Electronic Journal of Statistics
Vol. 6 (2012) 760–782
ISSN: 1935-7524
DOI: 10.1214/12-EJS692

A nonparametric multivariate multisample test based on data depth

Shojaeddin Chenouri[*†] **and Christopher G. Small**[‡]

Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada,
e-mail: schenouri@uwaterloo.ca; cgsmall@uwaterloo.ca

Abstract: In this paper, we construct a family of nonparametric multivariate multisample tests based on depth rankings. These tests are of Kruskal-Wallis type in the sense that the samples are variously ordered. However, unlike the Kruskal-Wallis test, these tests are based upon a depth ranking using a statistical depth function such as the halfspace depth or the Mahalanobis depth, etc. The types of tests we propose are adapted to the depth function that is most appropriate for the application. Under the null hypothesis that all samples come from the same distribution, we show that the test statistic asymptotically has a chi-square distribution. Some comparisons of power are made with the Hotelling T^2, and the test of Choi and Marden (1997). Our test is particularly recommended when the data are of unknown distribution type where there is some evidence that the density contours are not elliptical. However, when the data are normally distributed, we often obtain high relative power.

Keywords and phrases: Data depth, multivariate nonparametric tests, Kruskal-Wallis test, depth-depth plot.

Received April 2011.

1. Introduction

Generalizations of the univariate procedures based on ranks and signs for testing the equality of two or more populations to multivariate framework have been an interesting and important problem in statistics. Over the years several generalizations have been proposed. Most papers extend univariate rank procedures to multivariate problems using componentwise signs or ranks.(e.g. Bennett (1962), Bickel (1965), Chatterjee (1966), Puri and Sen (1971)). These procedures are not affine invariant, or even rotationally invariant, and, as Bickel (1965) pointed out, performance can be low relative to the normal theory tests if the components are highly correlated. Subsequent papers developed rotationally invariant and affine invariant multivariate sign, signed rank, and rank tests. See Blumen (1958), Brown and Hettmansperger (1987), Brown and Hettmansperger (1989), Chaudhuri and Sengupta (1993), Choi and Marden (1997), Dietz (1982),

*Corresponding author.
†Research supported by a grant from the Natural Sciences and Engineering Research Council of Canada.
‡Research supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Hettmansperger et al. (1994), Hettmansperger and Oja (1994), Hettmansperger et al. (1997), Hettmansperger et al. (1998), Hodges (1955), Hössjer and Croux (1995), Liu and Singh (1993), Möttönen et al. (2003), Möttönen and Oja (1995), Oja (1999), Oja and Nyblom (1989), Peters and Randles (1990), Peters and Randles (1991), Randles (1989), Randles (2000), Randles and Peters (1990), Um and Randles (1998), Zuo and He (2006).

In Section 2 we discuss briefly different types of existing multivariate multisample tests based on different types of rank vectors. Section 3 is devoted to a discussion on the different notions of data depth. The depth ranking based on data depth is given in section 4. In Section 5 we construct a new multivariate multisample test based on the depth ranking discussed in Section 4 and investigate its properties. In Section 6.1, via simulation, we investigate the asymptotic distribution of the test statistic, and in Section 6.2, we compare power of the test with some of the other tests, including generalized T^2 Hotelling. Finally, in Section 7, we apply the proposed test to a real data set.

2. Multivariate Kruskal-Wallis type tests

Consider comparing t absolutely continuous multivariate distributions F_k, $k = 1, 2, \ldots, t$. The hypothesis to be tested, say H_0, specifies that

$$H_0 : \; F_1(\mathbf{x}) = F_2(\mathbf{x}) = \cdots = F_t(\mathbf{x}) \qquad \text{for all } \mathbf{x}\,.$$

Under H_0, the common distribution function shall be represented by F. The alternative hypothesis H_1 to H_0 says that H_0 does not hold. For $j = 1, \ldots, t$, assume that $\mathbf{X}_{ij}, i = 1, \ldots, n_j$, denotes a random sample of size n_j from a d-variate population F_j.

For the assumption that the distributions are d-variate normal with common unknown covariance matrix Σ and different mean vectors, Lawley (1938) and Hotelling (1951) introduced an affine equivariant test statistic which is well known as the generalized Hotelling's T^2. Hotelling (1951) showed that the asymptotic null distribution of Hotelling's T^2 is chi-square with $d(t-1)$ degree of freedom.

The standard univariate and *nonparametric* test for one way analysis of variance is the Kruskal-Wallis test (Kruskal (1952), Kruskal and Wallis (1952)). Puri and Sen (1971) proposed a multivariate extension of the Kruskal-Wallis test based on a component-wise ranking. However, unlike the Hotelling's T^2, it is not affine equivariant. It is only invariant under the coordinate-wise monotone transformation. Thus its performance will depend on the form of the covariance matrix and the direction of shift. The efficiency of the test based on the component-wise ranks becomes really poor in the case of highly correlated components. See Bickel (1965) for a general discussion on the procedures based on the component-wise ordering.

Choi and Marden (1997) developed a multivariate multi-sample test based on average gradient vectors of the Euclidean norm. For each d-variate observation

$\mathbf{X}_i$, its rank vector is defined by

$$\mathbf{R}(\mathbf{X}_i) = \frac{1}{n}\sum_{j\neq i}^{n} \frac{\mathbf{X}_i - \mathbf{X}_j}{||\mathbf{X}_i - \mathbf{X}_j||}\,.$$

Choi and Marden's test statistic based on gradients of the Euclidean norm are not affine equivariant, but they are equivariant under orthogonal transformations. As discussed in Brown (1983), and Choi and Marden (1997), the efficiency of a test based on these gradients is increasing with the dimension in the case of the multivariate spherical distributions. Re-scaling one of the components, i.e. moving from a spherical case to an elliptical case, may however highly reduce the efficiency (see Brown (1983), and Chakraborty et al. (1998)).

Hettmansperger et al. (1998) constructed an affine equivariant asymptotically distribution free multivariate multi-sample test by introducing a multivariate centered rank function based on the Oja (1983) criterion function and discussed its properties. For each d-variate observation $\mathbf{X}_i$, Oja's centered rank vector is defined by

$$\mathbf{R}(\mathbf{X}_i) = d!\binom{N}{d}^{-1} \sum_{i_1<\cdots<i_d} \bigtriangledown \text{volume}(\mathbf{X}_i, \mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_d})\,,$$

where $\bigtriangledown$ is the gradient operator, and $\text{volume}(\mathbf{x}, \mathbf{x}_1, \ldots, \mathbf{x}_d)$ is the volume of the simplex with vertices $\mathbf{x}$, $\mathbf{x}_1, \ldots, \mathbf{x}_d$ in $\mathbb{R}^d$.

Um and Randles (1998) developed an affine equivariant nonparametric multivariate multisample test based on *interdirections*. Interdirections were introduced in Randles (1989), Peters and Randles (1990), and Randles and Peters (1990) to construct distribution free multivariate signed, signed-rank, and two sample tests. Interdirections measure the angular distance between two observation vectors relative to the rest of the data. The tests based on interdirections are distribution free over the class of all elliptically symmetric distributions.

When there are only two populations, say F_1 and F_2, Liu and Singh (1993) introduced a quality index, $Q(F_1, F_2)$, based on data depth to measure the overall "outlyingness" of population F_2 relative to population F_1. They showed that the empirical quality index $Q(\widehat{F}_1, \widehat{F}_2)$ produces a Wilcoxon type rank sum testing procedure to test equality of F_1 and F_2 when they belong to a location and scale family. The asymptotic distribution of this rank sum statistic, which was conjectured in Liu and Singh (1993), is proved in Zuo and He (2006).

The tests listed above are distribution free for large samples. In small sample cases, the tests are conditionally distribution free, or distribution free over a class of elliptically symmetric distributions. In Section 5, we will construct another asymptotically, as well as conditionally distribution free multivariate multi-sample test. In addition, this test can be constructed so as to be affine equivariant. The test statistic has an asymptotic chi-square distribution with $t-1$ degrees of freedom in many standard settings. A graphical technique to compare two distributions known as the depth-depth plot, or DD-plot in short, is introduced in Liu et al. (1999).

In view of the large number of available tests, the introduction of another one needs some explanation. Our multisample test can be viewed as a formal multisample extension of the DD-plots of Liu et al. (1999). Our family of tests also has several desirable features. First, it allows the researcher to flexibly choose a depth function for the analysis that most appropriately combines the advantages of equivariance, robustness and computational convenience. In other words, there is no one prescription for all contexts, but rather an overall approach that can be tailored to the context. Secondly, the test is based upon ranking multidimensional data. Rank statistics have the advantage that they are essentially dimensionless, being positive integer values. This property of ranking is especially useful if a data set is pre-processed before analysis by a dimension reduction technique such as PCA. The effects of the PCA on the ranks of the data can be studied directly because they are dimension-free. Vector ranks are not dimensionless, making the effects of PCA and other dimension-changing methods more difficult to interpret. Like other generalizations of the Kruskal-Wallis test, the proposed multivariate depth rank based tests are sensitive to location shifts and also sensitive to other departures in varying degrees.

