Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Ball Covariance: A Generic Measure of Dependence in Banach Space

Wenliang Pan, Xueqin Wang, Heping Zhang, Hongtu Zhu & Jin Zhu

To cite this article: Wenliang Pan, Xueqin Wang, Heping Zhang, Hongtu Zhu & Jin Zhu (2020) Ball Covariance: A Generic Measure of Dependence in Banach Space, Journal of the American Statistical Association, 115:529, 307-317, DOI: 10.1080/01621459.2018.1543600

To link to this article: https://doi.org/10.1080/01621459.2018.1543600

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2020, VOL. 115, NO. 529, 307–317: Theory and Methods
https://doi.org/10.1080/01621459.2018.1543600

Taylor & Francis
Taylor & Francis Group

Check for updates

Ball Covariance: A Generic Measure of Dependence in Banach Space

Wenliang Pan[a], Xueqin Wang[b,c,d], Heping Zhang[e], Hongtu Zhu[f], and Jin Zhu[a]

[a]School of Mathematics, Department of Statistical Science, Sun Yat-Sen University, Guangzhou, China; [b]School of Mathematics, Department of Statistical Science, Southern China Center for Statistical Science, Sun Yat-Sen University, Guangzhou, China; [c]Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou, China; [d]Xinhua College, Sun Yat-Sen University, Guangzhou, 510520, China; [e]School of Public Health, Yale University, New Haven, CT; [f]University of Texas MD Anderson Cancer Center, Houston, TX

ABSTRACT
Technological advances in science and engineering have led to the routine collection of large and complex data objects, where the dependence structure among those objects is often of great interest. Those complex objects (e.g., different brain subcortical structures) often reside in some Banach spaces, and hence their relationship cannot be well characterized by most of the existing measures of dependence such as correlation coefficients developed in Hilbert spaces. To overcome the limitations of the existing measures, we propose Ball Covariance as a generic measure of dependence between two random objects in two possibly different Banach spaces. Our Ball Covariance possesses the following attractive properties: (i) It is nonparametric and model-free, which make the proposed measure robust to model mis-specification; (ii) It is nonnegative and equal to zero if and only if two random objects in two separable Banach spaces are independent; (iii) Empirical Ball Covariance is easy to compute and can be used as a test statistic of independence. We present both theoretical and numerical results to reveal the potential power of the Ball Covariance in detecting dependence. Also importantly, we analyze two real datasets to demonstrate the usefulness of Ball Covariance in the complex dependence detection. Supplementary materials for this article are avaiable online.

ARTICLE HISTORY
Received November 2017
Revised October 2018

KEYWORDS
Ball correlation; Ball covariance; Hoeffding's dependence measure; Rank; Shape analysis.

1. Introduction

The need to measure dependence or independence among complex objects arises in many domains, such as medical imaging, computational biology, and computer vision (Dryden and Mardia 1998; Grenander and Miller 2007; Fan and Lv 2008; Brunel et al. 2010; Younes 2010; Jung, Dryden, and Marron 2012; Li, Zhong, and Zhu 2012; Cornea et al. 2016; Banerjee et al. 2016; Srivastava and Klassen 2016). Examples of complex objects include the Grassmann manifolds, planar shapes, tree-structured data, matrix Lie groups, deformation fields, symmetric positive-definite (SPD) matrices, and the shape representations of cortical and subcortical structures. Most of these complex objects are in a non-Hilbert space, and thus they are inherently nonlinear and high-dimensional (or even infinite-dimensional). However, the traditional statistical techniques were developed in the Hilbert space and may not be directly applied to such complex objects. Therefore, analysis of objects possibly in a non-Hilbert space presents major mathematical and computational challenges.

Correlation, as a basic concept of dependence in statistics, has been widely developed for Hilbert spaces, such as Pearson correlation (Pearson 1895), Spearman's rho (Spearman 1904), and Kendall's tau (Kendall 1938). Besides, Hoeffding (1948) and Blum, Kiefer, and Rosenblatt (1961) proposed two Cramér–von Mises criterion dependence measures based on the empirical distribution function, which can also be regarded as two special types of correlation. Distance correlation, proposed by Székely, Rizzo, and Bakirov (2007), represents one of the most important breakthroughs in measuring the dependence between two random variables (or vectors). Zhu et al. (2017) also proposed a projection correlation in Euclidean space. These two measures are powerful in detecting nonlinear and nonmonotonic dependence between two random vectors of arbitrary dimension. More importantly, they satisfy the independence-zero equivalence property: they are equal to zero if and only if the two random vectors are independent with Euclidean metric. However, Lyons (2013) pointed out that this independence-zero equivalence property holds for distance correlation if and only if the metric spaces underlying two random vectors are of strong negative type. Unfortunately, many metric spaces are not of strong negative type, or even of negative type. In those cases, distance correlation does not have the independence-zero equivalence property. For instance, $\mathbb{R}^n$ with ℓ^p metric is not of negative type whenever $3 \leq n \leq \infty$ and $2 < p \leq \infty$. The geodesic metric, commonly used to evaluate the distance of the manifold structure, is of negative type, but not strong negative type. It is noteworthy that the cortical surface data which we will analyze below belong to a metric space that may not be of strong negative type. Furthermore, we give two simulation examples (Examples 4.3.7 and 4.3.8) to illustrate that the independence

CONTACT Heping Zhang heping.zhang@yale.edu School of Public Health, Yale University, New Haven, CT 06520, USA.
All authors contributed equally to this article, and the authors are listed in the alphabetic order.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JASA.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

test based on distance covariance totally loses its power when the distances are not of strong negative type.

It is fundamental to find a dependence measure, if exists, that enjoys the independence-zero equivalence property in a metric space, even when the metric space does not satisfy the necessary and sufficient conditions in Lyons (2013). We accomplish this goal in Banach spaces as well as metric spaces. Because Banach spaces have more algebraic structures than metric spaces which enables us to use more analytic techniques. Another reason we do so is that the Banach–Mazur theorem (Banach 1932; Kleiber and Pervin 1969) allows a metric space to isometrically embed to a Banach space. In this article, we propose a novel dependence measure: Ball Covariance, and show that it is equal to zero if and only if the two random objects in two possibly different Banach spaces are independent.

The rest of this article is organized as follows. In Section 2, we define Ball Covariance by using the projection-type method and Hoeffding's dependence measure (Hoeffding 1948) on the corresponding one-dimensional space of radial distance. Then, we show that a Ball Covariance has the independence-zero equivalence property for two random objects in two Banach spaces under some mild conditions. Basic properties such as the Cauchy-type inequality are also examined. Furthermore, we show that the Heller–Heller–Gorfine (HHG) measure (Heller, Heller, and Gorfine 2013) can be derived from Ball Covariance by choosing a proper weight. In Section 3, we introduce the empirical Ball Covariance as a new test statistic of independence. Its asymptotic distributions under both null and alternative hypotheses are derived. This new test statistic is further shown to be consistent against alternatives. Simulation studies are conducted to compare the performance of distance covariance and a few special forms of Ball Covariance in Section 4. We illustrate the usefulness of Ball Covariance in analyzing a shape dataset for neuroimaging phenotypes as well as another dataset on 16s rRNA gut microbiota in Section 5. We conclude this work with a few remarks in Section 6. Some of the technical details are deferred to the appendix.

2. Ball Covariance and Ball Correlation

2.1. Ball Covariance

Let $(\mathscr{X}, \rho)$ and $(\mathscr{Y}, \zeta)$ be two Banach spaces, where the norms ρ and ζ also represent their induced distances. Let θ be a Borel probability measure on $\mathscr{X} \times \mathscr{Y}$, μ, ν be two Borel probability measures on $\mathscr{X}, \mathscr{Y}$, and (X, Y) be a B-valued random variable defined on a probability space such that $(X, Y) \sim \theta, X \sim \mu$, and $Y \sim \nu$. Denote the closed ball with the center x_1 and the radius $\rho(x_1, x_2)$ in $\mathscr{X}$ as $\bar{B}(x_1, \rho(x_1, x_2))$ or $\bar{B}_\rho(x_1, x_2)$, and the closed ball with the center y_1 and the radius $\zeta(y_1, y_2)$ in $\mathscr{Y}$ as $\bar{B}(y_1, \zeta(y_1, y_2))$ or $\bar{B}_\zeta(y_1, y_2)$. Let $\{W_i = (X_i, Y_i), i = 1, 2, \ldots\}$ be an infinite sequence of iid samples of (X, Y), and $\omega = (\omega_1, \omega_2)$ be the positive weight function on the support set of θ.

