Sign Tests in Multidimension: Inference Based on the Geometry of the Data Cloud

Probal CHAUDHURI and Debapriya SENGUPTA*

Multivariate sign tests attracted several statisticians in the past, and it is evident from recent nonparametric literature that they still continue to draw attention. One of the most important features of the univariate sign test is that it does not involve much technical assumptions or complicacy, and this makes it quite popular among statistics users. In this article we have come up with a new method for constructing multivariate sign tests that have reasonable statistical properties and can be used conveniently to solve one-sample location problems. Our principal strategy here is to make a wise utilization of certain geometric structures in the constellation of data points for making inference about the location of their distribution. As we proceed with the development of a fairly broad and general methodology, we indicate its relationship with previous work done by others and sometimes attempt to unify some of the earlier ideas. In particular, we pick up some well-known tests for uniform distribution of directional data and integrate them into the technology of multivariate sign tests to synthesize useful new procedures. Our procedures enjoy affine invariance and the distribution-free property for elliptically symmetric models. We report several interesting results that provide powerful insights into certain critical aspects of the problem. What is most appealing is the fundamental dependence of our approach on the basic geometry of the data cloud formed by the observations. In this article our only key to unlock the information contained in the data is the spatial arrangement of data points in a d-dimensional Euclidean space.

KEY WORDS: Affine invariance; Beran's family of tests; Data-driven coordinates; Degenerate U statistics; Distribution-free tests; Elliptically symmetric distributions.

1. A REVIEW OF AVAILABLE TECHNIQUES

The simplicity and the distribution-free nature of the univariate sign test for one-sample location problems have motivated numerous authors to explore several multivariate versions of the test. Pioneering attempts to construct sign tests, which are applicable to bivariate data, were made by Blumen (1958) and Hodges (1955). These efforts continued in the 1960s when Bennett (1962) and Bickel (1965) developed some multivariate sign tests and Chatterjee (1966) introduced a bivariate sign test. A strong enthusiasm about the problem still persists, as is apparent in recent papers by Brown and Hettmansperger (1989), Brown, Hettmansperger, Nyblom, and Oja (1992), Dietz (1982), Oja and Nyblom (1989), Randles (1989), and others. The popularity of the univariate sign test has its root in its widespread applicability to solve a number of practical problems for which more sophisticated techniques cannot be used as they will frequently require technical assumptions that are hard to justify in practice. Bivariate and trivariate sign tests have been used to analyze several interesting data sets by Brown et al. (1992), Dietz (1982), Oja and Nyblom (1989), Randles (1989), and others. In addition to these tests' practical importance, several exceedingly challenging and stimulating theoretical issues revolve around multivariate extensions of the sign test. Relevant theoretical studies in the literature have focused on distribution-free property, affine invariance, and asymptotic efficiency. It is a well-known fact that the univariate sign test is actually a test for the median of the probability distribution generating the data. So it is natural to expect that a multivariate sign test will be a test for the median of the parent multivariate distribution from which the observations are drawn. Because there are several ways of defining the median of a multivariate distribution (see, for example, Small 1990), there is a wide range of possibilities for multivariate sign tests. This creates ample opportunity for statisticians to play with new ideas, try innovative techniques, and discover interesting facts.

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be iid observations in R^d with a common absolutely continuous distribution having median $\theta \in R^d$, where the median is defined in some appropriate way. The basic multivariate one-sample location problem considered in this article can be formulated in terms of testing the null hypothesis $H_0: \theta = 0$ against the alternative $H_A: \theta \neq 0$. For $d = 1$, the standard sign test is based on $\operatorname{sign}(\mathbf{X}_1), \operatorname{sign}(\mathbf{X}_2), \ldots, \operatorname{sign}(\mathbf{X}_n)$, which are iid random variables each taking the values $+1$ and -1 with equal probabilities under H_0. To get extensions of this test in dimensions $d \geq 2$, one must define the "sign" of a random vector in a suitable way. Hereafter, unless specified otherwise, every vector in this article will be a column vector, and T will denote the transpose of vectors and matrices. For $1 \leq j \leq d$, let X_{ij} be the jth real-valued component of the random vector $\mathbf{X}_i$. Then the most naive definition of $\operatorname{sign}(\mathbf{X}_i)$ is given by $\operatorname{sign}(\mathbf{X}_i) = (\operatorname{sign}(X_{i1}), \operatorname{sign}(X_{i2}), \ldots, \operatorname{sign}(X_{id}))^T$, and tests for H_0 can be constructed based on $\operatorname{sign}(\mathbf{X}_1), \operatorname{sign}(\mathbf{X}_2), \ldots, \operatorname{sign}(\mathbf{X}_n)$. Such tests have been extensively considered by Bennett (1962), Bickel (1965), Chatterjee (1966), Puri and Sen (1971), and others, and they can be viewed as tests for the median θ of $\mathbf{X}_i$, where the median of a random vector is defined as the vector of the medians of its real-valued components. Suppose now that the common density of the $\mathbf{X}_i$'s is of the form $f(\mathbf{x} - \theta)$, where f is a symmetric function whose value depends only on the absolute values of the coordinates of its d-di-

* Probal Chaudhuri and Debapriya Sengupta are members of the scientific staff, Division of Theoretical Statistics and Mathematics, Indian Statistical Institute, Calcutta 700035, India. Simulations were carried out on a VAX 8650 maintained by the Computer & Statistical Services Center (CSSC) at the Indian Statistical Institute, Calcutta. Chaudhuri's research was supported in part by a Wisconsin Alumni Research Foundation Grant from University of Wisconsin, Madison. The authors thank two anonymous referees and an anonymous associate editor who reviewed earlier drafts of the article with unusual care and interest; their constructive suggestions contributed greatly towards several improvements in the revision.

mensional argument $\mathbf{x} - \theta$. Then it is easy to see that the null distribution of $\text{sign}(\mathbf{X}_i)$ does not depend on f, and in fact it is uniform on the set of all 2^d d-tuples of $+1$'s and -1's. In particular, this ensures the distribution-free property of any test that depends solely on $\text{sign}(\mathbf{X}_1), \text{sign}(\mathbf{X}_2), \ldots, \text{sign}(\mathbf{X}_n)$.

