Projection correlation between two random vectors

BY LIPING ZHU
Institute of Statistics and Big Data, Renmin University of China, 59 Zhongguancun Avenue, Beijing 100872, China
zhu.liping@ruc.edu.cn

KAI XU
School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China
tjxxukai@163.com

RUNZE LI
Department of Statistics, Pennsylvania State University, University Park, Pennsylvania 16802, U.S.A.
rzli@psu.edu

AND WEI ZHONG
Wang Yanan Institute for Studies in Economics, School of Economics, Xiamen University, Fujian 361005, China
wzhong@xmu.edu.cn

SUMMARY
We propose the use of projection correlation to characterize dependence between two random vectors. Projection correlation has several appealing properties. It equals zero if and only if the two random vectors are independent, it is not sensitive to the dimensions of the two random vectors, it is invariant with respect to the group of orthogonal transformations, and its estimation is free of tuning parameters and does not require moment conditions on the random vectors. We show that the sample estimate of the projection correlation is \( n \)-consistent if the two random vectors are independent and root-\( n \)-consistent otherwise. Monte Carlo simulation studies indicate that the projection correlation has higher power than the distance correlation and the ranks of distances in tests of independence, especially when the dimensions are relatively large or the moment conditions required by the distance correlation are violated.

Some key words: Distance correlation; Projection correlation; Ranks of distance.

1. INTRODUCTION

Let \( X \in \mathbb{R}^p \) and \( Y \in \mathbb{R}^q \) be two random vectors. In this paper, we aim to test

\[ H_0 : X \text{ and } Y \text{ are independent} \quad \text{versus} \quad H_1 : \text{otherwise}. \]
Measuring and testing dependence between $X$ and $Y$ is a fundamental problem in statistics. The Pearson correlation is perhaps the first and the best-known quantity to measure the degree of linear dependence between two univariate random variables. Extensions including Spearman’s (1904) $\rho$, Kendall’s (1938) $\tau$, and those due to Hoeffding (1948) and Blum (1961) can be used to measure nonlinear dependence without moment conditions.

Testing independence has important applications. Two examples from genomics research are testing whether two groups of genes are associated and examining whether certain phenotypes are determined by particular genotypes. In social science research, scientists are interested in understanding potential associations between psychological and physiological characteristics. Wilks (1935) introduced a parametric test based on $|\Sigma_{X,Y}|/(|\Sigma_X| \cdot |\Sigma_Y|)$, where $\Sigma_{X,Y} = \text{var}(X^T, Y^T) \in \mathbb{R}^{(p+q) \times (p+q)}$, $\Sigma_X = \text{var}(X) \in \mathbb{R}^{p \times p}$ and $\Sigma_Y = \text{var}(Y) \in \mathbb{R}^{q \times q}$. Throughout $\Sigma_Z = \text{var}(Z)$ stands for the covariance matrix of $Z$ and $|\Sigma_Z|$ stands for the determinant of $\Sigma_Z$. Hotelling (1936) suggested the canonical correlation coefficient, which seeks $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$ such that the Pearson correlation between $\alpha^T X$ and $\beta^T Y$ is maximized. Both Wilks’s test and the canonical correlation can be used to test for independence between $X$ and $Y$ when they follow normal distributions. Nonparametric extensions of Wilks’s test were proposed by Puri & Sen (1971), Hettmansperger & Oja (1994), Gieser & Randles (1997), Taskinen et al. (2003) and Taskinen et al. (2005). These tests can be used to test for independence between $X$ and $Y$ when they follow elliptically symmetric distributions, but they are inapplicable when the normality or ellipticity assumptions are violated or when the dimensions of $X$ and $Y$ exceed the sample size. In addition, multivariate rank-based tests of independence are ineffective for testing nonmonotone dependence (Székely et al., 2007).

The distance correlation (Székely et al., 2007) can be used to measure and test dependence between $X$ and $Y$ in arbitrary dimensions without assuming normality or ellipticity. Provided that $E(||X|| + ||Y||) < \infty$, the distance correlation between $X$ and $Y$, denoted by $\text{dcorr}(X, Y)$, is nonnegative, and it equals zero if and only if $X$ and $Y$ are independent. Throughout, we define $||b|| = (b^T b)^{1/2}$ for a vector $b$. Székely & Rizzo (2013) observed that the distance correlation may be adversely affected by the dimensions of $X$ and $Y$, and proposed an unbiased estimator of it when $X$ and $Y$ are high-dimensional. In this paper, we shall demonstrate that the distance correlation may be less efficient in detecting nonlinear dependence when the assumption $E(||X|| + ||Y||) < \infty$ is violated. To remove this moment condition, Benjamini et al. (2013) suggested using ranks of distances, but this involves the selection of several tuning parameters, the choice of which is an open problem. The asymptotic properties of a test based on ranks of distances also need further investigation.

We propose using projection correlation to characterize dependence between $X$ and $Y$. Projection correlation first projects the multivariate random vectors into a series of univariate random variables, then detects nonlinear dependence by calculating the Pearson correlation between the dichotomized univariate random variables. The projection correlation between $X$ and $Y$, denoted by $\text{pcorr}(X, Y)$, is nonnegative and equals zero if and only if $X$ and $Y$ are independent, so it is generally applicable as an index for measuring the degree of nonlinear dependence without moment conditions, normality or ellipticity (Tracz et al., 1992). The projection correlation test for independence is consistent against all dependence alternatives. The projection correlation is free of tuning parameters and is invariant to orthogonal transformation. We shall show that the sample estimator of projection correlation is $n$-consistent if $X$ and $Y$ are independent and root-$n$-consistent otherwise. We conduct Monte Carlo studies to evaluate the finite-sample performance of the projection correlation test. The results indicate that the projection correlation is less sensitive to the dimensions of $X$ and $Y$ than the distance correlation and even its improved version (Székely & Rizzo, 2013), and is more powerful than both the distance correlation and
ranks of distances, especially when the dimensions of $X$ and $Y$ are relatively large or the moment conditions required by the distance correlation are violated.

