

## BROWNIAN DISTANCE COVARIANCE

BY GÁBOR J. SZÉKELY<sup>1</sup> AND MARIA L. RIZZO

*Bowling Green State University, Hungarian Academy of Sciences and  
 Bowling Green State University*

Distance correlation is a new class of multivariate dependence coefficients applicable to random vectors of arbitrary and not necessarily equal dimension. Distance covariance and distance correlation are analogous to product-moment covariance and correlation, but generalize and extend these classical bivariate measures of dependence. Distance correlation characterizes independence: it is zero if and only if the random vectors are independent. The notion of covariance with respect to a stochastic process is introduced, and it is shown that population distance covariance coincides with the covariance with respect to Brownian motion; thus, both can be called Brownian distance covariance. In the bivariate case, Brownian covariance is the natural extension of product-moment covariance, as we obtain Pearson product-moment covariance by replacing the Brownian motion in the definition with identity. The corresponding statistic has an elegantly simple computing formula. Advantages of applying Brownian covariance and correlation vs the classical Pearson covariance and correlation are discussed and illustrated.

**1. Introduction.** The importance of independence arises in diverse applications, for inference and whenever it is essential to measure complicated dependence structures in bivariate or multivariate data. This paper focuses on a new dependence coefficient that measures all types of dependence between random vectors  $X$  and  $Y$  in arbitrary dimension. *Distance correlation* and *distance covariance* (Székely, Rizzo, and Bakirov [28]), and *Brownian covariance*, introduced in this paper, provide a new approach to the problem of measuring dependence and testing the joint independence of random vectors in arbitrary dimension. The corresponding statistics have simple computing formulae, apply to sample sizes  $n \geq 2$  (not constrained by dimension), and do not require matrix inversion or estimation of parameters. For example, the distance covariance (dCov) statistic, derived in the

---

Received June 2009; revised October 2009.

Discussed in [10.1214/09-AOAS312A](#), [10.1214/09-AOAS312B](#), [10.1214/09-AOAS312C](#), [10.1214/09-AOAS312D](#), [10.1214/09-AOAS312E](#), [10.1214/09-AOAS312F](#) and [10.1214/09-AOAS312G](#); rejoinder at [10.1214/09-AOAS312REJ](#).

<sup>1</sup>Research supported in part by the NSF.

*Key words and phrases.* Distance correlation, dcor, Brownian covariance, independence, multivariate.

next section, is the square root of

$$\mathcal{V}_n^2 = \frac{1}{n^2} \sum_{k,l=1}^n A_{kl} B_{kl},$$

where  $A_{kl}$  and  $B_{kl}$  are simple linear functions of the pairwise distances between sample elements. It will be shown that the definitions of the new dependence coefficients have theoretical foundations based on characteristic functions and on the new concept of covariance with respect to Brownian motion. Our independence test statistics are consistent against all types of dependent alternatives with finite second moments.

Classical Pearson product-moment correlation ( $\rho$ ) and covariance measure linear dependence between two random variables, and in the bivariate normal case  $\rho = 0$  is equivalent to independence. In the multivariate normal case, a diagonal covariance matrix  $\Sigma$  implies independence, but is not a sufficient condition for independence in the general case. Nonlinear or nonmonotone dependence may exist. Thus,  $\rho$  or  $\Sigma$  do not characterize independence in general.

Although it does not characterize independence, classical correlation is widely applied in time series, clinical trials, longitudinal studies, modeling financial data, meta-analysis, model selection in parametric and nonparametric models, classification and pattern recognition, etc. Ratios and other methods of combining and applying correlation coefficients have also been proposed. An important example is maximal correlation, characterized by Rényi [22].

For multivariate inference, methods based on likelihood ratio tests (LRT) such as Wilks' Lambda [32] or Puri-Sen [20] are not applicable if dimension exceeds sample size, or when distributional assumptions do not hold. Although methods based on ranks can be applied in some problems, many classical methods are effective only for testing linear or monotone types of dependence.

There is much literature on testing or measuring independence. See, for example, Blomqvist [3], Blum, Kiefer, and Rosenblatt [4], or methods outlined in Hollander and Wolfe [16] and Anderson [1]. Multivariate nonparametric approaches to this problem can be found in Taskinen, Oja, and Randles [30], and the references therein.

Our proposed distance correlation represents an entirely new approach. For all distributions with finite first moments, *distance correlation*  $\mathcal{R}$  generalizes the idea of *correlation* in at least two fundamental ways:

- (i)  $\mathcal{R}(X, Y)$  is defined for  $X$  and  $Y$  in arbitrary dimension.
- (ii)  $\mathcal{R}(X, Y) = 0$  characterizes independence of  $X$  and  $Y$ .

The coefficient  $\mathcal{R}(X, Y)$  is a standardized version of distance covariance  $\mathcal{V}(X, Y)$ , defined in the next section. Distance correlation satisfies  $0 \leq \mathcal{R} \leq 1$ , and  $\mathcal{R} = 0$  only if  $X$  and  $Y$  are independent. In the *bivariate normal* case,  $\mathcal{R}$  is a deterministic function of  $\rho$ , and  $\mathcal{R}(X, Y) \leq |\rho(X, Y)|$  with equality when  $\rho = \pm 1$ .

Thus, distance covariance and distance correlation provide a natural extension of Pearson product-moment covariance  $\sigma_{X,Y}$  and correlation  $\rho$ , and new methodology for measuring dependence in all types of applications.

The notion of covariance of random vectors  $(X, Y)$  with respect to a stochastic process  $U$  is introduced in this paper. This new notion  $\text{Cov}_U(X, Y)$  contains as distinct special cases distance covariance  $\mathcal{V}^2(X, Y)$  and, for bivariate  $(X, Y)$ ,  $\sigma_{X,Y}^2$ . The title of this paper refers to  $\text{Cov}_W(X, Y)$ , where  $W$  is a Wiener process.

*Brownian covariance*  $\mathcal{W} = \mathcal{W}(X, Y)$  is based on Brownian motion or Wiener process for random variables  $X \in \mathbb{R}^p$  and  $Y \in \mathbb{R}^q$  with finite second moments. An important property of Brownian covariance is that  $\mathcal{W}(X, Y) = 0$  if and only if  $X$  and  $Y$  are independent.

A surprising result develops: the Brownian covariance is equal to the distance covariance. This equivalence is not only surprising, it also shows that distance covariance is a natural counterpart of product-moment covariance. For bivariate  $(X, Y)$ , by considering the simplest nonrandom function, identity ( $id$ ), we obtain  $\text{Cov}_{id}(X, Y) = \sigma_{X,Y}^2$ . Then by considering the most fundamental random processes, Brownian motion  $W$ , we arrive at  $\text{Cov}_W(X, Y) = \mathcal{V}^2(X, Y)$ . Brownian correlation is a standardized Brownian covariance, such that if Brownian motion is replaced with the identity function, we obtain the absolute value of Pearson's correlation  $\rho$ .

A further advantage of extending Pearson correlation with distance correlation is that while uncorrelatedness ( $\rho = 0$ ) can sometimes replace independence, for example, in proving some classical laws of large numbers, uncorrelatedness is too weak to imply a central limit theorem, even for strongly stationary summands (see Bradley [7–9]). On the other hand, a central limit theorem for strongly stationary sequences of summands follows from  $\mathcal{R} = 0$  type conditions (Székely and Bakirov [25]).

Distance correlation and distance covariance are presented in Section 2. Brownian covariance is introduced in Section 3. Extensions and applications are discussed in Sections 4 and 5.

**2. Distance covariance and distance correlation.** Let  $X$  in  $\mathbb{R}^p$  and  $Y$  in  $\mathbb{R}^q$  be random vectors, where  $p$  and  $q$  are positive integers. The lower case  $f_X$  and  $f_Y$  will be used to denote the characteristic functions of  $X$  and  $Y$ , respectively, and their joint characteristic function is denoted  $f_{X,Y}$ . In terms of characteristic functions,  $X$  and  $Y$  are independent if and only if  $f_{X,Y} = f_X f_Y$ . Thus, a natural approach to measuring the dependence between  $X$  and  $Y$  is to find a suitable norm to measure the distance between  $f_{X,Y}$  and  $f_X f_Y$ .

Distance covariance  $\mathcal{V}$  is a measure of the distance between  $f_{X,Y}$  and the product  $f_X f_Y$ . A norm  $\|\cdot\|$  and a distance  $\|f_{X,Y} - f_X f_Y\|$  are defined in Section 2.2. Then an empirical version of  $\mathcal{V}$  is developed and applied to test the hypothesis of independence

$$H_0 : f_{X,Y} = f_X f_Y \quad \text{vs} \quad H_1 : f_{X,Y} \neq f_X f_Y.$$

In Székely et al. [28] an omnibus test of independence based on the sample distance covariance  $\mathcal{V}$  is introduced that is easily implemented in arbitrary dimension without requiring distributional assumptions. In Monte Carlo studies, the distance covariance test exhibited superior power relative to parametric or rank-based likelihood ratio tests against nonmonotone types of dependence. It was also demonstrated that the tests were quite competitive with the parametric likelihood ratio test when applied to multivariate normal data. The practical message is that distance covariance tests are powerful tests for all types of dependence.

### 2.1. Motivation.

*Notation.* The scalar product of vectors  $t$  and  $s$  is denoted by  $\langle t, s \rangle$ . For complex-valued functions  $f(\cdot)$ , the complex conjugate of  $f$  is denoted by  $\overline{f}$  and  $|f|^2 = f\overline{f}$ . The Euclidean norm of  $x$  in  $\mathbb{R}^p$  is  $|x|_p$ . A primed variable  $X'$  is an independent copy of  $X$ ; that is,  $X$  and  $X'$  are independent and identically distributed (i.i.d.).

For complex functions  $\gamma$  defined on  $\mathbb{R}^p \times \mathbb{R}^q$ , the  $\|\cdot\|_w$ -norm in the weighted  $L_2$  space of functions on  $\mathbb{R}^{p+q}$  is defined by

$$(2.1) \quad \|\gamma(t, s)\|_w^2 = \int_{\mathbb{R}^{p+q}} |\gamma(t, s)|^2 w(t, s) dt ds,$$

where  $w(t, s)$  is an arbitrary positive weight function for which the integral above exists.

With a suitable choice of weight function  $w(t, s)$ , discussed below, we shall define a measure of dependence

$$(2.2) \quad \begin{aligned} \mathcal{V}^2(X, Y; w) &= \|f_{X,Y}(t, s) - f_X(t)f_Y(s)\|_w^2 \\ &= \int_{\mathbb{R}^{p+q}} |f_{X,Y}(t, s) - f_X(t)f_Y(s)|^2 w(t, s) dt ds, \end{aligned}$$

which is analogous to classical covariance, but with the important property that  $\mathcal{V}^2(X, Y; w) = 0$  if and only if  $X$  and  $Y$  are independent. In what follows,  $w$  is chosen such that we can also define

$$\begin{aligned} \mathcal{V}^2(X; w) &= \mathcal{V}^2(X, X; w) = \|f_{X,X}(t, s) - f_X(t)f_X(s)\|_w^2 \\ &= \int_{\mathbb{R}^{2p}} |f_{X,X}(t, s) - f_X(t)f_X(s)|^2 w(t, s) dt ds, \end{aligned}$$

and similarly define  $\mathcal{V}^2(Y; w)$ . Then a standardized version of  $\mathcal{V}(X, Y; w)$  is

$$\mathcal{R}_w = \frac{\mathcal{V}(X, Y; w)}{\sqrt{\mathcal{V}(X; w)\mathcal{V}(Y; w)}},$$

a type of unsigned correlation.