Definition 2.1.1. The Ball Covariance is defined as an integral of the Hoeffding's dependence measure on the coordinate of radius over poles as follows.

$$\mathbf{BCov}^2_\omega(X, Y) = \int [\theta - \mu \otimes \nu]^2 (\bar{B}_\rho(x_1, x_2) \times \bar{B}_\zeta(y_1, y_2)) \omega_1(x_1, x_2)\omega_2(y_1, y_2)\theta(dx_1, dy_1)\theta(dx_2, dy_2),$$

where $[\theta - \mu \otimes \nu]^2(A \times B) := [\theta(A \times B) - \mu(A)\nu(B)]^2$ for $A \in \mathscr{X}$ and $B \in \mathscr{Y}$.

In one-dimensional case, if we let $A = (-\infty, x]$, $B = (-\infty, y]$, and $\omega_1 = \omega_2 = 1$, then Ball Covariance reduces to the following simple form:

$$H = \int [\theta - \mu \otimes \nu]^2 (A \times B)\theta(dx, dy),$$

which is Hoeffding's dependence measure.

Remark 2.1.1. In Pan, Wang, and Zhang (2018), Ball divergence is defined as

$$D(\eta_1, \eta_2) = \iint_{\mathscr{X} \times \mathscr{X}} [\eta_1 - \eta_2]^2 (\bar{B}(u, \rho(u, v)))(\eta_1(du)\eta_1(dv) + \eta_2(du)\eta_2(dv)),$$

where η_1 and η_2 are the probability measures on Banach space $(\mathscr{X}, \rho)$. Ball Covariance and Ball Divergence are related, but fundamentally different. Ball divergence is the ball probability difference between η_1 and η_2 on the support set of η_1 and η_2, but Ball Covariance is the product ball probability difference between θ and $\mu \otimes \nu$ on the support set of θ.

Both Ball Covariance and Ball divergence can be defined in the general metric space. According to the generalized Banach-Mazur theorem (Kleiber and Pervin 1969) as stated in the supplementary material, we can isometrically embedding a metric space into a Banach space. Thus, the properties of Ball Covariance and Ball divergence can be studied in Banach spaces.

Next, we present the independence-zero equivalence property of the Ball Covariance, which requires the completeness because we need to ensure that the limit of the points on the centers and circumferences of the balls is contained in the space. It is easy to see that $\mathbf{BCov}_\omega(X, Y) \geq 0$ and the equality holds if X and Y are independent; that is, if $\theta = \mu \otimes \nu$, then $\mathbf{BCov}_\omega(X, Y) = 0$. The following theorem delineates what happens when $\mathbf{BCov}_\omega(X, Y) = 0$.

Theorem 2.1.1 (Independence-zero equivalence property). Let S_θ, S_μ and S_ν denote the support sets of θ, μ and ν, respectively. $\mathbf{BCov}_\omega(X, Y) = 0$ implies $\theta = \mu \otimes \nu$ if one of the following conditions is met.

(a) $\mathscr{X} \times \mathscr{Y}$ is a finite dimensional Banach space with $S_\theta = S_\mu \times S_\nu$.
(b) $\theta = a_1\theta_d + a_2\theta_a$, where a_1 and a_2 are positive constants, θ_d is a discrete measure, and θ_a is an absolutely continuous measure with a continuous Radon–Nikodym derivative with respect to the Gaussian measure.

2.2. Ball Correlation

In this section, we investigate the basic properties of Ball Covariance before defining Ball Correlation. Let $\delta^X_{ij,k} := I(X_k \in \bar{B}_\rho(X_i, X_j))$, which indicates whether X_k is located in the closed ball $\bar{B}_\rho(X_i, X_j)$, and $\delta^X_{ij,kl} = \delta^X_{ij,k}\delta^X_{ij,l}$, which is an indicator for whether both of X_k and X_l fall into the closed ball $\bar{B}_\rho(X_i, X_j)$. Also let $\xi^X_{ij,klst} = (\delta^X_{ij,kl} + \delta^X_{ij,st} - \delta^X_{ij,ks} - \delta^X_{ij,lt})/2$. Similarly, define the notations $\delta^Y_{ij,k}$, $\delta^Y_{ij,kl}$ and $\xi^Y_{ij,klst}$ for Y. The following proposition states that $\mathbf{BCov}^2_\omega(X,Y)$ is the expectation of the product of two weight functions: one with X's only and the other with Y's only.

Proposition 2.2.1 (Separability property). Let $(X_i, Y_i), i = 1, 2, \ldots, 6$ be *iid* samples from θ. We have

$$\mathbf{BCov}^2_\omega(X,Y) = E\{\xi^X_{12,3456}\xi^Y_{12,3456}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2)\}. \quad (1)$$

Distance covariance (Székely, Rizzo, and Bakirov 2007) and conditional distance covariance (Wang et al. 2015) also satisfy the separability property. This separability property leads to the Cauchy–Schwarz type inequality below.

Let

$$\mathbf{BCov}^2_\omega(X) := \mathbf{BCov}^2_\omega(X,X) = E(\xi^X_{12,3456}\omega_1(X_1,X_2))^2$$

and

$$\mathbf{BCov}^2_\omega(Y) := \mathbf{BCov}^2_\omega(Y,Y) = E(\xi^Y_{12,3456}\omega_2(Y_1,Y_2))^2,$$

respectively, then we have

Proposition 2.2.2 (Cauchy–Schwarz type inequality).

$$\mathbf{BCov}^2_\omega(X,Y) \le \mathbf{BCov}_\omega(X)\mathbf{BCov}_\omega(Y). \quad (2)$$

Furthermore, if θ is nondegenerated and $\delta^X_{12,3} = \delta^Y_{12,3}$, $\omega_1(X_1,X_2) = \omega_2(Y_1,Y_2)$ a.s., then the equality above holds.

Proposition 2.2.2 implies that we can standardize $\mathbf{BCov}^2_\omega(X,Y)$. Hence, we can define Ball Correlation as follows.

Definition 2.2.1. The Ball Correlation $\mathbf{BCor}_\omega(X,Y)$ is defined as the square root of

$$\mathbf{BCor}^2_\omega(X,Y) := \mathbf{BCov}^2_\omega(X,Y)/\sqrt{\mathbf{BCov}^2_\omega(X)\mathbf{BCov}^2_\omega(Y)},$$

if $\mathbf{BCov}_\omega(X)\mathbf{BCov}_\omega(Y) > 0$, or 0 otherwise.

Some additional properties of $\mathbf{BCov}_\omega(X,Y)$ and $\mathbf{BCor}_\omega(X,Y)$ shall be investigated and presented in part B of the supplementary material. In the special case where $\omega_1 = \omega_2 \equiv 1$, denote $\mathbf{BCov}_\omega(X,Y)$ and $\mathbf{BCor}_\omega(X,Y)$ by $\mathbf{BCov}(X,Y)$ and $\mathbf{BCor}(X,Y)$, respectively.

2.3. Empirical Ball Covariance and Ball Correlation

We now introduce the empirical Ball Covariance and Ball Correlation. Consider an observed random sample $(\mathbf{X},\mathbf{Y}) = \{(X_k, Y_k) : k = 1,\ldots,n\}$ from the joint distribution θ. Let $\hat\omega_{1,n}$ and $\hat\omega_{2,n}$ be the estimates of the weight functions ω_1 and ω_2, respectively. Denote

$$\Delta^{XY}_{ij,n} = \frac{1}{n}\sum_{k=1}^n \delta^X_{ij,k}\delta^Y_{ij,k},\ \Delta^X_{ij,n} = \frac{1}{n}\sum_{k=1}^n \delta^X_{ij,k},\ \Delta^Y_{ij,n} = \frac{1}{n}\sum_{k=1}^n \delta^Y_{ij,k},$$

then we can obtain the empirical Ball Covariance and Correlation as follows.

Definition 2.3.1. The empirical Ball Covariance $\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y})$ is defined as the square root of

$$\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y}) := \frac{1}{n^2}\sum_{i,j=1}^n (\Delta^{XY}_{ij,n} - \Delta^X_{ij,n}\Delta^Y_{ij,n})^2 \times \hat\omega_{1,n}(X_i,X_j)\hat\omega_{2,n}(Y_i,Y_j).$$

Note that $n\Delta^X_{ij,n}$ is the rank of $\rho(X_i,X_j)$ among $\{\rho(X_i,X_k), k = 1,\ldots,n\}$ and $n\Delta^Y_{ij,n}$ is the rank of $\zeta(Y_i,Y_j)$ among $\{\zeta(Y_i,Y_k), k = 1,\ldots,n\}$. Thus, our test belongs to a class of metric rank tests and possesses the properties of a general rank test such as robustness.

Recall that the weight is the function defined on the support set of θ. In applications, the weight may also directly depend on the sample size n. Below are some examples of the weight function.