2. Projection Correlation

2.1. Motivation

In this section, we propose a new measure of dependence between two random vectors. Testing that $X$ and $Y$ are independent is equivalent to testing whether $U = \alpha^T X$ and $V = \beta^T Y$ are independent for all unit vectors $\alpha$ and $\beta$. Let $F_{U,V}(u,v)$ denote the joint distribution of $(U, V)$, and let $F_U(u)$ and $F_V(v)$ denote the marginal distributions of $U$ and $V$. Given $\alpha$ and $\beta$, $U$ and $V$ are independent if and only if $F_{U,V}(u,v) - F_U(u)F_V(v) = \text{cov}(I(\alpha^T X \leq u), I(\beta^T Y \leq v)) = 0$, for all $u, v \in \mathbb{R}^1$. Therefore, testing whether $X$ and $Y$ are independent amounts to testing whether

$$\int \int \int \text{cov}^2\{I(\alpha^T X \leq u), I(\beta^T Y \leq v)\} \text{d}F_{U,V}(u,v) \text{d}\alpha \text{d}\beta = 0. \quad (1)$$

Suppose that $\{(X_i, Y_i), i = 1, \ldots, n\}$ is a random sample of $(X, Y)$. Using the first five independent copies of $(X, Y)$, we rewrite the left-hand side of (1) as

$$\int \int \int E^2\{I(\alpha^T X_1 \leq u)I(\beta^T Y_1 \leq v) - I(\alpha^T X_1 \leq u)I(\beta^T Y_2 \leq v)\} \text{d}F_{U,V}(u,v) \text{d}\alpha \text{d}\beta$$

$$= \int \int E\{\{I(\alpha^T X_1 \leq \alpha^T X_3)I(\beta^T Y_1 \leq \beta^T Y_3) - I(\alpha^T X_1 \leq \alpha^T X_3)I(\beta^T Y_2 \leq \beta^T Y_3)\}$$

$$\times \{I(\alpha^T X_4 \leq \alpha^T X_3)I(\beta^T Y_4 \leq \beta^T Y_3) - I(\alpha^T X_4 \leq \alpha^T X_3)I(\beta^T Y_5 \leq \beta^T Y_3)\} \text{d}\alpha \text{d}\beta.$$

Consequently, by Fubini’s theorem, $X$ and $Y$ are independent if and only if

$$E\left(\int \int \{\{I(\alpha^T X_1 \leq \alpha^T X_3)I(\beta^T Y_1 \leq \beta^T Y_3) - I(\alpha^T X_1 \leq \alpha^T X_3)I(\beta^T Y_2 \leq \beta^T Y_3)\}$$

$$\times \{I(\alpha^T X_4 \leq \alpha^T X_3)I(\beta^T Y_4 \leq \beta^T Y_3) - I(\alpha^T X_4 \leq \alpha^T X_3)I(\beta^T Y_5 \leq \beta^T Y_3)\} \text{d}\alpha \text{d}\beta\right) = 0. \quad (2)$$

In general, integration over the $(p + q - 2)$-dimensional space $\{(|\alpha| = ||\beta|| = 1)\}$ is not straightforward. Lemma 1 enables us to derive an explicit form for (2).

**Lemma 1** (Escanciano, 2006). For two arbitrary vectors $U_1, U_2 \in \mathbb{R}^p$, we have

$$\int_{||\alpha||=1} I(\alpha^TU_1 \leq 0)I(\alpha^TU_2 \leq 0) \text{d}\alpha = c(p) \left\{\pi - \arccos\left(\frac{U_1^T U_2}{||U_1|| ||U_2||}\right)\right\},$$

where $c(p) = \pi^{p/2-1}/\Gamma(p/2)$, $\Gamma(\cdot)$ is the gamma function and $\arccos$ is the inverse cosine function.

Lemma 1 yields an explicit formula for the left-hand side of (2). Ignoring the constants irrelevant to the joint distribution of $(X, Y)$, we define the resultant explicit formula as the squared
projection covariance between $X$ and $Y$. To be precise, define

$$\{\text{Pcov}(X, Y)\}^2 = S_1 + S_2 - 2S_3$$

$$= E \left[ \arccos \left( \frac{(X_1 - X_3)\T (X_4 - X_3)}{\|X_1 - X_3\| \|X_4 - X_3\|} \right) \arccos \left( \frac{(Y_1 - Y_3)\T (Y_4 - Y_3)}{\|Y_1 - Y_3\| \|Y_4 - Y_3\|} \right) \right]$$

$$+ E \left[ \arccos \left( \frac{(X_1 - X_3)\T (X_4 - X_3)}{\|X_1 - X_3\| \|X_4 - X_3\|} \right) \arccos \left( \frac{(Y_2 - Y_3)\T (Y_4 - Y_3)}{\|Y_2 - Y_3\| \|Y_4 - Y_3\|} \right) \right]$$

$$- 2E \left[ \arccos \left( \frac{(X_1 - X_3)\T (X_4 - X_3)}{\|X_1 - X_3\| \|X_4 - X_3\|} \right) \arccos \left( \frac{(Y_2 - Y_3)\T (Y_4 - Y_3)}{\|Y_2 - Y_3\| \|Y_4 - Y_3\|} \right) \right],$$

(3)

where $S_1$, $S_2$ and $S_3$ are defined in an obvious manner. We provide details of the derivation of (3) in the Appendix. A distinctive feature of $\text{Pcov}(X, Y)$ is that it uses only vectors of the form $(X_k - X_i)/\|X_k - X_i\|$ and $(Y_k - Y_i)/\|Y_k - Y_i\|$, whose second moments always equal unity, regardless of the dimensions of the random vectors. This indicates that the projection covariance removes the moment restrictions on $(X, Y)$ required by the distance correlation.

Define the projection correlation between $X$ and $Y$, denoted by $\text{pc}(X, Y)$, as the square root of

$$\{\text{pc}(X, Y)\}^2 = \frac{\{\text{Pcov}(X, Y)\}^2}{\text{Pcov}(X, X)\text{Pcov}(Y, Y)},$$

and set $\text{pc}(X, Y) = 0$ if $\text{Pcov}(X, X) = 0$ or $\text{Pcov}(Y, Y) = 0$. Proposition 1 presents the appealing properties of the projection correlation at the population level.

**Proposition 1.**

(i) In general, $0 \leq \text{pc}(X, Y) \leq 1$. In particular, $\text{pc}(X, Y) = 0$ if and only if $X$ and $Y$ are independent, and $\text{pc}(X, X) = 0$ if and only if $X = E(X)$ almost surely.

