

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

$H_0 : X$ and Y are independent versus $H_1 : \text{otherwise.}$

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 $|\Sigma_{X,Y}|/(|\Sigma_X| |\Sigma_Y|)$, where $\Sigma_{X,Y} = \text{var}\{(X^T, Y^T)^T\} \in \mathbb{R}^{(p+q) \times (p+q)}$, $\Sigma_X = \text{var}(X) \in \mathbb{R}^{p \times p}$ 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 Σ_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 $\text{dc}(X, Y)$, 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 $\text{pc}(X, Y)$, 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 α and β . 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 α and β , 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

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

Suppose that $\{(X_i, Y_i), i = 1, \dots, n\}$ 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

$$\begin{aligned} & \int \int \int E^2\{I(\alpha^T X_1 \leq u)I(\beta^T Y_1 \leq v) - I(\alpha^T X_1 \leq u)I(\beta^T Y_2 \leq v)\} dF_{U,V}(u, v) d\alpha d\beta \\ &= \int \int \int E\{[I(\alpha^T X_1 \leq \alpha^T X_3)I(\beta^T Y_1 \leq \beta^T Y_3) - I(\alpha^T X_1 \leq \alpha^T X_3)I(\beta^T Y_2 \leq \beta^T Y_3)] \\ & \quad \times [I(\alpha^T X_4 \leq \alpha^T X_3)I(\beta^T Y_4 \leq \beta^T Y_3) - I(\alpha^T X_4 \leq \alpha^T X_3)I(\beta^T Y_5 \leq \beta^T Y_3)]\} d\alpha d\beta. \end{aligned}$$

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

$$\begin{aligned} & E\left(\int \int \int [I(\alpha^T X_1 \leq \alpha^T X_3)I(\beta^T Y_1 \leq \beta^T Y_3) - I(\alpha^T X_1 \leq \alpha^T X_3)I(\beta^T Y_2 \leq \beta^T Y_3)] \right. \\ & \quad \left. [I(\alpha^T X_4 \leq \alpha^T X_3)I(\beta^T Y_4 \leq \beta^T Y_3) - I(\alpha^T X_4 \leq \alpha^T X_3)I(\beta^T Y_5 \leq \beta^T Y_3)] d\alpha d\beta\right) = 0. \end{aligned} \tag{2}$$

In general, integration over the $(p + q - 2)$ -dimensional space $\{(\alpha, \beta) : \|\alpha\| = \|\beta\| = 1\}$ 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

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

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

$$\begin{aligned} \{\text{Pcov}(X, Y)\}^2 &= S_1 + S_2 - 2S_3 \\ &= E \left[\arccos \left\{ \frac{(X_1 - X_3)^\top (X_4 - X_3)}{\|X_1 - X_3\| \|X_4 - X_3\|} \right\} \arccos \left\{ \frac{(Y_1 - Y_3)^\top (Y_4 - Y_3)}{\|Y_1 - Y_3\| \|Y_4 - Y_3\|} \right\} \right] \\ &\quad + E \left[\arccos \left\{ \frac{(X_1 - X_3)^\top (X_4 - X_3)}{\|X_1 - X_3\| \|X_4 - X_3\|} \right\} \arccos \left\{ \frac{(Y_2 - Y_3)^\top (Y_5 - Y_3)}{\|Y_2 - Y_3\| \|Y_5 - Y_3\|} \right\} \right] \\ &\quad - 2E \left[\arccos \left\{ \frac{(X_1 - X_3)^\top (X_4 - X_3)}{\|X_1 - X_3\| \|X_4 - X_3\|} \right\} \arccos \left\{ \frac{(Y_2 - Y_3)^\top (Y_4 - Y_3)}{\|Y_2 - Y_3\| \|Y_4 - Y_3\|} \right\} \right], \end{aligned} \tag{3}$$

