XI Congreso Galego de Estatística e Investigación de Operacións
A Coruña, 24–25–26 de outubro de 2013

GOODNESS–OF–FIT TESTS FOR DIRECTIONAL AND LINEAR DATA

Eduardo García–Portugués[1], Rosa M. Crujeiras[1] and Wenceslao González–Manteiga[1]

[1]Department of Statistics and Operations Research. University of Santiago de Compostela.

ABSTRACT

A central limit theorem for the integrated squared error of the directional–linear kernel density estimator is given in this work. The result is applied to derive a goodness–of–fit testing procedure for parametric families of directional–linear densities. The limit distribution of the proposed test statistic, together with a consistent bootstrap strategy for calibration, is developed for the directional–linear case and also for directional–directional data. Simulation results are provided to illustrate the finite sample performance of the testing methods for a collection of parametric models. Finally, the test is illustrated in two datasets from forestry and proteomics.

Keywords: Bootstrap; Directional–linear data; Goodness–of–fit; Kernel density estimator.

1. INTRODUCTION

Kernel density estimation is a classical topic in nonparametric statistics, with some comprehensive references in this area given by of Silverman (1986), Scott (1992) and Wand and Jones (1995), among others. Beyond the linear case, i.e. when the data lie in the real line, kernel density estimation has been also adapted to directional data, that is, data in the q–dimensional sphere. This kind of data arises, for example, in environmental science when measuring wind direction ($q = 1$), in astronomy in the study of stars position in the celestial sphere ($q = 2$) and in text mining for large q (see Mardia and Jupp (2000) for an exhaustive review). Some early works on kernel density estimation with directional data are the papers by Hall et al. (1987) and Bai et al. (1988), who defined two type of kernel estimators and gave their basic properties (bias, variance and uniformly strong consistency, among others). Later, Klemelä (2000) studied the estimation of the Laplacian of the density and generic derivatives and Zhao and Wu (2001) stated a Central Limit Theorem (CLT) for the Integrated Squared Error (ISE) of the directional kernel density estimator, based on the U–statistic martingale ideas introduced by Hall (1984). Some recent works focused on bandwidth selection for kernel density estimation with directional data include the papers by Taylor (2008), Oliveira et al. (2012) and García-Portugués (2013), where the first two ones are devoted to the circular case ($q = 1$) and the third one makes use of the exact risk ideas introduced by Marron and Wand (1992) to derive bandwidth selectors in the general q–dimensional setting. In addition, apart from the linear and directional cases, García-Portugués et al. (2013c) proposed a kernel density estimator for directional–linear data, deriving bias, variance and exact Mean Integrated Squared Error (MISE) computations.

Apart from estimation purposes, kernel density estimation has been used extensively for the development of goodness–of–fit tests, where the nonparametric estimators are used to check a parametric hypothesis. For example, Bickel and Rosenblatt (1973) and Fan (1994), provided goodness–of–fit tests for parametric families of densities for the linear case. Similarly, in the directional setting, Boente et al. (2013) presented a goodness–of–fit test with a bootstrap procedure based on the CLT introduced by Zhao and Wu (2001). On the other hand, kernel density estimation has also been used for constructing tests of independence between two random variates. The first works in this setting are the papers by Rosenblatt (1975) and Rosenblatt and Wahlen (1992), where the authors proposed a statistic based on the ISE of the joint kernel density estimator and the product of marginal kernel estimators. This idea was exploited recently in García-Portugués et al. (2013a) for deriving a bootstrap based independence test to study the relation in directional–linear variables, with an application to analyze the dependence between wildfire orientation and burnt area in Portugal.

The common key part of the goodness–of–fit and independence tests previously mentioned is the CLT for the ISE of the kernel estimator. The main aim of this work is to provide such a result for the directional–linear kernel estimator introduced by García-Portugués et al. (2013c) and use it to derive a goodness–of–fit test for parametric families of directional–linear densities. The CLT is obtained by proving an extended version of Theorem 1 in Hall (1984) that deals with modified U–statistics. The goodness-of–fit test follows by considering as a statistic the ISE between the joint kernel estimator and a smoothed parametric estimate of the unknown density, whose limit distribution is proved from a telescoping argument and the previous CLT. In addition to the limit distributions, bootstrap resampling strategies to calibrate the level of both tests are investigated. Closed expressions for the common case where the kernels are von Mises and Normal are also provided. In addition, the results obtained for the directional–linear case can be extended to the directional–directional context. The finite sample properties of the tests are illustrated with an extensive simulation study. Finally, the tests are applied to two datasets from forestry and proteomics.

This paper is organized as follows. Section 2 presents some background on kernel density estimation for linear, directional, directional–linear and directional–directional data. Section 3 presents the CLT for the ISE of the directional–linear estimator (Subsection 3.1) and the extensions to the directional–directional setting (Subsection 3.2). The goodness–of–fit test for the simple (Subsection 4.1) and composite null hypotheses (Subsection 4.2), its bootstrap calibration (Subsection 4.3) and extensions (Subsection 4.4) are given in Section 4. The empirical performance of the tests is illustrated with a simulation study in Section 5 and with applications to real datasets in Section 6. Some conclusions and final comments are given in Section 7 and, finally, the proofs of all the main results and technical lemmas are referred to the supplementary materials.

2. BACKGROUND

This section is devoted to a brief introduction on kernel density estimation for linear and directional data. For the sake of simplicity, f will denote the target density along the paper, which may be linear, directional, directional–linear or directional–directional, depending on the context.

Kernel density estimation for linear data was originally proposed by Parzen (1962) and Rosenblatt (1956) and has been widely studied in the statistical literature (see the aforementioned references for a review). Let Z denote a linear random variable with support $Supp(Z) \subseteq \mathbb{R}$ and density f and let $Z_1, \ldots, Z_n$ be a random sample of Z. The linear kernel density estimator, for $z \in \mathbb{R}$, is defined as

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

where K denotes the kernel, usually a symmetric density about the origin, and $g > 0$ is the bandwidth parameter, which controls the smoothness of the estimator. Specifically, large values of the bandwidth parameter will produce oversmoothed estimates of f, whereas small values will provide undersmoothed curves.

In the case of directional data, the directional kernel density estimation was introduced by Hall et al. (1987) and Bai et al. (1988). Let $\mathbf{X}$ denote a directional random variable with density f whose support is the q–dimensional sphere, denoted by $\Omega_q = \left\{\mathbf{x} \in \mathbb{R}^{q+1} : x_1^2 + \cdots + x_{q+1}^2 = 1\right\}$. The Lebesgue measure in Ω_q will be denoted by ω_q and, therefore, a directional density satisfies $\int_{\Omega_q} f(\mathbf{x})\, \omega_q(d\mathbf{x}) = 1$. Further, when there is no possible confusion, ω_q will also denote the surface area of Ω_q:

$$\omega_q = \omega_q\left(\Omega_q\right) = \frac{2\pi^{\frac{q+1}{2}}}{\Gamma\left(\frac{q+1}{2}\right)}, \quad q \geq 0,$$

where Γ represents the Gamma function defined as $\Gamma(p) = \int_0^\infty x^{p-1}e^{-x}\,dx$, for $p > -1$.

The directional kernel density estimator for a directional density f, based on a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ in the q–sphere, is defined as

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

where L is the directional kernel, $h > 0$ is the bandwidth parameter and $c_{h,q}(L)$ is a normalizing constant depending on the kernel L, the bandwidth h and the dimension q. The scalar product of two vectors, $\mathbf{x}$ and $\mathbf{y}$, is denoted by $\mathbf{x}^T\mathbf{y}$, where T is the transpose operator.

In this setting, directional kernels are not directional densities but functions of rapid decay. Therefore, to ensure that the resulting estimator is indeed a directional density, the normalizing constant $c_{h,q}(L)$ is needed. Specifically (Bai et al., 1988), the inverse of this normalizing constant for any $\mathbf{x} \in \Omega_q$ is given by

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

with $\lambda_{h,q}(L) = \omega_{q-1}\int_0^{2h^{-2}} L(r)r^{\frac{q}{2}-1}(2-rh^2)^{\frac{q}{2}-1}\,dr$, $\lambda_q(L) = 2^{\frac{q}{2}-1}\omega_{q-1}\int_0^\infty L(r)r^{\frac{q}{2}-1}\,dr$ and where $a_n \sim b_n$ denotes $\frac{a_n}{b_n} \to 1$ as $n \to \infty$ (see Lemma 1 in García-Portugués et al. (2013c) for a proof of this result). A common choice for the directional kernel is $L(r) = e^{-r}$, $r \geq 0$, also known as the von Mises kernel due to its relation with the von Mises–Fisher density (Watson, 1983). This density, denoted $vM(\boldsymbol{\mu}, \kappa)$, is given by

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

being $\boldsymbol{\mu} \in \Omega_q$ the directional mean, $\kappa > 0$ the concentration parameter around the mean and $\mathcal{I}_\nu$ the modified Bessel function of order ν (see equation 10.32.2 of Olver et al. (2010)),

$$\mathcal{I}_\nu(z) = \frac{\left(\frac{z}{2}\right)^\nu}{\pi^{1/2}\Gamma\left(\nu+\frac{1}{2}\right)}\int_{-1}^1 (1-t^2)^{\nu-\frac{1}{2}}e^{zt}\,dt.$$

The directional–linear kernel density 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 et al. (2013c). For $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$, it is defined as

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

where LK is a directional–linear kernel, g is the bandwidth parameter for the linear component, h the bandwidth parameter for the directional component and $c_{h,q}(L)$ the normalizing constant. For the sake of simplicity, the product kernel $LK(\cdot,\cdot) = L(\cdot) \times K(\cdot)$ is considered.