(ii) Let $C_1 \in \mathbb{R}^{p \times p}$ and $C_2 \in \mathbb{R}^{q \times q}$ be two orthonormal matrices, $a_1 \in \mathbb{R}^p$ and $a_2 \in \mathbb{R}^q$ be two vectors, and $a_1$ and $b_1$ be two scalars. Then $\text{pc}(X, Y) = \text{pc}(a_1 + b_1 C_1 X, a_2 + b_2 C_2 Y)$.

The first statement indicates that the projection correlation is generally applicable as an index to measure dependence. The second statement implies that, although it is not affine-invariant, the projection correlation is invariant with respect to the group of orthogonal transformations.

2-2. Asymptotic properties

We give two equivalent estimators for $\text{Pcov}(X, Y)$ and study their asymptotic properties. The first estimate is built upon the $V$-statistic (Serfling, 1980), given by the square root of

$$\{\hat{\text{Pcov}}_1(X, Y)\}^2 = \hat{S}_1 + \hat{S}_2 - 2\hat{S}_3$$

$$= n^{-3} \sum_{i,k,l=1}^n \left[ \arccos \left( \frac{(X_i - X_k)\T (X_l - X_k)}{\|X_i - X_k\| \|X_l - X_k\|} \right) \arccos \left( \frac{(Y_i - Y_k)\T (Y_l - Y_k)}{\|Y_i - Y_k\| \|Y_l - Y_k\|} \right) \right]$$

$$+ n^{-5} \sum_{i,j,k,l,r=1}^n \left[ \arccos \left( \frac{(X_i - X_k)\T (X_l - X_k)}{\|X_i - X_k\| \|X_l - X_k\|} \right) \arccos \left( \frac{(Y_j - Y_k)\T (Y_r - Y_k)}{\|Y_j - Y_k\| \|Y_r - Y_k\|} \right) \right]$$

$$- 2n^{-4} \sum_{i,j,k,l=1}^n \left[ \arccos \left( \frac{(X_i - X_k)\T (X_l - X_k)}{\|X_i - X_k\| \|X_l - X_k\|} \right) \arccos \left( \frac{(Y_j - Y_k)\T (Y_l - Y_k)}{\|Y_j - Y_k\| \|Y_l - Y_k\|} \right) \right].$$
Here $\hat{S}_1$, $\hat{S}_2$ and $\hat{S}_3$ are defined in an obvious fashion and are the estimates of $S_1$, $S_2$ and $S_3$, respectively. The $V$-statistic estimate appears natural, yet it is difficult to calculate (Szekely & Rizzo, 2010). Therefore, we give an equivalent form below. Define, for $k, l, r = 1, \ldots, n$,

\[
a_{klr} = \arccos \left\{ \frac{(X_k - X_r)^T (X_l - X_r)}{\|X_k - X_r\| \|X_l - X_r\|} \right\},
\]

\[
\bar{a}_{k,r} = n^{-1} \sum_{l=1}^{n} a_{klr}, \quad \bar{a}_{l,r} = n^{-1} \sum_{k=1}^{n} a_{klr}, \quad \bar{a}_{.,r} = n^{-2} \sum_{k=1}^{n} \sum_{l=1}^{n} a_{klr},
\]

\[
A_{klr} = a_{klr} - \bar{a}_{k,r} - \bar{a}_{l,r} + \bar{a}_{.,r},
\]

\[
b_{klr} = \arccos \left\{ \frac{(Y_k - Y_r)^T (Y_l - Y_r)}{\|Y_k - Y_r\| \|Y_l - Y_r\|} \right\},
\]

\[
\bar{b}_{k,r} = n^{-1} \sum_{l=1}^{n} b_{klr}, \quad \bar{b}_{l,r} = n^{-1} \sum_{k=1}^{n} b_{klr}, \quad \bar{b}_{.,r} = n^{-2} \sum_{k=1}^{n} \sum_{l=1}^{n} b_{klr},
\]

\[
B_{klr} = b_{klr} - \bar{b}_{k,r} - \bar{b}_{l,r} + \bar{b}_{.,r},
\]

To avoid possible confusion, we define $a_{klr} = b_{klr} = 0$ if $k = r$ or $l = r$. The second sample estimate of $P_{\text{cov}}(X, Y)$ is defined by

\[
\{P_{\hat{\text{cov}}}(X, Y)\}^2 = n^{-3} \sum_{k,l,r=1}^{n} A_{klr} B_{klr}.
\]

Accordingly, the sample estimate of $p_{C}(X, Y)$ is defined by the square root of

\[
\{\hat{p}_{C}(X, Y)\}^2 = \frac{\{P_{\hat{\text{cov}}}(X, Y)\}^2}{P_{\text{cov}}(X, Y) P_{\hat{\text{cov}}}(X, Y)}.
\]

In general, $P_{\hat{\text{cov}}}(X, Y)$ is easier to compute than $P_{\hat{\text{cov}}}(X, Y)$. Although it may not be immediately obvious that $P_{\text{cov}}(X, Y) \geq 0$, this fact will become clear from Theorem 1.

**Theorem 1.** For a given random sample $\{(X_i, Y_i), i = 1, \ldots, n\}$,

\[
\{P_{\hat{\text{cov}}}(X, Y)\}^2 = \{P_{\hat{\text{cov}}}(X, Y)\}^2,
\]

and both equal

\[
(nc(p)c(q))^{-1} \sum_{i=1}^{n} \left[ \int \int \{\hat{F}_{U',Y'}(\alpha^T X_i, \beta^T Y_i) - \hat{F}_{U}(\alpha^T X_i)\hat{F}_{Y}(\beta^T Y_i)\}^2 d\alpha d\beta \right],
\]

where $\hat{F}_{U,Y}$, $\hat{F}_{U}$ and $\hat{F}_{Y}$ stand for the empirical distributions of $(\alpha^T X, \beta^T Y)$, $(\alpha^T X)$ and $(\beta^T Y)$, respectively, $c(p) = \pi^{p/2-1}/\Gamma(p/2)$, and $c(q) = \pi^{q/2-1}/\Gamma(q/2)$.

The following theorems state the consistency of $P_{\hat{\text{cov}}}(X, Y)$ and $P_{\hat{\text{cov}}}(X, Y)$.

**Theorem 2.** For a given random sample $\{(X_i, Y_i), i = 1, \ldots, n\}$, $\lim_{n \to \infty} P_{\hat{\text{cov}}}(X, Y) = P_{\text{cov}}(X, Y)$ almost surely.
Theorem 3.
(i) If $X$ and $Y$ are independent, then as $n \to \infty$, $n \left\{ \hat{P}\text{cov}(X, Y) \right\}^2 / (\pi^2 - \hat{S}_2)$ converges in distribution to $\sum_{k=1}^{\infty} \lambda_k Z_k^2$, where the $\lambda_k$ depend on the distribution of $(X, Y)$ and are nonnegative with sum equal to one, and the $Z_k$ are independent standard normal random variables.

