pi \in S_n}$, with Y_π Poisson distributed with parameter $\lambda\theta(\pi)$. Here $\sum_{\pi \in S_n}\theta(\pi) = 1$ and $\lambda > 0$ are parameters. As is well known [see, e.g., Rao (1965), page 231], for $\lambda\theta(\pi)$ large, $\sqrt{Y_\pi}$ is approximately normal, with mean $\sqrt{\lambda\theta(\pi)}$ and variance $\frac{1}{4}$.

A linear analysis of the transformed, fully ranked data is presented in Table 13.

TABLE 13

Square roots of fully ranked data. $q = 4$, $n = 5738$

$$M^{1,1,1,1,1} = S^5 \oplus 4S^{4,1} \oplus 5S^{3,2} \oplus 6S^{3,1,1} \oplus 5S^{2,2,1} \oplus 4S^{2,1,1,1} \oplus S^{1,1,1,1,1}$$

Dim	120	1	16	25	36	25	16	1
SS/120		5311	146	212	25	9	4	0

First order effects (s.d. = ±2.2)

	Rank				
Candidate	1	2	3	4	5
1	−4.9	20.8	12.7	−8.6	−20.0
2	−25.1	−2.2	15.1	17.5	−4.7
3	27.0	−13.8	−24.3	−3.2	14.4
4	2.7	−11.6	−1.1	−0.4	10.4
5	1.0	6.8	−2.4	−5.4	−0.1

Second order effects (s.d. = ±3.9)

	Rank									
Candidate	12	13	14	15	23	24	25	34	35	45
12	−8.7	−0.7	1.2	8.1	6.2	1.7	0.8	0.2	−5.7	−3.1
13	29.1	−5.0	−11.4	−12.7	−7.8	−11.4	−9.8	6.6	6.2	16.3
14	−11.9	1.8	8.0	2.2	0.1	6.6	5.2	−4.5	2.7	−10.0
15	−8.5	3.9	2.2	2.4	1.5	3.1	3.9	−2.2	−3.2	−3.1
23	−2.3	3.8	1.0	−2.6	1.8	−1.2	1.7	−3.2	−2.5	3.4
24	10.6	−0.8	−3.7	−6.2	−1.4	−5.1	−4.2	0.2	1.9	8.5
25	0.3	−2.4	1.4	0.8	−6.6	4.5	1.8	2.8	6.2	−8.7
34	−16.9	0.8	4.9	11.2	1.1	1.4	6.4	0.8	−2.6	−15.0
35	−10.0	0.3	5.6	4.1	4.9	3.3	1.7	−4.2	−1.1	−4.7
45	18.1	−1.8	−9.2	−7.2	0.2	−10.1	−7.4	3.5	−2.0	16.6

REMARK 1. The overall picture gleaned from the square roots is the same as from the untransformed data. In the first order effects (projection onto $4S^{4,1}$) candidate 3 is most popular, but has some hate vote, etc. The sum of squares shows a large unordered pair effect. The second order structure seems the same: candidates 1 and 3 versus 4 and 5.

REMARK 2. The benefit of square roots is that all the numbers in a table have the same variability. The s.d.'s follow from the lengths of the projections given in Table 12: $2.2 = \sqrt{96/5 \cdot 1/4}$; $3.9 = \sqrt{60/4}$. Under the assumptions, they are appropriate measures of variability for each table entry. Of course, they are only valid marginally; the correlations associated with 100 entries and 25 degrees of freedom are still present.

REMARK 3. Two problems with the square root analysis: The square roots are hard to interpret. Also, for this data set, sampling variability seems far-fetched.

REMARK 4. The variance stabilizing transformation has the amazing property that all variables have variance $\frac{1}{4}$. It seems statistically natural to put a parameter σ^2 in place of $\frac{1}{4}$. This can then be estimated from the projections on the higher order subspaces. This way of thinking leads to analysis associated to generalized linear models (GLIM) which works directly with variations of Poisson likelihood. McCullagh and Nelder (1983) or Efron (1986) give useful treatments. I have not tried out this approach.

