Statistica Sinica Preprint No: SS-2021-0416

Title	Nonparametric Tests of Independence for Circular Data Based on Trigonometric Moments
Manuscript ID	SS-2021-0416
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0416
Complete List of Authors	Eduardo García-Portugués, Pierre Lafaye de Micheaux, Simos G. Meintanis and Thomas Verdebout
Corresponding Authors	Eduardo García-Portugués
E-mails	edgarcia@est-econ.uc3m.es

Notice: Accepted version subject to English editing.

Statistica Sinica

NONPARAMETRIC TESTS OF INDEPENDENCE FOR CIRCULAR DATA BASED ON TRIGONOMETRIC MOMENTS

Eduardo García-Portugués[1], Pierre Lafaye de Micheaux[2,3,4],

Simos G. Meintanis[5,6], Thomas Verdebout[7,8]

[1] *Department of Statistics, Carlos III University of Madrid*

[2] *School of Mathematics and Statistics, UNSW Sydney*

[3] *Desbrest Institute of Epidemiology and Public Health, Université de Montpellier*

[4] *AMIS, Université Paul Valéry Montpellier 3*

[5] *Department of Economics, National and Kapodistrian University of Athens*

[6] *Unit for Pure and Applied Analytics, North-West University*

[7] *Département de Mathématique, Université libre de Bruxelles*

[8] *ECARES, Université libre de Bruxelles*

Abstract: We introduce nonparametric tests of independence for bivariate circular data based on trigonometric moments. Our contributions lie in (i) proposing nonparametric tests that are locally and asymptotically optimal against bivariate cosine von Mises alternatives and (ii) extending these tests, via the empirical characteristic function, to obtain consistent tests against broader sets of alternatives, eventually being omnibus. In particular, one of such omnibus tests is a circular version of the celebrated distance-covariance test. We thus provide a collection of trigonometric-based tests of varying generality and known optimalities. The large-sample behavior of the tests under the null and alternative hypotheses are obtained, while simulations show that the new tests are competitive against previous proposals. Two data applications in astronomy and forest science illustrate the usage of the tests.

Key words and phrases: Characteristic function, directional data, independence, trigonometric moments.

1. Introduction

The goal of this paper is to introduce new tests of independence between two circular random variables $\vartheta^{(1)}$ and $\vartheta^{(2)}$ that are supported on $\mathbb{T} := [-\pi, \pi)$. Given an independent and identically distributed sample $\big(\vartheta_1^{(1)}, \vartheta_1^{(2)}\big), \ldots, \big(\vartheta_n^{(1)}, \vartheta_n^{(2)}\big)$, we wish to test the null hypothesis $\mathcal{H}_0$ of independence between $\vartheta^{(1)}$ and $\vartheta^{(2)}$ against the general alternative $\mathcal{H}_1$ consisting on the negation of $\mathcal{H}_0$. This fundamental testing problem has relevant applications in fields where circular data are common, such as in astronomy, biology, geology, and forest science, to name just a few.

A variety of tailored statistical methods for the analysis of data comprised by directions, such as circular data, have been developed in the last decades; see the general treatments of Mardia and Jupp (1999), Jammalamadaka and SenGupta (2001), and Ley and Verdebout (2017), as well as the recent review of Pewsey and García-Portugués (2021). In particular, the analysis of data on $\mathbb{T}^2$ that is generated by a pair of angular variables, referred to as "circular-circular" or "toroidal" data, has attracted a sizable number of modeling proposals in the recent years (Pewsey and García-Portugués, 2021, Section 3.2). This interest has been notably boosted by applications in bioinformatics, where a sequence of dihedral angles characterizes a protein's three-dimensional backbone (e.g., Boomsma et al., 2008). In addition, the development of toroidal distributions is intimately related with the design of models for circular time series (Wehrly and Johnson, 1980) that naturally appear in a variety of other fields such as astronomy and forest science; see Section 5.

Much of the modeling effort for toroidal data has been dominated by the search for bivariate extensions of the von Mises distribution, often regarded as the "circular Gaussian" distribution. The first of such proposals was the bivariate von Mises density of Mardia (1975), considered as an overparametrized model due to its eight parameters. This motivated the six-parameter submodel of Rivest (1988) and the five-parameter "sine" (Singh et al., 2002), "cosine" (Mardia et al., 2007), and "hybrid" (Kent et al., 2008) submodels. The properties of the last three were compared in Kent et al. (2008) and Mardia and Frellsen (2012). A different modeling pathway was initiated with the family of copula-structured toroidal densities by Wehrly and Johnson (1980), whose most successful representative is the bivariate wrapped Cauchy distribution (Kato and Pewsey, 2015).

Investigating relationships between variables is central to many scientific studies, and tests of independence typically precede any attempt at modeling association. Consequently, many contributions in directional statistics have been dealing with correlation, dependence, and tests for independence. Measures of circular correlation have been put forward by Watson and Beran (1967), Jupp and Mardia (1980), Shieh et al. (1994), and more recently by Zhan et al. (2019). In a different direction, Rothman (1971) introduced a version of the Cramér–von Mises test of independence. In parametric contexts related with the models of the previous paragraph, one may resort to the likelihood-based tests suggested by Mardia and Puri (1978), Puri and Rao (1977), and Shieh and Johnson (2005). Finally, for testing independence in data with mixed directional/linear components, smoothing-based tests have been proposed by García-Portugués et al. (2015). Unlike standard independence tests, the aforementioned tests honor the circular/directional nature of the random variables involved by being rotation invariant on them. A non-rotation-invariant test provides spurious decisions for assessing the independence of $(\vartheta^{(1)}, \vartheta^{(2)})$, as its p-value is dependent on the sample coordinates (e.g., representations on $[-\pi, \pi)^2$ or $[0, 2\pi)^2$ might yield different test decisions); see Section D in the Supplementary Materials (SM) for specific examples.

When testing independence, nonparametric methods based on the characteristic function have also been employed as alternatives to non-omnibus tests based on association coefficients and to smoothing-based tests that exhibit the familiar drawbacks of bandwidth selection and slow convergence. These tests exploit the factorization characterization of the joint characteristic function of independent random variables. This property propagated "Fourier"-type tests in the past, going back as far as Csörgő and Hall (1982) and Csörgő (1985). Since then, Fourier methods have enjoyed increasing popularity, finally reaching some sort of climax with the introduction of the novel notions of "distance covariance" and "distance correlation" (Székely et al., 2007), and beyond. Indicatively, we refer to the contributions by Gretton et al. (2005), Székely et al. (2007), Meintanis and Iliopoulos (2008), Hlávka et al. (2011), Fan et al. (2017), Chen et al. (2019), and Chakraborty and Zhang (2019), all of which propose tests of independence in varying settings and different levels of generality, but always with the characteristic function being the underlying notion. This popularity notwithstanding, and despite the fact that testing based on characteristic functions is not unfamiliar to circular data (Meintanis and Verdebout, 2019), the use of characteristic functions for testing independence of non-linear data remains substantially unexplored.

We introduce in this paper nonparametric tests of independence for toroidal data based on trigonometric moments. We first propose nonparametric tests using joint cosine moments that are locally and asymptotically optimal against sequences of bivariate cosine von Mises alternatives, and for which the powers of the tests are explicitly obtained. We then extend these tests, via the empirical characteristic function, to more general multiple-orders tests that merge cosine and sine moments, and that are consistent against broader sets of alternatives. We obtain usable asymptotic null distributions for all the test statistics, thus avoiding their calibration by resampling methods. We then propose two characteristic function-based omnibus tests with tractable computational forms that can be efficiently calibrated using permutations. The second one is a kind of circular distance-covariance test. Simulations corroborate the adequate finite-sample null and non-null behavior of the tests, as well as their competitiveness against other testing approaches based on association coefficients and smoothing. Two data applications are provided, one on the study on the serial dependence of long-period comet records and another in the evaluation of the dependence between the orientations of Portuguese wildfires.

2. A cosine test of independence

2.1 Genesis and null asymptotic distribution

Our objective is to test the null hypothesis $\mathcal{H}_0$ of independence between $\vartheta^{(1)}$ and $\vartheta^{(2)}$. Without loss of generality (see Proposition 3 below), we assume that $\vartheta^{(j)}$ is circularly centered, i.e., such that its circular mean $\mu^{(j)} := \operatorname{atan2}\big(\mathbb{E}\big[\sin\big(\vartheta^{(j)}\big)\big], \mathbb{E}\big[\cos\big(\vartheta^{(j)}\big)\big]\big)$ is zero, $j = 1, 2$, where $\operatorname{atan2}(y, x) \in \mathbb{T}$ is the argument of the complex number $x + \mathrm{i}y$. Given an independent and identically distributed sample $\big(\vartheta_1^{(1)}, \vartheta_1^{(2)}\big), \ldots, \big(\vartheta_n^{(1)}, \vartheta_n^{(2)}\big)$ from $\big(\vartheta^{(1)}, \vartheta^{(2)}\big)$, we consider the empirical versions

$$\hat{\mathcal{J}}_{jc}(r) := n^{-1} \sum_{i=1}^{n} \cos\big(r\vartheta_i^{(j)}\big), \quad \hat{\mathcal{J}}_{js}(r) := n^{-1} \sum_{i=1}^{n} \sin\big(r\vartheta_i^{(j)}\big), \quad j = 1, 2,$$
$$\hat{\mathcal{J}}_{c}(r_1, r_2) := n^{-1} \sum_{i=1}^{n} \cos\big(r_1\vartheta_i^{(1)} + r_2\vartheta_i^{(2)}\big), \quad \hat{\mathcal{J}}_{s}(r_1, r_2) := n^{-1} \sum_{i=1}^{n} \sin\big(r_1\vartheta_i^{(1)} + r_2\vartheta_i^{(2)}\big),$$

of the respective marginal "cosine" and "sine" population moments (as well as their "addition" forms) given by

$$\mathcal{J}_{jc}(r) := \mathbb{E}\big[\cos\big(r\vartheta^{(j)}\big)\big], \quad \mathcal{J}_{js}(r) := \mathbb{E}\big[\sin\big(r\vartheta^{(j)}\big)\big], \quad j = 1, 2,$$

$\mathcal{J}_c(r_1, r_2) := \mathbb{E}\big[\cos\big(r_1\vartheta^{(1)} + r_2\vartheta^{(2)}\big)\big], \quad \mathcal{J}_s(r_1, r_2) := \mathbb{E}\big[\sin\big(r_1\vartheta^{(1)} + r_2\vartheta^{(2)}\big)\big].$

Here r, r_1, and r_2 are reals, although we will soon restrict to integer numbers; see below (3.6).

