GRAPHICAL TECHNIQUES FOR RANKED DATA

by

G.L. Thompson

Southern Methodist University
Technical Report No. SMU/DS/TR-249
Department of Statistical Science

November 1990

Graphical Techniques for Ranked Data

G. L. Thompson
Department of Statistical Science
Southern Methodist University
Dallas, Texas

Graphical techniques for presenting data on finite groups that do not have linear orderings are almost nonexistent. In particular, exploratory graphical methods are critically needed for data on the symmetric group on n elements, S_n, and related groups of cosets such as the alternating group, A_n. Because, unlike points on the real line, the elements of the symmetric group have several reasonable measures of distance which induce partial, but not linear, orderings, graphical methods such as histograms and bar graphs are not available. Graphical techniques related to histograms and bar graphs can be developed for ranked data to illustrate the probability density function by using permutation polytopes. A polytope is the convex hull of a finite set of points in $\mathbf{R}^n$, and a permutation polytope is the convex hull of the n! permutations of n elements when regarded as vectors in $\mathbf{R}^n$ (see, for example, Yemelichev et. al. (1984)). This concept is closely related to the observation by McCullagh (1990) that the n! elements of S_n lie on the surface of a sphere in $\mathbf{R}^{n-1}$.

To illustrate a rudimentary version of the proposed graphical technique, we wlll consider the paired ranking data of Critchlow and Verducci (1989) in which n=4. This data consists of pairs of orderings in which 38 students have ranked 4 styles of literary criticism in their order of preference, both before and after taking a course. The question of interest is whether the students' preferences have moved toward the teacher's preferred ordering <p,c,a,t>, i.e, toward the ordering in which style p is ranked first, style c is ranked second, style a is ranked third, and style t is ranked fourth. As illustrated by Figure 1 of McCullagh (1990), the 24 possible orderings of 4 items form the vertices of an Achimedian solid, the truncated octahedron, which is the permutational polytope for n=4. The edges connect orderings that differ by exactly one pairwise transposition so that Kendall's τ is the minimum number of edges that must be traversed to get from one ordering to another. The straight line distance between two points is proportional to Spearman's ρ. In the proposed graphical technique, the frequency with which each ordering appears in a data set is visually indicated, for example, by the size of a dot, at the corresponding vertex. The frequencies of the 38 pre-rankings are shown in Figure 1 and the 38 post-rankings are shown in Figure 2; the radius of the circle is proportional to the frequency.

insert Figures 1 and 2 here

Although the bivariate nature of the data is lost, valuable insight into this data can be obtained from the above figures. (Similarly, with paired univariate data it is not at all unusual, and often fruitful, to do exploratory data analysis by comparing the histograms of the "before" and "after" components.) The most obvious observation is that the frequencies do change a great deal between the two sets of rankings. First, there seems to be a notable increase in the frequencies of the vertices of the hexagon corresponding to the 6 orderings beginning with c. In fact, one might hypothesize that the orderings have moved toward <c,p,t,a> because almost half of the post-rankings lie either on <c,p,t,a> or on the three vertices within one edge (pairwise transposition) of <c,p,t,a>. (This is not inconsistent, however, with the movement toward <p,c,a,t> indicated by the results of Critchlow and Verducci's test statistic.) Also, there is a decrease in the frequencies of the 6 vertices corresponding to orderings that end with c, and we might hypothesize that c has a higher level of preference in the post-rankings than in the pre-rankings. Other observations include 1) style a is rarely chosen as either first or second choice after the course is completed; 2) the incidence of style t as a first choice decreases; and 3) there does not seem to be any movement toward <a,c,p,t>, the ordering of the styles on the second questionaire. Different rotations of Figures 1 and 2 would make some of these observations more apparent.

