The Annals of Statistics
1997, Vol. 25, No. 2, 478–494

MULTIDIMENSIONAL MEDIANS ARISING FROM GEODESICS ON GRAPHS[1]

BY CHRISTOPHER G. SMALL

University of Waterloo

In this paper we introduce a depth function for distributions on graphs that is analogous to recent multivariate definitions. Using the property of geodesic convexity on graphs, a median-like center for distributions on graphs is constructed and applied to ranking data as well as multivariate data spanned by the minimal spanning tree.

1. Introduction. Recent work on the concept and theory of multidimensional medians has produced a proliferation of different constructions. In dimension 1, the median is characterized by a variety of equivalent definitions. However, in higher dimensions, these definitions cease to be equivalent, and produce different multidimensional medians which can all claim, to a greater or lesser degree, to be the natural extension of the one-dimensional definition. For example, in dimension 1, the median is that point about which the expected absolute deviation is minimized. In dimension 2 and higher, the point from which the expected distance to a random point of the distribution is smallest can be called the spatial median. The sample analog of this was applied by Brown (1983), who derived its first-order asymptotics. A variation of this definition by Oja (1983) and Oja and Niinimaa (1985) replaced the expected absolute deviation with the expected volume of a simplex. The similarity between the two definitions belies the differences in their properties, the former having 50 percent breakdown, and the latter 0 percent. See Niinimaa, Oja and Tableman (1990). The characterization of the median in dimension 1 as the deepest point in the distribution was generalized by Tukey (1975) to what is usually now called the Tukey depth median. See Donoho and Gasko (1992) for a study of its properties. The reader can find a survey of work up to 1990 in Small (1990).

Multidimensional medians have also been applied to directional statistics. The extension to directional data is a natural one because directional data on the sphere $\mathbf{S}^2$, say, can be considered as data in $\mathbf{R}^3$ having the appropriate constraint of lying at a unit distance from the origin. As a directional median is constrained to lie on a unit sphere, directional medians are distinct from their multidimensional counterparts. Nevertheless, each multidimensional median for $\mathbf{R}^3$ typically has an analog on the unit sphere $\mathbf{S}^2$. See

Received April 1994; revised December 1995.

[1]Research supported by an operating grant from the Natural Sciences and Engineering Council of Canada.

AMS 1991 *subject classifications*. 62H99, 60D05.

Key words and phrases. Median, graph, convex set, permutation, minimal spanning tree.

Fisher (1985), Ducharme and Milasevic (1987) and Liu and Singh (1992) for examples and theory of these.

In this paper we shall introduce a family of medians that are applicable to distributions and data on graphs. We shall call these the *geodesic convexity medians* or GCM's, for short. As many multivariate data sets can be regarded as generating a graph such as the minimal spanning tree, this median will be applicable to data sets in $\mathbf{R}^p$. The proposed median will also be applicable to data and distributions which lie naturally on a graph. As a special case of this, we propose *permutation medians* for the analysis of rank or permutation data. The graph median will be seen to be an extension of the univariate median, which will arise as the linear graph associated with the ordering of the data.

There are several reasons for constructing medians on graphs. In dimension 1, the most obvious characterization of the median is as the middle order statistic of the data. However, it is this most obvious interpretation which is most difficult to generalize; by contrast, absolute deviation arguments generalize immediately. By interpreting the median as the middle order statistic, we assume that distance relationships among points are only relevant to the extent that they characterize the ordering of the data. The median is thereby equivariant under monotone transformations which can distort distances considerably. Ordering of data values is a topological property that is akin to the more general property of the connectivity of a graph. Seen in this context, a linear ordering of points becomes a special case of a graph in which neighboring order statistics are joined by an edge. In some cases, such as permutation models considered below, distance values are as naturally measured through geodesic distance on a graph as through Euclidean distance. For example, as McCullagh (1993) has noted, Spearman's rank correlation is associated with Euclidean distance on a permutation polytope, whereas Kendall's rank correlation is naturally associated with the geodesic distance measure on the graph whose vertices and edges are the respective vertices and edges of the polytope. The question of which measure of distance is more natural is analogous to the choice between Spearman's and Kendall's coefficient. A second reason for considering medians on graphs is that geodesic distance on a finite graph is bounded. This implies a natural robustness of medians and depth measures based on geodesic distance. For example, we shall note that medians constructed for minimal spanning trees have 50 percent breakdown points.

2. Definitions and examples. By an (*undirected*) *graph* $\mathscr{G}$ we shall understand a finite collection of *vertices* $\mathscr{X} = \{x_i : i = 1, 2, \ldots, n\}$ together with a collection of *edges* $\mathscr{E} = \{(x_i, x_j) : 1 \le i \le j \le n\} \subset \mathscr{X} \times \mathscr{X}$. A sequence of vertices $x_{i_1}, x_{i_2}, \ldots, x_{i_j}$ is said to be a *path* of length j if each successive pair $x_{i_k}, x_{i_{k+1}}$ is joined by an edge from $\mathscr{E}$. In particular, if $i_1 = i_j$ and the path contains at least one edge, then the path is said to be a *cycle*. A graph is said to be a *tree* if it has no cycles. A graph is said to be *connected* if any two distinct vertices can be joined by a path. *Throughout this paper, all graphs will be connected.*

DEFINITION 2.1. A sequence of vertices $x_a = x_{i_1}, \ldots, x_{i_j} = x_b$ is said to be a *geodesic path* from x_a to x_b if it is of minimum length among all paths from x_a to x_b. We say that the length j of such a geodesic path is the *geodesic distance* from x_a to x_b, and we write $d(x_a, x_b) = j$. In particular, $d(x_a, x_b) = 0$ if and only if $x_a = x_b$. A set of vertices $\mathscr{A} \subset \mathscr{X}$ is said to be *geodesically convex* or *G-convex* if, for each x_a, $x_b \in \mathscr{A}$, the vertices of every geodesic path joining x_a to x_b are also in $\mathscr{A}$.

Note that the empty set, singleton sets of vertices and the set $\mathscr{X}$ are all G-convex. See Jamison-Waldner (1982) for more on the theory of geodesic convexity on graphs.

By a distribution μ on $\mathscr{G}$ we shall mean an assignment of nonnegative weights p_i to each $x_i \in \mathscr{X}$ such that $\sum p_i = 1$. A random vertex from $\mathscr{X}$ chosen with some probability distribution will typically be written as X.

