CENTRAL LIMIT THEOREMS FOR DIRECTIONAL AND LINEAR RANDOM VARIABLES WITH APPLICATIONS

Eduardo García-Portugués, Rosa M. Crujeiras and
Wenceslao González-Manteiga

University of Santiago de Compostela

Abstract: A central limit theorem for the integrated squared error of the directional-linear kernel density estimator is established. The result enables the construction and analysis of two testing procedures based on squared loss: a nonparametric independence test for directional and linear random variables and a goodness-of-fit test for parametric families of directional-linear densities. Limit distributions for both test statistics, and a consistent bootstrap strategy for the goodness-of-fit test, are developed for the directional-linear case and adapted to the directional-directional setting. Finite sample performance for the goodness-of-fit test is illustrated in a simulation study. This test is also applied to datasets from biology and environmental sciences.

Key words and phrases: Directional data, goodness-of-fit, independence test, kernel density estimation, limit distribution.

1. Introduction

Statistical inference on random variables comprises estimation and testing procedures that allow one to characterize the underlying distribution, regardless the variables nature and/or dimension. Specifically, density estimation stands out as a basic problem in statistical inference, for which parametric and nonparametric approaches have been explored. In nonparametrics, kernel density estimation (see Silverman (1986), Scott (1992), or Wand and Jones (1995), as comprehensive references for scalar random variables) provides a simple and intuitive way to explore and do inference on random variables. Among other contexts, kernel density estimation has been also adapted to directional data (see Mardia and Jupp (2000)). This data on the q-dimensional sphere arises, for example, in meteorology when measuring wind direction; in proteomics, when studying the angles in protein structure (circular data, $q = 1$, see Fernández-Durán (2007)); in astronomy, with the stars positions in the celestial sphere ($q = 2$, see García-Portugués (2013)); in text mining, when codifying documents in the vector space model (large q, see Chapter 6 in Srivastava and Sahami (2009)). Some early works on

kernel density estimation with directional data are the papers by Hall, Watson, and Cabrera (1987) and Bai, Rao, and Zhao (1988), who introduced kernel density estimators and their properties (bias, variance and uniformly strong consistency, among others). The estimation of the density derivatives was studied by Klemelä (2000), and Zhao and Wu (2001) stated a Central Limit Theorem (CLT) for the Integrated Squared Error (ISE) of the directional kernel density estimator. Some recent works deal with the bandwidth selection problem, such as Taylor (2008) and Oliveira, Crujeiras, and Rodríguez-Casal (2012), devoted to circular data and García-Portugués (2013), for a general dimension. In some contexts, joint density models for directional and linear random variables are useful (e.g. for describing wind direction and SO_2 concentration García-Portugués, Crujeiras, and González-Manteiga (2013a)). In this setting, a kernel density estimator for directional-linear data was proposed and analysed by García-Portugués, Crujeiras, and González-Manteiga (2013b).

Regardless of estimation purposes, kernel density estimators have been extensively used for the development of goodness-of-fit tests (see González-Manteiga and Crujeiras (2013) for a review) and independence tests. For example, Bickel and Rosenblatt (1973) and Fan (1994) provided goodness-of-fit tests for parametric densities for real random variables. Similarly, in the directional setting, Boente, González-Manteiga, and Rodríguez (2014) presented a goodness-of-fit test for parametric directional densities. For assessing independence between two linear random variables, Rosenblatt (1975) proposed a test statistic based on the squared difference between the joint kernel density estimator and the product of the marginal ones (see also Rosenblatt and Wahlen (1992)). This idea was adapted to the directional-linear setting by García-Portugués et al. (2014), who derived a permutation independence test and compared its performance with the testing proposals given by Mardia (1976), Johnson and Wehrly (1978), and Fisher and Lee (1981) in this context.

The main device for the goodness-of-fit and independence tests is the CLT for the ISE of the kernel density estimator, and the aim of this work is to provide such a result for the directional-linear kernel estimator, and use it to derive a goodness-of-fit test for parametric families of directional-linear densities and an independence test for directional and linear variables. The CLT is obtained by proving an extended version of Theorem 1 in Hall (1984). The goodness-of-fit test follows by taking the ISE between the joint kernel estimator and a smoothed parametric estimate of the unknown density as a test statistic. For the independence test, the test statistic introduced in García-Portugués et al. (2014) is considered and its asymptotic properties are studied. Jointly with the asymptotic distribution, a bootstrap resampling strategy to calibrate the goodness-of-fit test is investigated. Finite sample performance of the goodness-of-fit test is checked through an extensive simulation study, and this methodology is

applied to analyse datasets from forestry and proteomics. In addition, the results obtained for the directional-linear case are adapted to the directional-directional context.

The rest of this paper is organized as follows. Section 2 presents some background on kernel density estimation for directional and linear random variables. Section 3 includes the CLT for the ISE of the directional-linear estimator and its extension to the directional-directional setting. The independence test for directional and linear variables is presented in Section 4. The goodness-of-fit test for simple and composite null hypotheses, its bootstrap calibration and extensions are given in Section 5. The empirical performance of the goodness-of-fit test is illustrated with a simulation study in Section 6 and with applications to datasets in Section 7. Appendix A collects the outline of the main proofs. Technical lemmas and further details on simulations and data analysis are provided as supplementary material, as well as the extensions of the independence test.

2. Background

For simplicity, f denotes the target density along the paper, which may be linear, directional, directional-linear, or directional-directional, depending on the context.

Let Z denote a linear random variable with support $\text{supp}(Z) \subseteq \mathbb{R}$ and density f, and let $Z_1, \ldots, Z_n$ be a random sample of Z. The linear kernel density estimator is defined as

$$\hat{f}_g(z) = \frac{1}{ng} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{g}\right), \quad z \in \mathbb{R},$$

where K denotes the kernel function and $g > 0$ is the bandwidth parameter, which controls the smoothness of the estimator (see Silverman (1986), among others).

Let $\mathbf{X}$ denote a directional random variable with density f and support the q-dimensional sphere, denoted by $\Omega_q = \left\{\mathbf{x} \in \mathbb{R}^{q+1} : x_1^2 + \cdots + x_{q+1}^2 = 1\right\}$. Lebesgue measure in Ω_q is denoted by ω_q and, therefore, a directional density satisfies $\int_{\Omega_q} f(\mathbf{x}) \, \omega_q(d\mathbf{x}) = 1$. When there is no possible confusion, ω_q will also denote the surface area of Ω_q: $\omega_q = \omega_q(\Omega_q) = 2\pi^{(q+1)/2} / \Gamma((q+1)/2)$. The directional kernel density estimator introduced by Hall, Watson, and Cabrera (1987) and Bai, Rao, and Zhao (1988) for a directional density f, based on a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ in the q-sphere, is

$$\hat{f}_h(\mathbf{x}) = \frac{c_{h,q}(L)}{n} \sum_{i=1}^{n} L\left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2}\right), \text{ for } \mathbf{x} \in \Omega_q,$$

where L is the directional kernel, $h > 0$ is the bandwidth parameter and the scalar product of two vectors, $\mathbf{x}$ and $\mathbf{y}$, is denoted by $\mathbf{x}^T\mathbf{y}$, where $\mathbf{x}^T$ is the transpose of the column vector $\mathbf{x}$. $c_{h,q}(L)$ is a normalizing constant depending on the kernel L, the bandwidth h and the dimension q. Specifically, Bai, Rao, and Zhao (1988) has the inverse of the normalizing constant as

$$c_{h,q}(L)^{-1} = \lambda_{h,q}(L)h^q \sim \lambda_q(L)h^q, \tag{2.1}$$

where $\lambda_{h,q}(L) = \omega_{q-1}\int_0^{2h^{-2}} L(r)r^{q/2-1}(2 - rh^2)^{q/2-1}\,dr$ and $\lambda_q(L) = 2^{q/2-1}\omega_{q-1}$ $\int_0^\infty L(r)r^{q/2-1}\,dr$. The notation $a_n \sim b_n$ means that $a_n = b_n(1 + o(1))$.

A usual choice for the directional kernel is $L(r) = e^{-r}$, also known as the von Mises kernel due to its relation with the von Mises-Fisher density (Watson (1983)), $\mathrm{vM}(\boldsymbol{\mu}, \kappa)$, given by

$$f_{\mathrm{vM}}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = C_q(\kappa)\exp\left\{\kappa\mathbf{x}^T\boldsymbol{\mu}\right\}, \quad C_q(\kappa) = \frac{\kappa^{(q-1)/2}}{(2\pi)^{(q+1)/2}\mathcal{I}_{(q-1)/2}(\kappa)},$$

where $\boldsymbol{\mu} \in \Omega_q$ is the directional mean, $\kappa > 0$ is the concentration parameter around the mean, and $\mathcal{I}_\nu$ is the modified Bessel function of order ν.

The kernel estimator for a directional-linear density f based on a random sample $(\mathbf{X}_1, Z_1), \ldots, (\mathbf{X}_n, Z_n)$, with $(\mathbf{X}_i, Z_i) \in \Omega_q\times\mathbb{R}$, $i = 1, \ldots, n$, was proposed by García-Portugués, Crujeiras, and González-Manteiga (2013b):

$$\hat{f}_{h,g}(\mathbf{x}, z) = \frac{c_{h,q}(L)}{ng}\sum_{i=1}^n LK\left(\frac{1 - \mathbf{x}^T\mathbf{X}_i}{h^2}, \frac{z - Z_i}{g}\right), \quad (\mathbf{x}, z) \in \Omega_q \times \mathbb{R}, \tag{2.2}$$

where LK is a directional-linear kernel, h and g are the bandwidths for the directional and the linear components, respectively, and $c_{h,q}(L)$ is the normalizing constant. For simplicity, the product kernel $LK(\cdot, \cdot) = L(\cdot) \times K(\cdot)$ is considered. To quantify the error of the density estimator, the ISE,

$$\mathrm{ISE}\left[\hat{f}_{h,g}\right] = \int_{\Omega_q\times\mathbb{R}}\left(\hat{f}_{h,g}(\mathbf{x}, z) - f(\mathbf{x}, z)\right)^2 dz\,\omega_q(d\mathbf{x}),$$

can be used. In this expression, the integral is taken with respect to the product measure $\omega_q \times m_{\mathbb{R}}$, with $m_{\mathbb{R}}$ denoting the usual Lebesgue measure in $\mathbb{R}$.