Remark 2.3.1. (i) If $\hat\omega_{1,n} \equiv \hat\omega_{2,n} \equiv 1$, we simplify the notation $\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y})$ as $\mathbf{BCov}_n(\mathbf{X},\mathbf{Y})$.
(ii) If $\hat\omega_{1,n} = (\rho(X_i,X_j))^\alpha$ and $\hat\omega_{2,n} = (\zeta(Y_i,Y_j))^\beta$, where $\alpha,\beta \in \mathbb{R}$, $\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y})$ is referred to as $\mathbf{BCov}_{d,n}(\mathbf{X},\mathbf{Y})$.
(iii) If $\hat\omega_{1,n} = (\Delta^X_{ij,n})^\alpha$ and $\hat\omega_{2,n} = (\Delta^Y_{ij,n})^\beta$, $\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y})$ is referred to as $\mathbf{BCov}_{\Delta,n}(\mathbf{X},\mathbf{Y})$.
(iv) If $\hat\omega_{1,n} = \{\Delta^X_{ij,n}(1-\Delta^X_{ij,n})\}^{-1}$ and $\hat\omega_{2,n} = \{\Delta^Y_{ij,n}(1-\Delta^Y_{ij,n})\}^{-1}$, $\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y})$ is asymptotically equivalent to HHG; See Proposition B.1 in the supplementary material.

Remark 2.3.2. It is natural to extend $\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y})$ to measure mutual independence between random vectors. For example, we can redefine

$$\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y},\mathbf{Z}) = \frac{1}{n^2}\sum_{i,j=1}^n (\Delta^{XYZ}_{ij,n} - \Delta^X_{ij,n}\Delta^Y_{ij,n}\Delta^Z_{ij,n})^2 \times \hat\omega_{1,n}(X_i,X_j)\hat\omega_{2,n}(Y_i,Y_j)\hat\omega_{3,n}(Z_i,Z_j)$$

to measure whether $P_{XYZ} = P_XP_YP_Z$ without increasing the computational complexity.

Similarly, we introduce $\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{X})$ and $\mathbf{BCov}^2_{\omega,n}(\mathbf{Y},\mathbf{Y})$, which are the empirical versions of $\mathbf{BCov}^2_\omega(\mathbf{X},\mathbf{X})$ and $\mathbf{BCov}^2_\omega(\mathbf{Y},\mathbf{Y})$. Now, we define the empirical Ball Correlation as follows.

Definition 2.3.2. The empirical Ball Correlation $\mathbf{BCor}_{\omega,n}(\mathbf{X},\mathbf{Y})$ is defined as the square root of

$$\mathbf{BCor}^2_{\omega,n}(\mathbf{X},\mathbf{Y}) := \mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y})/\sqrt{\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{X})\mathbf{BCov}^2_{\omega,n}(\mathbf{Y},\mathbf{Y})},$$

if $\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{X})\mathbf{BCov}^2_{\omega,n}(\mathbf{Y},\mathbf{Y}) > 0$, or 0 otherwise.

Proposition 2.2.1 implies that we can also take

$$\mathcal{V}_n(\mathbf{X},\mathbf{Y}) := \frac{1}{n^6}\sum_{i,j,k,l,u,v=1}^n \xi^X_{ij,kluv}\xi^Y_{ij,kluv}\hat\omega_{1,n}(X_i,X_j)\hat\omega_{2,n}(Y_i,Y_j)$$

as the sample version for $\mathbf{BCov}^2_\omega(X,Y)$. In fact, we have the following proposition.

Proposition 2.3.1. $\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y}) = \mathcal{V}_n(\mathbf{X},\mathbf{Y})$.

Using Proposition 2.3.1, we can obtain the asymptotic consistency of $\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y})$ and $\mathbf{BCor}_{\omega,n}(\mathbf{X},\mathbf{Y})$ under the following condition (**C**) on the weights:

(**C**)$\hat{\omega}_{1,n}$ and $\hat{\omega}_{2,n}$ uniformly converge to ω_1 and ω_2 with $E(\omega_1\omega_2) < \infty$, respectively.

Theorem 2.3.1. Under Condition (**C**), we have

$$\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y}) \xrightarrow[n\to\infty]{a.s.} \mathbf{BCov}_{\omega}(X,Y) \text{ and}$$
$$\mathbf{BCor}_{\omega,n}(\mathbf{X},\mathbf{Y}) \xrightarrow[n\to\infty]{a.s.} \mathbf{BCor}_{\omega}(X,Y).$$

3. Test of Independence

3.1. Asymptotics of Test Statistic

First, we derive the asymptotic distributions of $\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y})$ under the null and alternative hypotheses by using the theory of V-statistics. Let

$$\psi(W_1,\ldots,W_6) = \xi^X_{12,3456}\xi^Y_{12,3456}\omega_1(X_1,X_2)\omega_2(Y_1,Y_2),$$
$$\bar{\psi}(W_1,\ldots,W_6) = \frac{1}{6!}\sum \psi(W_{\sigma(1)},\ldots,W_{\sigma(6)})$$

in which the sum extends over all permutations $(\sigma(1),\ldots,\sigma(6))$ of distinct integers chosen from $\{1,\ldots,6\}$, and $\Sigma = 36\mathrm{Var}(E(\bar{\psi}(W_1,\ldots,W_6)|W_1))$. Furthermore, denote λ_v and f_v as the eigenvalues and eigenfunctions of the symmetric function $E[\bar{\psi}(W_1,\ldots,W_6)|W_1,W_2]$, where $E[\bar{\psi}(W_1,\ldots,W_6)|W_1,W_2]$ has the following spectral decomposition:

$$E[\bar{\psi}(W_1,\ldots,W_6)|W_1,W_2] = \sum_{v=1}^{\infty}\lambda_v f_v(W_1)f_v(W_2).$$

We further investigate the asymptotic properties of $\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y})$ under the null and alternative hypotheses.

Theorem 3.1.1. Under Condition (**C**) and the null hypothesis, we have

$$n\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y}) \xrightarrow[n\to\infty]{d} \sum_{v=1}^{\infty}\lambda_v Z_v^2,$$

where Z_v are independent standard normal random variables.

Theorem 3.1.2. Under Condition (**C**) and the alternative hypothesis, we have

$$\sqrt{n}(\mathbf{BCov}^2_{\omega,n}(\mathbf{X},\mathbf{Y}) - \mathbf{BCov}^2_{\omega}(X,Y)) \xrightarrow[n\to\infty]{d} N(0,\Sigma).$$

Theorem 3.1.1 implies that we cannot directly use the null distribution for Ball Covariance test since λ_v depends on the unknown null distribution. Thus, a permutation procedure is needed for evaluating the p-values in general. Next, applying Theorem 2.1.1, we can easily obtain Theorem 3.1.3 which implies that Ball Covariance test is consistent against the dependence alternatives without any moment conditions.

Theorem 3.1.3. Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be iid. samples of (X,Y). Under Condition (**C**) and the conditions of Theorem 2.1.1, $\mathbf{BCov}_{\omega,n}(\mathbf{X},\mathbf{Y})$ can serve as a test statistic for independence which is consistent against the alternatives.

4. Simulation Studies

In this section, we assess the Type I errors and the powers of four independence tests: $\mathbf{BCov}_n$, $\mathbf{BCov}_{\Delta,n}$ with $\alpha = \beta = -1$, HHG (Heller, Heller, and Gorfine 2013) and dCov (Székely, Rizzo, and Bakirov 2007). In all of the following settings, we repeat each example 1000 times and set the sample size from 30 to 100. The nominal significance level is at 0.05.

4.1. Multivariate Case

We begin with the comparison of Type I error rates in Examples 4.1.1–4.1.4 and evaluate the power of four methods in Examples 4.1.5–4.1.10.