Based on the form of the "cosine addition moment", we have that, under the null the hypothesis of independence,

$$\mathcal{J}_c(r_1, r_2) = \mathcal{J}_{1c}(r_1)\mathcal{J}_{2c}(r_2) - \mathcal{J}_{1s}(r_1)\mathcal{J}_{2s}(r_2). \tag{2.1}$$

Based on (2.1), it is very natural to consider tests that reject $\mathcal{H}_0$ for large absolute values of the statistic

$$D_c^{(n)}(r_1, r_2) := \hat{\mathcal{J}}_c(r_1, r_2) - \hat{\mathcal{J}}_{1c}(r_1)\hat{\mathcal{J}}_{2c}(r_2) + \hat{\mathcal{J}}_{1s}(r_1)\hat{\mathcal{J}}_{2s}(r_2), \tag{2.2}$$

since, for any $(r_1, r_2) \in \mathbb{R}^2$, $D_c^{(n)}(r_1, r_2)$ will be close to zero under $\mathcal{H}_0$. The following proposition provides the asymptotic distribution of $D_c^{(n)}(r_1, r_2)$ under $\mathcal{H}_0$. Its proof is relegated to Section A in the SM, where all the results of the paper are proved.

Proposition 1. *Fix* $(r_1, r_2) \in \mathbb{R}^2$. *Under* $\mathcal{H}_0$, $\sqrt{n}D_c^{(n)}(r_1, r_2)$ *converges weakly as* $n \to \infty$ *to a Gaussian random variable with mean zero and variance* $V(r_1, r_2) := \mathbb{E}\big[\big\{\cos\big(r_1\vartheta^{(1)}+r_2\vartheta^{(2)}\big)-\mathcal{J}_{2c}(r_2)\cos\big(r_1\vartheta^{(1)}\big)-\mathcal{J}_{1c}(r_1)\cos\big(r_2\vartheta^{(2)}\big)+ \mathcal{J}_{2s}(r_2)\sin\big(r_1\vartheta^{(1)}\big) + \mathcal{J}_{1s}(r_1)\sin\big(r_2\vartheta^{(2)}\big)\big\}^2\big]$.

The asymptotic normality of $\sqrt{n}D_c^{(n)}(r_1, r_2)$ does not require any assumption on the distribution of the pair of random angles $(\vartheta^{(1)}, \vartheta^{(2)})$. A purely nonparametric test of independence can therefore be obtained on the basis of the Proposition 1. Indeed, we can consider tests $\phi_c^{(n)}(r_1, r_2)$ rejecting the null hypothesis of independence at the asymptotic level α when

$$T_n(r_1, r_2) := \frac{n\big(D_c^{(n)}(r_1, r_2)\big)^2}{\hat{V}_n(r_1, r_2)} > \chi^2_{1;1-\alpha}, \tag{2.3}$$

where $\chi^2_{1;\nu}$ denotes the νth (lower) quantile of the chi-square distribution with one degree of freedom and $\hat{V}_n(r_1, r_2)$ is a consistent estimator of the variance term $V(r_1, r_2)$ defined in Proposition 1, such as its direct empirical version. Although being purely nonparametric, as no assumption on the data generating process is imposed, the tests $\phi_c^{(n)}(r_1, r_2)$ with $r_1 = 1$ and $r_2 = \pm 1$ will enjoy certain local and asymptotic optimality properties.

2.2 Optimality and power against bivariate von Mises alternatives

Consider the bivariate cosine von Mises model of Mardia et al. (2007), characterized by densities of the form

$$\left(\vartheta^{(1)}, \vartheta^{(2)}\right) \mapsto C(\kappa_1, \kappa_2, \kappa_3) \exp\left\{\kappa_1 \cos\left(\vartheta^{(1)}\right) + \kappa_2 \cos\left(\vartheta^{(2)}\right) + \kappa_3 \cos\left(\vartheta^{(1)} - \vartheta^{(2)}\right)\right\}, \tag{2.4}$$

where $\kappa_1, \kappa_2 \geq 0$ are concentration parameters, $\kappa_3 \in \mathbb{R}$ is a parameter controlling the dependence, and $C(\kappa_1, \kappa_2, \kappa_3)$ is a normalizing constant. Note that, for the ease of our derivations, we flip the sign of $\kappa_3 \in \mathbb{R}$ in (2.4) with respect to the original model parametrization. Following the terminology in Mardia and Frellsen (2012), density (2.4) is called the bivariate cosine model with *positive interaction*. The same model with *negative interaction* is obtained by replacing $\cos\left(\vartheta^{(1)} - \vartheta^{(2)}\right)$ with $\cos\left(\vartheta^{(1)} + \vartheta^{(2)}\right)$ in (2.4). As stated in Mardia et al. (2007), both models capture the correlations between the cosines and sines of the circular variables, though none is strictly associated with positive or negative correlations between angles. Indeed, the sign of "angular correlations" depends on κ_3, which affects asymmetrically the kind of dependence induced by (2.4): positive values of κ_3 guarantee unimodality, with positive/negative angular correlation depending on the positive/negative interaction (Theorem 6.2 in Mardia and Frellsen (2012); third column of Figure 1); negative κ_3 may generate bimodality distributed in an opposite correlation pattern to that of $\kappa_3 > 0$. Shifting of (2.4) can be achieved by replacing $\vartheta^{(j)}$ with $\vartheta^{(j)} - \mu^{(j)}$, for $\mu^{(j)} \in \mathbb{T}$, $j = 1, 2$. Location parameters do not affect the dependence form of (2.4), yet they make it more cumbersome.

When $\kappa_3 = 0$, the marginals of (2.4) are independent centered von Mises distributions with concentrations κ_1 and κ_2, thus testing independence in this model reduces to testing $\mathcal{H}_0 : \kappa_3 = 0$ against $\mathcal{H}_1 : \kappa_3 \neq 0$. We show in Proposition 2 that the tests $\phi_c^{(n)}(1,1)$ and $\phi_c^{(n)}(1,-1)$ are locally and asymptotically maximin (see Ley and Verdebout, 2017, Section 5, for a definition) for testing $\mathcal{H}_0 : \kappa_3 = 0$ against $\mathcal{H}_1 : \kappa_3 \neq 0$ within sequences of bivariate cosine models with negative and positive interaction, respectively. Recall that a test ϕ^* is called maximin in the class $\mathcal{C}_\alpha$ of level-α tests for some null hypothesis $\mathcal{H}_0$ against the alternative $\mathcal{H}_1$ if: (i) ϕ^* has level α; (ii) the power of ϕ^* is such that

$$\inf_{\mathrm{P}\in\mathcal{H}_1} \mathbb{E}_{\mathrm{P}}[\phi^*] \geq \sup_{\phi\in\mathcal{C}_\alpha} \inf_{\mathrm{P}\in\mathcal{H}_1} \mathbb{E}_{\mathrm{P}}[\phi].$$

We denote by $\mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,\kappa_3);-}$ and $\mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,\kappa_3);+}$ the joint distributions of an independent and identically distributed sample $\big(\vartheta_1^{(1)},\vartheta_1^{(2)}\big),\ldots,\big(\vartheta_n^{(1)},\vartheta_n^{(2)}\big)$ from distribution (2.4), respectively with negative and positive interaction. Obviously, $\mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,0);-} = \mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,0);+}$, which is simply denoted as $\mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,0)}$.

Proposition 2. *Letting τ_n be a bounded real sequence, the test $\phi_c^{(n)}(1,1)$ is locally and asymptotically maximin for testing $\mathcal{H}_0 : \cup_{\kappa_1\geq 0} \cup_{\kappa_2\geq 0} \mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,0)}$ against $\mathcal{H}_1 : \cup_{\kappa_1\geq 0} \cup_{\kappa_2\geq 0} \mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,n^{-1/2}\tau_n);-}$, while the test $\phi_c^{(n)}(1,-1)$ is locally and asymptotically maximin for testing $\mathcal{H}_0 : \cup_{\kappa_1\geq 0} \cup_{\kappa_2\geq 0} \mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,0)}$ against $\mathcal{H}_1 : \cup_{\kappa_1\geq 0} \cup_{\kappa_2\geq 0} \mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,n^{-1/2}\tau_n);+}$.*

The nonparametric tests $\phi_c^{(n)}(1,1)$ and $\phi_c^{(n)}(1,-1)$ therefore enjoy some parametric optimality properties for testing $\mathcal{H}_0 : \kappa_3 = 0$ against $\mathcal{H}_1 : \kappa_3 \neq 0$. Although the tests $\phi_c^{(n)}(r_1, r_2)$, $(r_1, r_2) \in \mathbb{R}^2$ do not enjoying such local and asymptotic optimality, it is easy to show that they enjoy non-trivial power against the contiguous alternatives $\mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,n^{-1/2}\tau_n);+}$ and $\mathrm{P}^{(n)}_{(\kappa_1,\kappa_2,n^{-1/2}\tau_n);-}$, and can therefore be considered as reasonable tests for such alternatives.