It is possible to define a directional-directional kernel density estimator at $(\mathbf{x}, \mathbf{y}) \in \Omega_{q_1}\times\Omega_{q_2}$ from a random sample $(\mathbf{X}_1, \mathbf{Y}_1), \ldots, (\mathbf{X}_n, \mathbf{Y}_n)$, with $(\mathbf{X}_i, \mathbf{Y}_i) \in \Omega_{q_1} \times \Omega_{q_2}$, $i = 1, \ldots, n$, that comes from a directional-directional density f:

$$\hat{f}_{h_1,h_2}(\mathbf{x}, \mathbf{y}) = \frac{c_{h_1,q_1}(L_1)c_{h_2,q_2}(L_2)}{n}\sum_{i=1}^n L_1\left(\frac{1 - \mathbf{x}^T\mathbf{X}_i}{h_1^2}\right) \times L_2\left(\frac{1 - \mathbf{y}^T\mathbf{Y}_i}{h_2^2}\right).$$

To fix notation, $R(\varphi)$ denotes the integral of the squared function φ^2 along its domain. The following integrals are needed:

$$\mu_2(K) = \int_{\mathbb{R}} z^2 K(z)\, dz, \quad b_q(L) = \frac{\int_0^\infty L(r) r^{q/2}\, dr}{\int_0^\infty L(r) r^{q/2-1}\, dr}.$$

Density derivatives of different orders are denoted as follows:

$$\boldsymbol{\nabla} f(\mathbf{x}, z) = \left(\frac{\partial f(\mathbf{x}, z)}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x}, z)}{\partial x_{q+1}}, \frac{\partial f(\mathbf{x}, z)}{\partial z} \right)^T = (\boldsymbol{\nabla}_{\mathbf{x}} f(\mathbf{x}, z), \nabla_z f(\mathbf{x}, z))^T,$$

$$\boldsymbol{\mathcal{H}} f(\mathbf{x}, z) = \left(\begin{array}{c|c} \left(\frac{\partial^2 f(\mathbf{x}, z)}{\partial x_i \partial x_j} \right) & \frac{\partial^2 f(\mathbf{x}, z)}{\partial \mathbf{x} \partial z} \\ \hline \frac{\partial^2 f(\mathbf{x}, z)}{\partial z \partial \mathbf{x}^T} & \frac{\partial^2 f(\mathbf{x}, z)}{\partial z^2} \end{array} \right) = \left(\begin{array}{c|c} \boldsymbol{\mathcal{H}}_{\mathbf{x}} f(\mathbf{x}, z) & \boldsymbol{\mathcal{H}}_{\mathbf{x}, z} f(\mathbf{x}, z) \\ \hline \boldsymbol{\mathcal{H}}_{\mathbf{x}, z} f(\mathbf{x}, z)^T & \mathcal{H}_z f(\mathbf{x}, z) \end{array} \right).$$

3. Central Limit Theorem for the Integrated Squared Error

Our main result is the CLT for the ISE of the kernel density estimator (2.2).

3.1. Main result

We need the following conditions.

A1 If f is extended from $\Omega_q \times \mathbb{R}$ to $\mathbb{R}^{q+2} \setminus \{ (\mathbf{0}, z) : z \in \mathbb{R} \}$ as $f(\mathbf{x}, z) \equiv f(\mathbf{x}/\|\mathbf{x}\|, z)$ for all $\mathbf{x} \neq \mathbf{0}$ and $z \in \mathbb{R}$, f and its first three derivatives are bounded and uniformly continuous with respect to the product Euclidean norm in $\Omega_q \times \mathbb{R}$, $\|(\mathbf{x}, z)\| = \sqrt{\|\mathbf{x}\|^2 + |z|^2}$.

A2 $L : [0, \infty) \to [0, \infty)$ and $K : \mathbb{R} \to [0, \infty)$ are continuous and bounded; L is nonincreasing such that $0 < \lambda_q(L), \lambda_q(L^2) < \infty, \forall q \geq 1$ and K is a linear density, symmetric around zero and with $\mu_2(K) < \infty$.

A3 $h = h_n$ and $g = g_n$ are sequences of positive numbers such that $h_n \to 0$, $g_n \to 0$, and $n h_n^q g_n \to \infty$ as $n \to \infty$.

The uniform continuity and boundedness up to the second derivatives of f is a common assumption that appears, among others, in Hall (1984) and Rosenblatt and Wahlen (1992), while the assumption on the third derivatives is needed for uniform convergence. The assumption of compact support for the directional kernel L, stated in Zhao and Wu (2001), is replaced by the nonincreasing requirement and the finiteness of $\lambda_q(L)$ and $\lambda_q(L^2)$. These two conditions are less restrictive and allow for consideration of the von Mises kernel. We provide the limit distribution of the ISE for (2.2). The proof is based on a generalization of Theorem 1 in Hall (1984), stated as Lemma 1 in Appendix A.

Theorem 1 (CLT for the directional-linear ISE). *Denote the ISE of $\hat{f}_{h,g}$ by I_n. If* **A1**$-$**A3** *hold, then*

(i) $n^{1/2}\phi(h,g)^{-1/2}\left(I_n - \mathbb{E}[I_n]\right) \xrightarrow{d} \mathcal{N}(0,1)$, *if* $n\phi(h,g)h^q g \to \infty$,

(ii) $n(h^q g)^{1/2}\left(I_n - \mathbb{E}[I_n]\right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2\right)$, *if* $n\phi(h,g)h^q g \to 0$,

(iii) $n(h^q g)^{1/2}\left(I_n - \mathbb{E}[I_n]\right) \xrightarrow{d} \mathcal{N}\left(0, \delta + 2\sigma^2\right)$, *if* $n\phi(h,g)h^q g \to \delta$,

where $0 < \delta < \infty$ *and*

$$\phi(h,g) = \frac{4b_q(L)^2}{q^2}\sigma_{\mathbf{X}}^2 h^4 + \mu_2(K)^2\sigma_Z^2 g^4 + \frac{4b_q(L)\mu_2(K)}{q}\sigma_{\mathbf{X},Z}h^2 g^2,$$

with $\sigma_{\mathbf{X},Z} = \mathbb{C}\mathrm{ov}[\mathrm{tr}[\boldsymbol{\mathcal{H}}_{\mathbf{x}}f(\mathbf{X},Z)], \mathcal{H}_z f(\mathbf{X},Z)]$, $\sigma_{\mathbf{X}}^2 = \mathbb{V}\mathrm{ar}\left[\mathrm{tr}[\boldsymbol{\mathcal{H}}_{\mathbf{x}}f(\mathbf{X},Z)]\right]$ *and* $\sigma_Z^2 = \mathbb{V}\mathrm{ar}\left[\mathcal{H}_z f(\mathbf{X},Z)\right]$. *The remaining constants are*

$$\sigma^2 = R(f) \times \gamma_q \lambda_q(L)^{-4} \int_0^\infty r^{q/2-1}\left\{\int_0^\infty \rho^{q/2-1}L(\rho)\varphi_q(r,\rho)\,d\rho\right\}^2 dr$$

$$\times \int_{\mathbb{R}}\left\{\int_{\mathbb{R}} K(u)K(u+v)\,du\right\}^2 dv,$$

$$\varphi_q(r,\rho) = \begin{cases} L\left(r + \rho - 2(r\rho)^{1/2}\right) + L\left(r + \rho + 2(r\rho)^{1/2}\right), q = 1, \\ \int_{-1}^1 \left(1 - \theta^2\right)^{(q-3)/2}L\left(r + \rho - 2\theta(r\rho)^{1/2}\right)\,d\theta, q \geq 2, \end{cases}$$

$$\gamma_q = \begin{cases} 2^{-1/2}, & q = 1, \\ \omega_{q-1}\omega_{q-2}^2 2^{3q/2-3}, & q \geq 2. \end{cases}$$

The same limit distributions hold in (i)$-$(iii) *if* $\mathbb{E}[I_n]$ *is replaced by*

$$\int_{\Omega_q \times \mathbb{R}} \left(\mathbb{E}[\hat{f}_{h,g}(\mathbf{x},z)] - f(\mathbf{x},z)\right)^2 dz\,\omega_q(d\mathbf{x}) + \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g}.$$

Bearing in mind the CLT result in Hall (1984) for the linear case, a bandwidth-free rate of convergence should be expected in (iii). Nevertheless, when $n\phi(h,g)h^q g \to \delta$, the analytical difficulty of joining the two rates of convergence of the dominant terms forces the normalizing rate to be $n(h^q g)^{1/2}$, although the sequence of bandwidths is restricted to satisfy the constraint $n\phi(h,g)h^q g \to \delta$. To clarify this point, a corollary presents a special case with proportional bandwidth sequences where the rate of convergence can be analytically stated in a bandwidth-free form.

Corollary 1. *Under* **A1**$-$**A3**, *and assuming* $g_n = \beta h_n$ *for a fixed* $\beta > 0$ *and* $0 < \delta < \infty$,

(i) $n^{1/2}h^{-2}\left(I_n - \mathbb{E}[I_n]\right) \xrightarrow{d} \mathcal{N}(0, \phi(1,\beta))$, *if* $nh^{q+5} \to \infty$,

(ii) $nh^{(q+1)/2}(I_n - \mathbb{E}[I_n]) \xrightarrow{d} \mathcal{N}(0, 2\sigma^2)$, if $nh^{q+5} \to 0$,

(iii) $n^{(q+9)/2(q+5)}(I_n - \mathbb{E}[I_n]) \xrightarrow{d} \mathcal{N}(0, \phi(1,\beta)\delta^{4/(q+5)} + 2\sigma^2\delta^{-(q+1)/(q+5)})$, if $nh^{q+5} \to \delta$.