Figures 1 and 2 are fairly elementary examples of the potential of permutation polytopes in developing graphical techniques for ranked data. Significant improvements are immediately possible in the following three areas. First, the availabilty of information would be greatly enhanced if the truncated octohedrons could be arbitrarily rotated about any axis. This can be accomplished via interactive software, especially if the coordinates of the permutational polytope are known. For S_n the coordinates of the vertices of the permutation polytope in $\mathbf{R}^{n-1}$ with center at $\underset{\sim}{0}$ are found by using the Helmert transformation. Let $\underset{\sim}{\pi}=(\pi_1,\pi_2,\ldots,\pi_n)' \varepsilon S^n$ be any ranking (not ordering) of n items. This means that item i has ranking π_i. Let $\underset{\sim}{r}$ be the n dimensional column vector in which every element equals $(n+1)/2$, and let H be the Helmert transformation which maps the hyperplane $\sum_{i=1}^{n} x_i = 0$ onto the hyperplane $x_n = 0$. Note that H is orthogonal and preserves Euclidian distances. The coordinates of the vertices of the permutation polytope in $\mathbf{R}^{n-1}$ are $H(\underset{\sim}{\pi}-\underset{\sim}{r})$, $\underset{\sim}{\pi}\,\varepsilon S^n$. At the same time capabilities are introduced to rotate the polytopes, the frequencies associated with each vertex could be color coded on a scale chosen to highlight desired features. With color and rotational capabilities, it also might be worthwhile to experiment with illustrating the observed frequencies on the duals of the polytopes. For n=4, the dual of the truncated octohedron is the tetrakis hexahedron. It has 24 faces which, instead of the 24 vertices of the truncated octahedron, would correspond to the 24 possible orderings. See Cundy and Rollett (1951) for a discussion of the duals of Achimedian solids.

Second, these graphical techniques can be extended to the case where n is greater than 4 even though the dimension of the corresponding permutational polytope is greater than three. One approach to this is interactive software to successively view adjacent three dimensional faces. Possibly, the faces of greatest interest are those generated by the set of orderings in which four items are permuted while the remaining n-4 items remain fixed. Characterizations of all of the faces of permutational polytopes are given by Yemelichev et. al. (1984). The application of other new graphical methods of viewing higher dimensions may be very useful.

Third, the above arguments can be extended to partially ranked data by considering the multiset $M=(1^{a_1}, 2^{a_2}, \ldots, k^{a_k})$. This multiset has k levels and a_i items are ranked together in the ith level. It is assumed that $a_i \varepsilon \mathbf{Z}^+$ and $\sum_{i=1}^{k} a_i = n$. Stanley (1986) discusses multisets. The set of permutations of M, when regarded as vectors in $\mathbf{R}^n$, become the vertices of an integral polytope. These vertices can be shown to lie on an $n-1$ dimensional sphere in $\mathbf{R}^n$. Just as with fully ranked data, software to rotate the vertices and techniques to view the three dimensional faces for $n>4$ are needed.

The characterizations of the faces of integral polytopes formed by permutations of multisets is a straightforward generalization of the resullts in Yemelichev et. al. (1984) for permutation polytopes. These integral polytopes can also be thought of as Cayley diagrams (see Coxeter (1980)) and have interesting connections with the generators of the corresponding group $S_n/\Pi S_{a_i}$. Furthermore, they induce an interesting extension of Kendall's τ to partially ranked data that has properties quite different from both the Haussdorf metric (Critchlow (1985)) and the metric $i(\pi)$ discussed by Diaconis (1988). Similarly, the straight line distance between vertices in $\mathbb{R}^n$ is an extension of Spearman's ρ .

REFERENCES

Coxeter, H.S.M. and Moser, W. O. (1980), *Generators and Relations for Discrete Groups* (4th ed.), Berlin: Springer-Verlag.

Critchlow, D. E. (1985), *Metric Methods for Analyzing Partially Ranked Data*, Berlin: Springer-Verlag.

Critchlow, D. E. and Verducci, J. S. (1989), "Detecting a Trend in Paired Rankings," Technical Report #418, Department of Statistics, The Ohio State University.