We shall now introduce the GCM through a measure of depth that is analogous to Tukey's depth measure in the standard Euclidean setting. First, however, we need to introduce the concept of a half set on a graph. A set $\mathscr{A} \subset \mathscr{X}$ is said to be a *half set* if $\mathscr{A}$ and its complement $\mathscr{A}^c = \{x \in \mathscr{X}: x \notin \mathscr{A}\}$ are both G-convex. Note that the set $\mathscr{X}$ is a half set because the empty set is always G-convex. The terminology "half set" is apparently new in this graph-theoretic context. However, it is well known for the theory of convex sets in Euclidean spaces, where geodesic convexity on a graph is replaced with the standard convexity of $\mathbf{R}^p$. Application of this definition to usual convexity in Euclidean space leads to the class of half spaces. See Benson (1966) for a number of results on convex sets and half spaces in $\mathbf{R}^p$.

Of particular interest are those graphs satisfying the following separation condition.

DEFINITION 2.2. A graph $\mathscr{G}$ with vertex set $\mathscr{X}$ is said to satisfy the *half set separation condition* if, for all vertices $x \in \mathscr{X}$ and for all G-convex sets $\mathscr{B} \subset \mathscr{X}$ for which $x \notin \mathscr{B}$, there exists a half set $\mathscr{A} \subset \mathscr{X}$ such that $x \notin \mathscr{A}$ and $\mathscr{B} \subset \mathscr{A}$.

A property that is equivalent to the half set separation condition is that any G-convex set can be written as an intersection of half sets. (The set $\mathscr{X}$ is the empty intersection.) This is a well-known property of convexity in Euclidean space. Here, we adapt this property for use within graphs. *Henceforth, all graphs that we shall consider will be assumed to satisfy the half set separation condition.* Typically, to verify the condition, it is necessary to characterize the half sets of the graph and check the condition directly.

To illustrate the concept of G-convexity, consider the graph on the univariate order statistics $x_{(1)}, \ldots, x_{(n)}$ constructed by joining all neighboring order statistics $x_{(i)}$, $x_{(i+1)}$ by an edge. A subset of these statistics is G-convex provided it is of the form $\{x_{(i)}, x_{(i+1)}, \ldots, x_{(i+k)}\}$. It is a half set provided $i = 1$ or $i + k = n$. As we can write $\{x_{(i)}, \ldots, x_{(i+k)}\}$ as the intersection of $\{x_{(1)}, \ldots, x_{(i+k)}\}$ and $\{x_{(i)}, \ldots, x_{(n)}\}$, it follows that this graph satisfies the half set separation property.

DEFINITION 2.3. Let $\mathscr{G}$ be a graph with vertex set $\mathscr{X}$ and let μ be a distribution on $\mathscr{G}$. We define the *geodesic convexity depth* or GC-depth of a vertex x of $\mathscr{G}$ to be the minimum value of $\mu(\mathscr{A})$, where $\mathscr{A}$ ranges over the class of all half sets of $\mathscr{G}$ such that $x \in \mathscr{A}$. We shall write the GC-depth of x as $\mathrm{GCD}(x, \mu)$.

We now have the following definition.

DEFINITION 2.4. We shall say that $x \in \mathscr{X}$ is a *geodesic convexity median* (GC-median, or GCM) for μ if $\mathrm{GCD}(y, \mu)$ is maximized over $y \in \mathscr{X}$ at $y = x$. Let $\mathrm{GCM}(\mu)$ be the set of all such GC-medians for μ.

The following proposition shows that the GC-depth function is concave.

PROPOSITION 2.5. *The set of all $x \in \mathscr{X}$ such that $GCD(x, \mu) \geq p$ for real p is a convex set.*

PROOF. The proof of this proposition uses the property that

$$\{x \in \mathscr{X} : \mathrm{GCD}(x, \mu) \geq p\} = \bigcap \{\mathscr{A} : \mu(\mathscr{A}) > 1 - p\},$$

where $\mathscr{A}$ denotes a half set. To prove this identity, consider a vertex x which is an element of the left-hand side. Then $\mathrm{GCD}(x, \mu) \geq p$. This will be true if and only if $x \in \mathscr{A}$ implies that $\mu(\mathscr{A}) \geq p$. Equivalently, we can write that $\mu(\mathscr{A}) < p$ implies that $x \in \mathscr{A}^c$. In turn, this is the case if and only if $x \in \cap\{\mathscr{A}^c : \mu(\mathscr{A}^c) > 1 - p\}$. As the class of half sets is closed under complementation, this set is simply the right-hand side of the identity. Proposition 2.5 now follows by noting that the intersection of half sets is G-convex. □

So, in particular, the set $\mathrm{GCM}(\mu)$ is a G-convex set. Proposition 2.5 has particular consequences for the search for the GCM on a graph. As every G-convex set is connected, it follows that any local maximum of the function $\mathrm{GCD}(\cdot, \mu)$ must also be a global maximum. That is, if $\mathrm{GCD}(y, \mu) < \mathrm{GCD}(x, \mu)$ for every y that shares an edge with x, then x is the unique GC-median of μ. Thus a basic but reliable way to find the GCM of μ when the GCD function is strictly convex proceeds as follows. First calculate $\mathrm{GCD}(x, \mu)$ for some initial starting point $x \in \mathscr{X}$ and compare this value with the GC-depth of neighboring vertices. For convenience, let us write the set of such neighbors y of x as $\mathscr{N}(x)$. In fact, we can do slightly better than searching all of $\mathscr{N}(x)$. In calculating $\mathrm{GCD}(x, \mu)$ we have found a half set $\mathscr{A}$ such that $x \in \mathscr{A}$ and $\mathrm{GCD}(x, \mu) = \mu(\mathscr{A})$. It follows that any neighbor y of x whose GC-depth is greater than x must be an element of $\mathscr{A}^c$. Thus we can restrict the search for a vertex of greater GC-depth to $\mathscr{N}(x) \cap \mathscr{A}^c$.

We close this section by considering the analysis of paired data on graphs. Paired data of the form (x_r, y_s) can be regarded as an element of the Cartesian product $\mathscr{X}^2 = \mathscr{X} \times \mathscr{X}$, or more generally $\mathscr{X} \times \mathscr{Y}$. By a *paired GCM* we mean a GC-median for distributions on the sets of vertices of $\mathscr{X} \times \mathscr{Y}$. To construct this,

some appropriate set of edges should be imposed on the Cartesian product to make $\mathscr{X} \times \mathscr{Y}$ the set of vertices of a graph. Many choices are available here. A simple choice is to join vertices so as to make the geodesic distance on $\mathscr{X} \times \mathscr{Y}$ into a "Manhattan metric." (Of course, this should not be given priority over what is statistically sensible.) To construct a graph whose geodesic distance is the Manhattan metric, we join vertex (x_r, y_s) to (x_t, y_u) if $x_r = x_t$ and y_s is a neighbor of y_u in $\mathscr{Y}$ or if $y_s = y_u$ and x_r is a neighbor of x_t in $\mathscr{X}$. The resulting G-convex subsets of $\mathscr{X} \times \mathscr{Y}$ will be Cartesian products $\mathscr{A} \times \mathscr{B}$, where $\mathscr{A}$ is a G-convex subset of $\mathscr{X}$ and $\mathscr{B}$ is a G-convex subset of $\mathscr{Y}$. Suppose we have some distribution μ on the vertices of $\mathscr{X} \times \mathscr{Y}$. It can be seen that the GC-depth of a point $(x, y) \in \mathscr{X} \times \mathscr{Y}$ is given by

$$\mathrm{GCD}\,[(x, y), \mu] = \min\,[\mathrm{GCD}(x, \mu_1), \mathrm{GCD}(y, \mu_2)]\,,$$

where μ_1 is the marginal distribution on $\mathscr{X}$ and μ_2 is the marginal distribution on $\mathscr{Y}$. Thus the analysis of the depth of paired data can be reduced to consideration of the marginal distributions of the vertices. While this makes the GC-depth function easy to compute, it also shows the limitation of the Manhattan metric on $\mathscr{X} \times \mathscr{Y}$. Most multivariate data sets cannot be fully analyzed by an examination of their marginal distributions.