3.2. Extensions of Theorem 1

The previous results can be adapted to other contexts involving directional variables, such as directional-directional or directional-multivariate random vectors. Once the common structure and the effects of each component are determined, it is easy to reproduce the computations duplicating a certain component or modifying it. This will be used to derive the directional-directional versions of the most relevant results along the paper. By considering a single bandwidth for the estimator defined in $\mathbb{R}^p$ (as in Hall (1984), for example), Theorem 1 can be easily adapted to account for a multivariate component.

Considering the directional-directional estimator $\hat{f}_{h_1,h_2}$, the corresponding analogues of conditions **A1**−**A3** are obtained (extending f from $\Omega_{q_1} \times \Omega_{q_2}$ to $\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{q_1+q_2+2} : \mathbf{x} \neq \mathbf{0}, \mathbf{y} \neq \mathbf{0}\}$ and assuming $nh_{1,n}^{q_1}h_{2,n}^{q_2} \to \infty$). Then, it is possible to derive a directional-directional version of Theorem 1.

Corollary 2 (CLT for the directional-directional ISE). *Denote the ISE of $\hat{f}_{h_1,h_2}$ by $I_n = \int_{\Omega_{q_1} \times \Omega_{q_2}} (\hat{f}_{h_1,h_2}(\mathbf{x},\mathbf{y}) - f(\mathbf{x},\mathbf{y}))^2 \omega_{q_2}(d\mathbf{y})\,\omega_{q_1}(d\mathbf{x})$. Then, under the directional-directional analogues of* **A1**−**A3**,

(i) $n^{1/2}\phi(h_1,h_2)^{-1/2}(I_n - \mathbb{E}[I_n]) \xrightarrow{d} Z$, if $n\phi(h_1,h_2)h_1^{q_1}h_2^{q_2} \to \infty$,

(ii) $n(h_1^{q_1}h_2^{q_2})^{1/2}(I_n - \mathbb{E}[I_n]) \xrightarrow{d} 2^{1/2}\sigma Z$, if $n\phi(h_1,h_2)h_1^{q_1}h_2^{q_2} \to 0$,

(iii) $n(h_1^{q_1}h_2^{q_2})^{1/2}(I_n - \mathbb{E}[I_n]) \xrightarrow{d} (\delta + 2\sigma^2)^{1/2} Z$, if $n\phi(h_1,h_2)h_1^{q_1}h_2^{q_2} \to \delta$,

where $0 < \delta < \infty$ and

$$\phi(h_1,h_2) = \frac{4b_{q_1}(L_1)^2}{q_1^2}\sigma_\mathbf{X}^2 h_1^4 + \frac{4b_{q_2}(L_2)^2}{q_2^2}\sigma_\mathbf{Y}^2 h_2^4 + \frac{8b_{q_1}(L_1)b_{q_2}(L_2)}{q_1 q_2}\sigma_{\mathbf{X},\mathbf{Y}} h_1^2 h_2^2.$$

$$\sigma^2 = R(f) \times \gamma_{q_1}\lambda_{q_1}(L_1)^{-4}\int_0^\infty r^{q_1/2-1}\left\{\int_0^\infty \rho^{q_1/2-1}L_1(\rho)\varphi_{q_1}(r,\rho)\,d\rho\right\}^2 dr$$
$$\times \gamma_{q_2}\lambda_{q_2}(L_2)^{-4}\int_0^\infty r^{q_2/2-1}\left\{\int_0^\infty \rho^{q_2/2-1}L_2(\rho)\varphi_{q_2}(r,\rho)\,d\rho\right\}^2 dr,$$

with $\sigma_{\mathbf{X},\mathbf{Y}} = \mathbb{C}\mathrm{ov}[\mathrm{tr}[\mathcal{H}_\mathbf{x}f(\mathbf{X},\mathbf{Y})], \mathrm{tr}[\mathcal{H}_\mathbf{y}f(\mathbf{X},\mathbf{Y})]]$, $\sigma_\mathbf{X}^2 = \mathbb{V}\mathrm{ar}[\mathrm{tr}[\mathcal{H}_\mathbf{x}f(\mathbf{X},\mathbf{Y})]]$ and $\sigma_\mathbf{Y}^2 = \mathbb{V}\mathrm{ar}[\mathrm{tr}[\mathcal{H}_\mathbf{y}f(\mathbf{X},\mathbf{Y})]]$. The same limit distributions hold in (i)−(iii) *if $\mathbb{E}[I_n]$ is replaced by*

$$\int_{\Omega_{q_1} \times \Omega_{q_2}} \left(\mathbb{E}\left[\hat{f}_{h_1,h_2}(\mathbf{x},\mathbf{y})\right] - f(\mathbf{x},\mathbf{y})\right)^2 \omega_{q_2}(\mathbf{y})\,\omega_{q_1}(\mathbf{x}) + \frac{\lambda_{q_1}(L_1^2)\lambda_{q_2}(L_2^2)}{\lambda_{q_1}(L_1)^2\lambda_{q_2}(L_2)^2 nh_1^{q_1}h_2^{q_2}}.$$

4. Testing Independence with Directional Random Variables

Given a random sample $(\mathbf{X}_1, Z_1), \ldots, (\mathbf{X}_n, Z_n)$ from a directional-linear variable $(\mathbf{X}, Z)$, one may be interested in the assessment of independence between components. If such a hypothesis is rejected, the joint kernel density estimator may give an idea of the dependence structure between them.

Let $f_{(\mathbf{X},Z)}$ denote the directional-linear density of $(\mathbf{X}, Z)$, with $f_{\mathbf{X}}$ and f_Z the directional and linear marginal densities. In this setting, the null hypothesis of independence is stated as $H_0 : f_{(\mathbf{X},Z)}(\mathbf{x}, z) = f_{\mathbf{X}}(\mathbf{x})f_Z(z)$, $\forall (\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$, and the alternative as $H_1 : f_{(\mathbf{X},Z)}(\mathbf{x}, z) \neq f_{\mathbf{X}}(\mathbf{x})f_Z(z)$, for some $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$. A statistic to test H_0 can be constructed considering the squared distance between the nonparametric estimator of joint density, denoted in this setting by $\hat{f}_{(\mathbf{X},Z);h,g}$, and the product of the corresponding marginal kernel estimators, denoted by $\hat{f}_{\mathbf{X},h}$ and $\hat{f}_{Z,g}$,

$$T_n = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{(\mathbf{X},Z);h,g}(\mathbf{x}, z) - \hat{f}_{\mathbf{X};h}(\mathbf{x}) \hat{f}_{Z;g}(z) \right)^2 dz\, \omega_q(d\mathbf{x}).$$

This type of test was introduced by Rosenblatt (1975) and Rosenblatt and Wahlen (1992) for bivariate random variables, considering the same bandwidths for smoothing both components. The directional-linear context requires an assumption on the degree of smoothness in each component.

A4 $h_n^q g_n^{-1} \to c$, with $0 < c < \infty$, as $n \to \infty$.

Theorem 2 (Directional-linear independence test). *Under* **A1**$-$**A4** *and the null hypothesis of independence,*

$$n(h^q g)^{1/2} (T_n - A_n) \xrightarrow{d} \mathcal{N}(0, 2\sigma_I^2),$$

where

$$A_n = \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(f_Z)}{nh^q} - \frac{R(K)R(f_{\mathbf{X}})}{ng},$$

and σ_I^2 is defined as σ^2 in Theorem 1, but with $R(f) = R(f_{\mathbf{X}})R(f_Z)$.

Since the leading term is the same as in Theorem 1 for $n\phi(h, g)h^q g \to 0$, the asymptotic variance is also the same. As in the CLT for the ISE, the effect of the components can be disentangled in the asymptotic variance and in the bias term. The a priori complex contribution of the directional part in Theorems 1 and 4 is explained for a particular scenario in the supplementary material, together with some numerical experiments for illustrating Theorem 2.

5. Goodness-of-fit Test with Directional Random Variables

Testing methods for a specific parametric directional-linear density (simple H_0) or for a parametric family (composite H_0) are presented in this section.

5.1. Testing a simple null hypothesis

Given a random sample $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ from an unknown directional-linear density f, the simple null hypothesis testing problem is stated as $H_0 : f = f_{\boldsymbol{\theta}_0}$, $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, where $f_{\boldsymbol{\theta}_0}$ is a certain parametric density with known parameter $\boldsymbol{\theta}_0$ belonging to the parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^p$, with $p \geq 1$. The alternative hypothesis is taken as $H_1 : f(\mathbf{x}, z) \neq f_{\boldsymbol{\theta}_0}(\mathbf{x}, z)$, for some $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$ in a set of positive measure. The proposed test statistic is

$$R_n = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g}f_{\boldsymbol{\theta}_0}(\mathbf{x}, z) \right)^2 \, dz \, \omega_q(d\mathbf{x}), \qquad (5.1)$$

where $LK_{h,g}f_{\boldsymbol{\theta}_0}(\mathbf{x}, z)$ represents the expected value of $\hat{f}_{h,g}(\mathbf{x}, z)$ under H_0. In general, for a function f, this expected value is

$$LK_{h,g}f(\mathbf{x}, z) = \frac{c_{h,q}(L)}{g} \int_{\Omega_q \times \mathbb{R}} LK \left(\frac{1 - \mathbf{x}^T\mathbf{y}}{h^2}, \frac{z - t}{g} \right) f(\mathbf{y}, t) \, dt \, \omega_q(d\mathbf{y}). \quad (5.2)$$

Smoothing the parametric density was considered by Fan (1994), in the linear setting, to avoid the bias effects in the integrand of the square error between the nonparametric estimator under the alternative and the parametric estimate under the null. A modification of the smoothing proposal was used by Boente, González-Manteiga, and Rodríguez (2014) for the directional case.