Cundy,H. M. and Rollett, A. P. (1951), *Mathematical Models*, London: Oxford University Press.

Diaconis, P. (1988), *Group Representations in Probability and Statistics*, Hayward, California: Institute of Mathematical Statistics.

McCullagh, P. (1990) "Models on Spheres and Models on Permutations," Technical Report, Department of Statistics, University of Chicago.

Stanley, R. (1986), *Enumerative Combinatorics*, Monterey, California: Wadsworth & Brooks/Cole.

Yemelichev, V. A., Kovalev, M.M., and Kravtsov, M.K. (1984), *Polytopes, Graphs and Optimisation*, Cambridge: Cambridge University Press.

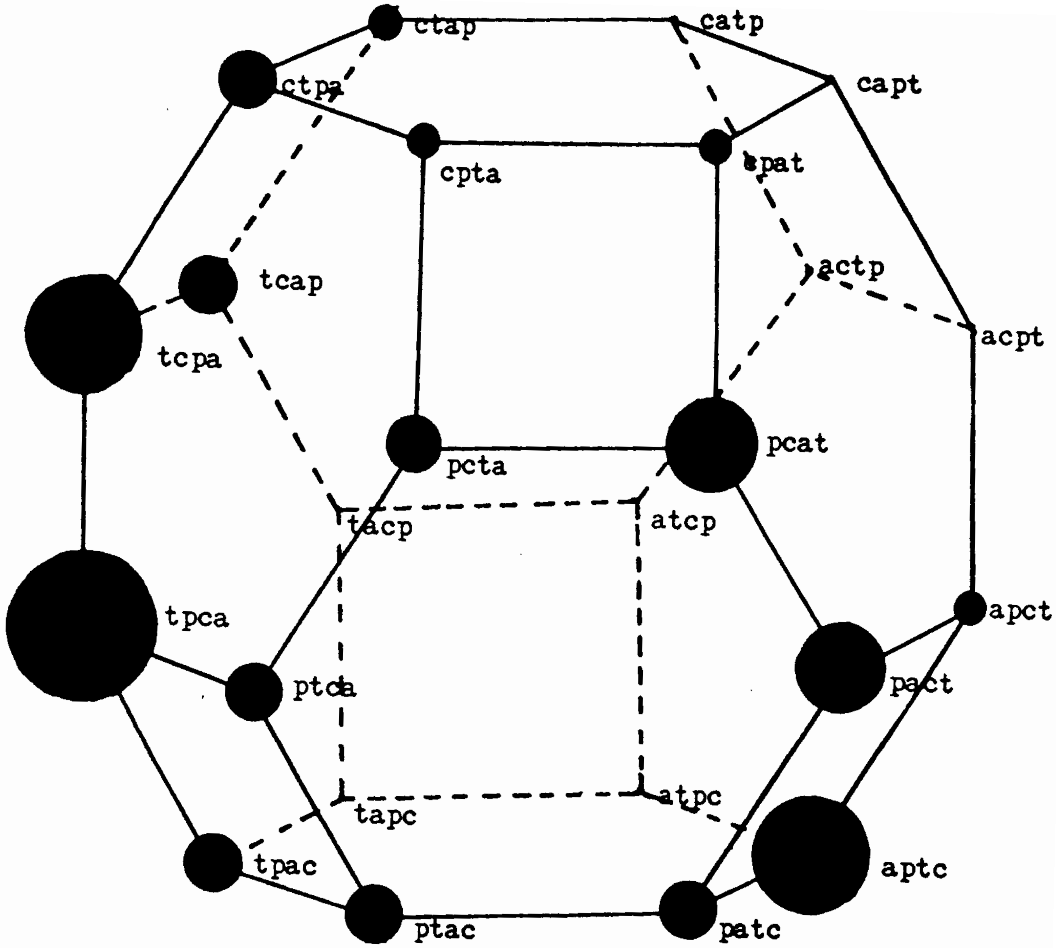
ctap
catp
ctpa
capt
cpta
cpat
tcap
actp
acpt
tcpa
pcat
pcta
tacp
atcp
apct
tpca
ptca
pact
atpc
tapc
tpac
aptc
ptac
patc

