Statistical Methods in Medical Research
http://smm.sagepub.com/

Multivariate tests based on interpoint distances with application to magnetic resonance imaging

Marco Marozzi

Stat Methods Med Res published online 16 April 2014
DOI: 10.1177/0962280214529104

The online version of this article can be found at:
http://smm.sagepub.com/content/early/2014/04/09/0962280214529104

Published by:
SAGE
http://www.sagepublications.com

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–18

DOI: 10.1177/0962280214529104
smm.sagepub.com
SAGE

Multivariate tests based on interpoint distances with application to magnetic resonance imaging

Marco Marozzi

Abstract

The multivariate location problem is addressed. The most familiar method to address the problem is the Hotelling test. When the hypothesis of normal distributions holds, the Hotelling test is optimal. Unfortunately, in practice the distributions underlying the samples are generally unknown and without assuming normality the finite sample unbiasedness of the Hotelling test is not guaranteed. Moreover, high-dimensional data are increasingly encountered when analyzing medical and biological problems, and in these situations the Hotelling test performs poorly or cannot be computed. A test that is unbiased for non-normal data, for small sample sizes as well as for two-sided alternatives and that can be computed for high-dimensional data has been recently proposed and is based on the ranks of the interpoint Euclidean distances between observations. Five modifications of this test are proposed and compared to the original test and the Hotelling test. Unbiasedness and consistency of the tests are proven and the problem of power computation is addressed. It is shown that two of the modified interpoint distance-based tests are always more powerful than the original test. Particularly, the modified test based on the Tippett criterium is suggested when the assumption of normality is not tenable and/or in case of high-dimensional data with complex dependence structure which are typical in molecular biology and medical imaging. A practical application to a case-control study where functional magnetic resonance imaging is used is discussed.

Keywords

high-dimensional data, case-control study, hypothesis testing, nonparametric tests, interpoint distance

1 Introduction

Multivariate data are very often considered in practice. For example, in medical studies an array of p measurements is generally collected from each subject to describe him/her/its health status. In dealing with such data, we should pay special attention because if each measurement of the array is only studied marginally, then important features of the data may not be detected. Another example

Department of Economics, Statistics and Finance, University of Calabria, Rende, Italy

Corresponding author:
Marco Marozzi, Department of Economics, Statistics and Finance, University of Calabria, Via Bucci 0C, 87036 Rende (CS), Italy.
Email: marco.marozzi@unical.it

is forensic research where a sample of items (for example glass fragments) from the crime scene has been measured using an elemental composition technique on p elements. A multivariate test, generally the Hotelling test,[1] is used to gather evidence against the null hypothesis that glass fragments recovered from a suspect and a control sample of glass fragments from the crime scene share a common source. Unfortunately, the Hotelling test has several practical drawbacks when the normal assumption is not tenable and when analyzing high-dimensional data which are increasingly encountered in many biological and medical applications such as microarray and functional magnetic resonance (MR) imaging. In these situations, the Hotelling test performs poorly or cannot even be computed.

In this paper, we wish to test the null hypothesis that two samples of p-dimensional observations come from the same distribution, continuous but unknown. Some authors assessed this problem by considering the ranks of geometric entities of the data set.[2–4] Other authors proposed permutation tests.[5,6] Tests based on distances between observations have been proposed for example by Baringhaus and Franz,[7] Rosenbaum[8] and Jureckova and Kalina.[9]

Let ψ be a test of the null hypothesis

$$H_0 = \{\text{distribution } G \text{ of a random vector } \mathbf{X} \text{ belongs to the set } \mathcal{G}\} \tag{1}$$

against the alternative hypothesis

$$H_1 = \{G \in \mathcal{F}\}.$$

Let ψ be of size α that is

$$\sup_{G \in \mathcal{G}} E_G[\psi(\mathbf{X})] \leq \alpha, \quad 0 \leq \alpha \leq 1.$$

A natural property is the unbiasedness. ψ is unbiased if its power is not smaller than α, the nominal significance level

$$\inf_{G \in \mathcal{F}} E_G[\psi(\mathbf{X})] \geq \alpha.$$

Jureckova and Kalina[9] emphasize that while many tests are asymptotically unbiased, their finite sample unbiasedness is still an open question in particular when the alternative is two-sided. An example is the univariate two-sample Wilcoxon test: while it is always (provided that distributions may differ only in location) unbiased against one-sided alternatives, it is generally not unbiased against two-sided alternatives. An unbiased test can be found only when the distribution functions behind the data satisfy special conditions. In the case of the Wilcoxon test, it is a symmetric density. However, with unknown G, it is very difficult to assess whether the necessary special conditions hold.

In this paper, we propose five modifications of a multivariate test for two-sample location problems based on the ranks of interpoint distances of observations proposed by Jureckova and Kalina[9] which can be applied to non-normal as well as high-dimensional data with complex dependence structure. The original test as well as the modifications are described in Section 2 with formal proofs of unbiasedness and consistency given in the Appendix. Section 3 addresses the problem of power estimation of the tests. Size and power of the new tests are compared in Section 4 with those of the original interpoint distance test and of the well-known Hotelling test. A practical application to a case-control study where MR imaging is used is discussed in Section 5. Some concluding remarks are given in Section 6.

2 Interpoint distance-based tests

Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)$ and $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ be two independent random samples from p-variate populations with continuous cumulative distribution functions $G(\mathbf{Z})$ and $G(\mathbf{Z} - \boldsymbol{\mu})$, respectively, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ is the location shift parameter with $-\infty < \mu_h < \infty$, $h = 1, \ldots, p$. We wish to test

$$\text{the null hypothesis } H_0 : \boldsymbol{\mu} = \mathbf{0}$$

against

$$\text{the two-sided alternative } H_1 : \boldsymbol{\mu} \neq \mathbf{0},$$

where $\mathbf{0}$ denotes a vector of length p of all zeros. Let $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_N)$ be the pooled sample with $\mathbf{Z}_i = \mathbf{X}_i = (X_{i1}, \ldots, X_{ip})', i = 1, \ldots, m$ and $\mathbf{Z}_{m+j} = \mathbf{Y}_j = (Y_{j1}, \ldots, Y_{jp})', j = 1, \ldots, n$ with $N = m + n$. The most familiar method for testing H_0 against H_1 is the Hotelling HT test that rejects H_0 for large values of

$$HT = (\overline{\mathbf{X}} - \overline{\mathbf{Y}})' \mathbf{V}^{-1} (\overline{\mathbf{X}} - \overline{\mathbf{Y}})$$

where $\overline{\mathbf{X}} = (\bar{X}_1, \ldots, \bar{X}_p)', \bar{X}_h = \frac{1}{m} \sum_{i=1}^{m} X_{ih}, \overline{\mathbf{Y}} = (\bar{Y}_1, \ldots, \bar{Y}_p)', \bar{Y}_h = \frac{1}{n} \sum_{j=1}^{n} Y_{jh}, h = 1, \ldots, p,$

$$\mathbf{V} = \left(\frac{1}{m} + \frac{1}{n} \right) \frac{(m-1)\mathbf{V}_1 + (n-1)\mathbf{V}_2}{N-2},$$

$$\mathbf{V}_1 = \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})',$$

$$\mathbf{V}_2 = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{Y}_j - \overline{\mathbf{Y}})(\mathbf{Y}_j - \overline{\mathbf{Y}})'.$$

Assuming that G is the multivariate normal distribution, under the null hypothesis

$$\frac{N-p-1}{(N-2)\,p} HT \sim F_{p,\, N-p-1}$$

where $F_{p,\, N-p-1}$ is the F distribution with p and $N-1-p$ degrees of freedom. Therefore, H_0 is rejected at the α nominal level of significance if $HT > F_{p,\, N-p-1}(1-\alpha)$ where $F_{p,\, N-p-1}(1-\alpha)$ is the $(1-\alpha)100$th percentile of the F distribution with p and $N-1-p$ degrees of freedom. Under normality, the test is the optimal unbiased test. Under non-normality, its finite sample unbiasedness is not guaranteed. Two more severe practical drawbacks of the Hotelling test are as follow:

(1) It shows poor performance for high-dimensional data. Bai and Saranadasa[10] showed that its power decreases as p/N gets larger with $p/N \to d \in [0, 1)$. The reason is the presence of the inverse of the $\mathbf{V}$ matrix in the HT statistic. When p and N are of the same order, the sample covariance matrix may not converge to the population covariance matrix and therefore

standardizing by the covariance, which is good for low-dimensional data, is bad for high-dimensional data;
(2) It cannot be computed when $p > N - 2$ because $\mathbf{V}$ cannot be inverted.