Hitherto, we have assumed the sample comes from a circularly-centered random vector. Otherwise, the test statistic $T_n(r_1, r_2)$ in (2.3) has to be computed from the centered data $\vartheta_i^{(j)} - \mu^{(j)}$, $i = 1, \ldots, n$, $j = 1, 2$; Proposition 1 then holds replacing the $\vartheta_i^{(j)}$'s and the $\vartheta^{(j)}$'s by $\vartheta_i^{(j)} - \mu^{(j)}$ and $\vartheta^{(j)} - \mu^{(j)}$, respectively, $i = 1, \ldots, n$, $j = 1, 2$. Moreover, the local and asymptotic optimality obtained in Proposition 2 also holds in the unspecified location case. Of course, the location parameters $\mu^{(1)}$ and $\mu^{(2)}$ are rarely known in practice so they have to be estimated. This can be done using the sample circular means

$$\hat{\mu}^{(j)} := \mathrm{atan2}\left(\frac{1}{n}\sum_{i=1}^{n} \sin\left(\vartheta_i^{(j)}\right), \frac{1}{n}\sum_{i=1}^{n} \cos\left(\vartheta_i^{(j)}\right)\right), \quad j = 1, 2.$$

This estimation produces the centered sample

$$\left(\vartheta_1^{(1)} - \hat{\mu}^{(1)}, \vartheta_1^{(2)} - \hat{\mu}^{(2)}\right), \ldots, \left(\vartheta_n^{(1)} - \hat{\mu}^{(1)}, \vartheta_n^{(2)} - \hat{\mu}^{(2)}\right). \tag{2.5}$$

When computed from this centered sample, the test statistic $T_n(r_1, r_2)$ in

(2.3) is rotation invariant, which is a highly desirable property in the present toroidal context. We moreover have the following result.

Proposition 3. *Denote by* $\hat{D}_c^{(n)}(r_1, r_2)$ *and* $D_c^{(n)}(r_1, r_2)$ *the quantities defined in* (2.2)*, but computed from the samples* (2.5) *and*

$$\big(\vartheta_1^{(1)} - \mu^{(1)}, \vartheta_1^{(2)} - \mu^{(2)}\big), \ldots, \big(\vartheta_n^{(1)} - \mu^{(1)}, \vartheta_n^{(2)} - \mu^{(2)}\big),$$

respectively. Then, provided that $\sqrt{n}\big(\hat{\mu}^{(j)} - \mu^{(j)}\big) = O_{\mathrm{P}}(1)$ *as* $n \to \infty$*,* $j = 1, 2$*,* $\sqrt{n}\big(\hat{D}_c^{(n)}(r_1, r_2) - D_c^{(n)}(r_1, r_2)\big)$ *is* $o_{\mathrm{P}}(1)$ *as* $n \to \infty$*.*

Classical arguments similarly show that, provided that the data generating process ensures that $\sqrt{n}\big(\hat{\mu}^{(j)} - \mu^{(j)}\big) = O_{\mathrm{P}}(1)$ as $n \to \infty$, $j = 1, 2$, the centering has no asymptotic effect on $\hat{V}_n(r_1, r_2)$ in (2.3). Consequently, the centering step does not affect the asymptotic null distribution of $T_n(r_1, r_2)$ in (2.3). Note that the same holds under contiguous alternatives. Since the centering of the sample is innocuous in terms of the asymptotic behavior of (2.3) and it makes the test rotation invariant, this centering is implicitly assumed henceforth when applying the $\phi_c^{(n)}(r_1, r_2)$ test.

We conclude the section by pointing out that, while being of a nonparametric nature, the tests $\phi_c^{(n)}(r_1, r_2)$ are clearly designed to detect certain types of dependence (and not any kind of dependence): as seen in Proposition 2, the tests $\phi_c^{(n)}(1, \pm 1)$ are particularly well-adapted to bivariate cosine von Mises alternatives that feature reflective symmetric marginal distributions. Working along the same lines, one could consider tests based on the sine empirical moments and show that some of their versions are locally and asymptotically optimal within specific parametric models. Rather than moving in this direction, in the following section we proceed towards tests of independence that are able to detect *arbitrary* types of dependence.

3. Omnibus tests

The well-known factorization property of characteristic functions entails that the null hypothesis of independence may equivalently be stated as

$$\varphi(r_1, r_2) = \varphi_1(r_1)\varphi_2(r_2), \quad \text{for all } (r_1, r_2) \in \mathbb{Z}^2, \tag{3.6}$$

where $\varphi(r_1, r_2) := \mathbb{E}\big[e^{\mathrm{i}(r_1\vartheta^{(1)}+r_2\vartheta^{(2)})}\big]$, $\mathrm{i} := \sqrt{-1}$, is the joint characteristic function and $\varphi_j(r_j) := \mathbb{E}\big[e^{\mathrm{i}r_j\vartheta^{(j)}}\big]$ stands for the marginal characteristic function of $\vartheta^{(j)}$, $j = 1, 2$. Recall that, for random variables on the real line, (3.6) needs to be considered for all $(r_1, r_2) \in \mathbb{R}^2$ while, due to periodicity, in the case of circular random variables, it is sufficient to consider the characteristic functions only for integer arguments. This is because the joint distribution of $\big(\vartheta^{(1)}, \vartheta^{(2)}\big)$ is identical to that of $\big(\vartheta^{(1)} + 2\pi, \vartheta^{(2)}\big)$ and thus we have $\varphi(r_1, r_2) = e^{\mathrm{i}2\pi r_1}\varphi(r_1, r_2)$, hence r_1 must be an integer, and likewise for r_2 (Jammalamadaka and SenGupta, 2001, Section 2.1).

Based on $\left(\vartheta_1^{(1)}, \vartheta_1^{(2)}\right), \ldots, \left(\vartheta_n^{(1)}, \vartheta_n^{(2)}\right)$, the classical estimator of the joint characteristic function is

$$\hat{\varphi}(r_1, r_2) := \frac{1}{n} \sum_{i=1}^{n} e^{\mathrm{i}(r_1 \vartheta_i^{(1)} + r_2 \vartheta_i^{(2)})}, \tag{3.7}$$

while the corresponding empirical marginals, say $\hat{\varphi}_1$ (respectively, $\hat{\varphi}_2$), can be obtained by setting $r_2 = 0$ ($r_1 = 0$) in (3.7). Then, in view of (3.6), it is natural to consider the test statistics

$$D^{(n)}(r_1, r_2) := \hat{\varphi}(r_1, r_2) - \hat{\varphi}_1(r_1)\hat{\varphi}_2(r_2), \quad (r_1, r_2) \in \mathbb{Z}^2, \tag{3.8}$$

as diagnostic components for independence. Notice that the quantity $D_c^{(n)}(r_1, r_2)$ defined in (2.2) is just the real part of $D^{(n)}(r_1, r_2)$, and consequently an extension of the tests studied in Section 2 may be obtained by considering both the real and imaginary parts of $D^{(n)}(r_1, r_2)$ for multiple arguments $(r_1, r_2) \in \mathbb{Z}^2$. To this end, we define the vector

$$\boldsymbol{\Delta}_n\left(\boldsymbol{r}^{(c)}, \boldsymbol{r}^{(s)}\right) := \left(D_c^{(n)}\left(r_{11}^{(c)}, r_{12}^{(c)}\right), \ldots, D_c^{(n)}\left(r_{J1}^{(c)}, r_{J2}^{(c)}\right), D_s^{(n)}\left(r_{11}^{(s)}, r_{12}^{(s)}\right), \ldots, D_s^{(n)}\left(r_{K1}^{(s)}, r_{K2}^{(s)}\right)\right)',$$

where $D_c^{(n)}(r_1, r_2)$ and $D_s^{(n)}(r_1, r_2)$ stand for the real and imaginary parts, respectively, of $D^{(n)}(r_1, r_2)$. Using similar arguments as those in Section 2, it may be shown that $\sqrt{n}\boldsymbol{\Delta}_n\left(\boldsymbol{r}^{(c)}, \boldsymbol{r}^{(s)}\right)$ is asymptotically a zero-mean multivariate Gaussian with some covariance matrix $\boldsymbol{\Sigma}$ that is easily computable; see Section B in the SM. As a result, letting $\hat{\boldsymbol{\Sigma}}$ be an invertible and consistent estimator of $\mathbf{\Sigma}$, a very natural test $\phi^{(n)}\big(\boldsymbol{r}^{(c)}, \boldsymbol{r}^{(s)}\big)$ rejects $\mathcal{H}_0$ for large values of $n\big(\mathbf{\Delta}_n\big(\boldsymbol{r}^{(c)}, \boldsymbol{r}^{(s)}\big)\big)'\hat{\mathbf{\Sigma}}^{-1}\mathbf{\Delta}_n\big(\boldsymbol{r}^{(c)}, \boldsymbol{r}^{(s)}\big)$. Note that some choices of $\boldsymbol{r}^{(c)} = \big(r_{11}^{(c)}, r_{12}^{(c)}, \ldots, r_{J1}^{(c)}, r_{J2}^{(c)}\big)' \in \mathbb{Z}^{2J}$ and $\boldsymbol{r}^{(s)} = \big(r_{11}^{(s)}, r_{12}^{(s)}, \ldots, r_{K1}^{(s)}, r_{K2}^{(s)}\big)' \in \mathbb{Z}^{2K}$ yield matrices $\mathbf{\Sigma}$ that are invertible, some not. Note also that the particular case obtained by putting $J = 2$ with $\big(r_{11}^{(c)}, r_{12}^{(c)}, r_{21}^{(c)}, r_{22}^{(c)}\big) = (1, -1, 1, 1)$ and $K = 0$ (so that there is no "sine part" in $\mathbf{\Delta}_n\big(\boldsymbol{r}^{(c)}, \boldsymbol{r}^{(s)}\big)$) yields a test that combines the two test statistics that are locally and asymptotically optimal against contiguous cosine von Mises alternatives with positive and negative dependence. An implicit centering of the sample is also assumed when applying $\phi^{(n)}\big(\boldsymbol{r}^{(c)}, \boldsymbol{r}^{(s)}\big)$ as, analogously to the $\phi_c^{(n)}(r_1, r_2)$ test, this centering step is innocuous in terms of the asymptotic behavior of the test and makes it rotation invariant.

