Journal of Multivariate Analysis **69**, 167–192 (1999)
Article ID jmva.1998.1808, available online at http://www.idealibrary.com on IDEAL®

# Permutation Tests for Multivariate Location Problems

Georg Neuhaus

*University of Hamburg, Hamburg, Germany*

and

Li-Xing Zhu*

*Chinese Academy of Sciences and The University of Hong Kong, Hong Kong*

Received August 20, 1997; revised September 16, 1998

The paper presents some permutation test procedures for multivariate location. The tests are based on projected univariate versions of multivariate data. For one-sample cases, the tests are affine invariant and strictly distribution-free for the symmetric null distribution with elliptical direction and their permutation counterparts are conditionally distribution-free when the underlying null distribution of the sample is angularly symmetric. For multi-sample cases, the tests are also affine invariant and permutation counterparts of the tests are conditionally distribution-free for any null distribution with certain continuity. Hence all of the tests in this paper are exactly valid. Furthermore, the equivalence, in the large sample sense, between the tests and their permutation counterparts are established. The power behavior of the tests and of their permutation counterparts under local alternative are investigated. A simulation study shows the tests to perform well compared with some existing tests in the literature, particularly when the underlying null distribution is symmetric whether light-tailed or heavy-tailed. For revealing the influence of data sparseness on the effect of the test, some simulations with different dimensions are also performed. © 1999 Academic Press

AMS 1991 subject classifications: 62G10, 62G20, 62G09.

Key words and phrases: empirical process; multivariate location; permutation test; sign test; validity of test; Wilcoxon type test.

## 1. INTRODUCTION

Several methods have been proposed to develop test statistics which are multivariate affine invariant analogs of univariate sign and rank tests in the

* Supported by the NNSF of China and a fellowship of the Max-Planck Gesellschaft zur Förderung der Wissenschaften of Germany, while on leave from Institute of Applied Mathematics, Chinese Academy of Sciences, at University of Hamburg. The revision was done at the University of Hong Kong.



0047-259X/99 $30.00

one-sample as well as in the multi-sample cases. For one-sample case, Randles (1989) proposed an affine invariant one-sample sign test which is strictly distribution free over a broad class of distributions with elliptical directions. This class includes all elliptical distributions and also some skewed distributions. Liu (1990) suggested a test based on simplicial depth of the data and investigated the limiting behavior of the test. Peters and Randles (1990) constructed a modified version of that in Randles (1989). Hettmansperger *et al.* (1994) introduced an affine invariant sign test which is conditionally distribution-free if the underlying distribution of variable, say $x$, is reflectedly (or diagonally) symmetric about a center point, say $\theta$, that is, $x-\theta$ has the same distribution as $-(x-\theta)$. For two-sample case, Liu and Singh (1993) suggested affine invariant two-sample rank tests bases on the notions of data depths. When the underlying distributions, say $F$ and $G$, of two samples are both not completely known, they derived, among others, some asymptotic properties of the tests. Piterbarg and Tyurin (1993) adopted a projected Wilcoxon type test which is based on rankings of all possible univariate data obtainable as projections of the multivariate samples onto various directions. They studied the behavior, especially the tail probability, of limiting null distribution of the test statistic. Hettmansperger and Oja (1994) introduced another sign test for multi-sample cases which are also based on Oja's multivariate median. They showed, assuming Oja's multivariate median is unique, that the test statistic converges in distribution, under the null hypothesis, to a chi-squared variable. Related works are Blumen (1958), Brown and Hettmansperger (1987, 1989), Brown *et al.* (1992), Hettmansperger (1984), Hodges (1955), Maritz (1981), Oja and Nyblom (1989), Peters and Randles (1991), Puri and Sen (1971), Randles and Peters (1990) and the references therein.

In this paper, we maintain our interest in the projection pursuit idea (or Tukey's depth, 1975) and investigate some affine invariant sign and rank tests. In one-sample case, the tests are strictly distribution-free over the class of distributions with elliptical directions. For the broader class of the distributions, we suggest permutation test procedures for the purpose of computing the critical values. The permutation counterparts of the tests are conditionally distribution-free over a broader class of distributions which includes the angularly symmetric distributions (see Liu and Singh, 1993). In multi-sample case, the tests are also affine invariant and their permutation counterparts are conditionally distribution-free over the class consisting of all distributions whose univariate marginals are continuous in certain sense. This class of distributions includes all absolute continuous distributions. Hence all tests investigated in this paper are exactly valid. Furthermore, it will be shown that the tests are asymptotically equivalent to their conditional counterparts. We shall also show that the proposed tests can detect local alternative converging to the null as fast as square root $n$, a parametric rate. In order to demonstrate how the tests work, some simulation experiments are performed. The tests, especially the integration type tests defined in Section 2, perform well compared with some existing tests in the literature, particularly for the symmetric null distribution whether light-tailed or heavy-tailed. Furthermore, some simulation is also performed to evidence how the data sparseness affects the performance of the tests.

This paper is organized in the following way: Section 2 contains the test statistics and associated small sample properties. The limiting behavior of the tests, including power study, is investigated in Section 3. Section 4 contains some simulation experiments. A concluding remark is put in Section 5 in which we discuss a possible application of the test in the nonparametric regression setting. Section 6 contains the proofs of the theorems.

## 2. CONSTRUCTION OF TESTS

We first adopt some definitions of symmetry of a random variable so that the description of the results is convenient. Let $x$ be a $d$-variate variable and $\|\cdot\|$ mean the $L^2$-norm.

DEFINITION 1. $x$ is elliptically symmetric if there exists a $d \times d$ matrix $A$ and a vector $\theta$ such that $A(x-\theta)/\|A(x-\theta)\|$ is distributed uniformly on $S^d$ and is independent of $\|A(x-\theta)\|$.

DEFINITION 2 (Randles, 1989). $x$ is symmetric with elliptical direction if there exist a $d \times d$ matrix $A$ and a vector $\theta$ such that $A(x-\theta)/\|A(x-\theta)\|$ is distributed uniformly on $S^d$.

DEFINITION 3. $x$ is reflectedly symmetric if there exists a vector $\theta$ such that $(x-\theta)$ and $-(x-\theta)$ have the same distribution.

DEFINITION 4 (Liu and Singh, 1993, p. 253). $x$ is angularly symmetric if there exist a $d \times d$ matrix $A$ and a vector $\theta$ such that $A(x-\theta)/\|A(x-\theta)\|$ is reflectedly symmetric.

Note that the variable $x$ satisfying any kind of symmetry defined above will have the following property which will be the basis of constructing the tests below: for any projection direction $a \in S^d = \{a\colon \|a\| = 1, a \in R^d\}$

$$P\{a'(x-\theta) \leqslant 0\} = 1/2. \tag{2.1}$$

*Remark* 2.1. From definitions above, it is easy to see the following relation among the classes of distributions: elliptically symmetric $\subset$ symmetric with elliptical direction or reflectedly symmetric $\subset$ angularly symmetric.

In the following context of this paper, we assume that $A$ is $d \times d$ full-rank matrix.

### 2.1. *One-sample Case*

Let $x_1, ..., x_m$ be an *iid* sample from $F$. We now want to test

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta \neq \theta_0 \tag{2.2}$$

for a known point $\theta_0$ in $R^d$. Without loss of generality, assume $\theta_0 = 0$. For any projection direction $a \in S^d = \{a \colon \|a\| = 1, a \in R^d\}$, the projected sample $a'x_1, ..., a'x_m$ is a univariate one. Hence a sign test can be applied to test this location shift hypothesis:

$$W_m(a) = (1/\sqrt{m}) \sum_{i=1}^{m} (I(a'x_i \leqslant 0) - 1/2). \tag{2.3}$$

For the hypothesis $H_0$, the supremum or quadratic functional of $W_m(a)$ over $a \in S^d$ can be used as a test statistic

$$W_{m1} = \sup_a |W_m(a)|, \tag{2.4}$$

or

$$W_{m2} = \int (W_m(a))^2 \, d\mu(a), \tag{2.5}$$

where $\mu(\cdot)$ is the uniform distribution of $S^d$. Clearly, these two test statistics are both affine invariant. It is known that the sampling null distribution of a test is crucial, especially in small sample cases, for determining the critical value of the test. The following proposition shows that the tests defined above are strictly distribution-free if the underlying distribution of the variable $x$ is symmetric with elliptical direction.