In the case $d = 1$ and $\mathbf{X}_i \neq 0$, if we recognize the fact that $\text{sign}(\mathbf{X}_i) = |\mathbf{X}_i|^{-1}\mathbf{X}_i$, we have another way of constructing multivariate analog of the univariate sign test. For $d \geq 2$, let us define the unit direction vector $U(\mathbf{X}_i)$ in the direction of $\mathbf{X}_i (\neq 0)$ as $U(\mathbf{X}_i) = |\mathbf{X}_i|^{-1}\mathbf{X}_i$, where $|\cdot|$ denotes the standard Euclidean norm. If the distribution of $\mathbf{X}_i$ is spherically symmetric around θ, then it is obvious that, under the hypothesis H_0, $U(\mathbf{X}_i)$ will have a uniform distribution on the unit sphere $S^{(d-1)} = \{\mathbf{x}: \mathbf{x} \in R^d \text{ and } |\mathbf{x}| = 1\}$. Hence, in that case, a test based on $U(\mathbf{X}_1), U(\mathbf{X}_2), \ldots, U(\mathbf{X}_n)$ is going to be distribution free, and such a test can be looked upon as a test for the L_1 median of $\mathbf{X}_i$, where the L_1 median θ of $\mathbf{X}_i$ is defined as $E(|\mathbf{X}_i - \theta|) = \min_{\phi \in R^d} E(|\mathbf{X}_i - \phi|)$ assuming that the expectations exist. This definition of the median of a random vector has been used by Haldane (1948), Brown (1983), and many others, and an interesting discussion about it can be found in the review paper by Small (1990). It will be appropriate to note here that there are numerous nonparametric tests developed in the literature to test the uniformity of distributions on circles and spheres. Extensive review of such tests can be found in Jammalamadaka (1984), Jupp and Mardia (1989), and Mardia (1972). Any of those tests for testing uniform distribution of directional data can be used to construct multivariate sign tests using the direction vectors $U(\mathbf{X}_1), U(\mathbf{X}_2), \ldots, U(\mathbf{X}_n)$, assuming spherical symmetry of the probability law followed by the data.

It is obvious that a test based on $U(\mathbf{X}_i)$'s will be invariant under orthogonal transformations of the data. But neither the $\text{sign}(\mathbf{X}_i)$'s nor the $U(\mathbf{X}_i)$'s are invariant under an arbitrary affine transformation of the $\mathbf{X}_i$'s. Hodges (1955) considered a bivariate sign test that rejects H_0 for large values of the statistic $\sup_{\lambda \in R^2} |\sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle)|$, where $\langle \cdot, \cdot \rangle$ denotes the usual Euclidean inner product. Here, for the sake of completeness, we adopt the convention that $\text{sign}(0) = 0$. This test is clearly affine-invariant and reminds us about the well-known "union intersection criterion" introduced by Roy (1953, 1957) to develop multivariate extensions of several standard univariate tests. Further, this bivariate sign test has a natural multivariate extension, and we will demonstrate later (in Sec. 3.2) that such a test has some fundamental connections with the "half space median" and a related notion of "data depth" in Tukey (1975). Blumen (1958) introduced another affine-invariant bivariate sign test, which was extended to a multivariate sign test by Randles (1989) using the concept of "interdirections" (see Sec. 3.1). Puri and Sen (1971, p. 148) criticized Blumen's test and Hodges's test on the grounds that they use statistics with very complex distributions that are difficult to work with in practice. But with recent developments in resampling techniques like the bootstrap and the rapid emergence of powerful computing facilities capable of doing intensive simulations, the distributional complexity of a test statistic is no longer a serious practical problem. Recently, Brown and Hettmansperger (1989) and Brown et al. (1992) investigated a bivariate sign test that is also affine-invariant. This test is naturally associated with Oja's simplex median (see Oja 1983), and it does not have the finite sample distribution-free property.

In this article we develop and study certain sign tests that exploit some very fundamental geometric structures in multivariate data sets. Our principal goal here is to form a basic tool kit for constructing affine-invariant multivariate sign tests. In the process of achieving this target, we construct new tests, extend existing procedures, and enrich the available methodology by blending it with new ideas.

The article is organized as follows. In Section 2 we introduce a data-driven coordinate system under which the transformed coordinates of the observations become affine-invariant. We show that the new set of transformed coordinates is a maximal invariant under nonsingular linear transformations, and that the "sign" vectors associated with transformed observations have the distribution-free property under models that are elliptically symmetric around the origin. We also develop a new family of multivariate sign tests using the "signs" of the invariant coordinates. In Section 3 we explore some intriguing relationships that exist among various tests considered in the literature and the tests proposed by us. In Section 4 we investigate the asymptotic properties of the new class of sign tests and present some finite sample results based on simulations. In Section 5 we conclude the article with a discussion of some of the open questions and related issues. All technical proofs are relegated to the Appendix.

2. CONSTRUCTION OF AFFINE-INVARIANT MULTIVARIATE SIGN TESTS: A NEW APPROACH

Let us now assume that the iid observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ in R^d are obtained from an elliptically symmetric distribution with density $|\Sigma|^{-1/2} f\{(\mathbf{x} - \theta)^T \Sigma^{-1} (\mathbf{x} - \theta)\}$, where Σ is a $d \times d$ symmetric positive-definite matrix and $|\Sigma|$ is its determinant. As before, we will be interested in testing H_0: $\theta = 0$ against H_A: $\theta \neq 0$. Because Σ will typically be unknown in practice, we cannot reduce the problem to a spherically symmetric situation. Also, note that if the $\mathbf{X}_i$'s get transformed to $\mathbf{A}\mathbf{X}_i$'s, where $\mathbf{A}$ is a $d \times d$ nonsingular matrix, then θ gets transformed into $\mathbf{A}\theta$ and Σ into $\mathbf{A}\Sigma\mathbf{A}^T$. Because $\mathbf{A}\theta = 0$ if and only if $\theta = 0$ in view of the nonsingularity of $\mathbf{A}$, the problem is intrinsically affine-invariant, and it will be most natural to use statistics that possess affine invariance to carry out the test. We will now present some techniques that can be used to construct multivariate sign tests that are affine-invariant. These techniques are quite easy to implement in practice and have very appealing geometric interpretations. In Section 3 we will demonstrate that Randles's test (Randles 1989) and the obvious multivariate extension of Hodges's bivariate sign test (Hodges 1955) are interesting specialized applications of a more general methodology developed here.

2.1 Data-Driven Coordinate Systems

We begin by introducing some notation. Define $S_n = \{\alpha \mid \alpha \subseteq \{1, 2, \ldots, n\} \text{ and } |\alpha| = d\}$, which is the collection of all subsets of size d of $\{1, 2, \ldots, n\}$. For a fixed $\alpha \in S_n$, let $\mathbf{X}(\alpha)$ be the $d \times d$ matrix whose columns are the random vectors $\mathbf{X}_i$'s with $i \in \alpha$. Here we are assuming that the elements of α are naturally ordered, and in view of the absolute continuity of the common distribution of the $\mathbf{X}_i$'s, $\mathbf{X}(\alpha)$ must be an invertible matrix with probability 1. We will treat $\mathbf{X}(\alpha)$ as the basis matrix for a data-based coordinate system. In terms of this basis matrix, a data point $\mathbf{X}_i$ such that $i \notin \alpha$ can be represented as $\mathbf{Y}_i^{(\alpha)} = \{\mathbf{X}(\alpha)\}^{-1}\mathbf{X}_i$. Note that here we are trying to view the data cloud from a data-centric reference frame created by the basis matrix $\mathbf{X}(\alpha)$. The transformed coordinates of the $\mathbf{X}_i$'s for $i \in \alpha$ are given by the standard basis vectors in R^d and hence they are non-informative. A simple but crucial fact can be stated as follows.

Proposition 2.1. Fix an $\alpha \in S_n$, and let the common distribution of the iid observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be absolutely continuous in R^d. Then the transformed data points $\mathbf{Y}_i^{(\alpha)}$'s with $1 \le i \le n$ and $i \notin \alpha$ form a maximal invariant with respect to the group of nonsingular linear transformations on R^d.

