Statistica Sinica **32** (2022), 417-433
doi:https://doi.org/10.5705/ss.202019.0457

POWER ANALYSIS OF PROJECTION-PURSUIT INDEPENDENCE TESTS

Kai Xu[1] and Liping Zhu[2,3]

[1]*Anhui Normal University,* [2]*Renmin University of China and* [3]*Zhejiang Gongshang University*

Abstract: Three important projection-pursuit correlations, namely, the distance, projection, and multivariate Blum–Kiefer–Rosenblatt (BKR) correlations, have been proposed in the literature to test for independence between two random vectors in arbitrary dimensions. In this study, we compare the asymptotic power performance of independence tests built upon these three projection-pursuit correlations, in a uniform sense. We show that in the presence of outliers, the projection and multivariate BKR correlation tests are still powerful, whereas the distance correlation test may lose power. We also analyze the minimax optimality of these independence tests. We show that their minimum separation rates are of order n^{-1}, where n stands for the sample size, and that this minimax optimal rate is tight in terms of the projection, distance, and multivariate BKR correlations.

Key words and phrases: Distance correlation, independence test, minimax optimality, projection correlation, power function, robustness.

1. Introduction

Many important applications require quantifying the degree of nonlinear dependence between two random vectors. For example, in genomics research, one may be interested in testing whether certain diseases are associated with mutations of a particular group of genes. In economic studies, one may wish to evaluate the nonlinear dependence between the stock market and real estate returns. In brain sciences, one may expect to discover whether two sets of voxels measured over time in different parts of brain are functionally related. We formulate these applications into problems of testing independence. Let $\mathbf{x} = (X_1, \ldots, X_p)^{\mathrm{T}} \in \mathbb{R}^p$ and $\mathbf{y} = (Y_1, \ldots, Y_q)^{\mathrm{T}} \in \mathbb{R}^q$ be two random vectors. We assume throughout that $p > 1$ and $q > 1$, unless stated otherwise. The goal of an independence test is to test

$$H_0 : \mathbf{x} \text{ and } \mathbf{y} \text{ are statistically independent; } H_1 : \mathbf{x} \text{ and } \mathbf{y} \text{ are dependent.} \quad (1.1)$$

Corresponding author: Liping Zhu, Center for Applied Statistics and Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, P. R. China. E-mail: zhu.liping@ruc.edu.cn.

Testing for independence has a long history in the literature. The Pearson correlation is perhaps the first and one of the most important metrics to test for independence between two univariate random variables (i.e., $p = q = 1$). Extensions within the univariate case include, but are not limited to those of Hoeffding (1948), Blum, Kiefer and Rosenblatt (1961), and Bergsma and Dassios (2014). These extensions are based on ranks of observations and, thus, cannot be used if either $\mathbf{x}$ or $\mathbf{y}$ is multivariate (i.e., $p > 1$ or $q > 1$). In the multivariate case, where $\mathbf{x}$ and $\mathbf{y}$ both follow jointly normal or elliptically symmetric distributions, testing for independence amounts to testing whether they are linearly uncorrelated (Oja (2010)). Important examples along this line include the likelihood ratio test (Wilks (1935)) and canonical correlation coefficient (Hotelling (1936)). Interested readers may refer to Puri and Sen (1971), Hettmansperger and Oja (1994), and Taskinen, Oja and Randles (2005) for extensions of the likelihood ratio test.

In the past two decades, numerous efforts have been made to relax the distributional assumptions; see, for example, Kankainen (1995) and Bakirov, Rizzo and Szekely (2006). Gretton et al. (2005) proposed an independence criterion based on the entire eigen-spectrum of covariance operators in reproducing kernel Hilbert spaces. Székely, Rizzo and Bakirov (2007) and Székely and Rizzo (2009) made important advances by proposing using a distance correlation to test for independence between two random vectors in arbitrary dimensions. A distance correlation is well defined by assuming the first moments of both $\mathbf{x}$ and $\mathbf{y}$ are finite, and is generalized by Sejdinovic et al. (2013), Pan et al. (2019), and Shen et al. (2019), from different perspectives. Heller, Heller and Corfine (2013) pointed out that if the moment conditions are violated, for example, if the underlying distribution of either $\mathbf{x}$ or $\mathbf{y}$ is heavy tailed or the observations contain outliers, the distance correlation test may suffer from low power. Given that outlying observations arise frequently in practice with high-dimensional data, it is highly desirable to develop robust alternatives to using a distance correlation. To this end, Zhu et al. (2017) proposed a projection correlation that removes the moment conditions required by a distance correlation. The projection correlation is, in spirit, a multivariate version of Hoeffding (1948). Kim, Balakrishnan and Wasserman (2018) suggested a projection-averaging approach to classic two-sample test problems, stating that their approach can be readily generalized to test for independence between two random vectors. We follow Kim, Balakrishnan and Wasserman (2018) by extending the Blum–Kiefer–Rosenblatt (BKR) correlation to the multivariate case. Neither the projection correlation nor the multivariate BKR correlation requires a moment condition on either $\mathbf{x}$ or $\mathbf{y}$. We show that the distance correlation and the projection correlation are both based on the integrated squared distance between the joint distribution of the projections and the product of their marginal distributions over unit spheres. The independence tests built upon the distance, projection, and multivariate BKR correlations are all of the projection-pursuit type.

We compare the power performance of the aforementioned three projection-pursuit independence tests because they share many similarities. In particular, projection, distance, and multivariate BKR correlations have closed-form expressions and require no tuning parameters, and all tests are consistent against all fixed alternatives. More importantly, all three tests can be represented by integrals of the distance between the joint distribution function of $(\mathbf{x}, \mathbf{y})$ and the product of the marginal distribution functions of $\mathbf{x}$ and $\mathbf{y}$. They differ only in the weights. To elaborate, we define $\mathcal{S}^{d-1} \stackrel{\text{def}}{=} \{\boldsymbol{\alpha} \in \mathbb{R}^d : \|\boldsymbol{\alpha}\| = 1\}$, where $\|\cdot\|$ is the Euclidean norm. Then, $F_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}}(s) \stackrel{\text{def}}{=} \mathrm{pr}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x} \le s)$, $F_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}}(t) \stackrel{\text{def}}{=} \mathrm{pr}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y} \le t)$, and $F_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}}(s,t) \stackrel{\text{def}}{=} \mathrm{pr}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x} \le s, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y} \le t)$, for $\boldsymbol{\alpha} \in \mathcal{S}^{p-1}$, $\boldsymbol{\beta} \in \mathcal{S}^{q-1}$, $s \in \mathbb{R}^1$, and $t \in \mathbb{R}^1$. Here, $(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ and $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})$ are the respective projections of $\mathbf{x}$ and $\mathbf{y}$. In the Supplementary Material, we show that the squared distance covariance can be represented as

$$\begin{aligned}\mathrm{DC}(\mathbf{x}, \mathbf{y}) = (c_p c_q)^{-1} \int_{\boldsymbol{\alpha} \in \mathcal{S}^{p-1}} \int_{\boldsymbol{\beta} \in \mathcal{S}^{q-1}} \int_{t \in \mathbb{R}^1} \int_{s \in \mathbb{R}^1} \\ \left\{F_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}}(s,t) - F_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}}(s) F_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}}(t)\right\}^2 (ds\ dt) d\boldsymbol{\beta} d\boldsymbol{\alpha},\end{aligned} \tag{1.2}$$

and the squared projection covariance can be represented as

$$\begin{aligned}\mathrm{PC}(\mathbf{x}, \mathbf{y}) = (\gamma_p \gamma_q)^{-1} \int_{\boldsymbol{\alpha} \in \mathcal{S}^{p-1}} \int_{\boldsymbol{\beta} \in \mathcal{S}^{q-1}} \int_{t \in \mathbb{R}^1} \int_{s \in \mathbb{R}^1} \\ \left\{F_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}}(s,t) - F_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}}(s) F_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}}(t)\right\}^2 dF_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}}(s,t) d\boldsymbol{\beta} d\boldsymbol{\alpha}.\end{aligned} \tag{1.3}$$