While the tests $\phi^{(n)}\big(\boldsymbol{r}^{(c)}, \boldsymbol{r}^{(s)}\big)$, with $\boldsymbol{r}^{(c)} \in \mathbb{Z}^{2J}$ and $\boldsymbol{r}^{(s)} \in \mathbb{Z}^{2K}$, are expected to have good power properties beyond the class of von Mises distributions for which $\phi_c^{(n)}(1, \pm 1)$ is locally and asymptotically maximin, these tests are not "omnibus", i.e., they may potentially have trivial power against certain alternatives. In order to have an omnibus test, the uniqueness property of characteristic functions dictates that we must take into account all possible pairs $(r_1, r_2) \in \mathbb{Z}^2$. Consequently, we define a test criterion that rejects $\mathcal{H}_0$ for large values of

$$T_{n,w} := n \sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} \left| D^{(n)}(r_1, r_2) \right|^2 w(r_1, r_2), \tag{3.9}$$

where $|\cdot|$ denotes the modulus of a complex number and $w : \mathbb{Z}^2 \to [0, \infty)$ is a weight function specified below. The following proposition formalizes the limit behavior of $T_{n,w}$ against arbitrary deviations from the null hypothesis of independence.

Proposition 4. *Assume that w in (3.9) satisfies $\sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} w(r_1, r_2) < \infty$. Then,*

$$\frac{T_{n,w}}{n} \to \mathcal{T}_w := \sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} |\varphi(r_1, r_2) - \varphi_1(r_1)\varphi_2(r_2)|^2 w(r_1, r_2) \tag{3.10}$$

almost surely as $n \to \infty$. Moreover, $\mathcal{T}_w$ is strictly positive unless $\mathcal{H}_0$ holds true, a fact which entails strong consistency of the test that rejects $\mathcal{H}_0$ for large values of $T_{n,w}$.

While L_2-type test statistics such as $T_{n,w}$ are omnibus, they typically have highly non-trivial asymptotic null distributions that essentially prevent their use as test criteria. We refer to Puri and Rao (1977), Shieh et al. (1994), and Watson and Beran (1967) for analogous results; see also Jammalamadaka and SenGupta (2001, Section 8.9). Nevertheless, it is straightforward to implement a permutation version of a test based on $T_{n,w}$.

The application of the test statistic would be further advanced if $T_{n,w}$ could be computed analytically. To this end, consider a weight function decomposed as $w(r_1, r_2) = v(r_1)v(r_2)$, with v being a symmetric function about zero. Then, (3.9) may be rewritten as (see Section A in the SM)

$$T_{n,w} = \frac{1}{n}\sum_{j,k=1}^{n} \mathcal{J}_c^{(v)}\big(\vartheta_{jk}^{(1)}\big)\mathcal{J}_c^{(v)}\big(\vartheta_{jk}^{(2)}\big) + \frac{1}{n^3}\Bigg[\sum_{j,k=1}^{n} \mathcal{J}_c^{(v)}\big(\vartheta_{jk}^{(1)}\big)\Bigg]\Bigg[\sum_{j,k=1}^{n} \mathcal{J}_c^{(v)}\big(\vartheta_{jk}^{(2)}\big)\Bigg] - \frac{2}{n^2}\sum_{j,k,\ell=1}^{n} \mathcal{J}_c^{(v)}\big(\vartheta_{jk}^{(1)}\big)\mathcal{J}_c^{(v)}\big(\vartheta_{j\ell}^{(2)}\big), \tag{3.11}$$

where

$$\mathcal{J}_c^{(v)}(\vartheta) := \sum_{r=-\infty}^{\infty} \cos(r\vartheta)v(r), \tag{3.12}$$

with $\vartheta_{jk}^{(m)} := \vartheta_j^{(m)} - \vartheta_k^{(m)}$, $j, k = 1, \ldots, n$, $m = 1, 2$. Since $T_{n,w}$ only depends on the distances between observations, it is rotation invariant without requiring a prior centering of the sample.

Moreover, if we consider any probability mass function on the non-negative integers and set v equal to the symmetrized version of this function, then the series figuring in (3.12) equals the real part of the characteristic function of that probability mass function, evaluated at ϑ. A standard option is to choose the Poisson distribution, in which case

$$\mathcal{J}_c^{(v)}(\vartheta) = \cos(\lambda \sin\vartheta)e^{\lambda(\cos\vartheta - 1)}, \tag{3.13}$$

where λ is the Poisson parameter. Choosing $\lambda \in (0, \pi/2]$ guarantees the non-negativity of the kernel (3.13) for any $\vartheta \in \mathbb{T}$ (and also if $0 < |\lambda| \leq \pi/2$). We denote by $T_{n,\lambda}$ the statistic (3.11) based on (3.13). The test $\phi^{(n)}(\lambda)$ that rejects $\mathcal{H}_0$ for large values of $T_{n,\lambda}$ is implemented with a permutation approach that is described in Section C in the SM.

The weight specification $w_{\mathrm{dc}}(r_1, r_2) = (r_1 r_2)^{-2} 1_{\{r_1 \neq 0, r_2 \neq 0\}}$ in (3.9) yields the "distance-covariance" test statistic

$$S_{n,\mathrm{dc}} := n \sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} \left| D^{(n)}(r_1, r_2) \right|^2 w_{\mathrm{dc}}(r_1, r_2); \tag{3.14}$$

its definition is driven by the connection of a distance covariance statistic with the characteristic function. Notice that in (3.14) we exclude the origin, which is anyway uninformative regarding independence. Also, $S_{n,\mathrm{dc}}$ clearly satisfies the global consistency property of Proposition 4.

Carrying out analogous computations as in (3.11) and denoting

$$\mathcal{I}(\vartheta) := \sum_{\substack{r=-\infty \\ r \neq 0}}^{\infty} \frac{\cos(r\vartheta)}{r^2}, \tag{3.15}$$

we obtain from (3.14) that

$$\begin{aligned} S_{n,\mathrm{dc}} = \frac{1}{n} \sum_{j,k=1}^{n} \mathcal{I}\big(\vartheta_{jk}^{(1)}\big) \mathcal{I}\big(\vartheta_{jk}^{(2)}\big) + \frac{1}{n^3} \Bigg[\sum_{j,k=1}^{n} \mathcal{I}\big(\vartheta_{jk}^{(1)}\big) \Bigg] \Bigg[\sum_{j,k=1}^{n} \mathcal{I}\big(\vartheta_{jk}^{(2)}\big) \Bigg] \\ - \frac{2}{n^2} \sum_{j,k,\ell=1}^{n} \mathcal{I}\big(\vartheta_{jk}^{(1)}\big) \mathcal{I}\big(\vartheta_{j\ell}^{(2)}\big). \end{aligned}$$

Unlike $T_{n,\lambda}$, though, the computation of $S_{n,\mathrm{dc}}$ is less straightforward, as it requires evaluating (3.15), which can also be expressed as $\mathcal{I}(\vartheta) = \mathrm{Li}_2(e^{-\mathrm{i}\vartheta}) + \mathrm{Li}_2(e^{\mathrm{i}\vartheta})$, with $\mathrm{Li}_2(x) := \sum_{m=1}^{\infty} m^{-2} x^m$ being the dilogarithm function. Since $\mathcal{O}(Bn^2)$ evaluations of the kernel (3.15) are required when evaluating the test based on $S_{n,\mathrm{dc}}$ (henceforth denoted $\phi_{\mathrm{dc}}^{(n)}$) with B permutations, the increased computational burden with respect to the $\phi^{(n)}(\lambda)$ test is significant.

As all the tests introduced in this paper, that based on $S_{n,\mathrm{dc}}$ honors the circularity of the variables $\vartheta^{(1)}$ and $\vartheta^{(2)}$ for testing their independence. Section D in the SM exemplifies the important practical issues of applying an independence test that is unaware of the circular nature of $\vartheta^{(1)}$ and $\vartheta^{(2)}$, such as a standard distance-covariance test.