We can appreciate the invariance property of the transformed observations $\mathbf{Y}_i^{(\alpha)}$'s more if we look at them with a physicist's eyes. We can identify a nonsingular linear transformation as some kind of a motion on the data cloud and the data-based coordinate system with the basis matrix $\mathbf{X}(\alpha)$ as a reference frame for an observer. Because the data-centric reference frame moves with the motion of the data cloud, the data points will appear to be stationary (i.e., invariant) to the observer. The following theorem reveals a fundamental statistical property of the transformed observations.

Theorem 2.2. Recall from the beginning of this section that the $\mathbf{X}_i$'s are iid observations with a common elliptically symmetric density $|\mathbf{\Sigma}|^{-1/2}f\{(\mathbf{x}-\theta)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\theta)\}$. Then, under the hypothesis $H_0: \theta = 0$, the joint distribution of the random vectors $\text{sign}(\mathbf{Y}_i^{(\alpha)})$'s, where the index i runs in the complement of α in $\{1, 2, \ldots, n\}$ and α runs in S_n, depends neither on $\mathbf{\Sigma}$ nor on f. Here the "sign" of a vector is defined in the same way as in section 1.

We will now use $\text{sign}(\mathbf{Y}_i^{(\alpha)})$'s as basic building blocks for the construction of various affine-invariant versions of multivariate sign tests.

2.2 A New Family of Multivariate Sign Tests

Because the choice of a particular $\alpha \in S_n$ is quite subjective, the tests considered in this article will involve all of the vectors $\text{sign}(\mathbf{Y}_i^{(\alpha)})$'s with α varying in the set S_n and i taking its integral values between 1 and n in the complement of α. Moreover, our test statistics will depend on the α's and the i's in a perfectly symmetric way. In view of Proposition 2.1 and Theorem 2.2, any reasonable test for H_0 against H_A, which is based solely on the vectors $\text{sign}(\mathbf{Y}_i^{(\alpha)})$'s, will be affine-invariant, distribution-free, and consequently a legitimate multivariate analog of the univariate sign test. A candidate that arises naturally is the test that rejects H_0 for large values of the statistic

$$W_n = \sum_{\alpha \in S_n} \left| \sum_{i \notin \alpha} \text{sign}(\mathbf{Y}_i^{(\alpha)}) \right|^2$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \sum_{\alpha \subset S_n, \alpha \not\ni i, j} \langle \text{sign}(\mathbf{Y}_i^{(\alpha)}), \text{sign}(\mathbf{Y}_j^{(\alpha)}) \rangle \right\}.$$

We will gradually see that for $d = 2$, W_n is related in a very interesting way to a statistic used by Ajne (1968) for testing the uniformity of a circular distribution. Further, in the case $d = 3$, W_n is related in the same way to a spherical extension of Ajne's test considered by Beran (1968). Beran (1968, 1969) (see also Jammalamadaka 1984; Mardia 1972) developed and investigated a very important class of tests for uniform distribution of circular data. This family includes famous tests, like Ajne's A_n test (Ajne 1968), Rayleigh's test (see Mardia 1972), and Watson's test (Watson 1961) as particular cases, and any test in the family has an obvious and natural extension to a test for uniform distribution on the unit sphere $S^{(d-1)}$. The vectors $\text{sign}(\mathbf{Y}_i^{(\alpha)})$'s can be used together with Beran's fundamental idea to generate an interesting family of useful multivariate sign tests, as follows.

For $1 \le i \ne j \le n$, let us define the quantity $\hat{\psi}_n(i, j)$ as

$$\hat{\psi}_n(i, j) = \frac{\pi}{2}\left[1 - \frac{(n-d-2)!(d-1)!}{(n-2)!} \times \sum_{\alpha \in S_n, \alpha \not\ni i, j} \langle \text{sign}(\mathbf{Y}_i^{(\alpha)}), \text{sign}(\mathbf{Y}_j^{(\alpha)}) \rangle \right],$$

and adopt the convention that $\hat{\psi}_n(i, i) = 0$. Hereafter, we will use $\hat{\psi}_n(i, j)$ as a nonparametric estimate of the geodesic angle $\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j) \rangle)$ between the data points $\mathbf{X}_i$ and $\mathbf{X}_j$, and a formal justification for it will be given in Sections 3.1 and 4. Recall that $U(\mathbf{X}_i)$ and $U(\mathbf{X}_j)$ are unit direction vectors along the directions of $\mathbf{X}_i$ and $\mathbf{X}_j$, as defined in Section 1. When $d = 2$, Ajne's A_n test (Ajne 1968) for testing the uniformity of the distribution of $U(\mathbf{X}_i)$'s on the unit circle $S^{(1)}$ uses a statistic equivalent to $\sum_{i=1}^n \sum_{j=1}^n \cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j) \rangle)$ (see Mardia 1972, pp. 191–192). It is now obvious that if we replace $\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j) \rangle)$ by $\hat{\psi}_n(i, j)$ in Ajne's A_n statistic and in Beran's spherical extension (Beran 1968) of the A_n statistic, we obtain test statistics equivalent to W_n in dimensions $d = 2$ and $d = 3$. For testing the uniform distribution of the $U(\mathbf{X}_i)$'s on the unit sphere $S^{(d-1)}$ in R^d, a natural extension of a typical member in Beran's family of test statistics (Beran 1968, 1969) has the form $T_n = \sum_{i=1}^n \sum_{j=1}^n h\{\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j) \rangle)\}$, where h is an appropriate real-valued kernel defined on the interval $[0, \pi]$. If we replace $\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j) \rangle)$ by $\hat{\psi}_n(i, j)$ in the expression defining T_n, we get the statistic $T_n^* = \sum_{i=1}^n \sum_{j=1}^n h\{\hat{\psi}_n(i, j)\}$. Note that for $d = 1$ and a pair of distinct indices i and j, $\hat{\psi}_n(i, j)$ equals 0 if $\text{sign}(\mathbf{X}_i) = \text{sign}(\mathbf{X}_j)$, and it equals π if $\text{sign}(\mathbf{X}_i) \ne \text{sign}(\mathbf{X}_j)$. Hence if $h(0) \ne h(\pi)$, then a test based on T_n^* will be equivalent to the standard univariate sign test for the two-sided alternative. Also, in view of our preceding results and observations, T_n^* is affine-invariant in dimensions $d \ge 2$, and it will lead to tests with distribution-free property under the assumption of elliptic symmetry of the common distribution of the $\mathbf{X}_i$'s. It is easy to check that an appropriate kernel corresponding to Ajne's A_n test is $h(\psi) = (\pi/2) - \psi$ and the associated T_n^* is equivalent to W_n. When

$$h(\psi) = 2\left\{\frac{1}{6} - \frac{\psi}{2\pi} + \left(\frac{\psi}{2\pi}\right)^2\right\},$$

T_n becomes Watson's statistic (Watson 1961), and

$$T_n^* = \sum_{i=1}^{n} \sum_{j=1}^{n} 2\left\{\frac{1}{6} - \frac{\hat{\psi}_n(i,j)}{2\pi} + \left(\frac{\hat{\psi}_n(i,j)}{2\pi}\right)^2\right\}$$

yields another multidimensional extension of the sign test.