- Example 4.1.1: X, Y are independent from the standard normal distribution $N(0,1)$.
- Example 4.1.2: X, Y are independent from the binomial distribution $B(10,0.5)$.
- Example 4.1.3: Z_1,Z_2 are independent from the binomial distribution $B(10,0.5)$, and
$$X = (Z_1)^2 + Z_1,\ Y = (Z_2)^2 + Z_2.$$
- Example 4.1.4: (X,Y) are from the 10-dimensional multivariate normal distribution with $\mu = \mathbf{0}$, $\mathrm{cov}(X_i,X_i) = \mathrm{cov}(Y_i,Y_i) = 1, i = 1,\ldots,5$, $\mathrm{cov}(X_i,X_j) = \mathrm{cov}(Y_i,Y_j) = 0.2$, $1 \le i < j \le 5$ and $\mathrm{cov}(X_i,Y_j) = 0, i,j = 1,\ldots,5$.
- Example 4.1.5: $Z_1,\ldots,Z_4$ are independent from the binomial distribution $B(10,0.5)$, $Z_5 \sim B(20,0.5)$,
$$Y = \sin(Z_1)(Z_2+Z_5)^2/(Z_3+1) + \log(Z_4+Z_5+1),$$
where $X = (Z_1,\ldots,Z_4)$.
- Example 4.1.6: Z_1,Z_2,Z_3,Z_4 are independent from the standard normal distribution $N(0,1)$, and
$$Y_1 = [4-(Z_1)^2]Z_3Z_4 + Z_2(Z_4)^2,$$
$$Y_2 = [4-(Z_2)^2]Z_3Z_4 + Z_1(Z_4)^2,$$
where $X = (Z_1,Z_2,Z_3)$, $Y = (Y_1,Y_2)$.
- Example 4.1.7: X, Y are from the 10-dimensional multivariate normal distribution with mean $\mu = \mathbf{0}$, $\mathrm{cov}(X_i,X_i) = \mathrm{cov}(Y_i,Y_i) = 1, i = 1,\ldots,5$, $\mathrm{cov}(X_i,X_j) = \mathrm{cov}(Y_i,Y_j) = 0.2$, $1 \le i < j \le 5$ and $\mathrm{cov}(X_i,Y_j) = 0.2, i,j = 1,\ldots,5$.
- Example 4.1.8: $Z_1,\ldots,Z_6$ are independent from the standard normal distribution $N(0,1)$,
$$Y_1 = \sqrt{|Z_1+(Z_4)^2|} + (Z_5)^2,$$
$$Y_2 = Z_2Z_5 + \tanh(Z_6),$$
where $X = (Z_1,Z_2,Z_3,Z_4,Z_6)$, $Y = (Y_1,Y_2)$ and $\tanh(\cdot)$ is the hyperbolic tangent function.
- Example 4.1.9: $Z_1,\ldots,Z_5$ are independent from the t distribution with 3 degrees of freedom and
$$Y_1 = (Z_1+Z_2)^3 + (Z_5)^2,$$
$$Y_2 = Z_3Z_4 + \tanh(Z_5),$$
$$Y_3 = (Z_1)^2 + Z_5,$$
where $X = (Z_1,Z_5)$, $Y = (Y_1,Y_2,Y_3)$ and $\tanh(\cdot)$ is the hyperbolic tangent function.

Table 1. Type I Error rates of four independence tests for multivariate cases.

	Example 4.1.1				Example 4.1.2			
n	30	50	70	100	30	50	70	100
dCov	0.045	0.051	0.044	0.045	0.043	0.044	0.056	0.051
HHG	0.052	0.050	0.040	0.051	0.044	0.047	0.052	0.055
$\mathbf{BCov}_n$	0.051	0.045	0.040	0.054	0.056	0.054	0.058	0.055
$\mathbf{BCov}_{\Delta,n}$	0.049	0.049	0.038	0.050	0.050	0.044	0.046	0.045
	Example 4.1.3				Example 4.1.4			
n	30	50	70	100	30	50	70	100
dCov	0.043	0.042	0.056	0.051	0.051	0.040	0.050	0.046
HHG	0.052	0.051	0.053	0.051	0.061	0.054	0.059	0.042
$\mathbf{BCov}_n$	0.060	0.047	0.060	0.054	0.052	0.050	0.050	0.042
$\mathbf{BCov}_{\Delta,n}$	0.054	0.043	0.050	0.045	0.047	0.050	0.049	0.045

Table 2. Powers of four independence tests for multivariate cases.

	Example 4.1.5				Example 4.1.6			
n	30	50	70	100	30	50	70	100
dCov	0.626	0.921	0.994	1.000	0.397	0.684	0.854	0.956
HHG	0.836	0.988	1.000	1.000	0.381	0.621	0.814	0.938
$\mathbf{BCov}_n$	0.867	0.992	1.000	1.000	0.484	0.742	0.870	0.966
$\mathbf{BCov}_{\Delta,n}$	0.933	0.998	1.000	1.000	0.587	0.900	0.980	0.999
	Example 4.1.7				Example 4.1.8			
n	30	50	70	100	30	50	70	100
dCov	0.693	0.908	0.980	0.999	0.639	0.909	0.986	1.000
HHG	0.415	0.701	0.865	0.970	0.605	0.879	0.976	0.998
BCov	0.457	0.710	0.868	0.966	0.727	0.940	0.992	1.000
$\mathbf{BCov}_{\Delta}$	0.320	0.547	0.732	0.901	0.740	0.959	0.999	1.000
	Example 4.1.9				Example 4.1.10			
n	30	50	70	100	30	50	70	100
dCov	0.397	0.598	0.739	0.884	0.144	0.170	0.201	0.242
HHG	0.408	0.689	0.821	0.965	0.298	0.542	0.723	0.862
$\mathbf{BCov}_n$	0.527	0.784	0.884	0.976	0.451	0.668	0.821	0.909
$\mathbf{BCov}_{\Delta,n}$	0.528	0.826	0.926	0.994	0.491	0.765	0.906	0.978

- Example 4.1.10: $Z_1, \ldots, Z_7$ are independent from the t distribution with 1 degree of freedom and

$$Y_1 = \tanh(Z_1 Z_2 Z_6),$$
$$Y_2 = \cos(I(Z_4 > 0)Y_1 + I(Z_4 < 0)Z_2),$$

  where $X = (Z_1, \ldots, Z_7), Y = (Y_1, Y_2)$ and $I(\cdot)$ is the indicator function.

Table 1 confirms that all methods can control the Type I error rates well. The power comparison is presented in Table 2. dCov has the lowest power in all cases except Example 4.1.7, but the power estimates of all methods generally get closer quickly and approach to 1. The performance of $\mathbf{BCov}_n$ or $\mathbf{BCov}_{\Delta,n}$ is better than dCov and HHG in all cases except Example 4.1.7. Moreover, $\mathbf{BCov}_n$, $\mathbf{BCov}_{\Delta,n}$, and HHG perform well when the distributions are heavy tail, but dCov loses the power under such distributions. $\mathbf{BCov}_n$, $\mathbf{BCov}_{\Delta,n}$ and HHG retain the robustness of common rank tests because they can be expressed in a form of ranks.

4.2. Infinite-Dimensional Case

In this subsection, we compare the performance of $\mathbf{BCov}_n$ and $\mathbf{BCov}_{\Delta,n}$ with dCov and HHG for functional data. To generate the functional data, we consider a process for the ith subject

$$Z_i(t) = \sum_{j=1}^{K} c_{ij} e_j(t), \quad i = 1, \ldots, n,$$

where $\{e_j(t), j = 1, \ldots, K\}$ are the basis functions of the Banach space. In our simulation studies, we generate each sample with 30–100 pairs of observed random processes $X(t)$ and $Y(t)$ with 50 observed points on each observed process where the observed points are equally spaced at interval $[-2, 2]$. Let $K = 25$ and set the basis functions to be B-spline functions with order 4. $\{c_{ij}, i = 1, \ldots, n; j = 1, \ldots, K\}$ are independent from the standard normal distribution.

We first construct Examples 4.2.1–4.2.2 to examine the Type I error of $\mathbf{BCov}_n$, $\mathbf{BCov}_{\Delta,n}$, dCov, and HHG. Then, we compare the power of the four methods in Examples 4.2.3–4.2.6.

- Example 4.2.1:

$$X(t) = Z^{(1)}(t), Y(t) = Z^{(2)}(t).$$

- Example 4.2.2:

$$X(t) = Z^{(1)}(t)Z^{(2)}(t)Z^{(3)}(t), Y(t) = Z^{(4)}(t)Z^{(5)}(t)Z^{(6)}(t).$$

Table 3. Type I error rates of four independence tests for Infinite dimension and manifold cases.

	Example 4.2.1				Example 4.2.2			
n	30	50	70	100	30	50	70	100
dCov	0.039	0.043	0.042	0.040	0.039	0.039	0.048	0.043
HHG	0.054	0.060	0.043	0.041	0.039	0.040	0.050	0.053
$\mathbf{BCov}_n$	0.046	0.045	0.042	0.049	0.054	0.048	0.040	0.039
$\mathbf{BCov}_{\Delta,n}$	0.039	0.047	0.035	0.054	0.043	0.047	0.028	0.041
	Example 4.3.1				Example 4.3.4			
n	30	50	70	100	30	50	70	100
dCov	0.053	0.040	0.051	0.039	0.049	0.039	0.040	0.042
HHG	0.053	0.044	0.046	0.045	0.052	0.046	0.050	0.035
$\mathbf{BCov}_n$	0.053	0.048	0.055	0.050	0.053	0.046	0.047	0.044
$\mathbf{BCov}_{\Delta,n}$	0.056	0.050	0.038	0.041	0.046	0.043	0.046	0.038

Table 4. Powers of four independence tests for infinite dimension and manifold cases.