3. Statistical depth functions

Let $\mathcal{F}$ be the class of all (Borel measurable) distributions on $\mathbb{R}^d$. In the definitions of some depth functions below, it will be assumed that certain moments exist. In these case, an additional appropriate restriction must be placed on $\mathcal{F}$ to make the definition meaningful.

A typical depth function will be a bounded nonnegative function of the form $D : \mathbb{R}^d \times \mathcal{F} \to \mathbb{R}$. For any random vector $\mathbf{Y}$, let $F_{\mathbf{Y}}$ denote the distribution of $\mathbf{Y}$. A depth function D will be said to be *affine equivariant* if $D(A\mathbf{x} + \mathbf{b}, F_{A\mathbf{X}+\mathbf{b}}) = D(\mathbf{x}, F_{\mathbf{X}})$ holds for any random vector $\mathbf{X}$ in $\mathbb{R}^d$, any $d \times d$ nonsingular matrix A, and any vectors $\mathbf{x}$ and $\mathbf{b}$ in $\mathbb{R}^d$. In most, but not all cases, we shall expect $D(\mathbf{x}, F)$ to be a *quasi-concave* function of $\mathbf{x}$, maximised somewhere in the center of the distribution F and vanishing at infinity. The sample version of $D(\,\cdot\,, F)$ is denoted by $D(\mathbf{x}, \widehat{F})$, where $\widehat{F}$ is the empirical distribution based on the sample. Let $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$ be a random sample from $d-$dimensional distribution F, and $\widehat{F}$ its empirical distribution. We use P_F to represent the probability measure corresponding to the distribution F, and E_F to represent expectation with respect to that probability measure. A typical depth function will be continuous in the sense that $D(\mathbf{x}, \widehat{F}) \to D(\mathbf{x}, F)$ whenever the sample size goes to infinity under random sampling from F. However, if empirical convergence is simply replaced by weak convergence, many depth functions are not continuous in this (now stronger) sense. We may loosely define the *robustness* of a depth function to be a measure of its continuity under various types of convergence of $\widehat{F}$ to F. The more that D is continuous under a variety of types of convergence, the more robust D is.

The *Mahalanobis depth* is perhaps the most common depth function associated with multivariate normal theory. Following Rao (1988), for a positive

definite $d \times d$ matrix M, we shall define the *Mahalanobis norm* (Mahalanobis (1936)) $||\cdot||_M$ as

$$||\mathbf{x}||_M = \sqrt{\mathbf{x}' M^{-1} \mathbf{x}}, \quad \text{for all } \mathbf{x} \in \mathbb{R}^d. \tag{1}$$

The Mahalanobis depth is defined as

$$MD(\mathbf{x}, F) = \frac{1}{1 + ||\mathbf{x} - \mu(F)||^2_{\Sigma(F)}},$$

where F is a given distribution and $\mu(F)$ and $\Sigma(F)$ are any corresponding location and covariance measures, respectively. The Mahalanobis depth $MD(\cdot, \widehat{F})$ is not robust. Robust alternatives to the usual mean vector and covariance matrix are available such as Rousseeuw's minimum covariance determinant (MCD), or the minimum volume ellipsoid (MVE) estimates (see Rousseeuw (1983) and Rousseeuw and Leroy (1987)). If the robustness of the Mahalanobis depth is desired, we can define a robust version $RMD(\cdot, \widehat{F})$ by using these robust estimates.

The *spatial depth* (Gower (1974), Brown (1983)) of a point $\mathbf{x}$ in $\mathbb{R}^d$ is defined as

$$SD(\mathbf{x}, F) = \frac{1}{1 + E_F||\mathbf{x} - \mathbf{X}||}. \tag{2}$$

It is easy to see that $SD(\mathbf{x}, F)$ is invariant under orthogonal transformations. A modification of the Euclidean norm used to define the spatial depth function yields an affine invariant version. The spatial depth function may be modified to an affine invariant version,

$$ASD(\mathbf{x}, F) = \frac{1}{1 + E_F[||\mathbf{x} - \mathbf{X}||_{\Sigma(F)}]}, \tag{3}$$

where $\Sigma(F)$ is the covariance matrix of F and $||\cdot||_{\Sigma(F)}$ denotes the Mahalanobis norm given by (1). Once again, if the robustness of the affine spatial depth is desired, we can define a robust version $RASD(\mathbf{x}, \widehat{F})$ by replacing $\widehat{\Sigma}$ with a robust estimate such as the MCD, or MVE.

Another depth function which is used in this paper is the *halfspace depth*. The halfspace depth (Hodges (1955), Tukey (1975), Donoho (1982)) at a point $\mathbf{x} \in \mathbb{R}^d$ with respect to F is defined to be

$$HSD(\mathbf{x}, F) = \inf_{\mathbf{u} \in \mathbb{R}^d} P_F(H_{\mathbf{x},\mathbf{u}}) = \inf_{||\mathbf{u}||=1} P_F(H_{\mathbf{x},\mathbf{u}}),$$

where $H_{\mathbf{x},\mathbf{u}} = \{\mathbf{y} \in \mathbb{R}^d; \mathbf{u}'\mathbf{y} \geq \mathbf{u}'\mathbf{x}\}$ is a closed halfspace. Several properties of the halfspace depth discussed by Donoho and Gasko (1992), Rousseeuw and Ruts (1999), Small (1987), Zuo and Serfling (2000). The computational issues are discussed in Rousseeuw and Ruts (1996), Rousseeuw and Ruts (1998), Rousseeuw and Struyf (1998), and Ruts and Rousseeuw (1996).

For an overview on other depth functions, we refer the reader to Small (1990) and Zuo and Serfling (2000)

4. Depth based ranking

For a univariate sample the ranking of data is clear and unambiguous. If $X_1, \ldots, X_n$ is a random sample from a continuous random variable X, we can place them in increasing order, as $X_{(1)} < \cdots < X_{(n)}$. The rank of X_i is the number of data points less than or equal X_i, that is

$$R(X_i) = \#\{X_j : \ X_j \leq X_i\}\,,$$

where $\#A$ represents the cardinality or the number of elements in the set A. These are *linear ordering and ranking*, respectively. An alternative method of ordering and ranking the sample is to order the observations in relation to their absolute deviation, or distance, from some reference point, M. If $M < X_{(1)}$ the ordering is the same as that just described. If M is in the body of the data, e.g. at the *median*, quite a different ordering arises. Such *distance ranking* (Barnett (1976)) is seldom directly considered for univariate data, but it has an appeal in the multivariate context.

For a d-variate ($d \geq 2$) sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$, there is no natural linear ranking, but one can use the notion of data depth to introduce depth ranking. Given a depth function D and a distribution function F, one can compute the depths of all the sample points $\mathbf{X}_1, \ldots, \mathbf{X}_n$, namely

$$D(\mathbf{X}_1,\ F), \ldots, D(\mathbf{X}_n,\ F)$$

and order them according to increasing depth values. This gives a depth ranking of the sample points. The rank of $\mathbf{X}_i$ is

$$R(\mathbf{X}_i) = \#\{j\,;\, D(\mathbf{X}_j,\ F) \leq D(\mathbf{X}_i,\ F)\}. \tag{4}$$

The implication of (4) is that a *larger rank* is associated with a *more central position* with respect to the data cloud.

With the exception of Mahalanobis depth and spatial depth, many empirical depth functions can have ties in ranks. Traditional solutions to ties in ranks are available here and will not affect the asymptotic properties which will be investigated later. It is necessary that for large sample sizes, the proportion of data tied at a given value goes to zero. We may choose to break ties at random or to assign an average rank simultaneously to all tied values. Both methods have merits and difficulties that we shall not pursue here. We refer the reader to Lehmann and D'abrera (2006), and Hollander and Wolfe (1999) for the theory of averaged ranks.

Obviously, different distribution functions yield different rankings. Of particular interest is the depth-based ranking using the empirical depth $D(\,\cdot\,, \widehat{F})$. Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is a random sample from some unknown distribution F. We shall write R_i and R_i^* for the ranks of $\mathbf{X}_i$ based upon the empirical depth $D(\,\cdot\,, \widehat{F})$ and theoretical depth $D(\,\cdot\,, F)$, respectively. We naturally expect R_i to be an empirical approximation to R_i^*. However, this will not be true without some additional regularity, which we shall discuss later.