Kim, Balakrishnan and Wasserman (2018) wrote the multivariate BKR correlation coefficient as

$$\begin{aligned}\mathrm{mBKR}(\mathbf{x}, \mathbf{y}) = (\gamma_p \gamma_q)^{-1} \int_{\boldsymbol{\alpha} \in \mathcal{S}^{p-1}} \int_{\boldsymbol{\beta} \in \mathcal{S}^{q-1}} \int_{t \in \mathbb{R}^1} \int_{s \in \mathbb{R}^1} \\ \left\{F_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x},\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}}(s,t) - F_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}}(s) F_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}}(t)\right\}^2 dF_{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}}(s) dF_{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}}(t) d\boldsymbol{\beta} d\boldsymbol{\alpha}.\end{aligned} \tag{1.4}$$

In the above three expressions, $c_p \stackrel{\text{def}}{=} \{2\pi^{(p-1)/2}/(p-1)\}/\Gamma\{(p-1)/2\}$, $\gamma_p \stackrel{\text{def}}{=} \pi^{p/2-1}/\Gamma(p/2)$, and $\Gamma(\cdot)$ is a gamma function. These expressions differ in terms of how they average over s and t. In particular, in (1.2), the uniform weights are given on the $\mathbb{R}^1 \otimes \mathbb{R}^1$ space, and in (1.3) and (1.4), more weight is given on higher density regions. It is thus anticipated that the projection correlation test and the multivariate BKR correlation test are more robust to extreme observations than is the distance correlation test. The projection correlation uses the joint density of $(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ and $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{y})$ as a weight function, whereas the multivariate BKR correlation uses the product of their marginal densities.

The asymptotic null distributions of the above projection-pursuit independence tests depend on the joint distribution of $\mathbf{x}$ and $\mathbf{y}$, which are, in general, unknown in practice. To approximate the asymptotic null distributions, random permutations are widely used in these independence tests. However, few studies have examined the consistency of random permutations. In the present context, we show that the permutation procedure provides a reasonable approximation of the asymptotic null distributions, without exhausting all possible permutations. As a by-product, this allows us to carry out a power analysis of the projection-pursuit independence tests. We show that in the presence of outliers, the permutation test based on either the projection correlation or the multivariate BKR correlation is very powerful, while that based on the distance correlation may lose power. To gain more insight into their asymptotic behaviors, we analyze the minimax optimality of these projection-pursuit independence tests over a wide class of distributions using Le Cam's lemma (Baraud (2002)). We show that their minimum separation rates are all of order n^{-1}, where n stands for the sample size. The minimum separation rate is a lower bound that characterizes the separation boundary between the testable and non-testable regions. Furthermore, the rate n^{-1} is tight in terms of the projection, distance, and multivariate BKR correlations.

2. Some Preliminaries

2.1. The computational complexities

We provide explicit forms for (1.2), (1.3), and (1.4) first. Suppose $\{(\mathbf{x}_i, \mathbf{y}_i)$, for $i = 1, \ldots, 6\}$, are six independent copies of $(\mathbf{x}, \mathbf{y})$. Let $\mathbf{z}$ be either $\mathbf{x}$ or $\mathbf{y}$. We define $a(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \mathbf{z}_5) \stackrel{\text{def}}{=} \operatorname{ang}(\mathbf{z}_1 - \mathbf{z}_5, \mathbf{z}_2 - \mathbf{z}_5) + \operatorname{ang}(\mathbf{z}_3 - \mathbf{z}_5, \mathbf{z}_4 - \mathbf{z}_5) - \operatorname{ang}(\mathbf{z}_1 - \mathbf{z}_5, \mathbf{z}_3 - \mathbf{z}_5) - \operatorname{ang}(\mathbf{z}_2 - \mathbf{z}_5, \mathbf{z}_4 - \mathbf{z}_5)$, where $\operatorname{ang}(\mathbf{a}, \mathbf{b}) \stackrel{\text{def}}{=} \arccos\{(\mathbf{a}^{\mathrm{T}}\mathbf{b})/(\|\mathbf{a}\|\|\mathbf{b}\|)\}$ stands for the angle between the two vectors $\mathbf{a}$ and $\mathbf{b}$, and $\arccos(\cdot)$ is the inverse cosine function. If $\mathbf{z}_i$, $\mathbf{z}_j$, and $\mathbf{z}_k$ are all distinctive, $\operatorname{ang}(\mathbf{z}_i - \mathbf{z}_k, \mathbf{z}_j - \mathbf{z}_k)$ is well defined and ranges from zero to π. Following Escanciano (2006) and Zhu et al. (2017), we define $\operatorname{ang}(\mathbf{z}_i - \mathbf{z}_k, \mathbf{z}_j - \mathbf{z}_k) = 0$ if $\mathbf{z}_i = \mathbf{z}_j \neq \mathbf{z}_k$, $\mathbf{z}_i = \mathbf{z}_k \neq \mathbf{z}_j$, or $\mathbf{z}_j = \mathbf{z}_k \neq \mathbf{z}_i$, and $\operatorname{ang}(\mathbf{z}_i - \mathbf{z}_k, \mathbf{z}_j - \mathbf{z}_k) = -\pi$ if $\mathbf{z}_i = \mathbf{z}_j = \mathbf{z}_k$. We further define $b(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4) \stackrel{\text{def}}{=} \|\mathbf{z}_1 - \mathbf{z}_2\| + \|\mathbf{z}_3 - \mathbf{z}_4\| - \|\mathbf{z}_1 - \mathbf{z}_3\| - \|\mathbf{z}_2 - \mathbf{z}_4\|$. Székely, Rizzo and Bakirov (2007) and Székely and Rizzo (2009) showed that

