

# Goodness-of-fit Test for Directional Data

GRACIELA BOENTE and DANIELA RODRIGUEZ

*Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET*

WENCESLAO GONZÁLEZ MANTEIGA

*Departamento de Estadística e Investigación Operativa, Universidad de Santiago de Compostela*

**ABSTRACT.** In this paper, we study the problem of testing the hypothesis on whether the density  $f$  of a random variable on a sphere belongs to a given parametric class of densities. We propose two test statistics based on the  $L^2$  and  $L^1$  distances between a non-parametric density estimator adapted to circular data and a smoothed version of the specified density. The asymptotic distribution of the  $L^2$  test statistic is provided under the null hypothesis and contiguous alternatives. We also consider a bootstrap method to approximate the distribution of both test statistics. Through a simulation study, we explore the moderate sample performance of the proposed tests under the null hypothesis and under different alternatives. Finally, the procedure is illustrated by analysing a real data set based on wind direction measurements.

*Key words:* asymptotic properties, bootstrap tests, density estimation, hypothesis testing, maximum likelihood estimators, spherical data, von Mises distribution

## 1. Introduction

The problem of testing the hypothesis that a random sample  $\mathbf{x}_1, \dots, \mathbf{x}_n$  is generated by a specific distribution or by some set of distributions has been widely studied in the literature. To our knowledge, the first authors who tackled the problem of goodness of fit when  $\mathbf{x}_i \in \mathbb{R}$  using non-parametric density estimators were Bickel & Rosenblatt (1973). See also Rosenblatt (1975, 1991), Ahmad & Cerrito (1993) and Fan (1994, 1998) for some results in the multidimensional case with Euclidean data, that is, when the observations belong to an open subset of  $\mathbb{R}^d$ . All these papers used as criterion the  $L^2$  distance between parametric and non-parametric (kernel-based) estimators of the unknown density. On the other hand, the literature with  $L^1$  distance is more scarce, and we refer to Cao & Lugosi (2005) for some development.

In many applications, as is the case when dealing with directional data, the variables under study have an additional structure, and this structure needs to be taken into account in both the estimation and inference procedures. Several authors such as Hall *et al.* (1987), Fisher *et al.* (1993) and Mardia & Jupp (2000) discussed estimation methods for spherical and circular data. Beran (1979) considered the situation of exponential models for directional data and a goodness-of-fit test for nested models. Related to the problem of predicting the average wind speed to harvest electricity from wind energy, Hering & Genton (2010) showed the advantage of treating the wind direction as a circular variable (also, Genton & Hering, 2007). In this setting, modelling the wind direction distribution using parametric density families is an important issue, and so the aim of our work is to consider the problem of testing the hypothesis on whether the density  $f$  of a random variable on a sphere belongs to a given class of densities.

For this purpose, the problem of testing whether  $f$  belongs to a parametric class of densities is considered. Our test statistic is based on  $L^p$  distances between a non-parametric density estimator of  $f(\mathbf{x})$  adapted to circular data and a smoothed version of a parametric estimator of the data density. In particular, we study the simple null hypothesis, that is, the situation in which we want to know if the sample density equals a completely specified density function  $f_0$ . In Section 2, we define the test statistics for each of the hypothesis to be considered.

In Section 3, for the composite hypothesis, that is, for the problem of testing whether  $f$  belongs to a parametric class of densities being considered, the asymptotic distribution of the  $L^2$  test, under the null hypothesis and under a set of contiguous alternatives, is obtained. A discussion of the particular case of the  $L^1$  distance test, including the possible difficulties and some open problems, is provided in Section 4. Besides, for the simple null hypothesis, we give a heuristic argument regarding the asymptotic properties for the  $L^p$  test statistic under the null hypothesis. The study of the asymptotic properties of the test based on the  $L^p$  distances, when  $p \neq 2$ , for the composite hypothesis is much more delicate and is unknown even in the Euclidean setting. Bootstrap procedures and their validation for the  $L^2$  distance are studied in Section 5. In Section 6, through a simulation study, we explore the performance of the test procedures introduced in this paper, for moderate sample sizes, under null hypothesis and under a set of alternatives. Some of the obtained results can be seen in the online Supporting Information on the journal website. Finally, Section 7 presents a real data example based on wind direction measurements. Some conclusions are discussed in Section 8. Proofs are relegated to the Appendix.