5. A depth based Kruskal-Wallis type test

Liu et al. (1999) introduced the depth-depth (DD) plot as a two dimensional graphical technique for visual comparison of two d-dimensional distributions based on their respective samples. They showed that different distributional differences, such as difference in location, scale, skewness and kurtosis can be associated with different patterns in DD-plots. Koshevoy (2001), and Struyf and Rousseeuw (1999) showed that some depth functions characterize distributions. These results motivate formal depth based test statistics for comparing two or more multivariate distributions.

We can tabulate data as in Table 1.

As earlier, under H_0 we shall let the common distribution to be F. Let $\widehat{F}_j$ denote the empirical distribution of the j-th sample $\mathcal{X}_j = \{\mathbf{X}_{1\,j}, \ldots, \mathbf{X}_{n_j\,j}\}$, so that $\widehat{F} = n^{-1}\sum_{j=1}^{t} n_j\,\widehat{F}_j$, where $n = \sum n_j$. Also, $\mathcal{X} = \bigcup_{j=1}^{t} \mathcal{X}_j$. Let $D_{ij} = D(\mathbf{X}_{ij}, \widehat{F})$ be the depth of $\mathbf{X}_{ij}$ for $i = 1, 2, \ldots, n_j$ and $j = 1, 2, \ldots, t$ with respect to $\widehat{F}$. Similarly, let $D_{ij}(k) = D(\mathbf{X}_{ij}\,, \widehat{F}_k)$. Finally, write $D^*_{ij} = D(\mathbf{X}_{ij}, F)$.

Let R_{ij} be the depth rank of $\mathbf{X}_{ij}$ with respect to the depth function $D(\,\cdot\,, \widehat{F})$. These ranks are summarized in Table 2.

Under the null hypothesis $H_0 : F_1 = \cdots = F_t = F$, assuming no ties, we have

$$\Pr_{H_0}(R_{ij} = r) = \begin{cases} n^{-1} & \text{if} \quad r = 1, \ldots, n \\ 0 & \text{otherwise,} \end{cases}$$

and for $(i,\, j) \neq (i',\, j')$

$$\Pr_{H_0}(R_{ij} = r, R_{i'\,j'} = r') = \begin{cases} \frac{1}{n(n-1)} & \text{if} \quad r \neq r' \\ 0 & \text{otherwise.} \end{cases}$$

TABLE 1
Data from t d-variate distributions

Treatments			
1	2	...	t
$\mathbf{X}_{11}$	$\mathbf{X}_{12}$	...	$\mathbf{X}_{1t}$
$\mathbf{X}_{21}$	$\mathbf{X}_{22}$	...	$\mathbf{X}_{2t}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\mathbf{X}_{n_1 1}$	$\mathbf{X}_{n_2 2}$	...	$\mathbf{X}_{n_t t}$

TABLE 2
Depth ranks based on pooled sample

Treatments			
1	2	...	t
R_{11}	R_{12}	...	R_{1t}
R_{21}	R_{22}	...	R_{2t}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$R_{n_1 1}$	$R_{n_2 2}$	...	$R_{n_t t}$

Now let

$$R_{\cdot j} = \sum_{i=1}^{n_j} R_{ij} \qquad \text{and} \qquad \overline{R}_{\cdot j} = \frac{R_{\cdot j}}{n_j}, \tag{5}$$

so it can be easily shown that

$$\begin{aligned} E_{H_0}(R_{\cdot j}) &= \frac{n_j(n+1)}{2}, \\ Var_{H_0}(R_{\cdot j}) &= \frac{n_j(n-n_j)(n+1)}{12}, \\ Cov_{H_0}(R_{\cdot j}, R_{\cdot j'}) &= \frac{-n_j n_{j'}(n+1)}{12}, \qquad j \neq j'. \end{aligned}$$

Now consider

$$\begin{aligned} K_j &= \frac{\left[\overline{R}_{\cdot j} - E_{H_0}(\overline{R}_{\cdot j})\right]^2}{Var_{H_0}(\overline{R}_{\cdot j})} \\ &= \frac{12 n_j}{(n-n_j)(n+1)} \left(\overline{R}_{\cdot j} - \frac{n+1}{2}\right)^2. \end{aligned} \tag{6}$$

Similar to the univariate Kruskal-Wallis test, we consider the test statistic to be the weighted average of K_j's, i.e.

$$\begin{aligned} K &= \sum_{j=1}^{t} \left(1 - \frac{n_j}{n}\right) K_j \\ &= \frac{12}{n(n+1)} \sum_{j=1}^{t} \frac{R_{\cdot j}^2}{n_j} - 3(n+1). \end{aligned} \tag{7}$$

Note that, under the null hypothesis K is distribution free. Unfortunately, the test based on K is not powerful against location shift alternatives. See Chenouri (2004), Liu and Singh (2006), and Chenouri et al. (2011). In order to construct a powerful test for location shift alternatives, we must consider ranking methods which have better sensitivity when samples are separated by a location shift. As an alternative, let us consider the depths $D_{ij}(k) = D(\mathbf{X}_{ij}, \widehat{F}_k)$ and ranks $R_{ij}(k)$ for $i = 1, 2, \ldots, n_j$ and $j, k = 1, 2, \ldots, t$, as described above. Table 3 summarizes the ranks $R_{ij}(k)$, for any $k = 1, \ldots, t$.

TABLE 3
Depth ranks based on the sample k

Treatments			
1	2	...	t
$R_{11}(k)$	$R_{12}(k)$	...	$R_{1t}(k)$
$R_{21}(k)$	$R_{22}(k)$	...	$R_{2t}(k)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$R_{n_1 1}(k)$	$R_{n_2 2}(k)$	...	$R_{n_t t}(k)$

Unlike the ranks R_{ij}, the ranks $R_{ij}(k)$ have some power to distinguish the observations in sample k from those in sample j where $j \neq k$. Thus it seems reasonable to construct a test statistic $H(k)$, say, based upon the ranks $R_{ij}(k)$ in a manner analogous to the construction of K as in formulas (6) and (7) above. We can then try to pool the test statistics $H(1), \ldots, H(t)$ together so as to create some overall test statistic for the null hypothesis that all samples have the same distribution. This suggests that we define

$$H_j(k) = \frac{12n_j}{(n-n_j)(n+1)} \left(\bar{R}_{\cdot j}(k) - \frac{n+1}{2} \right)^2 \tag{8}$$

by analogy with formula (6). Then

$$\begin{aligned} H(k) &= \sum_{j=1}^{t} \left(1 - \frac{n_j}{n}\right) H_j(k) \\ &= \frac{12}{n(n+1)} \sum_{j=1}^{t} n_j \left(\bar{R}_{\cdot j}(k) - \frac{n+1}{2} \right)^2 \end{aligned} \tag{9}$$

could be considered as a test statistic, but it depends on the choice of k. To overcome this problem, define

$$\begin{aligned} H &= \frac{1}{t} \sum_{k=1}^{t} H(k) \\ &= \frac{12}{n(n+1)t} \sum_{k=1}^{t} \sum_{j=1}^{t} n_j \left(\bar{R}_{\cdot j}(k) - \frac{n+1}{2} \right)^2 \\ &= \frac{12}{n(n+1)t} \sum_{k=1}^{t} \sum_{j=1}^{t} \frac{R_{\cdot j}^2(k)}{n_j} - 3(n+1)\,. \end{aligned} \tag{10}$$

The null hypothesis is to be rejected when the value of the test statistic H is very large.

Now we shall investigate the asymptotic behavior of the test statistic H under the null hypothesis. We will need some regularity conditions which we shall now consider.

Assumption 1. *Let G be any distribution, and let $\widehat{G}$ be the empirical distribution based upon any random sample of size m from G. As m goes to infinity, $D(\mathbf{x}, \widehat{G})$ converges to $D(\mathbf{x}, G)$ uniformly in $\mathbf{x} \in \mathbb{R}^d$ in probability in the sense that*

$$\sup_{\mathbf{x} \in \mathbb{R}^d} \left| D(\mathbf{x}, \widehat{G}) - D(\mathbf{x}, G) \right| \xrightarrow{p} 0$$

as $n \to \infty$.

In particular, we may have $m = n$ or n_k, and $\widehat{G} = \widehat{F}$ or $\widehat{F}_k$, respectively. This assumption is true for many depth functions such as the spatial, Mahalanobis, halfspace, simpilicial, projection depths.

