Statistical Science
2004, Vol. 19, No. 4, 598–605
DOI 10.1214/088342304000000558

Multivariate Nonparametric Tests

Hannu Oja and Ronald H. Randles

Abstract. Multivariate nonparametric statistical tests of hypotheses are described for the one-sample location problem, the several-sample location problem and the problem of testing independence between pairs of vectors. These methods are based on affine-invariant spatial sign and spatial rank vectors. They provide affine-invariant multivariate generalizations of the univariate sign test, signed-rank test, Wilcoxon rank sum test, Kruskal–Wallis test, and the Kendall and Spearman correlation tests. While the emphasis is on tests of hypotheses, certain references to associated affine-equivariant estimators are included. Pitman asymptotic efficiencies demonstrate the excellent performance of these methods, particularly in heavy-tailed population settings. Moreover, these methods are easy to compute for data in common dimensions.

Key words and phrases: Affine invariance, spatial rank, spatial sign, Pitman efficiency, robustness.

1. INTRODUCTION

Modern data collection settings often involve collecting information on multiple attributes of each object (person, animal) in the study. In health studies, for example, each observation on a patient is actually a whole array of measurements which together describe the health status of the person at a particular point in time. Thus we are naturally led to consider vector-valued observations in dealing with data from these settings. There are special needs and concerns when dealing with multivariate data. If each component of the vectors is only studied marginally, then certain outliers, strongly influential points and useful relationships among variables may not be detected. Thus a multivariate examination of the data is very appropriate and important. Describe each observation as a vector $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ of dimension p. The components $x_{i1}, \dots, x_{ip}$ usually (but not always) represent different types of measurements made on one experimental unit. In our discussions, we consider each component to be continuous (or at least fairly continuous) in nature. This paper examines a number of hypothesis testing problem settings for multivariate data.

Hannu Oja is Professor, Department of Mathematics and Statistics, University of Jyväskylä, FIN-40351 Jyväskylä, Finland (e-mail: ojahannu@cc.jyu.fi). Ronald H. Randles is Professor, Department of Statistics, University of Florida, Gainesville, Florida 32611-8545, USA (e-mail: rrandles@stat.ufl.edu).

2. ONE-SAMPLE LOCATION PROBLEM

2.1 Hotelling's T^2 Test

Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be i.i.d. from $F(\mathbf{x} - \boldsymbol{\theta})$, where $F(\cdot)$ represents a continuous p-dimensional distribution "located" at the vector parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$. We wish to test the hypotheses

$$H_0 : \boldsymbol{\theta} = \mathbf{0} \quad \text{vs.} \quad H_a : \boldsymbol{\theta} \neq \mathbf{0}.$$

Note that the zero vector, $\mathbf{0}$, is used without loss of generality, because to test $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs. $H_a : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, we substitute $\mathbf{x}_i - \boldsymbol{\theta}_0$ in place of $\mathbf{x}_i$ in the tests described below.

The classical parametric test, *Hotelling's* T^2, rejects H_0 if

$$T^2 = n\bar{\mathbf{X}}^T S^{-1} \bar{\mathbf{X}} \geq \frac{np}{n-p} F_{p,n-p}(\alpha),$$

where $\bar{\mathbf{X}} = \operatorname{ave}\{\mathbf{X}_i\}$ and $S = \operatorname{ave}\{(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T\}$ are the sample mean vector and sample covariance matrix, respectively, and $F_{\nu_1,\nu_2}(\alpha)$ is the upper αth quantile of an F distribution with ν_1 and ν_2 degrees of freedom. Notation "ave" means the average taken over all observations $i = 1, \dots, n$. This test assumes that the underlying population is multivariate normal with mean vector $\boldsymbol{\theta}$ and variance–covariance matrix Σ. Hotelling's T^2 test is also *asymptotically nonparametric* in the sense that if the random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is from any p-variate population with mean vector $\mathbf{0}$ and finite second moments, then

$$T^2 \xrightarrow{d} \chi_p^2$$

and therefore the quantiles of the chi-squared distribution give large sample cutoff values in nonnormal cases.

Let A_x be any nonsingular $p \times p$ matrix such that $A_x^T A_x = S^{-1}$. The matrix A_x may be an upper triangular matrix obtained from a Choleski factorization of S^{-1} or the symmetric square root matrix $A_x = S^{-1/2}$, for example. Then

$$T^2 = n\bar{\mathbf{Y}}^T\bar{\mathbf{Y}} = n\|\bar{\mathbf{Y}}\|^2,$$

where $\mathbf{Y}_i = A_x\mathbf{X}_i$, $i = 1, \ldots, n$. Thus T^2 is *n times the squared length of the average* (*mean vector*) *of the transformed data points*.