We now consider some examples of medians on graphs.

3. Permutation medians and depth ranking methods. In this section we consider data and distributions on the set of permutations $(i_1, i_2, \ldots, i_n)$ of the integers $1, 2, \ldots, n$. Permutation data typically arise in statistics from studies in which different individuals are asked to rank a set of n items, which are labeled from 1 up to n for convenience. For example, if m individuals are each asked to rank three types of coffee according to preferences, then the results of this study can be recorded as a data set of m elements on the set of permutations of the integers 1, 2 and 3. Note that we shall only consider complete rankings of the items here. Two notations are standard. Suppose, for example, that an individual is asked to rank the quality of three brands of coffee, the brands having labels 1, 2 and 3. The notation $\langle 3, 1, 2\rangle$ refers to *ordering*, and denotes the fact that the individual ranks brand 3 first, brand 1 second and brand 2 third. The notation $(3, 1, 2)$ refers to *ranking*, and denotes that brand 1 is ranked third, brand 2 is ranked first and brand 3 is ranked second. We shall use both notations, with an emphasis on the rank notation.

Let $\mathscr{X}_n$ be the set of all rankings $(i_1, i_2, \ldots, i_n)$ of the first n positive integers. A large number of metrics are available for $\mathscr{X}_n$ including the Hamming distance, Ulam's distance and Cayley's distance. For information about these and related measures, the reader is referred to Diaconis (1988). Closely related to the family of metrics on $\mathscr{X}_n$ are the measures of association, of which Spearman's rho and Kendall's tau are the best known. Of particular interest here is the metric induced by *Kendall's tau*. The distance between two rankings $(i_1, i_2, \ldots, i_n)$ and $(j_1, j_2, \ldots, j_n)$ can be defined as the number of pairs (r, s), $1 \le r, s \le n$, such that $i_r < i_s$ and $j_r > j_s$. Kendall's tau, as a

measure of association between rankings, is an affine transformation of this metric. We will say that a pair (r, s) of integers between 1 and n is *discordant between two rankings* if this condition is satisfied, and *concordant between the two rankings* otherwise. Thus Kendall's tau metric counts the number of discordant pairs between two permutations.

To interpret this metric in terms of graph theory, we note that the distance between $(i_1, i_2, \ldots, i_n)$ and $(j_1, j_2, \ldots, j_n)$ is the minimum number of transpositions of adjacent ranks necessary to permute $(i_1, i_2, \ldots, i_n)$ into $(j_1, j_2, \ldots, j_n)$. In this form, the metric is recognizable as a geodesic distance on a graph whose vertices are the permutations of the integers up to n. Two permutations are connected by an edge if their rankings differ by a single interchange of adjacent ranks. For example, the orderings $\langle 3, 1, 6, 2, 4, 5\rangle$ and $\langle 3, 6, 1, 2, 4, 5\rangle$ are adjacent. In ranking notation, we can write these equivalently as (2, 4, 1, 5, 6, 3) and (3, 4, 1, 5, 6, 2), respectively. This graph, for which Kendall's tau is the geodesic distance, also has a representation using the *permutation polytope*. See Yemelichev, Kovalev and Kravtsov (1984) and Thompson (1993). The permutation polytope is defined as the convex hull in $\mathbf{R}^n$ of the set of all rankings $(i_1, i_2, \ldots, i_n)$ considered as elements of $\mathbf{R}^n$. Any two vertices of the permutation polytope will be joined by an edge if and only if they differ by a single adjacent transposition. Thus Kendall's tau metric is the shortest number of steps required, moving via neighboring vertices connected by edges of the polytope. See also Mallows (1957) and McCullagh and Ye (1993).

As the geodesic distance between two rankings can be computed from Kendall's tau statistic without direct reference to any geodesic path between them, there is any easy test to check whether a third ranking lies on a geodesic path between any two permutations. A ranking $(k_1, k_2, \ldots, k_n)$ will lie on some geodesic path from $(i_1, i_2, \ldots, i_n)$ to $(j_1, j_2, \ldots, j_n)$ provided the geodesic distance from $(i_1, i_2, \ldots, i_n)$ to $(j_1, j_2, \ldots, j_n)$ is the sum of the geodesic distances from $(i_1, i_2, \ldots, i_n)$ to $(k_1, k_2, \ldots, k_n)$ and $(k_1, k_2, \ldots, k_n)$ to $(j_1, j_2, \ldots, j_n)$. This will be true provided $(k_1, k_2, \ldots, k_n)$ shares all concordant pairs between $(i_1, i_2, \ldots, i_n)$ and $(j_1, j_2, \ldots, j_n)$. More precisely, if r and s are integers such that $i_r < i_s$ and $j_r < j_s$, then it follows that $k_r < k_s$. We can also define the concept of concordance between points and sets.

DEFINITION 3.1. Let $(k_1, k_2, \ldots, k_n)$ be any ranking and let $\mathscr{A}$ be any set of rankings. We shall say that $(k_1, k_2, \ldots, k_n)$ is *concordant* with $\mathscr{A}$ provided that, for all $1 \le r, s \le n$, statement (A) implies statement (B) below:

(A) $i_r < i_s$ for all $(i_1, i_2, \ldots, i_n) \in \mathscr{A}$;
(B) $k_r < k_s$.

Note that this property can be vacuously satisfied if there exist no r and s satisfying statement (A). For example, every ranking is concordant with $\mathscr{X}_n$. On the other hand, no ranking is concordant with the empty set $\varnothing$. Note also that $(k_1, k_2, \ldots, k_n)$ lies on a geodesic path from $(i_1, i_2, \ldots, i_n)$ to

$(j_1, j_2, \ldots, j_n)$ provided that $(k_1, \ldots, k_n)$ is concordant with the set

$$\{(i_1, i_2, \ldots, i_n), (j_1, j_2, \ldots, j_n)\}.$$

Using Definition 3.1, we can obtain the following property of G-convex sets.