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 $(X_k - X_l)/\|X_k - X_l\|$ and $(Y_k - Y_l)/\|Y_k - Y_l\|$, 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

$$\begin{aligned} \{\widehat{\text{Pcov}}_1(X, Y)\}^2 &= \widehat{S}_1 + \widehat{S}_2 - 2\widehat{S}_3 \\ &= n^{-3} \sum_{i,k,l=1}^n \left[\arccos \left\{ \frac{(X_i - X_k)^\top (X_l - X_k)}{\|X_i - X_k\| \|X_l - X_k\|} \right\} \arccos \left\{ \frac{(Y_i - Y_k)^\top (Y_l - Y_k)}{\|Y_i - Y_k\| \|Y_l - Y_k\|} \right\} \right] \\ &\quad + n^{-5} \sum_{i,j,k,l,r=1}^n \left[\arccos \left\{ \frac{(X_i - X_k)^\top (X_l - X_k)}{\|X_i - X_k\| \|X_l - X_k\|} \right\} \arccos \left\{ \frac{(Y_j - Y_k)^\top (Y_r - Y_k)}{\|Y_j - Y_k\| \|Y_r - Y_k\|} \right\} \right] \\ &\quad - 2n^{-4} \sum_{i,j,k,l=1}^n \left[\arccos \left\{ \frac{(X_i - X_k)^\top (X_l - X_k)}{\|X_i - X_k\| \|X_l - X_k\|} \right\} \arccos \left\{ \frac{(Y_j - Y_k)^\top (Y_l - Y_k)}{\|Y_j - Y_k\| \|Y_l - Y_k\|} \right\} \right]. \end{aligned}$$

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 (Székely & Rizzo, 2010). Therefore, we give an equivalent form below. Define, for $k, l, r = 1, \dots, n$,

$$\begin{aligned}
 a_{klr} &= \arccos \left\{ \frac{(X_k - X_r)^T (X_l - X_r)}{\|X_k - X_r\| \|X_l - X_r\|} \right\}, \\
 \bar{a}_{k \cdot r} &= n^{-1} \sum_{l=1}^n a_{klr}, \quad \bar{a}_{\cdot lr} = n^{-1} \sum_{k=1}^n a_{klr}, \quad \bar{a}_{\cdot \cdot r} = n^{-2} \sum_{k=1}^n \sum_{l=1}^n a_{klr}, \\
 A_{klr} &= a_{klr} - \bar{a}_{k \cdot r} - \bar{a}_{\cdot lr} + \bar{a}_{\cdot \cdot r}, \\
 b_{klr} &= \arccos \left\{ \frac{(Y_k - Y_r)^T (Y_l - Y_r)}{\|Y_k - Y_r\| \|Y_l - Y_r\|} \right\}, \\
 \bar{b}_{k \cdot r} &= n^{-1} \sum_{l=1}^n b_{klr}, \quad \bar{b}_{\cdot lr} = n^{-1} \sum_{k=1}^n b_{klr}, \quad \bar{b}_{\cdot \cdot r} = n^{-2} \sum_{k=1}^n \sum_{l=1}^n b_{klr}, \\
 B_{klr} &= b_{klr} - \bar{b}_{k \cdot r} - \bar{b}_{\cdot lr} + \bar{b}_{\cdot \cdot r},
 \end{aligned}$$

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

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

Accordingly, the sample estimate of $\text{pc}(X, Y)$ is defined by the square root of

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

In general, $\hat{\text{Pcov}}(X, Y)$ is easier to compute than $\hat{\text{Pcov}}_1(X, Y)$. Although it may not be immediately obvious that $\hat{\text{Pcov}}(X, Y) \geq 0$, this fact will become clear from Theorem 1.