$\mathrm{DC}(\mathbf{x},\mathbf{y}) = E\left\{b(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4)b(\mathbf{y}_1,\mathbf{y}_2,\mathbf{y}_3,\mathbf{y}_4)\right\}/4$. By Theorem 1 of Zhu et al. (2017), the explicit form of the projection correlation is given by $\mathrm{PC}(\mathbf{x},\mathbf{y}) = E\{a(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4,\mathbf{x}_5)a(\mathbf{y}_1,\mathbf{y}_2,\mathbf{y}_3,\mathbf{y}_4,\mathbf{y}_5)\}/4$. Kim, Balakrishnan and Wasserman (2018, Theorem 7.2) derived that the multivariate BKR correlation has the form of $\mathrm{mBKR}(\mathbf{x},\mathbf{y}) = E\{a(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4,\mathbf{x}_5)a(\mathbf{y}_1,\mathbf{y}_2,\mathbf{y}_3,\mathbf{y}_4,\mathbf{y}_6)\}/4$. With a random sample of size n, say, $\{(\mathbf{x}_i,\mathbf{y}_i), i=1,\ldots,n\}$, we estimate $\mathrm{DC}(\mathbf{x},\mathbf{y})$, $\mathrm{PC}(\mathbf{x},\mathbf{y})$, and $\mathrm{mBKR}(\mathbf{x},\mathbf{y})$ using U-statistic theory. In particular,

$$\widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y}) \stackrel{\text{def}}{=} \{4(n)_4\}^{-1}\sum_{(i,j,k,l)}^{n} b(\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_k,\mathbf{x}_l)b(\mathbf{y}_i,\mathbf{y}_j,\mathbf{y}_k,\mathbf{y}_l),$$
$$\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) \stackrel{\text{def}}{=} \{4(n)_5\}^{-1}\sum_{(i,j,k,l,r)}^{n} a(\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_k,\mathbf{x}_l,\mathbf{x}_r)a(\mathbf{y}_i,\mathbf{y}_j,\mathbf{y}_k,\mathbf{y}_l,\mathbf{y}_r),$$

and

$$\widehat{\mathrm{mBKR}}(\mathbf{x},\mathbf{y}) \stackrel{\text{def}}{=} \{4(n)_6\}^{-1}\sum_{(i,j,k,l,r,s)}^{n} a(\mathbf{x}_i,\mathbf{x}_j,\mathbf{x}_k,\mathbf{x}_l,\mathbf{x}_r)a(\mathbf{y}_i,\mathbf{y}_j,\mathbf{y}_k,\mathbf{y}_l,\mathbf{y}_s),$$

where $(n)_m \stackrel{\text{def}}{=} n(n-1)\cdots(n-m+1)$. The summations

$$\sum_{(i,j,k,l)}^{n}, \quad \sum_{(i,j,k,l,r)}^{n}, \quad \text{and} \quad \sum_{(i,j,k,l,r,s)}^{n}$$

are taken over different indexes.

Next, we compare the computational complexity of calculating $\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})$, $\widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y})$, and $\widehat{\mathrm{mBKR}}(\mathbf{x},\mathbf{y})$. The sample distance covariance is a U-statistic of order four, the sample projection covariance is a U-statistic of order five, and the sample multivariate BKR correlation is a U-statistic of order six. Székely and Rizzo (2013) and Yao, Zhang and Shao (2018) stated that

$$\begin{aligned}\widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y}) = \{n(n-3)\}^{-1}\Big[&\mathrm{tr}(\widetilde{\mathbf{A}}\widetilde{\mathbf{B}}) \\ &+\{(n-1)_2\}^{-1}\mathbf{1}_n^{\mathrm{T}}\widetilde{\mathbf{A}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\widetilde{\mathbf{B}}\mathbf{1}_n - 2(n-2)^{-1}\mathbf{1}_n^{\mathrm{T}}\widetilde{\mathbf{A}}\widetilde{\mathbf{B}}\mathbf{1}_n\Big],\end{aligned} \tag{2.1}$$

where $\mathbf{1}_n \in \mathbb{R}^n$ is a vector of ones, $\widetilde{\mathbf{A}} = (\|\mathbf{x}_i-\mathbf{x}_j\|)_{n\times n} \in \mathbb{R}^{n\times n}$, and $\widetilde{\mathbf{B}} = (\|\mathbf{y}_i-\mathbf{y}_j\|)_{n\times n} \in \mathbb{R}^{n\times n}$. That is, the computational complexity of $\widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y})$ is of order $O(n^2)$. To calculate $\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})$ and $\widehat{\mathrm{mBKR}}(\mathbf{x},\mathbf{y})$, we define $\mathbf{A}_k \stackrel{\text{def}}{=} (a_{ijk}) \in \mathbb{R}^{(n-1)\times(n-1)}$ and $\mathbf{B}_k \stackrel{\text{def}}{=} (b_{ijk}) \in \mathbb{R}^{(n-1)\times(n-1)}$, where $a_{ijk} \stackrel{\text{def}}{=} \mathrm{ang}(\mathbf{x}_i-\mathbf{x}_k,\mathbf{x}_j-\mathbf{x}_k)$ and, $b_{ijk} \stackrel{\text{def}}{=} \mathrm{ang}(\mathbf{y}_i-\mathbf{y}_k,\mathbf{y}_j-\mathbf{y}_k)$, for $i\neq k$, $j\neq k$, and $k=1,\ldots,n$. With some straightforward algebraic calculations, it can be verified that

$$\begin{aligned}
\sum_{(i,j,k)}^{n} a_{ijk}b_{ijk} &= \sum_{k=1}^{n} \mathrm{tr}(\mathbf{A}_k\mathbf{B}_k), \\
\sum_{(i,j,k,l)}^{n} a_{ijl}b_{ikl} &= \sum_{l=1}^{n} \{\mathbf{1}_{(n-1)}^{\mathrm{T}}\mathbf{A}_l\mathbf{B}_l\mathbf{1}_{(n-1)} - \mathrm{tr}(\mathbf{A}_l\mathbf{B}_l)\}, \\
\sum_{(i,j,k,l,r)}^{n} a_{ijr}b_{klr} &= \sum_{r=1}^{n} \{\mathbf{1}_{(n-1)}^{\mathrm{T}}\mathbf{A}_r\mathbf{1}_{(n-1)}\mathbf{1}_{(n-1)}^{\mathrm{T}}\mathbf{B}_r\mathbf{1}_{(n-1)} \\
&\quad -4\mathbf{1}_{(n-1)}^{\mathrm{T}}\mathbf{A}_r\mathbf{B}_r\mathbf{1}_{(n-1)} + 2\mathrm{tr}(\mathbf{A}_r\mathbf{B}_r)\}.
\end{aligned}$$

Collecting these results, we have

$$\begin{aligned}
\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) = \{n(n-1)(n-4)\}^{-1}\sum_{r=1}^{n}\Big[&\mathrm{tr}(\mathbf{A}_r\mathbf{B}_r) \\
&+\{(n-2)_2\}^{-1}\mathbf{1}_{(n-1)}^{\mathrm{T}}\mathbf{A}_r\mathbf{1}_{(n-1)}\mathbf{1}_{(n-1)}^{\mathrm{T}}\mathbf{B}_r\mathbf{1}_{(n-1)} \\
&-2(n-3)^{-1}\mathbf{1}_{(n-1)}^{\mathrm{T}}\mathbf{A}_r\mathbf{B}_r\mathbf{1}_{(n-1)}\Big].
\end{aligned} \tag{2.2}$$

Thus, the computational complexity of $\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})$ is of order $O(n^3)$. Similarly, we can verify that

$$\begin{aligned}
\widehat{\mathrm{mBKR}}(\mathbf{x},\mathbf{y}) = \{n(n-1)(n-2)(n-5)\}^{-1}\sum_{r\neq s}^{n}\Big[&\mathrm{tr}(\mathbf{A}_r\mathbf{B}_s) \\
&+\{(n-3)_2\}^{-1}\mathbf{1}_{(n-1)}^{\mathrm{T}}\mathbf{A}_r\mathbf{1}_{(n-1)}\mathbf{1}_{(n-1)}^{\mathrm{T}}\mathbf{B}_s\mathbf{1}_{(n-1)} \\
&-2(n-4)^{-1}\mathbf{1}_{(n-1)}^{\mathrm{T}}\mathbf{A}_r\mathbf{B}_s\mathbf{1}_{(n-1)}\Big],
\end{aligned} \tag{2.3}$$

indicating that estimating the multivariate BKR correlation requires $O(n^4)$ operations. Calculating the distance correlation has the smallest complexity.