REMARK 5. If square roots are admitted as useful transformations, why not logs or nonparametric functions chosen to give best fits to linearity. The ACE algorithm [Breiman and Friedman (1985)] and generalized additive models [Stone (1985) and Hastie and Tibshirani (1986)] offer convenient technology for trying out such an analysis.

4F. *Final comments.* There is much to be said for an analysis based on directly interpretable averages which can be related to other information and covariates. This is the main appeal of spectral analysis.

The conditional uniformity assumption introduced in Sections 4B and C is a frail straw man. It allows conclusions like "if votes were randomly allocated, these differences would be surprising." Since there is no reason to think people vote randomly, this cannot hold much force beyond protecting us from finding patterns in noise.

The classical approach—sampling variability—is treated in Section 4E. It can be valid and useful but it is easily abused. Data are often not reasonably regarded as a sample from a population of interest. On the other hand, the sampling analysis is equivalent to the conditional uniform analysis through consideration of Section 4D above.

Ranked data have a built-in high dimensionality which is not apparent in the example used here. If 10 items are ranked, it is rare that there will be a sample of

sufficient size to give good estimates of individual ranking frequency. Low-dimensional models may be the only viable route to comparison and analysis.

The inferential developments in Sections 4A–E are a central part of statistics. In the example, the various $\pm$ numbers don't go much beyond what one gets by direct inspection of the tables to see the general size of the numbers and how they compare. It *is* reassuring when several different routes through the maze lead to similar answers. The comparison of observed and expected sum of squares summarized in Table 11 does suggest some further structure is present in the $S^{3,1,1}$ and $S^{2,2,1}$ spaces. This can be seen by going back to the original data in Table 1: The two largest counts 186 and 172 suggest an interaction between candidates 1, 2 and 3 that has not been explored.

5. General spectral analysis. The techniques of the previous sections are natural extensions of analysis of variance and spectral analysis of time series. More generally, let X be a finite set. Let G be a finite group operating transitively on X. Let $L(X)$ be the space of all functions on X with values in $\mathbb{R}$ (or C). This is a vector space on which G acts linearly as a group of transformations $[gf(x) = f(g^{-1}x)]$. Serre (1977), Theorem 1 asserts that $L(X)$ decomposes into a direct sum of invariant irreducible subspaces

$$(5.1) \qquad L(X) = V_0 \oplus V_1 \oplus \cdots \oplus V_k.$$

Let $f(x)$ be a data set (the number of times x appears in the sample). *Spectral analysis* is the projection of f onto the invariant subspaces and the approximation of f by as many pieces as required to give a reasonable fit.

In time series, X is the integers mod n, $L(X)$ is all complex functions. The space $L(X)$ splits into n one-dimensional irreducibles, the jth being spanned by $x \to e^{2\pi ijx/n}$. The projection is summarized by the discrete Fourier transform,

$$f(x) = \frac{1}{n} \sum_{j=0}^{n-1} \hat{f}(j) e^{-2\pi ijx/n}, \qquad \hat{f}(j) = \sum_{k=0}^{n-1} f(k) e^{2\pi ijk/n}.$$

Examining the spectrum $\hat{f}(j)$ is a familiar activity.

In analysis of variance X labels the underlying units involved in the experiment and G is an appropriately chosen group of symmetries. For example, for data in a two way array, with I rows, J columns and no repeated observations, $X = \{(i, j): 1 \le i \le I, 1 \le j \le J\}$. The group $S_I \times S_J$ operates transitively on X by $(\pi, \eta)(i, j) = (\pi(i), \eta(j))$. $L(X)$ can be taken as the real valued functions on X. It decomposes as

$$
\begin{array}{ccccccc}
L(X) & = & V_0 & \oplus & V_1 & \oplus & V_2 & \oplus & V_3 \\
\text{Dim } I \times J & & 1 & & I-1 & & J-1 & & (I-1)(J-1)
\end{array}.
$$

The projections are the usual grand mean, row effects, column effects and residuals.