**2. The test statistics**

Let  $\mathbf{x}_1, \dots, \mathbf{x}_n$  be independent observations of a random variable  $\mathbf{x}$  taking values in the  $d$ -dimensional unit sphere  $\mathcal{S}_d$  in  $\mathbb{R}^{d+1}$  with probability density function  $f(\mathbf{x})$  on  $\mathcal{S}_d$  such that  $\int_{\mathcal{S}_d} f(\mathbf{x}) \omega_d(d\mathbf{x}) = 1$ , where  $\omega_d$  is the rotation-invariant measure on the sphere. We begin by considering the situation in which the null hypothesis is completely specified by a fixed density. In Section 2.2, we extend the procedure to adapt to the case of the composite null hypothesis.

*2.1. Testing a simple null hypothesis*

In this section, we study the problem of testing the hypothesis

$$H_0 : f(\mathbf{x}) = f_0(\mathbf{x}) \quad \text{against} \quad H_1 : f(\mathbf{x}) \neq f_0(\mathbf{x}) \tag{1}$$

at a specified significance level  $\alpha$ , where  $f_0$  is a fixed density function. A natural approach is to consider as a measure of discrepancy to build the test statistic the  $L^2$  or, more generally,  $L^p$  distance between the target density  $f_0$  and a non-parametric estimator  $f_n$ , for example, a kernel estimator. Because the kernel estimator is biased, this measure needs to be modified. Instead of a comparison of the kernel estimators  $f_n$  and  $f_0$ , the idea is to compare  $f_n$  with its expected value under the null hypothesis. In the context of spherical data, kernel density estimators need to be adapted to the structure beyond the data. We will consider the kernel density estimator  $f_n$  suggested by Bai *et al.* (1988) and Hall *et al.* (1987) and defined as

$$f_n(\mathbf{x}) = \frac{c(h)}{n} \sum_{i=1}^n K\left(\frac{1 - \mathbf{x}^T \mathbf{x}_i}{h^2}\right), \tag{2}$$

where  $c(h)$  is a normalizing constant given by  $c(h)^{-1} = \int_{\mathcal{S}_d} K((1 - \mathbf{x}^T \mathbf{y})/h^2) \omega_d(d\mathbf{x})$  and  $h = h_n$  stands for the smoothing parameter. Hence, the test statistic  $T_{0,n}^{(p)}$  based on the  $L^p$  distance is defined as

$$T_{0,n}^{(p)} = \int |f_n(\mathbf{x}) - K_h f_0(\mathbf{x})|^p \omega_d(d\mathbf{x}), \tag{3}$$

where  $K_h g(\mathbf{x}) = c(h) \int K((1 - \mathbf{x}^T \mathbf{y})/h^2) g(\mathbf{y}) \omega_d(d\mathbf{y})$  and  $f_n$  is the kernel estimator defined through ‘(2)’. In particular, we will denote as  $T_{0,n}$  the test statistic  $T_{0,n}^{(2)}$  based on the  $L^2$  distance.

Zhao & Wu (2001) studied the asymptotic behaviour of  $I_n^{(2)} = \int (f_n(\mathbf{x}) - f_o(\mathbf{x}))^2 \omega_d(\mathbf{x})$ . Thus, as mentioned before, one may consider  $I_n^{(2)}$  instead of  $T_{o,n}$ . However, as pointed out for Euclidean data by Fan (1994), the bias introduced in kernel density estimation has a significant influence on  $I_n^{(2)}$ . In particular, theorem 1 in Zhao & Wu (2001) entails that depending on the rate of the bandwidth, that is, the relation between bias and variance, three different rates of convergence are attained for  $I_n^{(2)}$ . To be more precise, if  $nh^{d+4} \rightarrow \infty$ , that is, if the data are oversmoothed, the bias will be large relative to variance, so  $I_n^{(2)}$  converges at the rate of  $\sqrt{nh}^{-2}$ , whereas if the data are undersmoothed ( $nh^{d+4} \rightarrow 0$ ), the rate of convergence is  $nh^{d/2}$ . Finally, if  $nh^{d+4} \rightarrow \delta$ , that is, if the bias is balanced with the variance, the rate of convergence is still  $n^{(d+8)/(2d+8)} = nh^{d/2} = \sqrt{nh}^{-2}$ . As for the Euclidean case, the tests derived from  $I_n^{(2)}$  may have trivial power against certain Pitman alternatives when oversmoothing. For this reason, the statistic  $T_{o,n}$  should be preferred because it removes the bias inherent in kernel density estimation. This idea was used by Härdle & Mammen (1993) to construct tests for a parametric regression model and also by Fan (1994) for testing the goodness of fit for a parametric density family. In particular, as shown in theorem 1, the asymptotic distribution of  $T_{o,n}$  is the same whether the data are oversmoothed, optimally smoothed, or undersmoothed, and it is the same as that of  $I_n^{(2)}$  for undersmoothed data. As stated in theorem 2, the test based on  $T_{o,n}$ , similar to that based on  $I_n^{(2)}$ , detects Pitman alternatives with order of convergence  $n^{1/2}h^{d/4}$ ; however, it should be preferred to the test based on  $I_n^{(2)}$  because it allows a wider range of smoothing parameters. It is also worth noticing that when  $T_{o,n}$  is considered, only the continuity of the density function  $f$  is needed.