	Example 4.2.3				Example 4.2.4			
n	30	50	70	100	30	50	70	100
dCov	0.104	0.149	0.208	0.314	0.103	0.155	0.211	0.332
HHG	0.217	0.517	0.766	0.960	0.229	0.527	0.778	0.964
$\mathbf{BCov}_n$	0.322	0.631	0.837	0.971	0.329	0.647	0.847	0.975
$\mathbf{BCov}_{\Delta,n}$	0.450	0.815	0.963	1.000	0.454	0.830	0.965	1.000
	Example 4.2.5				Example 4.2.6			
n	30	50	70	100	30	50	70	100
dCov	0.374	0.417	0.559	0.641	0.432	0.617	0.732	0.874
HHG	0.465	0.694	0.861	0.960	0.420	0.673	0.820	0.938
$\mathbf{BCov}_n$	0.608	0.791	0.921	0.979	0.523	0.744	0.862	0.957
$\mathbf{BCov}_{\Delta,n}$	0.586	0.805	0.932	0.986	0.447	0.702	0.845	0.953
	Example 4.3.2				Example 4.3.3			
n	30	50	70	100	30	50	70	100
dCov	0.062	0.070	0.063	0.071	0.105	0.267	0.413	0.669
HHG	0.104	0.151	0.219	0.325	0.317	0.644	0.793	0.943
$\mathbf{BCov}_n$	0.119	0.170	0.237	0.372	0.476	0.688	0.804	0.950
$\mathbf{BCov}_{\Delta,n}$	0.093	0.132	0.182	0.321	0.511	0.716	0.854	0.959
	Example 4.3.5				Example 4.3.6			
n	30	50	70	100	30	50	70	100
dCov	0.357	0.649	0.837	0.951	0.259	0.483	0.631	0.829
HHG	0.362	0.677	0.883	0.977	0.300	0.603	0.807	0.965
$\mathbf{BCov}_n$	0.380	0.695	0.887	0.980	0.319	0.620	0.786	0.956
$\mathbf{BCov}_{\Delta,n}$	0.424	0.751	0.926	0.991	0.356	0.683	0.864	0.985
	Example 4.3.7				Example 4.3.8			
n	30	50	70	100	30	50	70	100
dCov	0.053	0.063	0.053	0.040	0.040	0.048	0.049	0.054
HHG	1.000	1.000	1.000	1.000	0.967	1.000	1.000	1.000
$\mathbf{BCov}_n$	1.000	1.000	1.000	1.000	0.760	1.000	1.000	1.000
$\mathbf{BCov}_{\Delta,n}$	1.000	1.000	1.000	1.000	0.985	1.000	1.000	1.000

- Example 4.2.3:
$$X(t) = Z^{(1)}(t), Y(t) = 2\cos[\pi X(t)] + \varepsilon(t),$$
where $\varepsilon(t)$ is a Gaussian process.
- Example 4.2.4:
$$X(t) = Z^{(1)}(t), Y(t) = 2\cos[\pi X(t)] + \frac{1}{20}X(t) + \varepsilon(t),$$
where $\varepsilon(t)$ is a Gaussian process.
- Example 4.2.5:
$$X(t) = [Z^{(1)}(t)]^3 + Z^{(1)}(t) + 2Z^{(2)}(t)Z^{(3)}(t),$$
$$Y(t) = [Z^{(2)}(t)]^4 + \cos[Z^{(2)}(t)] + Z^{(1)}(t)Z^{(2)}(t).$$
- Example 4.2.6:
$$X(t) = [Z^{(1)}(t) + Z^{(2)}(t)]^3 + [Z^{(3)}(t)]^2,$$
$$Y(t) = [Z^{(1)}(t)]^2 + Z^{(3)}(t).$$

Table 3 presents the Type I error rates of the four methods. $\mathbf{BCov}_n$, $\mathbf{BCov}_{\Delta,n}$, dCov and HHG control the Type I errors well around 0.05. Table 4 compares the power for Examples 4.2.3–4.2.6. For most of the nonlinear cases, $\mathbf{BCov}_n$ and $\mathbf{BCov}_{\Delta,n}$ are more powerful than HHG and dCov.

4.3. Manifold-Valued Data

In this subsection, we use manifold examples to illustrate the performance of $\mathbf{BCov}_n$ and $\mathbf{BCov}_{\Delta,n}$, relative to dCov and

HHG. Apart from Examples 4.3.1 and 4.3.4, all examples address various dependence cases.

First, we consider the spherical coordinate of unit sphere S^2. Let ϑ and ϕ represent the inclination (or elevation) and azimuth, respectively, where $\vartheta \in (-\infty, \infty)$ and $\phi \in (-\infty, \infty)$. We calculate the spherical distance for Y and Euclidean distance for X in Examples 4.3.1–4.3.2. Specifically, the simulation data $\{(X, Y) \in R^d \times R^3\}$ are generated as follows:

- Example 4.3.1: $Z_i, i = 1, \ldots, 5$ are independent from the uniform distribution $U(-\pi, \pi)$,

$$\vartheta \sim U(-\pi,\pi), \phi \sim U(-\pi,\pi),$$
$$X = (Z_1, \ldots, Z_5), Y = (\sin(\phi)\cos(\vartheta), \sin(\phi)\sin(\vartheta), \cos(\phi)).$$

- Example 4.3.2: $Z_i, i = 1, \ldots, 8$ are independent from the uniform distribution $U(-\pi, \pi)$,

$$\vartheta \sim U(-\pi,\pi), \phi = Z_3 I(Z_1 Z_2 < 0),$$
$$X = (Z_1, \ldots, Z_8),$$
$$Y = (\sin(\phi)\cos(\vartheta), \sin(\phi)\sin(\vartheta), \cos(\phi)).$$

Next, we consider the von Mises-Fisher distribution $M_p(u, \kappa)$, for which the data can be spherical or hyper-spherical with direction u and concentration degree κ. In Example 4.3.3, we generate u from the multivariate normal distribution $N(0, I_6)$ and calculate the spherical distance for Y and X, respectively.

- Example 4.3.3:

$$(Z_1, \ldots, Z_6) \sim M_6(u, 15),$$
$$\vartheta = \cos(Z_1)\sqrt{|Z_2 - Z_1|}, \phi = ((Z_1)^2 Z_2 + \tanh(Z_4) Z_3)_+^2,$$
$$X = (Z_1, \ldots, Z_6),$$
$$Y = (\sin(\phi)\cos(\vartheta), \sin(\phi)\sin(\vartheta), \cos(\phi)).$$

Now, we focus on the performances of four methods with spherical data and symmetric positive matrix in Examples 4.3.4-4.3.6 where Riemannian and great-circle distances are used to measure the dissimilarities of two symmetric positive matrices and two spherical or hyper-spherical observations. In the first two case, we consider the 3×3 symmetric positive matrix whose nondiagonal elements are all equal to γ and spherical data from the von Mises–Fisher distribution whose directional and concentration parameters are $u = (1, 0, 0)$ and $\kappa = 2$. And we construct the dependence relationship between spherical data and symmetric positive matrix by parameter γ.

- Example 4.3.4:

$$X \sim M_3(u, 2), \gamma \sim U(-0.99, 0.99).$$

- Example 4.3.5:

$$(Z_1, Z_2, Z_3) \sim M_3(u, 2), \epsilon \sim N(0, 1)$$
$$\vartheta = \arctan(Z_1/Z_2), X = (Z_1, Z_2, Z_3), \gamma = \frac{1}{1 + e^{-(\vartheta+\epsilon)}},$$

where $\arctan(\cdot)$ is the inverse tangent function.

Furthermore, we consider another form of symmetric positive matrix as follows:

$$\operatorname{diag}(A, A, A, A, A)_{10\times 10}, A = \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix}_{2\times 2}.$$

Hyper-spherical data follow von Mises–Fisher distribution with $u = (1, 0, 0, 0, 0)$ and $\kappa = 2$. Similarly, we construct the dependence relationship between spherical data and symmetric positive matrix by parameter γ. Specifically, the data $\{(X, Y) \in R^5 \times R^{10\times 10}\}$ are simulated as follows:

- Example 4.3.6:

$$(Z_1, \ldots, Z_5) \sim M_5(u, 2), \epsilon_1, \epsilon_2 \overset{\text{iid}}{\sim} N(0, 1),$$
$$\phi_1 = Z_1 Z_2 + \frac{1}{10}\epsilon_1, \phi_2 = Z_3 Z_4 + \frac{1}{10}\epsilon_2,$$
$$X = (Z_1, \ldots, Z_5), \gamma = \frac{1}{2}[\phi_1 I(\phi_1 > 0) + \phi_2 I(\phi_2 > 0)].$$

Finally, we consider two simple cases where distance covariance totally loses the efficiency but Ball Covariance is still able to detect the dependence. In the following examples, we calculate the spherical distance for Y and the Euclidean distance for X.