In the definition of the norm (2.1) there are more than one potentially interesting and applicable choices of weight function  $w$ , but not every  $w$  leads to a dependence

measure that has desirable statistical properties. Let us now discuss the motivation for our particular choice of weight function leading to distance covariance.

At least two conditions should be satisfied by the standardized coefficient  $\mathcal{R}_w$ :

- (i)  $\mathcal{R}_w \geq 0$  and  $\mathcal{R}_w = 0$  only if independence holds.
- (ii)  $\mathcal{R}_w$  is scale invariant, that is, invariant with respect to transformations  $(X, Y) \mapsto (\epsilon X, \epsilon Y)$ , for  $\epsilon > 0$ .

However, if we consider *integrable* weight function  $w(t, s)$ , then for  $X$  and  $Y$  with finite variance

$$\lim_{\epsilon \rightarrow 0} \frac{\mathcal{V}^2(\epsilon X, \epsilon Y; w)}{\sqrt{\mathcal{V}^2(\epsilon X; w)\mathcal{V}^2(\epsilon Y; w)}} = \rho^2(X, Y).$$

The above limit is obtained by considering the Taylor expansions of the underlying characteristic functions. Thus, if the weight function is integrable,  $\mathcal{R}_w$  can be arbitrarily close to zero even if  $X$  and  $Y$  are dependent. By using a suitable *non-integrable* weight function, we can obtain an  $\mathcal{R}_w$  that satisfies both properties (i) and (ii) above.

Considering the operations on characteristic functions involved in evaluating the integrand in (2.2), a promising solution to the choice of weight function  $w$  is suggested by the following lemma.

LEMMA 1. *If  $0 < \alpha < 2$ , then for all  $x$  in  $\mathbb{R}^d$*

$$\int_{\mathbb{R}^d} \frac{1 - \cos\langle t, x \rangle}{|t|_d^{d+\alpha}} dt = C(d, \alpha) |x|_d^\alpha,$$

where

$$C(d, \alpha) = \frac{2\pi^{d/2}\Gamma(1 - \alpha/2)}{\alpha 2^\alpha \Gamma((d + \alpha)/2)},$$

and  $\Gamma(\cdot)$  is the complete gamma function. The integrals at 0 and  $\infty$  are meant in the principal value sense:  $\lim_{\epsilon \rightarrow 0} \int_{\mathbb{R}^d \setminus \{\epsilon B + \epsilon^{-1} B^c\}}$ , where  $B$  is the unit ball (centered at 0) in  $\mathbb{R}^d$  and  $B^c$  is the complement of  $B$ .

A proof of Lemma 1 is given in Székely and Rizzo [27]. Lemma 1 suggests the weight functions

$$(2.3) \quad w(t, s; \alpha) = (C(p, \alpha)C(q, \alpha)|t|_p^{p+\alpha}|s|_q^{q+\alpha})^{-1}, \quad 0 < \alpha < 2.$$

The weight functions (2.3) result in coefficients  $\mathcal{R}_w$  that satisfy the scale invariance property (ii) above.

In the simplest case corresponding to  $\alpha = 1$  and Euclidean norm  $|x|$ ,

$$(2.4) \quad w(t, s) = (c_p c_q |t|_p^{1+p} |s|_q^{1+q})^{-1},$$

where

$$(2.5) \quad c_d = C(d, 1) = \frac{\pi^{(1+d)/2}}{\Gamma((1+d)/2)}.$$

(The constant  $2c_d$  is the surface area of the unit sphere in  $\mathbb{R}^{d+1}$ .)

REMARK 1. Lemma 1 is applied to evaluate the integrand in (2.2) for weight functions (2.3) and (2.4). For example, if  $\alpha = 1$  (2.4), then by Lemma 1 there exist constants  $c_p$  and  $c_q$  such that for  $X$  in  $\mathbb{R}^p$  and  $Y$  in  $\mathbb{R}^q$ ,

$$\begin{aligned} \int_{\mathbb{R}^p} \frac{1 - \exp\{i\langle t, X \rangle\}}{|t|_p^{1+p}} dt &= c_p |X|_p, & \int_{\mathbb{R}^q} \frac{1 - \exp\{i\langle s, Y \rangle\}}{|s|_q^{1+q}} ds &= c_q |Y|_q, \\ \int_{\mathbb{R}^p} \int_{\mathbb{R}^q} \frac{1 - \exp\{i\langle t, X \rangle + i\langle s, Y \rangle\}}{|t|_p^{1+p} |s|_q^{1+q}} dt ds &= c_p c_q |X|_p |Y|_q. \end{aligned}$$

Distance covariance and distance correlation are a class of dependence coefficients and statistics obtained by applying a weight function of the type (2.3),  $0 < \alpha < 2$ . This type of weight function leads to a simple product-average form of the covariance (2.8) analogous to Pearson covariance. Other interesting weight functions could be considered (see, e.g., Bakirov, Rizzo and Székely [2]), but only the weight functions (2.3) lead to distance covariance type statistics (2.8).

In this paper we apply weight function (2.4) and the corresponding weighted  $L_2$  norm  $\|\cdot\|$ , omitting the index  $w$ , and write the dependence measure (2.2) as  $\mathcal{V}^2(X, Y)$ . Section 4.1 extends our results for  $\alpha \in (0, 2)$ .

For finiteness of  $\|f_{X,Y}(t, s) - f_X(t)f_Y(s)\|^2$ , it is sufficient that  $E|X|_p < \infty$  and  $E|Y|_q < \infty$ .

## 2.2. Definitions.

DEFINITION 1. The *distance covariance* (dCov) between random vectors  $X$  and  $Y$  with finite first moments is the nonnegative number  $\mathcal{V}(X, Y)$  defined by

$$(2.6) \quad \begin{aligned} \mathcal{V}^2(X, Y) &= \|f_{X,Y}(t, s) - f_X(t)f_Y(s)\|^2 \\ &= \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \frac{|f_{X,Y}(t, s) - f_X(t)f_Y(s)|^2}{|t|_p^{1+p} |s|_q^{1+q}} dt ds. \end{aligned}$$

Similarly, distance variance (dVar) is defined as the square root of

$$\mathcal{V}^2(X) = \mathcal{V}^2(X, X) = \|f_{X,X}(t, s) - f_X(t)f_X(s)\|^2.$$

By definition of the norm  $\|\cdot\|$ , it is clear that  $\mathcal{V}(X, Y) \geq 0$  and  $\mathcal{V}(X, Y) = 0$  if and only if  $X$  and  $Y$  are independent.

DEFINITION 2. The *distance correlation* ( $dCor$ ) between random vectors  $X$  and  $Y$  with finite first moments is the nonnegative number  $\mathcal{R}(X, Y)$  defined by

$$(2.7) \quad \mathcal{R}^2(X, Y) = \begin{cases} \frac{\mathcal{V}^2(X, Y)}{\sqrt{\mathcal{V}^2(X)\mathcal{V}^2(Y)}}, & \mathcal{V}^2(X)\mathcal{V}^2(Y) > 0; \\ 0, & \mathcal{V}^2(X)\mathcal{V}^2(Y) = 0. \end{cases}$$

Several properties of  $\mathcal{R}$  analogous to  $\rho$  are given in Theorem 3. Results for the special case of bivariate normal  $(X, Y)$  are given in Theorem 6.

The distance dependence statistics are defined as follows. For a random sample  $(\mathbf{X}, \mathbf{Y}) = \{(X_k, Y_k) : k = 1, \dots, n\}$  of  $n$  i.i.d. random vectors  $(X, Y)$  from the joint distribution of random vectors  $X$  in  $\mathbb{R}^p$  and  $Y$  in  $\mathbb{R}^q$ , compute the Euclidean distance matrices  $(a_{kl}) = (|X_k - X_l|_p)$  and  $(b_{kl}) = (|Y_k - Y_l|_q)$ . Define

$$A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot}, \quad k, l = 1, \dots, n,$$

where

$$\bar{a}_{k\cdot} = \frac{1}{n} \sum_{l=1}^n a_{kl}, \quad \bar{a}_{\cdot l} = \frac{1}{n} \sum_{k=1}^n a_{kl}, \quad \bar{a}_{\cdot\cdot} = \frac{1}{n^2} \sum_{k,l=1}^n a_{kl}.$$

Similarly, define  $B_{kl} = b_{kl} - \bar{b}_{k\cdot} - \bar{b}_{\cdot l} + \bar{b}_{\cdot\cdot}$ , for  $k, l = 1, \dots, n$ .

DEFINITION 3. The nonnegative sample distance covariance  $\mathcal{V}_n(\mathbf{X}, \mathbf{Y})$  and sample distance correlation  $\mathcal{R}_n(\mathbf{X}, \mathbf{Y})$  are defined by

$$(2.8) \quad \mathcal{V}_n^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{k,l=1}^n A_{kl} B_{kl}$$

and

$$(2.9) \quad \mathcal{R}_n^2(\mathbf{X}, \mathbf{Y}) = \begin{cases} \frac{\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y})}{\sqrt{\mathcal{V}_n^2(\mathbf{X})\mathcal{V}_n^2(\mathbf{Y})}}, & \mathcal{V}_n^2(\mathbf{X})\mathcal{V}_n^2(\mathbf{Y}) > 0; \\ 0, & \mathcal{V}_n^2(\mathbf{X})\mathcal{V}_n^2(\mathbf{Y}) = 0, \end{cases}$$

respectively, where the sample distance variance is defined by

$$(2.10) \quad \mathcal{V}_n^2(\mathbf{X}) = \mathcal{V}_n^2(\mathbf{X}, \mathbf{X}) = \frac{1}{n^2} \sum_{k,l=1}^n A_{kl}^2.$$

The nonnegativity of  $\mathcal{R}_n^2$  and  $\mathcal{V}_n^2$  may not be immediately obvious from the definitions above, but this property as well as the motivation for the definitions of the statistics will become clear from Theorem 1 below.

2.3. *Properties of distance covariance.* Several interesting properties of distance covariance are obtained. Results in this section are summarized as follows:

- (i) Equivalent definition of  $\mathcal{V}_n$  in terms of empirical characteristic functions and norm  $\|\cdot\|$ .
- (ii) Almost sure convergence  $\mathcal{V}_n \rightarrow \mathcal{V}$  and  $\mathcal{R}_n^2 \rightarrow \mathcal{R}^2$ .
- (iii) Properties of  $\mathcal{V}(X, Y)$ ,  $\mathcal{V}(X)$ , and  $\mathcal{R}(X, Y)$ .
- (iv) Properties of  $\mathcal{R}_n$  and  $\mathcal{V}_n$ .
- (v) Weak convergence of  $n\mathcal{V}_n^2$ , the limit distribution of  $n\mathcal{V}_n^2$ , and statistical consistency.
- (vi) Results for the bivariate normal case.

Many of these results were obtained in Székely et al. [28]. Here we give the proofs of new results and readers are referred to [28] for more details and proofs of our previous results.