The next notation is necessary to simplify the asymptotic expressions of the directional–linear estimator:

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

$$\mathcal{H}f(\mathbf{x}, z) = \left(\begin{array}{ccc|c} \frac{\partial^2 f(\mathbf{x},z)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\mathbf{x},z)}{\partial x_1 \partial x_{q+1}} & \frac{\partial^2 f(\mathbf{x},z)}{\partial x_1 \partial z} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2 f(\mathbf{x},z)}{\partial x_{q+1} \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x},z)}{\partial x_{q+1}^2} & \frac{\partial^2 f(\mathbf{x},z)}{\partial x_{q+1} \partial z} \\ \hline \frac{\partial^2 f(\mathbf{x},z)}{\partial z \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x},z)}{\partial z \partial x_{q+1}} & \frac{\partial^2 f(\mathbf{x},z)}{\partial z^2} \end{array}\right) = \left(\begin{array}{c|c} \mathcal{H}_{\mathbf{x}} f(\mathbf{x}, z) & \mathcal{H}_{\mathbf{x},z} f(\mathbf{x}, z) \\ \hline \mathcal{H}_{\mathbf{x},z} f(\mathbf{x}, z)^T & \mathcal{H}_z f(\mathbf{x}, z) \end{array}\right),$$

$$\Psi_{\mathbf{x}}(f,\mathbf{x},z) = -\mathbf{x}^T\nabla_{\mathbf{x}} f(\mathbf{x},z) + q^{-1}\left(\nabla^2_{\mathbf{x}} f(\mathbf{x},z) - \mathbf{x}^T\mathcal{H}_{\mathbf{x}} f(\mathbf{x},z)\mathbf{x}\right), \quad \nabla^2_{\mathbf{x}} f(\mathbf{x},z) = \sum_{i=1}^{q+1}\frac{\partial^2 f(\mathbf{x},z)}{\partial \mathbf{x}_i^2},$$

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

Also, $R(\phi)$ will denote the integration of the square of a generic function ϕ along its domain.

The most common measure of overall performance of a kernel estimator is the MISE, defined as the expectation of the ISE. Under similar conditions to the ones presented in the next section, García-Portugués et al. (2013c) derived an expression for the MISE of the kernel estimator (2):

$$\begin{aligned}\mathrm{MISE}\left[\hat{f}_{h,g}(\cdot,\cdot)\right] =& b_q(L)^2 R\left(\Psi_{\mathbf{x}}(f,\cdot,\cdot)\right) h^4 + \frac{1}{4}\mu_2(K)^2 R\left(\mathcal{H}_z f(\cdot,\cdot)\right) g^4 \\ &+ b_q(L)\mu_2(K)\int_{\Omega_q\times\mathbb{R}} \Psi_{\mathbf{x}}(f,\mathbf{x},z)\mathcal{H}_z f(\mathbf{x},z)\,\omega_q(d\mathbf{x})\,dz h^2 g^2 \\ &+ \frac{c_{h,q}(L)}{ng} d_q(L) R(K) + o\left((nh^q g)^{-1} + h^4 + g^4\right),\end{aligned} \tag{3}$$

where the integrals are taken with respect to the product measure of $\omega_q \times m_{\mathbb{R}}$, with $m_{\mathbb{R}}$ denoting the usual Lebesgue measure in $\mathbb{R}$. When considering the parametrization $g = \beta h$, the previous expressions become simpler and there is a closed expression for the optimal pair of bandwidths $(h,g)_{\mathrm{AMISE}} = (h_{\mathrm{AMISE}}, \beta h_{\mathrm{AMISE}})$, where

$$h_{\mathrm{AMISE}} = \left[\frac{(q+1)d_q(L)R(K)}{4\beta\lambda_q(L)R\big(b_q(L)\Psi_{\mathbf{x}}(f,\cdot,\cdot) + \frac{\beta^2}{2}\mu_2(K)\mathcal{H}_z f(\cdot,\cdot)\big)n}\right]^{\frac{1}{5+q}}.$$

For the circular–linear data case ($q = 1$), the parameter β is given by:

$$\beta = \left(\frac{\frac{1}{4}\mu_2(K)^2 R\left(\mathcal{H}_z f(\cdot,\cdot)\right)}{b_q(L)^2 R\left(\Psi_{\mathbf{x}}(f,\cdot,\cdot)\right)}\right)^{\frac{1}{4}}.$$

Finally, 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:

$$\hat{f}_{h_1,h_2}(\mathbf{x},\mathbf{y}) = \frac{c_{h_1,q_1}(L_1)c_{h_2,q_2}(L_2)}{n}\sum_{i=1}^n L_1\left(\frac{1-\mathbf{x}^T\mathbf{X}_i}{h_1^2}\right)\times L_2\left(\frac{1-\mathbf{y}^T\mathbf{Y}_i}{h_2^2}\right), \tag{4}$$

and whose properties are similar to the previously defined directional–linear estimator. Specifically, under analogue directional–directional conditions to the ones presented in the next section, the MISE expansion is

$$\begin{aligned}\mathrm{MISE}\left[\hat{f}_{h_1,h_2}(\cdot,\cdot)\right] =& b_{q_1}(L_1)^2 R\left(\Psi_{\mathbf{x}}(f,\cdot,\cdot)\right) h_1^4 + b_{q_2}(L_2)^2 R\left(\Psi_{\mathbf{y}}(f,\cdot,\cdot)\right) h_2^4 \\ &+ b_{q_1}(L_1)b_{q_2}(L_2)\int_{\Omega_{q_1}\times\Omega_{q_2}} \Psi_{\mathbf{x}}(f,\mathbf{x},\mathbf{y})\Psi_{\mathbf{y}}(f,\mathbf{x},\mathbf{y})\,\omega_{q_1}(d\mathbf{x})\,\omega_{q_2}(d\mathbf{y}) h_1^2 h_2^2 \\ &+ \frac{c_{h_1,q_1}(L_1)c_{h_2,q_2}(L_2)}{n} d_{q_1}(L_1)d_{q_2}(L_2) + o\left((nh_1^{q_1}h_2^{q_2})^{-1} + h_1^4 + h_2^4\right).\end{aligned}$$

3. CENTRAL LIMIT THEOREM FOR THE INTEGRATED SQUARED ERROR

The main result of this paper, which is the basis for the application of the next section, is the CLT for the ISE of the kernel estimator (2). The result relies on an extended version of the key theorem in Hall (1984) of U–statistics and degenerate martingales.

3.1. Main result

The assumptions required to obtain the theoretical results will be introduced along the manuscript when needed. To begin with, the following conditions are needed to derive the asymptotic distribution for the ISE of the estimator (2):

KD1. Extend f from $\Omega_q \times \mathbb{R}$ to $\mathbb{R}^{q+2}\backslash A$, $A = \left\{(\mathbf{x}, z) \in \mathbb{R}^{q+2} : \mathbf{x} = \mathbf{0}\right\}$, by defining $f(\mathbf{x}, z) \equiv f\left(\frac{\mathbf{x}}{||\mathbf{x}||}, z\right)$ for all $\mathbf{x} \neq \mathbf{0}$ and $z \in \mathbb{R}$, where $||\cdot||$ denotes the Euclidean norm. Assume that f is bounded and differentiable up to third order, with its first three derivatives integrable and bounded.

KD2. Assume that $L : [0, \infty) \to [0, \infty)$ is a bounded and integrable function with exponential decay:

$$L(r) \leq Me^{-\alpha r},\ \forall r \in [0, \infty),\ \text{with } M,\ \alpha > 0.$$

Assume also that the kernel K is a bounded linear density function symmetric around zero with finite second order moment.

KD3. Assume that $h = h_n$ and $g = g_n$ are sequences of positive numbers such that $h_n \to 0$, $g_n \to 0$ and $nh_n^q g_n \to \infty$ as $n \to \infty$.

Remark 1. The continuity and boundedness of f and their first and second derivatives is a common assumption that appears, among others, in Hall (1984) and Rosenblatt and Wahlen (1992).

Remark 2. The assumption of compact support for the directional kernel L, stated in Zhao and Wu (2001), is not needed. Instead of, an exponential decay is required. This ensures that

$$0 < \int_0^\infty L^k(r) r^{\frac{q}{2}-1}\, dr < \infty,\ \forall q \geq 1,\ k = 1, 2,$$

conditions that are necessary to compute the bias and the variance, respectively, of the directional and directional–linear estimator (García-Portugués et al., 2013c). On the other hand, the exponential decay assumption is less restrictive than the compact support one and allows for the consideration of the von Mises kernel.

The next lemma is crucial for proving the main result of this section. The result is a generalization of the Theorem 1 of Hall (1984) for degenerate U–statistics that, up to the author's knowledge, was first stated by Zhao and Wu (2001), but without providing a formal proof (their version presents differences in the conditions from the one stated here). The interesting fact is that this lemma serves to prove both the cases when the sequence of bandwidths undersmooths the data (the variance is large relative to the bias, $n(h^4 + g^4)h^q g \to 0$) and when the bias is balanced with the variance ($n(h^4 + g^4)h^q g \to \delta$, $0 < \delta < \infty$). The lemma is presented in a generic notation of random variates, from which the adaptation to the directional–linear setting is straightforward.

Lemma 1. *Let $\{X_i\}_{i=1}^n$ be a sequence of independent and identically distributed random variables. Suppose that the function $H_n(x, y)$ is symmetric in x and y, with*

$$\mathbb{E}\left[H_n\left(X_1, X_2\right)|X_1\right] = 0 \text{ almost surely.} \tag{5}$$

and with $\mathbb{E}\left[H_n^4(X_1, X_2)\right] < \infty$, $\forall n$. Define the functions

$$G_n\left(x, y\right) = \mathbb{E}\left[H_n\left(x, X_1\right) H_n\left(y, X_1\right)\right]$$

and φ_n, satisfying $\mathbb{E}\left[\varphi_n(X_1)\right] = 0$ and $\mathbb{E}\left[\varphi_n^4(X_1)\right] = 0 < \infty$. Define also:

$$\begin{aligned} M_n(X_1) &= \mathbb{E}\left[\varphi_n(X_2) H_n(X_1, X_2)|X_1\right], \\ A_n &= n\mathbb{E}\left[\varphi_n^4(X_1)\right] + n^2\mathbb{E}\left[M_n^2(X_1)\right] + n^3\mathbb{E}\left[H_n^4(X_1, X_2)\right] + n^4\mathbb{E}\left[G_n^2(X_1, X_2)\right], \\ B_n &= n\mathbb{E}\left[\varphi_n^2(X_1)\right] + \frac{1}{2}n^2\mathbb{E}\left[H_n^2(X_1, X_2)\right]. \end{aligned}$$

Under these conditions, if $A_n B_n^{-2} \to 0$ as $n \to \infty$, then

$$U_n = \sum_{i=1}^{n} \varphi_n(X_i) + \sum_{1 \le i < j \le n} H_n(X_i, X_j) \xrightarrow{d} \mathcal{N}(0, B_n),$$

where $\xrightarrow{d}$ denotes the convergence in distribution.