PROPOSITION 2.1. *Assume that the distribution of the variable $x$ is symmetric with elliptical direction and $A$ is a $d \times d$ nonsingular matrix. The tests defined in* (2.4) *and* (2.5) *are then strictly distribution-free.*

Actually, it is easy to see that

$$W_m(a) = (1/\sqrt{m}) \sum_{i=1}^{m} I((a'A^{-1}/\|a'A^{-1}\|)\, Ax_i/\|Ax_i\| \leqslant 0) - 1/2$$

$$=: (1/\sqrt{m}) \sum_{i=1}^{m} I(b'z_i \leqslant 0) - 1/2 =: W'_m(b), \tag{2.6}$$

and then

$$W_{m1} = \sup_b |W'_m(b)| \tag{2.7}$$

and

$$W_{m2} = \int (W'_m(b))^2\, d\mu(b), \tag{2.8}$$

where $b' = a'A^{-1}/\|a'A^{-1}\| \in S^d$, and $z_1, ..., z_m$ are *iid* with the common uniform distribution on $S^d$.

From this proposition, we realize that although the closed form of the distribution of $W_{m1}(W_{m2})$ has not yet derived, the Monte Carlo approximation is clearly available because the random variables with a uniform distribution on $S^d$ can be generated by computer. Hence, the distribution of $W_{m1}(W_{m2})$ can be approximated at $\sqrt{r}$ rate by the empirical distribution of $W^{(i)}_{m1}$'s ($W^{(i)}_{m2}$'s) which are calculated basing on *iid* sets of simulated data $z^{(i)}_1, ..., z^{(i)}_m$, $i = 1, ..., r$. The approximation will be accurate enough as long as the number of replications $r$ is large enough.

When the variable $x$ is not symmetric with elliptical direction, but angularly or reflectedly symmetric, the test will no longer be strictly distribution-free, which causes the difficulty of determining the critical values. For solving this problem, the permutation tests will be defined in the following which are conditionally distribution-free.

Let $a \cdot b$ mean that every component of the vector $b$ is multiplied by a common univariate variable $a$. Let $e_1, ..., e_m$ be *iid* permutation variables, that is, $e_i = \pm 1$, $i = 1, ..., m$ with probability values one half and let $E_m = (e_1, ..., e_m)$. Define a permutation test statistic by

$$W_m(E_m, a) = (1/\sqrt{m}) \sum_{i=1}^{m} (I(a'e_i \cdot x_i \leqslant 0) - 1/2), \tag{2.9}$$

The resulting tests are then

$$W_{m1}(E_m) = \sup_a |W_m(E_m, a)| \tag{2.10}$$

and

$$W_{m2}(E_m) = \int (W_m(E_m, a))^2 \, d\mu(a), \tag{2.11}$$

PROPOSITION 2.2. *Assume that the distribution of the variable $x$ is angularly symmetric. Let $E_m^{(1)}, ..., E_m^{(l)}$ be iid copies of $E_m$. Then for any $0 < \alpha < 1$, $i = 1, 2$,*

$$P\{\#\{W_{mi} > W_{mi}(E_m^{(j)})' s\} > l - [l\alpha]\} \leqslant ([l\alpha] + 1)/(l+1), \tag{2.12}$$

*where the notation $[c]$ means the largest integer part of $c$.*

This proposition means that the tests are strictly valid in small sample case.

*Remark* 2.2. Actually, for a broader class of distributions, Proposition 2.2 still holds. The variable $x$ is said to be projectedly symmetric if there exists a real function $g(\cdot)$ on a symmetric subset of $R^d$, say $D$, which satisfies that (1) $g(x) > 0$ for $x \neq 0$; and (2) $g(x) = g(-x)$ and there exist a $d \times d$ matrix $A$ and a vector $\theta$ such that $A(x-\theta)/g(A(x-\theta))$ is reflectedly symmetric. The class of projectedly symmetric distributions contains the one of angularly symmetric distributions. The class of $L_\alpha$-symmetric distributions with density function $f(\sum_{j=1}^d |x^{(j)}|^\alpha)$ $0 < \alpha \leqslant 2$ is also a subclass where $x^{(j)}$'s are the components of $x$. As we know, $\alpha = 2$ corresponds to spherical distribution and $\alpha = 1$ to the distribution which is a multivariate extension of the Laplace distribution. In this case, $g(x) = (\sum_{j=1}^d |x^{(j)}|^\alpha)^{1/\alpha}$.

*Remark* 2.3. The above conditional tests are not the classical permutation test procedures. We name them as the permutation tests since in the spirit, they are similar.

### 2.2. *Two-sample Cases*

Following the procedure in the one-sample case, the test statistics for two samples can be constructed. Let $x_1, ..., x_m$ and $y_1, ..., y_n$ be two samples with the distributions $F(\cdot)$ and $F(\cdot - \Delta)$ respectively. Let $N = m + n$.

Consider two sample problems:

$$H_0: \Delta = 0 \qquad \text{versus} \qquad H_1: \Delta \neq 0. \tag{2.13}$$

For a projection direction $a \in S^d$, two projected samples are

I. $a'x_1, ..., a'x_m$;

II. $a'y_1, ..., a'y_n$.

Then a two-sample Wilcoxon-type test statistic which is almost the same as that defined by Piterbarg and Tyurin (1993) is

$$V_N(a) = \sqrt{nm/N}\left\{(1/mn)\sum_{i=1}^{m}\sum_{j=1}^{n} I(a'x_i \leqslant a'y_j) - 1/2\right\}. \tag{2.14}$$

The resulting test statistics are

$$V_{N,1} = \sup_a |V_N(a)| \tag{2.15}$$

and

$$V_{N,2} = \int (V_N(a))^2 \, d\mu(a). \tag{2.16}$$

Clearly, these two test statistics are both affine invariant.

For the same reason as that in the one-sample case, it is necessary to calculate (or approximate) the sampling null distribution of the test statistic. Piterbarg and Tyurin studied the tail probability of the limiting null distribution of the statistic when assuming the underlying distribution of the variable does not have heavy tail in certain sense, see Piterbarg and Tyurin (1993). It is not clear whether the tail probability estimation is available for determining the critical values in finite sample cases. Taking into account this question, the permutation technique is here recommended for small sample cases. We will see that the permutation tests will be exactly valid. Let $R = (r_1, ..., r_N)$ be a random permutation of $(1, ..., N)$. Denote $z_i = x_i$, for $1 \leqslant i \leqslant m$; $z_{m+j} = y_j$, for $1 \leqslant j \leqslant n$. Define $F^R_{ma}(t) = (1/m)\sum_{i=1}^{m} I(a'z_{r_i} \leqslant t)$ and $G^R_{na}(t) = (1/n)\sum_{j=1}^{n} I(a'z_{r_{m+j}} \leqslant t)$.

$$V_N(R, a) = \int F^R_{ma}(t)\, dG^R_{na}(t) - 1/2 \tag{2.17}$$

The resulting permutation tests are

$$V_{N,1}(R) = \sup_a |V_N(R, a)|, \tag{2.18}$$

and

$$V_{N,2}(R) = \int (V_N(R, a))^2 \, d\mu(a). \tag{2.19}$$

Parallel to Proposition 2.2, we have

PROPOSITION 2.3. *Assume that $P\{a'x \leqslant t\}$ is continuous with respect to $a$ and $t$. Let $R^{(1)}, ..., R^{(l)}$ be iid permutations of $(1, ..., N)$. Then for any $\alpha > 0$, $i = 1, 2$,*

$$P\{\#\{V_{N,i} > V_{N,i}(R^{(j)})'s\} > l - [l\alpha]\} \leqslant ([l\alpha] + 1)/(l+1). \tag{2.20}$$

## 2.3. *General Cases*

Since the conclusions for general cases parallel essentially to those in two-sample cases, hence we only sketch the development of the test statistics.