*An equivalent definition of  $\mathcal{V}_n$ .* The coefficient  $\mathcal{V}(X, Y)$  is defined in terms of characteristic functions, thus, a natural approach is to define the statistic  $\mathcal{V}_n(\mathbf{X}, \mathbf{Y})$  in terms of empirical characteristic functions. The joint empirical characteristic function of the sample,  $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ , is

$$f_{X,Y}^n(t, s) = \frac{1}{n} \sum_{k=1}^n \exp\{i\langle t, X_k \rangle + i\langle s, Y_k \rangle\}.$$

The marginal empirical characteristic functions of the  $X$  sample and  $Y$  sample are

$$f_X^n(t) = \frac{1}{n} \sum_{k=1}^n \exp\{i\langle t, X_k \rangle\}, \quad f_Y^n(s) = \frac{1}{n} \sum_{k=1}^n \exp\{i\langle s, Y_k \rangle\},$$

respectively. Then an empirical version of distance covariance could have been defined as  $\|f_{X,Y}^n(t, s) - f_X^n(t)f_Y^n(s)\|$ , where the norm  $\|\cdot\|$  is defined by the integral as above in (2.1). Theorem 1 establishes that this definition is equivalent to Definition 3.

**THEOREM 1.** *If  $(\mathbf{X}, \mathbf{Y})$  is a sample from the joint distribution of  $(X, Y)$ , then*

$$\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y}) = \|f_{X,Y}^n(t, s) - f_X^n(t)f_Y^n(s)\|^2.$$

The proof applies Lemma 1 to evaluate the integral  $\|f_{X,Y}^n(t, s) - f_X^n(t)f_Y^n(s)\|^2$  with  $w(t, s) = \{c_p c_q |t|_p^{1+p} |s|_q^{1+q}\}^{-1}$ . An intermediate result is

$$(2.11) \quad \|f_{X,Y}^n(t, s) - f_X^n(t)f_Y^n(s)\|^2 = T_1 + T_2 - 2T_3,$$



where

$$\begin{aligned} T_1 &= \frac{1}{n^2} \sum_{k,l=1}^n |X_k - X_l|_p |Y_k - Y_l|_q, \\ T_2 &= \frac{1}{n^2} \sum_{k,l=1}^n |X_k - X_l|_p \frac{1}{n^2} \sum_{k,l=1}^n |Y_k - Y_l|_q, \\ T_3 &= \frac{1}{n^3} \sum_{k=1}^n \sum_{l,m=1}^n |X_k - X_l|_p |Y_k - Y_m|_q. \end{aligned}$$

Then the algebraic identity  $T_1 + T_2 - 2T_3 = \mathcal{V}_n^2(\mathbf{X}, \mathbf{Y})$ , where  $\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y})$  is given by Definition 3, is established to complete the proof.

As a corollary to Theorem 1, we have  $\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y}) \geq 0$ . It is also easy to see that the statistic  $\mathcal{V}_n(\mathbf{X}) = 0$  if and only if every sample observation is identical. If  $\mathcal{V}_n(\mathbf{X}) = 0$ , then  $A_{kl} = 0$  for  $k, l = 1, \dots, n$ . Thus,  $0 = A_{kk} = -\bar{a}_{k\cdot} - \bar{a}_{\cdot k} + \bar{a}_{\cdot\cdot}$  implies that  $\bar{a}_{k\cdot} = \bar{a}_{\cdot k} = \bar{a}_{\cdot\cdot}/2$ , and

$$0 = A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot} = a_{kl} = |X_k - X_l|_p,$$

so  $X_1 = \dots = X_n$ .

REMARK 2. The simplicity of formula (2.8) for  $\mathcal{V}_n$  in Definition 3 has practical advantages. Although the identity (2.11) in Theorem 1 provides an alternate computing formula for  $\mathcal{V}_n$ , the original formula in Definition 3 is simpler and requires less computing time (1/3 less time per statistic on our current machine, for sample size 100). Reusable computations and other efficiencies possible using the simpler formula (2.8) execute our permutation tests in 94% to 98% less time, which depends on the number of replicates. It is straightforward to apply resampling procedures without the need to recompute the distance matrices. See Example 5, where a jackknife procedure is illustrated.

THEOREM 2. *If  $E|X|_p < \infty$  and  $E|Y|_q < \infty$ , then almost surely*

$$\lim_{n \rightarrow \infty} \mathcal{V}_n(\mathbf{X}, \mathbf{Y}) = \mathcal{V}(X, Y).$$

COROLLARY 1. *If  $E(|X|_p + |Y|_q) < \infty$ , then almost surely*

$$\lim_{n \rightarrow \infty} \mathcal{R}_n^2(\mathbf{X}, \mathbf{Y}) = \mathcal{R}^2(X, Y).$$

THEOREM 3. *For random vectors  $X \in \mathbb{R}^p$  and  $Y \in \mathbb{R}^q$  such that  $E(|X|_p + |Y|_q) < \infty$ , the following properties hold:*

- (i)  $0 \leq \mathcal{R}(X, Y) \leq 1$ , and  $\mathcal{R} = 0$  if and only if  $X$  and  $Y$  are independent.

(ii)  $\mathcal{V}(a_1 + b_1 C_1 X, a_2 + b_2 C_2 Y) = \sqrt{|b_1 b_2|} \mathcal{V}(X, Y)$ , for all constant vectors  $a_1 \in \mathbb{R}^p$ ,  $a_2 \in \mathbb{R}^q$ , scalars  $b_1, b_2$  and orthonormal matrices  $C_1, C_2$  in  $\mathbb{R}^p$  and  $\mathbb{R}^q$ , respectively.

(iii) If the random vector  $(X_1, Y_1)$  is independent of the random vector  $(X_2, Y_2)$ , then

$$\mathcal{V}(X_1 + X_2, Y_1 + Y_2) \leq \mathcal{V}(X_1, Y_1) + \mathcal{V}(X_2, Y_2).$$

Equality holds if and only if  $X_1$  and  $Y_1$  are both constants, or  $X_2$  and  $Y_2$  are both constants, or  $X_1, X_2, Y_1, Y_2$  are mutually independent.

(iv)  $\mathcal{V}(X) = 0$  implies that  $X = E[X]$ , almost surely.

(v)  $\mathcal{V}(a + bCX) = |b| \mathcal{V}(X)$ , for all constant vectors  $a$  in  $\mathbb{R}^p$ , scalars  $b$ , and  $p \times p$  orthonormal matrices  $C$ .

(vi) If  $X$  and  $Y$  are independent, then  $\mathcal{V}(X + Y) \leq \mathcal{V}(X) + \mathcal{V}(Y)$ . Equality holds if and only if one of the random vectors  $X$  or  $Y$  is constant.

Proofs of statements (iii) and (vi) are given in the [Appendix](#).

THEOREM 4.

(i)  $\mathcal{V}_n(\mathbf{X}, \mathbf{Y}) \geq 0$ .

(ii)  $\mathcal{V}_n(\mathbf{X}) = 0$  if and only if every sample observation is identical.

(iii)  $0 \leq \mathcal{R}_n(\mathbf{X}, \mathbf{Y}) \leq 1$ .

(iv)  $\mathcal{R}_n(\mathbf{X}, \mathbf{Y}) = 1$  implies that the dimensions of the linear subspaces spanned by  $\mathbf{X}$  and  $\mathbf{Y}$  respectively are almost surely equal, and if we assume that these subspaces are equal, then in this subspace

$$\mathbf{Y} = a + b\mathbf{X}C$$

for some vector  $a$ , nonzero real number  $b$ , and orthogonal matrix  $C$ .

Theorem 3 and the results below for the dCov test can be applied in a wide range of problems in statistical modeling and inference, including nonparametric models, models with multivariate response, or when dimension exceeds sample size. Some applications are discussed in Section 5.

*Asymptotic properties of  $n\mathcal{V}_n^2$ .* A multivariate test of independence is determined by  $n\mathcal{V}_n^2$  or  $n\mathcal{V}_n^2/T_2$ , where  $T_2 = \bar{a}.. \bar{b}..$  is as defined in Theorem 1. If we apply the latter version, it normalizes the statistic so that asymptotically it has expected value 1. Then if  $E(|X|_p + |Y|_q) < \infty$ , under independence,  $n\mathcal{V}_n^2/T_2$  converges in distribution to a quadratic form

$$(2.12) \quad Q \stackrel{\mathcal{D}}{=} \sum_{j=1}^{\infty} \lambda_j Z_j^2,$$

where  $Z_j$  are independent standard normal random variables,  $\{\lambda_j\}$  are nonnegative constants that depend on the distribution of  $(X, Y)$ , and  $E[Q] = 1$ . A test of independence that rejects independence for large  $n\mathcal{V}_n^2/T_2$  (or  $n\mathcal{V}_n^2$ ) is statistically consistent against all alternatives with finite first moments.

In the next theorem we need only assume finiteness of first moments for weak convergence of  $n\mathcal{V}_n^2$  under the independence hypothesis.

**THEOREM 5 (Weak convergence).** *If  $X$  and  $Y$  are independent and  $E(|X|_p + |Y|_q) < \infty$ , then*

$$n\mathcal{V}_n^2 \xrightarrow[n \rightarrow \infty]{D} \|\zeta(t, s)\|^2,$$

where  $\zeta(\cdot)$  is a complex-valued zero mean Gaussian random process with covariance function

$$R(u, u_0) = (f_X(t - t_0) - f_X(t)\overline{f_X(t_0)})(f_Y(s - s_0) - f_Y(s)\overline{f_Y(s_0)}),$$

for  $u = (t, s)$ ,  $u_0 = (t_0, s_0) \in \mathbb{R}^p \times \mathbb{R}^q$ .

**COROLLARY 2.** *If  $E(|X|_p + |Y|_q) < \infty$ , then*

- (i) *If  $X$  and  $Y$  are independent, then  $n\mathcal{V}_n^2/T_2 \xrightarrow[n \rightarrow \infty]{D} Q$  where  $Q$  is a nonnegative quadratic form of centered Gaussian random variables (2.12) and  $E[Q] = 1$ .*
- (ii) *If  $X$  and  $Y$  are independent, then  $n\mathcal{V}_n^2 \xrightarrow[n \rightarrow \infty]{D} Q_1$  where  $Q_1$  is a nonnegative quadratic form of centered Gaussian random variables and  $E[Q_1] = E|X - X'|E|Y - Y'|$ .*
- (iii) *If  $X$  and  $Y$  are dependent, then  $n\mathcal{V}_n^2/T_2 \xrightarrow[n \rightarrow \infty]{P} \infty$  and  $n\mathcal{V}_n^2 \xrightarrow[n \rightarrow \infty]{P} \infty$ .*

Corollary 2(i), (ii) guarantees that the dCov test statistic has a proper limit distribution under the hypothesis of independence for all  $X$  and  $Y$  with finite first moments, while Corollary 2(iii) shows that under any dependent alternative, the dCov test statistic tends to infinity (stochastically). Thus, the dCov test of independence is statistically consistent against all types of dependence.

The dCov test is easy to implement as a permutation test, which is the method that we applied in our examples and power comparisons. For the permutation test implementation one can apply test statistic  $n\mathcal{V}_n^2$ . Large values of  $n\mathcal{V}_n^2$  (or  $n\mathcal{V}_n^2/T_2$ ) are significant. The dCov test and test statistics are implemented in the *energy* package for R in functions *dcov.test*, *dcov*, and *dcor* [21, 23].

We have also obtained a result that gives an asymptotic critical value applicable to arbitrary distributions. If  $Q$  is a quadratic form of centered Gaussian random variables and  $E[Q] = 1$ , then

$$P\{Q \geq \chi_{1-\alpha}^2(1)\} \leq \alpha$$

for all  $0 < \alpha \leq 0.215$ , where  $\chi_{1-\alpha}^2(1)$  is the  $(1 - \alpha)$  quantile of a chi-square variable with 1 degree of freedom. This result follows from a theorem of Székely and Bakirov [26], page 181.