The transformation A_x makes the transformed points appear to have come from a population with variance–covariance matrix $\Sigma = I$, because the matrix S computed on the $\mathbf{Y}_i$'s is I. However, the fundamental purpose of the transformation A_x is to give the test statistic the following property: A test statistic $T(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ for testing $H_0 : \boldsymbol{\theta} = \mathbf{0}$ is said to be *affine invariant* if

$$T(D\mathbf{x}_1, \ldots, D\mathbf{x}_n) = T(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

for every $p \times p$ nonsingular matrix D and for every p-variate data set $\mathbf{x}_1, \ldots, \mathbf{x}_n$. In the current problem T^2 is affine invariant. This property ensures that its performance is consistent over all possible choices of the coordinate system.

2.2 Multivariate Sign Test

In one dimension, the sign of an observation is basically its direction ($+1$ or -1) from the origin. In higher dimensions, in this spirit, the *spatial sign function* is defined as

$$\mathbf{S}(\mathbf{x}) = \begin{cases} \|\mathbf{x}\|^{-1}\mathbf{x}, & \mathbf{x} \neq \mathbf{0}, \\ \mathbf{0}, & \mathbf{x} = \mathbf{0}, \end{cases}$$

where $\|\mathbf{x}\|$ is the L_2 norm (Euclidean distance of $\mathbf{x}$ from $\mathbf{0}$). The function value is thus just a direction (a point on the unit p sphere) whenever $\mathbf{x} \neq \mathbf{0}$.

To create an affine-invariant sign test, we apply the spatial sign function to transformed data points. Define the *spatial signs* to be

$$\mathbf{S}_i = \mathbf{S}(A_x\mathbf{X}_i) \quad \text{for } i = 1, \ldots, n, \tag{1}$$

where A_x is now the data driven transformation proposed by Tyler (1987). *Tyler's shape matrix* V_x is the positive definite symmetric $p \times p$ matrix with $\text{trace}(V_x) = p$ such that, for any A_x with $A_x^T A_x = V_x^{-1}$,

$$p\,\text{ave}\{\mathbf{S}_i\mathbf{S}_i^T\} = I_p.$$

The matrix A_x is then called *Tyler's transformation*. It is remarkable that Tyler's transformation A_x as well as the spatial signs $\mathbf{S}_i$, $i = 1, \ldots, n$, then depend on the data cloud only through directions $\|\mathbf{X}_i\|^{-1}\mathbf{X}_i$, $i = 1, \ldots, n$.

Tyler's transformation A_x is thus the transformation that makes the *sign covariance matrix* equal to $[1/p]I_p$, the variance–covariance matrix of a vector that is uniformly distributed on the unit p sphere. Since $\mathbf{S}_i$ and $-\mathbf{S}_i$ contribute identically to the sample covariance matrix, the Tyler transformation may be viewed as an attempt to make the signs (directions) of the transformed data points $\pm A_x\mathbf{X}_i$, $i = 1, \ldots, n$, appear as though they are uniformly distributed on the unit p sphere.

Matrix functions in modern computer programming languages have made Tyler's shape matrix and Tyler's transformation surprisingly easy to compute. Its iterative construction may begin with $V = I_p$ and use an iteration step that transforms from one V to the next via

$$V \leftarrow pV^{1/2}\,\text{ave}\{\mathbf{S}_i\mathbf{S}_i^T\}V^{1/2}.$$

When $\|p\,\text{ave}\{\mathbf{S}_i\mathbf{S}_i^T\} - I_p\|$ is sufficiently small, stop and set $V_x = [p/\text{trace}(V)]V$. Choose A_x so that $A_x^T A_x = V_x^{-1}$. Here, the matrix norm $\|A\| = \sqrt{\text{trace}(A^T A)}$.

Having found the spatial signs described in (1), the multivariate sign test then rejects H_0 in favor of H_a for large values of

$$Q^2 = np\bar{\mathbf{S}}^T\bar{\mathbf{S}} = np\|\bar{\mathbf{S}}\|^2, \tag{2}$$

which is simply np times the *squared length of the average direction vector of the transformed data points*. This test was developed by Randles (2000).

Appropriate cutoff values for conducting this test depend on the assumptions made about the underlying distribution $F(\mathbf{x} - \boldsymbol{\theta})$. The underlying distribution is said to be *elliptically symmetric* if its density takes the form

$$f(\mathbf{x} - \boldsymbol{\theta}) = |\Sigma|^{-1/2} g\big((\mathbf{x} - \boldsymbol{\theta})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\theta})\big)$$

with symmetry center $\boldsymbol{\theta}$ and positive definite symmetric $p \times p$ scatter matrix Σ. The contours of these densities form concentric ellipses centered at $\boldsymbol{\theta}$. The multivariate normal and multivariate t distributions, for example, are both members of this broad class. The test statistic Q^2 is *strictly distribution-free* over the class of elliptically symmetric distributions (and a somewhat larger class). Thus α-level cutoffs $Q \geq q_{n,p}(\alpha)$ could be established based on the elliptical distribution class.