Theorem 3. *Under* **A1**$-$**A3** *and the simple null hypothesis* $H_0 : f = f_{\boldsymbol{\theta}_0}$, *with* $\boldsymbol{\theta}_0 \in \Theta$ *known*,

$$n(h^q g)^{1/2} \left(R_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} \right) \xrightarrow{d} \mathcal{N} \left(0, 2\sigma_{\boldsymbol{\theta}_0}^2 \right),$$

where $\sigma_{\boldsymbol{\theta}_0}^2$ *follows from replacing* $f = f_{\boldsymbol{\theta}_0}$ *in* σ^2 *from Theorem 1.*

5.2. Composite null hypothesis

Consider the testing problem $H_0 : f \in \mathcal{F}_{\boldsymbol{\Theta}} = \{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, where $\mathcal{F}_{\boldsymbol{\Theta}}$ is a class of parametric densities indexed by the p-dimensional parameter $\boldsymbol{\theta}$, vs. $H_1 : f \notin \mathcal{F}_{\boldsymbol{\Theta}}$. Under H_0, a parametric density estimator $f_{\hat{\boldsymbol{\theta}}}$ can be obtained by Maximum Likelihood (ML). The next conditions are required.

A5 The function $f_{\boldsymbol{\theta}}$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$, with derivatives that are bounded and uniformly continuous for $(\mathbf{x}, z)$.

A6 There exists $\boldsymbol{\theta}_1 \in \Theta$ such that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_1 = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$, and if $H_0 : f = f_{\boldsymbol{\theta}_0}$ holds for a $\boldsymbol{\theta}_0 \in \Theta$, then $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$.

A5 is a regularity assumption on the parametric density, whereas **A6** states that the estimation of the unknown parameter must be $\sqrt{n}$-consistent in order to ensure that the effects of parametric estimation can be neglected. The $\sqrt{n}$-consistency is required under H_0 (for Theorem 4) and H_1 (for Theorem 6), which is satisfied by the ML estimator. The test statistic is an adaptation of (5.1), but plugging-in the estimator of the unknown parameter $\boldsymbol{\theta}_0$ under H_0 in the test statistic expression:

$$R_n = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right)^2 dz \, \omega_q(d\mathbf{x}). \tag{5.3}$$

Theorem 4 (Goodness-of-fit test for directional-linear densities). *Under* **A1**−**A3**, **A5**−**A6** *and the composite null hypothesis* $H_0 : f = f_{\boldsymbol{\theta}_0}$, *with* $\boldsymbol{\theta}_0 \in \Theta$ *unknown*,

$$n(h^q g)^{1/2} \left(R_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2} R(K)}{n h^q g} \right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_{\boldsymbol{\theta}_0}^2\right).$$

Families of Pitman alternatives are a common way to measure power for tests based on kernel smoothers (e.g., Fan (1994)). For the directional-linear case, these alternatives can be written as

$$H_{1P} : f(\mathbf{x}, z) = f_{\boldsymbol{\theta}_0}(\mathbf{x}, z) + \left(n h^{q/2} g^{1/2}\right)^{-1/2} \Delta(\mathbf{x}, z), \tag{5.4}$$

where $\Delta(\mathbf{x}, z) : \Omega_q \times \mathbb{R} \to \mathbb{R}$ is such that $\int_{\Omega_q \times \mathbb{R}} \Delta(\mathbf{x}, z) \, dz \, \omega_q(d\mathbf{x}) = 0$. A necessary condition to derive the limit distribution of R_n under H_{1P} is that the estimator $\hat{\boldsymbol{\theta}}$ is a $\sqrt{n}$-consistent estimator for $\boldsymbol{\theta}_0$.

A7 For the family of alternatives (5.4), $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$.

Theorem 5 (Local power under Pitman alternatives). *Under* **A1**−**A3**, **A5**−**A7** *and the alternative hypothesis* (5.4),

$$n(h^q g)^{1/2} \left(R_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2} R(K)}{n h^q g} \right) \xrightarrow{d} \mathcal{N}\left(R(\Delta), 2\sigma_{\boldsymbol{\theta}_0}^2\right).$$

5.3. Calibration in practise

In order to effectively calibrate the proposed test, a parametric bootstrap procedure is investigated. The bootstrap statistic is defined as

$$R_n^* = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}^*(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol{\theta}}^*}(\mathbf{x}, z) \right)^2 dz \, \omega_q(d\mathbf{x}),$$

where the superscript $*$ indicates that the estimators are computed from the bootstrap sample $\{(\mathbf{X}_i^*, Z_i^*)\}_{i=1}^n$ obtained from the density $f_{\hat{\boldsymbol{\theta}}}$, with $\hat{\boldsymbol{\theta}}$ computed from the original sample. The bootstrap procedure, considering the composite null hypothesis testing problem, is detailed in an algorithm. Calibration for the simple null hypothesis test can be done replacing $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^*$ by $\boldsymbol{\theta}_0$.

Algorithm 1 (Testing procedure). *Let $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ be a random sample from f. To test $H_0 : f = f_{\boldsymbol{\theta}_0}$, with $\boldsymbol{\theta}_0 \in \Theta$ unknown, proceed as follows.*

1. *Obtain $\hat{\boldsymbol{\theta}}$, a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}_0$.*
2. *Compute $R_n = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right)^2 dz\, \omega_q(d\mathbf{x})$.*
3. *Bootstrap strategy. For $b = 1, \ldots, B$:*
   (a) *Obtain a random sample $\{(\mathbf{X}_i^*, Z_i^*)\}_{i=1}^n$ from $f_{\hat{\boldsymbol{\theta}}}$.*
   (b) *Compute $\hat{\boldsymbol{\theta}}^*$ as in Step 1, from the bootstrap sample in (a).*
   (c) *Compute $R_n^{*b} = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}^*(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol{\theta}}^*}(\mathbf{x}, z) \right)^2 dz\, \omega_q(d\mathbf{x})$, where $\hat{f}_{h,g}^*$ is obtained from the bootstrap sample in (a).*
4. *Approximate the p-value of the test as p-value $\approx \# \left\{ R_n \le R_n^{*b} \right\} / B$.*

The consistency of this testing procedure is proved here, using the bootstrap analogue of **A6**.

A8 $\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}} = \mathcal{O}_{\mathbb{P}^*}\left(n^{-1/2}\right)$, where $\mathbb{P}^*$ represents the probability of $(\mathbf{X}^*, Z^*)$ conditioned on the sample $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$.

Theorem 6 (Bootstrap consistency). *Under $\mathbf{A1} - \mathbf{A3}$, $\mathbf{A5} - \mathbf{A6}$ and $\mathbf{A8}$, and conditionally on $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$,*

$$n(h^q g)^{1/2} \left(R_n^* - \frac{\lambda_q(L^2)\lambda_q(L)^{-2} R(K)}{n h^q g} \right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_{\boldsymbol{\theta}_1}^2\right) \text{ in probability.}$$

Then, the probability distribution function (pdf) of R_n^* conditioned on the sample converges in probability to a Gaussian pdf, regardless of whether H_0 holds or not. The asymptotic distribution coincides with the one of R_n if H_0 holds ($\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$).

5.4. Extensions to directional-directional models

The directional-directional versions of the previous results follow under analogous assumptions (modifying **A5**, (5.2) and (5.4) accordingly). The directional-directional test statistic for the composite hypothesis testing problem is

$$R_n = \int_{\Omega_{q_1} \times \Omega_{q_2}} \left(\hat{f}_{h_1, h_2}(\mathbf{x}, \mathbf{y}) - L_1 L_{2, h_1, h_2} f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, \mathbf{y}) \right)^2 \omega_{q_2}(d\mathbf{y})\, \omega_{q_1}(d\mathbf{x}).$$

Corollary 3 (Goodness-of-fit test for directional-directional densities)**.** *Under the directional-directional analogues of* **A1**−**A3**, **A5**−**A6** *and the composite null hypothesis* $H_0 : f = f_{\boldsymbol{\theta}_0}$, *with* $\boldsymbol{\theta}_0 \in \Theta$ *unknown,*

$$n(h_1^{q_1} h_2^{q_2})^{1/2} \left(R_n - \frac{\lambda_{q_1}(L_1^2)\lambda_{q_1}(L_1)^{-2}\lambda_{q_2}(L_2^2)\lambda_{q_2}(L_2)^{-2}}{nh_1^{q_1} h_2^{q_2}} \right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_{\boldsymbol{\theta}_0}^2\right).$$

6. Simulation Study

The finite sample performance of the directional-linear and directional-directional goodness-of-fit tests is illustrated in this section for a variety of models, sample sizes, and bandwidth choices. The study considered circular-linear and circular-circular scenarios, although these tests can be easily applied in higher dimensions, such as spherical-linear or spherical-circular, due to their general definition and resampling procedures. Details on simulated models and further results are included as supplementary material.

Circular-Linear (CL) and Circular-Circular (CC) parametric scenarios were considered. Figures 1 and 2 show the density contours in the cylinder (CL) and in the torus (CC) for the different models. The detailed description of each model is given in the supplementary material. Deviations from the composite null hypothesis $H_0 : f \in \mathcal{F}_\Theta$ were obtained by mixing the true density $f_{\boldsymbol{\theta}_0}$ with a density Δ such that the resulting density does not belong to $\mathcal{F}_\Theta$: $H_\delta : f = (1 - \delta)f_{\boldsymbol{\theta}_0} + \delta\Delta$, $0 \le \delta \le 1$. The goodness-of-fit tests were applied using the bootstrap strategy, for the whole collection of models, sample sizes $n = 100, 500, 1,000$ and deviations $\delta = 0, 0.10, 0.15$ ($\delta = 0$ for the null hypothesis). The number of bootstrap and Monte Carlo replicates was 1,000.