- Example 4.3.7: $X \sim B(1, 0.5)$, $Y = (r, \phi)$ is a two-dimensional polar coordinate and $P(r, \phi|X)$ is the conditional probability given X,

$$P((r,\phi) = (1,0)|X = 1) = P((r,\phi) = (1,\pi)|X = 1) = \frac{1}{2},$$
$$P((r,\phi) = (1,\frac{\pi}{2})|X = 0) = P((r,\phi) = (1,\frac{3}{2}\pi)|X = 0) = \frac{1}{2}.$$

- Example 4.3.8: $X \sim B(1, 0.5)$, $Y = (r, \phi)$ is a two-dimensional polar coordinate and $f(r, \phi|X)$ is the conditional density function given X,

$$f(r,\phi|X = 1) = \frac{1}{\pi}, \quad r = 1, 0 \le \phi \le \frac{\pi}{2}, \pi \le \phi \le \frac{3}{2}\pi,$$
$$f(r,\phi|X = 0) = \frac{1}{\pi}, \quad r = 1, \frac{\pi}{2} \le \phi \le \pi, \frac{3}{2}\pi \le \phi \le 2\pi.$$

The Type I error rates for the manifold-valued data are summarized in Table 3 and the empirical power results are given in Table 4. We can see from Table 3 that the Type I error rates of $\mathbf{BCov}_n$, $\mathbf{BCov}_{\Delta,n}$, dCov, and HHG are under reasonable control. Table 4 suggests that the four tests can detect the relationship in a manifold space. Either $\mathbf{BCov}_n$ or $\mathbf{BCov}_{\Delta,n}$ performs relatively better than HHG and dCov. More importantly, dCov totally loses its power in Examples 4.3.7 and 4.3.8, which is not surprising because the distances are not of strong negative type.

Overall, our simulation results demonstrate that the independence test based on Ball Covariance and the existing tests have comparable Type I error rates and power, and in most cases, the empirical results from all tests appear in a reasonable range. However, our proposed test is generally more powerful, and hence is expected to have a practical advance over the existing tests.

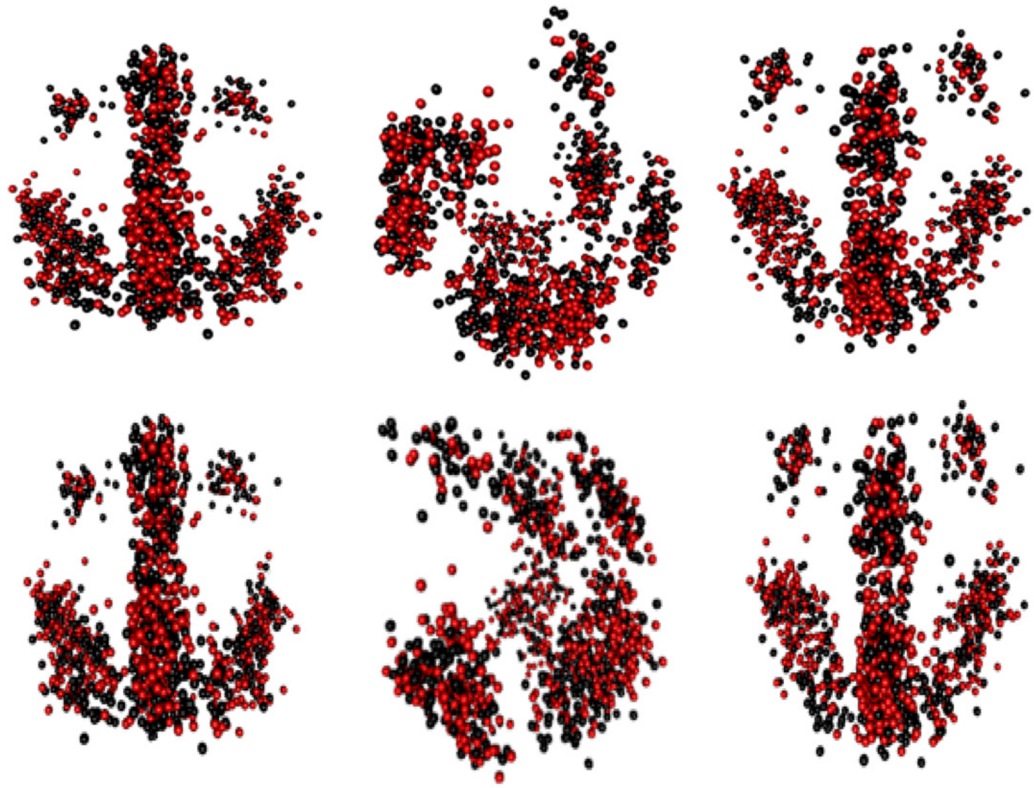

Figure 1. 3D scatterplot of cortical surface shape data for different ages (top) or handedness (bottom) : front view (left), side view (middle), and top view (right), where the red points correspond to the older or left handedness subjects, and the black points to the younger or right handedness subjects.

5. Real Data Analysis

5.1. Landmark-Based Cortical Surface Data

In this subsection, we demonstrate how to apply our method to study normal adult human brains. Using landmarks distributed over the cortical surface, we are interested in identifying factors that influence the cortical surface shape. We reanalyze a public dataset reported in Free et al. (2001), which is available in R package *shapes*.

The 3D landmark-based morphometric data contain 58 normal control subjects without a history of neurological or psychiatric illness. Each subject has a cortical surface with 24 landmarks and four covariates, including gender, age, handedness, and the interaction of handedness and gender. Handedness is determined by a 13-point questionnaire about hand usage in common situations. Subjects are identified as right-handers or non-right-handers. Twenty-seven subjects are females (19 right-handed, 8 non-right-handed). Thirty-one subjects are males (24 right-handed, 7 non-right-handed). The age range of subjects is from 16 to 59 years (mean 33 years, median 31 years). Figure 1 shows the 3D scatterplot of the cortical surface shape data for different ages or handedness.

To apply our method, we follow shape theory (Kendall et al. 2009) and define a metric space for the cortical surface shapes as follows:

1. For each subject, let $x = (x_1, \dots, x_{24})$ be a 3×24 matrix, where x_i is the data for the ith landmark, $i = 1, \dots, 24$.
2. Let
$$z = (x_1 - \bar{x}, \dots, x_{24} - \bar{x})/||(x_1 - \bar{x}, \dots, x_{24} - \bar{x})||,$$
where $\bar{x} = \frac{1}{24}\sum_{k=1}^{24} x_k$ and $\|\cdot\|$ is the Frobenius norm. This preshape process is to remove the effect of translation and scaling. Let S_3^{24} contain all z's defined here. This is the so-called preshape space.
3. Let $\pi(x) = \{Az : A \in SO(3)\}$, where $SO(3)$ is the set of all orthogonal 3×3 matrices with determinant being 1. $\pi(x)$ is called the shape of x. Let Σ_3^{24} contain all $\pi(x)$'s.

Now, for any x, y with their preshapes z, w, the geodesic distance between two shapes $\pi(x)$ and $\pi(y)$ is given by

$$d_{gs}(\pi(x), \pi(y)) = \arccos(\max_{A \in SO(3)} \mathrm{Trace}(Awz')).$$

Then, Σ_3^{24} is a metric space. As for the covariates, Euclidean distance can characterize the difference between the samples.

The p-values and the adjusted p-values with Benjamini–Hochberg (BH) procedure of the four methods are presented in Table 5. With BH procedure, dCov detected no covariates associated with the surface shape, but HHG and $\mathbf{BCov}_n$ revealed the association between handedness and the surface shape, as also reported in Free et al. (2001). Interestingly, $\mathbf{BCov}_n$ and $\mathbf{BCov}_{\Delta,n}$ also identified the association between age and the surface shape, which was not discovered by Free et al. (2001). It is important to note that age is well-documented in the literature (Passe et al. 1997; Grady et al. 2006; Watanabe et al. 2013) to cause changes to the brain size, vasculature and cognition. Specifically, the brain shrinks at all levels from molecules to morphology with the age increasing. The result underscores the usefulness of our new approach.

To further understand our results, we visualize the shape difference at different ages or handedness. First, we divide the data

Table 5. The p-values (adjusted p-values under BH procedure) of four independence tests for 3D landmark-based cortical surface data.