3. GEOMETRIC CONFIGURATION OF POINTS IN A DATA CLOUD AND RELATIONSHIPS AMONG MULTIVARIATE SIGN TESTS

We will now try to explore some more useful geometric structures in a multivariate data cloud and relate them to data-driven coordinates $\mathbf{Y}_i^{(\alpha)}$'s. Let Q_n be the collection of all subsets of size $d-1$ of $\{1, 2, \ldots, n\}$, so that $Q_n = \{\beta: \beta \subseteq \{1, 2, \ldots, n\}$ and $|\beta| = d-1\}$. For any $\beta \in Q_n$, we will denote by $H(\beta)$ the unique hyperplane in R^d containing the origin and the data points $\mathbf{X}_i$'s with $i \in \beta$. Clearly, $H(\beta)$ will split R^d into two disjoint parts, which are nothing but the two sides of this hyperplane. We assume that the elements of β are ordered in a natural way, and for $1 \le j \le n$ and $j \notin \beta$, let us define $\Delta(\beta, j)$ to be the determinant of the $d \times d$ matrix whose first $d-1$ columns are the $\mathbf{X}_i$'s for which $i \in \beta$ and the dth column is $\mathbf{X}_j$. Then $\text{sign}\{\Delta(\beta, j)\}$ is an indicator of the side of $H(\beta)$ in which $\mathbf{X}_j$ falls. The quantities $\text{sign}\{\Delta(\beta, j)\}$'s, where j varies in the complement of β in $\{1, 2, \ldots, n\}$ and β runs in Q_n, contain some fundamental information regarding the spatial configuration of the data points $\mathbf{X}_i$'s shaping the geometry of the data cloud in R^d. In particular, note that $|\sum_{j \notin \beta} \text{sign}\{\Delta(\beta, j)\}|$ gives the difference between the number of data points that fall in one side of $H(\beta)$ and the number of data points that fall in its other side. Hereafter, we will denote the quantity $|\sum_{j \notin \beta} \text{sign}\{\Delta(\beta, j)\}|$ by $\Phi_n(\beta)$. Elementary matrix algebra and the properties of the determinant of a matrix lead to the following important fact.

Fact 3.1. As before, assume that the $\mathbf{X}_i$'s are iid random vectors with a common elliptically symmetric density $|\mathbf{\Sigma}|^{-1/2} f\{(\mathbf{x}-\theta)^T \mathbf{\Sigma}^{-1}(\mathbf{x}-\theta)\}$. The quantities $\Phi_n(\beta)$'s with $\beta \in Q_n$ are invariant under any nonsingular linear transformation of the data points $\mathbf{X}_i$'s. Further, under the hypothesis H_0: $\theta = 0$, the joint distribution of $\Phi_n(\beta)$'s does not depend on f or $\mathbf{\Sigma}$.

This fact enables us to construct affine-invariant multivariate sign tests using $\Phi_n(\beta)$'s as the key quantities.

3.1 The Notion of Interdirections and Randles's Test

For $d \ge 2$, let us consider the test that rejects the null hypothesis H_0 for large values of the statistic $\sum_{\beta \in Q_n} \{\Phi_n(\beta)\}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} [\sum_{\beta \in Q_n, \beta \not\ni i,j} \text{sign}\{\Delta(\beta, i)\} \text{sign}\{\Delta(\beta, j)\}]$. In view of the definition of the transformed observation $\mathbf{Y}_i^{(\alpha)}$, we can view it as the solution to the system of linear equations $\mathbf{X}(\alpha)\mathbf{Y}_i^{(\alpha)} = \mathbf{X}_i$. Some elementary algebra, using Cramer's rule for solving linear equations and simple counting, now yield the following fact, which relates the vectors $\text{sign}(\mathbf{Y}_i^{(\alpha)})$'s to the quantities $\text{sign}\{\Delta(\beta, j)\}$'s in an interesting way.

Fact 3.2. For $d \ge 2$ and any pair of indices i and j such that $1 \le i \ne j \le n$, we have $\sum_{\alpha \in S_n, \alpha \not\ni i,j} \langle \text{sign}(\mathbf{Y}_i^{(\alpha)}), \text{sign}(\mathbf{Y}_j^{(\alpha)}) \rangle = (n-d-1) \sum_{\beta \subset Q_n, \beta \not\ni i,j} \text{sign}\{\Delta(\beta, i)\} \times \text{sign}\{\Delta(\beta, j)\}$.

This fact immediately implies that for $1 \le i \ne j \le n$, $\hat{\psi}_n(i, j)$, which is defined in Section 2.2, can be written in terms of $\Delta(\beta, j)$'s as

$$\hat{\psi}_n(i,j) = \frac{\pi}{2}\left[1 - \frac{(n-d-1)!(d-1)!}{(n-2)!} \times \sum_{\beta \subset Q_n, \beta \not\ni i,j} \text{sign}\{\Delta(\beta, i)\} \text{sign}\{\Delta(\beta, j)\}\right].$$

It is now obvious that the statistic $\sum_{\beta \in Q_n} \{\Phi_n(\beta)\}^2$ is actually equivalent to W_n defined in Section 2.2. Randles (1989) introduced the concept of "interdirections" and used it to develop an affine-invariant and distribution-free version of the sign test in multidimension (see also Peters and Randles 1990). For a pair of distinct data points $\mathbf{X}_i$ and $\mathbf{X}_j$, let C_{ij} denote the number of hyperplanes $H(\beta)$'s such that $\beta \in Q_n$, $\beta \not\ni i, j$ and $\mathbf{X}_i$ and $\mathbf{X}_j$ are in opposite sides of $H(\beta)$, so that $\text{sign}\{\Delta(\beta, i)\} \text{sign}\{\Delta(\beta, j)\} < 0$. Then the counts C_{ij}'s are nothing but "interdirections," and $\hat{\psi}_n(i, j)$ is a nonparametric estimate of the geodesic angle between the data points $\mathbf{X}_i$ and $\mathbf{X}_j$ based on "interdirections." Recall now T_n and T_n^* defined in Section 2.2. In the special case $h(\psi) = \cos\psi$, T_n is the Rayleigh's statistic, which is quite well known among analysts of circular and spherical data (see Mardia 1972), and T_n^* turns out to be equivalent to the test statistic used by Randles (1989). Hereafter, we will denote the statistic $\sum_{i=1}^{n} \sum_{j=1}^{n} \cos\{\hat{\psi}_n(i, j)\}$ by R_n. It will be appropriate to note here that Randles (1989) was partly motivated by Blumen's bivariate sign test (Blumen 1958) and applied a finite sample correction to his test so that it coincides with Blumen's test in the bivariate case.