These two drawbacks, not to mention that the HT test requires the assumption of normality, make it unuseful in addressing many medical and biological problems where p is comparable with or even (much) larger than N. Examples are genomics, transcriptomics, and proteomics data in molecular biology, clinical trials where many blood chemistry measurements are taken on each subject and medical imaging (like functional MR).

Jureckova and Kalina[9] proposed a test for testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ against $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$ which is finite sample distribution free, unbiased and that can be computed also when $p > N$. These features make it of particular interest in medicine and molecular biology. The test is based on the ranks of interpoint distances of observations. To perform the test, choose at random one of the observations of the first sample and let $\mathbf{X}_i$ be it. Compute the distances between $\mathbf{X}_i$ and the other observations of the pooled sample $\mathbf{Z}$ obtaining the vector $\mathbf{L}_i(\mathbf{Z})$ of $N-1$ interpoint distances $l_{ik}; k \neq i$. Consider the Euclidean distance, that is

$$l_{ik} = ||\mathbf{X}_i - \mathbf{Z}_k|| = \sqrt{\sum_{h=1}^{p} (X_{ih} - Z_{kh})^2}.$$

Compute the ranks R_{ik} of l_{ik} $k = 1, \ldots, N, k \neq i$. Finally, compute the test statistic as

$$JK_i = \sum_{k=m+1}^{N} R_{ik}.$$

Large values of JK_i are evidence against the null hypothesis. Note that the JK_i test is the one-sided Wilcoxon rank sum test applied to the two samples of interpoint distances of observations with fixed $\mathbf{X}_i$ (that is conditionally given $\mathbf{X}_i$): the first sample is l_{ik} $k = 1, \ldots, m, k \neq i$ and the second sample is l_{ik} $k = m + 1, \ldots, N$. Let QJK_i be the p-value of the JK_i test, which can be easily computed using, for example, the R software, and the null hypothesis is rejected at the α nominal level of significance if $QJK_i < \alpha$. Note that QJK_i as well as JK_i can be used as the test statistic because they are one to one decreasingly related (large values of JK_i speak against H_0 as small values of QJK_i do). It is important to note that the JK_i test is one-sided because under the alternative hypothesis that $\boldsymbol{\mu} \neq \mathbf{0}$, the interpoint distances between $\mathbf{X}_i$ and the second sample observations are expected to be stochastically greater than the interpoint distances between $\mathbf{X}_i$ and the first sample observations. Therefore, the test is unbiased, see also Theorem 1 in the Appendix. Moreover, it is also distribution-free, that is the null distribution of the test statistic does not depend on the unknown distribution underlying the samples. Note that if H_0 is true, the distributions underlying $\mathbf{X}$ and $\mathbf{Y}$ are equal and then also the hypothesis $\tilde{H}_0$ that the distribution of the interpoint distances are the same. When H_0 is false, then also $\tilde{H}_0$ is false.

Jureckova and Kalina[9] emphasized that every convenient homogeneous combination of statistics $JK_1, \ldots, JK_m$ leads to a rank test which is distribution-free. They choose the simplest combination: the randomization of $JK_1, \ldots, JK_m$, that is to choose at random one of $JK_1, \ldots, JK_m$. The aim of

this paper is to see whether it is possible to obtain more powerful tests than the original JK_i test, while retaining unbiasedness and distribution freeness, by a more efficient use of the data. Before doing this, we propose a very simple modification of the JK_i test, its permutation version: the PJK_i test. The steps for performing it are as follow:

(1) Choose at random one of the observations of the first sample, and let $\mathbf{X}_i$ be it;
(2) Compute the p-value QJK_i of the corresponding JK_i test, this is the observed value of the test statistic: ${}_0PJK_i = QJK_i$;
(3) Randomly permute the elements of $\mathbf{Z}$, obtaining the first permuted pooled sample ${}_1\mathbf{Z} = (Z_{u_1^*}, \ldots, Z_{u_N^*})$ where $u_1^*, \ldots, u_N^*$ is a random permutation of unit labels $1, \ldots, N$. Compute the p-value $QJK_{u_i^*}$ of the JK test corresponding to the $JK_{u_i^*}$ statistic; this is the first permutation value of the test statistic: ${}_1PJK_i^* = QJK_{u_i^*}$;
(4) Repeat step 3 for the $B - 1$ other permutations of $\mathbf{Z}$, where $B = N!/(m!n!)$;
(5) Reject the null hypothesis if the p-value

$$QPJK_i = \frac{1}{B}\sum_{b=1}^{B} I({}_bPJK_i^* < {}_0PJK_i)$$

is less than α, where $I(.)$ denotes the indicator function.

Note that the permutation test is justified because under the null hypothesis that the distributions to be compared are the same, sample observations (the column vectors of matrix $\mathbf{Z}$) are exchangeable. The test is unbiased as well as consistent, see Theorems 1 and 2, respectively, in the Appendix.

To modify the JK_i test, we first apply the algorithm for the PJK_i test for all $i = 1, \ldots, m$ obtaining an array of m tests: $PJK_1, \ldots, PJK_m$ whose corresponding test statistics are equally distributed for $i = 1, \ldots, m$ under the null hypothesis and under the alternative hypothesis although not independent. The corresponding results are arranged in the following matrix

$$\begin{bmatrix} {}_0QPJK_1 & \ldots & {}_0QPJK_i & \ldots & {}_0QPJK_m \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ {}_bQPJK_1^* & \ldots & {}_bQPJK_i^* & \ldots & {}_bQPJK_m^* \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ {}_BQPJK_1^* & \ldots & {}_BQPJK_i^* & \ldots & {}_BQPJK_m^* \end{bmatrix}.$$

Note that under the null hypothesis, all permutations of the order of the columns of the matrix are equally likely. To combine the m tests, we combine each row of the matrix according to a certain proper criterium M obtaining the observed value of the modified JK test statistic

$${}_0M_{JK} = M({}_0QPJK_1, \ldots, {}_0QPJK_m)$$

and the B permutation values of the modified JK test statistic

$${}_bM_{JK}^* = M({}_bQPJK_1^*, \ldots, {}_bQPJK_m^*), \quad b = 1, \ldots, B.$$

The p-value of the modified JK test is

$$QM_{JK} = \frac{1}{B}\sum_{b=1}^{B} I({}_bM_{JK}^* < {}_0M_{JK}).$$

We consider the median, Tippett, Liptak and Fisher criteria which respectively lead to

$$ {}_cMED_{JK} = \text{median}({}_cQPJK_1, \ldots, {}_cQPJK_m) = \begin{cases} {}_cQPJK_{(\frac{m+1}{2})} & \text{if } m \text{ is odd} \\ \frac{1}{2}({}_cQPJK_{(\frac{m}{2})} + {}_cQPJK_{(\frac{m}{2}+1)}) & \text{if } m \text{ is even} \end{cases} $$

$$ {}_cTIP_{JK} = \max_{i=1,\ldots,m}(1 - {}_cQPJK_i) = {}_cQPJK_{(1)} $$

$$ {}_cLIP_{JK} = \sum_{i=1}^{m} \Phi^{-1}({}_cQPJK_i) $$

$$ {}_cFIS_{JK} = \sum_{i=1}^{m} \log({}_cQPJK_i) $$

where ${}_cQPJK_{(i)}$ is the i-th order statistic among ${}_cQPJK_1, \ldots, {}_cQPJK_m$, with $c = 0$ for the observed value of the test statistic and $c = b$ for the b-th permutation value, $b = 1, \ldots, B$. From Theorems 1 and 2, it easily follows that the combined tests are unbiased and consistent.