2.2. The permutation procedure

Zhu et al. (2017) and Székely, Rizzo and Bakirov (2007) showed that the U-statistic estimates $\widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y})$ and $\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})$ are n-consistent under H_0 and root-n-consistent under fixed alternatives, respectively. Consequently, $n\ \widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y})$ and $n\ \widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})$ converge in distribution to their respective nondegenerate limits under H_0, and diverge to infinity under fixed alternatives. Following Zhu et al. (2017), we establish the distribution theory for $\widehat{\mathrm{mBKR}}$ under both the null and the alternative hypotheses. Specifically, $\widehat{\mathrm{mBKR}}$ is n-consistent under

H_0 and root-n-consistent under fixed alternatives. Therefore, we reject H_0 when $n\,\widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y})$, $n\,\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})$, and $n\,\widehat{\mathrm{mBKR}}$ are greater than or equal to certain critical values. However, the asymptotic null distributions of $n\,\widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y})$, $n\,\widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})$, and $n\,\widehat{\mathrm{mBKR}}$ are not tractable when $p>1$ or $q>1$. To address this issue, Zhu et al. (2017) and Székely, Rizzo and Bakirov (2007) suggested approximating the critical values adaptively using the following random permutation approach:

1. Suppose $\{i_1,i_2,\ldots,i_n\}$ and $\{j_1,j_2,\ldots,j_n\}$ are two random permutations of $\{1,2,\ldots,n\}$. Define $\mathbf{x}_k^b \stackrel{\text{def}}{=} \mathbf{x}_{i_k}$ and $\mathbf{y}_k^b \stackrel{\text{def}}{=} \mathbf{y}_{j_k}$, for $k=1,\ldots,n$. Re-estimate $\mathrm{DC}(\mathbf{x},\mathbf{y})$, $\mathrm{PC}(\mathbf{x},\mathbf{y})$, and $\mathrm{mBKR}(\mathbf{x},\mathbf{y})$ using $\{(\mathbf{x}_k^b,\mathbf{y}_k^b),k=1,\ldots,n\}$. Denote the resulting estimates by $\widehat{\mathrm{DC}}(\mathbf{x}^b,\mathbf{y}^b)$, $\widehat{\mathrm{PC}}(\mathbf{x}^b,\mathbf{y}^b)$, and $\widehat{\mathrm{mBKR}}(\mathbf{x}^b,\mathbf{y}^b)$, respectively. Replicate this permutation procedure B times, say, $B=1{,}000$, to approximate the asymptotic null distributions of $\widehat{\mathrm{DC}}(\mathbf{x}^b,\mathbf{y}^b)$, $\widehat{\mathrm{PC}}(\mathbf{x}^b,\mathbf{y}^b)$, and $\widehat{\mathrm{mBKR}}(\mathbf{x}^b,\mathbf{y}^b)$.

2. Denote the observations $\mathcal{D}_n \stackrel{\text{def}}{=} \{(\mathbf{x}_i,\mathbf{y}_i),i=1,\ldots,n\}$. We define the critical values at the significance level α by

$$q_{\alpha,n}^{DC} \stackrel{\text{def}}{=} \inf\left[t\in\mathbb{R}:1-\alpha\le \mathrm{pr}\{n\,\widehat{\mathrm{DC}}(\mathbf{x}^b,\mathbf{y}^b)\le t\mid \mathcal{D}_n\}\right], \tag{2.4}$$

$$q_{\alpha,n}^{PC} \stackrel{\text{def}}{=} \inf\left[t\in\mathbb{R}:1-\alpha\le \mathrm{pr}\{n\,\widehat{\mathrm{PC}}(\mathbf{x}^b,\mathbf{y}^b)\le t\mid \mathcal{D}_n\}\right], \tag{2.5}$$

$$q_{\alpha,n}^{mBKR} \stackrel{\text{def}}{=} \inf\left[t\in\mathbb{R}:1-\alpha\le \mathrm{pr}\{n\,\widehat{\mathrm{mBKR}}(\mathbf{x}^b,\mathbf{y}^b)\le t\mid \mathcal{D}_n\}\right]. \tag{2.6}$$

We approximate $\mathrm{pr}\{n\,\widehat{\mathrm{DC}}(\mathbf{x}^b,\mathbf{y}^b)\le t\mid\mathcal{D}_n\}$, $\mathrm{pr}\{n\,\widehat{\mathrm{PC}}(\mathbf{x}^b,\mathbf{y}^b)\le t\mid\mathcal{D}_n\}$, and $\mathrm{pr}\{n\,\widehat{\mathrm{mBKR}}(\mathbf{x}^b,\mathbf{y}^b)\le t\mid\mathcal{D}_n\}$ using empirical probabilities

$$B^{-1}\sum_{b=1}^{B} I\left\{n\,\widehat{\mathrm{DC}}(\mathbf{x}^b,\mathbf{y}^b)\le t\right\},\quad B^{-1}\sum_{b=1}^{B} I\left\{n\,\widehat{\mathrm{PC}}(\mathbf{x}^b,\mathbf{y}^b)\le t\right\},$$

and

$$B^{-1}\sum_{b=1}^{B} I\left\{n\,\widehat{\mathrm{mBKR}}(\mathbf{x}^b,\mathbf{y}^b)\le t\right\},$$

respectively. This, in spirit, approximates the asymptotic null distributions of $n\,\widehat{\mathrm{DC}}(\mathbf{x}^b,\mathbf{y}^b)$, $n\,\widehat{\mathrm{PC}}(\mathbf{x}^b,\mathbf{y}^b)$, and $n\,\widehat{\mathrm{mBKR}}(\mathbf{x}^b,\mathbf{y}^b)$, respectively.

This random permutation procedure is intuitively valid, and thus widely used in multiple testing problems and independence tests. A random permutation procedure is said to be consistent if it provides a reasonable approximation to the asymptotic null distribution. The consistency of random permutations has been studied extensively by Romano and Wolf (2005) in the context of multiple testing problems. However, its consistency is rarely discussed in the context of independence tests. In Theorem 1, we show that this permutation procedure is consistent in all three independence tests. Detailed proofs are given in the Supplementary Material. Throughout, $\mathrm{pr}(\cdot \mid H_0)$ and $\mathrm{pr}(\cdot \mid H_1)$ stand for the respective probabilities that a random event occurs under H_0 and H_1. They are not conditional probabilities.