Potentially weaker assumptions about $F(\cdot)$ include symmetry (under which $\mathbf{X} - \boldsymbol{\theta}$ has the same distribution as $\boldsymbol{\theta} - \mathbf{X}$) or directional symmetry [under which $(\mathbf{X} - \boldsymbol{\theta})/\|\mathbf{X} - \boldsymbol{\theta}\|$ has the same distribution as $(\boldsymbol{\theta} - \mathbf{X})/\|\boldsymbol{\theta} - \mathbf{X}\|$]. Since symmetry implies directional symmetry, the latter is a weaker assumption about $F(\cdot)$. Under the assumption of directional symmetry, a *conditional distribution-free* p *value* is found via

$$E_{\boldsymbol{\delta}}[I\{Q_{\boldsymbol{\delta}}^2 \geq Q^2\}],$$

where $\boldsymbol{\delta}$ is uniformly distributed over the 2^n p-dimensional vectors with each component a $+1$ or -1 and $Q_{\boldsymbol{\delta}}^2$ is the value of the test statistic for the data set $\delta_1\mathbf{X}_1, \ldots, \delta_n\mathbf{X}_n$. Since A_x does not depend on the signs of the $\mathbf{X}_i$'s it is sufficent to replace each $\mathbf{S}_i$ with $\delta_i\mathbf{S}_i$ in the computation of $Q_{\boldsymbol{\delta}}^2$.

Finally note that, for large sample sizes, a cutoff can be obtained by using the fact that when the underlying distribution is directionally symmetric and H_0 holds, then

$$Q^2 \xrightarrow{d} \chi_p^2.$$

A multivariate median estimating a directional center of the population and corresponding to the sign test based on Q^2 in the Hodges–Lehmann sense was developed by Hettmansperger and Randles (2002). This median is called the *transformation–retransformation spatial median*. The tranformation–retransformation technique was described by Chakraborty, Chaudhuri and Oja (1998), for example.

2.3 Multivariate Rank Methods

Multivariate ranks are constructed using the signs of transformed differences

$$\mathbf{S}_{ij} = \mathbf{S}\big(A_x(\mathbf{X}_i - \mathbf{X}_j)\big), \quad i, j = 1, \ldots, n,$$

again with a data based transformation A_x. This leads to the concept of a *centered rank*

$$\mathbf{R}_i = \text{ave}_j\{\mathbf{S}_{ij}\}$$

with the property $\text{ave}\{\mathbf{R}_i\} = \mathbf{0}$. To see that this is an extension of the univariate centered rank, consider univariate data. With univariate data, A_x can be taken to be a positive scalar and hence $S_{ij} = S(A_x(X_i - X_j)) = \text{sign}(X_i - X_j)$, that is, A_x plays no role. If no ties exist,

$$R_i = \frac{2}{n}\left[\text{Rank}(X_i) - \frac{n+1}{2}\right],$$

where $\text{Rank}(X_i)$ denotes the usual univariate rank of X_i among $X_1, \ldots, X_n$, ranking from smallest to largest. Since $(n+1)/2$ is the mean of $\text{Rank}(X_i)$, we see that R_i is $2/n$ times the regular rank centered at its mean.

In multivariate settings, the data based transformation A_x is chosen to make the rank procedures affine invariant. A natural choice of A_x is the transformation needed so that the ranks satisfy the property

$$p\,\text{ave}\{\mathbf{R}_i\mathbf{R}_i^T\} = \text{ave}\{\mathbf{R}_i^T\mathbf{R}_i\}I_p.$$

This transformation then makes the *rank covariance matrix* equal to a scalar times the identity matrix, that is, $\text{ave}\{\mathbf{R}_i\mathbf{R}_i^T\} = [c_x^2/p]I_p$, where $c_x^2 = \text{ave}\{\|\mathbf{R}_i\|^2\}$. The ranks of the transformed points thus behave as though they are spherically distributed in the unit p sphere. The iterative construction is as in the case of Tyler's shape matrix: One can again start with $V = I_p$ and use an iteration step

$$V \leftarrow \frac{p}{\text{ave}\{\mathbf{R}_i^T\mathbf{R}_i\}} V^{1/2}\,\text{ave}\{\mathbf{R}_i\mathbf{R}_i^T\}V^{1/2},$$

where the $\mathbf{R}_i$ are calculated from the $V^{-1/2}\mathbf{X}_i$. In the end, $V_x = [p/\text{trace}(V)]V$ and the transformation A_x is given by $A_x^TA_x = V_x^{-1}$. Unfortunately, there is no proof of the convergence of the algorithm so far, but in practice it seems always to converge.

The centered ranks are clearly invariant under location shifts and $\text{ave}\{\mathbf{R}_i\} = \mathbf{0}$. The ranks $\mathbf{R}_i$ lie in the unit p sphere; the direction of $\mathbf{R}_i$ roughly points outward from the center (spatial median) of the data cloud and its length (in a sense) tells how far away this point is from the center.