Remark 3. Note that when $\varphi_n \equiv 0$, Theorem 1 in Hall (1984) is a particular case of Lemma 1 and U_n is a U–statistic, something which it is not true in general.

The next theorem states the limit distribution of the ISE for the estimator (2). The theorem mixes up the previous results from the works by Hall (1984) (linear component) and Zhao and Wu (2001) (directional component). Its proof uses Lemma 1 and some other techniques developed in both of these papers.

Theorem 1 (CLT for the directional–linear ISE)**.** *Denote the ISE of (2) by*

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

*Then, under conditions **KD1**–**KD3**, it holds that*

$$d_n (I_n - C_n) \xrightarrow{d} \begin{cases} Z, & n\phi(h,g)h^q g \to \infty, \\ 2^{\frac{1}{2}} \sigma Z, & n\phi(h,g)h^q g \to 0, \\ \left(\delta + 2\sigma^2\right)^{\frac{1}{2}} Z, & n\phi(h,g)h^q g \to \delta,\ 0 < \delta < \infty, \end{cases}$$

where Z is $\mathcal{N}(0,1)$ distributed, with

$$C_n = \int_{\Omega_q \times \mathbb{R}} \left(\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x}, z) \right] - f(\mathbf{x}, z) \right)^2 \omega_q(d\mathbf{x})\, dz + \frac{\lambda_q(L^2)\lambda_q(L)^{-2} R(K)}{n h^q g}. \tag{6}$$

The normalizing rates are

$$d_n = \begin{cases} n^{\frac{1}{2}} \phi(h,g)^{-\frac{1}{2}}, & n\phi(h,g)h^q g \to \infty, \\ n(h^q g)^{\frac{1}{2}}, & n\phi(h,g)h^q g \to 0, \\ n(h^q g)^{\frac{1}{2}}, & n\phi(h,g)h^q g \to \delta,\ 0 < \delta < \infty, \end{cases}$$

where

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

with $\sigma_{\mathbf{X}}^2 = \mathbb{V}\mathrm{ar}\left[\Psi_{\mathbf{x}}(f, \cdot, \cdot)\right]$, $\sigma_Z^2 = \mathbb{V}\mathrm{ar}\left[\mathcal{H}_z f(\cdot, \cdot)\right]$ and $\sigma_{\mathbf{X},Z} = \mathbb{C}\mathrm{ov}\left[\Psi_{\mathbf{x}}(f, \cdot, \cdot), \mathcal{H}_z f(\cdot, \cdot)\right]$. The rest of constants related to this result are:

$$\begin{aligned} \sigma^2 = {} & \int_{\Omega_q \times \mathbb{R}} f(\mathbf{x}, z)^2 \, \omega_q(d\mathbf{x})\, dz \\ & \times \gamma_q \lambda_q(L)^{-4} \int_0^\infty r^{\frac{q}{2}-1} \left\{ \int_0^\infty \rho^{\frac{q}{2}-1} L(\rho) \varphi_q(r, \rho)\, d\rho \right\}^2 dr \\ & \times \int_{\mathbb{R}} \left\{ \int_{\mathbb{R}} K(u) K(u+v)\, du \right\}^2 dv, \end{aligned}$$

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

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

Remark 4. The asymptotic bias (6) can be written also using the MISE expression (3), yielding

$$C_n = R\left(b_q(L)\Psi_{\mathbf{x}}(f,\mathbf{x},z)h^2 + \frac{1}{2}\mu_2(K)\mathcal{H}_z f(\mathbf{x},z)g^2\right) + \frac{\lambda_q(L^2)\lambda_q(L)^{-2}R(K)}{nh^q g} + o\left(h^4+g^4\right). \quad (7)$$

Theorem 1 includes as marginal cases the results by Hall (1984) for the linear part and Zhao and Wu (2001) for the directional one. The effect of both components can be seen in the form of the asymptotic bias and the asymptotic variance. This is clearly shown in Corollary 2, which contains the expressions of the asymptotic bias and variance for the von Mises and Normal kernels.

An unpleasant feature of Theorem 1 is the rate of convergence when $n\phi(h,g)h^q g \to \delta$. Instead of having a bandwidth–free rate of convergence, the rate appears to be $n(h^q g)^{\frac{1}{2}}$, something different from the CLT of Hall (1984). This issue is related to the technical difficulty of joining the rates of h and g together and does not have to be interpreted as that the *real* rate of convergence in that case is $n(h^q g)^{\frac{1}{2}}$, since both sequences of bandwidths have to satisfy that $n\phi(h,g)h^q g \to \delta$. To clarify this issue, the next corollary presents a special situation when the two sequence of bandwidths are proportional and the rate of convergence is the one given by Hall (1984) for dimension $q' = q+1$.

Corollary 1. *Under conditions **KD1**–**KD3**, and assuming $g_n = \beta h_n$ for a fixed $\beta > 0$,*

$$d_n\left(I_n - A_n\right) \xrightarrow{d} \begin{cases} \phi(1,\beta)^{\frac{1}{2}} Z, & nh^{q+5} \to \infty, \\ 2^{\frac{1}{2}}\sigma Z, & nh^{q+5} \to 0, \\ \left(\phi(1,\beta)\delta^{\frac{4}{q+5}} + 2\sigma^2\delta^{-\frac{q+1}{q+5}}\right)^{\frac{1}{2}} Z, & nh^{q+5} \to \delta,\ 0<\delta<\infty, \end{cases}$$

where Z is $\mathcal{N}(0,1)$ distributed and

$$d_n = \begin{cases} n^{\frac{1}{2}}h^{-2}, & nh^{q+5} \to \infty, \\ nh^{\frac{q+1}{2}}, & nh^{q+5} \to 0, \\ n^{\frac{q+9}{2(q+5)}}, & nh^{q+5} \to \delta,\ 0<\delta<\infty. \end{cases}$$

The remaining constants are given in Theorem 1.

The a priori complex effect of the directional part of Theorems 1 and 3, specially in the asymptotic variance, contrasts with the well known contribution of the linear one. However, despite this first impression, the contributions of both parts when K is a normal density and L is the von Mises kernel (the most usual choices) are quite similar. In that situation the kernel functionals are easy to compute:

$$R(K) = \left(2\pi^{\frac{1}{2}}\right)^{-1}, \quad \lambda_q(L^2)\lambda_q(L)^{-2} = \left(2\pi^{\frac{1}{2}}\right)^{-q}, \quad \lambda_q(L) = (2\pi)^{\frac{q}{2}}.$$

Further, and more interesting, it is possible to compute exactly the form of the contributions of these two kernels to the asymptotic variance, resulting in

$$\int_{\mathbb{R}}\left\{\int_{\mathbb{R}} K(u)K(u+v)\,du\right\}^2 dv = (8\pi)^{-\frac{1}{2}},$$

$$\gamma_q\lambda_q(L)^{-4}\int_0^\infty r^{\frac{q}{2}-1}\left\{\int_0^\infty \rho^{\frac{q}{2}-1}L(\rho)\varphi_q(r,\rho)\,d\rho\right\}^2 dr = (8\pi)^{-\frac{q}{2}}.$$

It is interesting to note the symmetric role of the two variables for the case $q=1$ and the effect of the dimension q. These findings are summarized in the next result.

Corollary 2. *If $L(r) = e^{-r}$ and K is a normal density, then the asymptotic bias and variance in Theorem 1 are*

$$C_n = \int_{\Omega_q\times\mathbb{R}}\left(\mathbb{E}\left[\hat{f}_{h,g}(\mathbf{x},z)\right] - f(\mathbf{x},z)\right)^2 \omega_q(d\mathbf{x})\,dz + \frac{1}{2^{q+1}\pi^{\frac{q+1}{2}}nh^q g}, \quad \sigma_I^2 = (8\pi)^{-\frac{q+1}{2}}R(f).$$

Further, if $f = f_{\mathbf{X}} \times f_Z$, with $f_{\mathbf{X}} = f_{vM}(\cdot;\boldsymbol{\mu},\kappa)$ and f_Z the density of a $\mathcal{N}(m,\sigma^2)$, then $R(f) = R(f_{\mathbf{X}})R(f_Z)$, with $R(f_{\mathbf{X}}) = \left(2\pi^{\frac{q+1}{2}}\right)^{-1}\kappa^{\frac{q-1}{2}}\mathcal{I}_{\frac{q-1}{2}}(2\kappa)\mathcal{I}_{\frac{q-1}{2}}(\kappa)^{-2}$ and $R(f_Z) = \left(2\pi^{\frac{1}{2}}\sigma\right)^{-1}$.

3.2 Extensions of Theorem 1

An interesting advantage of developing theory for the directional–linear setting is that it can be easily extended to other contexts such as directional–directional or directional–multivariate. The reason is that once the common structure and the effects of each components are determined, it is easy to replicate the computations duplicating one part (directional–directional) or modifying it (directional–multivariate). This will be exploited to present the directional–directional versions of the most relevant results along the paper, but also a directional–multivariate extension is feasible by considering a single bandwidth for the multivariate estimator defined in $\mathbb{R}^p$ (as in Hall (1984), for example).

Starting from estimator (4), the directional–directional analogues of conditions **KD1**–**KD3** are obtained by duplicating the directional conditions (the main changes are 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 generalize the previous theorem to the directional–directional version using similar computations.