Recall that as $n_k \to \infty$, $k = 1, \ldots, t$ (and thus $n \to \infty$), the Glivenko-Cantelli theorem (see Pollard 1984) implies that

$$\sup_{\mathbf{x} \in \mathbb{R}^d} |\widehat{F}_k(\mathbf{x}) - \widehat{F}(\mathbf{x})| \xrightarrow{a.s.} 0\,.$$

Therefore under the null hypothesis, from the Assumption 1 we have

$$\sup_{\mathbf{x} \in \mathbb{R}^d} \left| D(\mathbf{x}, \widehat{F}) - D(\mathbf{x}, \widehat{F}_k) \right| \xrightarrow{p} 0\,.$$

This motivates the following assumption.

Assumption 2. *For any* $k = 1, \ldots, t$, $H(k) - K \xrightarrow{p} 0$.

In section 6 we will check Assumption 2 through simulations.

The Assumptions 1 and 2 in turn imply that $H - K \xrightarrow{p} 0$. Since $K \sim \chi^2_{(t-1)}$, thus in turn we have the following proposition, where we state that the asymptotic distribution of H is $\chi^2(t-1)$, i.e. H is asymptotically distribution free.

Proposition 1. *Suppose for all* i *that* $n_i/n \to \lambda_i$ *as* $n \to \infty$ *for some numbers* $\lambda_i > 0$, *then under the null hypothesis and Assumptions 1 and 2*

$$H = \frac{1}{t}\sum_{k=1}^{t} H(k) = \frac{12}{n(n+1)t}\sum_{k=1}^{t}\sum_{j=1}^{t}\frac{R^2_{\cdot j}(k)}{n_j} - 3(n+1).$$

has a large sample chi-square distribution with $t-1$ *degrees of freedom.*

In small samples, although the test statistic H is not distribution free under the null hypothesis, we can still do a permutation test of Fisher to approximate p-values (see Efron and Tibshirani, 1993, and Ernst, 2004) based on the test statistic H. In what follows we explain how to use this procedure.

1. Compute H based on the data $\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_t$ and denote it by H_{obs}, where
$$\mathbb{X}_j = \{\mathbf{X}_{1\,j}, \mathbf{X}_{2\,j}, \ldots, \mathbf{X}_{n_j\,j}\}\,, \quad j = 1, \ldots, t$$
2. Permute the combined sample $\mathbb{Y} = \mathbb{X}_1 \cup \mathbb{X}_2 \cdots \cup \mathbb{X}_t$ a total of B times, for B sufficiently large.
3. For all of B permutations, treat the first n_1 vectors as the $\mathbb{X}_1$ sample, the next n_2 vectors as the $\mathbb{X}_2$ sample, ..., and the last n_t vectors as the $\mathbb{X}_t$ sample. Let $\mathbb{X}^*_j(b) = \{\mathbf{X}^*_{1\,j}(b), \ldots, \mathbf{X}^*_{n_j\,j}(b)\}$, $j = 1, \ldots, t$ be the bth permutation for $b = 1, 2, \ldots, B$.
4. Compute the value of the test statistic H for each permutation. Denote these by H^*_b, $b = 1, 2, \ldots, B$. Thus the empirical distribution of H^*_bs can be used to approximate the null distribution of H.

Note that to approximate the p-value $P_{H_o}(H \geq H_{obs})$ one can use

$$\widehat{P_{H_o}}(H \geq H_{obs}) = \frac{1}{B}\sum_{b=1}^{B} I(H^*_b \geq H_{obs})\,. \tag{11}$$

Thus the power functional of this α-level permutation test is

$$\Pi(G) = P_G\left(\widehat{P_{H_o}}(H \geq H_{obs}) \leq \alpha\right) \tag{12}$$

where G is the joint distribution of the t populations. Under H_o, $\Pi(G)$ approximates the type I error and for an specific alternative $\Pi(G)$ provides an approximation of power at G. For a given G one can apply a Monte Carlo method to estimate $\Pi(G)$.

To end this section it is worthwhile to mention that, H is affine invariant, if the underlying depth function is affine equivariant. The test statistic H is also robust against outliers, especially when the underlying depth function is robust.

6. Monte Carlo studies

6.1. Investigating the asymptotic distribution of H

In Table 4 we have simulation studies of Assumption 2. We have considered two sample problems with sample sizes $n_1 = n_2 = 25, 50, 100, 200, \ldots, 1000$. The following five bivariate models have been considered:

- Model (I) refers to the case that both samples have be drawn from the standard bivariate normal distribution $N_2((0,0)', \mathbf{I})$, where $\mathbf{I}$ refers to the identity matrix.
- Model (II) is a mixture of two bivariate normal distributions with mixing probability $\epsilon = 0.80$. Both samples have been taken from $0.80\, N_2((0,0)', \mathbf{I}) + 0.20\, N_2((6,6)', \mathbf{I})$.
- Model (III) is a mixture of two bivariate normal distributions with mixing probability $\epsilon = 0.80$. Both samples have been taken from $0.80\, N_2((0,0)', \mathbf{I}) + 0.20\, N_2((6,6)', \mathbf{\Sigma}_a)$, where

$$\mathbf{\Sigma}_a = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}.$$

- Model (IV) is the bivariate t distribution $t((0,0)', \mathbf{I}, 3)$, Johnson and Kotz (1972).
- Model (V) is the bivariate t distribution $t((0,0)', \mathbf{I}, 1)$ or $C((0,0)', \mathbf{I})$.

TABLE 4
Monte Carlo estimates of $E_{H_0}|H - K|$ under different models and depth functions, for $d = 2$

	Depth function		
Model	*RMD*	*RASD*	*HSD*
(I)	$10.13\, n^{-0.85}$	$13.25\, n^{-0.92}$	$43.97\, n^{-0.91}$
(II)	$7.78\, n^{-0.95}$	$7.81\, n^{-0.94}$	$38.02\, n^{-0.86}$
(III)	$6.79\, n^{-0.92}$	$7.55\, n^{-0.94}$	$41.63\, n^{-0.89}$
(IV)	$6.37\, n^{-0.81}$	$8.18\, n^{-0.88}$	$41.81\, n^{-0.94}$
(V)	$4.24\, n^{-0.81}$	$4.71\, n^{-0.86}$	$37.90\, n^{-0.98}$

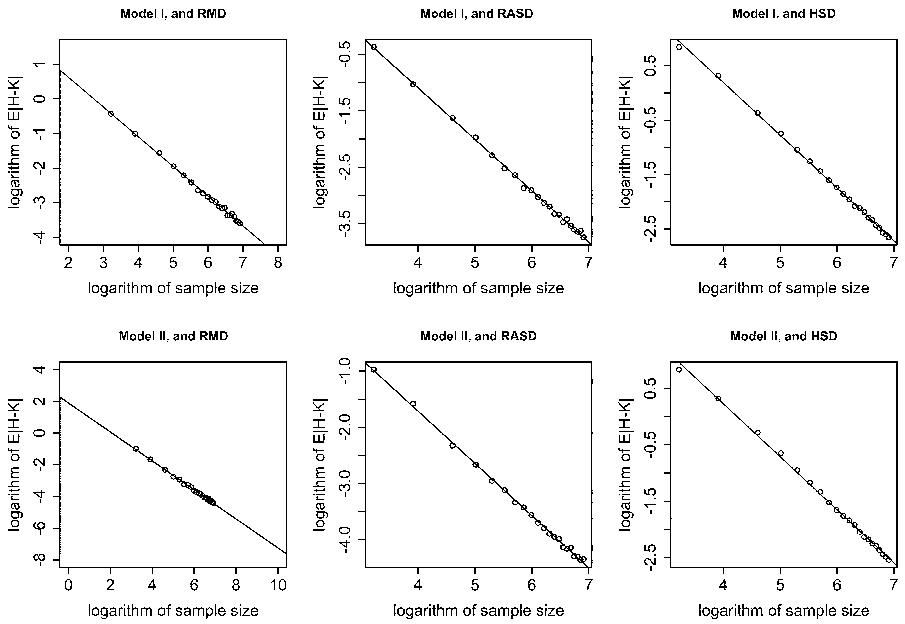

FIG 1. *Scatter plots of* $\log(\widehat{E}_{H_0}|H-K|)$ *versus* $\log(n)$ *for Models (I) and (II) based on RMD and RASD (using the MCD estimates with breakdown point of* 0.5*) and HSD.*

The underlying depth functions in our simulation study are robust Mahalonobis depth (RMD), robust affine spatial depth (RASD), and the halfspace or Tukey depth (HSD).

Figure 1 shows that the logarithm of $\widehat{E}_{H_0}|H-K|$ goes to $-\infty$ roughly linearly in the logarithm of the sample size, indicating that $E_{H_0}|H-K|$ roughly obeys a power law in n. Table 4 provides the exponent in this power law for each model and depth function. As we see in this table, when the sample sizes are increasing, $\widehat{E}_{H_0}|H-K| \to 0$. This implies that $H - K \xrightarrow{p} 0$.