Thus, a test that rejects independence if  $n\mathcal{V}_n^2/T_2 \geq \chi_{1-\alpha}^2(1)$  has an asymptotic significance level at most  $\alpha$ . This test criterion could be quite conservative for many distributions. Although this critical value is conservative, it is a sharp bound; the upper bound  $\alpha$  is achieved when  $X$  and  $Y$  are independent Bernoulli variables.

*Results for the bivariate normal distribution.* When  $(X, Y)$  has a bivariate normal distribution, there is a deterministic relation between  $\mathcal{R}$  and  $|\rho|$ .

**THEOREM 6.** *If  $X$  and  $Y$  are standard normal, with correlation  $\rho = \rho(X, Y)$ , then:*

- (i)  $\mathcal{R}(X, Y) \leq |\rho|$ ,
- (ii)  $\mathcal{R}^2(X, Y) = \frac{\rho \arcsin \rho + \sqrt{1-\rho^2} - \rho \arcsin(\rho/2) - \sqrt{4-\rho^2} + 1}{1+\pi/3-\sqrt{3}}$ ,
- (iii)  $\inf_{\rho \neq 0} \frac{\mathcal{R}(X, Y)}{|\rho|} = \lim_{\rho \rightarrow 0} \frac{\mathcal{R}(X, Y)}{|\rho|} = \frac{1}{2(1+\pi/3-\sqrt{3})^{1/2}} \approx 0.89066$ .

The relation between  $\mathcal{R}$  and  $\rho$  for a bivariate normal distribution is shown in Figure 1.

**3. Brownian covariance.** To introduce the notion of Brownian covariance, let us begin by considering the squared product-moment covariance. Recall that

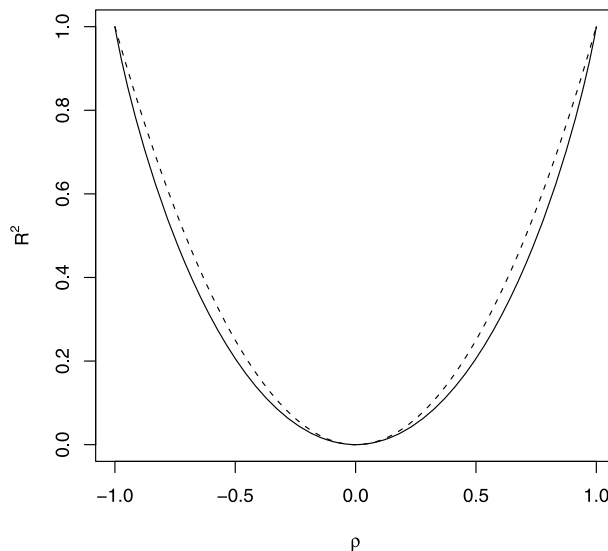


FIG. 1. Dependence coefficient  $\mathcal{R}^2$  (solid line) and correlation  $\rho^2$  (dashed line) in the bivariate normal case.

a primed variable  $X'$  denotes an i.i.d. copy of the unprimed symbol  $X$ . For two real-valued random variables, the square of their classical covariance is

$$(3.1) \quad \begin{aligned} & E^2[(X - E(X))(Y - E(Y))] \\ &= E[(X - E(X))(X' - E(X'))(Y - E(Y))(Y' - E(Y'))]. \end{aligned}$$

Now we generalize the squared covariance and define the square of conditional covariance, given two real-valued stochastic processes  $U(\cdot)$  and  $V(\cdot)$ . We obtain an interesting result when  $U$  and  $V$  are independent Wiener processes.

First, to center the random variable  $X$  in the conditional covariance, we need the following definition. Let  $X$  be a real-valued random variable and  $\{U(t) : t \in \mathbb{R}^1\}$  a real-valued stochastic process, independent of  $X$ . The  $U$ -centered version of  $X$  is defined by

$$(3.2) \quad X_U = U(X) - \int_{-\infty}^{\infty} U(t) dF_X(t) = U(X) - E[U(X)|U],$$

whenever the conditional expectation exists.

Note that if  $id$  is identity, we have  $X_{id} = X - E[X]$ . The important examples in this paper apply Brownian motion/Wiener processes.

**3.1. Definition of Brownian covariance.** Let  $W$  be a two-sided one-dimensional Brownian motion/Wiener process with expectation zero and covariance function

$$(3.3) \quad |s| + |t| - |s - t| = 2 \min(s, t), \quad t, s \geq 0.$$

This is twice the covariance of the standard Wiener process. Here the factor 2 simplifies the computations, so throughout the paper, covariance function (3.3) is assumed for  $W$ .

**DEFINITION 4.** The Brownian covariance or the Wiener covariance of two real-valued random variables  $X$  and  $Y$  with finite second moments is a non-negative number defined by its square

$$(3.4) \quad \mathcal{W}^2(X, Y) = \text{Cov}_W^2(X, Y) = E[X_W X'_W Y_{W'} Y'_{W'}],$$

where  $(W, W')$  does not depend on  $(X, Y, X', Y')$ .

Note that if  $W$  in  $\text{Cov}_W$  is replaced by the (nonrandom) identity function  $id$ , then  $\text{Cov}_{id}(X, Y) = |\text{Cov}(X, Y)| = |\sigma_{X,Y}|$ , the absolute value of Pearson's product-moment covariance. While the standardized product-moment covariance, Pearson correlation ( $\rho$ ), measures the degree of linear relationship between two real-valued variables, we shall see that standardized Brownian covariance measures the degree of *all kinds of possible relationships* between two real-valued random variables.

The definition of  $\text{Cov}_W(X, Y)$  can be extended to random processes in higher dimensions as follows. If  $X$  is an  $\mathbb{R}^p$ -valued random variable, and  $U(s)$  is a random process (random field) defined for all  $s \in \mathbb{R}^p$  and independent of  $X$ , define the  $U$ -centered version of  $X$  by

$$X_U = U(X) - E[U(X)|U],$$

whenever the conditional expectation exists.

DEFINITION 5. If  $X$  is an  $\mathbb{R}^p$ -valued random variable,  $Y$  is an  $\mathbb{R}^q$ -valued random variable, and  $U(s)$  and  $V(t)$  are arbitrary random processes (random fields) defined for all  $s \in \mathbb{R}^p$ ,  $t \in \mathbb{R}^q$ , then the  $(U, V)$  covariance of  $(X, Y)$  is defined as the nonnegative number whose square is

$$(3.5) \quad \text{Cov}_{U,V}^2(X, Y) = E[X_U X'_U Y_V Y'_V],$$

whenever the right-hand side is nonnegative and finite.

In particular, if  $W$  and  $W'$  are independent Brownian motions with covariance function (3.3) on  $\mathbb{R}^p$ , and  $\mathbb{R}^q$  respectively, the *Brownian covariance* of  $X$  and  $Y$  is defined by

$$(3.6) \quad \mathcal{W}^2(X, Y) = \text{Cov}_W^2(X, Y) = \text{Cov}_{W,W'}^2(X, Y).$$

Similarly, for random variables with finite variance define the *Brownian variance* by

$$\mathcal{W}(X) = \text{Var}_W(X) = \text{Cov}_W(X, X).$$

DEFINITION 6. The Brownian correlation is defined as

$$\text{Cor}_W(X, Y) = \frac{\mathcal{W}(X, Y)}{\sqrt{\mathcal{W}(X)\mathcal{W}(Y)}}$$

whenever the denominator is not zero; otherwise  $\text{Cor}_W(X, Y) = 0$ .

In the following sections we prove that  $\text{Cov}_W(X, Y)$  exists for random vectors  $X$  and  $Y$  with finite second moments, and derive the Brownian covariance in this case.

3.2. *Existence of  $\mathcal{W}(X, Y)$ .* In the following, the subscript on Euclidean norm  $|x|_d$  for  $x \in \mathbb{R}^d$  is omitted when the dimension is self-evident.

THEOREM 7. If  $X$  is an  $\mathbb{R}^p$ -valued random variable,  $Y$  is an  $\mathbb{R}^q$ -valued random variable, and  $E(|X|^2 + |Y|^2) < \infty$ , then  $E[X_W X'_W Y_{W'} Y'_{W'}]$  is nonnegative and finite, and

$$(3.7) \quad \begin{aligned} \mathcal{W}^2(X, Y) &= E[X_W X'_W Y_{W'} Y'_{W'}] \\ &= E|X - X'| |Y - Y'| + E|X - X'| E|Y - Y'| \\ &\quad - E|X - X'| |Y - Y''| - E|X - X''| |Y - Y'|, \end{aligned}$$

where  $(X, Y)$ ,  $(X', Y')$ , and  $(X'', Y'')$  are i.i.d.

PROOF. Observe that

$$\begin{aligned} E[X_W X'_W Y_{W'} Y'_{W'}] &= E[E(X_W Y_{W'} X'_W Y'_{W'} | W, W')] \\ &= E[E(X_W Y_{W'} | W, W') E(X'_W Y'_{W'} | W, W')] \\ &= E[E(X_W Y_{W'} | W, W')]^2, \end{aligned}$$

and this is always nonnegative. For finiteness, it is enough to prove that all factors in the definition of  $\text{Cov}_W(X, Y)$  have finite fourth moments. Equation (3.7) relies on the special form of the covariance function (3.3) of  $W$ . The remaining details are in the [Appendix](#).  $\square$

See Section 4.1 for definitions and extension of results for the general case of fractional Brownian motion with Hurst parameter  $0 < H < 1$  and covariance function  $|t|^{2H} + |s|^{2H} - |t - s|^{2H}$ .

### 3.3. The surprising coincidence: $\mathcal{W} = \mathcal{V}$ .

THEOREM 8. For arbitrary  $X \in \mathbb{R}^p, Y \in \mathbb{R}^q$  with finite second moments

$$\mathcal{W}(X, Y) = \mathcal{V}(X, Y).$$

PROOF. Both  $\mathcal{V}$  and  $\mathcal{W}$  are nonnegative, hence, it is enough to show that their squares coincide. Lemma 1 can be applied to evaluate  $\mathcal{V}^2(X, Y)$ . In the numerator of the integral we have terms like

$$E[\cos\langle X - X', t \rangle \cos\langle Y - Y', s \rangle],$$

where  $X, X'$  are i.i.d. and  $Y, Y'$  are i.i.d. Now apply the identity

$$\cos u \cos v = 1 - (1 - \cos u) - (1 - \cos v) + (1 - \cos u)(1 - \cos v)$$

and Lemma 1 to simplify the integrand. After cancelation in the numerator of the integrand, there remains to evaluate integrals of the type

$$\begin{aligned} &E \int_{\mathbb{R}^{p+q}} \frac{[1 - \cos\langle X - X', t \rangle][1 - \cos\langle Y - Y', s \rangle]}{|t|^{1+p}|s|^{1+q}} dt ds \\ &= E \left[ \int_{\mathbb{R}^p} \frac{1 - \cos\langle X - X', t \rangle}{|t|^{1+p}} dt \times \int_{\mathbb{R}^q} \frac{1 - \cos\langle Y - Y', s \rangle}{|s|^{1+q}} ds \right] \\ &= c_p c_q E|X - X'| E|Y - Y'|. \end{aligned}$$

Applying similar steps, after further simplification, we obtain

$$\begin{aligned} \mathcal{V}^2(X, Y) &= E|X - X'| |Y - Y'| + E|X - X'| E|Y - Y'| \\ &\quad - E|X - X'| |Y - Y''| - E|X - X''| |Y - Y'|, \end{aligned}$$

and this is exactly equal to the expression (3.7) obtained for  $\mathcal{W}(X, Y)$  in Theorem 7.  $\square$

As a corollary to Theorem 8, the properties of Brownian covariance for random vectors  $X$  and  $Y$  with finite second moments are therefore the same properties established for distance covariance  $\mathcal{V}(X, Y)$  in Theorem 3.