Spectral analysis has also been applied for data on Z_2^n—the group of binary n-tuples (test scores, panel study data)—by Bahadur (1961). Another classical example is harmonic analysis of data on the sphere. Geodesists regularly analyze

such data in a basis of spherical harmonics. Here, the straightforward extension from finite to compact is needed. Then $L(X)$ is replaced by the continuous functions and the direct sum (5.1) may be infinite, but each V_i is finite dimensional. Diaconis (1988), Chapter 8 contains further discussion and references.

There are also interesting problems in noncompact cases. For example, consider a set of studies of height and weight, each producing its own 2×2 covariance matrix. Nelder (personal communication) has described an ongoing study which currently has over 2000 such matrices. One is then in the business of analyzing data on the space of 2×2 positive definite symmetric matrices. This is a homogeneous space for GL_2. The set of all polynomial representations decomposes into a direct sum like (5.1) [see, e.g., James (1961) or Constantine (1963)]. Understanding low order terms seems like a worthwhile project. Similar analysis can be undertaken for data consisting of many lines in the plane under the action of the Euclidean group.

A second direction for generalization involves other orthonormal bases. There has been some work on spectral theory using Radamacher functions and the like. A more promising direction involves consideration of a Markov chain on the set X. The eigenfunctions of the transition matrix may offer a scientifically interesting spectral decomposition. This arises in the association schemes used to analyze designed experiments. Bailey (1985) is a useful review.

One cannot hope to go much beyond groups. Almost all naturally occurring orthogonal functions are connected to a group. Almost all explicitly diagonalizable Markov chains arise from random walks on groups. Finally, it is the desire for invariance under relabeling that makes spectral analysis natural.

The data-analytic approach suggested here proposes a set of linear functions or averages which have a good track record in applied problems. These can lead to novel analyses, even in well studied areas. Here are two examples.

In two-way ANOVA, the group $S_I \times S_J$ does not use the ordering of the rows or columns. Suppose the rows are months of the year, the columns are violent crimes and the entries are average crime rate. The group $Z_I \times S_J$ also operates transitively by $(x, \eta)(i, j) = (i + x, \eta(j))$. The space splits (over the complex numbers) as

$$L(X) = \left(\bigoplus_{j=0}^{I-1} \chi_j \right) \oplus \bigoplus_{j=0}^{I-1} \left(\chi_j \otimes S^{J-1,1} \right)$$

with χ_j the one-dimensional representations appearing in classical spectral analysis and $S^{J-1,1}$ the first order subspaces of Sections 2 and 3 above.

Fortini (1977) has given several examples of the group invariant approach illuminating classical analyses of designed experiments such as balanced incomplete blocks. Diaconis (1988), Chapter 8D gives an exposition of Fortini's results.

As a second example, it would be interesting to push through the inferential ideas of conditional uniformity, with the associated exponential families, in the classical time series context. The models are nonstandard and underscore the fact that the usual spectral estimates are not efficient estimates for parameters in any of the usual models.

Connections between ANOVA and time series as well as noncommutative extensions of the spectral representation of processes have been discussed by a number of authors. Hannan (1965) sketched out the ideas building on an early brief mention by James (1957). This has been carried forward in the mathematical arena by Yaglom (1961) and Ylinen (1986).

The point of view taken in the present paper is a generalization of the data-analytic hunt for periodicity. There is another tradition centered about components of variance, random effects and continuous spectra. This has been clearly discussed by Tukey (1961) and developed in a far ranging way by Bailey and Speed and their coworkers. Speed (1987) is a convenient reference.

A still richer possibility allows the variability of each $f(x)$ to vary and correlate. Then the problem can be approached by modern multivariate techniques as developed by the Danish school. Andersson (1987) or Perlman (1987) give a recent discussion.

The examples presented here show that linear analysis can be instructive. Modern statistics heads in nonlinear directions such as projection pursuit, recursive partitioning and ACE. All of these build on linearity: Projection pursuit fits a nonlinear function to the most informative linear projection; recursive partitioning fits linear models locally; ACE transforms to linearity. All of these ideas are applicable to the linear spectral analyses suggested here. Diaconis (1988), Chapter 8 gives references and discussion for a host of other innovations in ANOVA and time series methods (Bayesian methods, shrinking, robustness, missing data). All are worth carrying over to a genuine application.