The testing approach considered in this paper can be further modified by considering other distances than the Euclidean one as well as other rank tests than the Wilcoxon one. For example, one may study whether multivariate location/scale problems can be assessed using modified JK tests based on the Cucconi rank test for jointly testing for location and scale differences.[11] One might also study how to extend to the multivariate case the combined test for scale proposed by Marozzi.[12]

3 Estimation of the power function of the tests

It is very difficult, if not even impossible, to derive theoretically optimality properties of a nonparametric test without any assumptions on the population distributions.[13] To solve this problem, we can rely on Monte Carlo simulations to estimate the power function of the tests. Let pv be the true p-value of the test of interest which rejects H_0 at level α if $pv \leq \alpha$. Let MC be the number of Monte Carlo simulations. By generating MC independent data sets and computing the proportion of rejections of H_0, we can estimate the power function. The resulting root mean squared error $RMSE$ when the rejection probability 0.05 is estimated is

$$ RMSE = \sqrt{0.05(1 - 0.05)/MC}. $$

$RMSE$ is equal to 0.00308, 0.00487, and 0.00689, respectively, for $MC = 5000$, 2000, and 1000. $RMSE$ is bounded by

$$ \max RMSE = 0.5/\sqrt{MC} $$

that is reached when estimating a rejection probability of 0.5. $\max RMSE$ is equal to 0.00707, 0.01118, and 0.01581 respectively for $MC = 5000$, 2000, and 1000. In theory, the true p-value pv can be computed by considering all the possible permutations $B = \frac{N!}{n!m!}$ of the pooled sample. Unfortunately, this is rarely feasible in practice, unless the sample sizes are very small. Therefore, we consider a simple random sample of size C drawn with replacement from the B permutations of the pooled sample and then compute the estimated p-value $\widehat{pv}$ as the proportion of permutations

with test statistic greater than or equal to the observed one. As a consequence, the corresponding $RMSE$ will be greater than before, when the true p-value is used.

The aim is to find the optimal number of permutations to estimate the power function of the test given the Monte Carlo simulation size MC. Note that this estimation involves the generations of $MC \cdot C$ data sets because for every iteration of the Monte Carlo loop, we have a permutation loop of C iterations for computing $\widehat{pv}$. In addressing this problem, it is very important to emphasize that power estimation concerns the accept/reject decision on H_0 that is different from p-value estimation.

Let $W_{H_1}(\alpha) = \Pr(pv < \alpha)$ be the true power function of the test at an alternative H_1. For each Monte Carlo simulation, $B\widehat{pv}$ is a random variable with binomial distribution with parameters C, pv conditional on the Monte Carlo simulation, because the true p-value pv is estimated as $\widehat{pv}$ using a random sample of C permutations of $\mathbf{Z}$ drawn with replacement from the population of B permutations. Unconditionally, the estimated Monte Carlo power function is

$$\hat{W}_{H_1, MC, C}(\alpha) = \frac{1}{MC} \sum_{mc=1}^{MC} I(\widehat{pv}_{mc} \leq \alpha).$$

Its mean is

$$E(\hat{W}_{H_1, MC, C}(\alpha)) = W_{H_1, C}(\alpha) = \sum_{h=0}^{[\alpha C]} \binom{C}{h} \int_0^1 t^h (1-t)^{C-h} \mathrm{d}\, W_{H_1}(t),$$

where $[\alpha C]$ is the greatest integer part of αC.[14] If the p-values have a beta distribution function with parameters $a, b > 0$ then

$$W_{H_1, C}(\alpha) = \frac{1}{beta(a,b)} \sum_{h=0}^{[\alpha C]} \binom{C}{h} beta(h + a, C - h + b),$$

where $beta(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t$ is the beta function with parameters a, b. Note that $W_{H_1, C}(\alpha)$ is the distribution function of a beta-binomial random variable evaluated at $[\alpha C]$. The mean squared error (MSE) of $\hat{W}_{H_1, MC, C}(\alpha)$ is

$$MSE_{H_1}(\alpha) = [W_{H_1}(\alpha) - W_{H_1, C}(\alpha)]^2 + \frac{1}{MC} W_{H_1, C}(\alpha)[1 - W_{H_1, C}(\alpha)]$$

because unconditionally $MC\hat{W}_{H_1, MC, C}(\alpha)$ has a binomial distribution with parameters SIM, $W_{H_1, C}(\alpha)$. When the null hypothesis is true

$$W_{H_0, C}(\alpha) = \frac{[\alpha C] + 1}{C + 1}$$

because the p-values are uniformly distributed. Note that $W_{H_0, C}(\alpha)$ is the distribution function of a uniform distribution evaluated at $[\alpha C]$. Oden[14] showed that, given MC, the optimal C in the sense that it minimizes the maximum MSE over α under the null hypothesis (that is in the uniform case) is $C = 2\sqrt{MC}$. Under the alternative (that is the beta case), it is suggested that $C = 4\sqrt{MC}$. However, Boos and Zhang[15] suggest to use $C = 8\sqrt{MC}$ which leads to $C = 566$, 358, and 253, respectively, for $MC = 5000$, 2000, and 1000. In these cases, when estimating a rejection probability of rej, it is expected an $RMSE$ close to $1.2\sqrt{rej(1-rej)/MC}$. Therefore, $\max_{rej}(RMSE) = 0.6/\sqrt{MC}$. Note that the maximum $RMSE$ is $0.5/\sqrt{MC}$ when the p-value of the test is computed considering all

B permutations. Therefore, using $(MC, C) = (5000, 566)$, (2000, 358), and (1000, 253) Monte Carlo simulations, the maximum $RMSE$ is, respectively, 0.00849, 0.01342, and 0.01897, while the $RMSE$ under the null hypothesis is 0.00370, 0.00585, and 0.00827. If it were computationally feasible to consider all B permutations, the maximum $RMSE$ would be 0.005 and the $RMSE$ under the null hypothesis would be 0.00218. These considerations lead us to design a simulation study with $MC = 5000$, 2000, and 1000 and $C = 500$, see the next section.

4 Size and power of the tests

In this section, we study and compare empirical size and power of the tests described in the previous section, as well as the permutation version PHT of the Hotelling test which can be easily implemented using an algorithm similar to that for the permutation JK_i test. In the first round of simulations, we simulate rather simple situations with low-dimensional multivariate variables with independent components. We consider both bivariate ($p = 2$) and quadrivariate ($p = 4$) independent standard normal, Cauchy, and exponential distributions. The two samples are generated independently one from the other. The nominal significance level is set to $\alpha = .05$. The desired location shift is obtained by adding $\boldsymbol{\mu} = (\mu, \ldots, \mu)'$ to the second sample elements. μ has been set in order to obtain a power close to .25, .50, .75 for the Hotelling test in the various simulation settings. The sample sizes are $m = n = 10$, 30, and 50, respectively, for Monte Carlo simulations with sizes 5000, 2000, and 1000. Permutation p-values have been estimated using a random sample of 500 permutations taken from the population of B permutations (see the previous section). While $\mu = 0$ simulations estimate the size of the test, $\mu \neq 0$ simulations estimate the power. Results for the first round of simulations are reported in Tables 1–3.