With univariate data, the Wilcoxon signed-rank test statistic is essentially the sign test statistic applied to the Walsh sums (or averages) $x_i + x_j$ for $i \leq j$. Likewise, a *multivariate one-sample signed-rank test statistic* can be constructed using the signs of transformed Walsh sums (or averages), that is,

$$U^2 = \frac{np}{4c_x^2}\big\|\text{ave}\big\{\mathbf{S}\big(A_x(\mathbf{X}_i + \mathbf{X}_j)\big)\big\}\big\|^2, \tag{3}$$

where the average is over $i, j = 1, \ldots, n$. Here the transformation A_x is chosen to be the rank transformation and c_x^2 is the scalar described above. If, for example, $\mathbf{X}_1, \dots, \mathbf{X}_n$ is a random sample from an elliptically symmetric distribution with symmetry center $\boldsymbol{\theta} = \mathbf{0}$, then again

$$U^2 \xrightarrow{d} \chi_p^2$$

and approximate cutoffs can be obtained as quantiles of the chi-squared distribution.

The multivariate one-sample affine equivariant *Hodges–Lehmann estimate* is obtained as the transformed–retransformed (spatial) median of pairwise averages, that is, the value of $\boldsymbol{\theta}$ which would make $U^2 = 0$ when U^2 is computed replacing each $\mathbf{X}_i$ with $\mathbf{X}_i - \boldsymbol{\theta}$ for $i = 1, \dots, n$. For the noninvariant versions of the spatial tests and related estimators which do not utilize the auxiliary transformation A_x, see Möttönen and Oja (1995).

2.4 Efficiencies

The Pitman asymptotic efficiencies of the multivariate sign test and multivariate signed-rank test relative to Hotelling's T^2 when the underlying population is multivariate t were derived by Möttönen, Oja and Tienari (1997). Some efficiencies are displayed in Table 1. We see that as the dimension p increases and as the distribution gets heavier tailed (df gets smaller), the performance of Q^2 and U^2 improves relative to T^2. The sign test and the signed-rank test are clearly better than T^2 in heavy-tailed cases. For high dimensions and very heavy tails, the sign test is the more efficient test. Note that $df = +\infty$ is the multivariate normal. The efficiencies in this table also represent ratios of the asymptotic variances of the transformation–retransformation spatial median to the sample mean vector (sign test columns) and the Hodges–Lehmann estimator to the sample mean vector (signed-rank test columns).

TABLE 1

Asymptotic efficiencies of the multivariate sign test and the signed-rank test relative to Hotelling's T^2 under p-variate t distributions with ν degrees of freedom for selected values of p and ν

Dimension	**Sign test**			**Signed-rank test**		
p	$\nu = 3$	$\nu = 6$	$\nu = \infty$	$\nu = 3$	$\nu = 6$	$\nu = \infty$
1	1.62	0.88	0.64	1.90	1.16	0.95
2	2.00	1.08	0.78	1.95	1.19	0.97
4	2.25	1.22	0.88	2.02	1.21	0.98
10	2.42	1.31	0.95	2.09	1.22	0.99

2.5 Example

Merchant et al. (1975) studied changes in pulmonary function of 12 workers after 6 hours of cotton dust exposure. We examine the three-dimensional data produced by differences in forced vital capacity, forced expiratory volume and closing capacity. The concern in this problem is whether there is indication of pulmonary change. Thus we seek to test whether the three-dimensional population is located at $\mathbf{0}$ or not. Analyzing their data yields T^2 = 8.5265 with a p value = 0.166 (F), $Q^2 = 5.8345$ with a p value $= 0.120$ (χ_3^2) and $U^2 = 4.8169$ with a p value $= 0.186$ (χ_3^2).

3. SEVERAL-SAMPLES LOCATION PROBLEM

3.1 Classical Multivariate Analysis of Variance

Let

$$\mathbf{X}_1, \dots, \mathbf{X}_{N_1}; \mathbf{X}_{N_1+1}, \dots, \mathbf{X}_{N_2}; \cdots; \mathbf{X}_{N_{c-1}+1}, \dots, \mathbf{X}_{N_c}$$

be c independent random samples with sample sizes $n_1, \dots, n_c$, from p-variate distributions $F(\mathbf{x} - \boldsymbol{\theta}_1)$, $F(\mathbf{x} - \boldsymbol{\theta}_2), \dots, F(\mathbf{x} - \boldsymbol{\theta}_c)$ located at p-variate centers $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_c$, respectively. Here $N_i = n_1 + \cdots + n_i$ and $N_c = N$. Write also $N_0 = 0$. We wish to test the null hypothesis of no treatment difference, that is,

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \cdots = \boldsymbol{\theta}_c \quad \text{vs.} \quad H_a : \boldsymbol{\theta}_i\text{'s not all equal.}$$