2.2. Testing a null composite hypothesis

Let  $\mathcal{F} = \{f_{\beta}(\mathbf{x}) : \beta \in \mathcal{B}\}$  be a family of density functions parametrized with a vector of parameters  $\beta \in \mathcal{B}$ , where  $\mathcal{B}$  is a subset of  $\mathbb{R}^p$ . The parameter  $\beta$  will denote the indexing parameter of the family  $\mathcal{F}$ . We are interested in testing the composite hypothesis

$$H_o : f \in \mathcal{F} \quad \text{against} \quad H_1 : f \notin \mathcal{F} . \tag{4}$$

Under  $H_o$ , we have  $f = f_{\beta}$  for some  $\beta = \beta_o \in \mathcal{B}$ . In this case, a parametric estimator  $f_{\hat{\beta}}$  of the density  $f$  needs to be considered. One way to proceed is to measure the distance between  $f_{\hat{\beta}}$  and the non-parametric estimator  $f_n$  and to use this distance for testing the parametric model. However, as in Section 2.1 and because of the bias of the non-parametric estimator, a better approach is to consider the  $L^2$  distance between  $f_n$  and a smooth version of  $f_{\hat{\beta}}$ . Therefore, the test statistic based on the  $L^2$  distance is defined as

$$T_n = T_n^{(2)} = \int \left( f_n(\mathbf{x}) - K_h f_{\hat{\beta}}(\mathbf{x}) \right)^2 \omega_d(\mathbf{x}) , \tag{5}$$

whereas, more generally, that based on  $L^p$  distances is defined as

$$T_n^{(p)} = \int \left| f_n(\mathbf{x}) - K_h f_{\hat{\beta}}(\mathbf{x}) \right|^p \omega_d(\mathbf{x}) ,$$

where  $f_n$  and  $f_{\hat{\beta}}$  are the kernel estimator defined through ‘(2)’ and the parametric estimator, respectively, and  $K_h g(\mathbf{x}) = c(h) \int K((1 - \mathbf{x}^T \mathbf{y})/h^2) g(\mathbf{y}) \omega_d(\mathbf{y})$ , as before.

Typically, a root- $n$  estimator  $\hat{\beta}$  of  $\beta_o$  needs to be considered to ensure the proper rate of convergence. A good root- $n$  option for  $\hat{\beta}$  is the maximum likelihood estimator of  $\beta_o$ . It is well-known that, under regularity conditions, maximum likelihood estimators are

asymptotically normally distributed. In particular, when dealing with circular data, Cox (1974) studied the asymptotic properties of maximum likelihood estimators in the case of the von Mises distribution.

**3. Asymptotic behaviour of the statistic based on the  $L^2$  distance**

In this section, we derive the asymptotic distribution of the test statistic  $T_n$  defined in ‘(5)’, under the null hypothesis and under a sequence of regular contiguous alternatives. The proofs of these results are relegated to the Appendix.

It is worth noticing that, as the simple null hypothesis is a particular case of a composite hypothesis, the asymptotic null distribution of the test statistic  $T_{o,n}$ , defined in ‘(3)’ for  $p = 2$ , may be derived from theorem 1 whereas its behaviour under contiguous alternatives may be derived from theorem 2. However, it should be pointed out that when considering the simple null hypothesis defined in ‘(1)’, assumptions **A3** and **A5** are not needed.