Remark 1. The test statistic in (3.9) can be, heuristically, further scrutinized with regards to correlations between $\vartheta^{(1)}$ and $\vartheta^{(2)}$. Consider, for simplicity, its population counterpart from Proposition 4 and write it as

$$\mathcal{T}_w = \sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} \left|\mathbb{C}\mathrm{ov}\left[e^{\mathrm{i}r_1\vartheta^{(1)}}, e^{\mathrm{i}r_2\vartheta^{(2)}}\right]\right|^2 w(r_1, r_2),$$

where $\mathbb{C}\mathrm{ov}\left[\cdot,\cdot\right]$ denotes covariance. Now use the exponential function expansion $e^z = 1 + (z/1!) + (z^2/2!) + \cdots$, compute a few terms of the covariance thereof and after some simplification write

$$\mathbb{C}\mathrm{ov}\left[e^{\mathrm{i}r_1\vartheta^{(1)}}, e^{\mathrm{i}r_1\vartheta^{(2)}}\right] = -\, r_1 r_2 \mathbb{C}\mathrm{ov}\left[\vartheta^{(1)}, \vartheta^{(2)}\right] - \frac{\mathrm{i}}{2}\left(r_1 r_2^2 \mathbb{C}\mathrm{ov}\left[\vartheta^{(1)}, \vartheta^{(2)^2}\right] + r_1^2 r_2 \mathbb{C}\mathrm{ov}\left[(\vartheta^{(1)^2}, \vartheta^{(2)}\right]\right)$$

$$+\frac{r_1^2 r_2^2}{4}\mathbb{C}\mathrm{ov}\left[\vartheta^{(1)^2}, \vartheta^{(2)}\right] + \cdots .$$

Expanding $\left|\mathbb{C}\mathrm{ov}\left[e^{\mathrm{i}r_1\vartheta^{(1)}}, e^{\mathrm{i}r_2\vartheta^{(2)}}\right]\right|^2$, using that $w(r_1, r_2) = v(r_1)v(r_2)$, and letting $v(\cdot)$ be a symmetric around zero probability function, it follows that

$$\mathcal{T}_w = \mu_2^2 \mathbb{C}\mathrm{ov}^2\left[\vartheta^{(1)}, \vartheta^{(2)}\right] + \frac{\mu_4^2}{16}\mathbb{C}\mathrm{ov}^2\left[\vartheta^{(1)^2}, \vartheta^{(2)^2}\right] + \frac{\mu_2\mu_4}{4}\left(\mathbb{C}\mathrm{ov}^2\left[\vartheta^{(1)}, \vartheta^{(2)^2}\right] + \mathbb{C}\mathrm{ov}^2\left[\vartheta^{(1)^2}, \vartheta^{(2)}\right]\right) + \dots ,$$

where μ_m denotes the mth moment of $v(\cdot)$, which is assumed to exist. Consequently, $\mathcal{T}_w$ may be written as a weighted sum involving classical (squared) covariances between powers of $\vartheta^{(1)}$ and $\vartheta^{(2)}$. In this regard, the role of $v(\cdot)$ is to assign weights to these covariances via its population moments.

4. Simulation study

4.1 Toroidal distributions considered

To explore various shapes of dependence between $\vartheta^{(1)}$ and $\vartheta^{(2)}$, with a strength of dependence controlled by the value of a single parameter, we consider the four following joint parametric distributions of $\left(\vartheta^{(1)}, \vartheta^{(2)}\right)$, all supported on $\mathbb{T}^2$:

(*i*) The ParaBolic distribution PB(p), defined by $\vartheta^{(1)} \sim \mathrm{Unif}\left(\mathbb{T}\right)$ and $\vartheta^{(2)} = 2\left[p\left(\vartheta^{(1)}\right)^2 + (1-p)U^2\right]/\pi - \pi$, where $U \sim \mathrm{Unif}\left(\mathbb{T}\right)$ is independent of $\vartheta^{(1)}$ and $p \in [0, 1]$.

(*ii*) The (centered) Bivariate Wrapped Cauchy distribution as given in Pewsey and Kato (2016), denoted $\mathrm{BWC}(\rho_1, \rho_2, \rho)$ and with density being

$$\left(\vartheta^{(1)}, \vartheta^{(2)}\right) \mapsto c_0\big\{c_1 - c_2 \cos\left(\vartheta^{(1)}\right) - c_3 \cos\left(\vartheta^{(2)}\right) \\ - c_4 \cos\left(\vartheta^{(1)}\right) \cos\left(\vartheta^{(2)}\right) - c_5 \sin\left(\vartheta^{(1)}\right) \sin\left(\vartheta^{(2)}\right)\big\}^{-1},$$

where c_j, $j = 0, \ldots, 5$, are closed-form constants depending on $\rho_1, \rho_2, |\rho| \in [0, 1)$.

(*iii*) The (centered) Bivariate Cosine von Mises model with *positive* interaction, denoted $\mathrm{BCvM}(\kappa_1, \kappa_2, \kappa_3)$ and with density described in Equation (2.4).

(*iv*) The (centered) Bivariate von Mises by Shieh and Johnson (2005), denoted $\mathrm{BvM}(\kappa_1, \kappa_2, \mu_g, \kappa_g)$ and with density

$$\left(\vartheta^{(1)}, \vartheta^{(2)}\right) \mapsto f_1\left(\vartheta^{(1)}\right) f_2\left(\vartheta^{(2)}\right) f_g\left(2\pi\left\{F_1\left(\vartheta^{(1)}\right) - F_2\left(\vartheta^{(2)}\right)\right\}\right),$$

where f_j and F_j are respectively the marginal density and distribution functions of a zero-mean von Mises with concentration $\kappa_j \geq 0$, $j = 1, 2$, and the link density f_g is that of a von Mises with circular mean $\mu_g \in \mathbb{T}$ and concentration $\kappa_g \geq 0$.

The last parameter in each one of the four distributions controls the degree of dependence, with $p = \rho = \kappa_3 = \kappa_g = 0$ producing independence between $\vartheta^{(1)}$ and $\vartheta^{(2)}$.

Sampling from (i) is straightforward. For (iii), we used the function `rvmcos` from the `BAMBI` (v. 2.3.0) package (Chakraborty and Wong, 2019). One can simulate from (iv) using Algorithm A for von Mises marginals in Shieh and Johnson (2005). R codes for sampling (ii) and (iv) make use of package `circular` (v. 0.4-93) (Agostinelli and Lund, 2017) and were kindly provided by Arthur Pewsey. They are available from the authors. Figure 1 shows different scatterplots obtained from the considered distributions.

4.2 Empirical powers

We investigate the empirical size and power of our three families of tests. More specifically, we consider the tests based on statistics $T_n(\boldsymbol{r}_1)$ and $T_n(\boldsymbol{r}_2)$ with $\boldsymbol{r}_1 = (1, 1)$ and $\boldsymbol{r}_2 = (1, -1)$, $\boldsymbol{\Delta}_n \equiv \boldsymbol{\Delta}_n(\boldsymbol{r}^{(c)}, \boldsymbol{r}^{(s)})$ with $\boldsymbol{r}^{(c)} = (1, -1, 1, 1)$ and $K = 0$, and $T_{n,\lambda}$ for $\lambda \in \{0.1, 0.5, 1.0, 2.0\}$ as well as $S_{n,\mathrm{dc}}$. We also consider four competitors, namely the smoothing-based test of García-Portugués et al. (2015, Section C.2), denoted by G_n, the test based on the weighted U-statistic of Shieh et al. (1994, p. 737), denoted by U_n, the correlation test of Zhan et al. (2019, p. 1835) based on the statistic $\hat{\rho}_0$, and the omnibus test of Rothman (1971) based on the integrated empirical independence process denoted by C_n. For G_n, we set the bandwidths (h_1, h_2) respectively to $(1.00, 0.70)$, $(1.00, 1.00)$, $(0.50, 0.50)$ and $(0.55, 0.55)$ for the four scenarios of dependence considered. These bandwidths are sensible, as they are the empirical medians of 10^3 marginal "rule-of-thumb" bandwidths (García-Portugués, 2013) for $n = 20, 50$ and for each of the considered scenarios.

The empirical power of these tests is compared by generating $M = 10^5$ independent samples of sizes $n = 20$ and $n = 50$ from the distributions (i)–(iv), for varying dependence strengths. Results for a significance level $\alpha = 5\%$ are summarised in Table 1 for $n = 50$ below and in Table E.2 in the SM for $n = 20$. In these tables, the first row in each panel corresponds to the independence case, while subsequent rows represent increasing dependence strength. The extreme cases $p = 1$ and $\rho = 1$ give functional dependence. We proceed as follows to compute critical values under $\mathcal{H}_0$. For a given sample size n, and a given bivariate parametric alternative distribution $\mathcal{D}(\theta)$, we generate two independent samples $\big(\vartheta_1^{(1)}, \vartheta_1^{(2)}\big), \ldots, \big(\vartheta_n^{(1)}, \vartheta_n^{(2)}\big)$ and $\big(\tilde{\vartheta}_1^{(1)}, \tilde{\vartheta}_1^{(2)}\big), \ldots, \big(\tilde{\vartheta}_n^{(1)}, \tilde{\vartheta}_n^{(2)}\big)$ from $\mathcal{D}(\theta)$. Critical values are then obtained by computing empirical quantiles from the sample $\big(\vartheta_1^{(1)}, \tilde{\vartheta}_1^{(2)}\big), \ldots, \big(\vartheta_n^{(1)}, \tilde{\vartheta}_n^{(2)}\big)$. While this necessitates to generate two samples, it is much faster than rely-