PROPOSITION 3.2. *Let $\mathscr{A}$ be any G-convex set of rankings. Then any ranking which is concordant with $\mathscr{A}$ is an element of $\mathscr{A}$.*

PROOF. Suppose that $(i_1, \ldots, i_n) \notin \mathscr{A}$, and that $(i_1, \ldots, i_n)$ is concordant with $\mathscr{A}$. We will prove a contradiction. Let $(j_1, \ldots, j_n)$ be any ranking in $\mathscr{A}$ whose geodesic distance from $(i_1, \ldots, i_n)$ is minimum. Furthermore, let

$$(j_1, \ldots, j_n) \to (k_1, \ldots, k_n) \to \cdots \to (i_1, \ldots, i_n)$$

be a geodesic path from $(j_1, \ldots, j_n)$ to $(i_1, \ldots, i_n)$, where $(k_1, \ldots, k_n)$ differs from $(j_1, \ldots, j_n)$ by a single interchange of adjacent ranks. As $(j_1, \ldots, j_n)$ is the closest ranking in $\mathscr{A}$ to $(i_1, \ldots, i_n)$, it follows that $(k_1, \ldots, k_n) \notin \mathscr{A}$. The contradiction will be established if we can show that $(k_1, \ldots, k_n) \in \mathscr{A}$.

Now there exist integers $1 \le r, s \le n$ such that $j_r < j_s$, while $k_r > k_s$ and $i_r > i_s$. It is at this point that we use the fact that $(i_1, \ldots, i_n)$ is concordant with $\mathscr{A}$. This fact implies that there exists a ranking $(m_1, \ldots, m_n) \in \mathscr{A}$ such that $m_r > m_s$. [If not, we would have verified that $(i_1, \ldots, i_n)$ is not concordant with $\mathscr{A}$.] It can be checked that there exists a geodesic path

$$(j_1, \ldots, j_n) \to (k_1, \ldots, k_n) \to \cdots \to (m_1, \ldots, m_n).$$

As this is a geodesic whose endpoints lie in $\mathscr{A}$, the G-convexity of $\mathscr{A}$ implies that the rankings along this geodesic must lie in $\mathscr{A}$. In particular, $(k_1, \ldots, k_n) \in \mathscr{A}$ and the contradiction is established. □

Let us define subsets

$$\mathscr{A}(r, s) = \{(i_1, i_2, \ldots, i_n) : i_r < i_s\}.$$

Using Proposition 3.2, we can prove the following.

PROPOSITION 3.3. *The family of half sets consists of the empty set $\varnothing$, the set $\mathscr{X}_n$ and sets of the form $\mathscr{A}(r, s)$. Furthermore, the half set separation condition is satisfied.*

PROOF. It is immediate that $\varnothing$ and $\mathscr{X}_n$ are half sets. In addition, it is routine to show that sets of the form $\mathscr{A}(r, s)$ are G-convex. As this class is closed under complementation, these sets are seen to be half sets. So it suffices to show that there are no others.

Suppose $\mathscr{A}$ is a half set that is not of the form $\mathscr{A}(r, s)$. Then either $\mathscr{A}$ is not a subset of any set of the form $\mathscr{A}(r, s)$ or $\mathscr{A}^c$ is not a subset of any $\mathscr{A}(r, s)$. Without loss of generality, we can suppose that $\mathscr{A}$ is not a subset of any $\mathscr{A}(s, r)$. If this is the case, then condition (A) is vacuously satisfied for every

ranking $(k_1, \ldots, k_n)$. Moreover, since $\mathscr{A}$ is G-convex, it follows by Proposition 3.2 that every ranking $(k_1, \ldots, k_n)$ is an element of $\mathscr{A}$. So $\mathscr{A} = \mathscr{X}_n$, and $\mathscr{A}^c = \emptyset$.

To check that the half set separation condition is satisfied, we consider any G-convex set $\mathscr{A}$ and any ranking $(k_1, \ldots, k_n)$ which is not an element of $\mathscr{A}$. By Proposition 3.2, there must exist integers $1 \leq r, s \leq n$ such that $k_r > k_s$ and such that $i_r < i_s$ for all rankings $(i_1, \ldots, i_n)$ in $\mathscr{A}$. From this fact, we see that $(k_1, \ldots, k_n)$ is not an element of $\mathscr{A}(r, s)$ and that $\mathscr{A} \subset \mathscr{A}(r, s)$. This completes the proof. □

EXAMPLE 3.4. We consider the example of the literary criticism data set considered by Critchlow and Verducci (1992). This has also been studied by Thompson (1993) and McCullagh and Ye (1993). A group of 38 students in a course were asked to rank four types of literary criticism of a Faulkner novel in order of preference. These types were authorial (A), comparative (C), personal (P) and textual (T) (see Table 1).

This data set consists of paired values for each student, and therefore is naturally represented as an empirical distribution on $\mathscr{X}_4 \times \mathscr{X}_4$. However, as we noted in the previous section, the GC-depth on such a Cartesian product with

TABLE 1

Orderings	**Count**	**GC-depth**	**Count**	**GC-depth**
⟨A, C, P, T⟩	0	0.236842	0	0.184211
⟨A, C, T, P⟩	0	0.236842	1	0.184211
⟨A, T, C, P⟩	0	0.236842	0	0.184211
⟨A, T, P, C⟩	0	0.236842	1	0.184211
⟨A, P, T, C⟩	4	0.236842	1	0.184211
⟨A, P, C, T⟩	1	0.236842	0	0.184211
⟨C, A, P, T⟩	0	0.236842	1	0.289474
⟨C, A, T, P⟩	0	0.236842	2	0.289474
⟨C, T, P, A⟩	2	0.315789	3	0.421053
⟨C, T, A, P⟩	1	0.236842	4	0.289474
⟨C, P, A, T⟩	1	0.315789	5	0.578947
⟨C, P, T, A⟩	1	0.315789	4	0.578947
⟨P, C, A, T⟩	3	0.368421	2	0.394737
⟨P, C, T, A⟩	2	0.368421	4	0.394737
⟨P, T, A, C⟩	2	0.394737	0	0.184211
⟨P, T, C, A⟩	2	0.552632	2	0.289474
⟨P, A, T, C⟩	2	0.368421	2	0.184211
⟨P, A, C, T⟩	3	0.368421	1	0.184211
⟨T, A, C, P⟩	1	0.236842	1	0.184211
⟨T, A, P, C⟩	0	0.236842	0	0.184211
⟨T, C, P, A⟩	4	0.315789	2	0.289474
⟨T, C, A, P⟩	2	0.236842	0	0.289474
⟨T, P, A, C⟩	2	0.394737	0	0.184211
⟨T, P, C, A⟩	5	0.447368	2	0.289474

the Manhattan metric is a function of the marginal empirical distributions: in this case, the pre-course and post-course results. The pre-course orderings were compared with the post-course orderings.