The surprising result that Brownian covariance equals distance covariance  $\text{dCov}$ , exactly as defined in (2.6) for  $X \in \mathbb{R}^p$  and  $Y \in \mathbb{R}^q$ , parallels a familiar special case when  $p = q = 1$ . For bivariate  $(X, Y)$  we found that  $\mathcal{R}(X, Y)$  is a natural counterpart of the *absolute value of* the Pearson correlation. That is, if in (3.5)  $U$  and  $V$  are the simplest nonrandom function  $\text{id}$ , then we obtain the square of Pearson covariance  $\sigma_{X,Y}^2$ . Next, if we consider the most fundamental random processes,  $U = W$  and  $V = W'$ , we obtain the square of distance covariance,  $\mathcal{V}^2(X, Y)$ .

Interested readers are referred to Székely and Bakirov [25] for the background of the interesting coincidence in Theorem 8.

#### 4. Extensions.

4.1. *The class of  $\alpha$ -distance dependence measures.* In two contexts above we have introduced dependence measures based on Euclidean distance and on Brownian motion with Hurst index  $H = 1/2$  (self-similarity index). Our definitions and results can be extended to a one-parameter family of distance dependence measures indexed by a positive exponent  $0 < \alpha < 2$  on Euclidean distance, or equivalently by an index  $h$ , where  $h = 2H$  for Hurst parameters  $0 < H < 1$ .

If  $E(|X|_p^\alpha + |Y|_q^\alpha) < \infty$  define  $\mathcal{V}^{(\alpha)}$  by its square

$$\begin{aligned} \mathcal{V}^{2(\alpha)}(X, Y) &= \|f_{X,Y}(t, s) - f_X(t)f_Y(s)\|_\alpha^2 \\ &= \frac{1}{C(p, \alpha)C(q, \alpha)} \int_{\mathbb{R}^{p+q}} \frac{|f_{X,Y}(t, s) - f_X(t)f_Y(s)|^2}{|t|_p^{\alpha+p}|s|_q^{\alpha+q}} dt ds. \end{aligned}$$

Similarly,  $\mathcal{R}^{(\alpha)}$  is the square root of

$$\mathcal{R}^{2(\alpha)} = \frac{\mathcal{V}^{2(\alpha)}(X, Y)}{\sqrt{\mathcal{V}^{2(\alpha)}(X)\mathcal{V}^{2(\alpha)}(Y)}}, \quad 0 < \mathcal{V}^{2(\alpha)}(X), \mathcal{V}^{2(\alpha)}(Y) < \infty,$$

and  $\mathcal{R}^{(\alpha)} = 0$  if  $\mathcal{V}^{2(\alpha)}(X)\mathcal{V}^{2(\alpha)}(Y) = 0$ .

Now consider the Lévy fractional Brownian motion  $\{W_H^d(t), t \in \mathbb{R}^d\}$  with Hurst index  $H \in (0, 1)$ , which is a centered Gaussian random process with covariance function

$$E[W_H^d(t)W_H^d(s)] = |t|^{2H} + |s|^{2H} - |t - s|^{2H}, \quad t, s \in \mathbb{R}^d.$$



See Herbin and Merzbach [15].

In the following,  $(W_H, W_{H^*}')$  and  $(X, X', Y, Y')$  are supposed to be independent.

Using Lemma 1, it can be shown for Hurst parameters  $0 < H, H^* \leq 1, h := 2H$ , and  $h^* := 2H^*$ , that

$$\begin{aligned}
 & \text{Cov}_{W_H^p, W_{H^*}^{q'}}^2(X, Y) \\
 &= \frac{1}{C(p, h)C(q, h^*)} \int_{\mathbb{R}^p} \int_{\mathbb{R}^q} \frac{|f(t, s) - f(t)g(s)|^2 dt ds}{|t|_p^{p+h} |s|_q^{q+h^*}} \\
 (4.1) \quad &= E|X - X'|_p^h |Y - Y'|_q^{h^*} + E|X - X'|_p^h E|Y - Y'|_q^{h^*} \\
 &\quad - E|X - X'|_p^h |Y - Y''|_q^{h^*} - E|X - X''|_p^h |Y - Y'|_q^{h^*}.
 \end{aligned}$$

Here we need to suppose that  $E|X|_p^{2h} < \infty$ ,  $E|Y|_q^{2h^*} < \infty$ . Observe that when  $h = h^* = 1$ , (4.1) is equation (3.7) of Theorem 7.

The corresponding statistics are defined by replacing the exponent 1 with exponent  $\alpha$  (or  $h$ ) in the distance dependence statistics (2.8), (2.10), and (2.9). That is, in the sample distance matrices replace  $a_{kl} = |X_k - X_l|_p$  with  $a_{kl} = |X_k - X_l|_p^\alpha$ , and replace  $b_{kl} = |Y_k - Y_l|_q$  with  $b_{kl} = |Y_k - Y_l|_q^\alpha$ ,  $k, l = 1, \dots, n$ .

Theorem 2 can be generalized for  $\|\cdot\|_\alpha$  norms, so that almost sure convergence of  $\mathcal{V}_n^{(\alpha)} \rightarrow \mathcal{V}^{(\alpha)}$  follows if the  $\alpha$ -moments are finite. Similarly, one can prove the weak convergence and statistical consistency for  $\alpha$  exponents,  $0 < \alpha < 2$ , provided that  $\alpha$  moments are finite.

Note that the strict inequality  $0 < \alpha < 2$  is important. Although  $\mathcal{V}^{(2)}$  can be defined for  $\alpha = 2$ , it does not characterize independence. Indeed, the case  $\alpha = 2$  (squared Euclidean distance) leads to classical product-moment correlation and covariance for bivariate  $(X, Y)$ . Specifically, if  $p = q = 1$ , then  $\mathcal{R}^{(2)} = |\rho|$ ,  $\mathcal{R}_n^{(2)} = |\hat{\rho}|$ , and  $\mathcal{V}_n^{(2)} = 2|\hat{\sigma}_{xy}|$ , where  $\hat{\sigma}_{xy}$  is the maximum likelihood estimator of Pearson covariance  $\sigma_{x,y} = \sigma(X, Y)$ .

**4.2. Affine invariance.** Independence is preserved under affine transformations hence it is natural to consider dependence measures that are affine invariant. We have seen that  $\mathcal{R}(X, Y)$  is invariant with respect to orthogonal transformations

$$(4.2) \quad X \mapsto a_1 + b_1 C_1 X, \quad Y \mapsto a_2 + b_2 C_2 Y,$$

where  $a_1, a_2$  are arbitrary vectors,  $b_1, b_2$  are arbitrary nonzero numbers, and  $C_1, C_2$  are arbitrary orthogonal matrices. We can also define a distance correlation that is affine invariant. Define the scaled samples  $\mathbf{X}^*$  and  $\mathbf{Y}^*$  by

$$(4.3) \quad \mathbf{X}^* = \mathbf{X} S_X^{-1/2}, \quad \mathbf{Y}^* = \mathbf{Y} S_Y^{-1/2},$$

where  $S_X$  and  $S_Y$  are the sample covariance matrices of  $\mathbf{X}$  and  $\mathbf{Y}$  respectively. The sample vectors in (4.3) are not invariant to affine transformations, but the distances,

$|X_k^* - X_l^*|$  and  $|Y_k^* - Y_l^*|$ ,  $k, l = 1, \dots, n$ , are invariant to affine transformations. Thus, an affine distance correlation statistic can be defined by its square

$$\mathcal{R}_n^{*2}(\mathbf{X}, \mathbf{Y}) = \frac{\mathcal{V}_n^2(\mathbf{X}^*, \mathbf{Y}^*)}{\sqrt{\mathcal{V}_n^2(\mathbf{X}^*)\mathcal{V}_n^2(\mathbf{Y}^*)}}.$$

Theoretical properties established for  $\mathcal{V}_n$  and  $\mathcal{R}_n$  also hold for  $\mathcal{V}_n^*$  and  $\mathcal{R}_n^*$ , because the transformation simply replaces the original weight function  $\{c_p c_q |t|_p^{1+p} |s|_q^{1+q}\}^{-1}$  with  $\{c_p c_q |\Sigma_X^{1/2} t|_p^{1+p} |\Sigma_Y^{1/2} s|_q^{1+q}\}^{-1}$ .

**4.3. Rank test.** In the case of bivariate  $(X, Y)$  one can also consider a distance covariance test of independence for  $\text{rank}(X)$ ,  $\text{rank}(Y)$ , which has the advantage that it is distribution free and invariant with respect to monotone transformations of  $X$  and  $Y$ , but usually at a cost of lower power than the  $\text{dCov}(X, Y)$  test (see Example 1). The rank-dCov test can be applied to continuous or discrete data, but for discrete data it is necessary to use the correct method for breaking ties. Any ties in ranks should be broken randomly, so that a sample of size  $n$  is transformed to some permutation of the integers  $1:n$ . A table of critical values for the statistic  $n\mathcal{R}_n^2$ , based on Monte Carlo results, is provided in Table 2 in the [Appendix](#).

## 5. Applications.

**5.1. Nonlinear and nonmonotone dependence.** Suppose that one wants to test the independence of  $X$  and  $Y$ , where  $X$  and  $Y$  cannot be observed directly, but can only be measured with independent errors. Consider the following:

- (i) Suppose that  $X_i$  can only be measured through observation of  $A_i = X_i + \varepsilon_i$ , where  $\varepsilon_i$  are independent of  $X_i$ , and similarly for  $Y_i$ .
- (ii) One can only measure (non) random functions of  $X$  and  $Y$ , for example,  $A_i = \phi(X_i)$  and  $B_i = \psi(Y_i)$ .
- (iii) Suppose both (i) and (ii) for certain types of random  $\phi$  and  $\psi$ .

In all of these cases, even if  $(X, Y)$  were jointly normal, the dependence between  $(A, B)$  can be such that the correlation of  $A$  and  $B$  is almost irrelevant, but  $\text{dCor}(A, B)$  is obviously relevant.

In this section we illustrate a few of the many possible applications of distance covariance. The  $\text{dCov}$  test has been applied using the *dcov.test* function in the *energy* [23] package for R [21], where it is implemented as a permutation test.

### 5.2. Examples.

**EXAMPLE 1.** This example is similar to the type considered in (ii), with observed data from the NIST Statistical Reference Datasets (NIST StRD) for Non-linear Regression. The data analyzed is *Eckerle4*, data from an NIST study of

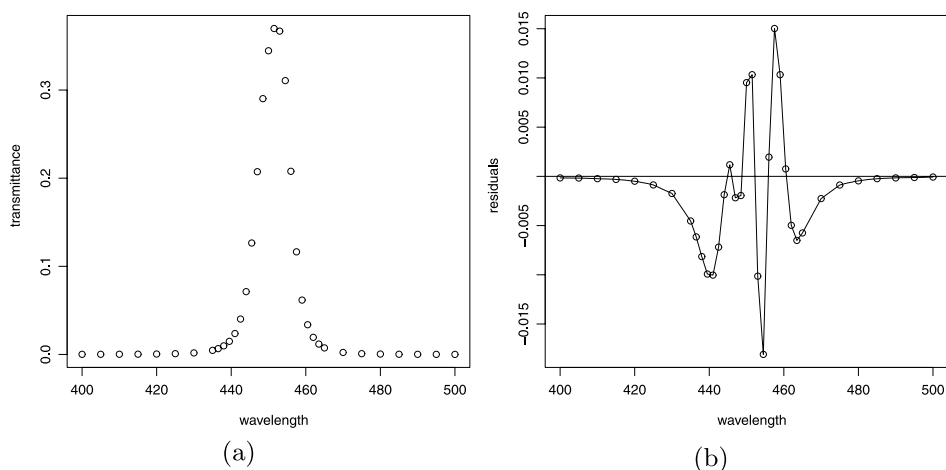


FIG. 2. The Eckerle4 data (a) and plot of residuals vs predictor variable for the NIST certified estimates (b), in Example 1.

circular interference transmittance [10]. There are 35 observations, the response variable is transmittance, and the predictor variable is wavelength. A plot of the data in Figure 2(a) reveals that there is a nonlinear relation between wavelength and transmittance. The proposed nonlinear model is

$$y = f(x; \beta) + \varepsilon = \frac{\beta_1}{\beta_2} \exp\left\{\frac{-(x - \beta_3)^2}{2\beta_2^2}\right\} + \varepsilon,$$

where  $\beta_1, \beta_2 > 0$ ,  $\beta_3 \in \mathbb{R}$ , and  $\varepsilon$  is random error. In the hypothesized model,  $Y$  depends on the density of  $X$ .