Note that under H_0, $\mathbf{X}_1, \dots, \mathbf{X}_N$ is a random sample from a common multivariate distribution. The classical multivariate analysis of variance (MANOVA) test statistic, *Hotelling's trace statistic*, is constructed as follows. First calculate the global mean vector $\bar{\mathbf{X}}$ and the within samples covariance matrix S. Then Hotelling's trace statistic is

$$T^2 = \sum_{i=1}^{c} n_i \|\bar{\mathbf{Y}}_i\|^2,$$

where

$$\bar{\mathbf{Y}}_i = \frac{1}{n_i} \sum_{j=N_{i-1}+1}^{N_i} \mathbf{Y}_j, \quad i = 1, \dots, c,$$

are the samplewise mean vectors of the transformed data points $\mathbf{Y}_i = A_x(\mathbf{X}_i - \bar{\mathbf{X}})$, with transformation A_x satisfying $A_x^T A_x = S^{-1}$. The T^2 test statistic is a *weighted sum of squared lengths of transformed distances of the sample averages from the grand average*. It thus measures the variability among the locations of the samples. If second moments exist, then under the null hypothesis,

$$T^2 \xrightarrow{d} \chi^2_{p(c-1)}.$$

Note that this is true also if the within covariance matrix is replaced by the regular combined sample covariance matrix.

Clearly the MANOVA statistic has the following desired affine invariance property: A test statistic $T(\mathbf{x}_1,\dots,\mathbf{x}_{N_1};\cdots;\mathbf{x}_{N_{c-1}+1},\dots,\mathbf{x}_{N_c})$ for testing H_0: $\boldsymbol{\theta}_1=\cdots=\boldsymbol{\theta}_c$ is said to be *affine invariant* if

$$
\begin{aligned}
&T(D\mathbf{x}_1+\mathbf{d},\dots,D\mathbf{x}_{N_1}+\mathbf{d};\cdots;\\
&\quad D\mathbf{x}_{N_{c-1}+1}+\mathbf{d},\dots,D\mathbf{x}_{N_c}+\mathbf{d})\\
&\quad\quad = T(\mathbf{x}_1,\dots,\mathbf{x}_{N_1};\cdots;\mathbf{x}_{N_{c-1}+1},\dots,\mathbf{x}_{N_c})
\end{aligned}
$$

for every $\mathbf{x}_1,\dots,\mathbf{x}_N$, $\mathbf{d}$ a $(p\times 1)$ vector and $\mathbf{D}$ a $(p\times p)$ nonsingular matrix. This affine invariance property ensures that the testing procedure is independent of the choice of the coordinate system and behaves consistently under different covariance structures. This property is attained because of the transformation A_x.

3.2 Several-Samples Rank Test

In the several-samples location problem, again consider the combined sample $\mathbf{X}_1,\dots,\mathbf{X}_N$. Form the signs of transformed differences

$$\mathbf{S}_{ij}=\mathbf{S}\big(A_x(\mathbf{X}_i-\mathbf{X}_j)\big),\quad i,j=1,\dots,N,$$

which lead to spatial centered ranks of each observation within the combined sample:

$$\mathbf{R}_i=\text{ave}_j\{\mathbf{S}_{ij}\},\quad i=1,\dots,N.$$

The data based transformation A_x is chosen to make the rank test affine invariant. It is determined by requiring, as before, that the ranks satisfy the property $p\,\text{ave}\{\mathbf{R}_i\mathbf{R}_i^T\}=\text{ave}\{\mathbf{R}_i^T\mathbf{R}_i\}I_p$. The scalar $c_x^2=\text{ave}\{\mathbf{R}_i^T\mathbf{R}_i\}$ depends on the data cloud.

Multivariate extensions of the two-sample Wilcoxon–Mann–Whitney test and the several-sample Kruskal–Wallis test are then obtained as follows. The *several-samples spatial rank test statistic* is

$$U^2=\frac{p}{c_x^2}\sum_{i=1}^{c}n_i\|\bar{\mathbf{R}}_i\|^2, \tag{4}$$

where $\bar{\mathbf{R}}_i$ for $i=1,\dots,c$ are samplewise mean vectors of the spatial centered ranks as defined above. The conditions under which the limiting null distribution of U^2 is the chi-squared distribution with $p(c-1)$ degrees of freedom are still to be settled. (The statistical properties of A_x are unknown.) Under these mild assumptions, the test statistic is thus *asymptotically distribution-free.*

The p value of a *conditionally distribution-free permutation test* based on U^2 is obtained via

$$E_{\boldsymbol{\gamma}}[I\{U_{\boldsymbol{\gamma}}^2\ge U^2\}],$$

where $\boldsymbol{\gamma}=(\gamma_1,\dots,\gamma_N)$ is uniformly distributed over the $N!$ permutations of $(1,\dots,N)$ and $U_{\boldsymbol{\gamma}}^2$ is the value of the test statistic for the permuted sample $\mathbf{X}_{\gamma_1},\dots,\mathbf{X}_{\gamma_N}$. Note that the A_x used to define the $\mathbf{R}_i$'s is invariant under permutations, so it is sufficient just to replace each spatial rank $\mathbf{R}_i$ with $\mathbf{R}_{\gamma_i}$ for $i=1,\dots,N$ when computing $U_{\boldsymbol{\gamma}}^2$ using (4).