Corollary 3 (CLT for the directional–directional ISE)**.** *Denote the ISE of (4) by*

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

*Then, under the directional–directional analogues of conditions **KD1**–**KD3**, it holds that*

$$d_n \left(I_n - C_n \right) \xrightarrow{d} \begin{cases} Z, & n\phi(h_1, h_2) h_1^{q_1} h_2^{q_2} \to \infty, \\ 2^{\frac{1}{2}} \sigma Z, & n\phi(h_1, h_2) h_1^{q_1} h_2^{q_2} \to 0, \\ \left(\delta + 2\sigma^2\right)^{\frac{1}{2}} Z, & n\phi(h_1, h_2) h_1^{q_1} h_2^{q_2} \to \delta,\ 0 < \delta < \infty, \end{cases}$$

where Z is $\mathcal{N}(0, 1)$ distributed, with

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

The normalizing rates are

$$d_n = \begin{cases} n^{\frac{1}{2}} \phi(h_1, h_2)^{-\frac{1}{2}}, & n\phi(h_1, h_2) h_1^{q_1} h_2^{q_2} \to \infty, \\ n(h_1^{q_1} h_2^{q_2})^{\frac{1}{2}}, & n\phi(h_1, h_2) h_1^{q_1} h_2^{q_2} \to 0, \\ n(h_1^{q_1} h_2^{q_2})^{\frac{1}{2}}, & n\phi(h_1, h_2) h_1^{q_1} h_2^{q_2} \to \delta,\ 0 < \delta < \infty, \end{cases}$$

where

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

with $\sigma_{\mathbf{X}}^2 = \mathbb{V}\mathrm{ar}\left[\Psi_{\mathbf{x}}(f, \cdot, \cdot)\right]$, $\sigma_{\mathbf{Y}}^2 = \mathbb{V}\mathrm{ar}\left[\Psi_{\mathbf{y}}(f, \cdot, \cdot)\right]$, $\sigma_{\mathbf{X},\mathbf{Y}} = \mathbb{C}\mathrm{ov}\left[\Psi_{\mathbf{x}}(f, \cdot, \cdot), \Psi_{\mathbf{y}}(f, \cdot, \cdot)\right]$ and

$$\begin{aligned} \sigma^2 = & \int_{\Omega_{q_1} \times \Omega_{q_2}} f(\mathbf{x}, \mathbf{y})^2\, \omega_{q_1}(d\mathbf{x})\, \omega_{q_2}(d\mathbf{y}) \\ & \times \gamma_{q_1} \lambda_{q_1}(L_1)^{-4} \int_0^\infty r^{\frac{q_1}{2} - 1} \left\{ \int_0^\infty \rho^{\frac{q_1}{2} - 1} L_1(\rho) \varphi_{q_1}(r, \rho)\, d\rho \right\}^2 dr \\ & \times \gamma_{q_2} \lambda_{q_2}(L_2)^{-4} \int_0^\infty r^{\frac{q_2}{2} - 1} \left\{ \int_0^\infty \rho^{\frac{q_2}{2} - 1} L_2(\rho) \varphi_{q_2}(r, \rho)\, d\rho \right\}^2 dr. \end{aligned}$$

4. GOODNESS–OF–FIT TESTS FOR MODELS WITH A DIRECTIONAL COMPONENT

When a random sample is given from an unknown directional–linear density, a first natural question one may ask is whether the data are componentwise independent. This hypothesis can be checked with the independence test of García-Portugués et al. (2013a). If the independence hypothesis is rejected, one may be interested in checking if a parametric model succeeds in describing the observed data in a satisfactory way. Two questions arise in this scenario: does a particular density explain well enough the data? and, more interesting, does a parametric family of densities explain well enough the data? These questions represent the *simple null hypothesis* and *composite null hypothesis*, respectively, and the present section is devoted to provide an answer to them using a goodness–of–fit test for parametric directional–linear densities.

4.1 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 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 a known parameter $\boldsymbol{\theta}_0$ that belongs to the parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^p$, with $p \geq 1$. Unlike parametric tests, that assume that the alternative to the null hypothesis belongs to a certain parametric family, the alternative here is a general hypothesis of the form $H_1 : f(\mathbf{x}, z) \neq f_{\boldsymbol{\theta}_0}(\mathbf{x}, z),$ for some $(\mathbf{x}, z) \in \Omega_q \times \mathbb{R}$ in a set of positive measure.

The statistic to check H_0 is then

$$R_n = \int_{\Omega_q\times\mathbb{R}} \left(\hat f_{h,g}(\mathbf{x}, z) - LK_{h,g} f_{\boldsymbol{\theta}_0}(\mathbf{x}, z)\right)^2 \omega_q(d\mathbf{x})\, dz. \tag{8}$$

The following quantity is of key importance in the test statistic. It represents the expectation of the kernel estimator $\hat f_{h,g}(\mathbf{x}, z)$ at a fixed point $(\mathbf{x}, z)$ with respect to the density f:

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

This smoothing in the parametric density was considered by Fan (1994), in the linear setting, to avoid the effects of bias in the integrand of the square error between the non parametric estimator under the alternative and the parametric estimate under the null. Later, a modification of the smoothing of Fan (1994) was employed in Boente et al. (2013) for the directional case.

Theorem 2. *Under conditions **KD1**–**KD3** and under the null hypothesis $H_0 : f = f_{\boldsymbol{\theta}_0}$, for a known $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$,*

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

where $\sigma^2_{\boldsymbol{\theta}_0}$ follows from replacing $f = f_{\boldsymbol{\theta}_0}$ in the definition of σ^2 in Theorem 1.

4.2 Composite null hypothesis

The simple null hypothesis has a limited application in practice, as one usually does not know a good candidate $f_{\boldsymbol{\theta}_0}$ for the density f. Instead of that, one is usually more interested in testing a more general hypothesis of the form

$$H_0 : f \in \mathcal{F}_{\boldsymbol{\Theta}} = \{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\},$$

where $\mathcal{F}_{\boldsymbol{\Theta}}$ is a class of parametric densities indexed by the p–dimensional parameter $\boldsymbol{\theta}$. This hypothesis is equivalent to $H_0 : f = f_{\boldsymbol{\theta}_0},\ \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, where $\boldsymbol{\theta}_0$ is *unknown*. If such a hypothesis holds, then the unknown density f has a parametric form and parametric inference can be applied, for instance Maximum Likelihood Estimation (MLE). In particular, if H_0 holds then $f_{\hat{\boldsymbol{\theta}}}$, where $\hat{\boldsymbol{\theta}}$ is an estimator of the true but unknown $\boldsymbol{\theta}_0$, will be a good estimator of $f_{\boldsymbol{\theta}_0}$.

To derive a goodness–of–fit test for the composite hypothesis, the following two conditions will be needed:

GF1. The function $f_{\boldsymbol{\theta}}$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$ and its partial derivatives are bounded and integrable with respect to $(\boldsymbol{\theta}, \mathbf{x}, z)$.

GF2. There exists a $\boldsymbol{\theta}_1$ such that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_1 = \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$ and if $f = f_{\boldsymbol{\theta}_0}$ holds for a certain $\boldsymbol{\theta}_0$, then $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$.

The first one is a regularity assumption on the parametric density, whereas the second states that the estimation of the unknown parameter must be $\sqrt{n}$–consistent in order to ensure that the effects of parametric estimation are not dominant. Note that the $\sqrt{n}$–consistency is required both when H_0 holds and not, something which is achieved for example, by MLE.

The test statistic is an adaptation of (8), but performing the estimation of the unknown parameter $\boldsymbol{\theta}_0$:

$$R_n = \int_{\Omega_q \times \mathbb{R}} \left(\hat{f}_{h,g}(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol{\theta}}}(\mathbf{x}, z) \right)^2 \omega_q(d\mathbf{x})\, dz. \tag{9}$$

The asymptotic normality of this statistic is given in the following theorem.

Theorem 3 (Goodness–of–fit test for directional–linear densities)**.** *Under conditions **KD1**–**KD3**, **GF1**–**GF2** and under the null hypothesis* $H_0 : f \in \mathcal{F}_{\boldsymbol{\Theta}}$*,*

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

Families of local alternatives, also known as Pitman alternatives, are a common way to measure the power in tests based on kernel smoothing (this was done, for example, in Fan (1994)). For the directional–linear case, these alternatives are of the form

$$H_{1P} : f(\mathbf{x}, z) = f_{\boldsymbol{\theta}_0}(\mathbf{x}, z) + \frac{1}{\sqrt{nh^{\frac{q}{2}} g^{\frac{1}{2}}}} \Delta(\mathbf{x}, z), \tag{10}$$

where $\Delta(\mathbf{x}, z) : \Omega_q \times \mathbb{R} \to \mathbb{R}$ is such that $\int_{\Omega_q \times \mathbb{R}} \Delta(\mathbf{x}, z)\, \omega_q(d\mathbf{x})\, dz = 0$. 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 the unknown parameter $\boldsymbol{\theta}_0$:

GF3. For the family of Pitman alternatives (10), it holds that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$.

Theorem 4 (Local power under Pitman alternatives)**.** *Under conditions **KD1**–**KD3**, **GF1**–**GF3** and under the alternative hypothesis* H_{1P}*,*

$$n(h^q g)^{\frac{1}{2}} \left(R_n - \frac{R(K)\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q g} \right) \xrightarrow{d} \mathcal{N}\left(\int_{\Omega_q \times \mathbb{R}} \Delta(\mathbf{x}, z)^2\, \omega_q(d\mathbf{x})\, dz, 2\sigma^2_{\boldsymbol{\theta}_0} \right).$$

4.3 Calibration in practice

As it usually happens with tests based on kernel smoothing, the asymptotic distribution is not an option in practice to calibrate the proposed goodness–of–fit test. For that reason, a parametric bootstrap calibration procedure is investigated. As in other bootstrap methods, the main idea is to approximate the distribution of R_n by the distribution of the bootstrap statistic R_n^*, which can be obtained by Monte Carlo. Then, it is necessary to define the bootstrap statistic

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

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 testing procedure in practice, together with the bootstrap strategy, are given in the next algorithm. It is stated only for the composite hypothesis, but the calibration for the simple hypothesis can be done just by replacing $\hat{\boldsymbol\theta}$ and $\hat{\boldsymbol\theta}^*$ by the parameter $\boldsymbol\theta_0$ under the null hypothesis.