Results of the dCov test of independence of wavelength and transmittance are

```
dCov test of independence
data: x and y
nV^2 = 8.1337, p-value = 0.021
sample estimates:
dCor
0.4275431
```

with  $\mathcal{R}_n \doteq 0.43$ , and dCov is significant ( $p$ -value = 0.021) based on 999 replicates. In contrast, neither Pearson correlation  $\hat{\rho} = 0.0356$ , ( $p$ -value = 0.839) nor Spearman rank correlation  $\hat{\rho}_s = 0.0062$  ( $p$ -value = 0.9718) detects the nonlinear dependence between wavelength and transmittance, even though the relation in Figure 2(a) appears to be nearly deterministic.

The certified estimates (best solution found) for the parameters are reported by NIST as  $\hat{\beta}_1 \doteq 1.55438$ ,  $\hat{\beta}_2 \doteq 4.08883$ , and  $\hat{\beta}_3 \doteq 451.541$ . The residuals of the fitted model are easiest to analyze when plotted vs the predictor variable as in Figure 2(b). Comparing residuals and transmittance,

```

dCov test of independence
data: y and res
nV^2 = 0.0019, p-value = 0.019
sample estimates:
dCor
0.4285534

```

we have  $\mathcal{R}_n \doteq 0.43$  and the dCov test is significant ( $p$ -value = 0.019) based on 999 replicates. Again the Pearson correlation is nonsignificant ( $\hat{\rho} \doteq 0.11$ ,  $p$ -value = 0.5378).

Although nonlinear dependence is clearly evident in both plots, note that the methodology applies to multivariate analysis as well, for which residual plots are much less informative.

**EXAMPLE 2.** In the model specification of Example 1, the response variable  $Y$  is assumed to be proportional to a normal density plus random error. For simplicity, consider  $(X, Y) = (X, \phi(X))$ , where  $X$  is standard normal and  $\phi(\cdot)$  is the standard normal density. Results of a Monte Carlo power comparison of the dCov test with classical Pearson correlation and Spearman rank tests are shown in Figure 3. The power estimates are computed as the proportion of significant tests out of 10,000 at 10% significance level.

In this example, where the relation between  $X$  and  $Y$  is deterministic but not monotone, it is clear that the dCov test is superior to product moment correla-

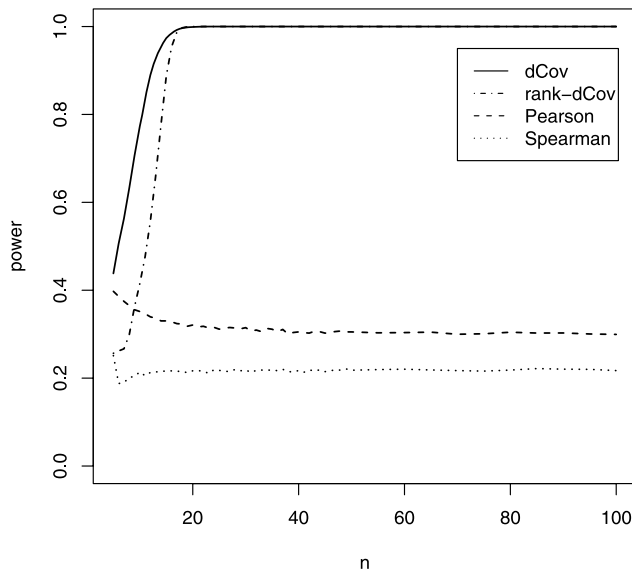


FIG. 3. Example 2: Empirical power at 0.1 significance and sample size  $n$ .

tion tests. Statistical consistency of the dCov test is evident, as its power increases to 1 with sample size, while the power of correlation tests against this alternative remains approximately level across sample sizes. We also note that distance correlation applied to ranks of the data is more powerful in this example than either correlation test, although somewhat less powerful than the dCov test on the original  $(X, Y)$  data.

**EXAMPLE 3.** The Saviotti aircraft data [24] record six characteristics of aircraft designs which appeared during the twentieth century. We consider two variables, wing span (m) and speed (km/h) for the 230 designs of the third (of three) periods. This example and the data (*aircraft*) are from Bowman and Azzalini [5, 6]. A scatterplot on log-log scale of the variables and contours of a nonparametric density estimate are shown in Figures 4(a) and 4(b). The nonlinear relation between speed and wing span is quite evident from the plots.

The dCov test of independence of  $\log(\text{Speed})$  and  $\log(\text{Span})$  in period 3 is significant ( $p$ -value = 0.001), while the Pearson correlation test is not significant ( $p$ -value = 0.8001).

```
dCov test of independence
data: logSpeed3 and logSpan3
nV^2 = 3.4151, p-value = 0.001
sample estimates:
      dCor
0.2804530
```

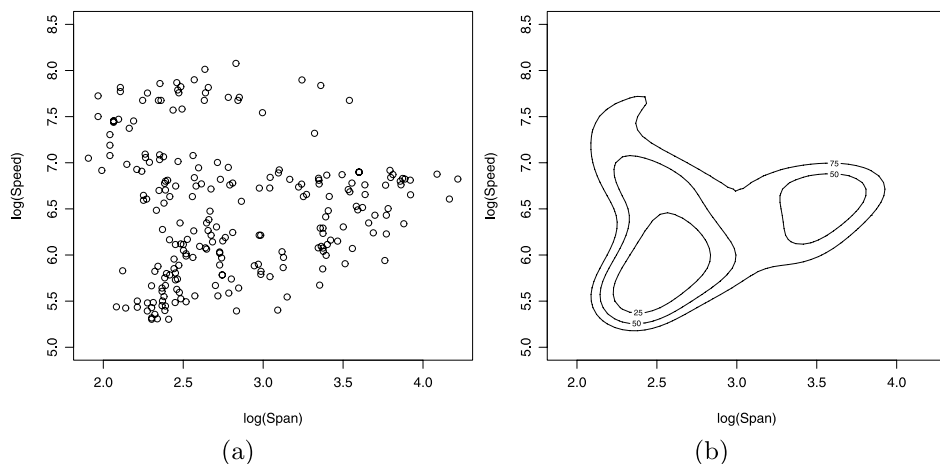


FIG. 4. Scatterplot and contours of density estimate for the aircraft speed and span variables, period 3, in Example 3.

```

Pearson's product-moment correlation
data: logSpeed3 and logSpan3
t = 0.2535, df = 228, p-value = 0.8001
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.1128179 0.1458274
sample estimates:
cor
0.01678556

```

The sample estimates are  $\hat{\rho} = 0.0168$  and  $\mathcal{R}_n = 0.2805$ . Here we have an example of observed data where two variables are nearly uncorrelated, but dependent. We obtained essentially the same results on the correlations of ranks of the data.

EXAMPLE 4. This example compares dCor and Pearson correlation in exploratory data analysis. Consider the Freedman [31, 13] data on crime rates in US metropolitan areas with 1968 populations of 250,000 or more. The data set is available from Fox [12], and contains four numeric variables:

*population* (total 1968, in thousands),  
*nonwhite* (percent nonwhite population, 1960),  
*density* (population per square mile, 1968),  
*crime* (crime rate per 100,000, 1969).

The 110 observations contain missing values. The data analyzed are the 100 cities with complete data. Pearson  $\hat{\rho}$  and dCor statistics  $\mathcal{R}_n$  are shown in Table 1. Note that there is a significant association between crime and population density measured by dCor, which is not significant when measured by  $\hat{\rho}$ .

Analysis of this data continues in Example 5.

EXAMPLE 5 (Influential observations). When  $\mathcal{V}_n$  and  $\mathcal{R}_n$  are computed using formula (2.8), it is straightforward to apply a jackknife procedure to identify possible influential observations or to estimate standard error of  $\mathcal{V}_n$  or  $\mathcal{R}_n$ . A ‘leave-one-out’ sample corresponds to  $(n - 1) \times (n - 1)$  matrices  $A_{(i)kl}$  and  $B_{(i)kl}$ , where

TABLE 1

Pearson correlation and distance correlation statistics for the Freedman data of Example 4. Significance at 0.05, 0.01, 0.001 for the corresponding tests is indicated by \*, \*\*, \*\*\*, respectively

|            | Pearson  |          |          | dCor     |          |          |
|------------|----------|----------|----------|----------|----------|----------|
|            | Nonwhite | Density  | Crime    | Nonwhite | Density  | Crime    |
| Population | 0.070    | 0.368*** | 0.396*** | 0.260*   | 0.615*** | 0.422**  |
| Nonwhite   |          | 0.002    | 0.294**  |          | 0.194    | 0.385*** |
| Density    |          |          | 0.112    |          |          | 0.250*   |

the subscript  $(i)$  indicates that the  $i$ th observation is left out. Then  $A_{(i)kl}$  is computed from distance matrix  $A = (a_{kl})$  by omitting the  $i$ th row and the  $i$ th column of  $A$ , and similarly  $B_{(i)kl}$  is computed from  $B = (b_{kl})$  by omitting the  $i$ th row and the  $i$ th column of  $B$ . Then

$$\mathcal{V}_{(i)}^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{(n-1)^2} \sum_{k,l \neq i} A_{(i)kl} B_{(i)kl}, \quad i = 1, \dots, n,$$

are the jackknife replicates of  $\mathcal{V}_n^2$ , obtained without recomputing matrices  $A$  and  $B$ . Similarly,  $\mathcal{R}_{(i)}^2$  can be computed from the matrices  $A$  and  $B$ . A jackknife estimate of the standard error of  $\mathcal{R}_n$  is thus easily obtained from the matrices  $A, B$  (on the jackknife, see, e.g., Efron and Tibshirani [11]).

The jackknife replicates  $\mathcal{R}_{(i)}$  can be used to identify potentially influential observations, in the sense that outliers within the sample of replicates correspond to observations  $X_i$  that increase or decrease the dependence coefficient more than other observations. These unusual replicates are not necessarily outliers in the original data.