Let $\{x_{11}, ..., x_{1n_1}\}, ..., \{x_{c1}, ..., x_{cn_c}\}$ be $c$ samples which are, respectively, from $F(\cdot - \theta - \Delta_i)$, $i = 1, ..., c$. We want to test the hypothesis $H_0\colon \Delta_1 = \Delta_2 \cdots = \Delta_c$, where $\Delta_1 = 0$. For every pair $(i, l)$, $1 \leqslant i < l \leqslant c$, define

$$V_{N,(i,l),1} = \sup_a \left| \sqrt{n_i n_l / N} \left\{ (1/n_i n_l) \sum_{j,m} (I(a'x_{ij} \leqslant a'x_{lm}) - 1/2) \right\} \right| \tag{2.21}$$

and

$$V_{N,(i,l),2} = \int \left( \sqrt{n_i n_l / N} \left\{ (1/n_i n_l) \sum_{j,m} (I(a'x_{ij} \leqslant a'x_{lm}) - 1/2) \right\} \right)^2 d\mu(a). \tag{2.22}$$

The resulting test statistics are

$$V_{N,1} = \frac{2}{c(c-1)} \sum_{j<l} V_{N,(i,l),1} \tag{2.23}$$

and

$$V_{N,2} = \frac{2}{c(c-1)} \sum_{j<l} V_{N,(i,l),2}. \tag{2.24}$$

In the next section, we will discuss the limit behavior of the test statistics.

## 3. SOME LIMITING PROPERTIES

In the previous section, some small sample properties of the tests are investigated. Making use of the permutation technique makes the tests exactly valid. On the other hand, one of the concerns is whether the tests and their permutation counterparts are asymptotically equivalent. We shall give a positive answer to this question below. It is also important to know that, in large sample sense, what kind of alternative could be detected by the tests. We shall present some asymptotic results about this.

For convenience of the representation of the results, a definition of $P$-Brownian bridge $B_P$ is adopted; for a full account see e.g. Gine and Zinn (1984, 1986) or Pollard (1984).

DEFINITION 5. A P-Brownian bridge $B_P$ is a zero-mean Gaussian process indexed by $\mathscr{F}$, a class of functions which is contained in $L_2(P)$, with covariance kernel

$$cov(B_P(f)\, B_P(g)) = E_P(fg) - E_P(f)\, E_P(g), \qquad f, g \in \mathscr{F} \tag{3.1}$$

and the process has bounded and $d_p$-uniformly continuous sample paths, where $d_p^2(f, g) = E_P(f-g)^2 - (E_P(f-g))^2$.

In order to investigate the limit behavior of the statistics under the null hypothesis and local alternatives, we shall describe it in a general framework. As we know, if under local alternative, the distribution of the random variables $x_1, ..., x_m$ is contiguous to a distribution $P$ say, it will be dependent on $m$, the size of the sample. Hence we here assume that the underlying probability measure of $x_i$ is $P^{(m)}$. Under the null hypothesis, $P^{(m)} = P$. The following states the asymptotic properties of $W_{m1}$ and $W_{m2}$.

THEOREM 3.1. *Assume that the underlying probability measure of $x_i$, $P^{(m)}$ and a probability measure $P$ satisfy the following*:

(a) *Both* $P^{(m)}(I(a'x<t))$ *and* $P(I(a'x<t))$ *are jointly continuous functions of* $(a, t)$;

(b) $\sup_{a,t} |P^{(m)}(I(a'x<t)) - P(I(a'x<t))| \to 0$.

*Then the empirical process* $\{W_m(a) - \sqrt{m}\,(P^{(m)}\{a'x \leqslant 0\} - 1/2) : a \in S^d\}$ *converges weakly to a P-Brownian bridge* $\{W_P(a) : a \in S^d\}$ *with covariance kernel*

$$k(a, b) = E_P(I(a'x \leqslant 0)\, I(b'x \leqslant 0)) - E_P\{a'x \leqslant 0\}\, E_P\{b'x \leqslant 0\}. \tag{3.2}$$

*Hence, under the null hypothesis* $H_0$, $\sqrt{m}\,(P^{(m)}\{a'x \leqslant 0\} - 1/2) = 0$ *and*

$$W_{m1} \Rightarrow \sup_a |W_P(a)|, \tag{3.3}$$

$$W_{m2} \Rightarrow \int (W_P(a))^2\, d\mu(a). \tag{3.4}$$

*Furthermore, under local alternative with* $\sqrt{m}\,(P^{(m)}\{a'x \leqslant 0\} - 1/2) \to f(a)$ *uniformly over* $a \in S^d$, *then* $f(\cdot)$ *is continuous and bounded and*

$$W_{m1} \Rightarrow \sup_a |W_P(a) + f(a)|, \tag{3.5}$$

$$W_{m2} \Rightarrow \int (W_P(a) + f(a))^2 \, d\mu(a). \tag{3.6}$$

*Remark* 3.1. In Theorem 3.1, we assume that every marginal distribution of $P$ (and of $P^{(m)}$) at the projection direction $a$ is continuous with respect to $a$ and $t$. This assumption is stronger than that $P$ (and $P^{(m)}$) is continuous. On the other hand, it is weaker than that $P$ (and $P^{(m)}$) is absolute continuous.

As to the permutation empirical process $\{W_m(E_m, a) : a \in S^d\}$, we have the following limiting property. Define the sample probability space $(\Omega, \mathscr{B}, \mathscr{P})$ and for $\omega \in \Omega$, write the permutation empirical process as $\{W_m(\omega, E_m, a) : a \in S^d\}$.

THEOREM 3.2. *Assume that the conditions in Theorem* 3.1 *hold and assume further that*

$$\sup_{a,\, b} |E_{P^{(m)}}(I(a'x \leqslant 0)(I(b'x \leqslant 0))) - E_P(I(a'x \leqslant 0)\, I(b'x \leqslant 0))| \to 0 \tag{3.7}$$

*Then there exists a measure-one subset of* $\Omega$, $\Omega_0$, *such that for any* $\omega \in \Omega_0$ *the permutation empirical process* $\{W_m(\omega, E_m, a) : a \in S^d\}$ *converges weakly to a P-Brownian bridge* $\{PW(a) : a \in S^d\}$ *with covariance kernel*

$$k(a, b) = E_P(I(a'x \leqslant 0) - 1/2)(I(b'x \leqslant 0) - 1/2). \tag{3.8}$$

*Consequently, under the null hypothesis* $H_0$, *for any* $\omega \in \Omega_0$

$$W_{m1}(\omega, E_m) \Rightarrow \sup_a |PW(a)|, \tag{3.9}$$

$$W_{m2}(\omega, E_m) \Rightarrow \int (PW(a))^2 \, d\mu(a). \tag{3.10}$$

*Remark* 3.2. Comparing Theorem 3.1 with Theorem 3.2, we learn that, under the null hypothesis, the test statistic $W_{m1}$ (and $W_{m2}$) and the associated permutation one have the same limit. The tests can detect, in the large sample sense, local alternatives converging to the null as fast as square root $n$, a parametric rate. Furthermore, the tests can not only be applied to test the location shift problem, but they can also test whether $P\{a'x \leqslant 0\} = 1/2$ for all $a \in S^d$, a more general hypothesis.

Following the above conclusions, the convergence of quantiles is established immediately.

Denote by $\lambda_m^{(i)}(\cdot, E_m)$, $\lambda_m^{(i)}$ and $\lambda^{(i)}$ the $1-\alpha$ quantiles of the distributions of $W_{mi}(\cdot, E_m)$ given $X_m = \{x_1, ..., x_m\}$, $W_{mi}$, and $W_i$ $i = 1, 2$ respectively.

Corollary 3.3. *Assume that the iid sample* $\{x_1, ..., x_m\}$ *is from a continuous distribution F satisfying condition* (a) *in Theorem* 3.1. *Then under the null hypothesis* $H_0$, *for any* $\omega \in \Omega_0$

$$\lambda_m^{(i)}(\omega, E_m) \to \lambda(\alpha) \qquad \text{in Probab.} \tag{3.11}$$

$$\lambda_m^{(i)} \to \lambda(\alpha) \qquad \text{in Probab.} \tag{3.12}$$

*as* $m \to \infty$.