Covariate	dCov	HHG	$\mathrm{BCov_n}$	$\mathrm{BCov}_{\Delta,n}$
Gender	0.666(0.666)	0.766(0.766)	0.777(0.777)	0.791(0.791)
Age	0.028(0.112)	0.040(0.080)	0.021(0.043)	0.011(0.044)
Handedness	0.339(0.666)	0.005(0.020)	0.007(0.028)	0.044(0.088)
Handedness × Gender	0.600(0.666)	0.119(0.159)	0.126(0.169)	0.203(0.271)

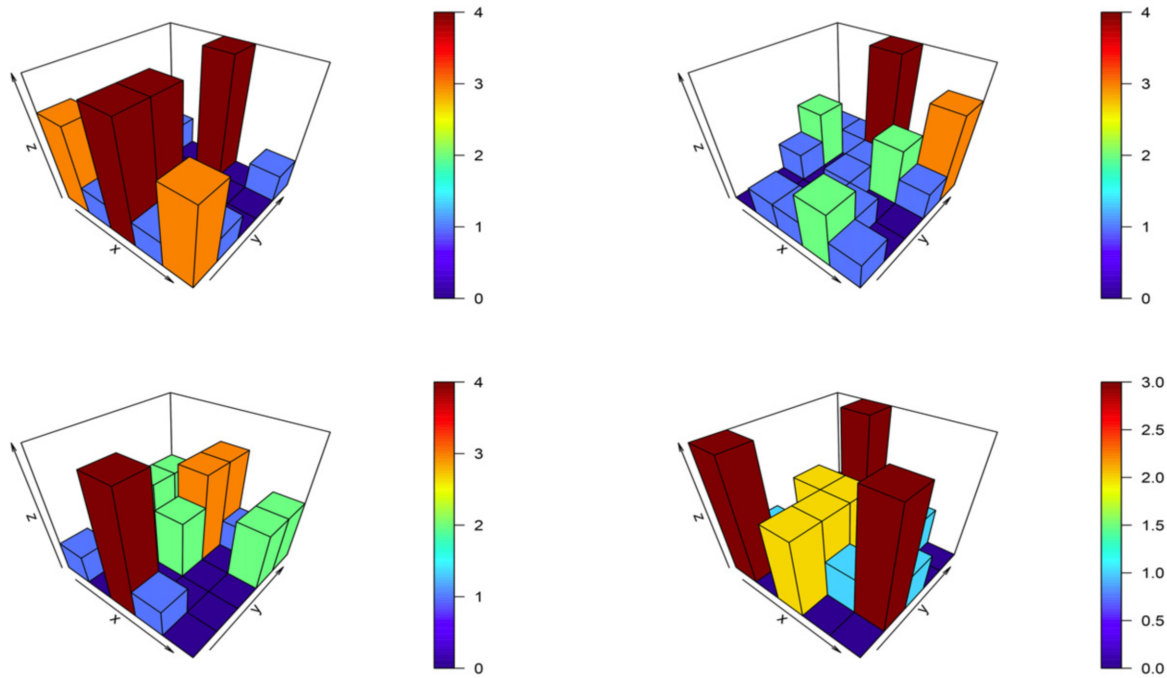

Figure 2. The top-left, top-right, bottom-left, and bottom-right correspond to younger, older, left-handedness, and right-handedness observations, where the x-axis and y-axis are the coordinates after applying the MLE transformation—a nonlinear dimension reduction procedure.

into two groups according to the median age or handedness. Second, we use modified locally linear embedding (MLLE), a manifold learning technique proposed by Zhang and Wang (2007), to project the 3D cortical surface data in each group to a two-dimensional space. Finally, we plot 2D histogram to visualize the MLLE processed data. These are done separately based on the median age and for handedness, giving rise to a total of four comparison groups. The four panels in Figure 2 visualize the 3D histograms of the samples in the four groups, where the top-left, top-right, bottom-left and bottom-right correspond to younger, older, left-handedness, right-handedness observations, respectively. The x-axis and y-axis in Figure 2 are the coordinates after we apply the MLLE transformation—a nonlinear dimension reduction procedure – to the cortical surface shape data. It is clear from Figure 2 that the histograms of the CS shapes are different between the two age groups and between the two handedness groups.

5.2. Gut Microbiota 16S rRNA Data

In this subsection, we reanalyze a processed 16S rRNA data reported by Guo et al. (2016), which were originally collected for discrimination analysis of gout patients and healthy humans. This dataset, including the sequence data, preprocessed sequence data as well as clinical information, is publically available in the MG-RAST database and nature website (*http://www.nature.com/srep*). We are interested in whether the intestinal microbiota structure impacts on body mass index (BMI) of healthy individuals.

The gut microbiota dataset contains data from 83 participants. Each participant had a total of 3684 Operational Taxonomic Units (OTUs) and twenty-two body measurements as covariates, such as disease status, BMI, gender, and age. Forty-one participants were gouty (24 males, 17 females) while forty-two participants healthy (17 males, 25 females). The ages of the participants varied from 27 to 75 years old (the mean and median were both 49 years). We extract the count of OTUs and BMI values of healthy participants for further analysis.

An important feature of the OTUs data is the frequent occurrences of zero. Thus, we define a metric space for the intestinal microbiota dataset in two steps.

1. For each participant, let $z = (z_1, \ldots, z_D)$, where z_i is the count of the ith OTU, and apply the transformation φ to z, where

$$\varphi : (z_1, \ldots, z_D) \to (u_1, \ldots, u_D),\ u_i = z_i / \sum_{i=1}^{D} z_i.$$

   Therefore, $u = (u_1, \ldots, u_D)$ becomes a typical D-part composition.
2. To deal with the frequent zero measurements, we use the square root transformation (Napier 2014) to map u onto the surface of the $(D - 1)$-dimensional spherical space Ω as

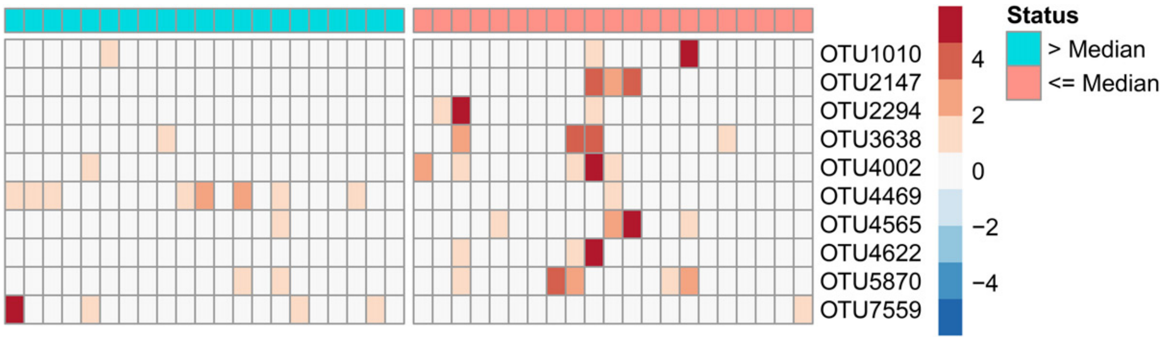

Figure 3. Low (left) and high (right) BMI group heatmap of the 16S rRNA data after feature screening procedure and scale transformation on each OTU, where each row and column corresponds to each OTU and observation, respectively.

follows:

$$u^* = s(u) = (\sqrt{u_1}, \ldots, \sqrt{u_D}).$$

After these two steps, for convenience, we refer the preprocessed compositions u^* as $s(\varphi(z))$.

Now, for participants x, y with their compositions z, w, respectively, the geodesic distance between the two preprocessed compositions $s(z), s(w)$ is given by

$$d_{gs}(z, w) = \arccos(|s(z)' s(w)|).$$

Then, Ω is a metric space. For BMI, Euclidean distance can characterize the difference between the samples.

Table 6 presents the p-values of the four methods. We can see that dCov and HHG fail to detect the association between BMI of healthy participants and the intestinal microbiota structure, but both $\mathbf{BCov}_n$ and $\mathbf{BCov}_{\Delta,n}$ reveal the relationship, which was previously discovered by Haro et al. (2016). To further understand our results, we select top 10 OTUs which show the strongest correlation with BMI (Kendall-Tau correlation test), and later, visualize the compositions difference by different BMI groups. We dichotomize the participants into two groups according to whether their BMI values exceed the median BMI. Furthermore, we graph the scaled component proportions of the two groups. It is clear from Figure 3 that the intestinal microbiota structures are different in the two BMI groups.

Table 6. The p-values of dCov, HHG, $\mathbf{BCov}_n$, and $\mathbf{BCov}_{\Delta,n}$ for 16S rRNA gut microbiota data.

Covariate	dCov	HHG	BCov_n	$\mathrm{BCov}_{\Delta,n}$
BMI	0.320	0.050	0.042	0.025

6. Conclusion

We proposed a novel nonparametric dependence measure: Ball Covariance for metric or Banach spaces. Our measure was defined by means of only metric or norm, which is very different from existing correlations in Hilbert spaces. Such a dependence measure is useful for shape analysis because the spaces underlying shape data are often not Hilbert spaces. We illustrated the usefulness of Ball Covariance with analyzing two real datasets. Our measure can also be extended to more complex situations, for example, to detect the mutual independence among more than two random objects.