The 24 orderings of the four types of criticism are shown in the first column, with the empirical counts before and after the course shown in columns 2 and 4, respectively. Column 3 shows the GC-depth function based on the pre-course empirical distribution found by dividing the entries of column 2 by 38. Finally, column 5 shows the corresponding GC-depth function for the post-course empirical distribution. To illustrate the GC-depth calculations, let us consider the ordering $\langle$P, T, C, A$\rangle$. We consider all half spaces which contain this ordering. Such half spaces correspond to orderings of pairs of criticism types which are compatible with $\langle$P, T, C, A$\rangle$. These are $\langle$P, T$\rangle$, $\langle$P, C$\rangle$, $\langle$P, A$\rangle$, $\langle$T, C$\rangle$, $\langle$T, A$\rangle$ and $\langle$C, A$\rangle$. In the pre-course empirical distribution, these have 21, 26, 29, 24, 24 and 23 students, respectively. The minimum of these is 21. Therefore, the GC-depth of the pre-course empirical distribution at $\langle$P, T, C, A$\rangle$ is $21/38 = 0.552632$, approximately.

As there are 38 observations among 24 categories, the empirical distribution is quite noisy. By contrast, the GC-depth takes on only a few values, and these are distributed so that neighboring permutations have similar depths. A simple interpretation can be drawn from the pre-course and post-course GC-depth functions. If we consider the four orderings whose pre-course depth is the highest, we see that these are the four which place textual (T) and personal (P) criticism above comparative (C) and authorial (A) criticism. Note also that the four orderings which place comparative and authorial criticism above textual and personal criticism are uniformly given the minimum GC-depth. If we compare this with the post-course GC-depth function in column 5, we see that the comparative style of criticism has been given a more favorable ranking than the pre-course GC-depth suggests. The three highest post-course GC-depth values place the comparative method first and the next two highest depth values place the comparative method second. These five orderings rank the authorial style third or fourth. This is supported by considering the minimum of the GC-depth as well, where orderings ranking the authorial style first have uniformly minimum depth.

This analysis suggests that the students moved from a preference for textual–personal criticism at the beginning of the course toward a preference for comparative criticism at the end of the course. This is compatible, although not identical, with the interpretation given by Thompson (1993). We must be careful not to read too much into the analysis using the GC-depth. The GC-depth serves as an exploratory tool, and cannot be used to make probability statements about full rankings. However, statements about pairs can be made. For example, the fact that the pre-course GC-depth of $\langle$P, T, C, A$\rangle$ is greater than 0.5 means that in the pre-course preferences, for any of the six paired orderings that are compatible with $\langle$P, T, C, A$\rangle$, the majority of students implicitly chose an ordering compatible with that pair. From this we cannot conclude anything directly about preferences for $\langle$P, T, C$\rangle$, say, or $\langle$P, C, A$\rangle$. To investigate these, we must return to the data.

It should be noted that the GCM's for both the pre-course and post-course empirical distributions are attained where the GC-depth is greater than 0.5. We now consider a class of models for permutation distributions which have this property. Two general classes of models have proved useful in analyzing permutation data. These are the Thurstone order statistics models and the Babington Smith models whose special cases have been developed by Mallows (1957). Of the two types, the Thurstone order statistics models are particularly useful for studying GC-depth functions, as paired comparisons in these models are straightforward. The general form of the model is as follows. Let F_1, $F_2, \ldots, F_n$ be n continuous cumulative distribution functions. We generate n independent random variables such that the ith variable has distribution F_i. Let i_k be the rank of the kth random variable. Then $(i_1, i_2, \ldots, i_n)$ is a random permutation of $(1, 2, \ldots, n)$, inducing a corresponding distribution $\mu = \mu(F_1, F_2, \ldots, F_n)$ on the set $\mathscr{X}_n$ of all permutations of the first n positive integers. We now have the following.

PROPOSITION 3.5. *Suppose that $F_1, F_2, \ldots, F_n$ are stochastically ordered in the sense that $F_i(t) > F_j(t)$ for all $-\infty < t < \infty$ whenever $i < j$. Then* GCM$(\mu) = (1, 2, \ldots, n)$. *Moreover, on this permutation the* GC*-depth is greater than or equal to* 0.5, *and is the only permutation for which this is so.*

PROOF. For $r > s$ the half sets $\mathscr{A}(r, s)$ have probability

$$\mu(\mathscr{A}_{rs}) = \int_{-\infty}^{+\infty} F_r(\xi)\, dF_s(\xi)$$
$$< \int_{-\infty}^{+\infty} F_s(\xi) dF_s(\xi) = \tfrac{1}{2}.$$

It follows easily from this that the identity permutation $(1, 2, \ldots, n)$ has GC-depth greater than one-half, in fact given by

$$1 - \max_{r>s} \mu\left(\mathscr{A}(r, s)\right).$$

For any other permutation $(i_1, i_2, \ldots, i_n)$ there must exist a pair $r > s$ such that $i_r < i_s$. For this case the probability of $\mathscr{A}(s, r)$ is greater than one-half, and the GC-depth of $(i_1, i_2, \ldots, i_n)$ is at most $1 - \mu(\mathscr{A}(s, r))$. □

The possibility that the GC-depth at the GCM can be less than 0.5 for certain distributions $F_1, \ldots, F_n$ other than the case above is a consequence of the intransitivity of ordering based on pairwise comparisons. Suppose we say that $F_r \prec F_s$ whenever $\int F_r dF_s > 0.5$. Then this ordering can be intransitive, so that we can find $F_1 \prec F_2 \prec \cdots \prec F_n \prec F_1$. For such examples, the GC-depth will be uniformly less than 0.5 for all vertices of the set of permutations. The intransitivity of orderings based on $\prec$ is known as the *voting paradox*. See Johnson and Kotz (1988), pages 325–328, for a description of the voting paradox and relevant references.

As an example, suppose $F_r(x) = \Phi[(x - \mu_r)/\sigma]$ is a normal distribution function with variance σ^2 centered at μ_r. Then we can find an expression for

the depth function. Suppose that the depth is evaluated at the permutation $(i_1, \ldots, i_n)$. Let $\delta = \min\{\mu_{i_r} - \mu_{i_s}; r > s\}$. Then the GC-depth evaluated at the permutation will be $\Phi[\delta/(\sigma\sqrt{2})]$.