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

1. *Compute $\hat{\boldsymbol\theta}$, a $\sqrt{n}$–consistent estimator of $\boldsymbol\theta_0$.*
2. *Compute $R_n = \int_{\Omega_q\times\mathbb{R}} \left(\hat f_{h,g}(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol\theta}}(\mathbf{x}, z)\right)^2 \omega_q(d\mathbf{x})\, dz$.*
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, but now from the bootstrap sample from (a).*
   (c) *Compute $R_n^{*b} = \int_{\Omega_q\times\mathbb{R}} \left(\hat f_{h,g}^*(\mathbf{x}, z) - LK_{h,g} f_{\hat{\boldsymbol\theta}^*}(\mathbf{x}, z)\right)^2 \omega_q(d\mathbf{x})\, dz$, where $\hat f_{h,g}^*$ is obtained from the bootstrap sample from (a).*
4. *Estimate the p–value of the test by p–value $\approx \frac{\#\left\{R_n^{*b}\leq R_n\right\}}{B}$.*

The consistency of this testing procedure is proved in the next result, based on the bootstrap analogue of assumption **GF2**:

GF4. $\hat{\boldsymbol\theta}^* - \hat{\boldsymbol\theta} = \mathcal{O}_{\mathbb{P}^*}\left(n^{-\frac{1}{2}}\right)$, where $\mathbb{P}^*$ stands for the conditional distribution of $\mathbf{X}_1^*, \ldots, \mathbf{X}_n^*$ on $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Theorem 5 (Bootstrap consistency)**.** *Under conditions **KD1**–**KD3**, **GF1**–**GF2** and **GF4**, for a $\boldsymbol\theta_1 \in \boldsymbol\Theta$,*

$$n(h^q g)^{\frac{1}{2}} \left(R_n^* - \frac{R(K)\lambda_q(L^2)\lambda_q(L)^{-2}}{nh^q g}\right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_{\boldsymbol\theta_1}^2\right) \text{ in probability.}$$

Recall that Theorem 5 holds regardless the null hypothesis is true or not, as the bootstrap resampling always happens under the null hypothesis. Finally, if H_0 is true, then $\boldsymbol\theta_1 = \boldsymbol\theta_0$ and the asymptotic distributions are the same.

4.4 Extensions to directional–directional models

The directional–directional versions of these results follows quite straightforwardly under the analogous assumptions (modifying **GF1** accordingly). For example, the directional–directional test statistic for the composite hypothesis will be:

$$R_n = \int_{\Omega_{q_1}\times\Omega_{q_2}} \left(\hat f_{h_1,h_2}(\mathbf{x}, \mathbf{y}) - L_1 L_{2,h_1,h_2} f_{\hat{\boldsymbol\theta}}(\mathbf{x}, \mathbf{y})\right)^2 \omega_{q_1}(d\mathbf{x})\, \omega_{q_2}(d\mathbf{y}). \tag{11}$$

Only the most important result of the previous subsection is stated for the directional–directional case: the asymptotic distribution of R_n and the testing procedure. The other ones can be also generalized.

Corollary 4 (Goodness–of–fit test for directional–directional densities)**.** *Under the directional–directional versions of conditions **KD1**–**KD3**, **GF1**–**GF2** and under the null hypothesis $H_0 : f = f_{\boldsymbol\theta_0}$, for an unknown $\boldsymbol\theta_0 \in \boldsymbol\Theta$,*

$$n(h_1^{q_1} h_2^{q_2})^{\frac{1}{2}} \left(R_n - \frac{\lambda_{q_1}(L_1^2)\lambda_{q_1}(L_1)^{-2}\lambda_{q_2}(L_2^2)\lambda_{q_2}(L_2)^{-2}}{nh_1^{q_1} h_2^{q_2}}\right) \xrightarrow{d} \mathcal{N}\left(0, 2\sigma_{\boldsymbol\theta_0}^2\right).$$

5. SIMULATION STUDY

The finite sample performance of the proposed directional–linear and directional–directional goodness–of–fit tests are illustrated in this section for a variety of models, sample sizes and bandwidth choices. The study is focused on the circular–linear and circular–circular cases, as these have been the cases more analyzed in the statistical literature. However, the tests can be easily applied in higher dimensions, such as spherical–linear or spherical–circular, due to their general definition and resampling procedure.

Two collections of Circular–Linear (CL) and Circular–Circular (CC) parametric scenarios are considered in the simulation study. Together with the expression of the parametric density, simulation and parametric estimation methods for all of the models have to be considered. Whereas for most of the models these issues have been worked out before, in some cases simulation or MLE methods have to be implemented from scratch and in that cases the techniques used are indicated. It is worth to mention that the use of copulas in some models (see Nelsen (2006) for a comprehensive review) simplifies notably their simulation and estimation, providing an easy way for dealing with circular–linear and circular–circular models. Specifically, copulas are involved in the models obtained from the papers of Johnson and Wehrly (1978), Wehrly and Johnson (1979), Kato (2009) and García-Portugués et al. (2013b). For most of the copula models, two–step MLE is used: the marginals are fitted first by MLE and then the copula is estimated by MLE using the pseudo–observations computed from the fitted marginals. Figures 1 and 2 show the density contours in the cylinder (CL) and in the torus (CC) of the simulation scenarios. A small description of the models is given in the following paragraphs.

For the circular–linear case, the first five models (CL1–CL5) contain parametric densities with independent components and different kinds of marginals, for which estimation and simulation are easily accomplished. The models employ von Mises, wrapped Cauchy, wrapped normal, normal, log–normal, gamma and mixtures of these densities. Models CL6–CL7 represent two parametric choices of the Mardia and Sutton (1978) model for cylindrical variables, which is constructed conditioning a normal density on a von Mises. Models CL8–CL9 contain two parametric densities of the semiparametric circular–linear model of Johnson and Wehrly (1978) (Theorem 5). This family is indexed by a circular density that defines the underlying circular–linear copula, allowing for flexibility both in the specification of the link density and of the marginals. Two–step MLE and simulation by the conditional and inversion methods are implemented for these models. CL10 is the model given in Theorem 1 of Johnson and Wehrly (1978), which considers an exponential density conditioned on a von Mises. For this model, MLE is done obtaining closed–form expressions. CL11 is the model given by the QS copula of García-Portugués et al. (2013b) implemented with cardioid and log–normal marginals, where closed expression for the MLE is derived. Finally, CL12 is the circular–circular copula of Kato (2009) with von Mises and log–normal marginals. This copula can be easily adapted to the circular–linear scenario.

The first models (CC1–CC5) of the circular–circular case contain also parametric densities with independent components and different kinds of marginals (von Mises, wrapped Cauchy, cardioid and mixtures of them). Models CC6–CC7 represent two parametric choices of the sine model given by Singh et al. (2002). This model introduces elliptical contours for bivariate circular densities and also allows for certain multimodality. Simulation has been implemented by the conditional and inversion methods. Models CC8–CC9 are two parametric densities of the semiparametric models of Wehrly and Johnson (1979), which are based on the previous work of Johnson and Wehrly (1978) and comprise as a particular case the bivariate von Mises model of Shieh and Johnson (2005). Again, two–step MLE and simulation by the conditional and inversion methods are implemented. Models CC10–CC11 are two parametric choices of the wrapped normal torus density given in Kent and Mardia (2009), a natural extension of the circular wrapped normal. Finally, CC12 is the copula of Kato (2009) with von Mises marginals.

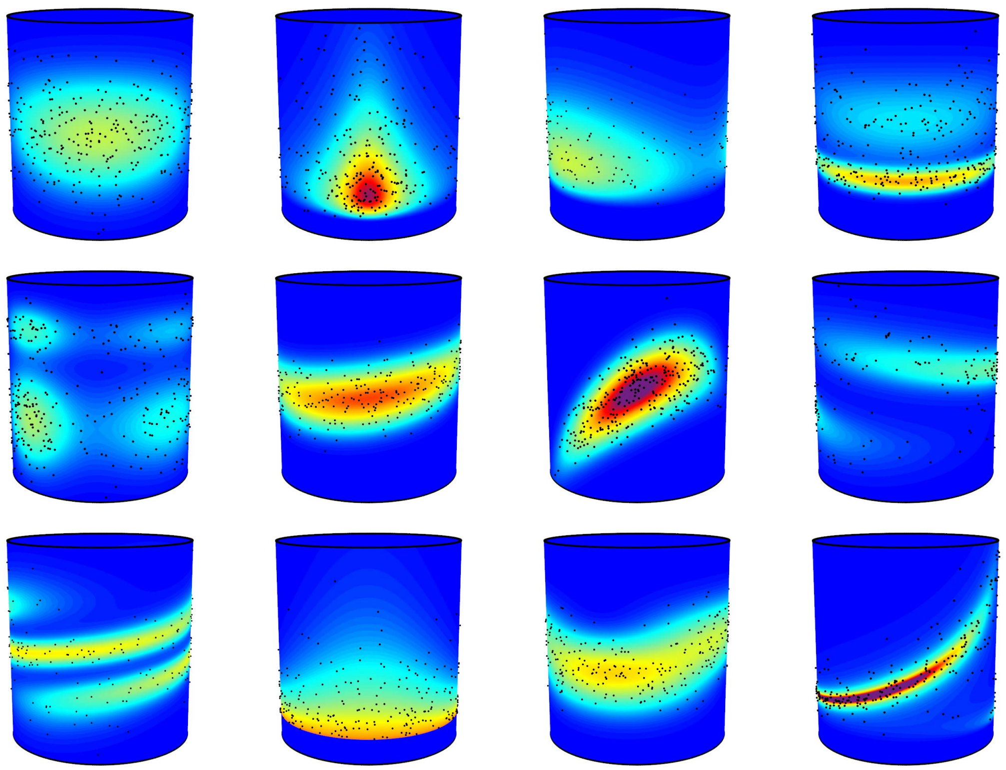

Figure 1: Density models for the simulation study in the circular–linear case. From left to right and up to down, models CL1 to CL12.

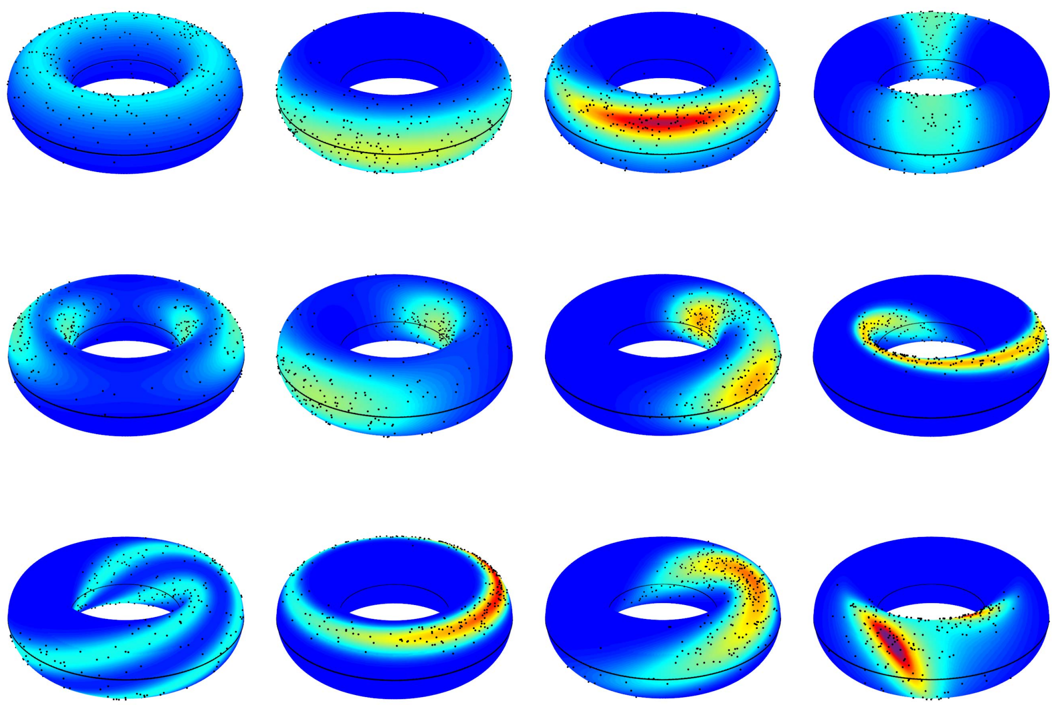

Figure 2: Density models for the simulation study in the circular–circular case. From left to right and up to down, models CC1 to CC12.