THEOREM 1. For a given random sample $\{(X_i, Y_i), i = 1, \dots, n\}$,

$$\{\hat{\text{Pcov}}_1(X, Y)\}^2 = \{\hat{\text{Pcov}}(X, Y)\}^2,$$

and both equal

$$\{nc(p)c(q)\}^{-1} \sum_{i=1}^n \left[\int \int \{\hat{F}_{U,V}(\alpha^T X_i, \beta^T Y_i) - \hat{F}_U(\alpha^T X_i)\hat{F}_V(\beta^T Y_i)\}^2 d\alpha d\beta \right],$$

where $\hat{F}_{U,V}$, \hat{F}_U and \hat{F}_V 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 $\hat{\text{Pcov}}(X, Y)$ and $\hat{\text{Pcov}}_1(X, Y)$.

THEOREM 2. For a given random sample $\{(X_i, Y_i), i = 1, \dots, n\}$, $\lim_{n \rightarrow \infty} \hat{\text{Pcov}}(X, Y) = \text{Pcov}(X, Y)$ almost surely.

THEOREM 3.

- (i) If X and Y are independent, then as $n \rightarrow \infty$, $n \{ \text{P}\hat{\text{cov}}(X, Y) \}^2 / (\pi^2 - \hat{S}_2)$ converges in distribution to $\sum_{k=1}^{\infty} \lambda_k Z_k^2$, where the λ_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 $n^{1/2} \{ \text{P}\hat{\text{cov}}(X, Y) - \text{Pcov}(X, Y) \}$ 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, $n \{ \text{P}\hat{\text{cov}}(X, Y) \}^2 / (\pi^2 - \hat{S}_2)$ diverges to ∞ .

The projection correlation test is built upon the test statistic $T_n = n \{ \text{P}\hat{\text{cov}}(X, Y) \}^2 / (\pi^2 - \hat{S}_2)$, which converges in distribution to the quadratic form if X and Y are independent and diverges to ∞ otherwise. Theorem 3 suggests that the projection correlation test is consistent against all dependence alternatives without requiring any moment conditions. Because the weights λ_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. Throughout our simulations, we use 2000 replications and obtain very good control of the Type I error rates. 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.ttest` 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 $X = (X_1, \dots, X_p)^T$ independently from a Cauchy distribution. Let $Y_i = \exp(X_i)$ ($i = 1, \dots, m$) and draw Y_i ($i = m + 1, \dots, q$) independently from a standard normal distribution.

Table 1. Empirical size and power (%) of the tests based on projection correlation, ranks of distances and distance correlation for different m in Example 1 with $p = q$ and $(n, p) = (30, 20)$.

All numbers reported in this table are multiplied by 100

Scenario	α (%)	Test	$m = 0$	$m = 2$	$m = 4$	$m = 6$	$m = 8$	$m = 10$
(1a)	1	Projection correlation	1.1	28.4	53.4	70.0	80.1	88.1
		Ranks of distances	0.9	2.7	8.1	20.1	36.2	52.3
		Distance correlation	1.3	2.9	3.4	4.1	4.2	4.4
	5	Projection correlation	4.9	52.9	75.1	87.7	93.0	96.0
		Ranks of distances	4.9	11.2	25.1	42.2	60.4	75.2
		Distance correlation	6.1	5.3	5.4	6.3	6.4	6.1
(1b)	1	Projection correlation	1.2	20.8	41.7	61.5	76.3	84.8
		Ranks of distances	0.9	2.3	8.0	19.5	35.3	51.3
		Distance correlation	1.4	3.0	4.0	4.2	4.5	5.0
	5	Projection correlation	5.2	39.9	64.8	81.4	89.5	94.6
		Ranks of distances	5.2	8.3	21.1	38.8	60.4	76.1
		Distance correlation	4.5	4.8	5.9	5.3	5.9	6.6
(1c)	1	Projection correlation	1.0	65.1	98.8	100	100	100
		Ranks of distances	0.9	20.3	58.9	87.1	96.6	99.5
		Distance correlation	0.9	67.0	98.3	100	100	100
	5	Projection correlation	4.5	82.0	99.6	100	100	100
		Ranks of distances	4.7	37.7	79.6	96.2	99.1	100
		Distance correlation	5.1	82.3	99.5	100	100	100

(1b) This is identical to scenario (1a), except that $Y_i, i = m + 1, \dots, q$, are sampled independently from the Cauchy distribution.