Theorem 1. *As $n \to \infty$, both*

$$\sup_{t\in\mathbb{R}} \left| \mathrm{pr}\{n\ \widehat{\mathrm{PC}}(\mathbf{x}^b, \mathbf{y}^b) \le t \mid \mathcal{D}_n\} - \mathrm{pr}\{n\ \widehat{\mathrm{PC}}(\mathbf{x}, \mathbf{y}) \le t \mid H_0\} \right|$$

and

$$\sup_{t\in\mathbb{R}} \left| \mathrm{pr}\{n\ \widehat{\mathrm{mBKR}}(\mathbf{x}^b, \mathbf{y}^b) \le t \mid \mathcal{D}_n\} - \mathrm{pr}\{n\ \widehat{\mathrm{mBKR}}(\mathbf{x}, \mathbf{y}) \le t \mid H_0\} \right|$$

converge in probability to zero. If we assume $E(\|\mathbf{x}\|^2) + E(\|\mathbf{y}\|^2) < \infty$, then

$$\sup_{t\in\mathbb{R}} \left| \mathrm{pr}\{n\ \widehat{\mathrm{DC}}(\mathbf{x}^b, \mathbf{y}^b) \le t \mid \mathcal{D}_n\} - \mathrm{pr}\{n\ \widehat{\mathrm{DC}}(\mathbf{x}, \mathbf{y}) \le t \mid H_0\} \right|$$

converges in probability to zero as $n \to \infty$.

We require the condition $E(\|\mathbf{x}\|^2) + E(\|\mathbf{y}\|^2) < \infty$ to ensure that the kernel of the U-statistic estimate $\widehat{\mathrm{DC}}(\mathbf{x}, \mathbf{y})$ is uniformly integrable. Theorem 1 guarantees that this random permutation procedure approximates the asymptotic null distributions precisely, as long as the sample size n is sufficiently large. In other words, the type-I error rates of all projection-pursuit independence tests are asymptotically controllable. This allows us to analyze the statistical power of these tests.

Exhausting all possible permutations is usually computationally prohibitive and practically infeasible. Therefore, we provide a random approximation in the above permutation procedure. Proposition 1 states that, as long as the number of random permutations, B, is sufficiently large, the random approximation is asymptotically valid.

Proposition 1. *Given the data $\mathcal{D}_n$,*

$$\sup_{t\in\mathbb{R}} \left| B^{-1} \sum_{b=1}^{B} I\left\{n\ \widehat{\mathrm{PC}}(\mathbf{x}^b, \mathbf{y}^b) \le t\right\} - \mathrm{pr}\{n\ \widehat{\mathrm{PC}}(\mathbf{x}^b, \mathbf{y}^b) \le t \mid \mathcal{D}_n\} \right|,$$

$$\sup_{t\in\mathbb{R}} \left| B^{-1} \sum_{b=1}^{B} I\left\{n\ \widehat{\mathrm{mBKR}}(\mathbf{x}^b, \mathbf{y}^b) \le t\right\} - \mathrm{pr}\{n\ \widehat{\mathrm{mBKR}}(\mathbf{x}^b, \mathbf{y}^b) \le t \mid \mathcal{D}_n\} \right|,$$

and

$$\sup_{t\in\mathbb{R}}\left|B^{-1}\sum_{b=1}^{B} I\left\{n\ \widehat{\mathrm{DC}}(\mathbf{x}^b,\mathbf{y}^b)\leq t\right\} - \mathrm{pr}\{n\ \widehat{\mathrm{DC}}(\mathbf{x}^b,\mathbf{y}^b)\leq t \mid \mathcal{D}_n\}\right|$$

converge in probability to zero, as $B\to\infty$.

3. Robustness Study

We first highlight the robustness of the projection correlation test and the multivariate BKR correlation test in a Huber contamination model. The following is an ϵ-contamination model:

$$(\mathbf{x},\mathbf{y})\sim F_{\mathbf{x},\mathbf{y}} = (1-\epsilon)F^{(1)}_{\mathbf{x},\mathbf{y}} + \epsilon H^{(n)}_{\mathbf{x},\mathbf{y}}, \tag{3.1}$$

where $F^{(1)}_{\mathbf{x},\mathbf{y}}$ and $H^{(n)}_{\mathbf{x},\mathbf{y}}$ are two distributional functions, with $F^{(1)}_{\mathbf{x},\mathbf{y}}$ fixed and $H^{(n)}_{\mathbf{x},\mathbf{y}}$ allowed to vary with n, and $0<\epsilon<1$. Note that $\mathbf{x}$ and $\mathbf{y}$ are dependent if $(\mathbf{x},\mathbf{y})\sim F^{(1)}_{\mathbf{x},\mathbf{y}}$, and independent if $(\mathbf{x},\mathbf{y})\sim H^{(n)}_{\mathbf{x},\mathbf{y}}$. We use the ϵ-contamination model (3.1) to evaluate whether an independence test can maintain adequate power when $H^{(n)}_{\mathbf{x},\mathbf{y}}$ has an adverse impact on its power performance. The test functions using the distance correlation, projection correlation, and multivariate BKR correlations are defined as

$$\begin{aligned}\Phi^{DC}_{\alpha} &\stackrel{\text{def}}{=} I\{n\ \widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y})\geq q^{DC}_{\alpha,n}\},\quad \Phi^{PC}_{\alpha}\stackrel{\text{def}}{=} I\{n\ \widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})\geq q^{PC}_{\alpha,n}\},\\ \Phi^{mBKR}_{\alpha} &\stackrel{\text{def}}{=} I\{n\ \widehat{\mathrm{mBKR}}(\mathbf{x},\mathbf{y})\geq q^{mBKR}_{\alpha,n}\},\end{aligned}$$

respectively, where $q^{DC}_{\alpha,n}$, $q^{PC}_{\alpha,n}$, and $q^{mBKR}_{\alpha,n}$ are the critical values defined in (2.4), (2.5), and (2.6), respectively, using random permutations, and $I(A)$ is an indicator function, which equals one if A is true, and zero otherwise. For all three projection-pursuit independence tests, we reject H_0 at the significance level α when the estimates of the projection-pursuit correlations are larger than their critical values, that is, when $n\ \widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y})\geq q^{DC}_{\alpha,n}$, $n\ \widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y})\geq q^{PC}_{\alpha,n}$ and $n\ \widehat{\mathrm{mBKR}}(\mathbf{x},\mathbf{y})\geq q^{mBKR}_{\alpha,n}$. We study the robustness of the projection-pursuit independence tests by comparing their power performance, because Theorem 1 ensures that one can always use random permutations to control the type-I error rate.

Theorem 2 states that the independence tests built on the projection correlation and the multivariate BKR correlation are uniformly powerful over different types of contaminations. In contrast, the distance correlation test becomes asymptotically powerless against certain contaminations.