It is important to note that all tests are robust in size according to the definition of robustness suggested by Marozzi[16] that the maximum estimated significance level (MESL) of a test must be less than 1.5 times the nominal significance level. Here, the MESLs of the tests are .060 (HT), .057 (PHT), .064 (JK_i), .063 (PJK_i), .058 (MED_{JK}), .061 (TIP_{JK}), .059 (LIP_{JK}), and .059 (FIS_{JK}). As expected, the power of the tests increases as the sample sizes increase and as the location shift increases. Table 1 shows without surprise that under normal distributions, the Hotelling test and its permutation version are the most powerful tests. Among the other tests, the best one is the modified JK test based on the Tippett criterium. Table 2 shows that under the Cauchy distribution, the modified JK test based on the Tippett criterium is markedly the most powerful test. In this case, the permutation version of the Hotelling test is preferable to the original one that is too conservative (this excess of conservativeness negatively affects its power). Table 3 shows that under exponential distributions, the modified JK test based on the Tippett criterium is once again markedly more powerful than the other tests. Note that results for $p = 2$ and $p = 4$ are very similar.

It is important to note that the sample size setting does not influence the results of the first round of simulations. Therefore, we may conclude that the different number of Monte Carlo simulations does not influence the results and that a larger number of simulations, although would lead to size and power estimates closer to the true values, is not necessary in our study.

It should be emphasized that the parametric Hotelling test and its permutation version are always equivalent with one exception: when data come from the heavy tailed Cauchy distribution, the Hotelling test is very conservative and this negatively affects its power, whereas the permutation Hotelling test has a size much closer to the nominal significance level and then it is more powerful.

In the second round of simulations, we simulate much more complex and real situations in the context of molecular biology or medical imaging with high-dimensional multivariate variables with dependent components. More precisely, we consider $p = 20$, 80, and 1000 and we generate

Table 1. Power of the tests under normal distributions, $p = 2, 4$.

	$p=2$				$p=4$			
	$m=n=10$				$m=n=10$			
μ	0	0.52	0.77	1.01	0	0.46	0.66	0.85
HT	0.049	0.256	0.503	0.745	0.049	0.255	0.497	0.733
PHT	0.048	0.250	0.498	0.741	0.047	0.255	0.497	0.731
JK_i	0.047	0.129	0.251	0.374	0.040	0.134	0.241	0.353
PJK_i	0.054	0.142	0.265	0.391	0.048	0.148	0.257	0.367
MED_{JK}	0.051	0.144	0.275	0.460	0.050	0.152	0.290	0.468
TIP_{JK}	0.056	0.193	0.409	0.626	0.053	0.208	0.410	0.635
LIP_{JK}	0.048	0.130	0.251	0.428	0.047	0.139	0.275	0.451
FIS_{JK}	0.048	0.157	0.325	0.533	0.047	0.165	0.331	0.539
	$m=n=30$				$m=n=30$			
μ	0	0.28	0.41	0.55	0	0.24	0.34	0.44
HT	0.053	0.252	0.492	0.751	0.060	0.264	0.479	0.740
PHT	0.052	0.245	0.487	0.746	0.057	0.260	0.480	0.731
JK_i	0.053	0.116	0.201	0.291	0.050	0.109	0.181	0.243
PJK_i	0.054	0.119	0.207	0.298	0.050	0.110	0.178	0.248
MED_{JK}	0.057	0.107	0.189	0.320	0.051	0.099	0.168	0.279
TIP_{JK}	0.051	0.180	0.376	0.618	0.055	0.184	0.355	0.577
LIP_{JK}	0.057	0.089	0.162	0.266	0.052	0.090	0.148	0.252
FIS_{JK}	0.054	0.125	0.257	0.443	0.048	0.129	0.226	0.401
	$m=n=50$				$m=n=50$			
μ	0	0.22	0.32	0.42	0	0.18	0.26	0.34
HT	0.045	0.259	0.492	0.769	0.043	0.242	0.526	0.765
PHT	0.046	0.264	0.489	0.767	0.041	0.244	0.525	0.757
JK_i	0.056	0.117	0.186	0.278	0.036	0.107	0.135	0.217
PJK_i	0.057	0.118	0.191	0.275	0.043	0.110	0.136	0.216
MED_{JK}	0.056	0.098	0.151	0.264	0.041	0.100	0.150	0.218
TIP_{JK}	0.061	0.196	0.331	0.621	0.044	0.178	0.374	0.575
LIP_{JK}	0.057	0.078	0.122	0.199	0.046	0.092	0.128	0.196
FIS_{JK}	0.058	0.128	0.219	0.414	0.042	0.123	0.216	0.347

multivariate variables with dependent components using copula theory. Since the focus of the second round of simulations is on high and very high-dimensional situations with complex dependence structures, we consider standard normal distributed marginals. To model the dependence, we use the most popular types of copulas: the elliptical and Archimedean ones. For simplicity, to describe the copulas, we consider the bivariate case. Without going into much details, according to the well-known Sklar theorem, for any joint distribution function $D(x_1,x_2)$ there exists a copula K that describes the dependence of the two random variables X_1 and X_2, with distributions G_1 and G_2, completely

$$D(x_1, x_2) = K(G_1(x_1), G_2(x_2)) = K(u_1, u_2).$$

A copula K is a distribution function $K : [0, 1]^2 \to [0, 1]$ with uniformly distributed margins.[17] An elliptical copula is the copula corresponding to an elliptical distribution through the Sklar theorem.

Table 2. Power of the tests under Cauchy distributions, $p = 2, 4$.

	$p=2$				$p=4$			
	$m=n=10$				$m=n=10$			
μ	0	1.95	3.25	5.45	0	1.65	2.5	3.7
HT	0.017	0.248	0.513	0.756	0.016	0.244	0.495	0.753
PHT	0.045	0.350	0.603	0.818	0.051	0.357	0.609	0.833
JK_i	0.047	0.268	0.486	0.718	0.050	0.157	0.269	0.436
PJK_i	0.055	0.286	0.502	0.730	0.055	0.172	0.288	0.455
MED_{JK}	0.052	0.309	0.593	0.844	0.050	0.173	0.319	0.535
TIP_{JK}	0.052	0.445	0.712	0.892	0.055	0.280	0.472	0.665
LIP_{JK}	0.049	0.267	0.516	0.759	0.048	0.154	0.280	0.482
FIS_{JK}	0.049	0.332	0.620	0.867	0.049	0.184	0.339	0.557
	$m=n=30$				$m=n=30$			
μ	0	1.8	3	5	0	1.50	2.25	3.3
HT	0.023	0.260	0.486	0.756	0.018	0.264	0.501	0.768
PHT	0.057	0.354	0.591	0.821	0.050	0.382	0.616	0.843
JK_i	0.048	0.476	0.701	0.847	0.049	0.286	0.467	0.664
PJK_i	0.049	0.479	0.702	0.847	0.056	0.285	0.477	0.667
MED_{JK}	0.045	0.634	0.939	0.999	0.058	0.286	0.605	0.887
TIP_{JK}	0.054	0.895	0.992	1	0.047	0.678	0.919	0.992
LIP_{JK}	0.046	0.434	0.825	0.981	0.058	0.218	0.426	0.738
FIS_{JK}	0.045	0.698	0.964	1	0.059	0.337	0.663	0.918
	$m=n=50$				$m=n=50$			
μ	0	1.8	2.95	4.9	0	1.45	2.2	3.25
HT	0.021	0.253	0.496	0.741	0.014	0.239	0.499	0.743
PHT	0.056	0.342	0.596	0.805	0.041	0.353	0.608	0.806
JK_i	0.041	0.543	0.730	0.826	0.052	0.337	0.560	0.721
PJK_i	0.042	0.541	0.726	0.824	0.055	0.338	0.555	0.726
MED_{JK}	0.040	0.784	0.997	1	0.041	0.371	0.761	0.979
TIP_{JK}	0.055	0.990	1	1	0.054	0.871	0.997	1
LIP_{JK}	0.040	0.562	0.933	1	0.037	0.253	0.553	0.872
FIS_{JK}	0.039	0.889	0.999	1	0.038	0.467	0.859	0.986

Here, we consider the normal copula and the Cauchy copula which corresponds to the Student copula with one degree of freedom. The normal copula is defined as

$$K_\rho^N(u_1, u_2) = \Phi_\rho(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$$

where Φ_ρ denotes the distribution function of a bivariate standard normal random variable with parameter ρ and Φ^{-1} denotes the inverse of the univariate standard normal distribution. By replacing the normal with the Cauchy, we have the Cauchy copula K_ρ^{CAU}. It is important to emphasize that the contour plot for density of bivariate distributions defined with normal copula is elliptically shaped, whereas that corresponding to the Cauchy copula is star shaped.