The simulation study is focused on the more challenging composite hypothesis. Then, for both circular–linear and circular–circular cases, the null hypothesis for model k will be denoted by

$$H_{k,0} : f \in \mathcal{F}^k_{\boldsymbol{\Theta}} = \left\{ f^k_{\boldsymbol{\theta}} : \boldsymbol{\theta}^k \in \boldsymbol{\Theta}^k \right\},$$

where $\mathcal{F}^k_{\boldsymbol{\Theta}}$ represents the k–th parametric family. The deviations from the null hypothesis are obtained by generating data from a density that does not belong to $\mathcal{F}^k_{\boldsymbol{\Theta}}$. This is done by mixing the true density $f^k_{\boldsymbol{\theta}_0}$ with a density Δ such that the resulting density is not in $\mathcal{F}^k_{\boldsymbol{\Theta}}$:

$$H_{k,\delta} : f = (1-\delta) f^k_{\boldsymbol{\theta}_0} + \delta \Delta, \quad 0 \leq \delta \leq 1.$$

Three mixing densities Δ are considered:

$$\begin{aligned} \Delta_1(\mathbf{x}, z) &= f_{vM}(\mathbf{x}; \mu_1, \kappa) \times \phi_{\sigma_1}(z - m_1), \\ \Delta_2(\mathbf{x}, z) &= f_{vM}(\mathbf{x}; \mu_1, \kappa) \times f_{LN}(z; m_2, \sigma_2), \\ \Delta_3(\mathbf{x}_1, \mathbf{y}) &= f_{vM}(\mathbf{x}; \mu_2, \kappa) \times f_{vM}(\mathbf{y}; \mu_1, \kappa), \end{aligned}$$

where $\mu_1 = (-1, 0)$, $\mu_2 = (1, 0)$, $\kappa = 3$, $m_1 = 2$, $\sigma_1 = 1$, $m_2 = \sigma_2 = \frac{1}{2}$ and $\phi_\sigma(\cdot - m)$ and $f_{LN}(\cdot; m, \sigma)$ denote the densities of $\mathcal{N}(0, \sigma^2)$ and $\mathcal{LN}(m, \sigma^2)$, respectively. To account for similar ranges in the linear data obtained under $H_{k,0}$ and under $H_{k,\delta}$, Δ_1 is used in models CL1, CL4–CL11 and CL13, whereas Δ_2 in the rest of models. In the circular–circular case, the deviation for all models is Δ_3.

The goodness–of–fit tests are applied in the way described in Subsections 4.3 and 4.4 to the whole collection of models, for sample sizes $n = 100, 500, 1000$ and deviations $\delta = 0.00, 0.10, 0.15$ ($\delta = 0.00$ for the null hypothesis). The number of bootstrap replicates is $B = 1000$ and, the number of Monte Carlo replicates to estimate the empirical level and power of the tests is $M = 1000$. The delicate issue of the bandwidth choice for the testing procedure is discussed as follows. Ideally, in a simulation study one would like to run the test in a grid of several bandwidths to see the dependence of the test on the bandwidth choice. However, the two bandwidths involved in the computation of the statistics (9) and (11) complicate notably the evaluation of the test in a relative dense grid (due to a high computational cost) and the posterior presentation of such results (lengthy outputs). Then, a reasonable option is to show the performance of the goodness–of–fit tests for a fixed, but *reasonable*, pair of bandwidths for different situations and show the performance of the tests in a grid of bandwidths for a couple of scenarios in order to illustrate the bandwidth dependence.

The *reasonable* pair of bandwidths that will be considered are computed from the Likelihood Cross Validation (LCV) bandwidths:

$$(h, g)_{\mathrm{LCV}} = \arg\max_{h,g>0} \sum_{i=1}^n \log f^{-i}_{h,g}(\mathbf{X}_i, Z_i), \quad (h_1, h_2)_{\mathrm{LCV}} = \arg\max_{h_1,h_2>0} \sum_{i=1}^n \log f^{-i}_{h_1,h_2}(\mathbf{X}_i, \mathbf{Y}_i), \quad (12)$$

where $f^{-i}_{\dots}$ denotes the kernel estimator computed without the i–th datum. The pair of bandwidths are chosen as the compontentwise median of bandwidths (12) obtained from $M = 1000$ random samples drawn for each combination of model, sample size and deviation. The use of LCV bandwidths with directional data has been extensively tested in García-Portugués (2013) and in García-Portugués et al. (2013a).

Remark 5. It is important to note that for each model, sample size and deviation, the bandwidths (h, g) (respectively, (h_1, h_2)) considered for the $M = 1000$ Monte Carlo replicates are fixed, i.e., for each combination there are M realizations of the statistic R_n. If a data–driven bandwidth $\widehat{(h, g)}$ was used, the empirical level and power obtained in the simulation study will not be the ones of the test statistic R_n, since for each replicate the bandwidths will be different. It is not the scope of this paper to develop theory and resampling methods for a test statistic with data–driven bandwidths $\widehat{(h, g)}$.

Tables 1 and 2 collect the results of the simulation study using the median of the (12) bandwidths for each combination of model (CL or CC), deviation (δ) and sample size (n). When the null hypothesis holds, the levels of the test are correctly attained for the significance levels $\alpha = 0.10, 0.05, 0.01$, for all the sample sizes, models and for the circular–linear and circular–circular cases. When the null hypothesis is false, the tests perform high satisfactorily, having both of them a quick detection of the alternative when only a 10% and a 15% of the data come from a density exogenous to the null parametric family. As expected, the rejection rates grow as the sample size and the deviation from the alternative do.

Model	Sample size n and significance level α								
	$n = 100$			$n = 500$			$n = 1000$		
	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01
$H_{1,0.00}$	0.111	0.051	0.010	0.107	0.052	0.013	0.102	0.048	0.013
$H_{2,0.00}$	0.094	0.051	0.013	0.096	0.049	0.010	0.107	0.050	0.009
$H_{3,0.00}$	0.095	0.048	0.014	0.101	0.046	0.014	0.090	0.050	0.009
$H_{4,0.00}$	0.102	0.045	0.009	0.096	0.039	0.011	0.102	0.045	0.008
$H_{5,0.00}$	0.094	0.049	0.009	0.102	0.049	0.009	0.101	0.041	0.009
$H_{6,0.00}$	0.095	0.039	0.010	0.104	0.043	0.010	0.110	0.050	0.015
$H_{7,0.00}$	0.086	0.042	0.013	0.093	0.043	0.008	0.091	0.049	0.016
$H_{8,0.00}$	0.095	0.049	0.011	0.108	0.050	0.003	0.108	0.044	0.006
$H_{9,0.00}$	0.106	0.062	0.016	0.086	0.043	0.010	0.104	0.064	0.015
$H_{10,0.00}$	0.094	0.045	0.007	0.103	0.056	0.018	0.097	0.045	0.005
$H_{11,0.00}$	0.102	0.059	0.009	0.104	0.056	0.010	0.113	0.056	0.013
$H_{12,0.00}$	0.120	0.073	0.020	0.113	0.054	0.013	0.109	0.051	0.010
$H_{1,0.10}$	0.665	0.552	0.355	1.000	0.997	0.981	1.000	1.000	1.000
$H_{2,0.10}$	0.361	0.244	0.107	0.885	0.805	0.579	0.995	0.982	0.898
$H_{3,0.10}$	0.185	0.107	0.032	0.502	0.362	0.166	0.775	0.659	0.421
$H_{4,0.10}$	0.255	0.172	0.060	0.687	0.568	0.322	0.927	0.868	0.697
$H_{5,0.10}$	0.416	0.272	0.087	0.987	0.972	0.894	1.000	1.000	0.999
$H_{6,0.10}$	0.997	0.996	0.988	1.000	1.000	1.000	1.000	1.000	1.000
$H_{7,0.10}$	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
$H_{8,0.10}$	0.325	0.204	0.069	0.940	0.893	0.723	1.000	1.000	0.983
$H_{9,0.10}$	0.947	0.914	0.796	1.000	1.000	1.000	1.000	1.000	1.000
$H_{10,0.10}$	0.340	0.218	0.089	0.829	0.723	0.481	0.962	0.944	0.838
$H_{11,0.10}$	0.618	0.510	0.296	0.996	0.993	0.963	1.000	1.000	1.000
$H_{12,0.10}$	0.230	0.152	0.057	0.788	0.655	0.442	0.991	0.969	0.895
$H_{1,0.15}$	0.883	0.822	0.621	1.000	1.000	1.000	1.000	1.000	1.000
$H_{2,0.15}$	0.650	0.525	0.311	1.000	0.997	0.977	1.000	1.000	1.000
$H_{3,0.15}$	0.281	0.163	0.055	0.776	0.682	0.420	0.970	0.940	0.860
$H_{4,0.15}$	0.399	0.297	0.127	0.910	0.869	0.724	0.998	0.993	0.981
$H_{5,0.15}$	0.663	0.514	0.235	0.999	0.999	0.999	1.000	1.000	1.000
$H_{6,0.15}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$H_{7,0.15}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$H_{8,0.15}$	0.522	0.379	0.168	0.999	0.997	0.976	1.000	1.000	1.000
$H_{9,0.15}$	0.996	0.989	0.962	1.000	1.000	1.000	1.000	1.000	1.000
$H_{10,0.15}$	0.505	0.378	0.154	0.988	0.975	0.893	1.000	1.000	0.996
$H_{11,0.15}$	0.838	0.763	0.567	1.000	1.000	1.000	1.000	1.000	1.000
$H_{12,0.15}$	0.373	0.254	0.114	0.989	0.967	0.872	1.000	1.000	1.000

Table 1: Empirical size and power of the circular–linear goodness–of–fit test for models CL1–CL12 with different sample sizes, deviations and significance levels.