Theorem 2. *Suppose* $\{(\mathbf{x}_i, \mathbf{y}_i), i = 1, \ldots, n\}$ *are generated independently from model (3.1) with the contamination ratio* $\epsilon = cn^{-1/2}$*, where* c *is a small positive constant not depending on* n*, and there exist three positive constants,* ϖ*,* ϖ'*, and* ϖ''*, such that* $\mathrm{PC}(\mathbf{x}, \mathbf{y}) \geq \varpi$*,* $\mathrm{DC}(\mathbf{x}, \mathbf{y}) \geq \varpi'$*, and* $\mathrm{mBKR}(\mathbf{x}, \mathbf{y}) \geq \varpi''$*, for sufficiently large* n*.*

1. *The projection correlation test and multivariate BKR correlation test are asymptotically powerful uniformly over* $H^{(n)}_{\mathbf{x},\mathbf{y}}$ *in the sense that*

$$\lim_{n\to\infty} \inf_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}(\Phi^{PC}_{\alpha} = 1 \mid H_1) = 1 \textit{ and } \lim_{n\to\infty} \inf_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}(\Phi^{mBKR}_{\alpha} = 1 \mid H_1) = 1.$$

2. *Assume* $E(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2) < \infty$ *if* $(\mathbf{x}, \mathbf{y}) \sim F^{(1)}_{\mathbf{x},\mathbf{y}}$*, and if* $(\mathbf{x}, \mathbf{y}) \sim H^{(n)}_{\mathbf{x},\mathbf{y}}$*,* $n\{\mathrm{var}(\|\mathbf{x}\|)\ \mathrm{var}(\|\mathbf{y}\|)\}^{-1/2} = o(1)$*. The distance correlation test is asymptotically powerless against such choices of* $H^{(n)}_{\mathbf{x},\mathbf{y}}$ *in the sense of*

$$\lim_{n\to\infty} \inf_{H^{(n)}_{\mathbf{x},\mathbf{y}}} \mathrm{pr}(\Phi^{DC}_{\alpha} = 1 \mid H_1) \leq \alpha.$$

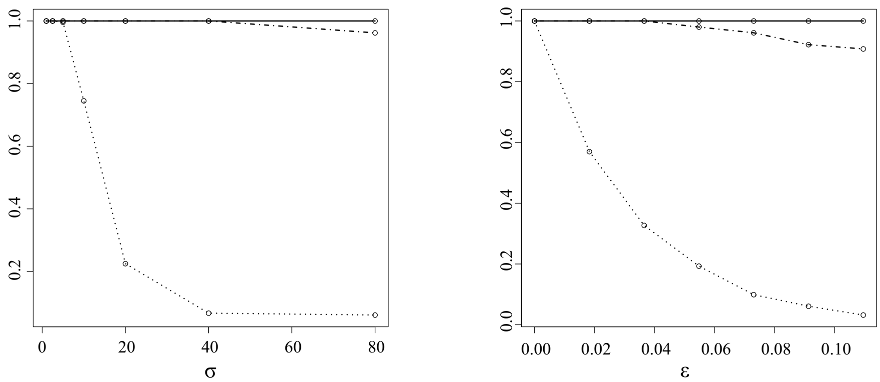
The first assertion of Theorem 2 implies that the projection and multivariate BKR correlation tests are insensitive to the presence of outliers. In the second statement of Theorem 2, we assume $n\{\mathrm{var}(\|\mathbf{x}\|)\ \mathrm{var}(\|\mathbf{y}\|)\}^{-1/2} = o(1)$ if $(\mathbf{x}, \mathbf{y}) \sim H^{(n)}_{\mathbf{x},\mathbf{y}}$, which allows $\mathrm{var}(\|\mathbf{x}\|)$ and $\mathrm{var}(\|\mathbf{y}\|)$ to be divergent, and accordingly, model (3.1) to yield outliers. We impose this condition to demonstrate that the distance correlation test might lose power in the presence of outliers.

We conduct simulations to illustrate Theorem 2 with a finite sample size. Following Davison and Hinkley (1997), we set $B = 1{,}000$ throughout our numerical studies.

Example 1. In the ϵ-contamination model (3.1), we consider an extreme case for $F^{(1)}_{\mathbf{x},\mathbf{y}}$: $\mathbf{x}$ follows a multivariate standard normal distribution, and $\mathbf{y}$ is equal to $\mathbf{x}$. This ensures that $\mathbf{x}$ and $\mathbf{y}$ are dependent. In other words, the observations are drawn under H_1. In addition, we set

$$H^{(n)}_{\mathbf{x},\mathbf{y}} = (2\pi\sigma^2)^{-p/2} \exp\left\{ -\frac{(\mathbf{x}^{\mathrm{T}}\mathbf{x})^2}{2\sigma^2} \right\} \prod_{k=1}^{p} I(0 \leq Y_k \leq 1).$$

We consider two scenarios for (ϵ, σ). In the first scenario, $\epsilon = 0.5n^{-1/2}$ and $\sigma = \{1, 2.5, 5, 10, 20, 40, 80\}$. In the second scenario, $\sigma = 100$ and $\epsilon = cn^{-1/2}$, for $c = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$. Both σ and ϵ control the degree of heavy-tailedness. As σ and c increase, the distance between H_0 and H_1 becomes smaller,



(a) $\epsilon = 0.5n^{-1/2}$ and $\sigma = \{1, 2.5, 5, 10, 20, 40, 80\}$

(b) $\sigma = 100$ and $\epsilon = cn^{-1/2}$, $c = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$

Figure 1. The empirical power of projection correlation test (solid line), distance correlation test (dotted line), and multivariate BKR correlation test (dotdash line) when the random sample is drawn from the ϵ-contamination model (3.1) with different ϵ and σ.

and the probabilities of observing extreme values from $H^{(n)}_{\mathbf{x},\mathbf{y}}$ increase as well. We fix $p = q = 10$ and $n = 30$, and decide the critical values with permutations at the significance level $\alpha = 0.05$. The simulations are replicated 1,000 times. The empirical power of the projection, distance and multivariate BKR correlation tests is summarized in Figure 1. It can be clearly seen that the empirical power of the projection and multivariate BKR correlation tests is very close to one throughout, indicating that these two tests are consistently robust to changes of σ and ϵ. In contrast, the empirical power of the distance correlation test drops quickly as σ and ϵ increase. The distance correlation test is powerless when σ or ϵ is sufficiently large.

Example 2. In the ϵ-contamination model (3.1), we set $H^{(n)}_{\mathbf{x},\mathbf{y}}$ to be the product of $(p+q)$ independent t-distributions with one degree of freedom, and set $F^{(1)}_{\mathbf{x},\mathbf{y}}$ to be the Dirac measure of the form $F^{(1)}_{\mathbf{x},\mathbf{y}} = I(\mathbf{x} = \kappa \mathbf{1}_p) I(\mathbf{y} = \mathbf{x})$, for $\kappa = \{5, 15\}$, where $\mathbf{1}_p$ is a p-vector of ones. Let $\epsilon = cn^{-1/2}$, for $c = \{0.2, 0.4, 0.6, 0.8, 1.0\}$. As c increases, the probabilities of observing extreme values from $H^{(n)}_{\mathbf{x},\mathbf{y}}$ increase as well, which, as stated in Theorem 2, may affect the power performance of the independence tests. Let $p = q = \{5, 10, 20\}$ and $n = 30$. The significance level is set to $\alpha = 0.05$.

The empirical power for each test is summarized in Tables 1 and 2, based on 1,000 replications. Following the suggestion of an anonymous reviewer, we

Table 1. The empirical power of the projection correlation test ("PC"), distance correlation test ("DC"), multivariate BKR correlation test ("mBKR"), and distance correlation t-test ("SR") in Example 2, with three different settings of dimension when $\kappa = 5$ and the nominal level is 0.05.