APPENDIX

On symmetry characterizations of the isotypic decomposition. Let G be a group, H a subgroup, $X = G/H$ the associated homogeneous space and $L(X)$ the space of all functions into the complex numbers. The space $L(X)$ decomposes into a direct sum of irreducible invariant subspaces. These can be grouped together into isomorphism classes to give the isotypic decomposition. There is one isotypic subspace V_χ for each irreducible character χ. The splitting is denoted

$$L(X) = \oplus V_\chi.$$

See Serre (1977), Theorem 8. This decomposition is computationally convenient. The projection of $f \in L(X)$ into $f_\chi \in V_\chi$ is given by

(A1) $$f_\chi(x) = \frac{d_\chi}{|G|} \sum_{s \in G} \chi(s^{-1}) f(s^{-1}x).$$

When $X = G$ (so $H = $ id) there is a simple characterization of the isotypics: A subspace V in $L(X)$ is a direct sum of isotypic irreducibles if and only if it is invariant under the action of G on both sides. In other words, $G \times G$ acts on functions by $(s, t)f(u) = f(s^{-1}ut)$. The V_χ are the minimal invariants under this action. This follows from the development below.

In the context of fully ranked data, this two-sided invariance amounts to considering only information that doesn't depend on the names of the items or

the order. This is sometimes natural: Consider an ESP experiment in which a psychic describes where a sender is standing on five occasions. A judge takes the five descriptions, visits the actual standing places and matches descriptions and places. With n judges, this gives n permutations. It seems natural to look at two-sided invariant analysis.

Usually, invariance is only natural on one side. Then, the decomposition is supplemented by projections onto known vectors which use the ordering as in Section 2.

It is worth recording a relation between Mallows' device, as used in Section 2, and group invariance. Let $\delta_{x,\,y}(s)$ be 1 if $sy = x$ and 0 elsewhere. There are the easily interpretable functions. The analysis of Section 2 projected a function f on G into one of the V_χ and then reported an $|X| \times |X|$ matrix N—the inner product of the projection with each $\delta_{x,\,y}(\cdot)$.

The $|X| \times |X|$ matrix $M_{xy} = \Sigma\, f(s)\delta_{x,\,y}(s)$ has a direct data-analytic interpretation. Also, it is the Fourier transform of f at the representation $L(X)$ of G. The group $G \times G$ acts on M by $(s, t)M_{xy} = M_{s^{-1}x,\,t^{-1}y}$. It can be shown using (A1) above, that the projection of M onto the $G \times G$ invariant subspace $V_\chi \otimes V_{\bar{\chi}}$ equals the matrix N described above.

A class of examples where isotypics are easy to describe arises when the pair (G, H) form a Gel'fand pair. Then each invariant irreducible subspace is unique so the isotypics and one-sided invariant decompositions coincide. Gel'fand pairs are characterized by this property. For a host of examples, references and other characterizations, see Letac (1981) or Diaconis (1988), Chapter 3G.

A natural Gel'fand pair occurs for partially ranked data with people choosing an unordered committee of k out of n. Then $G = S_n$ and $H = S_k \times S_{n-k}$ (the subgroups permuting the first k items among themselves and the last $n - k$ items among themselves). In the notation of Section 3, $L(X) = M^{k,\,n-k}$ and

$$L(X) = S^n \oplus S^{n-1,1} \oplus S^{n-2,2} \oplus \cdots \oplus S^{n-k,\,k}.$$

For partially ranked data, as in Section 3, $\lambda = n - k, k$ is the only example yielding a Gel'fand pair. Saxl (1981) determines all subgroups of S_n yielding Gel'fand pairs.