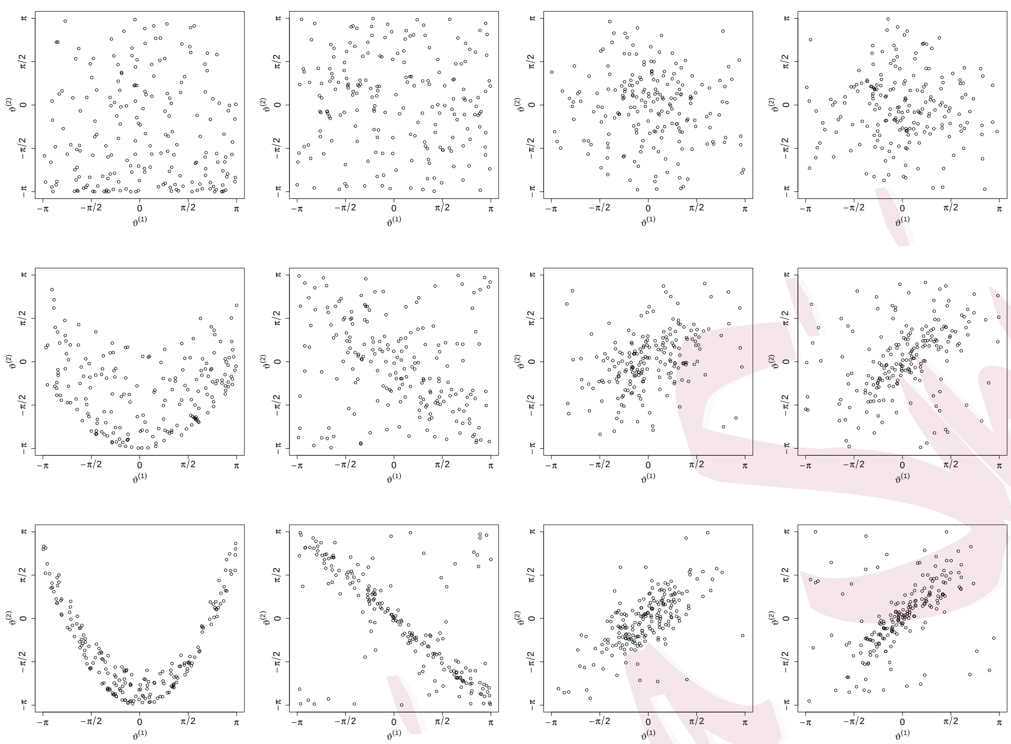

Figure 1: Scatterplots generated from the simulation scenarios considered in the simulation study. From left to right, columnwise: (i) PB(p) for $p = 0, 0.4, 0.8$ (top to bottom); (ii) BWC$(0.1, 0.1, -\rho)$ for $\rho = 0, 0.4, 0.8$; (iii) BCvM$(1, 1, \kappa_3)$ for $\kappa_3 = 0, 1, 2$; (iv) BvM$(1, 1, 0, \kappa_g)$ for $\kappa_g = 0, 1, 2$. The sample size considered is $n = 200$.

ing on a permutation approach. Moreover, this ensures that our empirical power values measure an ability to detect dependence by completely disregarding any potential marginal effect since the marginal distributions of

$\left(\vartheta^{(1)}, \tilde{\vartheta}^{(2)}\right)$ are the same as those of $\left(\vartheta^{(1)}, \vartheta^{(2)}\right)$, under the null and the alternative, respectively. We present in Section E of the SM an extensive simulation study showing that this much faster approach is equivalent, in terms of comparing the power values of the twelve tests under scrutiny, to obtaining by permutations the critical values. Both approaches lead to very close power values for all four scenarios considered.

Table 1: Empirical level and power (in %) for the distributions PB(p), BWC($0.1, 0.1, -\rho$), BCvM($1, 1, \kappa_3$), and BvM($1, 1, 0, \kappa_g$) (top to bottom), for $\alpha = 5\%$ and $n = 50$. On each row, the largest power value is in bold, and any other power value falling in the Wilson (1927)'s 95% binomial confidence interval for the theoretical power of this best test is also in bold.

		$T_n(\boldsymbol{r}_1)$	$T_n(\boldsymbol{r}_2)$	$\boldsymbol{\Delta}_n$	$T_{n,0.1}$	$T_{n,0.5}$	$T_{n,1.0}$	$T_{n,2.0}$	$S_{n,\text{dc}}$	G_n	U_n	$\hat{\rho}_0$	C_n
	0.0	5.00	5.05	4.95	4.95	4.97	4.99	5.00	4.97	4.98	4.74	4.86	4.79
	0.2	8.87	8.88	10.47	12.54	16.37	22.81	**30.48**	19.31	15.04	11.17	12.91	12.17
	0.4	39.03	38.88	47.93	75.72	82.01	**85.90**	79.04	83.18	79.30	18.07	25.01	21.26
p	0.6	69.00	68.85	87.13	**99.98**	**99.98**	**99.98**	99.66	**99.99**	**99.98**	33.51	46.85	52.92
	0.8	73.27	72.84	95.66	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	56.63	62.09	99.89
	1.0	66.87	66.93	83.73	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	71.16	70.48	**100.00**
	0.0	5.04	4.85	4.94	4.96	5.01	5.03	5.06	5.03	4.97	5.03	5.01	4.93
	0.2	30.15	4.74	23.26	**30.82**	30.03	26.60	14.84	26.78	**30.82**	18.63	20.68	3.53
ρ	0.4	74.85	4.96	69.17	**91.26**	90.70	87.91	66.01	90.55	91.23	77.59	80.21	28.50
	0.6	91.11	5.21	89.59	**99.98**	**99.98**	**99.97**	99.40	**99.98**	**99.98**	99.71	99.77	95.59
	0.8	97.55	5.14	97.17	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
	0.0	4.86	4.93	4.91	4.98	4.99	5.11	5.04	5.04	5.09	5.09	4.88	4.98
	0.5	10.07	**39.46**	27.43	32.74	30.37	24.92	13.22	30.06	26.57	13.20	30.75	20.10
	1.0	21.70	**88.71**	77.86	81.41	78.24	69.57	40.01	78.70	72.55	41.25	78.11	57.48
κ_3	1.5	35.22	**99.11**	96.89	97.35	96.42	93.08	71.47	97.02	94.39	72.01	96.65	84.89
	2.0	45.13	**99.94**	99.62	99.61	99.45	98.77	90.13	99.70	99.06	88.74	99.63	95.55
	3.0	56.22	**100.00**	**100.00**	99.99	99.98	99.96	99.23	**100.00**	99.97	98.47	99.99	99.66
	0.0	4.79	4.87	4.79	4.94	4.82	4.84	4.93	4.77	4.81	4.94	4.95	5.05
	0.5	6.49	**43.85**	31.75	34.98	40.05	42.11	32.66	40.87	41.60	27.76	43.16	41.95
	1.0	8.84	93.57	88.53	89.49	94.65	**96.08**	90.66	95.41	95.72	88.02	94.06	95.20
κ_g	1.5	10.85	99.74	99.41	99.68	99.96	**99.98**	99.86	**99.98**	99.97	99.71	99.90	99.96
	2.0	12.77	99.97	99.96	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
	3.0	15.27	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

A value of empirical level outside the interval $[4.86, 5.14]$ indicates that the nominal level (5%) does not fall within the corresponding realised 95% confidence interval. Given that 96 empirical levels were computed, a Bonferroni correction permits to extend the acceptable range to be within $[4.76, 5.24]$. The only observed marked discrepancy between nominal and empirical levels (i.e., 4.74%) occurs for the U_n-based test, the reason being that its statistic is a discrete random variable.

The following conclusions can be drawn from Table 1 and Table E.2 in the SM:

1. The optimality of $\phi_c^{(n)}(1,-1)$ is corroborated for alternatives (iii). In general, $\phi_c^{(n)}(1,-1)$ has a reasonable power against positive-correlation alternatives (iii) and (iv), while it has very low power against negative-correlation alternatives (ii). An opposite behavior for $\phi_c^{(n)}(1,1)$ is evidenced.

2. The $\phi^{(n)}((1,-1,1,1))$ test behaves as expected on merging the benefits of $\phi_c^{(n)}(1,-1)$ and $\phi_c^{(n)}(1,1)$, providing competitive powers (in particular, against the tests based on U_n, $\hat{\rho}_0$, and C_n) in all scenarios and against positive/negative correlation. It suffers a moderate loss of power with respect to the best-performing test among $\phi_c^{(n)}(1,-1)$ and $\phi_c^{(n)}(1,1)$.

3. $\phi^{(n)}(\lambda)$ is a very competitive test overall. For at least one choice of $\lambda \in \{0.1, 0.5, 1.0\}$ per simulation scenario, it dominates the rest of the tests for distributions (i), (ii), and (iv) or offers a competitive power for distribution (iii). In particular, $\phi^{(n)}(\lambda)$ dominates the three competing tests based on U_n, $\hat{\rho}_0$, and C_n, for at least one choice of $\lambda \in \{0.1, 0.5, 1.0\}$ and for all scenarios. This dominance is more marked when comparing to the only competing omnibus test, the one based on C_n.

4. The choice of λ is influential on the power of $\phi^{(n)}(\lambda)$. The choice $\lambda = 2$ is seen to be systematically worse, which might be explained by the fact that, in this case, the kernel (3.13) can be negative. Therefore, the power of $\phi^{(n)}(\lambda)$ might be drained by reducing the value of $T_{n,\lambda}$ for certain pairwise angles $\vartheta_{jk}^{(\ell)}$, $j, k = 1, \ldots, n$, $\ell = 1, 2$.

5. $\phi_{\mathrm{dc}}^{(n)}$ performs very similarly to $\phi^{(n)}(\lambda)$, and depending on the value of λ its power is above or below that of $\phi^{(n)}(\lambda)$.

6. The three competing tests based on U_n, $\hat{\rho}_0$, and C_n have a comparative poorer performance in scenario (i), which does not have a positive/negative-dependence pattern. In this case, our four tests clearly outperform the competition by a large margin.

7. The test based on G_n offers a comparable power to characteristic-function tests when the dependence is moderate to strong.

Overall, we recommend the use of the test $\phi^{(n)}(\lambda)$ for $\lambda \in \{0.1, 1.0\}$ given its similar performance to $\phi^{(n)}_{\mathrm{dc}}$ and its faster application. In particular, we corroborated that applying $\phi^{(n)}(\lambda)$ is five times faster than $\phi^{(n)}_{\mathrm{dc}}$ when $n = 50$ and $B = 10^4$ permutations are used.