(ii) If $X$ and $Y$ are not independent, then $n^{1/2} \left\{ \hat{P}\text{cov}(X, Y) - P\text{cov}(X, Y) \right\}$ converges in distribution to a normal distribution with mean zero and variance $\text{var}(N)$, where the random variable $N$ is defined in (A2). Consequently, $n \left\{ \hat{P}\text{cov}(X, Y) \right\}^2 / (\pi^2 - \hat{S}_2)$ diverges to $\infty$.

The projection correlation test is built upon the test statistic $T_n = n \left\{ \hat{P}\text{cov}(X, Y) \right\}^2 / (\pi^2 - \hat{S}_2)$, which converges in distribution to the quadratic form if $X$ and $Y$ are independent and diverges to $\infty$ otherwise. Theorem 3 suggests that the projection correlation test is consistent against all dependence alternatives without requiring any moment conditions. Because the weights $\lambda_k$ in the quadratic form are unknown, the asymptotic null distribution is intractable. To put the projection correlation test into practice, we approximate the asymptotic null distribution of $T_n$ through a random permutation method. Specifically, we calculate replicates of the test statistic under random permutations of the indices of the $X$ sample or, equivalently, the $Y$ sample. The $p$-value obtained from this permutation procedure is defined as the fraction of replicates of the test statistic under random permutations that are at least as large as the observed test statistic. The permutation procedure is computationally feasible owing to the simple form of the test statistic. Computer code for implementing the projection correlation test and the permutation procedure is available from the authors upon request.

3. Simulations

In this section, we conduct simulations to compare the performance of independence tests based on the projection correlation, the distance correlation and the ranks of distances (Benjamini et al., 2013). These three tests are consistent and suitable for arbitrary dimensions. Because the distance-correlation-based test is sensitive to the dimensions of random vectors, throughout our simulations we use its improved version recommended by Székely & Rizzo (2013).

We consider three simulated examples in which the dimensions of both $X$ and $Y$, denoted by $p$ and $q$, respectively, are relatively large for the sample size $n$. We set $p = q$ for simplicity. In Example 1, we set $n = 30$ and vary $p$ from 15 to 30. In Example 2, we set $p = 100$ and vary $n$ from 10 to 30. We also vary $p$ from 30 to 60 and $n$ from 10 to 30. In Example 3, we set $p = q = 1000$ and vary $n$ from 20 to 40. The dependence structure is monotone in Example 1 and nonmonotone in Example 2. The dependence structure is much more complicated in Example 3, where the random vectors are drawn from a mixture of distributions.

All simulations are implemented in R (R Development Core Team, 2017). We implement the test based on distance correlation by calling the dcor.test function in the energy package and the test based on ranks of distances by calling the hhg.test function in the HHG package. We repeat each setting 2000 times and report the size and power of the respective tests at significance levels $\alpha = 0.01$ and 0.05.

Example 1. We consider three scenarios in this example.

(1a) Draw $X = (X_1, \ldots, X_p)^T$ independently from a Cauchy distribution. Let $Y_i = \exp(X_i)$ ($i = 1, \ldots, m$) and draw $Y_i$ ($i = m + 1, \ldots, q$) independently from a standard normal distribution.
example tests. Projection correlation is much less sensitive to the increase of dimensions than the other two correlation. The test based on ranks of distances is slightly worse than our test based on projection correlation, especially in scenarios (1a) and (1b), where the power of all tests diminishes quickly as $\alpha$ and $\beta$ increase. Table 1 charts the empirical size and power of the tests based on projection correlation, ranks of distances, and distance correlation at significance levels $\alpha$ = 0.01 and 0.05. In all scenarios, the empirical sizes are very close to the significance levels, even when $\alpha = 0.01$. The test based on projection correlation has higher power than those based on distance correlation or ranks of distances, especially in scenarios (1a) and (1b), where the distributions of the random vectors are all heavy-tailed. The test based on distance correlation fails in the first two scenarios, partly because the moment restrictions required by this test are violated. The test based on ranks of distances is slightly worse than our test based on projection correlation.

In the above scenarios, we set $m = 0, 2, 4, 6, 8$ and 10, where $m = 0$ indicates that $X$ and $Y$ are independent. Table 1 charts the empirical size and power of the tests based on projection correlation, ranks of distances, and distance correlation at significance levels $\alpha = 0.01$ and 0.05. In all scenarios, the empirical sizes are very close to the significance levels, even when $\alpha = 0.01$. The test based on projection correlation has higher power than those based on distance correlation or ranks of distances, especially in scenarios (1a) and (1b), where the distributions of the random vectors are all heavy-tailed. The test based on distance correlation fails in the first two scenarios, partly because the moment restrictions required by this test are violated. The test based on ranks of distances is slightly worse than our test based on projection correlation.

We also vary the dimensions of $X$ and $Y$ from 15 to 30 and fix $m = 10$ in scenarios (1a) and (1b) and $m = 2$ in scenario (1c). The simulation results are summarized in Table 2. The power of all tests diminishes quickly as $p$ increases. Table 2 indicates that the test based on projection correlation is much less sensitive to the increase of dimensions than the other two tests.

Example 2. We draw $X_1, \ldots, X_p$ independently from the uniform distribution on $[0, 2\pi]$. We generate $Y_i = X_1 + 2X_2 + \cdots + iX_i + \{iX_1 + (i - 1)X_2 + \cdots + X_i\}^2 + \varepsilon_i (i = 1, \ldots, m)$, where the $\varepsilon_i$ are generated from the Cauchy distribution, and generate $Y_i (i = m + 1, \ldots, q)$ independently from the standard normal distribution. This model was also used in Escanciano (2006) for different purposes. In this example, $m = 0$ indicates that $X$ and $Y$ are independent.
We first fix $p = q = 100$ and vary $n$ from 10 to 30. The empirical size and power are displayed in Table 3 for $m=0, 5, 15$ and 25 and $n=10, 20$ and 30. All empirical sizes are close to the significance level. In this example, the moment conditions required by the test based on distance correlation are satisfied. The tests based on projection correlation and on distance correlation are more powerful than that based on ranks of distances, which appears to be very ineffective in this example, partly because the dependence structure is nonmonotone and the dependence strength is very weak.
Next we fix \( m = 25 \) and vary \( n \) from 10 to 30 and \( p \) from 30 to 60. Table 4 shows that, provided \( n \geq 30 \), say, the test based on projection correlation results in much less power loss across almost all scenarios as the dimensions of \( X \) and \( Y \) increase.