Finally, the effect of the bandwidths is explored in Figure 3. For models CL1, CL7, CC3 and CC8, the empirical size and power of the tests are computed on a logarithmic spaced 10×10 grid of bandwidths, for sample size $n = 100$ and deviations $\delta = 0.00$ (green, null hypothesis) and $\delta = 0.15$ (orange). As it can be seen, the test is correctly calibrated regardless the bandwidths value. In fact, for the all the bandwidths considered, the rejection rates obtained are inside the 95% confidence interval of the proportion $\alpha = 0.05$. However, the power is notably affected by the choice of the bandwidths, with different behaviours depending on the model and the alternative. Reasonable choices of the bandwidths, such a as the one obtained by the median of the LCV bandwidths (12), present a competitive power.

Model	Sample size n and significance level α								
	$n = 100$			$n = 500$			$n = 1000$		
	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01	α=0.10	α=0.05	α=0.01
$H_{1,0.00}$	0.102	0.061	0.016	0.094	0.047	0.004	0.103	0.048	0.008
$H_{2,0.00}$	0.094	0.054	0.007	0.100	0.043	0.011	0.096	0.056	0.012
$H_{3,0.00}$	0.103	0.061	0.009	0.096	0.042	0.011	0.113	0.058	0.011
$H_{4,0.00}$	0.094	0.049	0.010	0.089	0.048	0.008	0.108	0.052	0.016
$H_{5,0.00}$	0.117	0.059	0.011	0.091	0.050	0.003	0.090	0.051	0.009
$H_{6,0.00}$	0.101	0.069	0.055	0.082	0.045	0.009	0.074	0.034	0.009
$H_{7,0.00}$	0.095	0.048	0.010	0.100	0.059	0.014	0.105	0.044	0.005
$H_{8,0.00}$	0.094	0.043	0.014	0.100	0.054	0.013	0.097	0.050	0.011
$H_{9,0.00}$	0.094	0.043	0.009	0.104	0.057	0.017	0.098	0.042	0.012
$H_{10,0.00}$	0.095	0.047	0.005	0.096	0.041	0.006	0.088	0.042	0.010
$H_{11,0.00}$	0.088	0.041	0.008	0.096	0.047	0.010	0.108	0.053	0.013
$H_{12,0.00}$	0.117	0.062	0.023	0.116	0.058	0.013	0.092	0.048	0.016
$H_{1,0.10}$	0.587	0.456	0.240	0.996	0.995	0.961	1.000	1.000	1.000
$H_{2,0.10}$	0.634	0.506	0.300	0.998	0.994	0.976	1.000	1.000	1.000
$H_{3,0.10}$	0.786	0.706	0.466	1.000	1.000	1.000	1.000	1.000	1.000
$H_{4,0.10}$	0.890	0.837	0.665	1.000	1.000	1.000	1.000	1.000	1.000
$H_{5,0.10}$	0.601	0.431	0.176	1.000	1.000	0.999	1.000	1.000	1.000
$H_{6,0.10}$	0.237	0.123	0.059	0.875	0.759	0.503	0.982	0.958	0.859
$H_{7,0.10}$	0.210	0.112	0.025	0.838	0.724	0.429	0.996	0.989	0.916
$H_{8,0.10}$	0.794	0.693	0.480	1.000	1.000	1.000	1.000	1.000	1.000
$H_{9,0.10}$	0.471	0.325	0.112	1.000	1.000	1.000	1.000	1.000	1.000
$H_{10,0.10}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$H_{11,0.10}$	0.985	0.973	0.910	1.000	1.000	1.000	1.000	1.000	1.000
$H_{12,0.10}$	0.942	0.899	0.788	1.000	1.000	1.000	1.000	1.000	1.000
$H_{1,0.15}$	0.847	0.751	0.521	1.000	1.000	1.000	1.000	1.000	1.000
$H_{2,0.15}$	0.862	0.798	0.627	1.000	1.000	1.000	1.000	1.000	1.000
$H_{3,0.15}$	0.958	0.932	0.830	1.000	1.000	1.000	1.000	1.000	1.000
$H_{4,0.15}$	0.981	0.958	0.885	1.000	1.000	1.000	1.000	1.000	1.000
$H_{5,0.15}$	0.847	0.720	0.445	1.000	1.000	1.000	1.000	1.000	1.000
$H_{6,0.15}$	0.443	0.270	0.097	0.985	0.960	0.858	0.997	0.993	0.982
$H_{7,0.15}$	0.357	0.201	0.043	0.990	0.976	0.879	1.000	1.000	1.000
$H_{8,0.15}$	0.969	0.945	0.842	1.000	1.000	1.000	1.000	1.000	1.000
$H_{9,0.15}$	0.719	0.600	0.345	1.000	1.000	1.000	1.000	1.000	1.000
$H_{10,0.15}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$H_{11,0.15}$	1.000	1.000	0.993	1.000	1.000	1.000	1.000	1.000	1.000
$H_{12,0.15}$	0.999	0.993	0.975	1.000	1.000	1.000	1.000	1.000	1.000

Table 2: Empirical size and power of the circular–circular goodness–of–fit test for models CC1–CC12 with different sample sizes, deviations and significance levels.

6. DATA APPLICATION

The proposed goodness–of–fit tests are applied to study two real datasets. The first dataset comes from forestry and contains the orientations and log–burnt areas of 26870 fires occurred between 1985 and 2005 in Portugal. Due to the size of the dataset, the data has been aggregated in the watersheds of Portugal, resulting in 102 observations of the circular mean orientation and mean log–burnt area of the fires recorded for each watershed. Further details on the data acquisition procedure, the computation of the fires orientation and the watershed delimitation can be seen in Barros et al. (2012) and García-Portugués et al. (2013a). The model proposed by Mardia and

Sutton (1978), discussed previously in the simulation study, was tested for this dataset. The upper left panel of Figure 4 shows the fitted model to the data and the lower left panel shows the p–values of the test in a 10×10 logarithmic grid of bandwidths, obtained with $B = 1000$ bootstrap replicates for each bandwidth pair. The results show that there are not evidences against the null hypothesis.

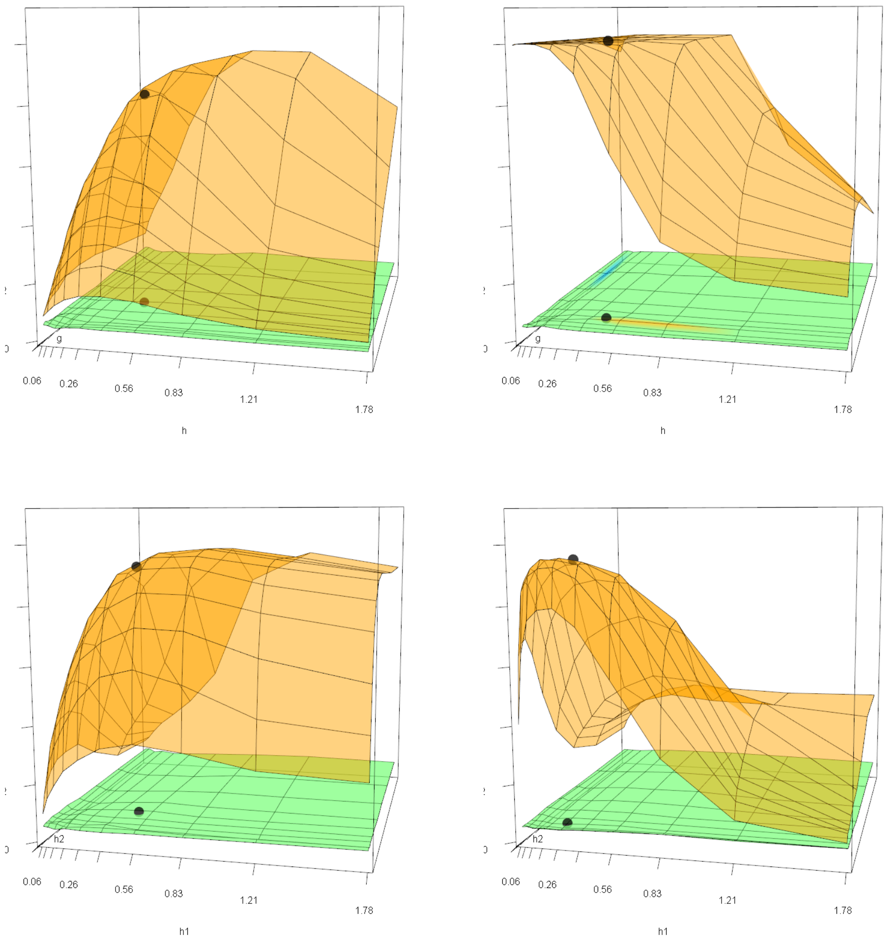

Figure 3: Empirical size and power of the circular–linear and circular–circular goodness–of–fit tests for a 10×10 logarithmic spaced grid. From left to right and up to down, models CL1, CL7, CC3 and CC8. Lower surface represents the empirical rejection rate under $H_{k,0.00}$ and upper surface under $H_{k,0.15}$. Green color represent that the empirical rejection is in the 95% confidence interval of $\alpha = 0.05$, blue that is lower and orange that is larger. Black points represent the empirical size and power obtained with the median of the LCV bandwidths.

The second dataset contains the pairs of dihedral angles of segments of the type alanine–alanine–alanine in alanine amino acids of 1932 proteins. The dataset, formed by 233 pairs of angles, 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) and is available in the `R` package `CircNNTSR` (Fernandez-Duran and Gregorio-Dominguez, 2013). The best model in terms of BIC described in Fernández-Durán (2007) was implemented using two–step MLE for the Wehrly and Johnson (1979) model and the tools of the `CircNNTSR` package for fitting the NTSS parametric densities. As it can be seen in Figure 4, the model is rejected for the significance level $\alpha = 0.05$ (and also for $\alpha = 0.01$). This may be explained by the lack of flexibility of the model of Wehrly and Johnson (1979) to capture the dependence structure between the two angles.