From now on, denote by  $\tau_d$  the surface area of  $S_d$ , that is,  $\tau_d = 2\pi^{(d+1)/2}/\Gamma((d + 1)/2)$ , for  $d \geq 1$ . Let  $\gamma_d$  and  $g_d(r)$  be defined as, respectively,

$$\gamma_d = \begin{cases} 2^{-1/2} & \text{when } d = 1 \\ \tau_{d-1}\tau_{d-2}^{2^{3(d/2-1)}} & \text{for } d > 1, \end{cases}$$

and

$$g_d(r) = \begin{cases} \int_0^\infty \rho^{-1/2} K(\rho)[K(r + \rho - 2(r\rho)^{1/2}) + K(r + \rho + 2(r\rho)^{1/2})] d\rho & \text{when } d = 1 \\ \int_0^\infty \rho^{d/2-1} K(\rho) \int_{-1}^1 (1 - \theta^2)^{(d-3)/2} K(r + \rho - 2\theta(r\rho)^{1/2}) d\theta d\rho & \text{for } d > 1. \end{cases}$$

To obtain the asymptotic distribution of the test statistics, we need the following assumptions:

- (A1) The kernel  $K : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$  is a bounded and integrable function with a compact support. Moreover, if  $d = 1$ ,  $\int_0^\infty K(r)r^{d/2-1}dr < \infty$ .
- (A2) The density function  $f$  is continuous on  $S_d$ .
- (A3) The function  $f_{\beta}(\mathbf{x})$  is twice continuously differentiable with respect to  $\beta$ , and its partial derivatives are bounded and uniformly continuous with respect to  $(\beta, \mathbf{x})$ .
- (A4) The sequence  $h$  satisfies  $nh^d \rightarrow \infty$  and  $h \rightarrow 0$  as  $n \rightarrow \infty$ .
- (A5) There exists  $\beta_1$  such that

- (a)  $\widehat{\beta} - \beta_1 = O_{\mathbb{P}}(n^{-1/2})$ , when  $\mathbf{x}_i \sim f$  (besides, when  $\mathbf{x}_i \sim f_{\beta_o}, \beta_1 = \beta_o$ ).
- (b)  $\widehat{\beta} - \beta_o = O_{\mathbb{P}}(n^{-1/2})$ , when  $\mathbf{x}_i \sim f_{\beta_o}(\mathbf{x}) + (1/\sqrt{nh^{d/2}}) \Delta(\mathbf{x})$ , where  $\int \Delta(\mathbf{x})\omega_d(d\mathbf{x}) = 0$ .

*Remark 1.* It is worth noting that **A4** together with **A1** entails that  $h^d c(h) \rightarrow \lambda^{-1}$  as  $n \rightarrow \infty$  with  $\lambda = 2^{d/2-1} \tau_{d-1} \int_0^\infty K(r)r^{d/2-1}dr$ , where  $\tau_d$  is the surface area of  $S_d$  (Bai et al., 1988). Assumptions **A1**, **A2**, and **A4** and the fact that  $h^d c(h) \rightarrow \lambda^{-1}$  were used by Zhao & Wu (2001) to derive the limiting distribution of the integrated squared error of the kernel density estimator of  $f(\mathbf{x})$ . As in Fan (1998), assumption **A5** is introduced to examine the effect of estimating  $\beta$  on the asymptotic distribution of  $T_n$ . If the parametric distribution  $f_{\beta_o}(x)$  is correctly specified, that is, if  $H_o$  holds, then  $\beta_1 = \beta_o$ , the true value of  $\beta$ . If  $H_o$  is not true, then  $\beta_1$  can be regarded as a *pseudotrue* value of  $\beta$ . As mentioned in Section 2.2, the validity of the assumption **A5** was verified, under  $H_o$ , in Cox (1974) under certain regularity conditions for the maximum likelihood estimator when considering the von Mises distribution. Beran (1979) extended these results for exponential families, which include the von Mises and

Bingham distributions and also provide a regression-based estimator, which turned out to be root- $n$  consistent.

**Theorem 1.** Assume that **A1–A4** and **A5(a)** hold. Then, under  $H_o$ , that is, when  $f = f_{\beta_o}$ , we have  $nh^{d/2} (T_n - b/(nh^d)) \xrightarrow{D} N(0, 2\lambda^{-4}\sigma_{\beta_o}^2)$ , where  $\lambda = 2^{d/2-1} \tau_{d-1} \int_0^\infty K(r)r^{d/2-1} dr$ ,  $\sigma_{\beta_o}^2 = \gamma_d \int_{S_d} f_{\beta_o}^2(\mathbf{x}) \omega_d(d\mathbf{x}) \int_0^\infty r^{d/2-1} g_d^2(r) dr$  and

$$b = \frac{\int_0^\infty K^2(r)r^{d/2-1} dr}{2^{d/2-1} \tau_{d-1} \left[ \int_0^\infty K(r)r^{d/2-1} dr \right]^2}.$$