(1c) This is identical to scenario (1a), except that X_i , for $i = 1, \dots, p$, are sampled independently from a standard normal distribution.

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, \dots, X_p independently from the uniform distribution on $[0, 2\pi]$. We generate $Y_i = X_1 + 2X_2 + \dots + iX_i + \{iX_1 + (i - 1)X_2 + \dots + X_i\}^2 + \varepsilon_i$ ($i = 1, \dots, m$), where the ε_i are generated from the Cauchy distribution, and generate Y_i ($i = m + 1, \dots, q$) 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.

Table 2. Power (%) of the tests based on projection correlation, ranks of distances and distance correlation for different p in Example 1 with $n = 30$ and $p = q$. All numbers reported in this table are multiplied by 100

Scenario	m	α (%)	Test	$p = 15$	$p = 20$	$p = 25$	$p = 30$
(1a)	10	1	Projection correlation	98.2	88.1	74.1	59.5
			Ranks of distances	81.6	52.3	35.6	24.2
			Distance correlation	7.9	4.4	4.6	2.7
	5	Projection correlation	99.8	96.0	89.4	79.3	
		Ranks of distances	93.7	75.2	60.6	46.8	
		Distance correlation	10.5	6.1	6.0	4.0	
(1b)	10	1	Projection correlation	97.0	84.8	70.7	54.7
			Ranks of distances	79.8	51.3	34.9	24.8
			Distance correlation	7.6	5.0	3.1	2.5
	5	Projection correlation	99.5	94.6	87.4	77.6	
		Ranks of distances	93.7	76.1	60.3	45.5	
		Distance correlation	9.7	6.6	4.5	3.7	
(1c)	2	1	Projection correlation	78.1	65.1	53.1	43.1
			Ranks of distances	31.7	20.3	13.7	9.2
			Distance correlation	79.5	67.0	54.1	43.4
	5	Projection correlation	89.7	82.0	72.2	63.3	
		Ranks of distances	54.2	37.7	30.4	24.6	
		Distance correlation	90.2	82.3	73.0	63.5	

Table 3. Empirical size and power (%) of the tests based on projection correlation, ranks of distances and distance correlation for different m in Example 2 with $p = q = 100$

n	α (%)	Test	$m = 0$	$m = 5$	$m = 15$	$m = 25$
10	1	Projection correlation	1.3	6.2	7.1	7.1
		Ranks of distances	1.2	1.7	2.2	1.9
		Distance correlation	1.3	6.2	6.5	5.7
	5	Projection correlation	5.4	18.8	20.7	20.4
		Ranks of distances	4.3	7.6	7.3	7.2
		Distance correlation	5.5	16.5	18.6	17.0
20	1	Projection correlation	1.2	21.5	26.3	28.2
		Ranks of distances	0.7	4.3	4.2	4.8
		Distance correlation	1.0	18.7	20.2	23.2
	5	Projection correlation	5.4	46.4	52.4	55.8
		Ranks of distances	4.4	13.4	13.1	12.5
		Distance correlation	5.2	40.5	43.3	45.2
30	1	Projection correlation	1.4	50.6	63.5	63.4
		Ranks of distances	0.9	8.2	9.3	9.5
		Distance correlation	1.5	42.4	51.3	51.5
	5	Projection correlation	5.1	78.4	87.0	85.9
		Ranks of distances	4.7	21.0	24.1	23.0
		Distance correlation	5.2	69.3	75.7	76.1

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.