A copula K^{ARCH} is called Archimedean if it can be represented as

$$K^{ARCH}(u_1, u_2) = \psi^{-1}(\psi(u_1) + \psi(u_2))$$

Table 3. Power of the tests under exponential distributions, $p = 2, 4$.

	$p=2$				$p=4$			
	$m=n=10$				$m=n=10$			
μ	0	0.46	0.69	0.94	0	0.41	0.59	0.78
HT	0.044	0.252	0.487	0.756	0.043	0.250	0.505	0.753
PHT	0.051	0.272	0.504	0.769	0.053	0.274	0.532	0.778
JK_i	0.048	0.213	0.352	0.495	0.052	0.185	0.299	0.453
PJK_i	0.056	0.228	0.369	0.512	0.057	0.199	0.316	0.470
MED_{JK}	0.050	0.189	0.401	0.650	0.052	0.144	0.311	0.578
TIP_{JK}	0.053	0.409	0.672	0.877	0.057	0.366	0.619	0.834
LIP_{JK}	0.050	0.211	0.466	0.753	0.051	0.160	0.342	0.644
FIS_{JK}	0.048	0.256	0.536	0.806	0.051	0.203	0.432	0.732
	$m=n=30$				$m=n=30$			
μ	0	0.27	0.4	0.53	0	0.23	0.33	0.43
HT	0.042	0.258	0.503	0.749	0.040	0.258	0.512	0.751
PHT	0.046	0.267	0.511	0.754	0.044	0.279	0.520	0.757
JK_i	0.053	0.228	0.360	0.450	0.050	0.173	0.256	0.378
PJK_i	0.052	0.230	0.365	0.454	0.052	0.178	0.261	0.381
MED_{JK}	0.051	0.193	0.397	0.660	0.055	0.104	0.197	0.426
TIP_{JK}	0.052	0.510	0.829	0.959	0.046	0.419	0.734	0.910
LIP_{JK}	0.052	0.231	0.481	0.743	0.055	0.135	0.240	0.491
FIS_{JK}	0.051	0.311	0.628	0.874	0.054	0.177	0.378	0.687
	$m=n=50$				$m=n=50$			
μ	0	0.21	0.31	0.41	0	0.17	0.25	0.33
HT	0.036	0.234	0.513	0.766	0.047	0.235	0.501	0.770
PHT	0.037	0.246	0.517	0.771	0.051	0.235	0.510	0.775
JK_i	0.048	0.219	0.345	0.440	0.064	0.138	0.254	0.356
PJK_i	0.044	0.219	0.350	0.437	0.063	0.132	0.255	0.354
MED_{JK}	0.056	0.164	0.371	0.661	0.049	0.077	0.153	0.340
TIP_{JK}	0.059	0.522	0.858	0.976	0.061	0.383	0.754	0.948
LIP_{JK}	0.059	0.224	0.433	0.722	0.053	0.113	0.208	0.415
FIS_{JK}	0.059	0.317	0.615	0.907	0.054	0.137	0.319	0.669

where ψ is called generator of the copula and is p-monotone. Archimedean copulas are familiar in practice because most of them admit an explicit formula for K and allow to model dependence with only one parameter. Here, we consider the Clayton copula which is generated by

$$\psi(t) = t^{-\vartheta} - 1, \quad \text{with } \vartheta \geq 0$$

and the Gumbel copula which is generated by

$$\psi(t) = (-\ln t)^{\xi}, \quad \text{with } \xi \geq 1.$$

The larger the parameters ϑ and ξ, the stronger the dependence. Note that the Clayton copula produces a tight correlation at the low end of each variable and that the contour plot for density is pear shaped. The Gumbel copula produces more correlation on the right tail than in the left tail. Note that Jureckova and Kalina[9] performed a very limited simulation study because they considered

Table 4. Power of the tests for high-dimensional data with complex dependence structure, $m = n = 20$, $p = 20$, 80, 1000.

	$p = 20$				$p = 80$				$p = 1000$			
μ	0	0.5	0.75	1	0	0.5	0.75	1	0	0.5	0.75	1
	Normal copula											
HT	0.056	0.113	0.232	0.407	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
PHT	0.055	0.111	0.229	0.403	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
JK_i	0.055	0.249	0.424	0.563	0.048	0.245	0.481	0.587	0.040	0.258	0.461	0.582
PJK_i	0.056	0.255	0.430	0.566	0.049	0.252	0.481	0.588	0.044	0.265	0.467	0.587
MED_{JK}	0.055	0.276	0.543	0.804	0.058	0.260	0.562	0.820	0.049	0.277	0.562	0.823
TIP_{JK}	0.055	0.388	0.736	0.951	0.060	0.402	0.782	0.958	0.043	0.424	0.802	0.961
LIP_{JK}	0.062	0.199	0.449	0.743	0.049	0.207	0.483	0.772	0.047	0.209	0.498	0.782
FIS_{JK}	0.057	0.322	0.668	0.911	0.054	0.331	0.704	0.921	0.046	0.344	0.708	0.925
	Cauchy copula											
HT	0.032	0.108	0.273	0.525	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
PHT	0.048	0.144	0.331	0.588	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
JK_i	0.051	0.258	0.462	0.634	0.056	0.267	0.470	0.626	0.055	0.272	0.490	0.617
PJK_i	0.052	0.266	0.469	0.637	0.058	0.273	0.475	0.631	0.055	0.279	0.496	0.622
MED_{JK}	0.054	0.248	0.542	0.837	0.056	0.261	0.578	0.837	0.055	0.274	0.563	0.827
TIP_{JK}	0.055	0.486	0.820	0.977	0.058	0.522	0.873	0.978	0.054	0.519	0.879	0.982
LIP_{JK}	0.057	0.185	0.445	0.753	0.059	0.205	0.479	0.751	0.051	0.212	0.456	0.738
FIS_{JK}	0.054	0.320	0.669	0.918	0.056	0.330	0.704	0.919	0.054	0.333	0.694	0.919
	Clayton copula											
HT	0.036	0.108	0.219	0.398	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
PHT	0.037	0.111	0.214	0.394	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
JK_i	0.043	0.232	0.421	0.561	0.044	0.237	0.434	0.582	0.053	0.240	0.448	0.578
PJK_i	0.049	0.236	0.427	0.565	0.047	0.242	0.436	0.586	0.061	0.247	0.451	0.584
MED_{JK}	0.043	0.240	0.534	0.777	0.049	0.262	0.519	0.789	0.061	0.216	0.475	0.772
TIP_{JK}	0.048	0.350	0.739	0.936	0.048	0.373	0.755	0.959	0.056	0.414	0.780	0.968
LIP_{JK}	0.049	0.145	0.392	0.669	0.047	0.172	0.378	0.687	0.057	0.146	0.365	0.672
FIS_{JK}	0.048	0.288	0.641	0.886	0.048	0.307	0.657	0.910	0.058	0.307	0.650	0.918
	Gumbel copula											
HT	0.047	0.102	0.218	0.373	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
PHT	0.048	0.110	0.228	0.382	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
JK_i	0.051	0.217	0.396	0.534	0.054	0.211	0.391	0.544	0.053	0.213	0.381	0.509
PJK_i	0.057	0.225	0.402	0.538	0.059	0.224	0.398	0.549	0.058	0.219	0.389	0.512
MED_{JK}	0.054	0.220	0.465	0.700	0.051	0.208	0.419	0.681	0.061	0.166	0.391	0.652
TIP_{JK}	0.054	0.350	0.716	0.924	0.047	0.362	0.728	0.935	0.061	0.394	0.756	0.957
LIP_{JK}	0.057	0.170	0.403	0.676	0.054	0.160	0.355	0.673	0.056	0.121	0.347	0.659
FIS_{JK}	0.052	0.279	0.615	0.851	0.049	0.272	0.561	0.849	0.058	0.235	0.567	0.845