		$c = 1.0$	$c = 0.8$	$c = 0.6$	$c = 0.4$	$c = 0.2$
$p = 5$	PC	0.152	0.391	0.635	0.808	0.907
	DC	0.056	0.069	0.104	0.158	0.219
	mBKR	0.114	0.303	0.534	0.725	0.841
	SR	0.059	0.102	0.247	0.386	0.457
$p = 10$	PC	0.198	0.458	0.701	0.851	0.946
	DC	0.051	0.063	0.083	0.123	0.160
	mBKR	0.164	0.397	0.616	0.798	0.884
	SR	0.058	0.143	0.295	0.479	0.542
$p = 20$	PC	0.231	0.544	0.777	0.897	0.955
	DC	0.053	0.061	0.078	0.089	0.109
	mBKR	0.182	0.410	0.647	0.841	0.916
	SR	0.074	0.221	0.346	0.552	0.703

also include the distance correlation-based t-test (Székely and Rizzo (2013)) in our comparison. We denote this test by SR. The SR test is asymptotically distribution-free. Therefore, we use its asymptotic null distribution directly to decide the critical values. We expect that the projection correlation test and the multivariate BKR correlation test to be significantly more powerful than the distance correlation test and the distance correlation t-test across all scenarios. When c decreases from 1 to 0.2, p and q increase from 5 to 20, or κ increases from 5 to 15, the deviation from H_0 is accumulating. The power of the projection and multivariate BKR correlation tests increases significantly. In contrast, the distance correlation test loses power completely when $\kappa = 5$. Because the SR test was developed specifically for large dimensions, it is more powerful than the distance correlation test, especially when $p = 20$. However, the SR test is still significantly inferior to the projection and multivariate BKR correlation tests in terms of power performance, particularly when p and c are relatively small.

4. Minimax Optimality

Next, we study the minimax optimality of the three tests. To simplify subsequent illustration, let Φ_α be a level-α test function, equal to one if we reject H_0, and zero otherwise. Denote by $\mathrm{pr}(\cdot \mid H_0)$ and $\mathrm{pr}(\cdot \mid H_1)$ the probabilities evaluated under H_0 and H_1, respectively. Accordingly, $\mathrm{pr}(\Phi_\alpha = 1 \mid H_0)$ is the type-I error rate, and $\mathrm{pr}(\Phi_\alpha = 0 \mid H_1)$ is the type-II error rate. We define the class of

Table 2. The empirical power of the projection correlation test ("PC"), distance correlation test ("DC"), multivariate BKR test ("mBKR"), and distance correlation t-test ("SR") in Example 2, with three different settings of dimension when $\kappa = 15$ and the nominal level is 0.05.

		$c = 1.0$	$c = 0.8$	$c = 0.6$	$c = 0.4$	$c = 0.2$
$p = 5$	PC	0.216	0.508	0.749	0.876	0.948
	DC	0.112	0.239	0.394	0.512	0.620
	mBKR	0.195	0.482	0.711	0.804	0.896
	SR	0.132	0.384	0.508	0.615	0.732
$p = 10$	PC	0.272	0.592	0.791	0.913	0.971
	DC	0.082	0.177	0.282	0.401	0.512
	mBKR	0.266	0.514	0.750	0.875	0.914
	SR	0.190	0.455	0.682	0.796	0.889
$p = 20$	PC	0.318	0.655	0.852	0.940	0.977
	DC	0.061	0.110	0.188	0.264	0.354
	mBKR	0.287	0.568	0.796	0.918	0.962
	SR	0.242	0.544	0.751	0.885	0.951

level-α test functions by $\mathcal{T}_\alpha \stackrel{\text{def}}{=} \{\Phi_\alpha : \text{pr}(\Phi_\alpha = 1 \mid H_0) \le \alpha\}$. We measure the dependence between $\mathbf{x}$ and $\mathbf{y}$ using a projection correlation, distance correlation, and multivariate BKR correlation. Define

$$\begin{aligned}
\mathcal{U}^{PC}(c) &\stackrel{\text{def}}{=} \{(\mathbf{x}, \mathbf{y}) : \text{PC}(\mathbf{x}, \mathbf{y}) \ge cn^{-1}\}, \\
\mathcal{U}^{DC}(c) &\stackrel{\text{def}}{=} \{(\mathbf{x}, \mathbf{y}) : \text{DC}(\mathbf{x}, \mathbf{y}) \ge cn^{-1}\}, \\
\mathcal{U}^{mBKR}(c) &\stackrel{\text{def}}{=} \{(\mathbf{x}, \mathbf{y}) : \text{mBKR}(\mathbf{x}, \mathbf{y}) \ge cn^{-1}\}.
\end{aligned}$$

If the degree of dependence between $\mathbf{x}$ and $\mathbf{y}$ is weak, it may be difficult to distinguish between H_0 and H_1. Theorem 3 states that, for all level-α tests, there exist $(\mathbf{x}, \mathbf{y}) \in \mathcal{U}^{PC}(c_0)$ for the projection correlation test, $(\mathbf{x}, \mathbf{y}) \in \mathcal{U}^{DC}(c_0)$ for the distance correlation test, and $(\mathbf{x}, \mathbf{y}) \in \mathcal{U}^{mBKR}(c_0)$ for the multivariate BKR correlation test, such that their type-II error rates, $\text{pr}(\Phi_\alpha = 0 \mid H_1)$, are not asymptotically negligible, even when $n \to \infty$. The specified constant c_0 quantifies the degree of deviation from H_0.

Theorem 3. *For any $0 < \xi < 1 - \alpha$, there exists $c_0 > 0$ such that the minimax type-II error rates are lower bounded as $n \to \infty$; that is,*

$$\begin{aligned}
\lim_{n\to\infty} \inf_{\Phi_\alpha \in \mathcal{T}_\alpha} \sup_{(\mathbf{x},\mathbf{y}) \in \mathcal{U}^{PC}(c_0)} \text{pr}(\Phi_\alpha^{PC} = 0 \mid H_1) &\ge \xi, \\
\lim_{n\to\infty} \inf_{\Phi_\alpha \in \mathcal{T}_\alpha} \sup_{(\mathbf{x},\mathbf{y}) \in \mathcal{U}^{mBKR}(c_0)} \text{pr}(\Phi_\alpha^{mBKR} = 0 \mid H_1) &\ge \xi,
\end{aligned}$$

$$\lim_{n\to\infty} \inf_{\Phi_\alpha \in \mathcal{T}_\alpha} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{DC}(c_0)} \mathrm{pr}(\Phi_\alpha^{DC} = 0 \mid H_1) \ \geq \xi.$$

Theorem 3 indicates that the projection-pursuit independence tests cannot maintain adequate power, even if the dependence between $\mathbf{x}$ and $\mathbf{y}$ is cn^{-1} apart in terms of $\mathrm{PC}(\mathbf{x},\mathbf{y})$, $\mathrm{mBKR}(\mathbf{x},\mathbf{y})$, or $\mathrm{DC}(\mathbf{x},\mathbf{y})$, for an arbitrarily small c. However, if we allow c to diverge to infinity, the type-II error rates of these independence tests shrink to zero as $n \to \infty$. This is formulated in Theorem 4. Define

$$\Phi_\alpha^{DC} \overset{\text{def}}{=} I\big\{n\ \widehat{\mathrm{DC}}(\mathbf{x},\mathbf{y}) \geq q_{\alpha,n}^{DC}\big\}, \quad \Phi_\alpha^{PC} \overset{\text{def}}{=} I\big\{n\ \widehat{\mathrm{PC}}(\mathbf{x},\mathbf{y}) \geq q_{\alpha,n}^{PC}\big\}$$
$$\Phi_\alpha^{mBKR} \overset{\text{def}}{=} I\big\{n\ \widehat{\mathrm{mBKR}}(\mathbf{x},\mathbf{y}) \geq q_{\alpha,n}^{mBKR}\big\},$$

where $q_{\alpha,n}^{DC}$, $q_{\alpha,n}^{PC}$, and $q_{\alpha,n}^{mBKR}$ are defined in (2.4), (2.5), and (2.6), respectively.