We now turn to the two-sample case. The following states the convergence of the empirical process $\{V_N(a) : a \in S^d\}$ and then of $V_{N,1}$ and of $V_{N,2}$.

Let $F^{(m)}$ and $G^{(n)}$ be, respectively, the distributions of $x_i's$ and $y_j's$ and let $F_a^{(n)}$ be the distribution of $a'x_i's$ and let $P^{(N)}$ be the empirical probability measure of $\{x_1, ..., x_m, y_1, ..., y_n\}$.

Theorem 3.4. *Assume that* $F_a^{(m)}(t)$, $G_a^{(n)}(t)$ *and two distributions* $F_a(t)$ *and* $G_a(t)$ *are continuous with respect to a and t and* $\lim_{m\to\infty} m/N = p$, $(0 < p < 1)$. *Assume further that*

$$\begin{aligned} &\sup_{a,t} |F_a^{(m)}(t) - F_a(t)| \to 0, \\ &\sup_{a,t} |G_a^{(n)}(t) - G_a(t)| \to 0 \end{aligned} \tag{3.13}$$

*as* $m, n \to \infty$. *Then the empirical process* $\{V_N(a) - \sqrt{nm/N}\,(P^{(N)}(a'x < a'y) - 1/2) : a \in S^d\}$ *converges weakly to a P-Brownian bridge* $\{V(a) : a \in S^d\}$ *with the covariance kernel*

$$\begin{aligned} k(a, b) = {} & (1-p)\, E_F(1 - G_a(a'x))(1 - G_b(a'x)) \\ & + pE_G(1 - F_a(a'y))(1 - F_b(a'y)). \end{aligned} \tag{3.14}$$

*Hence, under null hypothesis* $H_0$, $\sqrt{nm/N}\,(P^{(N)}(a'x < a'y) - 1/2) = 0$ *and*

$$V_{N,1} \Rightarrow \sup_a |V(a)| \tag{3.15}$$

*and*

$$V_{N,2} \Rightarrow \int (V(a))^2\, d\mu(a). \tag{3.16}$$

*Furthermore, under local alternative with* $\sqrt{nm/N}\,(P^{(N)}\{a'x \leqslant a'y\} - 1/2) \to g(a)$ *uniformly over* $a \in S^d$, *one has* $g(\cdot)$ *is continuous and bounded and*

$$V_{N,1} \Rightarrow \sup_a |V(a) + g(a)|, \tag{3.17}$$

$$V_{N,2} \Rightarrow \int (V(a) + g(a))^2 \, d\mu(a). \tag{3.18}$$

*Remark* 3.3. Piterbarg and Tyurin (1993, Theorem 1, p. 149) derived the limit null distribution of the test statistic when assuming the underlying continuous density is dominated by $c/(1 + \|x\|^r)$ for all $x \in R^d$ where $r \geqslant d$, the dimension of the variable $x$. We now need not assume this condition on the underlying distribution of the variable.

The limiting properties of the permutation empirical process $\{V_N(R_N, a): a \in S^d\}$ are stated as follows. Assume with no loss of generality that the random variables, $x_i$'s and $y_i$'s lie in a common sample probability space. Without confusion, the sample probability space is still defined by $(\Omega, \mathscr{B}, \mathscr{P})$.

THEOREM 3.5. *Assume the conditions in Theorem* 3.4 *hold. Then there exist a measure-one subset of* $\Omega$, $\Omega_0$ *such that for any* $\omega \in \Omega_0$, *the permutation empirical process* $\{V_N(\omega, R_N, a)\colon a \in S^d\}$ *converges weakly to a* $P$*-Brownian bridge* $\{RV(a) = \int B_H(a, a't)\, dH(t) : a \in S^d\}$, *where* $H(t) = dF + (1-d)\, G$. *Hence, under null hypothesis* $H_0$,

$$V_{N,1}(\omega, R) \Rightarrow \sup_a |RV(a)|, \tag{3.19}$$

$$V_{N,2}(\omega, R) \Rightarrow \int (RV(a))^2 \, d\mu(a). \tag{3.20}$$

Note that under the null hypothesis, the Gaussian processes in Theorem 3.4 and 3.5 are the same one. Following this conclusion, the convergence of the quantiles is established.

Denote by $\lambda_N^{(i)}(\cdot, R)$, $\lambda_N^{(i)}$ and $\lambda_V^{(i)}$ the $1-\alpha$ quantiles of the distributions of $V_{N,i}(\cdot, R)$ given $(X_m, Y_m)$, $V_{N,i}$, and $V_i$ $i = 1, 2$ respectively where $X_m = \{x_1, ..., x_m\}$ and $Y_n = \{y_1, ..., y_n\}$.

THEOREM 3.6. *Under the conditions in Theorem* 3.4, *for any* $\omega \in \Omega_0$

$$\lambda_N^{(i)}(\omega, R) \to \lambda_V^{(i)}(\alpha) \qquad \text{in Probab.} \tag{3.21}$$

$$\lambda_N^{(i)} \to \lambda_N^{(i)}(\alpha) \qquad \text{in Probab.} \tag{3.22}$$

*as* $n, m \to \infty$ *and* $m/N \to d$.

## 4. SIMULATIONS AND POWER COMPARISONS OF TESTS

In order to demonstrate the performance of the permutation tests, some small-sample simulation experiments and comparisons with other existing tests in the literature were performed. In order that the simulations are comparable with other tests, we used, in the one-sample cases, six different trivariate distributions and in the two-sample cases, four other trivariate distributions, which have been considered by Randles (1989, p. 1048) and Randles and Peters (1990, p. 4231–4233) already. On the other hand, in order to take into account the influence of the dimension, we perform some simulations with different dimensions of the variables. The results will be presented in the following subsections. There is one word about the calculating the test statistics. The statistics are all the supremum or the integration on the unit supersphere surface $S^d$. For calculating them, we randomly generate $l$ variables $a_i$, $i=1, ..., l$ distributed uniformly on $S^d$ and use the maximum or the average of the statistics at every direction $a_i$ instead of the test statistics in the practical use, that is, for instance, $W_{m1l}=\max_{1\leqslant i\leqslant l}|W_m(a_i)|$ is used instead of $W_{m1}$. In the simulation, we chosen $l=dn^2$.

### 4.1. *One-sample Cases*

In the simulation results reported in the Tables below, the sample size is $n=20$, the dimension of random variable, $x$, $di=3$, and the following distributions of the variable are investigated in which the first five distributions were located at $\mu=(t\theta, t\theta, t\theta)'$ for $t=0, 1, 2, 3$, (see Randles 1989):

Normal I—$x$ has the trivariate standard normal distribution $N(0, I_3)$, $\theta=0.15$;

Uniform—$x=zu^{1/3}/(z'z)^{1/2}$ where $z\sim N(0, I_3)$ and $u$ is independent and uniform (0, 1), $\theta=0.08$;

$T$—$x=z/(S/3)^{1/2}$ where $z\sim N(0, I_d)$ and $S$ is independent with a chi-square distribution with 3 df, $\theta=0.19$;

Cauchy—$x$ has the same distribution as a $T$ with 1 df, $\theta=0.21$;

Skewed I—$x=20z(1+z_1(z'z)^{-1/2})$ where $z\sim N(0, I_3)$ and $z_1$ is the first component of $z$, $\theta=0.8$;

Skewed II—the distribution of $x$ is the same as in Skewed I except that the direction of shift was changed to $\mu=(t\theta, 0, 0)'$, $\theta=4.0$.

The basic experiment was performed 1000 times. The nominal level was 0.05. Tables 1 and 2 show the proportion of times out of 1000 that each

TABLE 1

Simulation Power of Test $W_{m1}$ in One-sample Case, $m=20$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Uniform | 0.0430 | 0.3340 | 0.7650 | 0.9560 |
| Normal I | 0.0550 | 0.1030 | 0.3040 | 0.5850 |
| Cauchy | 0.0480 | 0.1240 | 0.3020 | 0.5480 |
| Student $t$ | 0.0540 | 0.1190 | 0.3480 | 0.6580 |
| Skewed I | 0.0520 | 0.0810 | 0.1750 | 0.3120 |
| Skewed II | 0.0450 | 0.0900 | 0.1480 | 0.2560 |

procedure rejected $H_0$ as the simulation power. Table 1 shows the simulated power of the test $W_{m1}$ and Table 2 for $W_{m2}$.