only the bivariate case with independent normal and Cauchy distributed components. In the second round of simulations, we consider $m = n = 20$, $MC = 2000$, $C = 500$ and always the same shifts $\mu = 0.5$, 0.75, and 1 because the marginal distributions are always standard normal. The copula parameters ρ, ϑ, ξ are set to 0.5, 1, and 1.5, respectively. The results are displayed in Table 4 and are very interesting. All the tests are robust in size because their MESLs are .056 (HT), .055 (PHT), .056 (JK_i), .061 (PJK_i), .061 (MED_{JK}), .061 (TIP_{JK}), .062 (LIP_{JK}), .058 (FIS_{JK}). Besides the HT and PHT

tests that cannot be computed when $p > N$, it is important to note that the number of variables does not affect test performance. The best test is the modified JK test based on the Tippett criterium which is always markedly the most powerful one followed by the other modified JK tests in this order: FIS_{JK}, MED_{JK} and LIP_{JK}. The Hotelling test and its permutation version (which can be computed only for $p = 20$) are, as expected, the worst tests. They behave very similarly except for the Cauchy copula case where the PHT test is more powerful than the HT test because the HT test is excessively conservative. The original JK test and its permutation version behave very similarly and have intermediate performance between the modified JK tests and the Hotelling type tests.

The simulation study shows that our strategy of searching for a more efficient use of that data with respect to Jureckova and Kalina[9] has paid off very well. The test of choice is the modified interpoint distance test based on the Tippett criterium which is particularly suitable for high-dimensional data with complex dependence structure. Note that in these cases, the Hotelling test has poor performance or cannot even be computed.

5 Application to a case-control study where MR imaging is used

Kato et al.[18] designed and performed a case-control study to determine whether phase-contrast (PC) cine MR imaging could detect altered myocardial blood flow (MBF) response to the cold pressor test (CPT) in smokers. The CPT test is a non-pharmacological stress test that induces endothelium dependent coronary vasodilatation. The dysfunction of the endothelium is considered to be the first stage of atherosclerosis and assessing endothelial function is important in predicting future cardiovascular events in particular in smokers. In fact, smoking is an important risk factor for coronary heart disease and several studies showed that there is an increased risk of mortality even in young smokers.[19] PC cine MR imaging can be used to assess the MBF response to the CPT without exposing patients to radiation or using intravenous injections of contrast medium or radioactive tracers. Ten healthy male non-smokers (mean age 28 ± 5 years) and 10 age-matched male smokers (smoking duration ≥ 5 years, mean age 28 ± 3 years) without cardiovascular risk factors except for smoking were examined. During the study, PC cine MR has been used also to measure several cardiovascular characteristics of the subjects:

- EDV = end-diastolic volume (mL)
- EDVI = end-diastolic volume index = EDV/BSA (mL/m^2), where BSA = body surface area (m^2)
- ESV = end-systolic volume (mL)
- ESVI = end-systolic volume index = ESV/BSA (mL/m^2)
- SV = stroke volume (mL)
- EF = ejection fraction (%)
- LVM = left ventricular mass (g)
- LVMD = left ventricular mass drain (g)
- LVMI = left ventricular mass index = LVM/BSA (g/m^2).

It is of interest to determine whether there were significant differences in cardiovascular characteristics between case and control subjects, and therefore we analyze the data presented in Table 5 using the tests compared in the previous section. The null hypothesis is that the multivariate distribution of cardiovascular characteristics is the same for smokers (cases) and non-smokers (controls) against the two-sided alternative that there is a systematic difference in location between the distributions. The Hotelling test is not suggested in this situation because $N = 20$ is small and especially because p/N is close to 0.5 which is a case where the Hotelling test is not much

Table 5. Cardiovascular characteristics of the subjects measured by PC cine MR.

	Cardiovascular characteristics								
Subjects	EDV	EDVI	ESV	ESVI	SV	EF	LVM	LVMD	LVMI
Cases									
1	120	68	38	22	83	68	120	93	68
2	140	79	68	39	72	51	110	93	62
3	149	84	66	37	83	56	132	112	74
4	124	72	39	23	84	68	123	104	71
5	152	75	52	26	100	66	122	103	60
6	145	84	61	35	85	58	123	104	71
7	128	79	58	36	71	55	89	76	55
8	150	84	73	41	68	56	105	89	58
9	126	79	50	31	76	60	104	88	65
10	116	67	42	59	74	64	102	87	59
Controls									
1	184	111	67	40	117	64	110	80	67
2	126	70	54	30	72	57	104	81	58
3	140	79	59	33	81	58	125	108	71
4	117	69	39	23	78	66	108	87	63
5	107	65	44	27	63	59	93	79	57
6	114	65	40	23	74	65	109	93	63
7	122	70	51	30	70	58	120	95	69
8	133	84	57	36	76	57	103	80	65
9	123	63	49	25	74	60	104	90	53
10	128	76	56	34	71	56	104	87	63

powerful (see Bai and Saranadasa[10] and the simulation results of the previous section when $p = 20$ and $N = 40$). Conversely, the tests based on the ranks of interpoint distances are suggested in a situation like the one at issue. Since the sample sizes are small, it is computationally feasible to consider all $N!/(m!n!) = 20!/(10!10!) = 184756$ permutations of the pooled sample when computing p-values of the permutation tests. Thus exact p-values are computed. The p-values are .35985 (JK_i), .35206 (PJK_i), .07135 (MED_{JK}), .11144 (TIP_{JK}), .16259 (LIP_{JK}), and .10578 (FIS_{JK}), and lead us to conclude that the null hypothesis of no difference between the distributions of cardiovascular characteristics of smokers and non-smokers is plausible.

6 Conclusion

We analyzed the two-sided two-sample location problem in the context of high-dimensional data with complex dependence structure within the nonparametric framework by modifying a rank test based on interpoint distances. The modified tests are distribution free, exact, unbiased, and consistent. Particularly, the modified test based on the Tippett criterium was shown to perform very well also for small sample size-very large dimension situations typical of molecular biology and medical imaging studies or when a long array of measurements is collected from each subject. Therefore, our strategy of searching for a more efficient use of that data respect to Jureckova and Kalina[9] paid off very well. Although interpoint distance-based tests have been discussed here for the

location problem, they are likely to be useful also for other testing problems. Another direction of future research might involve the use of distance other than the Euclidean one.

Acknowledgment

We are very grateful to Prof. Dr. K. Kitagawa for kindly providing the data employed in Section 5.