One approach to characterizing the isotypic decomposition by invariance considerations uses $\mathrm{End}_G(L(X))$, the linear maps $\phi: L(X) \to L(X)$ that commute with the action of G,

$$s\phi(f) = \phi(sf).$$

For finite groups, f is a function on cosets. Because G acts transitively, ϕ is determined by its action on H, say,

$$\phi(\delta_H) = \sum_x c(x)\delta_{xH}.$$

Acting by $h \in H$, we see $c(hx) = c(x)$ is required and any choice of coefficients satisfying this gives a suitable ϕ. Lifting $c(x)$ back to a function on G [by $c(xh) = c(x)$ as well], we see there is a 1-1 correspondence between bi-invariant

functions c on G and maps $\phi_c \in \text{End}(L(X))$, given by

(A2)
$$\phi_c f(x) = \sum_{y \in X} f(y) c(y^{-1}x).$$

The following lemma characterizes isotypic subspaces as minimal invariants under G and the ϕ_c.

LEMMA A1. *Let G be a finite group, H a subgroup, $X = G/H$ and $L(X)$ the complex functions on X. Let $V \subset L(X)$ be a subspace. Then V is invariant under G and all $\phi \in \text{End}_G(L(X))$ if and only if V is a direct sum of isotypic subspaces.*

PROOF. Choose a basis for $L(X)$ such that G acts as block diagonal matrices, one block for each irreducible representation. Schur's lemma [Serre (1977), Proposition 4] implies that any $\phi \in \text{End}_G(L(X))$ only maps isomorphic pieces among themselves. It follows that any such ϕ preserves isotypic pieces. Conversely, maps permuting isomorphic representations are in End_G, so any G invariant V contains all isomorphic copies of any irreducible it contains. $\square$

There is a geometrical description of invariance under End_G which has a direct, combinatorial flavor. It is stated here for partially ranked data, but similar results can be obtained in any circumstance where the $H \setminus G/H$ double cosets have a nice description. Recall that $H \times H$ acts on G by $(h_1, h_2)s = h_1^{-1} s h_2$. The orbits of this action are called *double cosets*. The indicator functions of these double cosets form a basis for the H bi-invariant functions that appear in (A2) above.

It follows that a subspace V is invariant under End_G if and only if it is invariant under convolving by the indicators of double cosets.

Consider the symmetric group S_n with a Young subgroup S_λ. Here $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r)$ is a partition of n. The homogeneous space $X = S_n/S_\lambda$ can be represented as the space of partial rankings: arrangements of $\{1, 2, \ldots, n\}$ into groups of size $\lambda_1, \lambda_2, \ldots, \lambda_r$, where order within a group doesn't matter and order between groups does matter. A matrix valued function $D: X \times X$ into $r \times r$ matrices, with row and column sums $(\lambda_1, \lambda_2, \ldots, \lambda_r)$ can be defined as

(A3)
$$D(x, y)_{ij} = |x_i \cap y_j|.$$

This is a measure of the discrepancy between the partial rankings x and y. It satisfies some useful properties.

LEMMA A2. *The discrepancy D defined in* (A3) *satisfies*:

  (i) $D(x, y) = D(sx, sy)$ *for $x, y \in X$, $s \in S_n$.*
  (ii) $D(x, x) = \text{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$.
  (iii) $D(x, y) = D(y, x)^t$.
  (iv) *If $x \neq hy$ for some $h \in S_\lambda$, $D(\text{id}, x) \neq D(\text{id}, y)$, so $D(\text{id}, x)$ determines the double coset containing x.*

(v) *Let D be a fixed matrix with nonnegative integer entries having row and column sums $\lambda_1, \lambda_2, \ldots, \lambda_r$.*

(A4) $\#\{x: D(\mathrm{id}, x) = D\} = \dfrac{\prod_{i=1}^{r} \lambda_i!}{\prod_{ij} D_{ij}!} =_d N(D).$