6.2. Monte Carlo power study

In this section we display results from a Monte Carlo power study for the cases $d = 2$, 6, 10. The underlying depth functions in our simulation study are RMD, RASD, and HSD. In addition to our proposed statistic H in Section 5, we examine the performance of Hotelling's T^2 statistic, the spatial statistic CM suggested by Choi and Marden (1997) and K. Note that the test statistics H, K, and T^2 are affine invariant, but CM is not. All the simulations are done for the two-sample case with equal numbers of multivariate observations for two groups. In tables 5 to 8, we present some Monte Carlo results based on 1,000 trials. Table 5 reports the results of permutation tests for $d = 2$ and small sample sizes. Tables 6, 7, 8 report the results using the asymptotic distribution for large sample sizes

TABLE 5
Permutations tests: observed relative frequency of rejecting H_0, $d = 2$

Scenario	$n_1 = n_2$	c	Test statistics			
			T^2	CM	HRMD	HRASD
(1)	25	0	0.042	0.040	0.048	0.043
		1	0.995	0.993	0.773	0.838
(2)	25	1	0.042	0.040	0.048	0.043
		2	0.038	0.038	0.474	0.482
(3)	25	0	0.042	0.040	0.048	0.043
		0.9	0.067	0.074	0.795	0.812
(4)	25	0	0.059	0.059	0.053	0.046
		2	0.704	0.951	0.993	0.986
(5)	25	0	0.059	0.059	0.053	0.046
		2	0.624	0.898	0.994	0.998
(6)	25	0	0.038	0.033	0.040	0.040
		2	0.282	0.435	0.686	0.813
(7)	25	0	0.056	0.050	0.058	0.055
		1	0.194	0.610	0.270	0.320
(8)	25	1	0.050	0.054	0.057	0.053
		3	0.065	0.077	0.334	0.334
(9)	25	∞	0.042	0.040	0.048	0.043
		1	0.057	0.066	0.529	0.514

in dimensions $d = 2, 6, 10$, respectively. For each trial, we generate a data set under the null hypothesis and a data set under the alternative hypothesis. Each data set consists of two samples which are generated from two d-dimensional distributions. For the null hypothesis these distributions are identical and for the alternative hypothesis they are different. Each sample from every distribution is of size $n = 25$ for permutation tests and $n = 50, 100$ for asymptotic tests in $d = 2$. The sample sizes for $d = 6, 10$ are given in Tables 7 and 8, respectively. The nominal significant level $\alpha = 0.05$ is used. In each case displayed in Tables 6, 7, 8, the cutoff point for the nominal significant level is calculated using the asymptotic distribution under the null hypothesis. So, for our proposed test H and also K we use the chi-square cutoff point, $\chi^2_{(1,\alpha)}$, and for CM, use $\chi^2_{(d,\alpha)}$, where $\chi^2_{(\mathrm{df},\alpha)}$ is the upper αth point of the $\chi^2_{(\mathrm{df})}$ distribution. For the Hotelling test, we use the cutoff value $((n-2)d/(n-1-d))F_{d,n-1-d,\alpha}$, where $F_{a,b,\alpha}$ is the upper αth point of the $F_{a,b}$ distribution. This cut-off value is the exact value under the multivariate normal assumption. The entries in each table are the proportions of times each statistic exceeded its asymptotic critical value.

In Table 5 to 8, different scenarios are considered.

- Scenario (1) refers to a multivariate normal distribution, where the first sample is taken from the standard multivariate normal distribution $N(\mathbf{0}, \mathbf{I})$ and the second sample is taken from the multivariate normal $N(c\mathbf{1}, \mathbf{I})$ with mean vector $c\mathbf{1}$, and identity covariance matrix $\mathbf{I}$, for $c \in \mathbb{R}$. Here $\mathbf{0} = (0, \ldots, 0)'$ and $\mathbf{1} = (1, \ldots, 1)'$, are the d-dimensional vectors of 0's and 1's, respectively. The null hypothesis corresponds to $c = 0$.

TABLE 6
Observed relative frequency of rejecting H_0, $d = 2$

Scenario	$n_1 = n_2$	c	T^2	CM	KRMD	KRASD	KHSD	HRMD	HRASD	HHSD
		0.00	0.039	0.034	0.047	0.047	0.049	0.049	0.055	0.131
	50	0.50	0.881	0.871	0.063	0.048	0.046	0.289	0.305	0.678
		0.75	0.996	0.994	0.048	0.037	0.025	0.737	0.79	0.968
(1)										
		0.00	0.049	0.047	0.054	0.056	0.052	0.052	0.054	0.092
	100	0.50	0.997	0.997	0.057	0.049	0.049	0.446	0.477	0.771
		0.75	1.000	1.000	0.064	0.042	0.024	0.951	0.971	0.999
		1.00	0.039	0.034	0.047	0.047	0.049	0.049	0.055	0.131
	50	0.60	0.059	0.055	0.569	0.569	0.548	0.536	0.555	0.682
		0.40	0.049	0.055	0.963	0.962	0.958	0.962	0.966	0.982
(2)										
		1.00	0.049	0.047	0.054	0.056	0.052	0.052	0.054	0.092
	100	0.60	0.039	0.040	0.879	0.881	0.880	0.863	0.860	0.897
		0.40	0.048	0.053	1.000	1.000	1.000	1.000	1.000	1.000
		0.00	0.039	0.034	0.047	0.047	0.049	0.049	0.055	0.131
	50	0.70	0.051	0.048	0.267	0.262	0.191	0.409	0.412	0.697
		0.90	0.054	0.062	0.917	0.919	0.725	0.979	0.982	0.998
(3)										
		0.00	0.049	0.047	0.054	0.056	0.052	0.052	0.054	0.092
	100	0.70	0.049	0.053	0.517	0.512	0.431	0.661	0.675	0.838
		0.90	0.062	0.066	0.998	0.998	0.974	1.000	1.000	1.000
		0.00	0.048	0.052	0.046	0.059	0.049	0.061	0.051	0.134
	50	1.00	0.407	0.920	0.071	0.284	0.418	0.789	0.794	0.983
		1.50	0.755	0.998	0.091	0.494	0.611	1.000	0.999	1.000
(4)										
		0.00	0.050	0.055	0.059	0.043	0.053	0.054	0.052	0.098
	100	1.00	0.727	0.999	0.080	0.521	0.734	0.989	0.987	1.000
		1.50	0.975	1.000	0.108	0.795	0.918	1.000	1.000	1.000
		0.00	0.048	0.052	0.046	0.059	0.049	0.061	0.051	0.134
	50	1.00	0.331	0.879	0.066	0.382	0.578	0.794	0.858	0.993
		1.50	0.642	0.984	0.095	0.587	0.841	0.998	1.000	1.000
(5)										
		0.00	0.050	0.055	0.059	0.043	0.053	0.054	0.052	0.098
	100	1.00	0.558	0.993	0.060	0.638	0.905	0.981	0.993	1.000
		1.50	0.925	1.000	0.088	0.867	0.987	1.000	1.000	1.000
		0.00	0.046	0.047	0.054	0.048	0.046	0.053	0.036	0.185
	50	1.00	0.165	0.373	0.081	0.632	0.538	0.290	0.846	0.944
		1.50	0.272	0.566	0.196	0.799	0.808	0.836	0.988	0.997
(6)										
		0.00	0.050	0.052	0.052	0.039	0.041	0.049	0.044	0.106
	100	1.00	0.258	0.652	0.074	0.907	0.850	0.563	0.992	0.992
		1.50	0.549	0.884	0.209	0.980	0.977	0.990	1.000	1.000
		0.00	0.014	0.045	0.052	0.053	0.054	0.043	0.043	0.075
	50	0.75	0.064	0.688	0.070	0.068	0.078	0.177	0.204	0.554
		1.00	0.129	0.923	0.075	0.079	0.079	0.513	0.549	0.831
(7)										
		0.00	0.018	0.049	0.052	0.057	0.053	0.043	0.042	0.062
	100	0.75	0.069	0.955	0.071	0.068	0.070	0.362	0.392	0.674
		1.00	0.134	1.000	0.065	0.058	0.062	0.863	0.868	0.932
		0.00	0.014	0.045	0.052	0.053	0.054	0.043	0.043	0.075
	50	2.00	0.013	0.051	0.321	0.319	0.306	0.290	0.286	0.396
		3.00	0.015	0.042	0.641	0.642	0.627	0.633	0.633	0.718
(8)										
		0.00	0.018	0.049	0.052	0.057	0.053	0.043	0.042	0.062
	100	2.00	0.013	0.036	0.530	0.535	0.517	0.546	0.548	0.587
		3.00	0.016	0.043	0.923	0.920	0.909	0.910	0.907	0.919
		∞	0.039	0.034	0.047	0.047	0.049	0.049	0.055	0.131
	50	3	0.051	0.047	0.218	0.224	0.206	0.217	0.215	0.310
		1	0.023	0.048	0.807	0.807	0.783	0.791	0.787	0.819
(9)										
		∞	0.049	0.047	0.054	0.056	0.052	0.052	0.054	0.092
	100	3	0.047	0.045	0.372	0.375	0.366	0.409	0.404	0.454
		1	0.028	0.053	0.971	0.969	0.966	0.978	0.980	0.979