Consider the crime data of Example 4. The studentized jackknife replicates  $\mathcal{R}_{(i)}/\widehat{se}(\mathcal{R}_{(i)})$ ,  $i = 1, \dots, n$ , are plotted in Figure 5(a). These replicates were computed on the pairs  $(x, y)$ , where  $x$  is the vector (*nonwhite*, *density*, *population*) and  $y$  is *crime*. The plot suggests that Philadelphia is an unusual observation. For comparison we plot the first two principal components of the four variables in Figure 5(b), but Philadelphia (PHIL) does not appear to be an unusual observation in this plot or other plots (not shown), including those where  $\log(\text{population})$  replaces

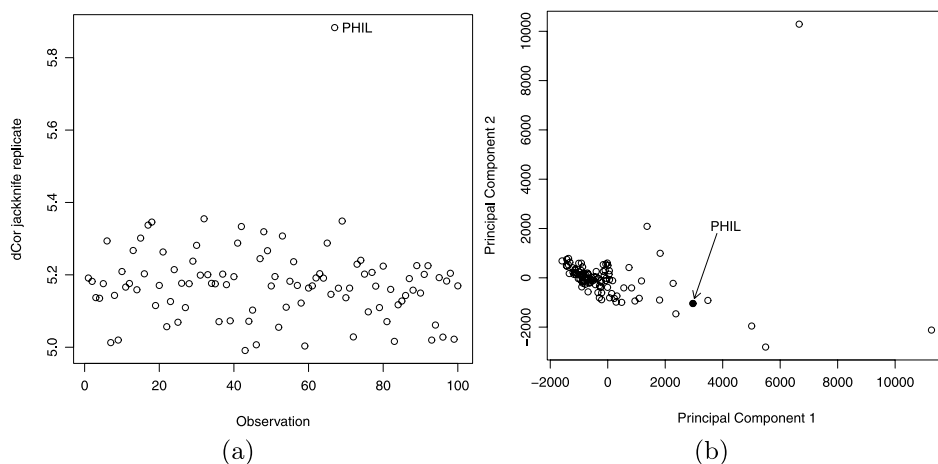


FIG. 5. Jackknife replicates of  $dCor$  (a) and principal components of Freedman data (b) in Example 5.

*population* in the analysis. One can see from comparing

|              | population | nonwhite | density | crime |
|--------------|------------|----------|---------|-------|
| PHILADELPHIA | 4829       | 15.7     | 1359    | 1753  |

with sample quartiles

|      | population | nonwhite | density  | crime   |
|------|------------|----------|----------|---------|
| 0%   | 270.00     | 0.300    | 37.00    | 458.00  |
| 25%  | 398.75     | 3.400    | 266.50   | 2100.25 |
| 50%  | 664.00     | 7.300    | 412.00   | 2762.00 |
| 75%  | 1167.75    | 14.825   | 773.25   | 3317.75 |
| 100% | 11551.00   | 64.300   | 13087.00 | 5441.00 |

that *crime* in Philadelphia is low while *population*, *nonwhite*, and *density* are all high relative to other cities. Recall that all Pearson correlations were positive in Example 4.

This example illustrates that having a single multivariate summary statistic  $dCor$  that measures dependence is a valuable tool in exploratory data analysis, and it can provide information about potential influential observations prior to model selection.

EXAMPLE 6. In this example we illustrate how to isolate the nonlinear dependence between random vectors to test for nonlinearity.

Gumbel's bivariate exponential distribution [14] has density function

$$f(x, y; \theta) = [(1 + \theta x)(1 + \theta y)] \exp(-x - y - \theta xy), \quad x, y > 0; 0 \leq \theta \leq 1.$$

The marginal distributions are standard exponential, so there is a strong nonlinear, but monotone dependence relation between  $X$  and  $Y$ . The conditional density is

$$f(y|x) = e^{-(1+\theta x)y} [(1 + \theta x)(1 + \theta y) - \theta], \quad y > 0.$$

If  $\theta = 0$ , then  $f_{X,Y}(x, y) = f_X(x)f_Y(y)$  and independence holds, so  $\rho = 0$ . At the opposite extreme, if  $\theta = 1$ , then  $\rho = -0.40365$  (see Kotz, Balakrishnan, and Johnson [18], Section 2.2). Simulated data was generated using the conditional distribution function approach outlined in Johnson [17]. Empirical power of  $dCov$  and correlation tests for the case  $\theta = 0.5$  are compared in Figure 6(a), estimated from 10,000 test decisions each for sample sizes  $\{10:100(10), 120:200(20), 250, 300\}$ . This comparison reveals that the correlation test is more powerful than  $dCov$  against this alternative, which is not unexpected because  $E[Y|X = x] = (1 + \theta + x\theta)/(1 + x\theta)^2$  is monotone.

While we cannot split the  $dCor$  or  $dCov$  coefficient into linear and nonlinear components, we can extract correlation first and then compute  $dCor$  on the residuals. In this way one can separately analyze the linear and nonlinear components of bivariate or multivariate dependence relations.



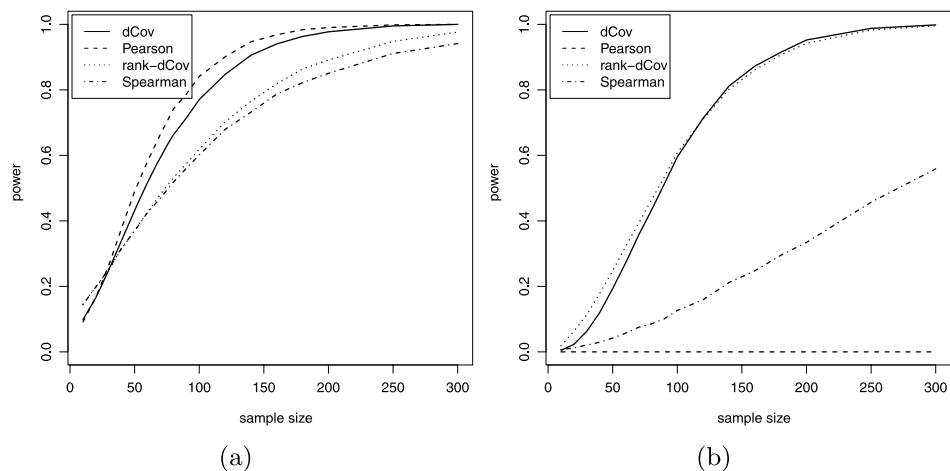


FIG. 6. Power comparison of dCov and correlation tests at 10% significance level for Gumbel's bivariate exponential distribution in Example 6.

To extract the linear component of dependence, fit a linear model  $Y = X\beta + \varepsilon$  to the sample  $(\mathbf{X}, \mathbf{Y})$  by ordinary least squares. It is not necessary to test whether the linear relation is significant. The residuals  $\hat{\varepsilon}_i = X_i\hat{\beta} - Y_i$  are uncorrelated with the predictors  $\mathbf{X}$ . Apply the dCov test of independence to  $(\mathbf{X}, \hat{\varepsilon})$ .

Returning to the Gumbel bivariate exponential example, we have extracted the linear component and applied dCov to the residuals of a simple linear regression model. Repeating the power comparison described above on  $(\mathbf{X}, \hat{\varepsilon})$  data, we obtained the power estimates shown in Figure 6(b). One can note that power of dCov tests is increasing to 1 with sample size, exhibiting statistical consistency against the nonlinear dependence remaining in the residuals of the linear model.

This procedure is easily applied in arbitrary dimension. One can fit a linear multiple regression model or a model with multivariate response to extract the linear component of dependence. This has important practical application for evaluating models in higher dimensions.

More examples, including Monte Carlo power comparisons for random vectors in dimensions up to  $p = q = 30$ , are given in Székely et al. [28].

**6. Summary.** Distance covariance and distance correlation are natural extensions and generalizations of classical Pearson covariance and correlation in at least two ways. In one direction we extend the ability to measure linear association to all types of dependence relations. In another direction we extend the bivariate measure to a single scalar measure of dependence between random vectors in arbitrary dimension. In addition to the obvious theoretical advantages, we have the practical advantages that the dCov and dCor statistics are computationally simple, and applicable in arbitrary dimension not constrained by sample size.

We cannot claim that dCov is the only possible or the only reasonable extension with the above mentioned properties, but we can claim that our extension is a natural generalization of Pearson's covariance in the following sense. We defined the covariance of random vectors with respect to a pair of random processes, and if these random processes are i.i.d. Brownian motions, which is a very natural choice, then we arrive at the distance covariance; on the other hand, if we choose the simplest nonrandom functions, a pair of identity functions (degenerate random processes), then we arrive at Pearson's covariance.

We have illustrated only a few of the many applications where distance correlation may provide additional information not measured by classical correlation or arrays of bivariate statistics. In exploratory data analysis, distance correlation has the flexibility to be applied as a multivariate measure of dependence, or measure of dependence among any of the lower dimensional marginal distributions.

The general linear model is fundamental in data analysis for several reasons, but often a linear model is not adequate. We can test for linearity using dCov as shown in Example 6. Although illustrated for simple linear regression, the basic method is applicable for all types of i.i.d. observations, including longitudinal data or other data with multivariate predictors and/or multivariate response.

In summary, distance correlation is a valuable, practical, and natural tool in data analysis and inference that extends the good properties of classical correlation to multivariate analysis and the general hypothesis of independence.

## APPENDIX A: PROOFS OF STATEMENTS

For  $\mathbb{R}^d$  valued random variables,  $|\cdot|_d$  denotes the Euclidean norm; whenever the dimension is self-evident we suppress the index  $d$ .

### A.1. Proof of Theorem 3(iii) and (vi).

PROOF. Starting with the left side of the inequality (iii),

$$\begin{aligned}
 & \mathcal{V}(X_1 + X_2, Y_1 + Y_2) \\
 &= \|f_{X_1+X_2, Y_1+Y_2}(t, s) - f_{X_1+X_2}(t)f_{Y_1+Y_2}(s)\| \\
 &= \|f_{X_1, Y_1}(t, s)f_{X_2, Y_2}(t, s) - f_{X_1}(t)f_{X_2}(t)f_{Y_1}(s)f_{Y_2}(s)\| \\
 (A.1) \quad &\leq \|f_{X_1, Y_1}(t, s)(f_{X_2, Y_2}(t, s) - f_{X_2}(t)f_{Y_2}(s))\| \\
 &\quad + \|f_{X_2}(t)f_{Y_2}(s)(f_{X_1, Y_1}(t, s) - f_{X_1}(t)f_{Y_1}(s))\| \\
 (A.2) \quad &\leq \|f_{X_2, Y_2}(t, s) - f_{X_2}(t)f_{Y_2}(s)\| + \|f_{X_1, Y_1}(t, s) - f_{X_1}(t)f_{Y_1}(s)\| \\
 &= \mathcal{V}(X_1, Y_1) + \mathcal{V}(X_2, Y_2).
 \end{aligned}$$

It is clear that if (a)  $X_1$  and  $Y_1$  are both constants, (b)  $X_2$  and  $Y_2$  are both constants, or (c)  $X_1, X_2, Y_1, Y_2$  are mutually independent, then we have equality

in (iii). Now suppose that we have equality in (iii), and thus we have equality above at (A.1) and (A.2), but neither (a) nor (b) hold. Then the only way we can have equality at (A.2) is if  $X_1, Y_1$  are independent and also  $X_2, Y_2$  are independent. But our hypothesis assumes that  $(X_1, Y_1)$  and  $(X_2, Y_2)$  are independent hence (c) must hold.

Finally, (vi) follows from (iii). In this special case  $X_1 = Y_1 = X$  and  $X_2 = Y_2 = Y$ . Now (a) means that  $X$  is constant, (b) means that  $Y$  is constant, and (c) means that both of them are constants, because this is the only case when a random variable can be independent of itself.  $\square$

**A.2. Existence of  $\mathcal{W}(X, Y)$ .** To complete the proof of Theorem 7, we need to show that all factors in the definition of  $\text{Cov}_W(X, Y)$  have finite fourth moments.