(vi) $n\text{-Tr}(D(x, y))$ *is a G invariant metric on X.*

PROOF. Claims (i)–(iii) are obvious. Claim (iv) is argued in James and Kerber (1981), page 101. For claim (v), an x with x_1 having D_{11} elements in common with $\{1, 2, \ldots, \lambda_1\}$, D_{12} elements in common with $\{\lambda_1 + 1, \ldots, \lambda_1 + \lambda_2\}$ and so on can be chosen in $\prod_{i=1}^{r} \binom{\lambda_i}{D_{1i}}$ ways. The next row of x can be chosen in $\prod_{i=1}^{r} \binom{\lambda_i - D_{1i}}{D_{2i}}$ ways and so on. This product equals the right side of (A4). For claim (vi) $n - D(x, y) = 0$ if and only if $|x_i \cap y_i| = \lambda_i$ for all n; symmetry comes from 3. The triangle inequality follows from the triangle inequalities for $\lambda_i - |x_i \cap y_i|$ for all i. This is well known; see, e.g., Diaconis (1988), Chapter 6D, Example 1 or Critchlow (1985). $\square$

The discrepancy D allows a geometric description of End_G invariance. Consider $f \in L(X)$. Given a fixed matrix D, let

(A5) $\phi_D f(x) = \dfrac{1}{N(D)} \displaystyle\sum_{D(x, y) = D} f(y)$

with $N(D)$ defined by (A4) above. Thus ϕ_D replaces f by its average over all y at discrepancy D. The remarks above may be summarized as follows.

PROPOSITION. *Let $X = S_n/S_\lambda$. An S_n invariant subspace $V \subset L(X)$ is a direct sum of isotypic subspaces if and only if V is invariant under all averages ϕ_D defined in (A5). Isotypics are the minimal invariants.*

REMARK. For $\lambda = (1, 1, \ldots, 1)$ and $S_\lambda = \{\mathrm{id}\}$, the matrices D are permutation matrices and ϕ_D invariance is equivalent to invariance under S_n acting on the right. For $\lambda = (k, n - k)$ it can be shown that for any $s \in S_n$, $s^{-1} \in S_\lambda s S_\lambda$. This implies that left S_n invariant subspaces are automatically invariant under convolution by S_λ bi-invariant functions. Here, the discrepancy is equivalent to the metric $k - |x_1 \cap y_1|$.

For more general λ, the combinatorial interpretation given here, while not totally transparent, seems like a good start at interpretation of the isotypic subspaces. It is instructive, even in the case of Gel'fand pairs.

Checking (A5) will be vastly easier if the following conjecture holds: There are two distance matrices D_1, D_2 such that if (A5) holds for D_1, D_2 it holds for all D. This seems plausible because most algebras are generated by two elements.

REFERENCES

ANDERSSON, S. (1987). The lattice structure of orthogonal linear models and orthogonal variance component models. Technical Report, Inst. Mathematical Statistics, Univ. Copenhagen.

BAHADUR, R. R. (1961). On classification based on responses to n dichotomous items. In *Studies in Item Analysis* (H. Solomon, ed.). Stanford Univ. Press, Stanford, Calif.

BAILEY, R. (1985). Partially balanced designs. *Encyclopedia of Statistical Sciences* **6** 593–610. Wiley, New York.

BATSELL, R. and POLKING, J. (1985). A new class of market share models. *Marketing Sci.* **4** 177–198.

BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.* **80** 580–597.

CARROLL, J. D. (1980). Models and methods for multi-dimensional analysis of preferential choice (or other dominance) data. In *Similarity and Choice* (E. Lanterman and H. Feger, eds.). Hans Huber, Bern.

COHEN, A. (1982). Analysis of large sets of ranking data. *Comm. Statist. A—Theory Methods* **11** 235–256.

COHEN, A. and MALLOWS, C. L. (1980). Analysis of ranking data. Technical Memorandum, Bell Laboratories, Murray Hill, N.J.

CONSTANTINE, A. G. (1963). Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.* **34** 1270–1285.

COOMBS, C. H., COHEN, J. L. and CHAMBERLAIN, J. (1984). An empirical study of some election systems. *Amer. Psychol.* **39** 140–157.

CRITCHLOW, D. (1985). *Metric Methods for Analyzing Partially Ranked Data*. Lecture Notes in Statist. **34**. Springer, Berlin.

DIACONIS, P. (1988). *Group Representations in Probability and Statistics*. IMS, Hayward, Calif.