For probability measures $\mu = \mu(F_1, \ldots, F_n)$, the asymptotic distribution of the GC-depth function can be analyzed. The reader should compare the analysis which follows with the work of Critchlow and Fligner (1993). For any distribution μ on $\mathscr{X}_n$ let μ^* be defined to be that element of the $n(n-1)/2$-dimensional cube $[0,1]^{n(n-1)/2}$ whose entries are $\mu(\mathscr{A}_{rs})_{1\le r<s\le n}$. For convenience, we abbreviate these entries as μ^*_{rs}. If $\hat{\mu}$ is an empirical distribution on $\mathscr{X}_n$ from a sample of m independent μ-variates, then $\hat{\mu}^*$ is a random point of $[0,1]^{n(n-1)/2}$ having an asymptotic normal distribution centered at μ^* and with covariance matrix $m^{-1}\Sigma$, where $\Sigma_{rs,\,tu} = 0$ when r, s, t and u are distinct integers, and $\Sigma_{rs,\,rs} = \mu^*_{rs}(1-\mu^*_{rs})$. Elements of the matrix of the form $\Sigma_{rs,\,st}$, and so on, where $r < s < t$, and so on, are expressible as $\mu^*_{rs,\,st}(1-\mu^*_{rs,st})$ where

$$\mu^*_{rs,\,st} = \int_{-\infty}^{+\infty} F_r(\xi)[1-F_t(\xi)]\,dF_s(\xi).$$

Also $\Sigma_{rs,\,rt} = \mu^*_{rs,\,rt}(1-\mu^*_{rs,\,rt})$, and $\Sigma_{rs,\,ts} = \mu^*_{rs,\,ts}(1-\mu^*_{rs,\,ts})$, where

$$\mu^*_{rs,\,rt} = \int_{-\infty}^{+\infty} [1-F_s(\xi)][1-F_t(\xi)]\,dF_r(\xi)$$

and

$$\mu^*_{rs,\,ts} = \int_{-\infty}^{+\infty} F_r(\xi)F_t(\xi)\,dF_s(\xi).$$

Now consider the problem of finding the distribution of the GC-depth function at some permutation. Without loss of generality, this can be assumed to be the identity permutation $x = (1, 2, \ldots, n)$ as other points can be transformed into this one by a relabeling. Evaluating at the identity permutation, we note that $\mathrm{GC}(x, \hat{\mu}) \le \mathrm{GC}(x, \mu) + d$ if and only if $\min_{r<s} \hat{\mu}^*_{rs} \le \mathrm{GC}(x, \mu) + d$. The calculation of the exact distribution of this minimum is difficult, requiring numerical tools such as simulation. However, we can use the fact that the random variables $\hat{\mu}^*_{rs}$ are dissociated asymptotically normal random variables. So an asymptotic form for the lower bound is given by

$$1 - \min_{r_k<s_k} \prod_{k=1}^{[n/2]} \Phi\left\{-\sqrt{\frac{m}{\mu^*_{r_k s_k}(1-\mu^*_{r_k s_k})}}\left[\mathrm{GC}(x,\mu) + d - \mu^*_{r_k s_k}\right]\right\}$$
$$\le P\left[\mathrm{GC}(x,\hat{\mu}) \le \mathrm{GC}(x,\mu) + d\right],$$

where again Φ is the standard normal distribution function. The minimum of the lower bound in the inequality is taken over all choices of $[n/2]$ disjoint pairs $r_1 < s_1, \ldots, r_{[n/2]} < s_{[n/2]}$. In the common circumstance where the expectations μ^*_{rs} are uniquely minimized at a particular choice of $r < s$ where

$\mu^*_{rs} = \mathrm{GC}(x, \mu)$, the lower bound given will be asymptotically tight. In this case, the asymptotic distribution of $\mathrm{GC}(x, \hat{\mu})$, will be equal to the distribution of μ^*_{rs}.

The problem of estimating parameters in the Thurstone order statistics model is discussed by Critchlow and Fligner (1993), page 10. The method that is most compatible with the construction of the depth function is to use an iteratively weighted least squares estimation based on the asymptotic normal approximation of $\hat{\mu}^*$. This is equation (8) of Critchlow and Fligner (1993). As the GC-depth is a function of $\hat{\mu}^*$, it is natural to fit parameters so that the estimating equations are those of the multivariate normal model for this vector. Maximization of the multivariate normal likelihood is equivalent to minimization of

$$\frac{n(n-1)}{2}\log\left(\det\sum\right) + (\hat{\mu}^* - \mu^*)m\overset{-1}{\sum}(\hat{\mu}^* - \mu^*)^T.$$

4. Minimal spanning trees. In this section we shall consider the use of GC-depth functions and GCM's on minimal spanning trees of multivariate data sets. A number of the results will generalize immediately to all graphs which are trees.

Let $x_1, x_2, \ldots, x_n$ be n data points in some Euclidean space $\mathbf{R}^p$, the points typically an independent sample from some general distribution on $\mathbf{R}^p$. We shall generally assume that the minimal spanning tree can be uniquely constructed. To ensure this, it is sufficient to suppose that the set of distances between points and their neighbors in the Delaunay triangulation are distinct. For convenience, we shall call this the *distinct neighbor* condition or DN condition. For vertices from some distribution on $\mathbf{R}^p$ which is absolutely continuous, this will hold with probability 1. In this context we shall suppose that the data themselves form the graph through their minimal spanning tree. Unless otherwise stated, the distribution μ under consideration is the empirical distribution which places mass $1/n$ on each vertex or data point.

A path of vertices within the minimal spanning tree will be a geodesic path if and only if the vertices are distinct. Any geodesic path between two distinct vertices is unique. It follows from this that the G-convex sets of the minimal spanning tree are the connected sets, that is, subsets of vertices for which any two distinct vertices can be joined by a path. The half sets of the minimal spanning tree are then seen to be the *branches*. There are a total of $2n - 2$ branches consisting of one of more vertices; these can be characterized as the connected components of the minimal spanning tree after the deletion of an arbitrary edge. The GC-depth of a vertex x can then be determined. If we delete x from the minimal spanning tree, along with the edges which are connected to it, the tree is broken up into k branches, say, having $n_1, n_2, \ldots, n_k$ vertices, respectively. Then the GC-depth at x is $1 - \max(n_1/n, n_2/n, \ldots, n_k/n)$. The GCM will be the vertex maximizing this quantity. It is possible to put bounds on the GC-depth at the GCM. In general, at the GCM of the tree in $\mathbf{R}^p$, we