Example 3. This example was used in Benjamini et al. (2013). We fix \( p = q = 1000 \) and vary \( n \) from 20 to 40. We draw \((X, Y)\) from a mixture distribution with 10 equally likely components. In the \( i \)th component, for \( i = 1, \ldots, 10 \), \((X, Y)\) are random vectors \( \{U_X(i) + \epsilon_X, U_Y(i) + \epsilon_Y\} \), where \( U_X(i) \) and \( U_Y(i) \) are sampled independently from a multivariate standard normal distribution, and \( (\epsilon_X, \epsilon_X) \) are sampled independently from a multivariate Cauchy or multivariate \( t \) distribution with three degrees of freedom and the identity correlation matrix. The dependence of \( X \) and \( Y \) is through the fixed pairs \( \{U_X(i), U_Y(i)\} \) \( i = 1, \ldots, 10 \), such that the data consist of ten clouds around these pairs.

The simulations are summarized in Table 5. The dependence of \( X \) and \( Y \) is through the ten equally likely components. The test based on projection correlation performs better than that based on distance correlation, especially when the moment requirements are not satisfied. The
improved version of the test based on distance correlation is designed for high dimensions, and its performance appears satisfactory when the moments exist. Again, for the multivariate Cauchy distribution, the test based on projection correlation outperforms that based on distance correlation significantly in this example.

In our simulations, the test based on projection correlation exhibits a good capability for testing monotone and nonmonotone dependence. Our limited experience indicates that it is very effective, even when the second moments are large or infinite, it is useful for limiting the power loss as the dimensions of random vectors increase, and it is suitable even in high-dimensional cases.

**ACKNOWLEDGEMENT**

The authors would like to thank Ms Amanda Applegate, the associate editor, and two reviewers for their constructive comments. Li and Zhong are the corresponding authors. This research was supported by the National Natural Science Foundation of China, National Science Foundation of USA, Chinese Ministry of Education Project of Key Research Institute of Humanities and Social Sciences at Universities, National Institute of Drug Abuse and National Institutes of Health of USA, and National Youth Top-notch Talent Support Program of China.

**APPENDIX**

**Proofs**

We first show that by invoking Lemma 1 repeatedly, the squared projection covariance \( \{\text{Pcov}(X, Y)\}^2 \) has an explicit form. In other words, we aim to show that

\[
\{\text{Pcov}(X, Y)\}^2 = (c(p)c(q))^{-1} \int_{|\alpha|=1} \int_{|\beta|=1} E \left( F_U(Y, (\alpha^T X_3, \beta^T Y_3) - F_U(\alpha^T X_3, \beta^T Y_3) \right)^2 d\alpha d\beta
\]

\[
= E \left[ \arccos \left( \frac{(X_1 - X_3)^T (X_4 - X_3)}{\|X_1 - X_3\| \|X_4 - X_3\|} \right) \arccos \left( \frac{(Y_1 - Y_3)^T (Y_4 - Y_3)}{\|Y_1 - Y_3\| \|Y_4 - Y_3\|} \right) \right]
\]

\[
+ E \left[ \arccos \left( \frac{(X_1 - X_3)^T (X_4 - X_3)}{\|X_1 - X_3\| \|X_4 - X_3\|} \right) \arccos \left( \frac{(Y_2 - Y_3)^T (Y_4 - Y_3)}{\|Y_2 - Y_3\| \|Y_4 - Y_3\|} \right) \right]
\]

\[
- 2E \left[ \arccos \left( \frac{(X_1 - X_3)^T (X_4 - X_3)}{\|X_1 - X_3\| \|X_4 - X_3\|} \right) \arccos \left( \frac{(Y_2 - Y_3)^T (Y_4 - Y_3)}{\|Y_2 - Y_3\| \|Y_4 - Y_3\|} \right) \right].
\]

For notational clarity, we define

\[
a_{ik} = \arccos \left( \frac{(X_k - X_i)^T (X_i - X_j)}{\|X_k - X_i\| \|X_i - X_j\|} \right), \quad b_{ik} = \arccos \left( \frac{(Y_k - Y_i)^T (Y_i - Y_j)}{\|Y_k - Y_i\| \|Y_i - Y_j\|} \right).
\]

All the indices in \( a_{ik} \) and \( b_{ik} \) may take value 1, 2, 3, 4 or 5. Invoking Lemma 1 repeatedly, we obtain

\[
\{c(p)c(q)\}^{-1} \int_{|\alpha|=1} \int_{|\beta|=1} E \left( F_U(Y, (\alpha^T X_3, \beta^T Y_3) - F_U(\alpha^T X_3, \beta^T Y_3) \right)^2 d\alpha d\beta
\]

\[
= E[(\pi - a_{135})(\pi - b_{135})] - E[(\pi - a_{135})(\pi - b_{145})] - E[(\pi - a_{135})(\pi - b_{235})]
\]

\[
+ E[(\pi - a_{135})(\pi - b_{245})]
\]

\[
= E(a_{135}b_{135}) + E(a_{135}b_{245}) - 2E(a_{135}b_{145})
\]

\[
= S_1 + S_2 - 2S_3.
\]

The last equality follows from \( E(b_{145} - b_{135} - b_{245} + b_{235}) = 0 \) and \( E(a_{135}b_{145}) = E(a_{135}b_{235}) \).
Proof of Proposition 1

We prove the first assertion. The statement that $0 \leq \text{PC}(X, Y) \leq 1$ is a direct consequence of the Cauchy–Schwarz inequality. By definition, $\text{PC}(X, Y) = 0$ indicates that $\alpha^T X$ and $\beta^T Y$ are independent for any $\|\alpha\| = 1$ and $\|\beta\| = 1$. In other words, $\text{PC}(X, Y) = 0$ if and only if $X$ and $Y$ are statistically independent. In addition, $\text{PC}(X, X) = 0$ indicates that $X$ must be a constant vector, because otherwise $X$ would not be independent of itself.