Table 4. The power (%) of the tests based on projection correlation, ranks of distances and distance correlation for different p in Example 2 with $p = q$

n	α (%)	Test	$p = 30$	$p = 40$	$p = 50$	$p = 60$
10	1	Projection correlation	18.5	13.2	11.1	9.3
		Ranks of distances	4.1	2.7	2.6	2.6
		Distance correlation	14.9	11.1	8.5	7.5
	5	Projection correlation	42.8	34.4	28.4	26.2
		Ranks of distances	13.1	10.6	8.8	8.7
		Distance correlation	34.9	27.6	23.7	21.7
20	1	Projection correlation	79.7	67.7	53.6	47.0
		Ranks of distances	20.1	15.9	10.8	8.7
		Distance correlation	66.7	55.8	42.4	37.6
	5	Projection correlation	95.2	88.9	80.2	74.3
		Ranks of distances	39.6	31.8	24.5	22.3
		Distance correlation	86.9	79.5	69.0	62.2
30	1	Projection correlation	99.6	97.5	93.8	86.9
		Ranks of distances	49.3	33.9	25.3	19.9
		Distance correlation	97.6	92.1	85.6	75.9
	5	Projection correlation	100	99.9	99.4	97.4
		Ranks of distances	72.6	55.9	45.8	38.6
		Distance correlation	99.7	98.1	96.1	92.2

Table 5. The power (%) of the tests based on projection correlation, ranks of distances and distance correlation for different n in Example 3 with $p = q = 1000$

n	Test	Cauchy		$t(3)$	
		$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 1\%$	$\alpha = 5\%$
20	Projection correlation	17.1	34.0	100	100
	Ranks of distances	1.5	6.3	5.9	16.4
	Distance correlation	10.2	18.0	100	100
30	Projection correlation	36.7	58.7	100	100
	Ranks of distances	1.9	8.8	10.0	30.5
	Distance correlation	10.3	18.2	100	100
40	Projection correlation	59.5	81.6	100	100
	Ranks of distances	2.1	9.2	15.3	42.4
	Distance correlation	9.5	17.1	100	100

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 $i = 1, \dots, 10$, (X, Y) are random vectors $\{U_X(i) + \varepsilon_X, U_Y(i) + \varepsilon_Y\}$, where $U_X(i)$ and $U_Y(i)$ are sampled independently from a multivariate standard normal distribution, and $(\varepsilon_X, \varepsilon_Y)$ 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 $\{U_X(i), U_Y(i)\}$ ($i = 1, \dots, 10$), 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 $\{Pcov(X, Y)\}^2$ has an explicit form. In other words, we aim to show that

$$\begin{aligned} \{Pcov(X, Y)\}^2 &= \{c(p)c(q)\}^{-1} \int_{\|\alpha\|=1} \int_{\|\beta\|=1} E\left\{F_{U,V}(\alpha^T X_3, \beta^T Y_3) - F_U(\alpha^T X_3)F_V(\beta^T Y_3)\right\}^2 d\alpha d\beta \\ &= E\left[\arccos\left\{\frac{(X_1 - X_3)^T(X_4 - X_3)}{\|X_1 - X_3\|\|X_4 - X_3\|}\right\} \arccos\left\{\frac{(Y_1 - Y_3)^T(Y_4 - Y_3)}{\|Y_1 - Y_3\|\|Y_4 - Y_3\|}\right\}\right] \\ &\quad + E\left[\arccos\left\{\frac{(X_1 - X_3)^T(X_4 - X_3)}{\|X_1 - X_3\|\|X_4 - X_3\|}\right\} \arccos\left\{\frac{(Y_2 - Y_3)^T(Y_5 - Y_3)}{\|Y_2 - Y_3\|\|Y_5 - Y_3\|}\right\}\right] \\ &\quad - 2E\left[\arccos\left\{\frac{(X_1 - X_3)^T(X_4 - X_3)}{\|X_1 - X_3\|\|X_4 - X_3\|}\right\} \arccos\left\{\frac{(Y_2 - Y_3)^T(Y_4 - Y_3)}{\|Y_2 - Y_3\|\|Y_4 - Y_3\|}\right\}\right]. \end{aligned}$$

For notational clarity, we define

$$a_{klr} = \arccos\left\{\frac{(X_k - X_r)^T(X_l - X_r)}{\|X_k - X_r\|\|X_l - X_r\|}\right\}, \quad b_{klr} = \arccos\left\{\frac{(Y_k - Y_r)^T(Y_l - Y_r)}{\|Y_k - Y_r\|\|Y_l - Y_r\|}\right\}.$$