have

$$\frac{[(n+1)/2]}{n} \leq \mathrm{GCD} \leq \frac{[(nK(p)+1-n)/K(p)]}{n},$$

where $[a]$ represents the greatest integer less than or equal to a, and the function $K(p)$ is the famous *kissing number* for dimension p. It is known that $K(1) = 2$, $K(2) = 6$ and $K(3) = 12$. However, even for dimensions as low as $p = 4$, the kissing number $K(4)$ is unknown, although it is known that $K(4) = 24$ or $K(4) = 25$. Thus a valid bound can be obtained by using the latter figure. The kissing number can be defined as the greatest number of unit spheres in dimension p which can be arranged around the outside of a given unit sphere so as to simultaneously be tangent to the given sphere. The case $p = 2$ is the well-known penny problem. The kissing number is also the largest number of edges that can be attached to a vertex of a minimal spanning tree. The upper bound on the GC-depth at the median is obtained by constructing the minimal spanning tree of the centers of the sphere arrangement attaining the kissing number, and attaching the remaining vertices using branches from the centers of the surrounding spheres. An application of the pigeonhole principle gives the upper bound of the GC-depth at the vertex corresponding to the central sphere. This is an upper bound on the GC-depth for any vertex of the minimal spanning tree, and *a fortiori* an upper bound on the GCM. For the special case $p = 2$, we can improve on this slightly by noting that the arrangement that attains the kissing number 6 violates the DN condition. Thus we can tighten the bound by replacing $K(2)$ by 5. However, in higher dimensions the arrangement that attains the kissing number will not generally be rigid, as some gaps will exist between neighboring spheres. For more work on kissing numbers, see Conway and Sloane (1988). Just as the upper bound is obtained by the maximum number of edges at the vertex, so the lower bound is obtained with the smallest possible number of edges at a GCM, namely 2.

The vertices of the minimal spanning tree of minimal depth are those points which have only one edge attached to them. These points are called *leaves* (i.e., the ends of the branches). Some specialized asymptotic results are available as to the proportion of leaves in a minimal spanning tree. See Steele, Shepp and Eddy (1987). In particular, the proportion can be shown to converge to a constant for a sample of n independent identically distributed points as $n \to \infty$. The value of this constant is generally difficult to obtain but is known to be approximately 2/9 for the minimal spanning tree of a set of uniform independent points in a square (dimension 2).

No matter what the dimension, the minimal spanning tree GCM shares the property of the one-dimensional sample median of being essentially unique. That is, the GC-depth will be maximized on one or at most two points of the minimal spanning tree. This follows from the fact that if x_1, x_2 and x_3 form a path of the tree, then the GC-depth at x_2 is strictly greater than the minimum of the GC-depths at x_1 and x_3. Thus these three vertices cannot all be GCMs. However, the set of all GCMs is connected, and this implies that its cardinality is at most 2.

Another property that the minimal spanning tree GCM shares with the univariate sample median is a 50 percent breakdown. The breakdown point of an estimator is the smallest proportion of the data whose perturbation to ∞ is capable of forcing the estimator to go to ∞ as well. If any proportion less than 0.5 of the vertices $x_1, x_2, \ldots, x_n \in \mathbf{R}^p$ is moved toward the point at ∞, deletion of any of these vertices and their associated edges will not disconnect the set of unmoved vertices. Thus the GC-depth at these vertices moved to ∞ will be less than $1/2$. Therefore, the GCM will be found among the remaining unperturbed vertices. This is a simple consequence of the property stated above, namely that the GC-depth at the GCM is bounded below by $[(n+1)/2]/n$. This value or higher cannot be attained at a vertex moved to ∞.

EXAMPLE 4.1. We now consider the application of GC-depth functions to some methods developed by Friedman and Rafsky (1979) for distribution-free nonparametric two-sample tests in $\mathbf{R}^p$ using minimal spanning trees. In particular, we consider the planar case $p = 2$. The null hypothesis of a two-sample test on a graph can be stated in graph theory terminology as saying that the partition of the graph into two sets is equivalent to a random two-coloring of the graph. Suppose $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ are two samples in $\mathbf{R}^2$. We draw the minimal spanning tree on the combined data and consider testing whether the two samples come from the same distribution versus a possible alternative such as a location, scale or other change. A natural class of tests is conditional on the topology of the graph. As an example of the use of minimal GC-depth for minimal spanning trees, we consider the data discussed by Hodder and Orton (1976). These data consist of the geographical locations, plotted in two dimensions, of a variety of pin and pendant types from the Middle Bronze Age. Figure 1 displays the minimal spanning tree for the distribution of pin type 76 and pendant type 128. Associated with each vertex, the GC-depth is given in Figure 1 as well. An examination of this figure shows clear evidence that the spatial distribution of type 128 differs from that of type 76. A chi-square test based on division of the plane into quadrats is possible. However, the dependence of the test on the method for grouping univariate data is exacerbated for bivariate or spatial data because there is an arbitrariness about the shape and orientation of the rectangles into which the region is subdivided. More successful are the methods of Friedman and Rafsky (1979). Friedman and Rafsky (1979) proposed a 2×2 contingency table test that is appropriate for scale alternatives to the null hypothesis. Suppose the data are divided into two groups based on the topology of the minimal spanning tree. Each vertex of the tree can be classified into a 2×2 contingency table based on its color (i.e., sample) and its group. As vertices are grouped without regard to color, the rows and columns will be independent under the null hypothesis. A chi-square test for independence can be run on this contingency table. Friedman and Rafsky noted that partitioning the tree into vertices that are leaves and remaining vertices has reasonable power for scale alternatives in higher dimensions. Its poor performance in lower dimensions can be seen clearly in dimension 1, where the minimal spanning tree has only two vertices which

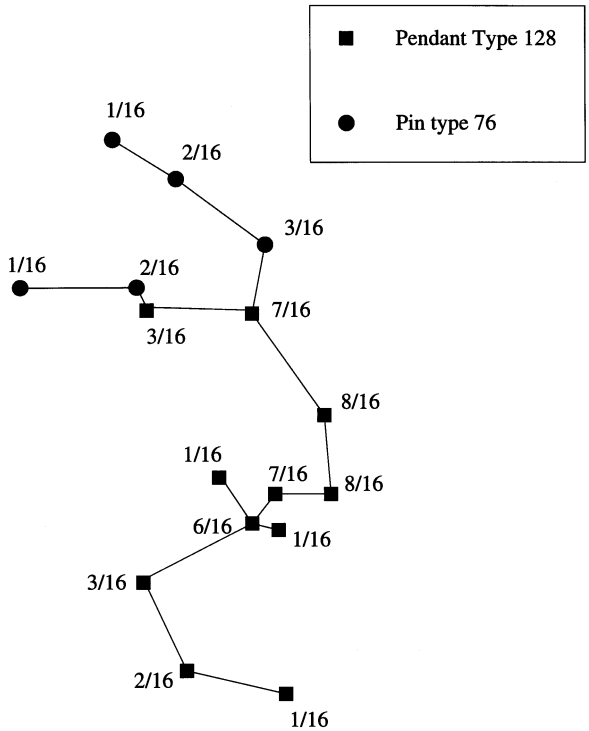

FIG. 1. *Pin and pendant types from Middle Bronze age.*

are leaves. We note here that the GC-depth function provides ample flexibility to partition the data into two groups based on the minimal spanning tree. An approach that is appropriate for the apparent location shift of our example is to sever the edge which joins the two vertices in Figure 1 having greatest depth. The result is a partition of the tree into two equal groups. The hypothesis that the two samples are from the same distribution is easily rejected by a test for independence on the 2×2 contingency table based on the partition. Fisher's exact test gives a p-value of 0.0128. Other partitions of a tree may be more appropriate to scale alternatives. For example, vertices can be partitioned into a group with high GC-depth and a group with low GC-depth. A special case of this partition leads to the contingency table test for scale alternatives proposed by Friedman and Rafsky: in this case, the group of vertices with smallest GC-depth consists of the leaves of the tree.