3.2 The Multivariate Extension of Hodges's Sign Test

A natural multivariate version of Hodges's statistic (Hodges 1955) is $\sup_{\lambda \in R^d} |\sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle)|$. Because for a fixed set of observations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, the quantity $|\sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle)|$ must be maximized at some $\lambda \in R^d$ such that $|\lambda| = 1$, one can restate this multivariate test statistic as $\max_{\lambda \in S^{(d-1)}} |\sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle)|$. Now, any unit vector $\lambda \in S^{(d-1)}$ determines a hyperplane $\{\mathbf{x}: \mathbf{x} \in R^d, \langle \lambda, \mathbf{x} \rangle = 0\}$ that passes through the origin in R^d, and $|\sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle)|$ is nothing but the difference between the number of data points that fall in one side of the plane determined by λ and the number of points that fall in the other side. Clearly, in view of the absolute continuity of the underlying distribution generating the observations, any hyperplane in R^d containing the origin can pass through at most $d-1$ data points. Also, any hyperplane through the origin and containing k $(0 < k \le d-1)$ data points can be rotated in appropriate direction(s) to bring it into a new position satisfying the following criteria:

1. The origin remains a fixed point on the plane during and after the rotation.
2. The points that were initially in a particular side of the plane remain together in one side during the rotation and when the plane reaches its final position.

3. After the rotation is complete, all of the k data points, which were originally on the plane, will fall in the side of the plane that initially contained larger number of data points compared to the other side.

In other words, if the unit vectors λ and λ' determine the initial and final positions of the plane respectively, we must have $|\sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle)| = |\sum_{i=1}^{n} \text{sign}(\langle \lambda', \mathbf{X}_i \rangle)| - k$. Similarly, it is easy to see that a hyperplane determined by the unit vector λ and containing the origin and k $(0 \le k < d - 1)$ data points can be rotated into a new position, so that it can be made to contain the origin and $d - 1$ data points including the ones that were originally on the plane. Further, we will have $|\sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle)| = |\sum_{i=1}^{n} \text{sign}(\langle \lambda', \mathbf{X}_i \rangle)| + d - 1 - k$, where the unit vector λ' corresponds to the final position of the rotating plane. Obviously, if we rotate a plane containing the origin as a fixed point on it, until the plane hits a data point that was previously not on the plane, or until a data point that was originally on the plane comes out of it and falls in one of its sides, the quantity $|\sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle)|$ remains unchanged for different values of λ corresponding to different positions of the rotating plane. Combining all these observations, we now have an interesting result.

Fact 3.3. The multivariate extension of Hodges's test statistic satisfies the identity

$$\sup_{\lambda \in R^d} \left| \sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle) \right| = \max_{\beta \in Q_n} \Phi_n(\beta) + d - 1.$$

This fact enables us to visualize the basic geometry behind extended Hodges's statistic. In a sense, this statistic measures the "depth" of the origin within the data cloud in a d-dimensional Euclidean space and uses it to make the decision about the null hypothesis that asserts the origin as a median of the probability distribution generating the data. It is now quite apparent that the sign test based on this statistic has a natural connection with Tukey's "half space median" (Tukey 1975; Donoho 1982). Further, Fact 3.3 provides a nice and convenient algorithm for computing $\sup_{\lambda \in R^d} |\sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle)|$. Combinatorial and algebraic computations using the idea of scanning data points on a circle using semicircular arcs can be found in Ajne (1968), Joffe and Klotz (1962), Klotz (1959), and others. In a similar setup, Rothman (1972) considered the maximum number of observations in an arc of any specified length on a circle. Finally, note that if $\alpha = \{i_1, i_2, \ldots, i_d\} \in S_n$, $\beta = \{i_1, \ldots, i_{k-1}, i_{k+1}, \ldots, i_d\} \in Q_n$, and $1 \le i_1 < i_2 < \cdots < i_d \le n$, it is a matter of straightforward algebra to verify (using the systems of linear equations $\mathbf{X}(\alpha)\mathbf{Y}_i^{(\alpha)} = \mathbf{X}_i$ with $i \notin \alpha$) that $\Phi_n(\beta)$ is nothing but the absolute value of the kth coordinate of $\sum_{i \notin \alpha} \text{sign}(\mathbf{Y}_i^{(\alpha)}) + \xi$, where ξ is the d-dimensional vector with each coordinate equal to 1. In other words, Fact 3.3 can be restated as

$$\sup_{\lambda \in R^d} \left| \sum_{i=1}^{n} \text{sign}(\langle \lambda, \mathbf{X}_i \rangle) \right| = \max_{\alpha \in S_n} \left\| \sum_{i \notin \alpha} \text{sign}(\mathbf{Y}_i^{(\alpha)}) + \xi \right\|_{\infty} + (d - 1),$$

where for $\mathbf{x} = (x_1, x_2, \ldots, x_d) \in R^d$, $\|\mathbf{x}\|_\infty = \max_{1 \le j \le d} |x_j|$.

4. BEHAVIOR OF PROPOSED TESTS: SOME ASYMPTOTICS AND SIMULATIONS

Randles (1989) established that when the null hypothesis is true, the normalized difference between his test statistic and Rayleigh's statistic converges to 0 in probability as the sample size n grows to infinity. Exploiting this critical observation, he showed that the asymptotic null distribution of his statistic is chi-squared with d (= dimension) degrees of freedom after appropriate normalization and worked out the Pitman efficiency of his test under contiguous alternatives for various models. If the kernel h that appears in the definitions of T_n and T_n^* (see Sec. 2.2) has a Fourier expansion of the form $h(\psi) = \sum_{k=1}^{\infty} a_k \cos k\psi$ (see Mardia's discussion of Beran's class of tests in Mardia 1972, pp. 190–191), where the a_k's are nonnegative and $\sum_{k=1}^{\infty} a_k < \infty$, it is easy to see that the conditional expectation $E[h\{\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j) \rangle)\} | \mathbf{X}_j]$ becomes identically a constant for any value of $\mathbf{X}_j$ whenever $U(\mathbf{X}_i)$ and $U(\mathbf{X}_j)$ are two independent and uniformly distributed random vectors on the unit sphere $S^{(d-1)}$. Hence the kernel h is degenerate in such a situation. It follows from a well-known asymptotic property of degenerate U statistics (see Serfling 1980, chap. 5, sec. 5.5.2) that as n tends to infinity, the limiting distribution of $n^{-1}T_n = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} h\{\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j) \rangle)\}$ will be same as that of a weighted sum of independent chi-squared random variables after appropriate centering. In the case $d = 2$, Beran (1969) explicitly derived the limit law of $n^{-1}T_n$, and it turns out to be the distribution of a weighted sum of independent chi-squared variables each with 2 degrees of freedom, where the weights are proportional to the coefficients a_k's that appear in the Fourier expansion of h; see Mardia (1972, p. 193) for some interesting heuristics. We now state a theorem that gives a useful generalization of one of Randles's fundamental observations.

Theorem 4.1. Assume that the $\mathbf{X}_i$'s are iid observations with a common absolutely continuous and spherically symmetric distribution around θ, and, as defined in Section 2.2, let $T_n = \sum_{i=1}^{n} \sum_{j=1}^{n} h\{\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j) \rangle)\}$ and $T_n^* = \sum_{i=1}^{n} \sum_{j=1}^{n} h\{\hat{\psi}_n(i, j)\}$, where h is a continuous function on $[0, \pi]$ satisfying $h(\psi) = -h(\pi - \psi)$. Then, under the hypothesis H_0: $\theta = 0$, the difference $n^{-1}T_n^* - n^{-1}T_n$ tends to 0 in probability as n tends to infinity.