In each case (model, sample size and deviation), the performance of the goodness-of-fit test is shown for a fixed pair of bandwidths, obtained from the median of 1,000 simulated Likelihood Cross Validation (LCV) bandwidths

$$\begin{aligned}
(h, g)_{\text{LCV}} &= \arg\max_{h,g>0} \sum_{i=1}^n \log \hat{f}_{h,g}^{-i}(\mathbf{X}_i, Z_i), \\
(h_1, h_2)_{\text{LCV}} &= \arg\max_{h_1,h_2>0} \sum_{i=1}^n \log \hat{f}_{h_1,h_2}^{-i}(\mathbf{X}_i, \mathbf{Y}_i),
\end{aligned} \tag{6.1}$$

where $\hat{f}_{\cdots}^{-i}$ denotes the kernel estimator computed without the ith datum. A deeper insight on the bandwidth effect is provided for some scenarios, where percentage of rejections are plotted for a grid of bandwidths (see Figure 3 for two cases, and supplementary material for extended results). The kernels considered were the von Mises and the normal ones.

Table 1 collects the results of the simulation study for each combination of model (CL or CC), deviation (δ) and sample size (n). When the null hypothesis holds, significance levels are correctly attained for $\alpha = 0.05$ (see supplementary
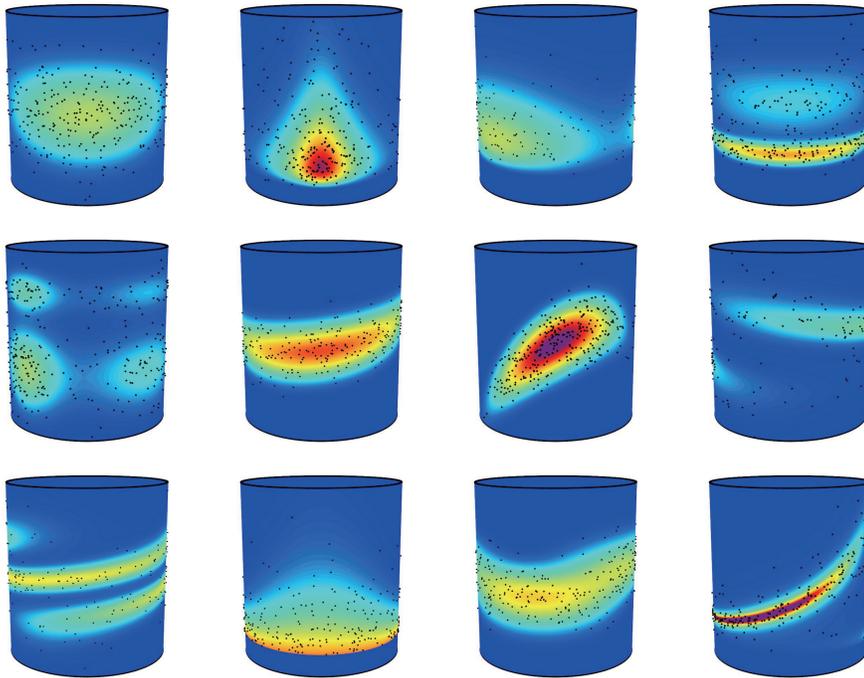
Figure 1. Density models for the simulation study in the circular-linear case. From left to right and up to down, models CL1 to CL12.
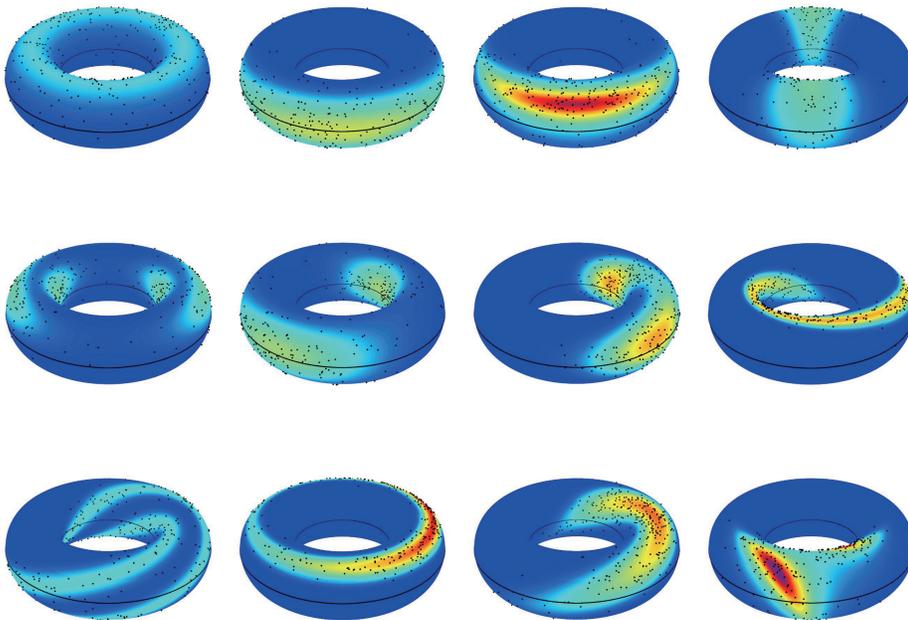

Figure 2. Density models for the simulation study in the circular-circular case. From left to right and up to down, models CC1 to CC12.

Table 1. Empirical size and power of the circular-linear and circular-circular goodness-of-fit tests for models CL1–CL12 and CC1–CC12 (respectively) with significance level $\alpha = 0.05$ and different sample sizes and deviations.

Model	Sample size n and deviation δ								
	$n = 100$			$n = 500$			$n = 1{,}000$		
	$\delta=0$	$\delta=0.10$	$\delta=0.15$	$\delta=0$	$\delta=0.10$	$\delta=0.15$	$\delta=0$	$\delta=0.10$	$\delta=0.15$
CL1	0.051	0.552	0.997	0.052	0.822	1.000	0.048	1.000	1.000
CL2	0.051	0.244	0.805	0.049	0.525	0.997	0.050	0.982	1.000
CL3	0.048	0.107	0.362	0.046	0.163	0.682	0.050	0.659	0.940
CL4	0.045	0.172	0.568	0.039	0.297	0.869	0.045	0.868	0.993
CL5	0.049	0.272	0.972	0.049	0.514	0.999	0.041	1.000	1.000
CL6	0.039	0.996	1.000	0.043	1.000	1.000	0.050	1.000	1.000
CL7	0.042	1.000	1.000	0.043	1.000	1.000	0.049	1.000	1.000
CL8	0.049	0.204	0.893	0.050	0.379	0.997	0.044	1.000	1.000
CL9	0.062	0.914	1.000	0.043	0.989	1.000	0.064	1.000	1.000
CL10	0.045	0.218	0.723	0.056	0.378	0.975	0.045	0.944	1.000
CL11	0.059	0.510	0.993	0.056	0.763	1.000	0.056	1.000	1.000
CL12	0.073	0.152	0.655	0.054	0.254	0.967	0.051	0.969	1.000
CC1	0.061	0.456	0.751	0.047	0.995	1.000	0.048	1.000	1.000
CC2	0.054	0.506	0.798	0.043	0.994	1.000	0.056	1.000	1.000
CC3	0.061	0.706	0.932	0.042	1.000	1.000	0.058	1.000	1.000
CC4	0.049	0.837	0.958	0.048	1.000	1.000	0.052	1.000	1.000
CC5	0.059	0.431	0.720	0.050	1.000	1.000	0.051	1.000	1.000
CC6	0.069	0.123	0.270	0.045	0.759	0.960	0.034	0.958	0.993
CC7	0.048	0.112	0.201	0.059	0.724	0.976	0.044	0.989	1.000
CC8	0.043	0.693	0.945	0.054	1.000	1.000	0.050	1.000	1.000
CC9	0.043	0.325	0.600	0.057	1.000	1.000	0.042	1.000	1.000
CC10	0.047	1.000	1.000	0.041	1.000	1.000	0.042	1.000	1.000
CC11	0.041	0.973	1.000	0.047	1.000	1.000	0.053	1.000	1.000
CC12	0.062	0.899	0.993	0.058	1.000	1.000	0.048	1.000	1.000

material for $\alpha = 0.10, 0.01$), for all sample sizes, models and deviations. When the null hypothesis does not hold, the tests perform satisfactorily, having in both cases a quick detection of the alternative when only a 10% and a 15% of the data come from a density out of the parametric family. As expected, the rejection rates grow as the sample size and the deviation from the alternative do.

Finally, the effect of the bandwidths is explored in Figure 3. For models CL1 and CC8, the empirical size and power of the tests are computed on a bivariate grid of bandwidths, for sample size $n = 100$ and deviations $\delta = 0$ (lower surface, null hypothesis) and $\delta = 0.15$ (upper surface). As it can be seen, the tests are correctly calibrated regardless of the choice of the bandwidths. However, the power is notably affected by the bandwidths, with different behaviours depending on the model and the alternative. Reasonable choices of the bandwidths, such
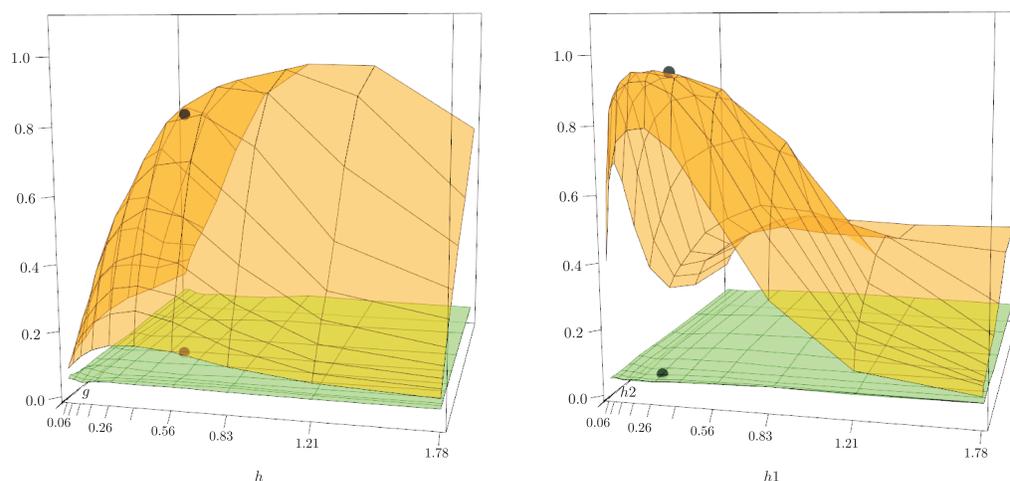
Figure 3. Empirical size and power of the circular-linear (left, model CL1) and circular-circular (right, model CC8) goodness-of-fit tests for a 10×10 logarithmic spaced grid. Lower surface represents the empirical rejection rate under $H_{0.00}$ and upper surface under $H_{0.15}$. Black points represent the empirical size and power obtained with the median of the LCV bandwidths.

as the median of the LCV bandwidths (6.1), present a competitive power. Further results supporting the same conclusions are available in the supplementary material.