Figure 1

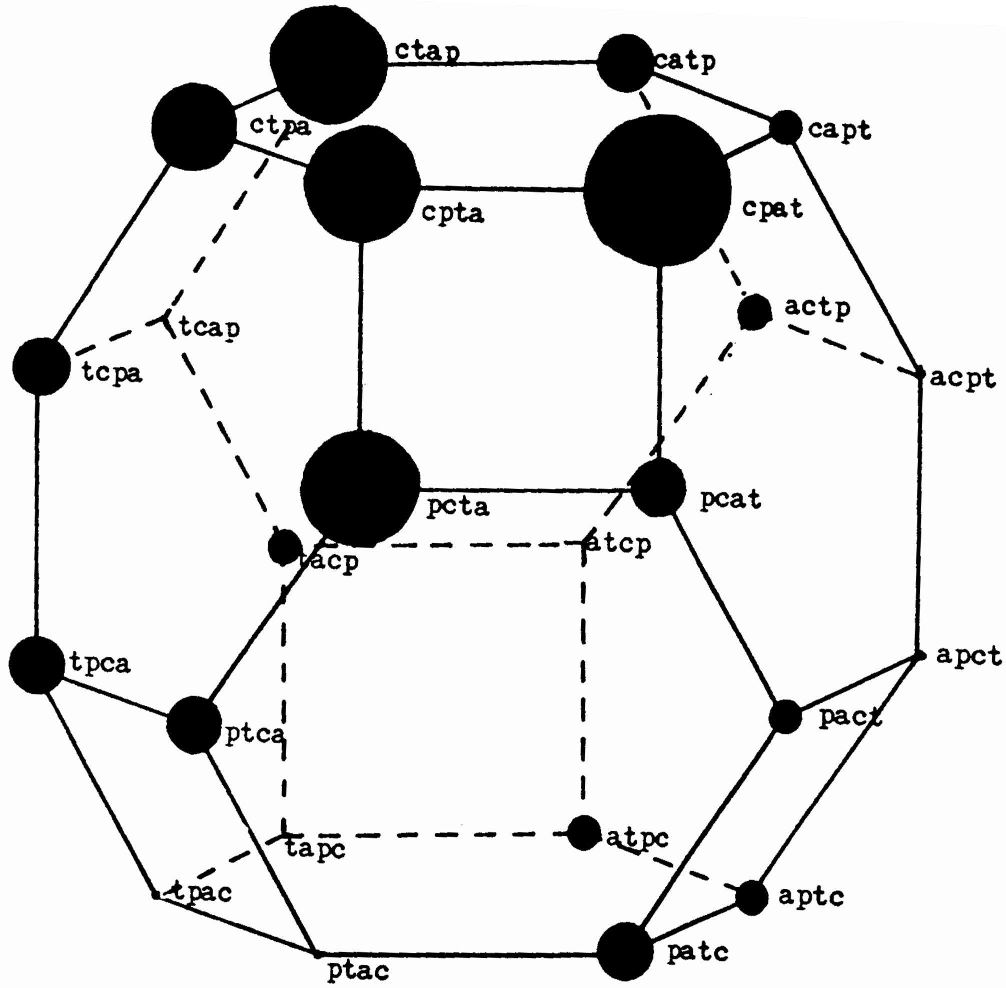
ctap
catp
ctpa
capt
cpta
cpat
actp
tcap
tcpa
acpt
pcta
pcat
atcp
tacp
tpca
apct
pact
ptca
atpc
tapc
aptc
tpac
patc
ptac

Figure 2