The several-samples multivariate sign tests—extensions of the univariate Mood test—could be defined as well. See Möttönen and Oja (1995) for noninvariant versions. The Pitman asymptotic relative efficiencies (ARE) of the several-samples multivariate spatial rank test relative to the classical MANOVA T^2 statistic are the same as the efficiencies of the multivariate signed-rank test relative to Hotelling's T^2; see Table 1.

3.3 An Example

Applying the methods of this section to the male Egyptian skull data found in Hand et al. (1994, page 299), we find that for these five samples of 30 observations in dimension 4, $T^2=52.643$ and $U^2=61.189$, which both yield tiny p values when compared to a chi-squared $(df=16)$ distribution.

4. TESTING FOR INDEPENDENCE

4.1 The Problem and Classical Test

It is often of interest to explore potential relationships among subsets of multiple measurements. Some measurements may represent attributes of psychological characteristics, while others represent attributes of physical characteristics. It may be of interest to determine whether there is a relationship between the psychological and the physical characteristics. This requires a test of independence between pairs of vectors, where the vectors potentially have different measurement scales and dimensions. Accordingly, let $\mathbf{X}_i^T=(\mathbf{X}_i^{(1)^T},\mathbf{X}_i^{(2)^T})$ for $i=1,\dots,n$ denote a random sample of vector pairs, where $\mathbf{X}_i^{(1)}$ and $\mathbf{X}_i^{(2)}$ are continuous vectors of dimensions p and q, respectively. We seek to test

$$H_0:\mathbf{X}_i^{(1)}\text{ and }\mathbf{X}_i^{(2)}\text{ are independent}\quad\text{vs.}$$

$$H_a:\text{they are dependent.}$$

In the multinormal case, Wilks (1935) derived the likelihood ratio criterion for detecting deviations from the hypothesis of independence. The Wilks test statistic can be expressed as

$$V^{n/2} = \det\big(\text{ave}\{\mathbf{Y}_i^{(1)}\mathbf{Y}_i^{(2)^T}\}\big),$$

where (as before) $\mathbf{Y}_i^{(v)} = A_x^{(v)}(\mathbf{X}_i^{(v)} - \bar{\mathbf{X}}^{(v)})$, $v = 1, 2$ and $i = 1, \dots, n$, with partitioned sample mean vectors $\bar{\mathbf{X}}^{(v)}$, sample covariance matrices $S^{(v)}$ and transformations $A_x^{(v)}$ such that $A_x^{(v)^T} A_x^{(v)} = (S^{(v)})^{-1}$ for $v = 1, 2$. An asymptotically equivalent test can be based on the statistic

$$W = npq \,\|\, \text{ave}\{\mathbf{Y}_i^{(1)}\mathbf{Y}_i^{(2)^T}\}\|^2.$$

The statistic W is seen to be npq times the sum of squares of covariances between elements of the transformed $\mathbf{X}_i^{(1)}$ with elements of the transformed $\mathbf{X}_i^{(2)}$ vectors. Under H_0, the limiting distribution of W is a chi-squared distribution with pq degrees of freedom.

Muirhead (1982) examined the effect of the group of transformations $\{\mathbf{x} \to D\mathbf{x} + \mathbf{d}\}$ on this problem. Here $\mathbf{d}$ is any $p + q$ vector and

$$D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$$

is any $(p+q) \times (p+q)$ nonsingular matrix of the form above with $p \times p$ matrix D_1 and $q \times q$ matrix D_2. The Wilks test is invariant under this group of transformations. Thus its value does not depend on the chosen marginal coordinate systems and its performance is consistent under different variance–covariance structures of either $\mathbf{X}_i^{(1)}$ or $\mathbf{X}_i^{(2)}$. This characteristic generally improves its power and control of α levels.

4.2 Rank Tests of Independence

To motivate multivariate nonparametric tests of independence, recall first the popular univariate ($p = q = 1$) nonparametric tests due to Kendall (1938) and Spearman (1904). Kendall's tau is a scalar multiple of

$$\text{ave}\{\text{sign}(X_i^{(1)} - X_j^{(1)})\,\text{sign}(X_i^{(2)} - X_j^{(2)})\}$$

or

$$\text{ave}\{S_{ij}^{(1)} S_{ij}^{(2)}\}$$

and Spearman's rho is a scalar multiple of

$$\text{ave}\left\{\left(\text{Rank}(X_i^{(1)}) - \frac{n+1}{2}\right)\left(\text{Rank}(X_i^{(2)}) - \frac{n+1}{2}\right)\right\}$$

or

$$\text{ave}\{R_i^{(1)} R_i^{(2)}\},$$

where $\text{Rank}(X_i^{(v)})$ is the usual univariate rank of $X_i^{(v)}$ among $X_1^{(v)}, \dots, X_n^{(v)}$ for $v = 1, 2$. Kendall's tau and Spearman's rho are correlations between signs of the pairwise differences and centered ranks, respectively.