For revealing the influence of the dimension of variable, the simulation with different dimensions was also performed. The underlying distribution of the variable was Normal I listed above. The dimensions conducted were from 3 to 8. The sample size was 20. The basic experiment was performed 1000 times. The nominal level was 0.05. The simulation results are presented in Tables 3 and 4 for the tests $W_{m1}$ and $W_{m2}$.

Look at Tables 1 and 2. The first finding is that $W_{m2}$ is better than $W_{m1}$ in most of the cases we conducted. When the underlying null distribution is elliptically symmetric, either light-tailed or heavy-tailed, we see that, comparing with three tests listed in Randles (1989, p. 1048), the performance of $W_{m2}$ is better than that of them. The performance of $W_{m1}$ is not encouraging except for the uniform case where $W_{m1}$ is better than all three in Randles (1989). On the other hand, when the distribution is skewed, the performance of the tests in Randles (1989) is better than both tests in this paper.

TABLE 2

Simulation Power of Test $W_{m2}$ in One-sample Case, $m=20$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Uniform | 0.0480 | 0.5260 | 0.9120 | 0.9960 |
| Normal I | 0.0460 | 0.1300 | 0.4120 | 0.7340 |
| Cauchy | 0.0520 | 0.1180 | 0.4120 | 0.7240 |
| Student $t$ | 0.0440 | 0.1380 | 0.4920 | 0.8380 |
| Skewed I | 0.0410 | 0.0940 | 0.1640 | 0.2840 |
| Skewed II | 0.0580 | 0.1060 | 0.3020 | 0.4480 |

TABLE 3

Simulation Power of Test $W_{m1}$ in One-sample Case, $m = 20$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| dim 3 | 0.0460 | 0.1300 | 0.4210 | 0.7340 |
| dim 4 | 0.0160 | 0.04100 | 0.2090 | 0.4950 |
| dim 5 | 0.0301 | 0.1020 | 0.2520 | 0.5530 |
| dim 6 | 0.0300 | 0.0830 | 0.3230 | 0.6980 |
| dim 7 | 0.0410 | 0.0870 | 0.4110 | 0.8170 |
| dim 8 | 0.0260 | 0.0940 | 0.3780 | 0.7790 |

From Table 4, it is somewhat surprised that the test $W_{m2}$ has higher power with higher dimension which lies in a moderate region and is able to hold the level as well. However with getting higher dimension (dimension is larger than or equal to 7) the test may not continue to maintain in this status. This is reasonable due to that sample size 20 is too small according to the dimension. For $W_{m1}$, the performance is not encouraging, even in the case of dimension 4, the significance level cannot be held. Hence in the practical use, the integration type statistic may be better to use.

## 4.2. *Two-sample Cases*

The Monte Carlo results presented are based on 1000 pairs of samples, each of size 15. The permutation procedure is performed 1000 times for determining the critical values. As before, the simulation power is the proportion of rejecting he null out of 1000.

The samples are from the following four different distributions which have been investigated by Randles and Peters (1990) already. The locations

TABLE 4

Simulation Power of Test $W_{m2}$ in One-sample Case, $m = 20$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| dim 3 | 0.0550 | 0.1030 | 0.3040 | 0.5850 |
| dim 4 | 0.0440 | 0.1420 | 0.5020 | 0.8560 |
| dim 5 | 0.0520 | 0.1720 | 0.6060 | 0.9220 |
| dim 6 | 0.0520 | 0.1520 | 0.6160 | 0.9580 |
| dim 7 | 0.0680 | 0.1820 | 0.6900 | 0.9740 |
| dim 8 | 0.0160 | 0.1040 | 0.5800 | 0.9800 |

TABLE 5

Simulation Power of Test $W_{N,1}$ in Two-sample Case, $m=n=15$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Normal II | 0.0330 | 0.2550 | 0.7510 | 0.9170 |
| Chauchy | 0.0310 | 0.1000 | 0.2110 | 0.2140 |
| Mixed I | 0.0400 | 0.1620 | 0.7600 | 0.9610 |
| Mixed II | 0.0300 | 0.0920 | 0.3940 | 0.6930 |

of distributions of $x$ and $y$ are $\theta=(0,0,0)$ and $\Delta=t(a,a,a)$ for the Cauchy, mixed I and mixed II below and $\Delta=t(a,a,0)$ for the normal II.

Normal II—$x$ has the trivariate normal distribution $N(0, V)$, where the element of $V$, say $v_{ij}=0.9$ for $i\neq j$; $=1$ for $i=j$ and $a=-0.2$;

Cauchy—$x$ has the Cauchy distribution listed in one-sample case and $a=0.2$;

Mixed I—$x$ has the distribution which is obtained by selecting the probability 0.9 from a $N(0, I_3)$ and with probability 0.1 from $N(0, 400\times I_3)$ and $a=0.4$;

Mixed II—$x$ has the distribution which is obtained by selecting the probability 0.9 from a $N(0, V)$ and with probability 0.1 from $N(0, 400\times V)$ where $V$ is the same as in Normal II in subsection 4.1, and $a=0.4$.

In two-sample cases we conducted, the permutation test $W_{N,2}$ can hold the significance level and it, comparing to the power of the tests investigated in Randles and Peters (1990), does well whether light-tailed or heavy-tailed symmetric distribution. For the mixed distributions, the

TABLE 6

Simulation Power of Test $W_{N,2}$ in Two-sample Case, $m=n=15$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Normal II | 0.0500 | 0.2210 | 0.7180 | 0.9370 |
| Chauchy | 0.0510 | 0.1160 | 0.2470 | 0.2690 |
| Mixed I | 0.0380 | 0.1900 | 0.7110 | 0.9480 |
| Mixed II | 0.0550 | 0.1410 | 0.5130 | 0.7800 |

TABLE 7

Simulation Power of Test $W_{N,1}$ in Two-sample Case, $m = n = 15$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| dim 3 | 0.0330 | 0.2550 | 0.7510 | 0.9170 |
| dim 4 | 0.0420 | 0.1310 | 0.4760 | 0.8660 |
| dim 5 | 0.0360 | 0.1240 | 0.4330 | 0.8130 |
| dim 6 | 0.0370 | 0.1090 | 0.3630 | 0.7550 |
| dim 7 | 0.0300 | 0.1140 | 0.3280 | 0.6810 |
| dim 8 | 0.0390 | 0.0820 | 0.3190 | 0.5970 |

permutation ones still do well in the mixed I case, but the power of them is in the mixed II case lower than those in Randles and Peters (1990). The performance of $W_{N1}$ is not so encouraging. The actual percentage of $H_0$ has been rejected is lower than the nominal level. (See Tables 5 and 6.)

We also performed some simulations to evidence how the influence of the data sparseness on the test effect is. The underlying distribution of the variable was Normal II listed above. The dimension conducted were from 3 to 8. The sample sizes were still both 15. The replication time and the nominal level were the same as before. The results are reported in Tables 7 and 8 for the tests $W_{N,1}$ and $W_{N,2}$.

Similar to one-sample case, the performance of integration type test $W_{N,2}$ is still encouraging. Even in the case of dimension 8, the test can still hold the level and have good power although it is getting slowly lower with increasing the dimension. But $W_{N,1}$ will not be able to hold the level.

TABLE 8

Simulation Power of Test $W_{N,2}$ in Two-sample Case, $m = n = 15$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| dim 3 | 0.0500 | 0.2210 | 0.7180 | 0.9370 |
| dim 4 | 0.0560 | 0.1740 | 0.6080 | 0.9050 |
| dim 5 | 0.0480 | 0.1280 | 0.4990 | 0.8830 |
| dim 6 | 0.0420 | 0.1350 | 0.4720 | 0.8520 |
| dim 7 | 0.0490 | 0.1070 | 0.3830 | 0.7520 |
| dim 8 | 0.0440 | 0.1360 | 0.3680 | 0.7310 |

## 5. A CONCLUDING REMARK

We would like to give a comment on a possible application to testing problem in regression.