5. Data applications

5.1 Wildfires

Barros et al. (2012) identified the existence of preferential orientations of wildfires on 102 characteristic watersheds of Portugal (see Figure 2) determined in a data-driven fashion. Their analysis quantified annual wildfire orientations through the axial direction (e.g., North–South) of the first principal component of a wildfire perimeter. These perimeters were obtained from Landsat imagery of Portugal after the end of wildfire season and were then assigned to different watersheds according to the position of their centroids. Wildfire orientation is likely explained by dominant weather during the Portuguese wildfire season (Barros et al., 2012) and is significantly associated with the size of burnt area (García-Portugués et al., 2014).

We aim to formally address the existence of significant long-term and short-term temporal patterns in the Portuguese wildfire orientations. As in García-Portugués et al. (2014), we restrict to the $26,870$ wildfires mapped in 1985–2005 due to the higher resolution of satellite imagery for that period (minimum mapping unit of 5 hectares). We then perform two data pre-processing steps. First, since a wildfire (axial) orientation is a π-periodic angular variable ϑ supported in $[0,\pi)$, we consider 2ϑ, a standard circular variable supported in $[0,2\pi)$. With this simple transformation, the angles $\{0,\pi/2,\pi,3\pi/2\}$ represent the $\{\text{E–W},\text{NE–SW},\text{N–S},\text{NW–SE}\}$ orientations, respectively. Second, we summarize the preferred orientation of the wildfires in each watershed by their weighted circular sample mean, with weights being the product between the proportion of explained variance and the burnt area of wildfire perimeter. The resulting dataset has 102 representative wildfire orientations, shown in Figure 2 for 1986–1995 and 1996–2005.

When applied to the datasets displayed in Figure 2, the tests $\phi_c^{(n)}(1,1)$, $\phi_c^{(n)}(1,-1)$, $\phi^{(n)}((1,-1,1,1))$, $\phi^{(n)}(0.1)$, and $\phi^{(n)}(1)$ yielded p-values 0.0593, 0.0798, 0.0730, 0, and 0.0003 (using 10^4 permutations for $\phi^{(n)}(\lambda)$), respectively. Therefore, significant long-term dependence is present in the orientation of wildfires. Short-term temporal dependence was also investigated by testing the null hypotheses of independence associated to the 20 consecutive pairs of years in 1985–2005 and applying Benjamini and Yekutieli

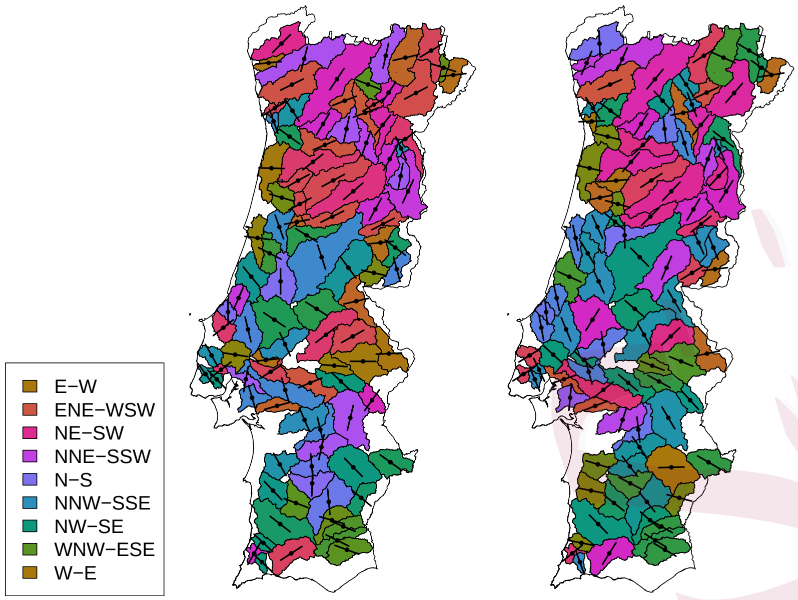

Figure 2: Weighted average orientations of the wildfires from 1986–1995 (left) and 1996–2005 (right), for each of the 102 watersheds determined in Barros et al. (2012).

(2001)'s correction procedure. None of the (corrected) p-values of the five tests were below the 5% significance level. For the 10% significance level, only three $\phi_c^{(n)}(1,-1)$ tests and one $\phi^{(n)}(0.1)$ test were significant. To investigate mid-term temporal dependence, we repeated the analysis for pairs of consecutive periods of 5-years (12 pairs) and 3-years (16 pairs). The proportion of (corrected) 5%-significant $\phi_c^{(n)}(1,-1)$ tests raised to 0.5 and 0.1875, respectively, while again no $\phi_c^{(n)}(1,1)$ tests were significant at any usual significance level. The corresponding proportions for the tests $\phi^{(n)}(0.1)$ and

$\phi^{(n)}(1)$ were 0.75 and 0.8333 (5-years), and 0 and 0.125 (3-years). In conclusion, significant positive dependence of the orientations of wildfires is present among spans of 10 and 5 years, while no significant dependence is found on consecutive years. Both conclusions support the existence of drivers of the orientations in the long-term, such as dominant weather during the wildfire season (Barros et al., 2012).

5.2 Long-period comets

Long-period comets are thought to originate in the Oort cloud, a widely accepted model posing the existence of a roughly spherical reservoir of icy planetesimals in the limits of the Solar System. It is believed that these icy planetesimals become long-period comets when randomly captured in heliocentric orbits due to the effect of several gravitational forces (see, e.g., Sections 5 and 7.2 in Dones et al., 2015). This conjectured origin explains the highly-characteristic nearly-isotropic distribution of the long-period comets' orbits (e.g., Wiegert and Tremaine, 1999). Such distribution is markedly different from that of short-period comets, who originate at the flattened Kuiper belt and whose orbits cluster about the ecliptic plane.

An orbit with *inclination* $i \in [0, \pi]$ and *longitude of the ascending node* $\Omega \in [0, 2\pi)$ has directed normal vector $(\sin(i)\sin(\Omega), -\sin(i)\cos(\Omega), \cos(i))'$

to the orbit's plane (e.g., Jupp et al., 2003). Using this parametrization, the Projected Cramér–von Mises, Projected Rothman, and Projected Anderson–Darling tests (García-Portugués et al., 2022) reject the uniformity of the orbits of long-period comets (p-values smaller than 0.0197) using the records of the JPL Small-Body Database Search Engine (`https://ssd.jpl.nasa.gov/tools/sbdb_query.html`) as of May 2022. The rejection may be driven by a truly non-uniform population or, according to the analysis in Jupp et al. (2003), by the existence of significant observational bias on the available records. As Jupp et al. (2003) explain, bias is induced by how comet search programs maximize success detection chances by preferentially exploring regions about the ecliptic plane, as those are where most asteroids and short-period comets cluster.

A possible manifestation of observational bias, both in long- and short-period comets, is in the appearance of serial dependence in the orbits of observed comets. To assess the existence of such serial dependence in a nonparametric way, we investigated the lag-1 dependence of the time series of Ω. We used the lagged samples (Ω_i, Ω_{i+1}), $i = 1, \ldots, n-1$, with $n = 623$ for long-period comets and $n = 905$ for short-period comets (see Figure 3), and applied to them several of the new tests of independence. The dataset is available through the `comets` object of the `sphunif` R package (v. 1.0.2)

(García-Portugués and Verdebout, 2022), and is sorted through the JPL's database ID, which is assigned chronologically based on the discovery of new comets.

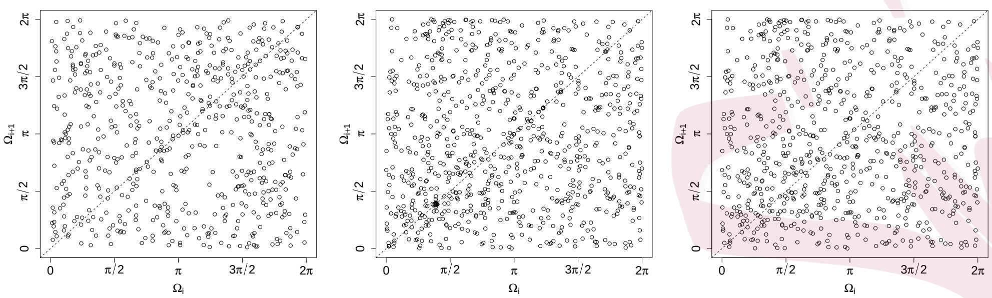

Figure 3: Scatterplots of (Ω_i, Ω_{i+1}) for long-period comets (left) and short-period comets (center and right). The clusters appearing on the diagonal of the central plot disappear once the fragments of disintegrating comets are removed from the dataset (right plot).

The tests $\phi_c^{(n)}(1,1)$, $\phi_c^{(n)}(1,-1)$, $\phi^{(n)}((1,-1,1,1))$, $\phi^{(n)}(0.1)$, and $\phi^{(n)}(1)$ yielded p-values 0.6063, 0.3710, 0.8941, 0.1745, and 0.3767, respectively, for the lagged sample of long-period comets. Therefore, no evidence against lag-1 independence on the series $\{\Omega_i\}_{i=1}^n$ for long-period comets is found, indicating that, if significant observational bias is present, it does not significantly induce the most obvious form of serial dependence on Ω. For short-period comets, the p-values were 0.0085, 4.8×10^{-8}, 2.9×10^{-7}, 0, and 0, thus signaling significant lag-1 dependence on the series of longitudes. A data inspection reveals that this rejection is a consequence of the clusters formed by fragments of disintegrating comets (see central plot of Figure 3). For example, there is a sequence of 68 records corresponding to fragments of the "73P/Schwassmann–Wachmann 3" comet. After removing 121 fragment records, the tests gave p-values 0.6702, 0.5066, 0.9609, 0.5608, and 0.6207, hence not rejecting lag-1 independence on the longitudes of non-disintegrating short-period comets. The same test decisions at the 5% significance level were obtained when using lags of order two and three in the whole analysis. The same test decisions were also obtained when, first, sorting the database records according to the dates of the first observations used in the fit of the orbits (gives a different chronological ordering) and, then, repeating the whole analysis while applying Benjamini and Yekutieli (2001)'s correction procedure.