7. Data Application

The proposed goodness-of-fit tests were applied to study two datasets (see supplementary material for further details). The first dataset comes from forestry and contains orientations and log-burnt areas of 26,870 wildfires occurred in Portugal between 1985 and 2005. Data was aggregated in watersheds, giving 102 observations of the circular mean orientation and mean log-burnt area for each watershed (circular-linear example). Further details on the data acquisition procedure, measurement of fires orientation and watershed delimitation can be seen in Barros, Pereira, and Lund (2012) and García-Portugués et al. (2014). The model proposed by Mardia and Sutton (1978) was tested for this dataset (Figure 4, left) using the LCV bandwidths and $B = 1,000$ bootstrap replicates, resulting a p-value of 0.156, showing no evidence against the null hypothesis.

The second dataset contains pairs of dihedral angles of segments of the type alanine-alanine-alanine in alanine amino acids in 1932 proteins. The dataset, formed by 233 pairs of angles (circular-circular), was studied by Fernández-Durán (2007) using Nonnegative Trigonometric Sums (NTSS) for the marginal and link function of the model of Wehrly and Johnson (1979). The best model
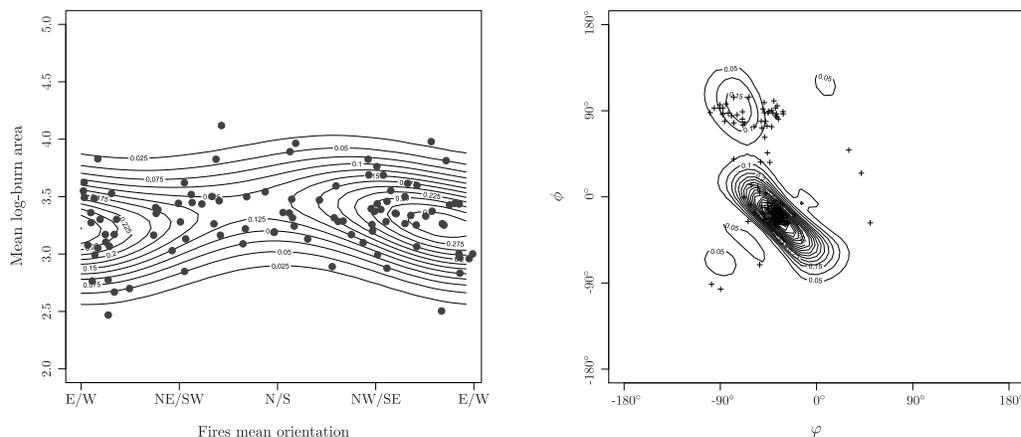
Figure 4. Left: parametric fit (model from Mardia and Sutton (1978)) to the circular mean orientation and mean log-burnt area of the fires in each of the 102 watersheds of Portugal. Right: parametric fit (model from Fernández-Durán (2007)) for the dihedral angles of the alanine-alanine-alanine segments.

in terms of BIC described in Fernández-Durán (2007) was implemented using a two-step Maximum Likelihood Estimation (MLE) procedure and the tools of the `CircNNTSR` package (Fernández-Durán and Gregorio-Domínguez (2013)) for fitting the NTSS parametric densities (Figure 4, right). The resulting p-value with the LCV bandwidths is 0.000, indicating that the dependence model of Wehrly and Johnson (1979) is not flexible enough to capture the dependence structure between the two angles. The reason for this lack of fit may be explained by a poor fit in a secondary cluster of data around $\Psi = 90°$, as can be seen in the contour plot in Figure 4.

Acknowledgement

Appendix A. Sketches of the Main Proofs

This section contains the sketches of the main proofs. Proofs for technical lemmas, complete numerical experiments and simulation results, and further details on data analysis are given in the supplementary material.

A.1. CLT for the integrated squared error

Proof of Theorem 1. The ISE can be decomposed into four addends, $I_n = I_{n_1} + I_{n_2} + I_{n_3} + I_{n_4}$:

$$I_{n_1} = 2\frac{c_{h,q}(L)}{ng} \sum_{i=1}^{n} \int_{\Omega_q \times \mathbb{R}} LK_n\left((\mathbf{x}, z), (\mathbf{X}_i, Z_i)\right) \left(\mathbb{E}[\hat{f}_{h,g}(\mathbf{x}, z)] - f(\mathbf{x}, z)\right) dz\, \omega_q(d\mathbf{x}),$$

$$I_{n_2} = \frac{c_{h,q}(L)^2}{n^2 g^2} \sum_{i=1}^{n} \int_{\Omega_q \times \mathbb{R}} LK_n^2\left((\mathbf{x}, z), (\mathbf{X}_i, Z_i)\right) dz\, \omega_q(d\mathbf{x}),$$

$$I_{n_3} = \frac{c_{h,q}(L)^2}{n^2 g^2} \sum_{1 \leq i < j \leq n} \int_{\Omega_q \times \mathbb{R}} LK_n\left((\mathbf{x}, z), (\mathbf{X}_i, Z_i)\right) LK_n\left((\mathbf{x}, z), (\mathbf{X}_j, Z_j)\right) dz\, \omega_q(d\mathbf{x}),$$

$$I_{n_4} = \int_{\Omega_q \times \mathbb{R}} \left(\mathbb{E}[\hat{f}_{h,g}(\mathbf{x}, z)] - f(\mathbf{x}, z)\right)^2 dz\, \omega_q(d\mathbf{x}),$$

where

$$LK_n\left((\mathbf{x}, z), (\mathbf{y}, t)\right) = LK\left(\frac{1 - \mathbf{x}^T \mathbf{y}}{h^2}, \frac{z - t}{g}\right) - \mathbb{E}\left[LK\left(\frac{1 - \mathbf{x}^T \mathbf{X}}{h^2}, \frac{z - Z}{g}\right)\right].$$

Except for the fourth term, which is deterministic, the CLT for the ISE is derived by examining the asymptotic behaviour of each addend. The first two can be written as $I_{n_1} = \sum_{i=1}^{n} I_{n_1}^{(i)}$ and $I_{n_2} = c_{h,q}(L)^2/(n^2 g^2) \sum_{i=1}^{n} I_{n_2}^{(i)}$, where $I_{n_1}^{(i)}$ and $I_{n_2}^{(i)}$ can be directly extracted from the previous expressions. Then, by Lemma 2,

$$n^{1/2} \phi(h, g)^{-1/2} I_{n_1} \xrightarrow{d} \mathcal{N}(0, 1) \tag{A.1}$$

and by Lemma 3,

$$I_{n_2} = \frac{\lambda_q(L^2)\lambda_q(L)^{-2} R(K)}{n h^q g} + \mathcal{O}_{\mathbb{P}}\left(n^{-3/2} h^{-q} g^{-1}\right). \tag{A.2}$$

The third term can be written as

$$I_{n_3} = 2\frac{c_{h,q}(L)^2}{n^2 g^2} \sum_{1 \leq i < j \leq n} H_n\left((\mathbf{X}_i, Z_i), (\mathbf{X}_j, Z_j)\right) = 2\frac{c_{h,q}(L)^2}{n^2 g^2} U_n, \tag{A.3}$$

where U_n is an U-statistic with kernel function H_n given in Lemma 4. U_n is degenerate since $\mathbb{E}[LK_n\left((\mathbf{x}, z), (\mathbf{X}, Z)\right)] = 0$.

In order to properly apply Lemma 1 for obtaining the asymptotic distribution of U_n in (A.3), Lemma 4 provides the explicit expressions for the required elements. Then, considering $\varphi_n \equiv 0$ in Lemma 1, condition $A_n B_n^{-2} \to 0$ is satisfied by **A3** and, as a consequence, $B_n^{-1/2} U_n \xrightarrow{d} \mathcal{N}(0,1)$. Since the variance of I_{n_3} is

$$\mathbb{V}\mathrm{ar}\,[I_{n_3}] = 4\frac{c_{h,q}(L)^4}{n^4 g^4}\mathbb{V}\mathrm{ar}\,[U_n] = 2\frac{\sigma^2}{n^2 h^q g}(1 + o(1)), \tag{A.4}$$

by Slutsky's theorem, (A.3) and (2.1),

$$n\left(h^q g\right)^{1/2} I_{n_3} \xrightarrow{d} \mathcal{N}\left(0, 2\sigma^2\right). \tag{A.5}$$

From (A.1), (A.2) and (A.5), it follows that:

$$I_n - \mathbb{E}[I_n] = n^{-1/2}\phi(h,g)^{1/2}N_{n_1} + \mathcal{O}_{\mathbb{P}}\left(n^{-3/2}h^{-q}g^{-1}\right)$$
$$+ 2^{1/2}\sigma n^{-1}(h^q g)^{-1/2}N_{n_3}, \tag{A.6}$$

where $N_{n_1}, N_{n_3} \xrightarrow{d} \mathcal{N}(0,1)$. By **A3**, $\left(n^{3/2}h^q g\right)^{-1} = o\left((nh^{q/2}g^{1/2})^{-1}\right)$ and the second addend I_{n_2} is asymptotically negligible compared with I_{n_3}. In order to determine dominance between I_{n_1} and I_{n_3}, the squared quotient between their orders is examined, being of order $n\phi(h,g)h^q g$. Then if $n\phi(h,g)h^q g \to \infty$ the last term on (A.6) is asymptotically negligible in comparison with the first, while if $n\phi(h,g)h^q g \to 0$, the first term is negligible in comparison with the last. By (A.2), (A.6) can be stated as

$$I_n - \left(\int_{\Omega_q \times \mathbb{R}}\left(\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right] - f(\mathbf{x},z)\right)^2 dz\,\omega_q(d\mathbf{x}) + \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g}\right)$$
$$= n^{-1/2}\phi(h,g)^{1/2}N_{n_1} + \mathcal{O}_{\mathbb{P}}\left(n^{-3/2}h^{-q}g^{-1}\right) + 2^{1/2}\sigma n^{-1}(h^q g)^{-1/2}N_{n_3}.$$