DIACONIS, P. and EFRON, B. (1985). Testing for independence in a two-way table: New interpretations of the chi-square statistic (with discussion). *Ann. Statist.* **13** 845–913.

DIACONIS, P. and FREEDMAN, D. (1980). Finite exchangeable sequences. *Ann. Probab.* **8** 745–764.

DIACONIS, P. and FREEDMAN, D. (1984). Partial exchangeability and sufficiency. In *Statistics: Applications and New Directions* (J. K. Ghosh and J. Roy, eds.) 205–236. Indian Statistical Institute, Calcutta.

DIACONIS, P. and FREEDMAN, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist.* **23** 397–423.

DIACONIS, P. and FREEDMAN, D. (1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. *J. Theoretical Probab.* **1** 381–410.

DIACONIS, P. and GANGOLLI, A. (1987). Algorithms for choosing at random from large finite sets. Unpublished.

DIACONIS, P. and ROCKMORE, D. (1988). Fast computation of projections onto invariant subspaces. Technical Report, Dept. Mathematics, Harvard Univ.

DIACONIS, P. and SMITH, L. (1988). Spectral analysis for panel study data. Technical Report, Dept. Mathematics, Harvard Univ.

EFRON, B. (1986). Double exponential families and their use in generalized linear regression. *J. Amer. Statist. Assoc.* **81** 709–721.

FEIGIN, P. D. and ALVO, M. (1986). Intergroup diversity and concordance for ranking data: An approach via metrics for permutations. *Ann. Statist.* **14** 691–707.

FISHBURN, P. (1973). *The Theory of Social Choice*. Princeton Univ. Press, Princeton, N.J.

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **222** 309–368.

FLIGNER, M. and VERDUCCI, J. (1986). Distance based ranking models. *J. Roy. Statist. Soc. Ser. B* **48** 359–369.

FORTINI, P. (1977). Representations of groups and the analysis of variance. Ph.D. dissertation, Dept. Statistics, Harvard Univ.

FRYDMAN, H. and SINGER, B. (1985). Assessing qualitative features of longitudinal data. In *Longitudinal Analysis of Labor Market Data* (J. Heckman and B. Singer, eds.) 308–324. Cambridge Univ. Press, Cambridge.

GREENE, C., NIJENHUIS, A. and WILF, H. (1979). A probabilistic proof of a formula for the number of Young tableaux of a given shape. *Adv. in Math.* **31** 104–109.

HAMMERSLEY, J. M. and HANSCOMB, D. C. (1964). *Monte Carlo Methods*. Chapman and Hall, London.

HANNAN, E. J. (1965). Group representations and applied probability. *J. Appl. Probab.* **2** 1–68.

HASTIE, T. and TIBSHIRANI, R. (1986). Generalized additive models (with discussion). *Statist. Sci.* **1** 297–318.

HOAGLIN, D., MOSTELLER, F. and TUKEY, J. (1985). *Exploring Data Tables, Trends, and Shapes*. Wiley, New York.

HSIAO, C. (1986). *Analysis of Panel Data*. Cambridge Univ. Press, Cambridge.

JAMES, A. T. (1957). The relationship algebra of an experimental design. *Ann. Math. Statist.* **28** 993–1002.

JAMES, A. T. (1961). Zonal polynomials of the real, positive definite matrices. *Ann. of Math.* (2) **74** 456–469.

JAMES, G. D. (1977). A characteristic-free approach to the representation theory of S_n. *J. Algebra* **46** 430–450.

JAMES, G. D. (1978). *The Representation Theory of the Symmetric Groups. Lecture Notes in Math.* **682**. Springer, New York.

JAMES, G. D. and KERBER, A. (1981). *The Representation Theory of the Symmetric Group*. Addison-Wesley, Reading, Mass.

KEROV, S. and VERSHICK, A. (1985). The characters of the infinite symmetric group and probability properties of the Robinson–Schensted–Knuth algorithm. *SIAM J. Algebraic Discrete Methods* **7** 116–124.

KNUTH, D. (1981). *The Art of Computer Programming* **2**, 2nd ed. Addison-Wesley, Reading, Mass.