PROOF. Note that  $E[W^2(t)] = 2|t|$ , so that  $E[W^4(t)] = 3(E[W^2(t)])^2 = 12|t|^2$  and, therefore,

$$E[W^4(X)] = E[E(W^4(X)|X)] = E[12|X|^2] < \infty.$$

On the other hand, by the inequality  $(a + b)^4 \leq 2^4(a^4 + b^4)$ , and by Jensen's inequality, we have

$$\begin{aligned} E(X_W)^4 &= E[W(X) - E(W(X)|W)]^4 \\ &\leq 2^4(E[W^4(X)] + E[E(W(X)|W)]^4) \\ &\leq 2^5 E[W^4(X)] = 2^5 12 E|X|^2 < \infty. \end{aligned}$$

Similarly, the random variables  $X'_{W'}$ ,  $Y_{W'}$ , and  $Y'_{W'}$  also have finite fourth moments, hence,

$$\begin{aligned} \mathcal{W}^2(X, Y) &= E[X_W X'_W Y_{W'} Y'_{W'}] \\ &\leq \frac{1}{4} E[(X_W)^4 + (X'_W)^4 + (Y_{W'})^4 + (Y'_{W'})^4] < \infty. \end{aligned}$$

Above we implicitly used the fact that  $E[W(X)|W] = \int_{\mathbb{R}^p} W(t) dF_X(t)$  exists a.s. This can easily be proved with the help of the Borel–Cantelli lemma, using the fact that the supremum of centered Gaussian processes have small tails (see [29, 19]).

Observe that

$$\begin{aligned} \mathcal{W}^2(X, Y) &= E[X_W X'_W Y_{W'} Y'_{W'}] \\ &= E[E(X_W X'_W Y_{W'} Y'_{W'} | X, X', Y, Y')] \\ &= E[E(X_W X'_W | X, X', Y, Y') E(Y_{W'} Y'_{W'} | X, X', Y, Y')]. \end{aligned}$$

Here

$$\begin{aligned} X_W X'_W &= \left\{ W(X) - \int_{\mathbb{R}^p} W(t) dF_X(t) \right\} \left\{ W(X') - \int_{\mathbb{R}^p} W(t) dF_X(t) \right\} \\ &= W(X)W(X') - \int_{\mathbb{R}^p} W(X)W(t) dF_X(t) \\ &\quad - \int_{\mathbb{R}^p} W(X')W(t) dF_X(t) + \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} W(t)W(s) dF_X(t) dF_X(s). \end{aligned}$$

By the definition of  $W(\cdot)$ , we have  $E[W(t)W(s)] = |t| + |s| - |t - s|$ , thus,

$$\begin{aligned} E[X_W X'_W | X, X', Y, Y'] &= |X| + |X'| - |X - X'| \\ &\quad - \int_{\mathbb{R}^p} (|X| + |t| - |X - t|) dF_X(t) \\ &\quad - \int_{\mathbb{R}^p} (|X'| + |t| - |X' - t|) dF_X(t) \\ &\quad + \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} (|t| + |s| - |t - s|) dF_X(t) dF_X(s). \end{aligned}$$

Hence,

$$\begin{aligned} E[X_W X'_W | X, X', Y, Y'] &= |X| + |X'| - |X - X'| \\ &\quad - (|X| + E|X| - E'|X - X'|) \\ &\quad - (|X'| + E|X| - E''|X' - X''|) \\ &\quad + (E|X| + E|X'| - E|X - X'|) \\ &= E'|X - X'| + E''|X' - X''| - |X - X'| - E|X - X'|, \end{aligned}$$

where  $E'$  denotes the expectation with respect to  $X'$  and  $E''$  denotes the expectation with respect to  $X''$ . A similar argument for  $Y$  completes the proof.  $\square$

## APPENDIX B: CRITICAL VALUES

Estimated critical values for  $n\mathcal{R}_n^2(\text{rank}(\mathbf{X}), \text{rank}(\mathbf{Y}))$  are summarized in Table 2 for 5% and 10% significance levels. The critical values are estimates of the 95th and 90th quantiles of the sampling distribution and were obtained by a large scale Monte Carlo simulation (100,000 replicates for each  $n$ ). For sample sizes  $n \leq 10$ , the probabilities were determined by generating all possible permutations of the ranks, so the achieved significance levels (ASL) reported for  $n \leq 10$  are exact. The rejection region is in the upper tail.

TABLE 2  
Critical values of  $n\mathcal{R}_n^2(\text{rank}(\mathbf{X}), \text{rank}(\mathbf{Y}))$ ; exact achieved significance level (ASL) for  $n \leq 10$ , and Monte Carlo estimates for  $n \geq 11$ . Reject independence if  $n\mathcal{R}_n^2$  is greater than or equal to the table value

| <i>n</i> | 10% (ASL)     | 5% (ASL)      | <i>n</i> | 10%  | 5%   | <i>n</i> | 10%  | 5%   |
|----------|---------------|---------------|----------|------|------|----------|------|------|
| 5        | 3.685 (0.100) | 4.211 (0.050) | 15       | 4.25 | 5.16 | 25       | 4.26 | 5.22 |
| 6        | 3.917 (0.097) | 4.699 (0.047) | 16       | 4.25 | 5.17 | 30       | 4.25 | 5.22 |
| 7        | 4.215 (0.098) | 4.858 (0.047) | 17       | 4.25 | 5.17 | 35       | 4.24 | 5.23 |
| 8        | 4.233 (0.099) | 4.995 (0.050) | 18       | 4.25 | 5.18 | 40       | 4.24 | 5.23 |
| 9        | 4.208 (0.100) | 5.072 (0.050) | 19       | 4.25 | 5.20 | 50       | 4.24 | 5.24 |
| 10       | 4.221 (0.100) | 5.047 (0.050) | 20       | 4.25 | 5.20 | 60       | 4.24 | 5.25 |
| 11       | 4.23          | 5.07          | 21       | 4.26 | 5.21 | 70       | 4.24 | 5.26 |
| 12       | 4.24          | 5.10          | 22       | 4.26 | 5.21 | 80       | 4.24 | 5.26 |
| 13       | 4.25          | 5.14          | 23       | 4.26 | 5.21 | 90       | 4.24 | 5.26 |
| 14       | 4.25          | 5.16          | 24       | 4.26 | 5.22 | 100      | 4.24 | 5.26 |

REFERENCES

[1] ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, New York. [MR1990662](#)

[2] BAKIROV, N. K., RIZZO, M. L. and SZÉKELY, G. J. (2006). A multivariate nonparametric test of independence. *J. Multivariate Anal.* **93** 1742–1756. [MR2298886](#)

[3] BLOMQVIST, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statist.* **21** 593–600. [MR0039190](#)

[4] BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.* **32** 485–498. [MR0125690](#)

[5] BOWMAN, A. and AZZALINI, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford Univ. Press, Oxford.

[6] BOWMAN, A. W. and AZZALINI, A. (2007). R package ‘[sm](#)’: Nonparametric smoothing methods (version 2.2).

[7] BRADLEY, R. C. (1981). Central limit theorem under weak dependence. *J. Multivariate Anal.* **11** 1–16. [MR0612287](#)

[8] BRADLEY, R. C. (1988). A Central Limit theorem for stationary  $\rho$ -mixing sequences with infinite variance. *Ann. Probab.* **16** 313–332. [MR0920274](#)

[9] BRADLEY, R. C. (2007). *Introduction to Strong Mixing Condition*, Vol. 1–3. Kendrick Press. [MR2325294](#)

[10] ECKERLE, K. and NIST (1979). Circular Interference Transmittance Study. Available at <http://www.itl.nist.gov/div898/strd/nls/data/eckerle4.shtml>.

[11] EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York. [MR1270903](#)

[12] FOX, J. (2009). *car*: Companion to Applied Regression. R package version 1.2-14.

[13] FREEDMAN, J. L. (1975). *Crowding and Behavior*. Viking Press, New York.

[14] GUMBEL, E. J. (1961). Multivariate exponential distributions. *Bulletin of the International Statistical Institute* **39** 469–475.

[15] HERBIN, E. and MERZBACH, E. (2007). The multiparameter fractional Brownian motion. In *Math. Everywhere* 93–101. Springer, Berlin. [MR2281427](#)

[16] HOLLANDER, M. and WOLFE, D. A. (1999). *Nonparametric Statistical Methods*, 2nd ed., Wiley, New York. [MR1666064](#)

- [17] JOHNSON, M. E. (1987). *Multivariate Statistical Simulation*. Wiley, New York.
- [18] KOTZ, S., BALAKRISHNAN, N. and JOHNSON, N. L. (2000). *Continuous Multivariate Distributions, Vol. 1*, 2nd ed. Wiley, New York. [MR1788152](#)
- [19] LANDAU, H. J. and SHEPP, L. A. (1970). On the supremum of a Gaussian process. *Sankya Ser. A* **32** 369–378. [MR0286167](#)
- [20] PURI, M. L. and SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York. [MR0298844](#)
- [21] R DEVELOPMENT CORE TEAM (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at <http://www.R-project.org>.
- [22] RÉNYI, A. (1959). On measures of dependence. *Acta. Math. Acad. Sci. Hungary* **10** 441–451. [MR0115203](#)
- [23] RIZZO, M. L. and SZÉKELY, G. J. (2008). Energy: E-statistics (energy statistics). R package version 1.1-0.
- [24] SAVIOTTI, P. P. (1996). *Technological Evolution, Variety and Economy*. Edward Elgar, Cheltenham.
- [25] SZÉKELY, G. J. and BAKIROV, N. K. (2008). Brownian covariance and CLT for stationary sequences. Technical Report No. 08-01. Dept. Mathematics and Statistics, Bowling Green State Univ., Bowling Green, OH.
- [26] SZÉKELY, G. J. and BAKIROV, N. K. (2003). Extremal probabilities for Gaussian quadratic forms. *Probab. Theory Related Fields* **126** 184–202. [MR1990053](#)
- [27] SZÉKELY, G. J. and RIZZO, M. L. (2005). Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. *J. Classification* **22** 151–183. [MR2231170](#)
- [28] SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing independence by correlation of distances. *Ann. Statist.* **35** 2769–2794. [MR2382665](#)
- [29] TALAGRAND, M. (1988). Small tails for the supremum of a gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.* **24** 307–315. [MR0953122](#)
- [30] TASKINEN, S., OJA, H. and RANGLES, R. H. (2005). Multivariate nonparametric tests of independence. *J. Amer. Statist. Assoc.* **100** 916–925. [MR2201019](#)
- [31] UNITED STATES BUREAU OF THE CENSUS (1970). Statistical Abstract of the United States.
- [32] WILKS, S. S. (1935). On the independence of  $k$  sets of normally distributed statistical variables. *Econometrica* **3** 309–326.

DEPARTMENT OF MATHEMATICS AND STATISTICS  
 BOWLING GREEN STATE UNIVERSITY  
 BOWLING GREEN, OHIO 43403  
 USA  
 AND  
 RÉNYI INSTITUTE OF MATHEMATICS  
 HUNGARIAN ACADEMY OF SCIENCES  
 BUDAPEST  
 HUNGARY  
 E-MAIL: [gszekely@nsf.gov](mailto:gszekely@nsf.gov)

DEPARTMENT OF MATHEMATICS AND STATISTICS  
 BOWLING GREEN STATE UNIVERSITY  
 BOWLING GREEN, OHIO 43403  
 USA  
 E-MAIL: [mrizzo@bgsu.edu](mailto:mrizzo@bgsu.edu)