Assume that when  $f \notin \mathcal{F}$ , **A5** holds, that is,  $\hat{\beta} \xrightarrow{a.s.} \beta_1$  for some  $\beta_1 \in \mathcal{B}$ , and furthermore that  $\sigma_{\hat{\beta}} \xrightarrow{a.s.} \sigma_{\beta_1} > 0$ . Hence, the dominated convergence theorem entails that  $T_n \xrightarrow{P} T = \int (f(\mathbf{x}) - f_{\beta_1}(\mathbf{x}))^2 \omega_d(d\mathbf{x}) > 0$ . For a given significance level  $\alpha$ , denote by  $z_\alpha$  the upper  $\alpha$ -quantile of the normal distribution, that is,  $\mathbb{P}(Z > z_\alpha) = \alpha$  with  $Z \sim N(0, 1)$ . Then, the test rejecting  $H_o$  when  $V_n = nh_n^{d/2} (T_n - b/(nh^d)) > \hat{v}_\alpha$  where  $\hat{v}_\alpha = \sqrt{2}\lambda^{-2}\sigma_{\hat{\beta}}z_\alpha$  provides a consistent test, as  $P(V_n > \hat{v}_\alpha) = P(T_n > \hat{v}_\alpha / (nh_n^{d/2}) + b/(nh^d)) \rightarrow P(T > 0) = 1$ . As is well-known, consistency is a desirable property for test statistics, as a consistent test will reject a false null hypothesis with probability 1 asymptotically. An important issue for consistent tests is to study their local power properties under sequences of local alternatives. To achieve this goal, theorem 2 studies the behaviour of  $T_n$  under the sequence of regular contiguous alternatives defined by

$$H_{1c} : f(\mathbf{x}) = f_{\beta_o}(\mathbf{x}) + \frac{1}{\sqrt{nh^{d/2}}} \Delta(\mathbf{x}), \tag{6}$$

where  $\int \Delta(\mathbf{x})\omega_d(d\mathbf{x}) = 0$ . Local power properties for this family of Pitman alternatives sequence were studied by Fan (1994) for Euclidean data.

**Theorem 2.** Assume that **A1–A4** and **A5(b)** hold. Then, under  $H_{1c}$  defined in ‘(6)’, we have  $nh^{d/2}(T_n - b/(nh^d)) \xrightarrow{D} N\left(\int \Delta^2(\mathbf{x})\omega_d(d\mathbf{x}), 2\lambda^{-4}\sigma_{\beta_o}^2\right)$ , where  $\lambda$ ,  $b$ , and  $\sigma_{\beta_o}$  are defined in theorem 1.

**4. Some remarks on the  $L^p$  distance test statistic when  $p \neq 2$**

As mentioned in Section 1, the  $L^1$  or, more generally, any  $L^p$  distance for  $p \neq 2$  can also be considered to measure the discrepancy between the kernel estimator and a smooth version of  $f_{\hat{\beta}}$ . To be more precise, if we want to test  $f = f_o$ , one may consider the statistic  $W_{o,n} = T_{o,n}^{(1)} = \int |f_n(\mathbf{x}) - K_h f_o(\mathbf{x})| \omega_d(d\mathbf{x})$ , whereas when composite hypothesis ‘(4)’ is tested, the  $L^1$  test statistics equal

$$W_n = T_n^{(1)} = \int \left| f_n(\mathbf{x}) - K_h f_{\hat{\beta}}(\mathbf{x}) \right| \omega_d(d\mathbf{x}), \tag{7}$$

where  $f_n$  and  $f_{\hat{\beta}}$  are as in Section 2.2.

It is worth noting that the asymptotic distribution of the test based on the  $L^p$  distance, for  $p \neq 2$ , is more delicate. In the case of real variables, that is, when  $\mathbf{x}_i \in \mathbb{R}$ , a central limit theorem for the  $L^p$  distance  $\int |f_n(\mathbf{x}) - f(\mathbf{x})|^p dx$  when  $p \geq 1$  is stated by Csörgö &

Horváth (1987), whereas the proofs of the given results are provided by Csörgö & Horváth (1988). In particular, they proved that, when  $h_n \rightarrow 0$ ,  $\sqrt{nh_n} \rightarrow \infty$ , and  $K$  satisfies standard conditions, the  $L^1$  distance between the kernel density estimator and its expected value is asymptotically normally distributed. More precisely, they obtained that, when  $\mathbf{x} \in \mathbb{R}$ ,  $\sqrt{n}(W_{o,n} - \mathbb{E}(W_{o,n})) \xrightarrow{D} N(0, \sigma^2(K))$ , where  $\sigma^2(K)$  is a constant depending on the kernel and  $W_{o,n}$  is the  $L^1$  test statistic adapted to one-dimensional data.