LANFORD, O. (1971). *Entropy and Equilibrium States in Classical Statistical Mechanics. Lecture Notes in Phys.* **20**. Springer, New York.

LAURITZEN, S. (1984). Extreme point models in statistics. *Scand. J. Statist.* **11** 63–91.

LEDERMANN, W. (1977). *Introduction to Group Characters*. Cambridge Univ. Press, Cambridge.

LETAC, G. (1981). Problèmes classiques de probabilité sur un couple de Gelfand. *Lecture Notes in Math.* **861** 93–120. Springer, New York.

LOGAN, B. and SHEPP, L. (1977). A variational problem for random Young tableaux. *Adv. in Math.* **26** 206–222.

MALLOWS, C. (1957). Non-null ranking models. I. *Biometrika* **44** 114–130.

MARTIN-LÖF, P. (1970). Statistika modeller. Mimeographed Lecture Notes, Stockholm Universitet.

MARTIN-LÖF, P. (1974a). The notion of redundancy and its use as a quantitative measure of the discrepancy between a statistical hypothesis and a set of observational data. *Scand. J. Statist.* **1** 3–10.

MARTIN-LÖF, P. (1974b). Repetitive structures and the relation between canonical and microcanonical distributions in statistics and statistical mechanics. In *Proc. Conf. Foundational Questions in Statistical Inference* (O. Barndorff-Nielsen, P. Blæsild and G. Schon, eds.) 271–294. Dept. Theoretical Statistics, Univ. Aarhus.

MARTIN-LÖF, P. (1975). Reply to Erling Sverdrup's polemical article "Tests without power". *Scand. J. Statist.* **2** 161–165.

MCCULLAGH, P. and NELDER, J. (1983). *Generalized Linear Models*. Chapman and Hall, London.

PERLMAN, M. (1987). Group symmetry covariance models. Comment on "A review of multivariate analysis", by M. J. Schervish. *Statist. Sci.* **2** 421–425.

PLACKETT, R. L. (1975). The analysis of permutations. *Appl. Statist.* **24** 193–202.

RAO, C. R. (1965). *Linear Statistical Inference and Its Applications*. Wiley, New York.

REMMEL, J. B. and WHITNEY, R. (1984). Multiplying Schur functions. *J. Algorithms* **5** 471–487.

SAXL, J. (1981). On multiplicity-free permutation representations. *Finite Geometrics and Designs.* *London Math. Soc. Lecture Note Ser.* **49** 337–353. Cambridge Univ. Press, Cambridge.

SERRE, J. P. (1977). *Linear Representations of Finite Groups.* Springer, New York.

SILVERBERG, A. (1980). Statistical models for q-permutations. Ph.D. dissertation, Dept. Statistics, Princeton Univ.

SPEED, T. (1987). What is an analysis of variance? (with discussion). *Ann. Statist.* **15** 885–941.

STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705.

TUKEY, J. (1961). Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics* **3** 191–219.

VERDUCCI, J. (1982). Discriminating between two probabilities on the basis of ranked preferences. Ph.D. dissertation, Dept. Statistics, Stanford Univ.

VERDUCCI, J. (1987). Probability models for permutations based on linear representations. Technical Report No. 359, Dept. Statistics, Ohio State Univ.

VERDUCCI, J. (1989). Minimum majorization decomposition of doubly stochastic matrices. In *Contributions to Probability and Statistics. Essays in Honor of Ingram Olkin* (L. J. Gleser, M. D. Perlman, S. J. Press and A. R. Sampson, eds.). Springer, New York.

YAGLOM, A. M. (1961). Second order homogeneous random fields. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **2** 593–622. Univ. California Press.

YLINEN, K. (1986). Random fields on noncommutative locally compact groups. *Lecture Notes in Math.* **1210** 365–386. Springer, Berlin.

DEPARTMENT OF MATHEMATICS
SCIENCE CENTER 325
HARVARD UNIVERSITY
ONE OXFORD STREET
CAMBRIDGE, MASSACHUSETTS 02138