Supplementary materials

The Supplementary Materials (SM) contain the proofs of all the stated results. In addition, they provide the derivation of the covariance matrix $\boldsymbol{\Sigma}$, detail the permutation algorithm applied to $\phi^{(n)}(\lambda)$, and give further simulation results.

Acknowledgments

E. García-Portugués acknowledges support by grants PGC2018-097284-B-100 and IJCI-2017-32005 from the Spain's Ministry of Economy and Competitiveness. The two grants were partially co-funded by the European Regional Development Fund. Part of this research was carried out while S. G. Meintanis was visiting P. Lafaye de Micheaux at UNSW, and hereby hospitality and financial support are sincerely acknowledged. T. Verdebout's research is supported by the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles. This research includes computations performed using the computational cluster Katana supported by Research Technology Services at UNSW Sydney. The comments by two anonymous referees are appreciated.

References

Agostinelli, C. and U. Lund (2017). *R package circular: Circular Statistics*. R package version 0.4-93.

Barros, A. M. G., J. Pereira, and U. J. Lund (2012). Identifying geographical patterns of wildfire orientation: a watershed-based analysis. *For. Ecol. Manag. 264*, 98–107.

Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat. 29*(4), 1165–1188.

Boomsma, W., K. V. Mardia, C. C. Taylor, J. Ferkinghoff-Borg, A. Krogh, and T. Hamelryck (2008). A generative, probabilistic model of local protein structure. *Proc. Natl. Acad. Sci. U.S.A. 105*(26), 8932–8937.

Chakraborty, S. and S. W. K. Wong (2019). *BAMBI: Bivariate Angular Mixture Models*. R package version 2.3.0.

Chakraborty, S. and X. Zhang (2019). Distance metrics for measuring joint dependence with application to causal inference. *J. Am. Stat. Assoc. 114*(528), 1638–1650.

Chen, F., S. G. Meintanis, and L. X. Zhu (2019). On some characterizations and multidimensional criteria for testing homogeneity, symmetry and independence. *J. Multivar. Anal. 173*, 125–144.

Csörgő, S. and P. Hall (1982). Estimable versions of Griffiths' measure of association. *Aust. J. Stat. 24*(3), 296–308.

Csörgő, S. (1985). Testing for independence by the empirical characteristic function. *J. Multivar. Anal. 16*(3), 290–299.

Dones, L., R. Brasser, N. Kaib, and H. Rickman (2015). Origin and evolution of the cometary reservoirs. *Space Sci. Rev 197*(1), 191–269.

Fan, Y., P. Lafaye de Micheaux, S. Penev, and D. Salopek (2017). Multivariate nonparametric test of independence. *J. Multivar. Anal. 153*, 189–210.

García-Portugués, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat. 7*, 1655–1685.

García-Portugués, E., A. M. G. Barros, R. M. Crujeiras, W. González-Manteiga, and J. Pereira (2014). A test for directional-linear independence, with applications to wildfire orientation and size. *Stoch. Environ. Res. Risk Assess. 28*(5), 1261–1275.

García-Portugués, E., R. M. Crujeiras, and W. González-Manteiga (2015). Central limit theorems for directional and linear random variables with applications. *Stat. Sin. 25*(3), 1207–1229.

García-Portugués, E., P. Navarro-Esteban, and J. A. Cuesta-Albertos (2022). On a projection-based class of uniformity tests on the hypersphere. *Bernoulli to appear*.

García-Portugués, E. and T. Verdebout (2022). *sphunif: Uniformity Tests on the Circle, Sphere, and Hypersphere*. R package version 1.0.2.

Gretton, A., R. Herbrich, A. Smola, O. Bousquet, and B. Schoelkopf (2005). Kernel methods for measuring independence. *J. Mach. Learn. Res. 6*(70), 2075–2129.

Hlávka, Z., M. Hušková, and S. G. Meintanis (2011). Testing independence in non-parametric regression models. *J. Multivar. Anal. 102*(7), 816–827.

Jammalamadaka, S. R. and A. SenGupta (2001). *Topics in Circular Statistics*, Volume 5 of *Series on Multivariate Analysis*. Singapore: World Scientific.

Jupp, P. E., P. T. Kim, J.-Y. Koo, and P. Wiegert (2003). The intrinsic distribution and selection bias of long-period cometary orbits. *J. Am. Stat. Assoc. 98*(463), 515–521.

Jupp, P. E. and K. V. Mardia (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika 67*(1), 163–173.

Kato, S. and A. Pewsey (2015). A Möbius transformation-induced distribution on the torus. *Biometrika 102*(2), 359–370.

Kent, J. T., K. V. Mardia, and C. C. Taylor (2008). Modelling strategies for bivariate circular data. In S. Barber, P. D. Baxter, A. Gusnanto, and K. V. Mardia (Eds.), *LASR 2008 – The Art & Science of Statistical Bioinformatics*, Leeds, pp. 70–73. Department of Statistics, University of Leeds.

Ley, C. and T. Verdebout (2017). *Modern Directional Statistics.* Chapman & Hall/CRC Interdisciplinary Statistics Series. Boca Raton: CRC Press.

Mardia, K. V. (1975). Statistics of directional data. *J. R. Stat. Soc. Ser. B Methodol. 37*(3), 349–393.

Mardia, K. V. and J. Frellsen (2012). Statistics of bivariate von Mises distributions. In T. Hamelryck, K. Mardia, and J. Ferkinghoff-Borg (Eds.), *Bayesian Methods in Structural Bioinformatics*, Statistics for Biology and Health, pp. 159–178. Berlin: Springer.

Mardia, K. V. and P. E. Jupp (1999). *Directional Statistics.* Wiley Series in Probability and Statistics. Chichester: Wiley.

Mardia, K. V. and M. L. Puri (1978). A spherical correlation coefficient robust against scale. *Biometrika 65*(2), 391–395.

Mardia, K. V., C. C. Taylor, and G. K. Subramaniam (2007). Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics 63*(2), 505–512.

Meintanis, S. and T. Verdebout (2019). Le Cam maximin tests for symmetry of circular data based on the characteristic function. *Stat. Sin. 29*(3), 1301–1320.

Meintanis, S. G. and G. Iliopoulos (2008). Fourier methods for testing multivariate independence. *Comput. Stat. Data Anal. 52*(4), 1884–1895.

Pewsey, A. and E. García-Portugués (2021). Recent advances in directional statistics. *Test 30*(1), 1–58.

Pewsey, A. and S. Kato (2016). Parametric bootstrap goodness-of-fit testing for Wehrly–Johnson bivariate circular distributions. *Stat. Comput. 26*(6), 1307–1317.

Puri, M. L. and J. S. Rao (1977). Problems of association for bivariate circular data and a new test of independence. In P. R. Krishnaiah (Ed.), *Multivariate Analysis IV*, Amsterdam, pp. 513–522. North-Holland.

Rivest, L.-P. (1988). A distribution for dependent unit vectors. *Commun. Stat. Theory Methods 17*(2), 461–483.

Rothman, E. D. (1971). Tests of coordinate independence for a bivariate sample on a torus. *Ann. Math. Stat. 42*(6), 1962–1969.

Shieh, G. S. and R. A. Johnson (2005). Inference based on a bivariate distribution with von Mises marginals. *Ann. Inst. Stat. Math. 57*(4), 789–802.

Shieh, G. S., R. A. Johnson, and E. W. Frees (1994). Testing independence of bivariate circular data and weighted degenerate U-statistics. *Stat. Sin. 4*(2), 729–747.

Singh, H., V. Hnizdo, and E. Demchuk (2002). Probabilistic model for two dependent circular variables. *Biometrika 89*(3), 719–723.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007). Measuring and testing dependence by correlation of distances. *Ann. Stat. 35*(6), 2769–2794.

Watson, G. S. and R. J. Beran (1967). Testing a sequence of unit vectors for serial correlation. *J. Geophys. Res. 72*(22), 5655–5659.

Wehrly, T. E. and R. A. Johnson (1980). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika 67*(1), 255–256.

Wiegert, P. and S. Tremaine (1999). The evolution of long-period comets. *Icarus 137*(1), 84–121.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *J. Am. Stat. Assoc. 22*(158), 209–212.

Zhan, X., T. Ma, S. Liu, and K. Shimizu (2019). On circular correlation for data on the torus. *Stat. Pap. 60*(6), 1827–1847.

Department of Statistics, Carlos III University of Madrid, Leganés, Spain

E-mail: edgarcia@est-econ.uc3m.es

School of Mathematics and Statistics, UNSW, Sydney, Australia

Desbrest Institute of Epidemiology and Public Health, Univ Montpellier, INSERM, Montpellier, France

AMIS, Université Paul Valéry Montpellier 3, France

E-mail: lafaye@unsw.edu.au

Department of Economics, National and Kapodistrian University of Athens, Athens, Greece

Unit for Pure and Applied Analytics, North-West University, Potchefstroom, South Africa

E-mail: simosmei@econ.uoa.gr

Département de Mathématique, Université libre de Bruxelles, Brussels, Belgium

ECARES, Université libre de Bruxelles, Brussels, Belgium

E-mail: tverdebo@ulb.ac.be