A multivariate extension of Kendall's tau is created by forming sign vectors $\mathbf{S}_{ij}^{(1)} = \mathbf{S}(A_x^{(1)}(\mathbf{X}_i^{(1)} - \mathbf{X}_j^{(1)}))$, where the transformation $A_x^{(1)}$ is chosen so that $p\,\text{ave}\{\mathbf{S}_{ij}^{(1)}\mathbf{S}_{ij}^{(1)^T}\} = I_p$. This is the transformation studied by Tyler (1987) but computed on differences $\mathbf{X}_i^{(1)} - \mathbf{X}_j^{(1)}$. The corresponding shape matrix V_x for which $A_x^T A_x = V_x^{-1}$ was introduced by Dümbgen (1998). Note that the $\mathbf{S}_{ij}^{(1)}$'s are invariant under location shifts. Similarly, q-dimensional sign vectors $\mathbf{S}_{ij}^{(2)}$ are formed based on differences among $A_x^{(2)}\mathbf{X}_1^{(2)}, \dots, A_x^{(2)}\mathbf{X}_n^{(2)}$ with a similar transformation $A_x^{(2)}$.

A multivariate version of the test based on Kendall's tau uses

$$\tau^2 = \frac{npq}{(2c_x^{(1)}c_x^{(2)})^2}\,\|\, \text{ave}\{\mathbf{S}_{ij}^{(1)}\mathbf{S}_{ij}^{(2)^T}\}\|^2$$

with data dependent constants $c_x^{(1)}$ and $c_x^{(2)}$ described below. Here the scalar multiple is chosen so that when the marginal distributions of $X_i^{(v)}$ are elliptically symmetric, $v = 1, 2$, and when H_0 is true, the limiting distribution of τ^2 is a chi-squared distribution with pq degrees of freedom.

The multivariate extension of Spearman's rho uses centered rank vectors $\mathbf{R}_i^{(1)}$ based on differences among the first components $A_x^{(1)}\mathbf{x}_1^{(1)}, \dots, A_x^{(1)}\mathbf{x}_n^{(1)}$ transformed by $A_x^{(1)}$ chosen so that

$$p\,\text{ave}\{\mathbf{R}_i^{(1)}\mathbf{R}_i^{(1)^T}\} = \text{ave}\{\mathbf{R}_i^{(1)^T}\mathbf{R}_i^{(1)}\}I_p.$$

With analogous descriptions of the q-dimensional rank vectors $\mathbf{R}_i^{(2)}$, a multivariate version of the test based on Spearman's rho uses

$$\rho^2 = \frac{npq}{(c_x^{(1)}c_x^{(2)})^2}\,\|\, \text{ave}\{\mathbf{R}_i^{(1)}\mathbf{R}_i^{(2)^T}\}\|^2$$

with

$$(c_x^{(1)})^2 = \text{ave}\{\mathbf{R}_i^{(1)^T}\mathbf{R}_i^{(1)}\}$$

and

$$(c_x^{(2)})^2 = \text{ave}\{\mathbf{R}_i^{(2)^T}\mathbf{R}_i^{(2)}\}.$$

Again the scalar multiple is chosen so that the limiting null distribution of ρ^2 is a chi-squared distribution with

pq degrees of freedom, under the same conditions described for τ^2. The statistics τ^2 and ρ^2 were proposed by Taskinen, Oja and Randles (2005).

The statistic τ^2 (ρ^2) is seen to be a scalar multiple of the sum of squares of covariances between elements of the sign-transformed differences $\mathbf{X}_i^{(1)} - \mathbf{X}_j^{(1)}$ (rank-transformed $\mathbf{X}_i^{(1)}$) and elements of the corresponding sign-transformed differences $\mathbf{X}_i^{(2)} - \mathbf{X}_j^{(2)}$ (rank-transformed $\mathbf{X}_i^{(2)}$). The statistics τ^2 and ρ^2 are easy to compute for data in common dimensions. This property makes them very practical. For small n, conditional p values can be generated via

$$E_{\boldsymbol{\gamma}}[I\{\tau_{\boldsymbol{\gamma}}^2 \geq \tau^2\}] \quad \text{and} \quad E_{\boldsymbol{\gamma}}[I\{\rho_{\boldsymbol{\gamma}}^2 \geq \rho^2\}],$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$ is uniformly distributed over the $n!$ permutations of the integers $(1, 2, \ldots, n)$ and

$$\tau_{\boldsymbol{\gamma}}^2 = \frac{npq}{(2c_x^{(1)}c_x^{(2)})^2}\|\operatorname{ave}\{\mathbf{S}_{ij}^{(1)}\mathbf{S}_{\gamma_i\gamma_j}^{(2)^T}\}\|^2,$$

$$\rho_{\boldsymbol{\gamma}}^2 = \frac{npq}{(c_x^{(1)}c_x^{(2)})^2}\|\operatorname{ave}\{\mathbf{R}_i^{(1)}\mathbf{R}_{\gamma_i}^{(2)^T}\}\|^2.$$

A multivariate analogue to the univariate Blomqvist (1950) quadrant test was developed by Taskinen, Kankainen and Oja (2003).