By definition, $(\text{Pcov}(X, Y))^2 = S_1 + S_2 - 2S_3$, and all the $S_k$ involve quantities of the form $a_{klr}$ and $b_{klr}$. It is easy to verify that both $a_{klr}$ and $b_{klr}$ are invariant with respect to orthogonal transformations, which completes the proof of the second assertion.

Proof of Theorem 1

We first prove that $\{\hat{\text{Pcov}}_1(X, Y)\}^2 = \{\text{Pcov}(X, Y)\}^2$. Recall the definitions of $a_{klr}$ and $b_{klr}$. Define

$$\bar{a}_{k,r} = n^{-1} \sum_{l=1}^{n} a_{klr}, \quad \bar{a}_{l,r} = n^{-1} \sum_{k=1}^{n} a_{klr}, \quad \bar{a}_{r} = n^{-1} \sum_{k=1}^{n} \sum_{l=1}^{n} a_{klr}, \quad \bar{b}_{k,r} = n^{-2} \sum_{k=1}^{n} \sum_{l=1}^{n} b_{klr},$$

$$A_{klr} = a_{klr} - \bar{a}_{k,r} - \bar{a}_{l,r} + \bar{a}_{r} , \quad B_{klr} = b_{klr} - \bar{b}_{k,r} - \bar{b}_{l,r} + \bar{b}_{r} .$$

We further define

$$I_1 = n^{-3} \sum_{i,j,r=1}^{n} a_{ijr} b_{ijr} , \quad I_2 = n^{-3} \sum_{i,j,r=1}^{n} a_{ijr} \bar{b}_{ijr} , \quad I_3 = n^{-3} \sum_{i,j,r=1}^{n} a_{ijr} \bar{b}_{ijr} ,$$

$$I_4 = n^{-1} \sum_{r=1}^{n} a_{r} \bar{b}_{r} , \quad I_5 = n^{-3} \sum_{i,j,r=1}^{n} \bar{a}_{ijr} b_{ijr} , \quad I_6 = n^{-2} \sum_{i,j,r=1}^{n} \bar{a}_{ijr} \bar{b}_{ijr} ,$$

$$I_7 = n^{-3} \sum_{i,j,r=1}^{n} \bar{a}_{ijr} \bar{b}_{ijr} , \quad I_8 = n^{-2} \sum_{i,j,r=1}^{n} \bar{a}_{ijr} \bar{b}_{ijr} , \quad I_9 = n^{-3} \sum_{i,j,r=1}^{n} \bar{a}_{ijr} b_{ijr} ,$$

$$I_{10} = n^{-3} \sum_{i,j,r=1}^{n} a_{ijr} \bar{b}_{ijr} , \quad I_{11} = n^{-2} \sum_{i,j,r=1}^{n} \bar{a}_{ijr} \bar{b}_{ijr} , \quad I_{12} = n^{-2} \sum_{i,j,r=1}^{n} \bar{a}_{ijr} \bar{b}_{ijr} ,$$

$$I_{13} = n^{-1} \sum_{r=1}^{n} a_{r} \bar{b}_{r} , \quad I_{14} = n^{-2} \sum_{i,j,r=1}^{n} a_{ijr} \bar{b}_{ijr} , \quad I_{15} = n^{-2} \sum_{i,j,r=1}^{n} a_{ijr} \bar{b}_{ijr} ,$$

$$I_{16} = n^{-1} \sum_{r=1}^{n} a_{r} \bar{b}_{r} .$$

It can be verified that $I_2 = I_5 = I_6 = I_9 = I_{11} = I_{14} = I_{15} = I_{16}$. It follows that

$$\{\text{Pcov}(X, Y)\}^2 = n^{-3} \sum_{k,j,r=1}^{n} A_{klr} B_{klr}$$

$$= n^{-3} \sum_{k,j,r=1}^{n} (a_{klr} - \bar{a}_{k,r} - \bar{a}_{l,r} + \bar{a}_{r})(b_{klr} - \bar{b}_{k,r} - \bar{b}_{l,r} + \bar{b}_{r})$$

$$= I_1 - I_2 + I_4 - I_5 + I_6 + I_7 - I_8 - I_9 + I_{10} + I_{11} - I_{12} + I_{13} - I_{14} - I_{15} + I_{16}$$

$$= I_1 - 2I_2 + I_4 = \hat{S}_1 + \hat{S}_2 - 2\hat{S}_3 = \{\text{Pcov}_1(X, Y)\}^2 ,$$

which completes the proof of the first part.
Next we prove that \( \{P_{\text{cov}}(X, Y)\}^2 \) is equal to

\[
\{nc(p)c(q)\}^{-1} \sum_{i=1}^{n} \left[ \int_{|\alpha| = 1} \int_{|\beta| = 1} \left\{ \hat{F}_{U,Y}(\alpha^T X_i, \beta^T Y_i) - \hat{F}_{U}(\alpha^T X_i)\hat{F}_{Y}(\beta^T Y_i) \right\}^2 \, d\alpha \, d\beta \right].
\]

Invoking Lemma 1, we have

\[
\int \left\{ I(\alpha^T X_i \leq \alpha^T X_r) - n^{-1} \sum_{i=1}^{n} I(\alpha^T X_i \leq \alpha^T X_r) \right\} \times \left\{ I(\alpha^T X_j \leq \alpha^T X_r) - n^{-1} \sum_{j=1}^{n} I(\alpha^T X_j \leq \alpha^T X_r) \right\} \, d\alpha
\]

\[
= -c(p)(a_{ij} - \bar{a}_{i} - \bar{a}_{j} + \bar{a}) = -c(p)A_{ij}.
\]

Following similar arguments, we obtain

\[
\int \left\{ I(\beta^T Y_i \leq \beta^T Y_r) - n^{-1} \sum_{i=1}^{n} I(\beta^T Y_i \leq \beta^T Y_r) \right\} \times \left\{ I(\beta^T Y_j \leq \beta^T Y_r) - n^{-1} \sum_{j=1}^{n} I(\beta^T Y_j \leq \beta^T Y_r) \right\} \, d\beta
\]

\[
= -c(q)(b_{ij} - \bar{b}_{i} - \bar{b}_{j} + \bar{b}) = -c(q)B_{ij}.
\]

The above two results yield

\[
\{nc(p)c(q)\}^{-1} \sum_{i=1}^{n} \left[ \int_{|\alpha| = 1} \int_{|\beta| = 1} \left\{ \hat{F}_{U,Y}(\alpha^T X_i, \beta^T Y_i) - \hat{F}_{U}(\alpha^T X_i)\hat{F}_{Y}(\beta^T Y_i) \right\}^2 \, d\alpha \, d\beta \right] = n^{-3} \sum_{k,l,r=1}^{n} A_{kl}B_{lr}.
\]

The proof of Theorem 1 is complete.