TABLE 7
Observed relative frequency of rejecting H_0, $d = 6$

Scenario	$n_1 = n_2$	c	T^2	CM	KRMD	KRASD	HRMD	HRASD
(1)	300	0.00	0.047	0.043	0.051	0.050	0.056	0.062
		0.25	1.000	1.000	0.048	0.046	0.526	0.606
		0.50	1.000	1.000	0.050	0.049	1.000	1.000
(2)	300	1.00	0.047	0.043	0.051	0.050	0.056	0.062
		0.60	0.056	0.048	1.000	1.000	1.000	1.000
(4)	200	0.00	0.044	0.049	0.044	0.046	0.050	0.061
		0.50	0.276	0.976	0.081	0.662	0.989	0.996
(5)	200	0.00	0.044	0.049	0.044	0.046	0.050	0.061
		0.50	0.190	0.960	0.058	0.762	0.988	0.998
(6)	200	0.00	0.037	0.039	0.060	0.058	0.054	0.040
		0.50	0.201	0.695	0.052	0.864	0.781	0.413
(7)	200	0.00	0.010	0.037	0.045	0.046	0.046	0.045
		0.25	0.027	0.756	0.061	0.060	0.061	0.061
		0.50	0.100	1.00	0.066	0.064	0.258	0.293
		0.75	0.253	1.000	0.081	0.078	0.993	0.994
(8)	200	1.00	0.010	0.037	0.045	0.046	0.046	0.045
		2.00	0.013	0.041	0.937	0.938	0.941	0.944
(9)	300	∞	0.047	0.043	0.051	0.050	0.056	0.062
		3	0.039	0.056	0.976	0.976	0.970	0.971
		1	0.026	0.041	1.000	1.000	1.000	1.000
(10)	200	1	0.010	0.037	0.045	0.046	0.046	0.045
		3	0.021	0.039	0.972	0.970	0.944	0.944
		∞	0.017	0.052	1.000	1.000	1.000	1.000

TABLE 8
Observed relative frequency of rejecting H_0, $d = 10$

Scenario	$n_1 = n_2$	c	T^2	CM	KRMD	KRASD	HRMD	HRASD
(1)	500	0	0.059	0.058	0.061	0.061	0.088	0.098
		0.25	1.000	1.000	0.051	0.051	0.999	1.000
(2)	500	1.00	0.059	0.058	0.061	0.061	0.088	0.098
		0.80	0.062	0.062	1.000	1.000	1.000	1.000
		0.60	0.052	0.052	1.000	1.000	1.000	1.000
(4)	500	0.00	0.052	0.057	0.055	0.051	0.049	0.051
		0.25	0.172	0.857	0.070	0.0752	0.727	0.967
(5)	500	0.00	0.052	0.057	0.055	0.051	0.049	0.051
		0.25	0.122	0.792	0.065	0.858	0.727	0.99
(6)	500	0.00	0.064	0.062	0.041	0.052	0.055	0.065
		0.25	0.101	0.426	0.058	0.758	0.552	0.350
(7)	500	0.00	0.011	0.024	0.047	0.047	0.048	0.049
		0.25	0.052	1.000	0.051	0.052	0.078	0.076
		0.50	0.234	1.000	0.059	0.058	0.953	0.963
(8)	500	1	0.011	0.024	0.047	0.047	0.048	0.049
		1.5	0.014	0.049	0.924	0.924	0.914	0.916
		2	0.015	0.046	1.000	1.000	1.000	1.000
(9)	500	∞	0.059	0.058	0.061	0.061	0.088	0.098
		3	0.046	0.045	1.000	1.000	0.999	0.999
		1	0.017	0.057	1.000	1.000	1.000	1.000
(10)	500	1	0.011	0.024	0.047	0.047	0.048	0.049
		3	0.019	0.058	1.000	1.000	1.000	1.000
		∞	0.018	0.057	1.000	1.000	1.000	1.000

- Scenario (2) refers again to a multivariate normal distribution, but this time the first sample is taken from the standard multivariate normal distribution $N(\mathbf{0}, \mathbf{I})$. The second sample is taken from the multivariate normal $N(\mathbf{0}, c\,\mathbf{I})$ with the same mean $\mathbf{0}$ but the covariance matrix is given by $c\,\mathbf{I}$, for $c > 0$. The null hypothesis corresponds to $c = 1$.
- Scenario (3) refers to a bivariate normal distribution, while the first sample is taken from the standard bivariate normal distribution $N((0,\, 0)', \mathbf{I})$. The second sample is taken from the bivariate normal with the same mean $(0, 0)'$ but the covariance matrix is given by

$$\begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}.$$

  The null hypothesis corresponds to $c = 0$.
- Scenario (4) refers to a mixture of two multivariate normal distributions with mixing probability $\epsilon = 0.8$. The first sample is taken from $0.8\,N(\mathbf{0}, \mathbf{I}) + 0.2\,N(6\mathbf{1}, \mathbf{I})$ and the second sample from $0.8\,N(c\mathbf{1}, \mathbf{I}) + 0.2\,N((6 + c)\,\mathbf{1}, \mathbf{I})$. The null hypothesis corresponds to $c = 0$.
- Scenario (5) refers to a mixture of two multivariate normal distributions with mixing probability $\epsilon = 0.8$. The first sample is taken from $0.8\,N(\mathbf{0}, \mathbf{I}) + 0.2\,N(6\mathbf{1}, \mathbf{I})$ and the second sample from $0.8\,N(c\mathbf{1}, \mathbf{I}) + 0.2\,N(6\mathbf{1}, \mathbf{I})$. The null hypothesis corresponds to $c = 0$.
- Scenario (6) refers to a mixture of two multivariate normal distributions with mixing probability $\epsilon = 0.6$. The first sample is taken from $0.6\,N(\mathbf{0}, \mathbf{I}) + 0.4\,N(6\mathbf{1}, \mathbf{I})$ and the second sample from $0.6\,N(c\mathbf{1}, \mathbf{I}) + 0.4\,N(6\mathbf{1}, \mathbf{I})$. The null hypothesis corresponds to $c = 0$.
- Scenario (7) refers to the multivariate t distribution, where the first sample is taken from $t(\mathbf{0},\, \mathbf{I},\, 1)$, which is multivariate Cauchy, and the second sample is from $t(c\mathbf{1},\, \mathbf{I},\, 1)$. The null hypothesis corresponds to $c = 0$.
- Scenario (8) refers to the multivariate t distribution, where the first sample is taken from $t(\mathbf{0},\, \mathbf{I},\, 1)$ and the second sample is from $t(\mathbf{0},\, c\mathbf{I},\, 1)$. The null hypothesis corresponds to $c = 1$.
- In scenario (9), the first sample is taken from the standard multivariate normal and the second sample is from $t(\mathbf{0},\, \mathbf{I},\, c)$. The null hypothesis corresponds to $c = \infty$.
- In scenario (10), the first sample is taken from $t(\mathbf{0},\, \mathbf{I},\, 1)$ and the second sample is from $t(\mathbf{0},\, \mathbf{I},\, c)$. The null hypothesis corresponds to $c = 1$.

In Table 5, which is the permutation test, the entries are calculated by random shuffling 1000 times. Therefore, the entries are typically close but not exactly at the nominal value of 0.05. In Table 6, under the null hypothesis, the asymptotic approximations to the distributions of the test statistics are satisfactory except for HHSD. For this particular case, the ranks of data points with tied depths have been averaged rather than randomly broken. This may affect the asymptotic calculations as seen. Therefore, we have chosen to omit HHSD in Tables 7 and 8.

All tests, except those for K (namely, KRMD, KRASD, KHSD), which are designed for scale shifts, perform well in the case of scenario (1) for all dimensions

studied. In this scenario, Hotelling's T^2 and CM, which are designed for location shifts, perform extremely well. As we see, depth-based tests (HRMD, HRASD, HHSD) using H perform well using different depth functions under variety of differently shaped distributions. While it does not always have the power that tests tailored to specific assumptions may have, it is reasonably robust against distribution location, shape and many other alternatives. Of course, this is true provided the central regions (as defined by the depth contours) do not overlap too much.