References

1. Campbell GP and Curran JM. The interpretation of elemental composition measurements from forensic glass evidence III. *Sci Justice* 2009; **49**: 2–7.
2. Choi K and Marden JI. Tests of multivariate linear models using spatial concordances. *J Nonparamet Stat* 2005; **17**: 167–183.
3. Randles RH and Peters D. Multivariate rank tests for the two–sample location problem. *Commun Stat Theory Meth* 1990; **19**: 4225–4238.
4. Topchii A, Tyurin Y and Oja H. Inference based on the affine invariant multivariate Mann–Whitney–Wilcoxon statistic. *J Nonparamet Stat* 2003; **14**: 403–414.
5. Hall P and Tajvidi N. Permutation tests for equality of distributions in high-dimensional settings. *Biometrika* 2002; **89**: 359–374.
6. Neuhaus G and Zhu LX. Permutation tests for multivariate location problems. *J Multivariate Anal* 1999; **69**: 167–192.
7. Baringhaus L and Franz C. On a new multivariate two–sample test. *J Multivariate Anal* 2004; **88**: 190–206.
8. Rosenbaum PR. An exact distribution free test comparing two multivariate distributions based on adjacency. *J Royal Stat Soc Ser B* 2005; **67**: 515–530.
9. Jureckova J and Kalina J. Nonparametric multivariate rank tests and their unbiasedness. *Bernoulli* 2012; **18**: 229–251.
10. Bai Z and Saranadasa H. Effect of high dimension: by an example of a two sample problem. *Stat Sinica* 1996; **6**: 311–329.
11. Marozzi M. Some notes on the location–scale Cucconi test. *J Nonparametr Stat* 2009; **21**: 629–647.
12. Marozzi M. A modified Hall–Padmanabhan test for the homogeneity of scales. *Commun Stat Theory Meth* 2012; **41**: 3068–3078.
13. Hajek J, Sidak Z and Sen PK. *Theory of rank tests*, 2nd ed. New York, NY: Academic Press, 1998.
14. Oden NL. Allocation of effort in Monte Carlo simulation for power of permutation tests. *J Am Stat Assoc* 1991; **86**: 1074–1076.
15. Boos DD and Zhang J. Monte Carlo evaluation of resampling–based hypothesis tests. *J Am Stat Assoc* 2000; **95**: 486–492.
16. Marozzi M. Levene type tests for the ratio of two scales. *J Stat Comput Simul* 2011; **81**: 815–826.
17. Nelsen RB. *An introduction to Copulas*, 2nd ed. New York, NY: Springer, 2006.
18. Kato S, Kitagawa K, Yoon YE, et al. Detection of diminished response to cold pressor test in smokers: Assessment using phase-contrast cine magnetic resonance imaging of the coronary sinus. *Magnet Res Imaging* 2014, http://dx.doi.org/10.1016/j.mri.2013.12.015).
19. Teo KK, Ounpuu S, Hawken S, et al. Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study. *Lancet* 2006; **368**: 647–658.

Appendix

Theorem 1. The PJK_i test is unbiased for testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ against $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$.

Proof. We consider the following additive model

$$\begin{cases} \mathbf{X}_i = \boldsymbol{\mu} + \mathbf{W}_i & i = 1, \ldots, m \\ \mathbf{Y}_{m+j} = \mathbf{W}_{m+j} & j = 1, \ldots, n \end{cases}$$

where Ws are independent and identically distributed random variables with null location and finite scale. Let

$$\mathbf{Z}(\mathbf{0}) = (\mathbf{Z}_i(\mathbf{0}), i = 1, \ldots, N) = (\mathbf{W}_i, i = 1, \ldots, N)$$

denote the pooled sample under the null hypothesis and let

$$\mathbf{Z}(\boldsymbol{\mu}) = (\mathbf{Z}_i(\boldsymbol{\mu}), i = 1, \ldots, N) = (\boldsymbol{\mu} + \mathbf{W}_i, i = 1, \ldots, m; \mathbf{W}_{m+j}, j = 1, \ldots, n)$$

denote the pooled sample under the alternative hypothesis. We must prove that

$$QT(\mathbf{Z}(\boldsymbol{\mu})) \leq QT(\mathbf{Z}(\mathbf{0}))$$

that is

$$\Pr(QT \leq q|\mathbf{Z}(\mathbf{0})) \leq \Pr(QT \leq q|\mathbf{Z}(\boldsymbol{\mu})), \quad \forall q \in]0, 1[$$

where QT denotes the p-value of the $T = PJK_i$ test. Let

$$_0T(\mathbf{0}) = \sum_{k=m+1}^{N} R_{ik}(\mathbf{0})$$

denote the observed value of the test statistic under H_0, where $R_{ik}(\mathbf{0})$ is the rank of $l_{ik}(\mathbf{0}) = ||\mathbf{W}_i - \mathbf{W}_k||$. Let

$$_0T(\boldsymbol{\mu}) = \sum_{k=m+1}^{N} R_{ik}(\boldsymbol{\mu})$$

denote the observed value of the test statistic under H_1, where $R_{ik}(\boldsymbol{\mu})$ is the rank of $l_{ik}(\boldsymbol{\mu}) = ||\boldsymbol{\mu} + \mathbf{W}_i - \mathbf{W}_k||$.

Assume that $0 \leq \tau^* \leq \min(n, m)$ observations are exchanged between samples. It is supposed, without loss of generality, that the exchanged observations are the first τ^*. Let $(u_k^*, k = 1, \ldots, N)$ be the corresponding permutation of $(1, \ldots, N)$. Let $T^*(\mathbf{0})$ denote the permutation value of the test statistic under H_0 with

$$T^*(\mathbf{0}) = \sum_{k=m+1}^{N} R_{ik}^*(\mathbf{0}) = \begin{cases} \sum_{k=m+1}^{m+\tau^*} R_{iu_k^*}(\mathbf{0}) + \sum_{k=m+\tau^*+1}^{N} R_{ik}(\mathbf{0}) & i > \tau^* \\ \sum_{k=m+1}^{m+\tau^*} R_{u_i^* u_k^*}(\mathbf{0}) + \sum_{k=m+\tau^*+1}^{N} R_{u_i^* k}(\mathbf{0}) & i \leq \tau^* \end{cases}$$

where $R_{iu_k^*}(\mathbf{0})$ is the rank of $l_{iu_k^*}(\mathbf{0}) = ||\mathbf{W}_i - \mathbf{W}_{u_k^*}||$, $R_{u_i^* u_k^*}(\mathbf{0})$ is the rank of $l_{u_i^* u_k^*}(\mathbf{0}) = ||\mathbf{W}_{u_i^*} - \mathbf{W}_{u_k^*}||$, and $R_{u_i^* k}(\mathbf{0})$ is the rank of $l_{u_i^* k}(\mathbf{0}) = ||\mathbf{W}_{u_i^*} - \mathbf{W}_k||$.

Let $T^*(\boldsymbol{\mu})$ denote the permutation value of the test statistic under H_1 with

$$T^*(\boldsymbol{\mu}) = \sum_{k=m+1}^{N} R_{ik}^*(\boldsymbol{\mu}) = \begin{cases} \sum_{k=m+1}^{m+\tau^*} R_{iu_k^*}(\boldsymbol{\mu}) + \sum_{k=m+\tau^*+1}^{N} R_{ik}(\boldsymbol{\mu}) & i > \tau^* \\ \sum_{k=m+1}^{m+\tau^*} R_{u_i^* u_k^*}(\boldsymbol{\mu}) + \sum_{k=m+\tau^*+1}^{N} R_{u_i^* k}(\boldsymbol{\mu}) & i \leq \tau^* \end{cases}$$

where $R_{iu_k^*}(\boldsymbol{\mu}) = R_{iu_k^*}(\mathbf{0})$ is the rank of $l_{iu_k^*}(\boldsymbol{\mu}) = ||\boldsymbol{\mu} + \mathbf{W}_i - \boldsymbol{\mu} - \mathbf{W}_{u_k^*}|| = l_{iu_k^*}(\mathbf{0})$, $R_{u_i^* u_k^*}(\boldsymbol{\mu})$ is the rank of $l_{u_i^* u_k^*}(\boldsymbol{\mu}) = ||\mathbf{W}_{u_i^*} - \boldsymbol{\mu} - \mathbf{W}_{u_k^*}||$, and $R_{u_i^* k}(\boldsymbol{\mu}) = R_{u_i^* k}(\mathbf{0})$ is the rank of $l_{u_i^* k}(\boldsymbol{\mu}) = ||\mathbf{W}_{u_i^*} - \mathbf{W}_k|| = l_{u_i^* k}(\mathbf{0})$.