Our Ball Covariance enjoys the following properties: (i) It is nonnegative, and satisfies the Cauchy-Schwarz type inequality; (ii) It is nonparametric and makes fewer restrictive data assumptions even without finite moment conditions; (iii) Its empirical version is feasible and can be used as a test statistic of independence with some desired test properties; and finally, (iv) it is interesting that the HHG dependence measure introduced by Heller, Heller, and Gorfine (2013) is a special case of Ball Covariance. HHG is known to be useful in detecting diverse relationships including linear and nonlinear monotonic/nonmonotonic functions, and nonfunctional relationships.

As shown by Lyons (2013), distance covariance satisfies the independence-zero equivalence property only for metric spaces which are of strong negative type. Developing a dependence measure with the independence-zero equivalence property in a general metric space is known to be challenging, but we overcame it by using the covering theorem (Bogachev 2007) and the Dynkin system (Riss 2006). Furthermore, our asymptotic results require advanced approximation techniques which involve V-statistics with random weights.

Some issues deserve further studies. First, the computational complexity of Ball Covariance is $O(n^2 \log n)$ in the general case, but it can be reduced to be $O(n^2)$ for univariate case. Second, our Ball Covariance can probably be generalized to semiparametric or more flexible forms, and deal with a wide range of problems such as variable selection, network construction, and signal detection.

Supplementary materials

The online supplementary materials contain some of the technical details and the additional properties.

Funding

Dr. Pan's research is partially supported by the National Natural Science Foundation of China (11701590), Natural Science Foundation of Guangdong Province of China (2017A030310053) and Young teacher program/Fundamental Research Funds for the Central Universities (17lgpy14). Dr. Wang's research is partially supported by the National Natural Science Foundation of China (11771462) and International Science & Technology Cooperation program of Guangdong (201634000424100001). Zhang's work was supported in part by the National Institutes of Health (Grant MH116527) and National Science Foundation (Grant

DMS-1722544). Dr. Zhu's work was partially supported by the U.S. National Institutes of Health (Grants MH086633 and R01MH116527).

References

Banach, S. (1932), "Théorie des Opérations Linéaires," Warszawa: Monografie Matematyczne. [308]

Banerjee, M., Chakraborty, R., Ofori, E., Okun, M. S., Viallancourt, D. E., and Vemuri, B. C. (2016), "A Nonlinear Regression Technique for Manifold Valued Data With Applications to Medical Image Analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4424–4432. [307]

Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961), "Distribution Free Tests of Independence Based on the Sample Distribution Function," *Annals of Mathematical Statistics*, 32, 485–498. [307]

Bogachev, V. I. (2007), *Measure Theory Volume I*, Berlin, Germany: Springer. [316]

Brunel, H., Gallardo-Chacón, J.-J., Buil, A., Vallverdú, M., Soria, J. M., Caminal, P., and Perera, A. (2010), "MISS: A Non-Linear Methodology Based on Mutual Information for Genetic Association Studies in Both Population and Sib-Pairs Analysis," *Bioinformatics*, 26, 1811–1818. [307]

Cornea, E., Zhu, H., Kim, P. T., and Ibrahim, J. G. (2016), "Regression models on Riemannian Symmetric Spaces," *Journal of the Royal Statistical Society*, Series B, 79, 463–482. [307]

Dryden, I., and Mardia, K. (1998), *Statistical Shape Analysis*, New York: Wiley. [307]

Fan, J., and Lv, J. (2008), "Sure Independence Screening for Ultrahigh Dimensional Feature Space," *Journal of the Royal Statistical Society*, Series B, 70, 849–911. [307]

Free, S. L., Ohiggins, P., Maudgil, D., Dryden, I. L., Lemieux, L., Fish, D. R., and Shorvon, S. D. (2001), "Landmark-Based Morphometrics of the Normal Adult Brain Using MRI," *NeuroImage*, 13, 801–813. [314]

Grady, C. L., Springer, M. V., Hongwanishkul, D., Mcintosh, A. R., and Winocur, G. (2006), "Age-Related Changes in Brain Activity Across the Adult Lifespan," *Journal of Cognitive Neuroscience*, 18, 227–241. [314]

Grenander, U., and Miller, M. I. (2007), *Pattern Theory From Representation to Inference*, Oxford, UK: Oxford University Press. [307]

Guo, Z., Zhang, J., Wang, Z., Ang, K. Y., Huang, S., Hou, Q., Su, X., Qiao, J., Zheng, Y., Wang, L., Koh, E., Danliang, H., Xu, J., Lee, Y. K., and Zhang H. (2016), "Intestinal Microbiota Distinguish Gout Patients From Healthy Humans," *Scientific Reports*, 6, 20602. [315]

Haro, C., Rangel-Zúñiga, O. A., Alcalá-Díaz, J. F., Gómez-Delgado, F., Pérez-Martínez, P., Delgado-Lista, J., Quintana-Navarro, G. M., Landa, B. B., Navas-Cortés, J. A., Tena-Sempere, M. et al. (2016), "Intestinal Microbiota is Influenced by Gender and Body Mass Index," *PloS One*, 11, e0154090. [316]

Heller, R., Heller, Y., and Gorfine, M. (2013), "A Consistent Multivariate Test of Association Based on Ranks of Distances," *Biometrika*, 100, 503–510. [308,310,316]

Hoeffding, W. (1948), "A Non-parametric Test of Independence," *The Annals of Mathematical Statistics*, 19, 546–557. [307,308]

Jung, S., Dryden, I. L., and Marron, J. S. (2012), "Analysis of Principal Nested Spheres," *Biometrika*, 99, 551–568. [307]

Kendall, D. G., Barden, D., Carne, T. K., and Le, H. (2009), *Shape and Shape Theory* (Vol. 500). New York: Wiley. [314]

Kendall, M. G. (1938), "A New Measure of Rank Correlation," *Biometrika*, 30, 81–93. [307]

Kleiber, M., and Pervin, W. J. (1969), "A Generalized Banach-Mazur Theorem," *Bulletin of The Australian Mathematical Society*, 1, 169–173. [308]

Li, R., Zhong, W., and Zhu, L. (2012), "Feature Screening Via Distance Correlation Learning," *Journal of the American Statistical Association*, 107, 1129–1139. [307]

Lyons, R. (2013), "Distance Covariance in Metric Spaces," *The Annals of Probability*, 41, 3284–3305. [307,308,316]

Napier, G. (2014), "A Bayesian Hierarchical Model of Compositional Data With Zeros: Classification and Evidence Evaluation of Forensic Glass," PhD thesis, University of Glasgow. [315]

Pan, W., Tian, Y., Wang, X., and Zhang, H. (2018), "Ball Divergence: Nonparametric Two Sample Test," *The Annals of Statistics*, 46, 1109–1137. [308]

Passe, T. J., Rajagopalan, P., Tupler, L. A., Byrum, C. E., Macfall, J. R., and Krishnan, K. R. R. (1997), "Age and Sex Effects on Brain Morphology," *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 21, 1231–1237. [314]

Pearson, K. (1895), "Note on Regression and Inheritance in the Case of Two Parents," *Proceedings of the Royal Society of London*, 58, 240–242. [307]

Riss, E. (2006), "Generating Borel Sets by Balls," *St. Petersburg Mathematical Journal*, 17, 683–698. [316]

Spearman, C. (1904), "The Proof and Measurement of Association Between Two Things," *American Journal of Psychology*, 15, 72–101. [307]

Srivastava, A., and Klassen, E. P. (2016), *Functional and Shape Data Analysis*, New York: Springer. [307]

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007), "Measuring and Testing Dependence by Correlation of Distances," *The Annals of Statistics*, 35, 2769–2794. [307,309,310]

Wang, X., Pan, W., Hu, W., Tian, Y., and Zhang, H. (2015), "Conditional Distance Correlation," *Journal of the American Statistical Association*, 110, 1726–1734. [309]

Watanabe, M., Sakai, O., Ozonoff, A., Kussman, S., and Jara, H. (2013), "Age-Related Apparent Diffusion Coefficient Changes in the Normal Brain," *Radiology*, 266, 575–582. [314]

Younes, L. (2010), *Shapes and Diffeomorphisms*, Berlin, Heidelberg: Springer. [307]

Zhang, Z., and Wang J. (2007), "MLLE: Modified Locally Linear Embedding using Multiple Weights," in *Advances in Neural Information Processing Systems*, Cambridge, MA: MIT Press, pp. 1593–1600. [315]

Zhu, L., Xu, K., Li, R., and Zhong, W. (2017), "Projection Correlation Between Two Random Vectors," *Biometrika*, 104, 829–843. [307]