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

$$\begin{aligned} &\{c(p)c(q)\}^{-1} \int_{\|\alpha\|=1} \int_{\|\beta\|=1} E\{F_{U,V}(\alpha^T X_3, \beta^T Y_3) - F_U(\alpha^T X_3)F_V(\beta^T Y_3)\}^2 d\alpha d\beta \\ &= E\{(\pi - a_{135})(\pi - b_{135})\} - E\{(\pi - a_{135})(\pi - b_{145})\} - E\{(\pi - a_{135})(\pi - b_{235})\} \\ &\quad + E\{(\pi - a_{135})(\pi - b_{245})\} \\ &= E(a_{135}b_{135}) + E(a_{135}b_{245}) - 2E(a_{135}b_{145}) \\ &= S_1 + S_2 - 2S_3. \end{aligned}$$

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

$$\begin{aligned} \bar{a}_{k..r} &= n^{-1} \sum_{l=1}^n a_{klr}, & \bar{a}_{.lr} &= n^{-1} \sum_{k=1}^n a_{klr}, & \bar{a}_{..r} &= n^{-2} \sum_{k=1}^n \sum_{l=1}^n a_{klr}, & \bar{b}_{..r} &= n^{-2} \sum_{k=1}^n \sum_{l=1}^n b_{klr}, \\ A_{klr} &= a_{klr} - \bar{a}_{k..r} - \bar{a}_{.lr} + \bar{a}_{..r}, & B_{klr} &= b_{klr} - \bar{b}_{k..r} - \bar{b}_{.lr} + \bar{b}_{..r}. \end{aligned}$$

We further define

$$\begin{aligned} I_1 &= n^{-3} \sum_{i,j,r=1}^n a_{ijr} b_{ijr}, & I_2 &= n^{-3} \sum_{i,j,r=1}^n a_{ijr} \bar{b}_{i..r}, & I_3 &= n^{-3} \sum_{i,j,r=1}^n a_{ijr} \bar{b}_{j..r}, \\ I_4 &= n^{-1} \sum_{r=1}^n \bar{a}_{..r} \bar{b}_{..r}, & I_5 &= n^{-3} \sum_{i,j,r=1}^n \bar{a}_{i..r} b_{ijr}, & I_6 &= n^{-2} \sum_{i,r=1}^n \bar{a}_{i..r} \bar{b}_{i..r}, \\ I_7 &= n^{-3} \sum_{i,j,r=1}^n \bar{a}_{i..r} \bar{b}_{j..r}, & I_8 &= n^{-2} \sum_{i,r=1}^n \bar{a}_{i..r} \bar{b}_{..r}, & I_9 &= n^{-3} \sum_{i,j,r=1}^n \bar{a}_{j..r} b_{ijr}, \\ I_{10} &= n^{-3} \sum_{i,j,r=1}^n \bar{a}_{j..r} \bar{b}_{i..r}, & I_{11} &= n^{-2} \sum_{j,r=1}^n \bar{a}_{j..r} \bar{b}_{j..r}, & I_{12} &= n^{-2} \sum_{j,r=1}^n \bar{a}_{j..r} \bar{b}_{..r}, \\ I_{13} &= n^{-1} \sum_{r=1}^n \bar{a}_{..r} \bar{b}_{..r}, & I_{14} &= n^{-2} \sum_{i,r=1}^n \bar{a}_{..r} \bar{b}_{i..r}, & I_{15} &= n^{-2} \sum_{j,r=1}^n \bar{a}_{..r} \bar{b}_{j..r}, \\ I_{16} &= n^{-1} \sum_{r=1}^n \bar{a}_{..r} \bar{b}_{..r}. \end{aligned}$$