7. Example

A data set given by Johnson and Wichern (1988, p. 261-262) is examined here using the statistics T^2, and our proposed depth-based test H using MD depth. The data set originally was taken from Jolicoeur and Mosimann (1960), who studied the relationship of size and shape for painted turtles. Table 9 contains their measurements on the carapace of $n_1 = 24$ female and $n_2 = 24$ male turtles. This data set was partially analyzed in Hettmansperger et al. (1998).

We treat observations as two independent samples from trivariate distributions. We consider a test for $H_0 : F_1 = F_2$ where F_1 and F_2 are the respective trivariate distributions. To see that whether the trivariate normality is a reasonable assumption for the female and male populations or not, we generate two

TABLE 9
Carapace measurements (mm) for painted turtles

Female			Male		
Length (X_{11})	Width (X_{12})	Height (X_{13})	Length (X_{21})	Width (X_{22})	Height (X_{23})
98	81	38	93	74	37
103	84	38	94	78	35
103	86	42	96	80	35
105	86	42	101	84	39
109	88	44	102	85	38
123	92	50	103	81	37
123	95	46	104	83	39
133	99	51	106	83	39
133	102	51	107	82	38
133	102	51	112	89	40
134	100	48	113	88	40
136	102	49	114	86	40
138	98	51	116	90	43
138	99	51	117	90	41
141	105	53	117	91	41
147	108	57	119	93	41
149	107	55	120	89	40
153	107	56	120	93	44
155	115	63	121	95	42
158	115	62	127	96	45
159	118	63	128	95	45
162	124	61	131	95	46
177	132	67	135	106	47

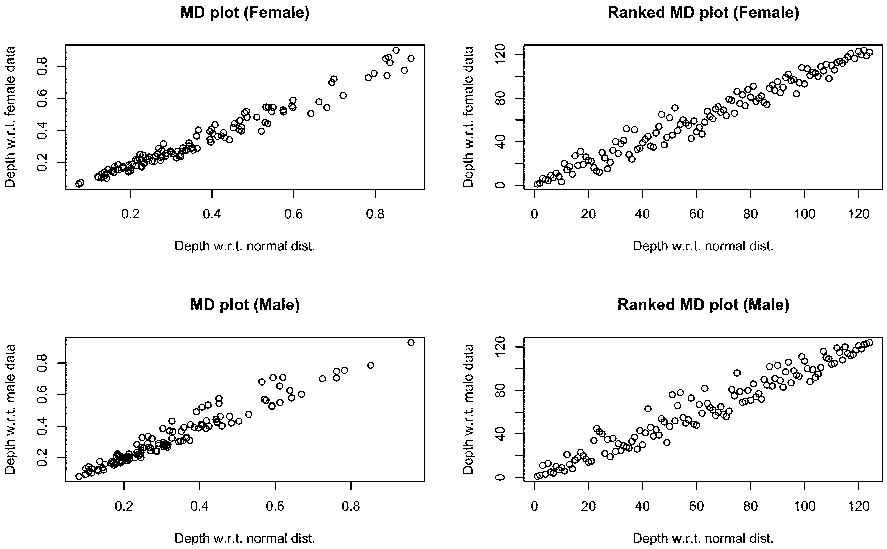

FIG 2. *Depth-Depth plots to justify the normality assumption of the female and male distributions.*

samples of size 100 from trivariate normal distributions with mean vectors and covariance matrices being estimated by the sample means and covariance matrices of the female and male data. We use the Mahalonobis depth to compute the depth values of the pooled data points. Figure 2 represents the depth-depth plots (see Liu et al. (1999)) and also ranked depth-depth plots for the female and male samples. In these plots, the x and y axes represent the depth values of the pooled data with respect to the generated samples and to the female and male data points respectively. We see that all plots are concentrated along the diagonal line. This indicates that the normality can be a reasonable assumption for both female and male samples. These elliptical structures of the two distributions are also clear from the pairwise scatter plots in Figure 3. Hence it is reasonable to use the Mahalanobis depth and not worry about tied ranks. An eyeball investigation of the pairwise scatter plots also reveals that, in addition to a location shift, two populations have different dispersion parameters.

To carry out the the test, we compute the depth of each data point in the pooled data set with respect to the female and then male samples. So we have

$$H(1) = \frac{12}{n(n+1)} \sum_{j=1}^{2} n_j \left(\bar{R}_{\cdot j}(1) - \frac{n+1}{2} \right)^2 = 12.14328$$

$$H(2) = \frac{12}{n(n+1)} \sum_{j=1}^{2} n_j \left(\bar{R}_{\cdot j}(1) - \frac{n+1}{2} \right)^2 = 28.30102$$

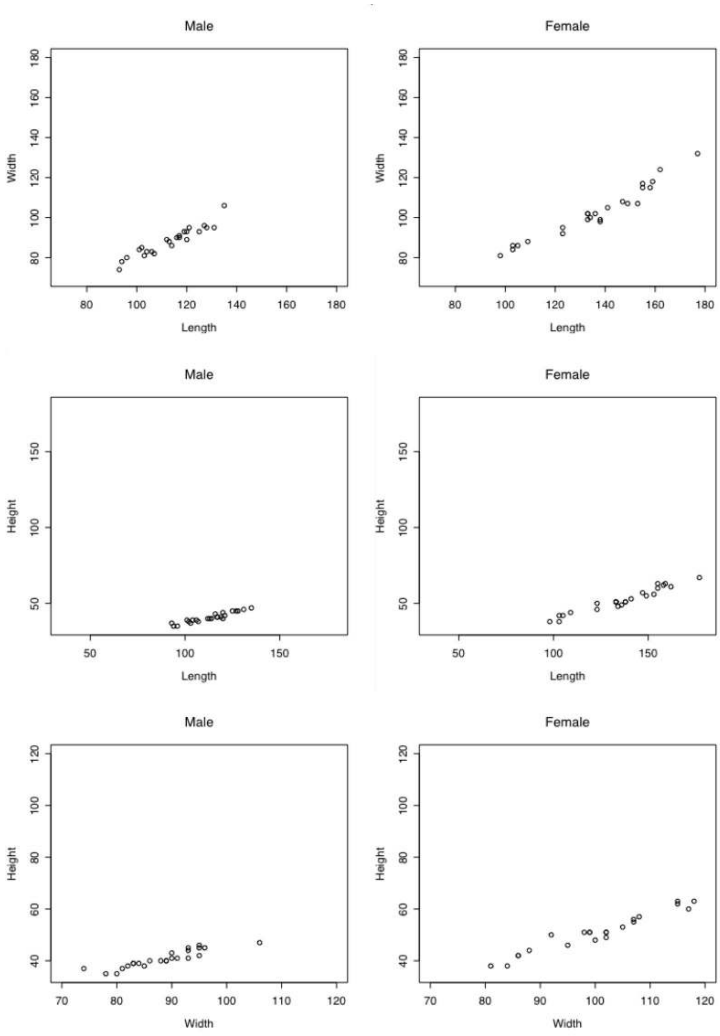

FIG 3. *Scatter plots of the carapace data male and female.*

and then finally $H = \frac{H(1)+H(2)}{2} = 20.22215$. Comparing with the percentiles of a chi-square distribution with $t-1=1$ degrees of freedom we see that the asymptotic P-value is 7×10^{-6}. We easily reject the null hypothesis $H_0 : F_1 = F_2$.

Based on 1,000,000 random permutations of the numbers $1, 2, \ldots, 48$, the P-value is estimated to be less than 10^{-6}. Figure 4 represents the depth-depth plot to compare the female and male populations, which visually justifies our finding based on the test statistic. Assuming multivariate normality, which seems reasonable from the depth-depth plots in Figure 2, the Hotelling's T^2 gives the P-value 0.