Let us try to see some of the crucial implications of Theorem 4.1. The elliptically symmetric density $|\Sigma|^{-1/2} f\{(\mathbf{x} - \theta)^T \Sigma^{-1} (\mathbf{x} - \theta)\}$ becomes spherically symmetric if and only if $\Sigma = c\mathbf{I}_d$, c being some positive constant and $\mathbf{I}_d$ being the $d \times d$ identity matrix. Recall also that the distribution of T_n^* does not depend on Σ or f when the iid observations $\mathbf{X}_i$'s come from an elliptically symmetric distribution around the origin. Combining this with the preceding theorem, we can now conclude that the asymptotic null distribution of $n^{-1}T_n^*$ under the assumption of elliptic symmetry of the parent distribution is actually same as the limiting distribution of $n^{-1}T_n$ based on iid observations coming from a spherically symmetric distribution around the origin. But note that a test based on T_n^* will be affine-invariant and will have the distribution-free property for arbitrary elliptically symmetric models, whereas a test based on T_n will possess neither of these features.

Table 1. Observed Power (In Percent) at the 5% Level

Test statistic	Value of noncentrality parameter $(\theta^T\Sigma^{-1}\theta)^{1/2}$						
	0	.25	.50	.75	1.00	1.25	1.50
				Normal distribution			
W_n	5.00	11.27	32.80	63.93	88.03	97.27	99.83
R_n	5.00	10.07	28.97	60.57	85.63	96.63	99.47
T^2	5.00	11.50	35.67	68.87	91.93	99.20	≈100
				Exponential power distribution with p = 1.0			
W_n	5.00	6.53	12.90	25.03	42.13	57.57	72.90
R_n	5.00	6.37	11.53	21.47	35.33	51.37	67.47
T^2	4.67	6.13	11.83	23.17	39.53	56.93	72.47
				Exponential power distribution with p = 3.0			
W_n	5.00	85.10	94.10	97.53	98.47	99.10	≈100
R_n	5.00	25.40	84.43	99.87	≈100	≈100	≈100
T^2	6.33	7.10	18.57	29.73	39.30	47.60	55.07
				Cauchy type distribution with q = 2.0			
W_n	5.00	8.57	19.67	36.33	54.47	68.63	79.13
R_n	5.00	6.93	14.93	27.53	41.20	54.33	65.63
T^2	2.77	3.50	5.83	12.03	20.13	28.73	38.37
				Cauchy type distribution with q = 2.5			
W_n	5.00	14.97	46.13	76.27	92.00	97.93	99.37
R_n	5.00	11.13	34.37	62.43	82.83	92.67	96.80
T^2	4.63	8.87	28.70	55.37	76.67	88.00	93.83

It is obvious that the kernels associated with Ajne's A_n statistic (which is equivalent to T_n with $h(\psi) = (\pi/2) - \psi$) and Rayleigh's test statistic (which is equivalent to T_n with $h(\psi) = \cos\psi$) will satisfy the condition $h(\psi) = -h(\pi - \psi)$. As a matter of fact, this condition will hold whenever $h(\psi)$ has a Fourier expansion of the form $h(\psi) = \sum_{k=1}^{\infty} a_{2k-1}\cos(2k - 1)\psi$. In particular, Theorem 4.1 can be used to derive asymptotic approximations for the null distributions of W_n (which is equivalent to T_n^* with $h(\psi) = (\pi/2) - \psi$) and the test statistic R_n (which is nothing but T_n^* with $h(\psi) = \cos\psi$, as described in Sec. 3.1). Exact asymptotic distribution of Ajne's A_n statistic for circular data can be found in Jammalamadaka (1984); Beran (1968) provided some useful approximations for the limiting distribution of the spherical extension of this statistic.

The statistic W_n has a simple and naturally appealing form, as we have observed in Sections 2.2 and 3.1. It is quite easy to compute, which is not true for many of the other T_n^*'s arising from different choices of the kernel function h. One can easily simulate the exact finite sample null distribution of W_n on a computer, and the critical values necessary for implementing the test can be conveniently estimated. In Table 1 we report some simulation results in an attempt to compare W_n with the statistic R_n, which is equivalent to the statistic proposed by Randles (1989), and Hotelling's T^2 statistic. In each of the cases reported, the sample size is 20, and we restricted ourselves to the case $d = 3$. In addition to normal distribution, we have considered elliptically symmetric exponential power distributions, with f having the form $f(\mathbf{z}) = a\exp(-|\mathbf{z}|^p)(p > 0)$, and elliptically symmetric Cauchy-type distributions with $f(\mathbf{z}) = b\{1 + |\mathbf{z}|^2\}^{-q}(q > d/2)$. We used random number generating routines available in IMSL; in each case the observed power is based on the outcomes of 3,000 Monte Carlo replications. Because W_n and R_n are affine-invariant distribution-free statistics, we used simple simulations to estimate their 5% critical values. For the T^2 statistic, the critical value was determined from the F-distribution table. It is quite apparent from the figures in Table 1 that W_n is a clear winner, outperforming both of R_n and T^2 for several nonnormal probability models. Even in the case of normal distribution, when T^2 is the most powerful invariant test, W_n falls behind T^2, but the race between them is quite close.

5. CONCLUDING REMARKS

1. Randles (1989) showed that for several elliptically symmetric models, the Pitman efficiency of his test increases with the dimension d. The multivariate L_1 median $\hat{\theta}_n$, defined as $\sum_{i=1}^{n} |\mathbf{X}_i - \hat{\theta}_n| = \min_{\phi \in R^d} \sum_{i=1}^{n} |\mathbf{X}_i - \phi|$, bears a close connection with Randles's test, and it will be appropriate to note here that the asymptotic efficiency of this location estimate in spherically symmetric models increases as the dimension d grows (see Brown 1983; Chaudhuri 1992). In view of Theorem 4.1 and the discussion preceding it, the statistic $n^{-1}T_n^*$ is approximable by a degenerate U statistic and has a limiting distribution, which is same as that of a weighted sum of independent chi-squared random variables. Interestingly, in the case of Randles's test (i.e., the test with $h(\psi) = \cos\psi$ as the kernel function), the test statistic has a limiting chi-squared distribution with d degrees of freedom (i.e., the weighted sum actually reduces to a single chi-square) after appropriate normalization. As a matter of fact, normalized Rayleigh's statistic, which approximates normalized Randles's statistic under the null hypothesis H_0 as the sample size grows, has the form $n^{-1}T_n = n^{-1}|\sum_{i=1}^{n} U(\mathbf{X}_i)|^2$. This greatly simplifies the efficiency computation for Randles's test via some standard contiguity based analysis. But when the limit law of $n^{-1}T_n$ (and hence that of $n^{-1}T_n^*$) is not so simple, the derivation of asymptotic efficiency of the associated test remains a highly nontrivial unsolved problem.