Theorem 4. *The minimax type-II error rate of the projection correlation test tends to zero uniformly over $\mathcal{U}^{PC}(c_n)$, with $c_n \to \infty$ as $n \to \infty$; that is,*

$$\lim_{n\to\infty} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{PC}(c_n)} \mathrm{pr}(\Phi_\alpha^{PC} = 0 \mid H_1) = 0.$$

The minimax type-II error rate of the multivariate BKR correlation test tends to zero uniformly over $\mathcal{U}^{mBKR}(c_n)$, with $c_n \to \infty$ as $n \to \infty$; that is,

$$\lim_{n\to\infty} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{mBKR}(c_n)} \mathrm{pr}(\Phi_\alpha^{mBKR} = 0 \mid H_1) = 0.$$

Furthermore, if $\|\mathbf{x}\|$ and $\|\mathbf{y}\|$ are squared integrable, the minimax type-II error rate of the distance correlation test tends to zero uniformly over $\mathcal{U}^{DC}(c_n)$, with $c_n \to \infty$ as $n \to \infty$; that is,

$$\lim_{n\to\infty} \sup_{(\mathbf{x},\mathbf{y})\in\mathcal{U}^{DC}(c_n)} \mathrm{pr}(\Phi_\alpha^{DC} = 0 \mid H_1) = 0.$$

Theorem 4, together with Theorem 3, indicates that the minimax lower bound of the minimum separation rate is n^{-1}. This lower bound is asymptotically tight for the projection and multivariate BKR correlation tests. If $\|\mathbf{x}\|$ and $\|\mathbf{y}\|$ are squared integrable, this lower bound is also asymptotically tight for the distance correlation test.

5. Discussion

We consider three projection-pursuit correlations, namely, the distance, projection, and multivariate BKR correlations. These correlations quantify the difference between the joint distribution function and the product of the marginal distribution functions. These three projection-pursuit correlations differ only in their weight functions. We investigate their robustness, and compare the power performance of independence tests built upon these correlations under a minimax framework. We also seek the conditions under which the projection-pursuit independence tests are minimax rate optimal.

It is practically interesting, yet theoretically challenging to characterize the exact value of c in $\mathcal{U}^{DC}(c)$, $\mathcal{U}^{PC}(c)$, and $\mathcal{U}^{mBKR}(c)$ that separates the testable region from the non-testable one. This is because the class of alternatives we are targeting is huge, owing to the existence of nonlinear dependence. This issue is beyond the scope of this study, and is left to future research.

Supplementary Material

The online Supplementary Material contains the proofs of (1.2), Proposition 1, and Theorems 1–4.

Acknowledgments

Xu's research was supported by the National Natural Science Foundation of China (11901006) and the Natural Science Foundation of Anhui Province (1908085QA06). Zhu is the corresponding author, and his research was supported by the Natural Science Foundation of Beijing (Z19J00009) and the National Natural Science Foundation of China (11731011, 11931014).

References

Bakirov, N., Rizzo, M. and Szekely, G. (2006). A multivariate nonparametric test of independence. *Journal of Multivariate Analysis* **97**, 1742–1756.

Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8**, 577–606.

Bergsma, W. and Dassios, A. (2014). A consistent test of independence based on a sign covariance related to kendall's tau. *Bernoulli* **20**, 1006–1028.

Blum, J., Kiefer, J. and Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics* **32**, 485–498.

Davison, A. and Hinkley, D. (1997). *Bootstrap Methods and Their Application.* Cambridge University Press, Cambridge.

Escanciano, J. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22**, 1030–1051.

Gretton, A., Bousquet, O., Smola, A. and Schölkopf, B. (2005). Measuring statistical dependence with hilbert-schmidt norms. In *International Conference on Algorithmic Learning Theory*, 63–77.

Heller, R., Heller, Y. and Corfine, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika* **100**, 503–510.

Hettmansperger, T. and Oja, H. (1994). Affine invariant multivariate multisample sign tests. *Journal of the Royal Statistical Society, Series B (Methodological)* **56**, 235–249.

Hoeffding, W. (1948). A non-parametric test of independence. *The Annals of Mathematical Statistics* **19**, 546–557.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* **28**, 321–377.

Kankainen, A. (1995). *Consistent Testing of Total Independence based on the Empirical Characteristic Function*. University of Jyväskylä, Jyväskylä.

Kim, I., Balakrishnan, S. and Wasserman, L. (2018). Robust multivariate nonparametric tests via projection-pursuit. https://arxiv.org/abs/1803.00715.

Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. Springer-Verlag, New York.

Pan, W., Wang, X., Zhang, H., Zhu, H. and Zhu, J. (2019). Ball covariance: A generic measure of dependence in banach space. *Journal of the American Statistical Association* **115**, 307–317.

Puri, M. and Sen, P. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

Romano, J. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association* **100**, 94–108.

Sejdinovic, D., Sriperumbudur, B., Gretton, A. and Fukumizu, K. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics* **41**, 2263–2291.

Shen, C., Priebe, C., Gretton, A. and Vogelstein, J. (2019). From distance correlation to multiscale graph correlation. *Journal of the American Statistical Association* **115**, 280–291.

Székely, G. and Rizzo, M. (2009). Brownian distance covariance. *The Annals of Applied Statistics* **3**, 1236–1265.

Székely, G. and Rizzo, M. (2013). The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis* **117**, 193–213.

Székely, G., Rizzo, M. and Bakirov, N. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35**, 2769–2794.

Taskinen, S., Oja, H. and Randles, R. (2005). Multivariate nonparametric tests of independence. *Journal of the American Statistical Association* **471**, 916–925.

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

Yao, S., Zhang, X. Y. and Shao, X. F. (2018). Testing mutual independence in high dimension via distance covariance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**, 455–480.

Zhu, L., Xu, K., Li, R. and Zhong, W. (2017). Projection correlation between two random vectors. *Biometrika* **104**, 829–843.

Kai Xu

School of Mathematics and Statistics, Anhui Normal University, Wuhu 241002, China.

E-mail: tjxxukai@163.com

Liping Zhu

Center for Applied Statistics and Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, P. R. China.

Zhijiang Institute of Big Data and Statistics, Zhejiang Gongshang University, Hangzhou 310018, P. R. China.

E-mail: zhu.liping@ruc.edu.cn

(Received March 2019; accepted July 2020)