The important property in this result is the rate of convergence. Instead of the rate  $nh^{1/2}$  attained by the  $L^2$  distance statistic when  $d = 1$  and  $\mathbf{x}_i \in \mathbb{R}$  (Fan, 1994, 1998), the rate  $\sqrt{n}$  is achieved. An extension of the results used by Csörgö & Horváth (1987) to the situation of bivariate vectors was provided by Horváth (1991).

As mentioned earlier, under the simple null hypothesis,  $H_o : f(\mathbf{x}) = f_o(\mathbf{x})$ , theorem 1 entails that  $nh^{d/2} (T_{o,n} - b/(nh^d))$  is asymptotically normally distributed. A systematic study of the asymptotic distribution of the  $L^p$  test statistic  $T_{o,n}^{(p)} = \int |f_n(\mathbf{x}) - K_h f_o(\mathbf{x})|^p \omega_d(d\mathbf{x})$ , under  $H_o$ , requires Poissonization techniques similar to those considered by Horváth (1991). This interesting topic may be the subject of future research. However, some standard computations allow us to heuristically derive that

$$V_{o,n}^{(p)} = \frac{(nh^d)^{p/2}}{h^{d/2}} \left( T_{o,n}^{(p)} - \frac{b_p}{(nh^d)^{p/2}} \right) = \frac{1}{h^{d/2}} \left( (nh^d)^{p/2} T_{o,n}^{(p)} - b_p \right)$$

is asymptotically normally distributed, where  $b_p = b^{p/2} \int_{S_d} f_{\beta_o}^{p/2}(\mathbf{x}) \omega_d(d\mathbf{x}) \mathbb{E}(|N|^p)$  with  $N \sim N(0, 1)$ .

In particular, for directional data, the test statistic  $W_{o,n}$  based on the  $L^1$  distance will also have a root- $n$  rate of convergence when testing the simple null hypothesis. We expect that the same rate is preserved when considering a composite parametric hypothesis if the estimator  $\hat{\beta}$  has a root- $n$  rate of convergence.

In the  $L^1$  situation, the asymptotic behaviour of the test statistic for the composite null hypothesis is in fact much more important and challenging, as usually the parametric estimator is also root- $n$  consistent. Hence, the asymptotic behaviour of both density estimators, the parametric and non-parametric ones, will be relevant for the asymptotic distribution of the test statistics  $W_n$ . As in the case of statistics based on the empirical process for real variables, such as the Kolmogorov–Smirnov or Anderson–Darling statistics, the asymptotic distribution will be much more complicated, and we argue that it will depend on the covariance between the expansion of the parametric component and the asymptotic component of the non-parametric one.

**5. The bootstrap test**

The discussion given in the previous sections motivates the use of bootstrap methods. In particular, for the  $L^2$  test statistic, the rate of convergence is  $nh^{d/2}$ , and so, as in other non-parametric situations, we may expect that the normal approximation will not work well for moderate sample sizes. We briefly discuss this fact in Section 6. To provide an alternative to the asymptotic distribution of test statistics defined in Section 2, we study a bootstrap procedure for each of the hypotheses considered. For this purpose, we will use a parametric bootstrap by generating independent bootstrap samples according to the density  $f_o$  or  $f_{\hat{\beta}}$  depending on the hypothesis. The use of the parametric bootstrap in goodness of fit dates back to Stute et al. (1993), who applied it to a test statistic based on the cumulative distribution. Recently, Genest & Rémillard (2008) extended the result to goodness-of-fit statistics derived from processes having a  $\sqrt{n}$  rate of convergence. Besides, for data belonging to an open subset of  $\mathbb{R}^d$ , Fan (1998) considered a parametric bootstrap goodness-of-fit test. His approach approximates

the critical values of a weighted  $L^2$  distance between the empirical characteristic function and its parametric estimate under the null model. Later on, Neumann & Paparoditis (2000) considered the independent and identically distributed parametric bootstrap to check parametric hypotheses about the stationary density of weakly dependent observations, based also on the  $L^2$  distance between non-parametric and smoothed versions of the parametric density estimator. Several authors, such as Berg (2009), Genest *et al.* (2009), and Bücher & Dette (2010), have investigated the performance of the parametric bootstrap, when considering goodness-of-fit tests for the parametric form of the copula, also based on  $L^2$  distances, showing that it yields reliable approximations to the nominal level.