It can be verified that $I_2 = I_3 = I_5 = I_6 = I_9 = I_{11}$ and $I_4 = I_7 = I_8 = I_{10} = I_{12} = I_{13} = I_{14} = I_{15} = I_{16}$. It follows that

$$\begin{aligned} \{\hat{\text{Pcov}}(X, Y)\}^2 &= n^{-3} \sum_{k,l,r=1}^n A_{klr} B_{klr} \\ &= n^{-3} \sum_{k,l,r=1}^n (a_{klr} - \bar{a}_{k..r} - \bar{a}_{.lr} + \bar{a}_{..r})(b_{klr} - \bar{b}_{k..r} - \bar{b}_{.lr} + \bar{b}_{..r}) \\ &= I_1 - I_2 - I_3 + I_4 - I_5 + I_6 + I_7 - I_8 - I_9 + I_{10} + I_{11} - I_{12} + I_{13} - I_{14} \\ &\quad - I_{15} + I_{16} \\ &= I_1 - 2I_2 + I_4 = \hat{S}_1 + \hat{S}_2 - 2\hat{S}_3 = \{\hat{\text{Pcov}}_1(X, Y)\}^2, \end{aligned}$$

which completes the proof of the first part.

Next we prove that $\{\widehat{\text{Pcov}}(X, Y)\}^2$ is equal to

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

Invoking Lemma 1, we have

$$\begin{aligned} & \int \left\{ I(\alpha^T X_i \leq \alpha^T X_r) - n^{-1} \sum_{i=1}^n I(\alpha^T X_i \leq \alpha^T X_r) \right\} \\ & \quad \times \left\{ I(\alpha^T X_j \leq \alpha^T X_r) - n^{-1} \sum_{i=1}^n I(\alpha^T X_i \leq \alpha^T X_r) \right\} d\alpha \\ & = -c(p)(a_{ijr} - \bar{a}_{i.r} - \bar{a}_{.jr} + \bar{a}_{..r}) = -c(p)A_{ijr}. \end{aligned}$$

Following similar arguments, we obtain

$$\begin{aligned} & \int \left\{ I(\beta^T Y_i \leq \beta^T Y_r) - n^{-1} \sum_{i=1}^n I(\beta^T Y_i \leq \beta^T Y_r) \right\} \\ & \quad \times \left\{ I(\beta^T Y_j \leq \beta^T Y_r) - n^{-1} \sum_{i=1}^n I(\beta^T Y_i \leq \beta^T Y_r) \right\} d\beta \\ & = -c(q)(b_{ijr} - \bar{b}_{i.r} - \bar{b}_{.jr} + \bar{b}_{..r}) = -c(q)B_{ijr}. \end{aligned}$$

The above two results yield

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

The proof of Theorem 1 is complete.

Proof of Theorem 2

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

Proof of Theorem 3

Define the empirical process

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

where $\alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^q, u \in \mathbb{R}$ 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

$$\begin{aligned} & E\{R(\alpha_1, \beta_1, u_1, v_1)R(\alpha_2, \beta_2, u_2, v_2)\} \\ & = E\{I(\alpha_1^T X \leq u_1, \alpha_2^T X \leq u_2)\}E\{I(\beta_1^T Y \leq v_1, \beta_2^T Y \leq v_2)\} - F_U(u_1)F_U(u_2)F_V(v_1)F_V(v_2). \end{aligned}$$

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 (α, β, u, v) with $\|\alpha\| = 1$ and $\|\beta\| = 1$. It is easy to verify that

$$\begin{aligned} &W_n(\alpha, \beta, u, v) - R_n(\alpha, \beta, u, v) \\ &= -n^{-1} \sum_{i=1}^n \{I(\alpha^T X_i \leq u - F_U(u))\} n^{-1/2} \sum_{i=1}^n \{I(\beta^T Y_i \leq v - F_V(v))\}. \end{aligned}$$