Suppose the underlying model is

$$Y = (y_1, ..., y_N)' = \Theta + T(X) + \varepsilon, \tag{5.1}$$

where $Y$ is the $N \times d$ data matrix of dependent variables, $X$ is a known full rank $N \times p$ design matrix, $T$ is an unknown smooth function of $X$, $\Theta = (\theta, ..., \theta)'$ is a $N \times d$ matrix of unknown location parameters and $\varepsilon$ is an $N \times d$ matrix of random errors. We here assume the rows of $\varepsilon$ are *iid* $d$-variate random vectors with the distribution $F$ satisfying that the marginal distribution $F_a(t)$ at the projection direction $a$ is continuous with respect to $a$ and $t$. The hypothesis to be tested is $H_0$: $T(\cdot) \equiv 0$. This is testing whether the predictor has no effect on a response variable. It is a multivariate extension of univariate problem investigated, for example, Manson and Jernigan (1989), Buckley (1991), Barry and Hartigen (1990), Eubankand Hart (1993), Stute (1997), Stute *et al.* (1998) and Zhu and Lam (1994). The following is a test constructed by the procedure in Section 2. Break $Y$ into two parts $Y_1 = (y_1, ..., y_{[N/2]})'$ and $Y_2 = (y_{[N/2]+1}, ..., y_N)'$ and then construct a test similar to that in Section 2. If the sample size $N$ is even, then let $z_j(y_j - y_{[N/2]+j})$

$$T_{N,1} = \int \max_{1 \leqslant i \leqslant [N/2]} \left( \sqrt{2/N} \sum_{j=1}^{i} (I(a'z_j \leqslant 0) - 1/2) \right)^2 d\mu(a). \tag{5.2}$$

If the size $N$ is odd, the test may be constructed as

$$T_{N,2} = \int \max_{1 \leqslant i \leqslant [N/2],\, 1 \leqslant l \leqslant [N/2]+1} \times \left( 1/\sqrt{N} \sum_{j=1}^{i} \sum_{m=1}^{l} (I(a'y_j \leqslant a'y_{[N/2]+m}) - 1/2) \right)^2 d\mu(a). \tag{5.3}$$

## 6. APPENDIX

### *Proofs of Theorems*

#### *Proofs for the Conclusions of Section* 2

*Proof of Proposition* 2.2. Recall the notation "$a \cdot b$" in Section 2. Without loss of generality, assume that $x$ is reflectedly symmetric, that is,

$x$ and $-x$ have the same distribution. Note that $x = e \cdot (e \cdot x) =: e \cdot x^*$ where $e = \pm 1$ with probability value one half. Via some elementary calculation, the following equivalence holds: that $x$, $e$ and $x^*$ are independent and $x$, $x^*$ have the same distribution is equivalent to that $x$ is reflectedly symmetric. Let $X_m^* = (x_1^*, ..., x_m^*)$, and define $E_m \circ E_m^{(1)} = (e_1 \cdot e_1^{(1)}, ..., e_m \cdot e_m^{(1)})'$. Then the set $\{W_{mi}(X_m) > l - [l\alpha]$ of $W_{mi}(E_m^{(j)}, X_m)'s\}$ equals exactly the set $\{W_{mi}(E_m, X_m^*) > l - [l\alpha]$ of $W_{mi}(E_m \circ E_m^{(j)}, X_m^*), j = 1, ..., l\}$. It is easy to check that $\{E_m, E_m \circ E_m^{(j)},\ j = 1, ..., l\}$ are *iid* $m$-dimensional variables. Indeed, for any $t$ and $s \in \mathscr{E}$, a set consisting of all $n$-dimensional variables of the form $(\pm 1, ..., \pm 1)$, say $\{t_1, ..., t_{2^m}\}$,

$$\begin{aligned} P\{E_m \circ E_m^{(1)} = t, E_m \circ E_m^{(2)} = s\} &= \frac{1}{2^m} \sum_{j=1}^{2^m} P\{E_m^{(1)} \circ t_j = t, E_m^{(2)} \circ t_j = s\} \\ &= \frac{1}{2^m} \sum_{j=1}^{2^m} \frac{1}{2^m} \frac{1}{2^m} \\ &= P\{E_m \circ E_m^{(1)} = t\} P\{E_m \circ E_m^{(2)} = s\}. \end{aligned}$$

The independence between $E_m$ and $E_m \circ E_m^{(j)}$ can be checked in the same way. Hence, when given $X_m^*$, $W_{mi}(E_m, X_m^*)$ and $W_{mi}(E_m \circ E_m^{(j)}, X_m^*)$, are $l+1$ *iid* variables, which implies that

$$P\{W_{mi}(E_m, X_m^*) > l - [l\alpha] \text{ of } W_{mi}(E_m \circ E_m^{(j)}, X_m^*)'\, s \mid X_m^*\} \leqslant \frac{[l\alpha]+1}{l+1}.$$

The proof is concluded from integrating $X_m^*$.

*Proof of Proposition* 2.3. Proving this conclusion is a standard argument (e.g. Bickel, 1969). Hence the detail of the proof is omitted.

*Proofs for the Theorems in Section* 3

*Proof of Theorem* 3.1. The conclusion is just a direct consequence of Corollary 2.7 of Gine and Zinn (1991, p. 771) since the conditions in that result can be satisfied in our case.

*Proof of Theorem* 3.2. It is known that (e.g. see Alexander, 1984)

$$\begin{aligned} \sup_{a,b} \Bigg| \frac{1}{m} \sum_{i=1}^{m} (I(a'x_i \leqslant 0) - 1/2)(I(b'x_i \leqslant 0) - 1/2) \\ - E_{P^{(m)}}(I(a'x_i \leqslant 0) - 1/2)(I(b'x_i \leqslant 0) - 1/2) \Bigg| \to 0 \quad \text{a.s.} \end{aligned}$$

as $m \to \infty$. Combining (3.7), we have

$$\sup_{a,b} \left| \frac{1}{m} \sum_{i=1}^{m} (I(a'x_i \leqslant 0) - 1/2)(I(b'x_i \leqslant 0) - 1/2) - E_P(I(a'x \leqslant 0) - 1/2)(I(b'x \leqslant 0) - 1/2) \right| \to 0 \quad \text{a.s.}$$

Consider the random variable $x$ as a function on the sample space $\Omega$. Let

$$\Omega_0 = \left\{ \omega \colon \sup_{a,b} \left| \frac{1}{m} \sum_{i=1}^{m} (I(a'x_i(\omega) \leqslant 0) - 1/2)(I(b'x_i(\omega) \leqslant 0) - 1/2) - E_P(I(a'x(\omega) \leqslant 0) - 1/2)(I(b'x(\omega) \leqslant 0) - 1/2) \right| \to 0 \right\}. \tag{6.1}$$

$\Omega_0$ is then a subset of the sample space $\Omega$ with probability measure one. In the following, we always assume without further mentioning that $\omega \in \Omega_0$ for the given $\{x_1(\omega), ..., x_m(\omega), ...\}$. Without confusion, we simply write $\{x_1, ..., x_m, ...\}$ for $\{x_1(\omega), ..., x_m(\omega), ...\}$. We need to prove the *fidis* convergence and the uniform tightness of the process. The *fidis* convergence can be easily achieved by the CLT together with (6.1). So omit the details. As to the uniform tightness, all we need to do is to show that for any $\eta > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ for which

$$\limsup_{m \to \infty} P\{\sup_{[\delta]} |W_m(E_m, a) - W_m(E_m, b)| > \eta \mid X_m\} < \varepsilon, \tag{6.2}$$

where $[\delta] = \{(a, b) \colon d(a, b) = \sqrt{E(I(a'x \leqslant 0) - I(b'x \leqslant 0))^2} \leqslant \delta\}$. Since the limiting properties are being investigated as $m \to \infty$, $m$ is always considered to be large enough below which simplifies some arguments of proof:

Note that $W_m(E_m, a) = (1/\sqrt{m}) \sum_{i=1}^{m} (I(a'e_i \cdot x_i \leqslant 0) - 1/2) = (1/\sqrt{m}) \sum_{i=1}^{m} e_i(I(a'x_i \leqslant 0) - 1/2)$. Write $P_m^\circ$ for the sighed measure that places mass $e_i/m$ at $x_i$. We then write the LHS of (6.2) in another form:

$$\limsup_{m \to \infty} P\{\sup_{[\delta]} \sqrt{m}\, |P_m^\circ(I(a'x \leqslant 0) - I(b'x \leqslant 0))| > \eta \mid X_m\}. \tag{6.3}$$

Let $\langle 2\delta \rangle = \{(a, b) \colon d_m(a, b) = \sqrt{(1/\sqrt{m}) \sum_{i=1}^{m} (I(a'x_i \leqslant 0) - I(b'x_i \leqslant 0))^2} \leqslant 2\delta\}$. By the uniform strong law of large numbers, we have

$$\sup_{a} \left| \frac{1}{m} \sum_{i=1}^{m} (I(a'x_i \leqslant 0) - I(b'x_i \leqslant 0)) - E(I(a'x \leqslant 0) - I(b'x \leqslant 0) \right| \to 0 \quad \text{a.s.} \tag{6.4}$$

as $m \to \infty$. Then

$$P\{[\delta] \in \langle 2\delta \rangle\} \to 0 \tag{6.5}$$

as $m \to \infty$. Consequently, the value of (6.3) can be bounded by

$$\limsup_{m \to \infty} P\{\sup_{\langle 2\delta \rangle} \sqrt{m}\, |P_m^\circ(I(a'x \leqslant 0) - I(b'x \leqslant 0)| > \eta \mid X_m\}. \tag{6.6}$$

Note that the class of functions $e(I(a'x \leqslant 0) - 1/2)$ over $S^d$ is a V-C class. Following almost the same argument of the equicontinuity lemma (see, e.g. Pollard 1984, pp. 150–151), (6.2) can be derived. We here omit the details and complete the proof.

In the following we first prove Theorem 3.5, the argument can be used to prove Theorem 3.4.

*Proof of Theorem* 3.5. Let $H_{Na}(t) = (m/N)\, F_{ma}(t) + (n/N)\, G_{na}(t))$ and $H_a(t) = pF_a(t) + (1-p)\, G_a(t))$. Applying Theorem 1 of Praestgaard (1995, p. 309), for almost all series $\{x_1, ..., x_m,\}$ and $\{y_1, ..., y_n\}$

$$\begin{aligned}
&\{\sqrt{nm/N}\,(F_{ma}^R(t) - G_{na}^R(t)) : a \in S^d,\ t \in R^1\} \\
&\quad = \{\sqrt{mN/n}\,(F_{ma}^R(t) - H_{Na}(t)) : a \in S^d,\ t \in R^1\} \\
&\quad \Rightarrow RV_H =: \{RV_H(a, t) : a \in S^d,\ t \in R^1\},
\end{aligned} \tag{6.7}$$

where $RV_H$ is a $P$-Brownian bridge where $H = pF + (1-p)\, G$. The convergence is convergence in distribution in $l^\infty(\mathscr{F})$ consisting of all of bounded, real-valued functions defined on $\mathscr{F}$ where $\mathscr{F}$ is the class of indicator functions of half spaces $\{a' \cdot \leqslant t\}$. As usual, (see, Dudley, 1978, p. 901 or Gine and Zinn, 1984, 1986), the supremum norm on this space is considered. Note that all of sample paths of $RV_H$ is contained in $C(\mathscr{F}, H)$, a sub-collection consisting of all bounded, uniformly continuous function under the semimetric defined in (3.1). It is known that $C(\mathscr{F}, H)$ is separable (e.g. see Pollard, 1984, p. 169, ex. 7). Furthermore, any point in $C(\mathscr{F}, H)$ can easily be showed to be completely regular (Pollard, 1984, p. 67). By the representation theory (e.g. Pollard, 1984, p. 71), we have, under uniform norm,

$$\begin{aligned}
&\{\sqrt{mN/n}\,(F_{ma}^R(t) - H_{Na}(t)) : a \in S^d,\ t \in R^1\} \\
&\quad \to \{RV_H(a, t) : a \in S^d,\ t \in R^1\} \quad \text{a.s.}
\end{aligned} \tag{6.8}$$

Consequently, recalling $(z_1, ..., z_N) = (x_1, ..., x_m, y_1, ..., y_n)$ and letting $H_N(\cdot) = 1/N \sum_{j=1}^N I(z_j \leqslant \cdot)$,

$$\begin{aligned}
\sqrt{nm/N} & \int F_{ma}^R(t)\, dG_{na}^R(t) - 1/2 \\
&= \sqrt{nm/N} \int (F_{ma}^R(t) - G_{na}^R(t))\, dG_{na}^R(t) + o(1) \\
&= \sqrt{mN/n} \int (F_{ma}^R(t) - H_{Na}(t))\, dG_{na}^R(t) + o(1) \\
&= \sqrt{mN/n}\,(N/n) \int (F_{ma}^R(t) - H_{Na}(t))\, dH_{Na}(t) \\
&\quad - \sqrt{mN/n}\,(m/n) \int (F_{ma}^R(t) - H_{Na}(t))\, dF_{ma}^R(t) + o(1) \\
&= \sqrt{mN/n}(N/n) \int (F_{ma}^R(t) - H_{Na}(t))\, dH_{Na}(t) \\
&\quad \text{(integrating by parts)} \\
&\quad - \sqrt{mN/n}\,(m/n)/2 + \sqrt{mN/n}\,(m/n) \\
&\quad - \sqrt{mN/n}\,(m/n) \int F_{ma}^R(t)\, dH_{Na}(t) + o(1) \\
&= \sqrt{mN/n} \int (F_{ma}^R(t) - H_{Na}(t))\, dH_{Na}(t) + o(1) \\
&= \int RV_H(\cdot, a, t)\, dH_{Na}(t) + o_p(1) \\
&= \int RV_H(\cdot, a, a's)\, dH_N(s) + o_p(1).
\end{aligned} \tag{6.9}$$

For each $\omega \in \Omega$, $RV_H(\omega, a, a's)$ is continuous on $S^d \times R^d$ and uniformly bounded. Let $\sup_{a, s} |RV_H(\omega, a, a's)| =: c$. Since $\sup_s |H_N(s) - H(s)| \to 0$ a.s. as $N \to \infty$, then for any cube $\delta \in R^d$, $|\int (I(z \in \delta))\, dH_N - E_H(I(z \in \delta))| \to 0$ a.s. Denote $\Delta_k = [-k, k]^d$ and $\Delta_k^c$ its complement set in $R^d$. For any $\varepsilon > 0$, there exists a constant $k$ such that

$$\sup_a \left| \int I(s \in \Delta_k^c)\, RV_H(\omega, a, a's)\, d(H_N(s) - H(s)) \right|$$

$$\leqslant c \int I(s \in \Delta_k^c)\, d(H_N(s) - H(s))| \leqslant \varepsilon. \tag{6.10}$$

Since $RV_H(\omega, a, a's)$ is uniformly continuous on $S^d \times \Delta_k$, there exist then a partition of $\Delta_k$, say, $B$ small cubes $\delta_1, ..., \delta_B$ such that for $s_l^* \in \delta_l$, $l = 1, ...B$

$$\sup_{s \in \delta_j} \sup_{a \in S^d} |RV_H(\omega, a, a's_j^*) - RV_H(\omega, a, a's)| \leqslant \varepsilon. \tag{6.11}$$

Hence for any $\omega \in \Omega_0$

$$\sup_a \left| \int I(s \in \Delta)\, RV_H(\omega, a, a's)\, d(H_N(s) - H(s)) \right|$$

$$\leqslant \sum_{l=1}^{B} \sup_a \left| \int (I(s \in \delta_l)\, RV_H(\omega, a, a's)\, d(H_N(s) - H(s)) \right.$$

$$\left. - I(s \in \delta_l)\, RV_H(\omega, a, a's_l^*)\, d(H_N(s) - H(s)) \right|$$

$$+ c \sum_{l=1}^{B} \left| \int (I(s \in \delta_l))\, dH_N - E_H(I(s \in \delta_l)) \right|$$

$$\leqslant \sum_{l=1}^{B} \sup_a \left| \int (I(s \in \delta_l)\, RV_H(\omega, a, a's)\, d(H_N(s) - H(s)) \right.$$

$$\left. - I(s \in \delta_l)\, RV_H(\omega, a, a's_l^*)\, d(H_N(s) - H(s)) \right| + o_p(1)$$

$$\leqslant \varepsilon + o_p(1). \tag{6.12}$$

The proof is concluded from combining (6.9).