5.1. Description of the bootstrap tests

Consider first the case of testing a simple null hypothesis. In this case, the distribution of the variables  $\mathbf{x}_1, \dots, \mathbf{x}_n$  is completely known under the null hypothesis. Then, we may consider the bootstrap statistic constructed by generating independent bootstrap samples according to the density  $f_o$ . The bootstrap procedure in this case is in fact a Monte Carlo approximation and can be described as follows:

- Step 1. Generate a random sample of size  $n$ ,  $\mathbf{x}_1^*, \dots, \mathbf{x}_n^*$ , from the distribution  $f_o$ .
- Step 2. Compute the  $L^1$  or  $L^2$  statistics with the bootstrap sample and denote them as  $W_{o,n}^* = T_{o,n}^{(1),*}$  and  $T_{o,n}^* = T_{o,n}^{(2),*}$ , respectively, with  $T_{o,n}^{(p),*} = \int |f_n^*(\mathbf{x}) - K_h f_o(\mathbf{x})|^p \omega_d(\mathbf{d}\mathbf{x})$ , where  $f_n^*$  is the kernel estimator based on  $\{\mathbf{x}_i^*\}_{1 \leq i \leq n}$ .
- Step 3. Repeat steps 1 and 2,  $B$  times. Let  $\widehat{F}_{T_{o,n}^*}$  and  $\widehat{F}_{W_{o,n}^*}$  stand for the empirical distributions of  $\{T_{o,n,1}^*, \dots, T_{o,n,B}^*\}$  and  $\{W_{o,n,1}^*, \dots, W_{o,n,B}^*\}$ , respectively. Compute  $t_{o,n,\alpha}$  (or  $w_{o,n,\alpha}$ ), the upper  $\alpha$ -percentile of  $\widehat{F}_{T_{o,n}^*}$  (or  $\widehat{F}_{W_{o,n}^*}$ ).

Then, the bootstrap procedure based on either the  $L^1$  or  $L^2$  distance rejects  $H_o$  if  $W_{o,n} > w_{o,n,\alpha}$  or  $T_{o,n} > t_{o,n,\alpha}$ , respectively.

When the composite null hypothesis ‘(4)’ is considered, the bootstrap procedure is defined by generating samples according to a parametric density estimator obtained using the original observations. Let  $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(\mathbf{x}_1, \dots, \mathbf{x}_n)$  be a root- $n$  estimator of the indexing parameter  $\boldsymbol{\beta}_o$  and denote by  $f_{\widehat{\boldsymbol{\beta}}}$  the parametric estimator of  $f$ , under  $H_o$ , obtained from it. The bootstrap procedure can be described as follows:

- Step 1. Generate a random sample of size  $n$ ,  $\mathbf{x}_1^*, \dots, \mathbf{x}_n^*$ , from the distribution  $f_{\widehat{\boldsymbol{\beta}}}$ .
- Step 2. Compute the  $L^1$  or  $L^2$  statistics defined through ‘(5)’ and ‘(7)’ with the bootstrap samples and denote them by  $W_n^* = T_n^{(1),*}$  and  $T_n^* = T_n^{(2),*}$ , respectively, where  $T_n^{(p),*} = \int |f_n^*(\mathbf{x}) - K_h f_{\widehat{\boldsymbol{\beta}}^*}(\mathbf{x})|^p \omega_d(\mathbf{d}\mathbf{x})$ , with  $f_n^*$  and  $\widehat{\boldsymbol{\beta}}^* = \widehat{\boldsymbol{\beta}}(\mathbf{x}_1^*, \dots, \mathbf{x}_n^*)$  as the kernel density and the indexing parameter estimator based on  $\{\mathbf{x}_i^*\}_{1 \leq i \leq n}$ , respectively.
- Step 3. Repeat steps 1 and 2,  $B$  times. Let  $\widehat{F}_{T_n^*}$  and  $\widehat{F}_{W_n^*}$  stand for the empirical distributions of  $\{T_{n,1}^*, \dots, T_{n,B}^*\}$  and  $\{W_{n,1}^*, \dots, W_{n,B}^*\}$ , respectively. Compute  $t_{n,\alpha}$  (or  $w_{n,\alpha}$ ), the upper  $\alpha$ -percentile of  $\widehat{F}_{T_n^*}$  (or  $\widehat{F}_{W_n^*}$ ).

Then, we will reject  $H_o$  if  $T_n > t_{n,\alpha}$  (or  $W_n > w_{n,\alpha}$ ).