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 (α, u) with $\|\alpha\| = 1$. Using Theorem 2.5.2 in [van der Vaart & Wellner \(1996\)](#), we can show that $n^{-1/2} \sum_{i=1}^n \{I(\beta^T Y_i \leq v) - F_V(v)\}$ converges to a Gaussian process with zero mean and variance-covariance function $E \{I(\beta_1^T Y \leq v_1, \beta_2^T Y \leq v_2) - F_V(v_1)F_V(v_2)\}$. 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

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

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 \(1975, Ch. 1, § 2\)](#),

$$\{c(p)c(q)\}^{-1} \int R^2(\alpha, \beta, u, v) F_{U,V}(du, dv) d\alpha d\beta \tag{A1}$$

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 λ_j^* depend on the distribution of (X, Y) .

Next we derive the sum of the λ_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 λ_j^* . If X and Y are independent, then

$$\begin{aligned} &E \left\{ \int R^2(\alpha, \beta, u, v) F_{U,V}(du, dv) d\alpha d\beta \right\} \\ &= \int_{\|\alpha\|=1} \int_{\|\beta\|=1} E \{F_U(\alpha^T X_r) F_V(\beta^T Y_r) - F_U^2(\alpha^T X_r) F_V^2(\beta^T Y_r)\} d\alpha d\beta \\ &= E \left[E \left\{ \int_{\|\alpha\|=1} I(\alpha^T X_i \leq \alpha^T X_r) d\alpha \mid X_r \right\} E \left\{ \int_{\|\beta\|=1} I(\beta^T Y_i \leq \beta^T Y_r) d\beta \mid Y_r \right\} \right] \end{aligned}$$

$$- E \left\{ \int_{\|\alpha\|=1} I(\alpha^T X_i \leq \alpha^T X_r, \alpha^T X_j \leq \alpha^T X_r) d\alpha \mid X_r \right\} \\ \times E \left\{ \int_{\|\beta\|=1} I(\beta^T Y_i \leq \beta^T Y_r, \beta^T Y_j \leq \beta^T Y_r) d\beta \mid Y_r \right\}.$$

Using Lemma 1, the right-hand side of the above equation is equal to

$$\pi^2 - E \left[\arccos \left\{ \frac{(X_1 - X_5)^T (X_3 - X_5)}{\|X_1 - X_5\| \|X_3 - X_5\|} \right\} \arccos \left\{ \frac{(Y_1 - Y_5)^T (Y_3 - Y_5)}{\|Y_1 - Y_5\| \|Y_3 - Y_5\|} \right\} \right] = \pi^2 - S_2.$$

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 \hat{S}_k with the U -statistics \hat{U}_k , 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 $K_1(X_i, Y_i; X_j, Y_j; X_k, Y_k) = 6^{-1} \sum_3 a_{i_1 i_2 i_3} b_{i_1 i_2 i_3} = k_1$, 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 $K_2(X_i, Y_i; X_j, Y_j; X_k, Y_k; X_l, Y_l; X_r, Y_r) = 120^{-1} \sum_5 a_{i_1 i_2 i_3} b_{i_3 i_4 i_5} = K_2$, 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 $K_3(X_i, Y_i; X_j, Y_j; X_k, Y_k; X_l, Y_l) = 24^{-1} \sum_4 a_{i_1 i_2 i_4} b_{i_1 i_3 i_4} = k_3$, where \sum_4 is the permutation of three distinct elements (i, j, k, l) . Using standard V - and U -statistic theory, we have

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

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

$$N_i = E\{(K_1 + K_2 - 2K_3) \mid X_i, Y_i\} - \{\text{Pcov}(X, Y)\}^2. \tag{A2}$$

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. & RANDLES, R. H. (1997). A nonparametric test of independence between two vectors. *J. Am. Statist. Assoc.* **92**, 561–7.
- HETTMANSBERGER, 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. & RANDLES, R. H. (2005). Multivariate nonparametric tests of independence. *J. Am. Statist. Assoc.* **100**, 916–25.
- TRACZ, S. M., ELMORE, P. B. & POHLMANN, 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. *Econometrica* **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]