**Proof of Theorem 2**

By definition, \( \{P_{\text{cov}}(X, Y)\}^2 = \hat{S}_1 + \hat{S}_2 - 2\hat{S}_3 \). By the strong law of large numbers for \( V \)-statistics (Serfling, 1980), it follows that almost surely \( \lim_{n \to \infty} \hat{S}_1 = S_1 \), \( \lim_{n \to \infty} \hat{S}_2 = S_2 \), and \( \lim_{n \to \infty} \hat{S}_3 = S_3 \). Therefore, \( \lim_{n \to \infty} \{P_{\text{cov}}(X, Y)\}^2 \) almost surely. This completes the proof.

**Proof of Theorem 3**

Define the empirical process

\[
R_n(\alpha, \beta, u, v) = n^{-1/2} \sum_{i=1}^{n} \left\{ I(\alpha^T X_i \leq u) - pr(\alpha^T X_i \leq u) \right\} \left\{ I(\beta^T Y_i \leq v) - pr(\beta^T Y_i \leq v) \right\},
\]

where \( \alpha \in \mathbb{R}^q, \beta \in \mathbb{R}^q, u \in \mathbb{R} \) and \( v \in \mathbb{R} \). When \( X \) and \( Y \) are independent, \( R_n(\alpha, \beta, u, v) \) converges in distribution to a zero-mean Gaussian random process \( R(\alpha, \beta, u, v) \) with covariance function

\[
E[R(\alpha_1, \beta_1, u_1, v_1)R(\alpha_2, \beta_2, u_2, v_2)]
\]

\[
= E[I(\alpha_1^T X \leq u_1, \alpha_2^T X \leq u_2)]E[I(\beta_1^T Y \leq v_1, \beta_2^T Y \leq v_2)] - F_U(u_1)F_U(u_2)F_V(v_1)F_V(v_2).
\]
Next we define an approximation of \( R_n(\alpha, \beta, u, v) \), denoted by \( W_n(\alpha, \beta, u, v) \), as follows:

\[
W_n(\alpha, \beta, u, v) = n^{1/2} \left\{ n^{-1} \sum_{i=1}^{n} I(\alpha^T X_i \leq u, \beta^T Y_i \leq v) - n^{-1} \sum_{i=1}^{n} I(\alpha^T X_i \leq u)n^{-1} \sum_{i=1}^{n} I(\beta^T Y_i \leq v) \right\}.
\]

We first prove that \( W_n(\alpha, \beta, u, v) - R_n(\alpha, \beta, u, v) = o(1) \) holds uniformly for \((\alpha, \beta, u, v)\) with \( \|\alpha\| = 1 \) and \( \|\beta\| = 1 \). It is easy to verify that

\[
W_n(\alpha, \beta, u, v) - R_n(\alpha, \beta, u, v) = -n^{-1} \sum_{i=1}^{n} \left\{ I(\alpha^T X_i \leq u) - F_U(u) \right\} n^{-1/2} \sum_{i=1}^{n} \left\{ I(\beta^T Y_i \leq v) - F_V(v) \right\}.
\]

Invoking the uniform law of large numbers of Jennrich (1969) or the generalization by Wolfowitz (1954) of the Glivenko–Cantelli theorem, we know that \( n^{-1} \sum_{i=1}^{n} I(\alpha^T X_i \leq u) - F_U(X) = o(1) \) uniformly for \((\alpha, u)\) with \( \|\alpha\| = 1 \). Using Theorem 2.5.2 in van der Vaart & Wellner (1996), we can show that \( n^{-1/2} \sum_{i=1}^{n} \left\{ I(\beta^T Y_i \leq v) - F_V(v) \right\} \) converges to a Gaussian process with zero mean and variance-covariance function

\[
E \left\{ I(\beta_1^* Y_1 \leq v_1, \beta_2^* Y_2 \leq v_2) - F_V(v_1)F_V(v_2) \right\}.
\]

Therefore, \( W_n(\alpha, \beta, u, v) - R_n(\alpha, \beta, u, v) = o(1) \) holds uniformly for \((\alpha, \beta, u, v) \in \mathbb{R}^{p+q+2} \).

Using Theorem 2.5.2 in van der Vaart & Wellner (1996) again, we can show that the finite-dimensional distributions of \( R_n(\cdot) \) converge to \( R(\cdot) \), which implies that \( R_n(\cdot) \) is asymptotically tight. Therefore, for a random continuous functional, Lemma 3.1 in Chang (1990) yields

\[
n \{P\text{cov}(X, Y)\}^2 = \{c(p)c(q)\}^{-1} \int W_n^2(\alpha, \beta, u, v) \hat{F}_{U,V}(du, dv) \, d\alpha \, d\beta
\]

and converges in distribution to \( \{c(p)c(q)\}^{-1} \int R^2(\alpha, \beta, u, v) F_{U,V}(du, dv) \, d\alpha \, d\beta \). When \( X \) and \( Y \) are independent, \( R(\cdot) \) is a zero-mean process. According to Kuo (1975, Ch. 1, § 2),

\[
\{c(p)c(q)\}^{-1} \int R^2(\alpha, \beta, u, v) F_{U,V}(du, dv) \, d\alpha \, d\beta
\]

(A1)

follows the same distribution as \( \sum_{j=1}^{\infty} \lambda_j^* Z_j^2 \), where the \( Z_j \) are independent standard normal random variables, and in general, the nonnegative constants \( \lambda_j^* \) depend on the distribution of \((X, Y)\).