4.3 Efficiencies and an Example

Using the model

$$\begin{pmatrix}\mathbf{X}^{(1)}\\ \mathbf{X}^{(2)}\end{pmatrix} = \begin{pmatrix}(1-\Delta)I_p & \Delta M_1\\ \Delta M_1^T & (1-\Delta)I_q\end{pmatrix}\begin{pmatrix}\mathbf{Z}^{(1)}\\ \mathbf{Z}_{(2)}\end{pmatrix},$$

where $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ are independent, Pitman AREs were developed by Taskinen, Oja and Randles (2005). Here M_1 denotes an arbitrary $p \times q$ matrix with $\|M_1\|^2 > 0$. Some AREs for contaminated normal $\mathbf{Z}^{(v)}$ are shown in Table 2, where $p = q$ and $\mathbf{Z}^{(v)} \sim (0.9)N(\mathbf{0}, I) + (0.1)N(0, cI)$. The efficiencies are relative to the classical parametric test with test statistic W. Again we observe the superiority of the spatial sign and rank based methods, particularly in higher dimensions and with heavier tailed populations.

Applying these methods to the head length and head breadth measurements on both first and second born sons in 25 families (see Hand et al., 1994, page 85), we test whether there is correlation among these paired bivariate measurements. The tests yield $W = 75.872$, $\tau^2 = 66.678$ and $\rho^2 = 26.914$. The p values are very small for all three tests based on comparison to a chi-squared ($df = 4$) distribution.

TABLE 2

Asymptotic efficiencies of the multivariate analogues to Spearman's rho and Kendall's tau tests at different contaminated normal distributions for $\varepsilon = 0.1$ and for selected values of c and selected dimensions $p = q$

Dimension	**Kendall and Spearman**		
$p = q$	$c = 1$	$c = 3$	$c = 6$
2	0.93	1.17	1.92
5	0.96	1.23	2.05
10	0.98	1.26	2.11

5. FINAL REMARKS

This paper describes only one possible approach to creating multivariate analogues to common univariate tests of hypotheses. Additional analogues based on alternative principles include the following. First, the so-called interdirection counts introduced by Randles (1989) can be used to construct nonparametric tests which are often asymptotically equivalent to the tests discussed here. Second, methods based on marginal signs and ranks were described by Puri and Sen (1971). Other methods, based on distances measured via volumes of simplices, were described by Oja (1999) and the references contained therein. Optimal signed-rank testing procedures based on interdirections and (univariate) ranks of the lengths of the residual vectors were developed by Hallin and Paindaveine (2002). Also different depth functions (Liu, Parelius and Singh, 1999; Zuo and Serfling, 2000) provide center-outward orderings or rankings of data points which can be used in test constructions.

The authors wish to thank Seija Sirkiä for her help in the implementation of the methods; the R functions used in the calculation can be found on her website: *http://www.maths.jyu.fi/~ssirkia/signrank/signrank.html*.

REFERENCES

BLOMQVIST, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statist.* **21** 593–600.

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

DÜMBGEN, L. (1998). On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.* **50** 471–491.

HALLIN, M. and PAINDAVEINE, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133.

HAND, D. J., DALY, F., LUNN, A. D., MCCONWAY, K. J. and OSTROWSKI, E., eds. (1994). *A Handbook of Small Data Sets*. Chapman and Hall, London.

HETTMANSPERGER, T. P. and RANDLES, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika* **89** 851–860.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika* **30** 81–93.

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

MERCHANT, J. A., HALPRIN, G. M., HUDSON, A. R., KILBURN, K. H., MCKENZIE, W. N., HURST, D. J. and BERMAZOHN, P. (1975). Responses to cotton dust. *Archives of Environmental Health* **30** 222–229.

MÖTTÖNEN, J. and OJA, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Statist.* **5** 201–213.

MÖTTÖNEN, J., OJA, H. and TIENARI, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.* **25** 542–552.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

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

PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

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

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

SPEARMAN, C. (1904). The proof and measurement of association between two things. *American J. Psychology* **15** 72–101.

TASKINEN, S., KANKAINEN, A. and OJA, H. (2003). Sign test of independence between two random vectors. *Statist. Probab. Lett.* **62** 9–21.

TASKINEN, S., OJA, H. and RANDLES, R. H. (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc.* **100**. To appear.

TYLER, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* **15** 234–251.

WILKS, S. S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica* **3** 309–326.

ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482.