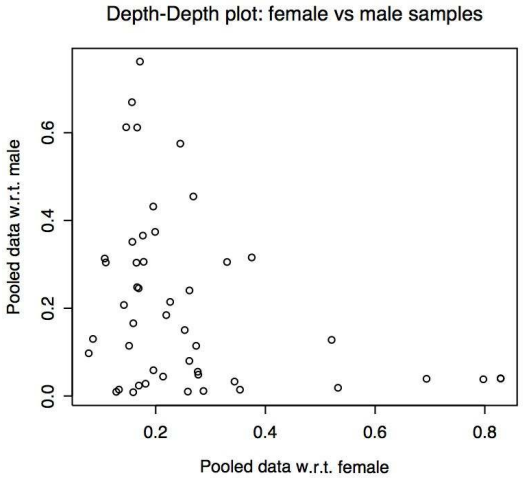

FIG 4. *Depth-Depth plot to compare the female and male populations.*

References

BARNETT, V. (1976). The ordering of multivariate data. *J. Roy. Statist. Soc. Ser. A*, 138:318–344. MR0445726

BENNETT, B. M. (1962). On multivariate sign tests. *J. Roy. Statist. Soc.*, 24:159–161. MR0138163

BICKEL, P. J. (1965). On some asymptotically non-parametric competitors of hotelling's t^2. *Ann. Math. Statist.*, 36:160–173. MR0177484

BLUMEN, I. (1958). A new bivariate sign test. *J. Amer. Statist. Assoc.*, 53:448–456.

BROWN, B. M. (1983). Statistical use of spatial median. *J. Roy. Statist. Soc.*, 45:23–30. MR0701072

BROWN, B. M. AND HETTMANSPERGER, T. P. (1987). Affine invariant rank methods and the bivariate location model. *J. Roy. Statist. Soc.*, 49:301–310. MR0928938

BROWN, B. M. AND HETTMANSPERGER, T. P. (1989). An affine invariant version of the sign test. *J. Roy. Statist. Soc.*, 51:117–125. MR0984998

CHAKRABORTY, B., CHAUDHURI, P., AND OJA, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statist. Sinica*, 8:767–784. MR1651507

CHATTERJEE, S. K. (1966). A bivariate sign test for location. *Ann. Math. Statist*, pages 1771–1780. MR0201017

CHAUDHURI, P. AND SENGUPTA, D. (1993). Sign tests in multidimension: inference based on the geometry of the data cloud. *J. Amer. Statist. Assoc.*, 88:1363–1370. MR1245371

Chenouri, S. (2004). *Multivariate robust nonparametric inference based on data depth*. PhD thesis, University of Waterloo, Waterloo, ON, CANADA.

Chenouri, S., Small, C. G., and Farrar, T. J. (2011). Data depth-based nonparametric scale tests. *Canad. J. Statist.*, 39:356–369. MR2839485

Choi, K. and Marden, J. (1997). An approach to multivariate rank tests in multivariate analysis of variance. *J. Amer. Statist. Assoc.*, 92:1581–1590. MR1615267

Dietz, E. J. (1982). Bivariate nonparametric tests for the one-sample location problem. *J. Amer. Statist. Assoc.*, 77:163–169. MR0648040

Donoho, D. (1982). Breakdown properties of multivariate location estimators. *PhD Qualifying paper, Harvard University, Boston*.

Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlingness. *Ann. Statist*, 20:1803–1827. MR1193313

Gower, J. C. (1974). Algorithm as 78: The mediancentre. *App. Statist.*, 23:466–470.

Hettmansperger, T. P., Möttönen, J., and Oja, H. (1997). Affine invariant multivariate one sample signed rank test. *J. Amer. Statist. Assoc.*, 92:1591–1600. MR1615268

Hettmansperger, T. P., Möttönen, J., and Oja, H. (1998). Affine invariant multivariate rank tests for several samples. *Statist. Sinica*, 8:765–800. MR1651508

Hettmansperger, T. P., Nyblom, J., and Oja, H. (1994). Affine invariant multivariate one sample sign tests. *J. Roy. Statist. Soc. Ser. B*, 56:221–234. MR1257809

Hettmansperger, T. P. and Oja, H. (1994). Affine invariant multivariate multisample sign tests. *J. Roy. Statist. Soc. Ser. B*, 56:235–249. MR1257810

Hodges, J. L. (1955). A bivariate sign test. *Ann. Math. Statist.*, 26:523–527. MR0070921

Hollander, M. and Wolfe, D. (1999). *Nonparametric Statistical Methods*. John Wiley, New York. MR1666064

Hössjer, O. and Croux, C. (1995). Generalizing univariate signed rank statistics for testing and estimating a multivariate location parameter. *J. Nonparametric Statist.*, 4:293–308. MR1366776

Hotelling, H. (1951). A generalized t test and measure of multivariate dispersion. *Proceeding of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 23–41. MR0044798

Johnson, N. L. and Kotz, S. (1972). *Distributions in statistics: continuous multivariate distributions*. John Wiley and Sons, New York. MR0418337

Koshevoy, G. (2001). Projections of lift zonoids, the oja depth and the tukey depth. *Unpublished manuscript*.

Kruskal, W. H. (1952). A nonparametric test for the several sample problem. *Ann. Math. Statist.*, 23:525–540. MR0050850

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one criterion variance analysis. *J. Amer. Statist. Assoc.*, 47:583–621. MR0577363

Lawley, D. N. (1938). A generalization of fisher's z-test. *Biometrika*, 30:180–187.

Lehmann, E. and D'abrera, H. (2006). *Nonparametrics: statistical methods based on ranks*. Springer, New York. MR2279708

Liu, R. and Singh, K. (2006). Rank tests for multivariate scale difference based on data depth. *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, DIMACS Series, AMS*, pages 17–36. MR2343110

Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Ann. Statist.*, 27:783–858. MR1724033

Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, 88:252–260. MR1212489

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proc. Nat. Acad. India*, 12:49–55.

Möttönen, J., Hüsler, J., and Oja, H. (2003). Multivariate nonparametric tests in a randomized complete block design. *J. Multiv. Analysis*, 85:106–129. MR1978180

Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametric Statist.*, 5:201–213. MR1346895

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Prob. Letters*, 1:327–333. MR0721446

Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: a review. *Scand. J. Statist.*, 26:319–343. MR1712063

Oja, H. and Nyblom, J. (1989). Bivariate sign tests. *J. Amer. Statist. Assoc.*, 84:249–259. MR0999686

Peters, D. and Randles, R. H. (1990). A multivariate signed-ranked test for the one-sample location problem. *J. Amer. Statist. Assoc.*, 85:552–557. MR1141757

Peters, D. and Randles, R. H. (1991). A bivariate signed rank test for the two-sample location problem. *J. Roy. Statist. Soc. Ser. B*, 53:493–504. MR1108344

Puri, M. L. and Sen, P. K. (1971). *Nonparametric methods in multivariate analysis*. John Wiley and Sons, New York. MR0298844

Randles, R. H. (1989). A distribution-free multivariate sign test based on interdirections. *J. Amer. Statist Assoc.*, 84:1045–1050. MR1134492

Randles, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *J. Amer. Statist. Assoc.*, 95:1263–1268. MR1792189

Randles, R. H. and Peters, D. (1990). Multivariate rank tests for the two sample location problem. *Comm. Statist. Theory Methods*, 19:4225–4238. MR1103009

Rao, C. R. (1988). Methodology based on the l^1 norm in statistical inference. *Sankhyā Ser. A*, 50:289–313. MR1065546

Rousseeuw, P. J. (1983). Multivariate estimation with high breakdown point. *Proc. of the 4th pannonian Symp.*

Rousseeuw, P. J. and Leroy, A. (1987). *Robust Regression and Outlier Detection*. Wiley, New York. MR0914792

Rousseeuw, P. J. and Ruts, I. (1996). Bivariate location depth. *Applied Statistics*, 45:519–526.

Rousseeuw, P. J. and Ruts, I. (1998). Constructing the bivariate tukey median. *Statist. Sinica*, 8:828–839. MR1651511

Rousseeuw, P. J. and Ruts, I. (1999). The depth function of a population distribution. *Metrika*, 49:213–244. MR1731769

Rousseeuw, P. J. and Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statist. Comput.*, 8:193–203.

Ruts, I. and Rousseeuw, P. J. (1996). Computing depth contours of bivariate point clouds. *Comput. Statist. data Analysis*, 23:153–168.

Small, C. G. (1987). Measures of centrality for multivariate and directional distributions. *Canad. J. Statist.*, 15:31–39. MR0887986

Small, C. G. (1990). A survey of multidimensional medians. *Intern. Statist. Inst. Rev.*, 58:263–277.

Struyf, A. and Rousseeuw, P. J. (1999). Halfspace depth and regression depth characterize the empirical distribution. *J. Multiv. Statist. Analysis.*, 69:135–153. MR1701410

Tukey, J. W. (1975). Mathematics and picturing data. *Proc. Intern. Congr. Math.*, 2:523–531. MR0426989

Um, Y. and Randles, R. H. (1998). Nonparametric tests for the multivariate multisample location problem. *Statist. Sinica*, 8:801–812. MR1651509

Zuo, Y. and He, X. (2006). On the limiting distributions of multivariate depth-based rank sum statistics and related tests. *Ann. Statist.*, 34:2879–2896. MR2329471

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.*, 28:461–482. MR1790005