2. The performance of Ajne's A_n test, Rayleigh's test, Watson's test, and, in general, Beran's class of tests for circular data has been extensively studied in the literature. Jammalamadaka (1984), Jupp and Mardia (1989), and Mardia (1972) have reviewed many results related to the finite sample and the asymptotic behavior of Beran's class of tests. In particular, Mardia (1972, chap. 7, sec. 7.2.5) gave an excellent summary of several optimal power properties of these tests. Though our principal goal in this article is to develop a vision encompassing various sign tests and their interrelations rather than recommending a particular test, it is appropriate to consider the question of how to choose the kernel h. Clearly, the answer to this question requires knowledge of the form of the density f, which is typically an unknown object in practice. In the case of tests for uniform distribution of directional data, it is well known that different T_n's (i.e., statistics in Beran's family) associated with different choices of h will be optimal (in terms of power) for different types of alternative hypotheses (i.e., hypotheses corresponding to different types of nonuniform distributions on circles and spheres). The nature of nonuniformity introduced in the distribution of $U(\mathbf{X}_i)$ due to departure from the null

hypothesis $H_0: \theta = 0$ depends critically on the form of the unknown density f. Recently, Oja and Nyblom (1989) explored asymptotic relative efficiencies of several bivariate sign tests that are affine-invariant and have the distribution-free property.

3. Joffe and Klotz (1962) and Klotz (1959, 1964) have thoroughly investigated the finite sample as well as the asymptotic performance of the bivariate version of Hodges's test (see also Ajne 1968; Bhattacharya and Johnson 1969; Rao 1969). Similar properties of the multivariate version of Hodges's test are yet to be worked out, and Fact 3.3, which expresses the statistic in terms of the maximum of a finite number of simple discrete random variables, can be quite useful in such investigations. But there is a practical problem in using this test statistic that must be pointed out here. When the sample size is not very large, the distribution of this statistic remains concentrated on very few point masses. As a result, an exact implementation of a 5% or a 1% test will require an excessive amount of randomization.

4. One encouraging feature of Table 1 is the monotonicity in the observed power with respect to the noncentrality parameter $\theta^T\Sigma^{-1}\theta$. We have run some simulations with sample sizes other than 20, and such a monotonicity is clearly visible in all cases. It makes one wonder about some formal analytic proofs for this particular characteristic of the power functions of the test statistics when the underlying probability distribution is elliptically symmetric. Along the same line, another interesting open problem is to determine the nature of dependence of the power functions on the parameters p and q, which control the tail behavior and the degree of concentration of masses in their respective distributions.

5. As we have observed in Section 3.2, Hodges's test has a natural connection with Tukey's "half space median" and an associated concept of "data depth" (see Tukey 1975). Liu (1990) introduced the notion of "simplicial data depth" and defined what is called the "simplicial median" for multivariate data. Oja and Nyblom (1989) introduced an affine-invariant bivariate sign test (see the test statistic U_n discussed in Oja and Nyblom 1989, pp. 250–251) and investigated its properties. Interestingly, this test statistic is equivalent to W_n in dimension $d = 2$, and both of them can be viewed as sign tests that naturally correspond to the "bivariate simplicial median." It is a matter of simple and straightforward algebra to verify that in the bivariate case, W_n is equivalent to the statistic that counts the number of triangles (simplices in the two-dimensional Euclidean plane), which are formed with the data points as their vertices and contain the origin $0 \in R^2$ as an interior point. But an analogous result fails to hold in any of the dimensions $d \geq 3$. The implications and the consequences of this intriguing observation are not fully understood at present.

APPENDIX: PROOFS

Proof of Proposition 2.1. Let $\mathbf{A}$ be a nonsingular $d \times d$ matrix and let $\mathbf{Z}_i = \mathbf{A}\mathbf{X}_i$ for $1 \leq i \leq n$. Then it is easy to see that $\mathbf{Z}(\alpha) = \mathbf{A}\mathbf{X}(\alpha)$. Hence $\{\mathbf{Z}(\alpha)\}^{-1}\mathbf{Z}_i = \{\mathbf{X}(\alpha)\}^{-1}\mathbf{A}^{-1}\mathbf{A}\mathbf{X}_i = \{\mathbf{X}(\alpha)\}^{-1}\mathbf{X}_i$ for all $1 \leq i \leq n$. This ensures the invariance of the $\mathbf{Y}_i^{(\alpha)}$'s under nonsingular linear transformations. Also, for two sets of data points $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$ and $\{\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n\}$, if we have $\{\mathbf{X}(\alpha)\}^{-1}\mathbf{X}_i = \{\mathbf{Z}(\alpha)\}^{-1}\mathbf{Z}_i$ for all $i \notin \alpha$, then we automatically have $\mathbf{X}_i = \mathbf{X}(\alpha)\{\mathbf{Z}(\alpha)\}^{-1}\mathbf{Z}_i$ for all i such that $1 \leq i \leq n$. Note that the equation is trivially true for $i \in \alpha$. Therefore, the data set $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$ is obtainable from $\{\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n\}$ by a nonsingular linear transformation and vice versa. This completes the proof by establishing the maximality of the invariant coordinates.

Proof of Theorem 2.2. Using the positive definiteness of Σ, let us write $\mathbf{Z}_i = \Sigma^{-1}\mathbf{X}_i$ for all $1 \leq i \leq n$. Then the $\mathbf{Z}_i$'s will be iid observations with a common spherically symmetric distribution. Define $r_i = |\mathbf{Z}_i|$ and $\mathbf{E}_i = \mathbf{Z}_i r_i^{-1}$. Clearly, when $\theta = 0$, the $\mathbf{E}_i$'s become iid random vectors that are uniformly distributed on the unit sphere $S^{(d-1)}$ in R^d. For $\alpha \in S_n$, we will denote by $\mathbf{E}(\alpha)$ the $d \times d$ matrix whose columns are the vectors $\mathbf{E}_i$'s and will denote by $\mathbf{R}(\alpha)$ the $d \times d$ diagonal matrix with diagonal entries r_i's, where $i \in \alpha$. Now in view of the definition of the basis matrix $\mathbf{X}(\alpha)$, we have $\mathbf{X}(\alpha) = \Sigma^{1/2}\mathbf{E}(\alpha)\mathbf{R}(\alpha)$. At this point, if we fix an $\alpha \in S_n$, an index i such that $i \notin \alpha$, and then look at the transformed observation $\mathbf{Y}_i^{(\alpha)}$, we get $\mathbf{Y}_i^{(\alpha)} = \{\mathbf{X}(\alpha)\}^{-1}\mathbf{X}_i = \{\mathbf{R}(\alpha)\}^{-1}\{\mathbf{E}(\alpha)\}^{-1}\Sigma^{-1/2} \times \Sigma^{1/2}\mathbf{E}_i r_i = \{\mathbf{R}(\alpha)\}^{-1}\{\mathbf{E}(\alpha)\}^{-1}\mathbf{E}_i r_i$. This immediately implies that $\text{sign}(\mathbf{Y}_i^{(\alpha)}) = \text{sign}(\{\mathbf{E}(\alpha)\}^{-1}\mathbf{E}_i)$, using the fact that each of the r_i's is positive. Because under H_0 the joint distribution of the $\mathbf{E}_i$'s is completely free from Σ and f, the same must be true for the joint distribution of $\text{sign}(\mathbf{Y}_i^{(\alpha)})$'s.