We close this section with a consideration of the computational aspects of GC-depth for minimal spanning trees. The computation of the minimal spanning tree itself has been well developed by Prim (1957). More efficient methods for constructing the tree are based on the fact that it is a subgraph of the Delaunay triangulation. See Preparata and Shamos (1985). We shall

only concern ourselves here with the calculation of the GC-depth function. The calculation of this function reduces to the problem of finding the size of the largest connected component of a given graph. This can be determined directly by classifying points iteratively into classes where any two points sharing a common edge are classified similarly. An alternative procedure is to construct a uniformized Markov chain and to determine its (approximate) equilibrium distribution. Let M be the $n \times n$ matrix whose ijth entry is 1 if the ith and jth vertices share a common edge, and is 0 if not. Let M_i be the $(n-1) \times (n-1)$ matrix of 0's and 1's obtained from M by deleting the ith row and the ith column. For $j = 1, \ldots, n-1$ let s_j be the sum of the elements of the jth row of M_i. We set $s = \max(s_1, s_2, \ldots, s_{n-1})$ and let N_i be the $(n-1) \times (n-1)$ matrix whose off-diagonal elements are identical to those of M_i and whose jth diagonal element is $s - s_j$. We construct a transition matrix P_i for a uniformized Markov chain on the vertices by dividing the elements of N_i by s. The Markov chain on the state space of $n-1$ vertices will have a limiting distribution that is uniform on the connected components of the graph. We write P_i in similar canonical form as $P_i U_i = U_i D_i$, where $D_i = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{n-1})$ is the diagonal matrix of eigenvalues of P_i. Let D_i^* be the diagonal matrix whose jth diagonal element is 1 if and only if the jth eigenvalue is 1, and is 0 otherwise. That is,

$$D_i^* = \text{diag}[1(\lambda_1 = 1), 1(\lambda_2 = 1), \ldots, 1(\lambda_{n-1} = 1)].$$

The GC-depth of the vertex x_i is determined as the reciprocal of the minimum of the elements on the main diagonal of the matrix $U_i D_i^* U_i^{-1}$.

5. Conclusions. We have illustrated the use of GC-depth and GCMs through two examples and data sets. Many other types of graphs remain to be considered although general results such as those of Section 2 hold for all.

Acknowledgments. I am grateful to an Associate Editor and two referees for detailed and insightful comments on this paper.

REFERENCES

BENSON, R. V. (1966). *Euclidean Geometry and Convexity*. McGraw-Hill, New York.

BROWN, B. M. (1983). Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B* **45** 25–30.

CONWAY, J. H. and SLOANE, N. J. A. (1988). *Sphere Packings, Lattices and Groups*. Springer, New York.

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

CRITCHLOW, D. E. and FLIGNER, M. A. (1993). Ranking models with item covariates. In *Probability Models and Statistical Analyses for Ranking Data. Lecture Notes in Statist.* **80** 1–19. Springer, New York.

CRITCHLOW, D. E. and VERDUCCI, J. S. (1992). Detecting a trend in paired rankings. *Appl. Statist.* **41** 17–30.

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

DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on half-space depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.

DUCHARME, G. R. and MILASEVIC, P. (1987). Spatial median and directional data. *Biometrika* **74** 212–215.

FISHER, N. I. (1985). Spherical medians. *J. Roy. Statist. Soc. Ser. B* **47** 342–348.

FRIEDMAN, J. H. and RAFSKY, L. C. (1979). Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7** 697–717.

GRIEG-SMITH, P. (1952). The use of random and contiguous quadrats in the study of the structure of plant communities. *Ann. Botany London* **16** 293–316.

HODDER, I. and ORTON, C. (1976). *Spatial Analysis in Archaeology* 198–223. Cambridge Univ. Press.

JAMISON-WALDNER, R. E. (1982). A perspective on abstract convexity: classifying alignments by varieties. In *Convexity and Related Combinatorial Geometry*. *Proceedings of the Second University of Oklahoma Conference* (D. C. Kay and M. Breen, eds.) 113–150.

JOHNSON, N. L. and KOTZ, S. (1988). *Encyclopedia of Statistical Sciences* **9** 507–509. Wiley, New York.

LIU, R. Y. and SINGH, K. (1992). Ordering directional data: concepts of data depth on circles and spheres. *Ann. Statist.* **20** 1468–1484.

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

MCCULLAGH, P. (1993). Models on spheres and models for permutations. In *Probability Models and Statistical Analyses for Ranking Data*. *Lecture Notes in Statist.* **80** 278–283. Springer, New York.

MCCULLAGH, P. and YE, J. (1993). Matched pairs and ranked data. In *Probability Models and Statistical Analyses for Ranking Data*. *Lecture Notes in Statist.* **80** 299–305. Springer, New York.

NIINIMAA, A., OJA, H. and TABLEMAN, M. (1990). The finite-sample breakdown point of the Oja bivariate median and of the corresponding half-samples version. *Statist. Probab. Lett.* **10** 325–328.

OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332.

OJA, H. and NIINIMAA, A. (1985). Asymptotic properties of the generalized median in the case of multivariate normality. *J. Roy. Statist. Soc. Ser. B* **47** 372–377.

PIELOU, E. C. (1969). *An Introduction to Mathematical Ecology.* Wiley, New York.

PREPARATA, F. and SHAMOS, M. (1985). *Computational Geometry: An Introduction.* Springer, New York.

PRIM, R. C. (1957). Shortest connection networks and some generalizations. *Bell System Tech. J.* **36** 1389–1401.

SMALL, C. G. (1990). A survey of multidimensional medians. *Internat. Statist. Rev.* **58** 263–277.

STEELE, J. M., SHEPP, L., and EDDY, W. (1987). On the number of leaves of a minimal spanning tree. *J. Appl. Probab.* **24** 809–826.

THOMPSON, G. L. (1993). Generalized permutation polytopes and exploratory graphical methods for ranked data. *Ann. Statist.* **21** 1401–1430.

TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians*, *Vancouver*, *1974* **2** 523–531.

YEMELICHEV, V. A., KOVALEV, M. M. and KRAVTSOV, M. K. (1984). *Polytopes*, *Graphs*, *and Optimisation.* Cambridge Univ. Press.

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO
CANADA N2L 3G1