The case where $n\phi(h,g)h^q g \to \delta$, $0 < \delta < \infty$, needs a special treatment because none of the terms can be neglected. In this case,

$$I_n - \mathbb{E}[I_n] = n^{-1/2}\phi(h,g)^{1/2}N_{n_1} + 2^{1/2}\sigma n^{-1}(h^q g)^{-1/2}N_{n_3} + \mathcal{O}_{\mathbb{P}}\left(n^{-3/2}h^{-q}g^{-1}\right)$$
$$= n^{-1}(h^q g)^{-1/2}\left(\delta^{1/2}N_{n_1} + 2^{1/2}\sigma N_{n_3}\right) + \mathcal{O}_{\mathbb{P}}\left(n^{-3/2}h^{-q}g^{-1}\right).$$

In order to apply Lemma 1, set $\widetilde{U}_n = I_{n_1} + I_{n_3}$ with

$$\widetilde{U}_n = \sum_{i=1}^{n}\varphi_n(\mathbf{X}_i, Z_i) + \sum_{1 \le i < j \le n}\widetilde{H}_n\left((\mathbf{X}_i, Z_i), (\mathbf{X}_j, Z_j)\right),$$

where $\varphi_n(\mathbf{X}_1, Z_1) = I_{n_1}^{(1)}$, $\widetilde{H}_n((\mathbf{x}, z), (\mathbf{y}, t)) = 2c_{h,q}(L)^2/(n^2g^2)H_n((\mathbf{x}, z), (\mathbf{y}, t))$, and $\widetilde{G}_n((\mathbf{x}, z), (\mathbf{y}, t)) = \mathbb{E}[\widetilde{H}_n((\mathbf{X}, Z), (\mathbf{x}, z))\widetilde{H}_n((\mathbf{X}, Z), (\mathbf{y}, t))]$.

By Lemma 4 and the definitions of $\widetilde{H}_n$, $\widetilde{G}_n$, φ_n and M_n,

$$\mathbb{E}[\widetilde{H}_n^2((\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2))] = 4n^{-4}h^{-q}g^{-1}\sigma^2(1 + o(1)),$$
$$\mathbb{E}[\widetilde{H}_n^4((\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2))] = \mathcal{O}(n^{-8}h^{-3q}g^{-3}),$$
$$\mathbb{E}[\widetilde{G}_n^2((\mathbf{X}_1, Z_1), (\mathbf{X}_2, Z_2))] = \mathcal{O}(n^{-8}h^{-q}g^{-1}),$$
$$\mathbb{E}[\varphi_n^2(\mathbf{X}_1, Z_1)] = n^{-2}\phi(h, g)(1 + o(1)),$$
$$\mathbb{E}[\varphi_n^4(\mathbf{X}_1, Z_1)] = \mathcal{O}(n^{-4}(h^8 + g^8)),$$
$$\mathbb{E}[M_n^2(\mathbf{X}_1, Z_1)] = \mathcal{O}(n^{-6}(h^4 + g^4)h^{-3q/2}g^{-3/2}).$$

Applying these orders and using $n\phi(h, g)h^q g \to \delta$,

$$\frac{A_n}{B_n^2} = \mathcal{O}(n^{-1}) + \mathcal{O}((nh^q g)^{-1}h^{q/2}g^{\frac{1}{2}}) + \mathcal{O}((nh^q g)^{-1}) + \mathcal{O}(h^q g).$$

Then, by **A3**, the four previous orders tend to zero and therefore $B_n^{-1/2}\widetilde{U}_n \xrightarrow{d} \mathcal{N}(0, 1)$, where $B_n \sim n^{-1}\phi(h, g) + 2n^{-2}(h^q g)^{-1}\sigma^2 \sim n^{-2}(h^q g)^{-1}(\delta + 2\sigma^2)$. Finally, $n(h^q g)^{1/2}2^{1/2}(\delta + 2\sigma^2)^{-1/2}(I_{n_1} + I_{n_3}) \xrightarrow{d} \mathcal{N}(0, 1)$ by Slutsky's Theorem.

Proof of Corollary 1. As $g = \beta h$, for a fixed $\beta > 0$, $n\phi(h, g)h^q g = \mathcal{O}(nh^{q+5})$ and the cases in Theorem 1 are given by the asymptotic behaviour of this sequence. When $nh^{q+5} \to \infty$ and $nh^{q+5} \to 0$, the result is obtained immediately, whereas for $nh^{q+5} \to \delta$, $0 < \delta < \infty$, Lemma 1 gives

$$B_n \sim \phi(1, \beta)n^{-1}h^4 + 2\sigma^2 n^{-2}h^{-(q+1)} \sim n^{-\frac{q+9}{q+5}}(\phi(1, \beta)\delta^{\frac{4}{q+5}} + 2\sigma^2\delta^{-\frac{q+1}{q+5}}).$$

Therefore, $n^{\frac{q+9}{2(q+5)}}(\phi(1, \beta)\delta^{\frac{4}{q+5}} + 2\sigma^2\delta^{-\frac{q+1}{q+5}})^{-1/2}(I_{n_1} + I_{n_3}) \xrightarrow{d} \mathcal{N}(0, 1)$.

Proof of Corollary 2. The proof follows from an adaptation of the proof of Theorem 1 to the directional-directional context.

A.2. Testing independence with directional data

Proof of Theorem 2. The test statistic is decomposed as $T_n = T_{n_1} + T_{n_2} + T_{n_3}$ taking into account that, under independence, $\mathbb{E}[\hat{f}_{h,g}(\mathbf{x}, z)] = \mathbb{E}[\hat{f}_h(\mathbf{x})]\mathbb{E}[\hat{f}_g(z)]$:

$$T_{n_1} = \int_{\Omega_q \times \mathbb{R}} (\hat{f}_{h,g}(\mathbf{x}, z) - \mathbb{E}[\hat{f}_{h,g}(\mathbf{x}, z)])^2 dz\,\omega_q(d\mathbf{x}),$$
$$T_{n_2} = \int_{\Omega_q \times \mathbb{R}} (\hat{f}_h(\mathbf{x})\hat{f}_g(z) - \mathbb{E}[\hat{f}_h(\mathbf{x})]\mathbb{E}[\hat{f}_g(z)])^2 dz\,\omega_q(d\mathbf{x}),$$

$$T_{n_3} = -2 \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - \mathbb{E}[\hat{f}_{h,g}(\mathbf{x}, z)] \right)$$

$$\times \left(\hat{f}_h(\mathbf{x}) \hat{f}_g(z) - \mathbb{E}[\hat{f}_h(\mathbf{x})] \mathbb{E}[\hat{f}_g(z)] \right) \, dz \, \omega_q(d\mathbf{x}).$$

By Chebychev's inequality and Lemmas 6 and 7, the sum of the second and third addends is $-\mathbb{E}[T_{n_2}] + \mathcal{O}_{\mathbb{P}}\left(n^{-1}(h^{-q} + g^{-1})^{1/2}\right)$. Considering the test statistic decomposition and using Lemma 5 yields

$$T_n = T_{n_1} - \mathbb{E}[T_{n_2}] + \mathcal{O}_{\mathbb{P}}\left(n^{-1}(h^{-q} + g^{-1})^{1/2}\right)$$

$$= \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} + 2^{1/2}\sigma n^{-1}(h^q g)^{-1/2}N_n - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(f_Z)}{nh^q}$$

$$- \frac{R(K)R(f_\mathbf{X})}{ng} + o(n^{-1}\left(h^{-q} + g^{-1}\right)) + \mathcal{O}_{\mathbb{P}}\left(n^{-1}(h^{-q} + g^{-1})^{1/2}\right).$$

Now, $\mathcal{O}_{\mathbb{P}}\left(n^{-1}(h^{-q} + g^{-1})^{1/2}\right)$ is negligible in comparison with the second addend by **A3** and the deterministic order $o(n^{-1}(h^{-q} + g^{-1}))$ is also negligible by **A3** and **A4**. Therefore, $n(h^q g)^{1/2}(T_n - A_n) \xrightarrow{d} \mathcal{N}(0, 2\sigma_I^2)$.

A.3. Goodness-of-fit test for models with directional data

Proof of Theorem 3. Under $H_0 : f = f_{\boldsymbol{\theta}_0}$, the test statistic $R_n = I_{n_2} + I_{n_3}$, where I_{n_2} and I_{n_3} are given by (A.2) and (A.5) in the proof of Theorem 1, so

$$n(h^q g)^{1/2}\left(I_{n_2} + I_{n_3} - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g}\right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_{\boldsymbol{\theta}_0}^2\right). \qquad (A.7)$$

Proof of Theorem 4. The test statistic is decomposed as $R_n = R_{n_1} + I_{n_2} + I_{n_3} + R_{n_4}$ by adding and subtracting $\mathbb{E}[\hat{f}_{h,g}(\mathbf{x}, z)] = LK_{h,g}f(\mathbf{x}, z)$, with

$$R_{n_1} = 2 \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g}f(\mathbf{x}, z) \right) LK_{h,g} \left(f(\mathbf{x}, z) - f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right) \, dz \, \omega_q(d\mathbf{x}),$$

$$R_{n_4} = \int_{\Omega_q \times \mathbb{R}} \left(LK_{h,g} \left(f(\mathbf{x}, z) - f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right) \right)^2 \, dz \, \omega_q(d\mathbf{x}).$$

The limit of $I_{n_2} + I_{n_3}$ is given by (A.7) whereas, by Lemma 8, R_{n_1} and R_{n_4} are negligible in probability. Then, the limit distribution of R_n is determined by $I_{n_2} + I_{n_3}$.