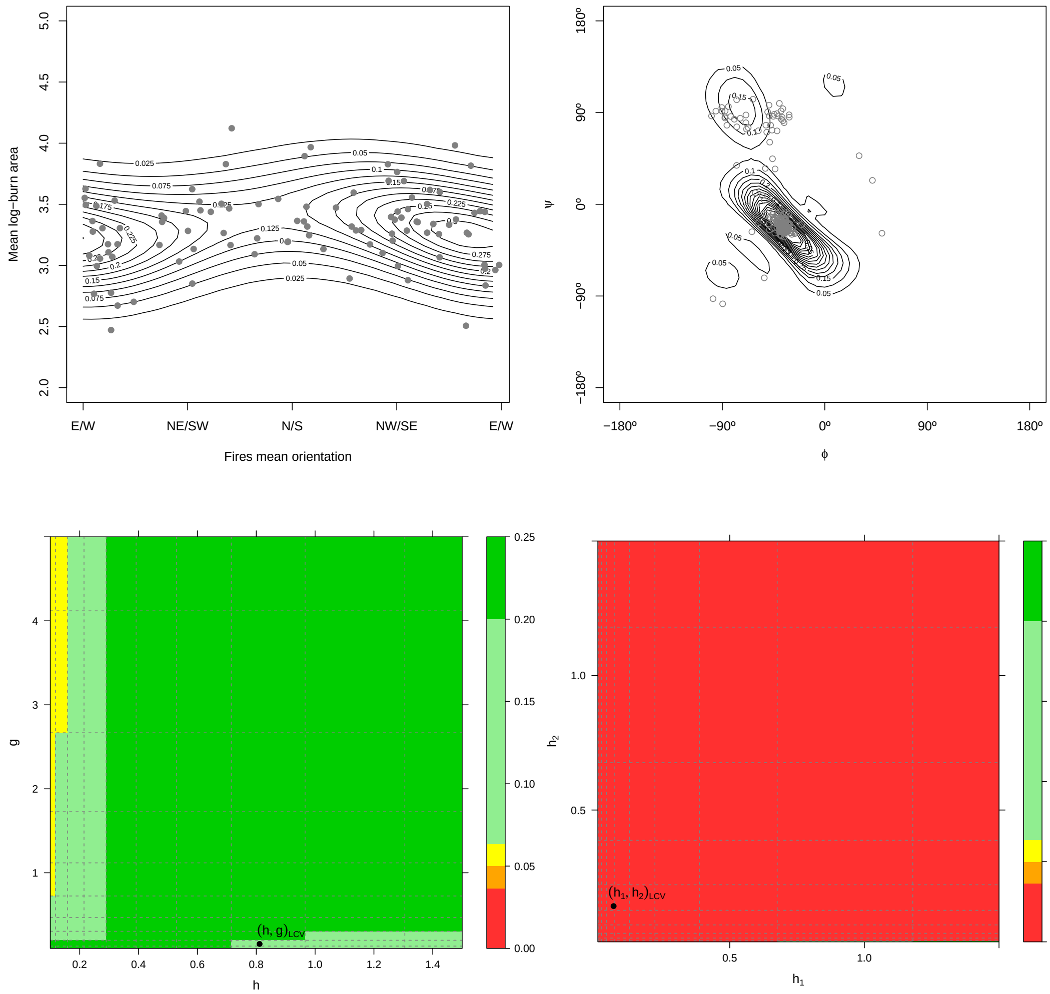

Figure 4: Upper row, from left to right: fitted model of Mardia and Sutton (1978) to the circular mean orientation and mean log–burnt area of the fires in each of the 102 watersheds of Portugal and fitted model for the dihedral angles of the alanine–alanine–alanine segments. Lower row: p–values of the goodness–of–fit tests for a 10×10 grid, with the LCV bandwidth for the data.

7. CONCLUSIONS

A CLT for the integrated squared error of a kernel density estimator for directional–linear data is obtained. The CLT is used to derive goodness–of–fit tests for parametric directional–linear densities. Specifically, simple and composite hypotheses are investigated and results for power under local alternatives and consistency of the bootstrap strategy are given. Further, all the results obtained for directional–linear data can be extended to directional–directional data and are coherent with the previous works in the area.

Apart from the theoretical results, the simulation study corroborates the well behaviour in practice of the goodness–of–fit testing procedure, both for directional–linear data and directional–directional data, that are illustrated with a complete simulation study in a collection of parametric models. Finally, the goodness–of–fit tests are applied to study two datasets coming from forestry and proteomics.

The methodology presented in this paper is based on kernel smoothing and therefore it involves the usual problem of bandwidth choice. Although the effect of the bandwidths is not important for the level of the test, as shown in the simulation study, the power is affected by the choice of the bandwidths. Therefore, further investigation is needed in this direction.

SUPPLEMENTARY MATERIALS

The proofs of all the results stated in this paper are available at http://eio.usc.es/pub/eduardo/. The user and password to access the contents, limited to the use of the jury, is SGAPEIO2013.

ACKNOWLEDGEMENTS

The authors acknowledge the support of Project MTM2008–03010, from the Spanish Ministry of Science and Innovation, Project 10MDS207015PR from Dirección Xeral de I+D, Xunta de Galicia and IAP network StUDyS, from Belgian Science Policy. Work of E. García–Portugués has been supported by FPU grant AP2010–0957 from the Spanish Ministry of Education. The authors also acknowledge the CESGA Supercomputing Center for computational time at the SVG Supercomputer

REFERENCES

Bai, Z. D., Rao, C. R., and Zhao, L. C. (1988) Kernel estimators of density function of directional data. Journal of Multivariate Analysis, 27(1), 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 Ecology and Management, 264, 98–107.

Bickel, P. J. and Rosenblatt, M. (1973) On some global measures of the deviations of density function estimates. The Annals of Statistics, 1(6), 1071–1095.

Boente, G., González-Manteiga, W., and Rodríguez, D. (2013) Goodness–of–fit test for directional data. Scandinavian Journal of Statistics, to appear.

Fan, Y. (1994) Testing the goodness of fit of a parametric density function by kernel method. Econometric Theory, 10(2), 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(2), 579–585.

Fernandez-Duran, J. J. and Gregorio-Dominguez, M. M. (2013) CircNNTSR: CircNNTSR: An R package for the statistical analysis of circular data using nonnegative trigonometric sums (NNTS) models. R package version 2.1.

García-Portugués, E. (2013) Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. Electronic Journal of Statistics, 7, 1655–1685.

García-Portugués, E., Barros, A., Crujeiras, R., González-Manteiga, W., and Pereira, J. (2013a) A test for directional-linear independence, with applications to wildfire orientation and size. Available at arXiv:1301.2550.

García-Portugués, E., Crujeiras, R., and González-Manteiga, W. (2013b) Exploring wind direction and SO_2 concentration by circular–linear density estimation. Stochastic Environmental Research and Risk Assessment, 27(5), 1055–1067.

García-Portugués, E., Crujeiras, R., and González-Manteiga, W. (2013c) Kernel density estimation for directional–linear data. Journal of Multivariate Analysis, to appear.

Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators. Journal of Multivariate Analysis, 14(1), 1–16.

Hall, P., Watson, G. S., and Cabrera, J. (1987) Kernel density estimation with spherical data. Biometrika, 74(4), 751–762.

Johnson, R. A. and Wehrly, T. E. (1978) Some angular–linear distributions and related regression models. Journal of the American Statistical Association, 73(363), 602–606.

Kato, S. (2009) A distribution for a pair of unit vectors generated by Brownian motion. Bernoulli, 15(3), 898–921.

Kent, J. T. and Mardia, K. V. (2009) Principal component analysis for the wrapped normal torus model. Proceedings of the LASR 2009 workshop on Statistical Tools for Challenges in Bioinformatics.

Klemelä, J. (2000) Estimation of densities and derivatives of densities with directional data. Journal of Multivariate Analysis, 73(1), 18–40.

Mardia, K. V. and Jupp, P. E. (2000) Directional statistics. John Wiley & Sons.

Mardia, K. V. and Sutton, T. W. (1978) A model for cylindrical variables with applications. Journal of the Royal Statistical Society: Series B, 40(2), 229–233.

Marron, J. S. and Wand, M. P. (1992) Exact mean integrated squared error. The Annals of Statistics, 20(2), 712–736.

Nelsen, R. B. (2006) An introduction to copulas. Springer, second edition.

Oliveira, M., Crujeiras, R., and Rodríguez-Casal, A. (2012) A plug-in rule for bandwidth selection in circular density estimation. Computational Statistics & Data Analysis, 56(12), 3898–3908.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F., and Clark, C. W., editors (2010) NIST handbook of mathematical functions. U.S. Department of Commerce National Institute of Standards and Technology.

Parzen, E. (1962) On estimation of a probability density function and mode. Annals of Mathematical Statistics, 33, 1065–1076.

Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics, 27, 832–837.

Rosenblatt, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence. The Annals of Statistics, 3, 1–14.

Rosenblatt, M. and Wahlen, B. E. (1992) A nonparametric measure of independence under a hypothesis of independent components. Statistics & Probability Letters, 15(3), 245–252.

Scott, D. W. (1992) Multivariate density estimation. John Wiley & Sons.

Shieh, G. S. and Johnson, R. A. (2005) Inferences based on a bivariate distribution with von Mises marginals. Annals of the Institute of Statistical Mathematics, 57(4), 789–802.

Silverman, B. W. (1986) Density estimation for statistics and data analysis. Chapman & Hall.

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

Taylor, C. C. (2008) Automatic bandwidth selection for circular density estimation. Computational Statistics & Data Analysis, 52(7), 3493–3500.

Wand, M. P. and Jones, M. C. (1995) Kernel smoothing. Chapman & Hall.

Watson, G. S. (1983) Statistics on spheres. John Wiley & Sons.

Wehrly, T. E. and Johnson, R. A. (1979) Bivariate models for dependence of angular observations and a related Markov process. Biometrika, 67(1), 255–256.

Zhao, L. and Wu, C. (2001) Central limit theorem for integrated square error of kernel estimators of spherical density. Science in China Series A, 44(4), 474–483.