Next we derive the sum of the \( \lambda_j^* \). In view of (A1), we easily find that

\[
\{c(p)c(q)\}^{-1} E \left\{ \int R^2(\alpha, \beta, u, v) F_{U,V}(du, dv) \, d\alpha \, d\beta \right\} = \sum_{j=1}^{\infty} \lambda_j^*.
\]

Next we calculate the sum of \( \lambda_j^* \). If \( X \) and \( Y \) are independent, then

\[
E \left\{ \int R^2(\alpha, \beta, u, v) F_{U,V}(du, dv) \, d\alpha \, d\beta \right\}
\]

\[
= \int_{\|\alpha\|=1} \int_{\|\beta\|=1} E \{F_U(\alpha^T X)F_V(\beta^T Y) - F_U^2(\alpha^T X)F_V^2(\beta^T Y)\} \, d\alpha \, d\beta
\]

\[
= E \left\{ \int_{\|\alpha\|=1} I(\alpha^T X_i \leq \alpha^T X_i) \, d\alpha \right\} E \left\{ \int_{\|\beta\|=1} I(\beta^T Y_i \leq \beta^T Y_i) \, d\beta \right\}.
\]
Using Lemma 1, the right-hand side of the above equation is equal to

\[ \pi^2 - E \left[ \arccos \frac{(X_1 - X_3)^\top (X_1 - X_3)}{\|X_1 - X_3\| \|X_3 - X_2\|} \right] \arccos \left[ \frac{(Y_1 - Y_3)^\top (Y_1 - Y_3)}{\|Y_1 - Y_2\| \|Y_3 - Y_2\|} \right] = \pi^2 - S_2. \]

By the strong law of large numbers for V-statistics, we complete the proof of the first part.

Next, we deal with the second part. We approximate Š1 with the U-statistics Ûk, which can be approximated with their projections. The projections of the U-statistics are averages of independent and identically distributed random variables, and thus the asymptotic normality follows. Define C(n, a) to be the number of a combinations from a set of n elements. Define the U-statistic

\[ \hat{U}_1 = C(n, 3)^{-1} \sum_{1 \leq i < j < k \leq n} K_1(X_i, Y_i; X_j, Y_j; X_k, Y_k) \]

with the kernel \( K_1(X_i, Y_i; X_j, Y_j; X_k, Y_k) = 6^{-1} \sum_3 a_{i1j2i3} b_{i1j2i3} = k_1 \), where \( \sum_3 \) is the permutation of three distinct elements (i, j, k). Define the U-statistic

\[ \hat{U}_2 = C(n, 5)^{-1} \sum_{1 \leq i < j < k < l < r \leq n} K_2(X_i, Y_i; X_j, Y_j; X_k, Y_k; X_l, Y_l; X_r, Y_r) \]

with the kernel \( K_2(X_i, Y_i; X_j, Y_j; X_k, Y_k; X_l, Y_l; X_r, Y_r) = 120^{-1} \sum_5 a_{i1j2i3} b_{i1j2i3} = k_2 \), where \( \sum_5 \) is the permutation of three distinct elements (i, j, k, l, r). Define the U-statistic

\[ \hat{U}_3 = C(n, 4)^{-1} \sum_{1 \leq i < j < k < l \leq n} K_3(X_i, Y_i; X_j, Y_j; X_k, Y_k; X_l, Y_l) \]

with the kernel \( K_3(X_i, Y_i; X_j, Y_j; X_k, Y_k; X_l, Y_l) = 24^{-1} \sum_4 a_{i1j2i3} b_{i1j2i3} = k_3 \), where \( \sum_4 \) is the permutation of three distinct elements (i, j, k, l). Using standard V- and U-statistic theory, we have

\[ n^{1/2} \left[ \left( \text{Pcov}(X, Y) \right)^2 - (\text{Pcov}(X, Y))^2 \right] = n^{1/2} \left[ \hat{S}_1 + \hat{S}_2 - 2 \hat{S}_1 \right] = n^{1/2} \left[ \hat{U}_1 + \hat{U}_2 - 2 \hat{U}_3 - (\text{Pcov}(X, Y))^2 \right] + o_p(1) = n^{1/2} \sum_{i=1}^n N_i + o_p(1), \]

where the \( N_i \) are the centralized projections of the U-statistics \( \hat{U}_k \), which are defined as

\[ N_i = E \{ (K_1 + K_2 - 2K_3) \mid X_i, Y_i \} - \{ \text{Pcov}(X, Y) \}^2. \]

(A2)

All the \( N_i \) are independent and identically distributed. The second part of Theorem 3 can be proved with the classical central limit theorem.
REFERENCES

BENJAMINI, Y., MADAR, V. & STARK, P. B. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika* **100**, 503–10.

BLUM, J. R. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32**, 485–98.

CHANG, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18**, 391–404.

ESCANCIANO, J. C. (2006). A consistent diagnostic test for regression models using projections. *Economet. Theory* **22**, 1030–51.

GIESER, P. W. & RANDLE, R. H. (1997). A nonparametric test of independence between two vectors. *J. Am. Statist. Assoc.* **92**, 561–7.

HETTMANSPERGER, T. P. & OJA, H. (1994). Affine invariant multivariate multisample sign tests. *J. R. Statist. Soc., B* **56**, 235–49.

HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Statist.* **19**, 546–57.

HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28**, 321–77.

JENNRICH, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40**, 633–43.

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

KUO, H. H. (1975). *Gaussian Measures in Banach Spaces*. Lecture Notes in Mathematics 463. Berlin: Springer.

PURI, M. & SEN, P. (1971). *Nonparametric Methods in Multivariate Analysis*. New York: Wiley.

R DEVELOPMENT CORE TEAM (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

SERFLING, R. L. (1980). *Approximation Theorems in Mathematical Statistics*. New York: Wiley.

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

SZÉKELY, G. J. & RIZZO, M. L. (2010). Brownian distance covariance. *Ann. Appl. Statist.* **3**, 1236–65.

SZÉKELY, G. J. & RIZZO, M. L. (2013). The distance correlation $t$-test of independence in high dimension. *J. Mult. Anal.* **117**, 193–213.

SZÉKELY, G. J., RIZZO, M. L. & BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35**, 2769–94.

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

TASKINEN, S., OJA, H. & RANDLE, R. H. (2005). Multivariate nonparametric tests of independence. *J. Am. Statist. Assoc.* **100**, 916–25.

TRACZ, S. M., ELMORE, P. B. & POHLZEN, J. T. (1992). Correlational meta-analysis: Independent and nonindependent cases. *Educ. Psychol. Meas.* **52**, 879–88.

VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

WILKS, S. S. (1935). On the independence of $k$ sets of normally distributed statistical variables. *Econometrika* **3**, 309–26.

WOLFOWITZ, J. (1954). Generalization of the theorem of Glivenko–Cantelli. *Ann. Math. Statist.* **25**, 131–8.

[Received on 10 December 2016. Editorial decision on 25 June 2017]