Proof of Theorem 5. As in the proof of Theorem 4, $R_n = \widetilde{R}_{n_1} + I_{n_2} + I_{n_3} + \widetilde{R}_{n_4}$, where $I_{n_2} + I_{n_3}$ behaves as (A.7). The asymptotic variance remains $\sigma_{\boldsymbol{\theta}_0}^2$ since

$$R(f) = R(f_{\boldsymbol{\theta}_0}) + \frac{R(\Delta)}{nh^{q/2}g^{1/2}} + \frac{\int_{\Omega_q \times \mathbb{R}} f(\mathbf{x}, z)\Delta(\mathbf{x}, z) \, dz \, \omega_q(d\mathbf{x})}{n^{1/2}h^{q/4}g^{1/4}}$$

and then the second and third addends are negligible with respect to the first by **A3**, leaving the same asymptotic variance. The terms $\widetilde{R}_{n_1} = R_{n_1} + \widetilde{R}_{n_1}^{(1)}$ and $\widetilde{R}_{n_4} = R_{n_4} + \widetilde{R}_{n_4}^{(1)} + \widetilde{R}_{n_4}^{(2)}$ are decomposed as

$$\widetilde{R}_{n_1}^{(1)} = \frac{2}{\sqrt{nh^{q/2}g^{1/2}}} \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g}f(\mathbf{x}, z) \right) LK_{h,g}\Delta(\mathbf{x}, z) \, dz \, \omega_q(d\mathbf{x}),$$

$$\widetilde{R}_{n_4}^{(1)} = \frac{1}{nh^{q/2}g^{1/2}} \int_{\Omega_q \times \mathbb{R}} (LK_{h,g}\Delta(\mathbf{x}, z))^2 \, dz \, \omega_q(d\mathbf{x}),$$

$$\widetilde{R}_{n_4}^{(2)} = \frac{2}{\sqrt{nh^{q/2}g^{1/2}}} \int_{\Omega_q \times \mathbb{R}} LK_{h,g} \left(f(\mathbf{x}, z) - f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right) LK_{h,g}\Delta(\mathbf{x}, z) \, dz \, \omega_q(d\mathbf{x}).$$

The remaining terms follow from Lemma 9.

Proof of Theorem 6. Similar to the proof of Theorem 4, $R_n^* = R_{n_1}^* + I_{n_2}^* + I_{n_3}^* + R_{n_4}^*$, where the terms involved are the bootstrap versions of the ones defined in the aforementioned proof:

$$R_{n_1}^* = 2 \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}^*(\mathbf{x}, z) - LK_{h,g}f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right) LK_{h,g} \left(f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) - f_{\hat{\boldsymbol{\theta}}^*}(\mathbf{x}, z) \right) \, dz \, \omega_q(d\mathbf{x}),$$

$$I_{n_2}^* = \frac{c_{h,q}(L)^2}{n^2 g^2} \sum_{i=1}^n \int_{\Omega_q \times \mathbb{R}} (LK_n^*((\mathbf{x}, z), (\mathbf{X}_i^*, Z_i^*)))^2 \, dz \, \omega_q(d\mathbf{x}),$$

$$I_{n_3}^* = \frac{c_{h,q}(L)^2}{n^2 g^2} \sum_{1 \le i < j \le n} \int_{\Omega_q \times \mathbb{R}} LK_n^*((\mathbf{x}, z), (\mathbf{X}_i^*, Z_i^*))$$
$$\times LK_n^*((\mathbf{x}, z), (\mathbf{X}_j^*, Z_j^*)) \, dz \, \omega_q(d\mathbf{x}),$$

$$R_{n_4}^* = \int_{\Omega_q \times \mathbb{R}} \left(LK_{h,g} \left(f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) - f_{\hat{\boldsymbol{\theta}}^*}(\mathbf{x}, z) \right) \right)^2 \, dz \, \omega_q(d\mathbf{x}),$$

with

$$LK_n^*((\mathbf{x}, z), (\mathbf{y}, t)) = LK \left(\frac{1 - \mathbf{x}^T \mathbf{y}}{h^2}, \frac{z - t}{g} \right) - \mathbb{E}^* \left[LK \left(\frac{1 - \mathbf{x}^T \mathbf{X}^*}{h^2}, \frac{z - Z^*}{g} \right) \right]$$

and where $\mathbb{E}^*$ represents the expectation with respect to $f_{\hat{\boldsymbol{\theta}}}$, which is obtained from the original sample.

Using the same arguments as in Lemma 8, but replacing assumption **A6** by **A8**, it follows that $n(h^q g)^{1/2} R_{n_1}^*$ and $n(h^q g)^{1/2} R_{n_4}^*$ converge to zero conditionally on the sample, that is, in probability $\mathbb{P}^*$. On the other hand, the terms $I_{n_2}^*$ and $I_{n_3}^*$ follow from considering similar arguments to the ones used for deriving (A.2) and (A.5), but conditionally on the sample. Specifically, it follows that $I_{n_2}^* = \lambda_q(L^2)\lambda_q(L)^{-2}R(K)/(nh^q g) + \mathcal{O}_{\mathbb{P}^*}\left(n^{-3/2}h^{-q}g^{-1} \right)$ and, for a certain $\boldsymbol{\theta}_1 \in \Theta$, $(nh^q g)^{1/2}I_{n_3}^* \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_{\boldsymbol{\theta}_1}^2 \right)$. The main difference with the proof of Theorem 4 concerns the asymptotic variance given by $n(h^q g)^{1/2}I_{n_3}^*$: $\mathbb{V}\mathrm{ar}^*\left[n(h^q g)^{1/2}I_{n_3}^* \right] \xrightarrow{p}$

$2\sigma_{\boldsymbol{\theta}_1}^2$, since by **A5**, $R(f_{\hat{\boldsymbol{\theta}}}) = R(f_{\boldsymbol{\theta}_1}) + \mathcal{O}_{\mathbb{P}}(n^{-1/2})$. Hence,

$$n(h^q g)^{1/2}\left(R_n^* - \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g}\right)$$
$$= o_{\mathbb{P}^*}(1) + \mathcal{O}_{\mathbb{P}^*}\left((nh^q g)^{-1/2}\right) + 2^{1/2}\sigma_{\boldsymbol{\theta}_1}N_n + o_{\mathbb{P}^*}(1)$$

and bootstrap consistency follows.

Proof of Corollary 3. The proof follows by adapting the proof of Theorem 4.

References

Bai, Z. D., Rao, C. R. and Zhao, L. C. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.* **27**, 24-39.

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

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071-1095.

Boente, G., González-Manteiga, W. and Rodríguez, D. (2014). Goodness-of-fit test for directional data. *Scand. J. Statist.* **41**, 259-275.

Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Economet. Theor.* **10**, 316-356.

Fernández-Durán, J. J. (2007). Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. *Biometrics* **63**, 579-585.

Fernández-Durán, J. J. and Gregorio-Domínguez, M. M. (2013). *CircNNTSR, An R Package for the Statistical Analysis of Circular Data using Nonnegative Trigonometric Sums (NNTS) Models.* R package version 2.1.

Fisher, N. I. and Lee, A. J. (1981). Nonparametric measures of angular-linear association. *Biometrika* **68**, 629-636.

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

García-Portugués, E., Crujeiras, R. and González-Manteiga, W. (2013a). Exploring wind direction and SO$_2$ concentration by circular-linear density estimation. *Stoch. Environ. Res. Risk Assess.* **27**, 1055-1067.

García-Portugués, E., Crujeiras, R. and González-Manteiga, W. (2013b). Kernel density estimation for directional-linear data. *J. Multivariate Anal.* **121**, 152-175.

González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST* **22**, 361-411.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14**, 1-16.

Hall, P., Watson, G. S. and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74**, 751-762.

Johnson, R. A. and Wehrly, T. (1978). Some angular-linear distributions and related regression models. *J. Amer. Statist. Assoc.* **73**, 602-606.

Klemelä, J. (2000). Estimation of densities and derivatives of densities with directional data. *J. Multivariate Anal.* **73**, 18-40.

Mardia, K. (1976). Linear-circular correlation coefficients and rhythmometry. *Biometrika* **63**, 403-405.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Wiley.

Mardia, K. V. and Sutton, T. W. (1978). A model for cylindrical variables with applications. *J. Roy. Statist. Soc. B* **40**, 229-233.

Oliveira, M., Crujeiras, R. and Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Comput. Statist. Data Anal.* **56**, 3898-3908.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3**, 1-14.

Rosenblatt, M. and Wahlen, B. E. (1992). A nonparametric measure of independence under a hypothesis of independent components. *Statist. Probab. Lett.* **15**, 245-252.

Scott, D. W. (1992). *Multivariate Density Estimation*. Wiley.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Srivastava, A. and Sahami, M., editors (2009). *Text Mining, Classification, Clustering, and Applications*. Chapman and Hall/CRC.

Taylor, C. C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.* **52**, 3493-3500.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall.

Watson, G. S. (1983). *Statistics on Spheres*. Wiley.

Wehrly, T. E. and Johnson, R. A. (1979). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika* **67**, 255-256.

Zhao, L. and Wu, C. (2001). Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China Ser. A* **44**, 474-483.

Department of Mathematical Sciences, University of Copenhagen, Denmark.

The Bioinformatics Centre, University of Copenhagen, Denmark.

E-mail: egarcia@math.ku.dk

Department of Statistics and Operations Research, University of Santiago de Compostela, Spain.

E-mail: rosa.crujeiras@usc.es

Department of Statistics and Operations Research, University of Santiago de Compostela, Spain.

E-mail: wenceslao.gonzalez@usc.es