*Proof of Theorem* 3.4. Applying Corollary 2.7 of Gine and Zinn (1991, p. 771), it is easy to see that $\{\sqrt{m}\,(F_{ma}(t) - F_a^{(m)}) : a \in S^d,\ t \in R^1\}$ and $\{\sqrt{n}\,(G_{na}(t) - G_a^{(n)}) : a \in S^d,\ t \in R^1\}$ converge weakly, respectively, to the $P$-Brownian bridges $B_F$ and $B_G$. Following the argument in proof of Theorem 3.5, we have

$$\sqrt{nm/N}\left\{\int F_{ma}(t)\,dG_{na}(t)-\int F_a^{(m)}(t)\,dG_a^{(n)}(t)\right\}$$

$$=\sqrt{nm/N}\int (F_{ma}(t)-F_a^{(m)}(t))\,dG_{na}(t)$$

$$-\sqrt{nm/N}\int (G_{na}(t)-G_a^{(n)}(t))\,dF_a^{(m)}(t)+o(1)$$

$$=\sqrt{1-p}\int (B_F(a,t))\,dG_{na}(t))$$

$$-\sqrt{p}\int (B_G(a,t))\,dF_a^{(m)}(t)+o(1)$$

$$=\sqrt{1-p}\int (B_F(a,t))\,dG_a(t))$$

$$-\sqrt{p}\int (B_G(a,t))\,dF_a(t)+o(1). \tag{6.13}$$

The proof is completed.

## ACKNOWLEDGMENT

We thank N. Henze who is at Karlsruhe University for carefully reading for our draft, showing us some references and kindly telling us that the reflected symmetry implies the angular symmetry. We also thank a referee whose suggestions improved the presentation of the paper.

## REFERENCES

1. K. S. Alexander, Probability inequalities for empirical processes and a law of the iterated logarithm, *Ann. Probab.* **12** (1984), 1041–1067.
2. D. Barry and J. A. Hartigen, An omnibus test for departures from constant mean, *Ann. Statist.* **18** (1990), 1340–1357.
3. P. J. Bickel, A distribution free version of the Smirnov two sample test in the p-variate case, *Ann. Math. Statist.* **40** (1969), 1–23.
4. I. Blumen, A new bivariate sign test for location, *J. Amer. Statist. Assoc.* **53** (1958), 448–456.
5. B. M. Brown and T. P. Hettmansperger, Affine invariant rank methods in the bivariate location model, *J. Roy. Statist. Soc. B* **45** (1987), 25–30.
6. B. M. Brown and T. P. Hettmannsperger, An affine invariant bivariate version of the sign test, *J. Roy. Statist. Soc. B* **51** (1989), 117–125.
7. B. M. Brown, T. P. Hettmannsperger, J. Nyblom, and H. Oja, On certain bivariate sign tests and medians, *J. Amer. Statist. Assoc.* **87** (1992), 127–135.

8. M. J. Buckley, Detecting a smooth signal: optimality of cusum based procedures, *Biometrika* **78** (1991), 252–262.
9. D. L. Donoho, "Breakdown Properties of Multivariate Location Estimators," qualifying paper, Harvard Univ., 1982.
10. D. L. Donoho and M. Gasko, Multivariate generalization of the median and trimmed means (I), Unpublished.
11. R. M. Dudley, Central limit theorems for empirical measures, *Ann. Probab.* **6** (1978), 899–929.
12. R. L. Eubank and J. D. Hart, Commonality of cusum, von Neumann and smoothing-based goodness-of-fit tests, *Biometrika* **80** (1993), 89–98.
13. E. Gine and J. Zinn, Some limit theorems for empirical processes, *Ann. Probab.* **12** (1984), 927–989.
14. E. Gine and J. Zinn, Lecture on the central limit theorem for empirical processes, *in* "Lecture Notes in Mathematics," Vol. 1221, pp. 50–113, Springer-Verlag, New York, 1986.
15. E. Gine and J. Zinn, Gaussian characterization of uniform Donsker classes of functions, *Ann. Probab.* **19** (1991), 900–929.
16. T. P. Hettmannsperger, "Statistical Inference Based on Ranks," Ch. 6, Wiley, New York, 1984.
17. T. P. Hettmannsperger, J. Nyblom, and H. Oja, Affine invariant multivariate one-sample sign tests, *J. Roy. Statist. Soc. B* **56** (1994), 221–234.
18. T. P. Hettmannsperger and H. Oja, Affine invariant multivariate multisample sign tests, *J. Roy. Statist. Soc. B* **56** (1994), 235–249.
19. J. L. Hodge, A bivariate sign test, *Ann. Math. Statist.* **26** (1955), 523–527.
20. R. Y. Liu, On a notion of data depth based upon random simplices, *Ann. Statist.* **18** (1990), 405–414.
21. R. Y. Liu and K. Singh, A quality index based on data depth and multivariate rank tests, *J. Amer. Statist. Assoc.* **88** (1993), 252–260.
22. P. J. Manson and R. W. Jernigan, A cubic spline extension of the Durbin-Watson test, *Biometrika* **76** (1989), 39–47.
23. J. S. Maritz, "Distribution-free Statistical Methods," Ch. 7, Chapman and Hall, London, 1981.
24. H. Oja and J. Nyblom, On bivariate sign tests, *J. Amer. Statist. Assoc.* **85** (1989), 249–259.
25. D. Peters and R. H. Randles, Multivariate signed-rank tests for the one-sample location problems, *J. Amer. Statist. Assoc.* **85** (1990), 552–557.
26. D. Peters and R. H. Randles, A bivariate signed rank test for the two-sample location problem, *J. Roy. Statist. Soc. B* (1991).
27. V. I. Piterbarg and Yu. N. Tyurin, Testing for homogeneity of two multivariate samples: A Gaussian field on a sphere, *Math. Methods Statist.* **2** (1993), 147–164.
28. D. Pollard, "Convergence of Stochastic Processes," Springer-Verlag, New York, 1984.
29. J. T. Praestgaard, Permutation and bootstrap Kolmogorov-Smirnov tests for the equality of two distributions, *Scand. J. Statist.* **22** (1995), 305–322.
30. M. L. Puri and P. K. Sen, "Nonparametric Methods in Multivariate Analysis," Ch. 4, Wiley, New York, 1971.
31. R. H. Randless, A distribution-free multivariate sign test based on interdirections, *J. Amer. Statist. Assoc.* **85** (1989), 1045–1050.
32. R. H. Randles and D. Peters, Multivariate rank tests for the two-sample location problem, *Communs. Statist. Theory Math.* **15** (1990), 4225–4238.
33. J. P. Romano, Bootstrap and randomization tests for some nonparametric hypotheses, *Ann. Statist.* **17** (1988), 141–159.

34. W. A. Stahel, "Robuste Schätzungen: In Finitesimale Optimalität und Schätzungen von Kovarianzmatrizen," ETH, Zürich, 1981.
35. W. Stute, Nonparametric model checks for regression, *Ann. Statist.* **25** (1997), 613–641.
36. W. Stute, W. G. Manteiga, and M. P. Quidimil, Bootstrap approximations in model checks for regression, *J. Amer. Statist. Assoc.* **93** (1998), 141–149.
37. J. W. Tukey, Mathematics and picturing data, *in* "Proceedings of International Congress of Mathematics, Vancouver," Vol. 2, pp. 525–531, 1975.
38. L. X. Zhu and Y. Lam, On Buckley's test for no effect of the predictor on a response variable, unpublished paper, 1994.