For $i > \tau^*$, it is

$$QT(\mathbf{Z}(\boldsymbol{\mu})) = \Pr(T^*(\boldsymbol{\mu}) \geq {}_0T(\boldsymbol{\mu})) = \Pr\left(\sum_{k=m+1}^{m+\tau^*} R_{iu_k^*}(\mathbf{0}) + \sum_{k=m+\tau^*+1}^{N} R_{ik}(\boldsymbol{\mu}) \geq \sum_{k=m+1}^{m+\tau^*} R_{ik}(\boldsymbol{\mu}) + \sum_{k=m+\tau^*+1}^{N} R_{ik}(\boldsymbol{\mu})\right)$$

which is less than or equal to

$$QT(\mathbf{Z}(\mathbf{0})) = \Pr(T^*(\mathbf{0}) \geq {}_0T(\mathbf{0})) = \Pr\left(\sum_{k=m+1}^{m+\tau^*} R_{iu_k^*}(\mathbf{0}) + \sum_{k=m+\tau^*+1}^{N} R_{ik}(\mathbf{0}) \geq \sum_{k=m+1}^{m+\tau^*} R_{ik}(\mathbf{0}) + \sum_{k=m+\tau^*+1}^{N} R_{ik}(\mathbf{0})\right)$$

because $l_{ik}(\boldsymbol{\mu}) > l_{ik}(\mathbf{0})\ \forall i,k$ and then $\sum_{k=m+1}^{m+\tau^*} R_{ik}(\boldsymbol{\mu}) \geq \sum_{k=m+1}^{m+\tau^*} R_{ik}(\mathbf{0})$.

For $i \leq \tau^*$, it is

$$QT(\mathbf{Z}(\boldsymbol{\mu})) = \Pr\left(\sum_{k=m+1}^{m+\tau^*} R_{u_i^* u_k^*}(\boldsymbol{\mu}) + \sum_{k=m+\tau^*+1}^{N} R_{u_i^* k}(\mathbf{0}) \geq \sum_{k=m+1}^{N} R_{ik}(\boldsymbol{\mu})\right)$$

which is less than or equal to

$$QT(\mathbf{Z}(\mathbf{0})) = \Pr\left(\sum_{k=m+1}^{m+\tau^*} R_{u_i^* u_k^*}(\mathbf{0}) + \sum_{k=m+\tau^*+1}^{N} R_{u_i^* k}(\mathbf{0}) \geq \sum_{k=m+1}^{N} R_{ik}(\mathbf{0})\right)$$

because $\sum_{k=m+1}^{N} R_{ik}(\boldsymbol{\mu}) - \sum_{k=m+1}^{N} R_{ik}(\mathbf{0}) \geq \sum_{k=m+1}^{m+\tau^*} R_{u_i^* u_k^*}(\boldsymbol{\mu}) - \sum_{k=m+1}^{m+\tau^*} R_{u_i^* u_k^*}(\mathbf{0})$ being $\tau^* \leq \min(n,m)$. ■

Theorem 2. The PJK_i test is consistent for testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ against $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$.

Proof. Let us consider the standardized $T = PJK_i$ test statistic

$$ST^*(\boldsymbol{\mu}) = \frac{T^*(\boldsymbol{\mu}) - E(T^*(\mathbf{0}))}{\sqrt{VAR(T^*(\mathbf{0}))}}$$

where $E(T^*(\mathbf{0})) = \frac{n(N+1)}{2}$ and $VAR(T^*(\mathbf{0})) = \frac{mn(N+1)}{12}$ denote, respectively, the mean and variance of the T^* statistic under the null hypothesis. To prove consistency, we have to show that

$$\Pr(ST^*(\boldsymbol{\mu}) \geq ST_{1-\alpha}(n)|\mathbf{Z}(\boldsymbol{\mu})) \to 1$$

as $m, n \to \infty$ with $\frac{m}{N} \to \lambda < 1$, where $ST_{1-\alpha}(n)$ is the critical value of the standardized test. Note that $ST_{1-\alpha}(n)$ is data-dependent and is calculated as the $(1-\alpha)100$th percentile of the permutation distribution of the ST^* statistic since the test is significant for large values of the test statistic. Note that $ST^*(\boldsymbol{\mu})$ is asymptotically distribution as a standard normal. If $m, n \to \infty$, then $ST_{1-\alpha}(n) \to z_{1-\alpha} < \infty$, where $z_{1-\alpha}$ is the $(1-\alpha)100$th percentile of the standard normal distribution. Now, as sample sizes increase

$$\Pr(ST^*(\boldsymbol{\mu}) \geq ST_{1-\alpha}(n)|\mathbf{Z}(\boldsymbol{\mu})) = \Pr\left(T^*(\boldsymbol{\mu}) \geq ST_{1-\alpha}(n)\sqrt{VAR(T^*(\mathbf{0}))} + E(T^*(\mathbf{0}))|\mathbf{Z}(\boldsymbol{\mu})\right)$$

approaches

$$1 - \Phi\left(\frac{z_{1-\alpha}\sqrt{VAR(T^*(\mathbf{0}))} + E(T^*(\mathbf{0})) - E(T^*(\boldsymbol{\mu}))}{\sqrt{VAR(T^*(\boldsymbol{\mu}))}}\right),$$

where Φ denotes the distribution function of the standard normal distribution,

$$E(T^*(\boldsymbol{\mu})) = E(T^*(\mathbf{0})) + \gamma mn$$

and

$$VAR(T^*(\boldsymbol{\mu})) = VAR(T^*(\mathbf{0})) + mn(-\gamma^2(N+1) + (\gamma-\delta)(m-1) + (\gamma-\eta)(n-1))$$

denote, respectively, the mean and variance of the T^* statistic under the alternative hypothesis, where $0 < \gamma < \frac{1}{2}, 0 < \delta < \frac{1}{3}, 0 < \eta < \frac{1}{3}$. Simple algebra shows that

$$\frac{\sqrt{VAR(T^*(\mathbf{0}))}}{\sqrt{VAR(T^*(\boldsymbol{\mu}))}} = \left(\sqrt{1 + 12\left(-\gamma^2\frac{N-1}{N+1} + (\gamma-\delta)\frac{m-1}{N+1} + (\gamma-\eta)\frac{n-1}{N+1}\right)}\right)^{-1}$$

which converges as $m, n \to \infty$ with $\frac{m}{N} \to \lambda < 1$ and that

$$\frac{E(T^*(\mathbf{0})) - E(T^*(\boldsymbol{\mu}))}{\sqrt{VAR(T^*(\boldsymbol{\mu}))}} = -\left(\sqrt{\frac{1}{12\gamma^2}\left(\frac{1}{n}+\frac{1}{m}+\frac{1}{mn}\right) - \left(\frac{1}{n}+\frac{1}{m}-\frac{1}{mn}\right) + \frac{\gamma-\delta}{\gamma^2}\left(\frac{1}{n}+\frac{1}{mn}\right) + \frac{\gamma-\eta}{\gamma^2}\left(\frac{1}{m}+\frac{1}{mn}\right)}\right)^{-1}$$

which diverges to $-\infty$. Therefore, $\Pr(ST^*(\boldsymbol{\mu}) \geq ST_{1-\alpha}(n)|\mathbf{Z}(\boldsymbol{\mu})) \to 1$ and the test is consistent. ■