Proof of Theorem 4.1. Using the condition $h(\psi) = -h(\pi - \psi)$ and the idea in lemma A.1 of Randles (1989), we first conclude that

$$E_{H_0}(n^{-1}T_n^* - n^{-1}T_n)^2$$

$$= n^{-2}E_{H_0}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}[h\{\hat{\psi}_n(i,j)\} - h\{\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j)\rangle)\}]\right]^2$$

$$= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}E_{H_0}[h\{\hat{\psi}_n(i,j)\} - h\{\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j)\rangle)\}]^2.$$

Then, in view of the iid nature of the $\mathbf{X}_i$'s assumed in the theorem and the fact that $h\{\hat{\psi}_n(i,i)\} = h\{\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_i)\rangle)\} = h(0)$ for $1 \leq i \leq n$, we must have

$$E_{H_0}(n^{-1}T_n^* - n^{-1}T_n)^2$$

$$= n^{-1}(n-1)E_{H_0}[h\{\hat{\psi}_n(i,j)\} - h\{\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j)\rangle)\}]^2$$

for any pair of distinct indices i and j. On the other hand, it is quite easy to see (see, for example, Randles 1989) that the difference $\hat{\psi}_n(i,j) - \cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j)\rangle)$, where i and j are fixed indices such that $1 \leq i \neq j \leq n$, will converge to 0 in probability as n tends to infinity if the hypothesis $H_0: \theta = 0$ is true. Hence an application of Lebesgue's dominated convergence theorem using the continuity of h yields

$$\lim_{n\to\infty} E_{H_0}[h\{\hat{\psi}_n(i,j)\} - h\{\cos^{-1}(\langle U(\mathbf{X}_i), U(\mathbf{X}_j)\rangle)\}]^2 = 0.$$

Consequently, we must have $\lim_{n\to\infty} E_{H_0}(n^{-1}T_n^* - n^{-1}T_n)^2 = 0$. This completes the proof of the theorem.

[Received August 1991. Revised December 1992.]

REFERENCES

Ajne, B. (1968), "A Simple Test for Uniformity of a Circular Distribution," *Biometrika*, 55, 343–354.

Bennett, B. M. (1962), "On Multivariate Sign Tests," *Journal of the Royal Statistical Society*, Ser. B, 24, 159–161.

Beran, R. J. (1968), "Testing for Uniformity on a Compact Homogeneous Space," *Journal of Applied Probability*, 5, 177–195.

——— (1969), "Asymptotic Theory of a Class of Tests for Uniformity of a Circular Distribution," *Annals of Mathematical Statistics*, 40, 1196–1206.

Bhattacharya, G. K., and Johnson, R. A. (1969), "On Hodges's Bivariate Sign Test and a Test for Uniformity of a Circular Distribution," *Biometrika*, 56, 446–449.

Bickel, P. J. (1965), "On Some Asymptotically Nonparametric Competitors of Hotelling's T^2," *Annals of Mathematical Statistics*, 36, 160–173.

Blumen, I. (1958), "A New Bivariate Sign Test," *Journal of the American Statistical Association*, 53, 448–456.

Brown, B. M. (1983), "Statistical Use of the Spatial Median," *Journal of the Royal Statistical Society*, Ser. B, 45, 25–30.

Brown, B. M., and Hettmansperger, T. P. (1989), "An Affine Invariant Bivariate Version of the Sign Test," *Journal of the Royal Statistical Society*, Ser. B, 51, 117–125.

Brown, B. M., Hettmansperger, T. P., Nyblom, J., and Oja, H. (1992), "On Certain Bivariate Sign Tests and Medians," *Journal of the American Statistical Association*, 87, 127–135.

Chatterjee, S. K. (1966), "A Bivariate Sign Test for Location," *Annals of Mathematical Statistics*, 37, 1771–1782.

Chaudhuri, P. (1992), "Multivariate Location Estimation Using Extension of *R* Estimates Through *U* Statistics-Type Approach," *The Annals of Statistics*, 20, 897–916.

Dietz, E. J. (1982), "Bivariate Nonparametric Tests for the One-Sample Location Problem," *Journal of the American Statistical Association*, 77, 163–169.

Donoho, D. (1982), "Breakdown Properties of Multivariate Locations Estimators," unpublished Ph.D. qualifying paper, Harvard University, Dept. of Statistics.

Haldane, J. B. S. (1948), "Note on the Median of a Multivariate Distribution," *Biometrika*, 35, 414–415.

Hodges, J. L. (1955), "A Bivariate Sign Test," *Annals of Mathematical Statistics*, 26, 523–527.

Jammalamadaka, S. R. (1984), "Nonparametric Methods in Directional Data Analysis," in *Handbook of Statistics*, Vol. 4, eds. P. R. Krishnaiah and P. K. Sen, Amsterdam: Elsevier, pp. 755–770.

Joffe, A., and Klotz, J. (1962), "Null Distribution and Bahadur Efficiency of the Hodges Bivariate Sign Test," *Annals of Mathematical Statistics*, 33, 803–807.

Jupp, P. E., and Mardia, K. V. (1989), "A Unified View of the Theory of Directional Statistics, 1975–1988," *International Statistical Review*, 57, 261–294.

Klotz, J. (1959), "Null Distribution of the Hodges Bivariate Sign Test," *Annals of Mathematical Statistics*, 30, 1029–1033.

——— (1964), "Small Sample Power of the Bivariate Sign Tests of Blumen and Hodges," *Annals of Mathematical Statistics*, 33, 1576–1582.

Liu, R. (1990), "On a Notion of Data Depth Based on Random Simplices," *The Annals of Statistics*, 18, 405–414.

Mardia, K. V. (1972), *Statistics of Directional Data*, London: Academic Press.

Oja, H., and Nyblom, J. (1989), "Bivariate Sign Tests," *Journal of the American Statistical Association*, 84, 249–259.

Peters, D., and Randles, R. H. (1990), "A Multivariate Signed Rank Test for the One-Sample Location Problem," *Journal of the American Statistical Association*, 85, 552–557.

Puri, M. L., and Sen, P. K. (1971), *Nonparametric Methods in Multivariate Analysis*, New York: John Wiley.

Randles, R. H. (1989), "A Distribution-Free Multivariate Sign Test Based on Interdirections," *Journal of the American Statistical Association*, 84, 1045–1050.

Rao, J. S. (1969), "Some Contributions to the Analysis of Circular Data," unpublished Ph.D. dissertation, Indian Statistical Institute, Calcutta.

Rothman, E. D. (1972), "Tests of Uniformity for Circular Distributions," *Sankhyā*, Ser. A, 34, 23–32.

Roy, S. N. (1953), "On a Heuristic Method of Test Construction and Its Use in Multivariate Analysis," *Annals of Mathematical Statistics*, 24, 220–238.

——— (1957), *Some Aspects of Multivariate Analysis*, New York: John Wiley.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: John Wiley.

Small, C. G. (1990), "A Survey of Multidimensional Medians," *International Statistical Review*, 58, 263–277.

Tukey, J. W. (1975) "Mathematics and Picturing Data," in *Proceedings of International Congress of Mathematics, Vancouver*, Vol. 2, pp. 523–531.

Watson, G. S. (1961), "Goodness of Fit Tests on a Circle," *Biometrika*, 48, 109–114.