Note that the same family of estimators of the parameter  $\boldsymbol{\beta}$  needs to be used both when generating the sample and in step 2. To be more precise, if, for instance, the maximum likelihood estimator is used to compute  $\widehat{\boldsymbol{\beta}}$ , then in step 2,  $\widehat{\boldsymbol{\beta}}^*$  is the maximum likelihood estimator of the indexing parameter based on the bootstrap sample  $\{\mathbf{x}_i^*\}_{1 \leq i \leq n}$ .

5.2. Validity of the bootstrap procedure

When the simple null hypothesis ‘(1)’ is considered, it follows immediately that the null distribution of  $T_{o,n}^{(p),*}$  is the same as that of  $T_{o,n}^{(p)}$ . This fact entails the validity of the bootstrap procedure for the  $L^p$  distance tests and, in particular, for both the  $L^1$  and  $L^2$  distance test statistics, when we consider a simple null hypothesis.

On the other hand, for the composite hypothesis, we mentioned, in Section 4, the difficulties arising when dealing with the  $L^p$  distance. Therefore, we will only study the null asymptotic properties of the test statistics based on the  $L^2$  distance,  $T_n^*$ , when  $H_o$  is defined through ‘(4)’. To derive the validity of the bootstrap method for composite null hypothesis, we will need the following additional assumption related to the asymptotic behaviour of the parametric bootstrap estimator.

**A6**  $\widehat{\beta}^* - \widehat{\beta} = O_{\mathbb{P}_*}(n^{-1/2})$ , where  $\mathbb{P}_*$  refers to the conditional distribution of  $\mathbf{x}_i^*$ ’s on  $\mathbf{x}_i$ ’s. Moreover,  $\widehat{\beta} \rightarrow \beta_1$  almost surely.

**Theorem 3.** Assume that **A1–A4**, **A5(a)**, and **A6** hold. Then, conditional on  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , we have  $nh^{d/2} (T_n^* - b/(nh^d)) \xrightarrow{D} N(0, 2\lambda^{-4}\sigma_{\beta_1}^2)$  in probability, where  $\lambda$ ,  $b$ , and  $\sigma^2$  are defined in theorem 1.

Note that theorem 3 holds regardless of whether the null hypothesis is true or not because the bootstrap samples  $\mathbf{x}_1^*, \dots, \mathbf{x}_n^*$  are generated according to  $f_{\widehat{\beta}} \in \mathcal{F}$ . Therefore, the bootstrap procedure leads to a consistent test. Effectively, the following occur:

- (i) Under  $H_o$ , the bootstrap distribution of  $T_n$  converges to the asymptotic null distribution of  $T_n$ , so the asymptotic significance level of the test  $T_n$  based on the bootstrap critical value  $t_{n,\alpha}$  is indeed  $\alpha$ .
- (ii) When the null hypothesis is false, that is, when  $\mathbf{x}_i \sim f \notin \mathcal{F}$ , we have  $T_n \xrightarrow{P} T = \int (f(\mathbf{x}) - f_{\beta_1}(\mathbf{x}))^2 \omega_d(d\mathbf{x}) > 0$ , so that the test statistic  $nh^{d/2} (T_n - b/(nh^d))$  converges to infinity, whereas asymptotically, the bootstrap critical value is still finite.

6. Simulations

This section contains the results of a simulation study in the circle ( $d = 1$ ), designed to evaluate the performance of the test procedures under  $H_o$  and under different alternatives, and is dedicated to exploring numerically different aspects regarding the finite sample performance of the proposed tests. The results of a Monte Carlo study conducted to study the performance of the  $L^2$  test statistics when using the asymptotic approximation derived in Section 3 can be seen in the online Supporting Information on the journal website. The results obtained therein show that approximations of the critical values, as those described in Section 5, are needed. For this reason, the main goal of this section is to validate numerically the good performance of the bootstrap procedure for both the  $L^2$  and  $L^1$  distances for the simple null hypothesis. A composite hypothesis was also considered, and the obtained results are given in the online Supporting Information.

The null hypothesis corresponds to a von Mises (circular normal) distribution. The von Mises distribution has a density function  $f_{\mu,\kappa}(\mathbf{x}) = \exp\{\kappa \mathbf{x}^T \boldsymbol{\mu}\} / (2\pi I_0(\kappa))$ , with  $\boldsymbol{\mu} = (\cos \mu, \sin \mu)$ , where  $0 \leq \mu < 2\pi$  is the mean parameter,  $\kappa > 0$  is the concentration parameter, and  $I_0(\kappa)$  stands for the modified Bessel function of the first kind and order zero, that is,  $I_0(\kappa) = (2\pi)^{-1} \int_0^{2\pi} \exp\{\kappa \cos(\theta)\} d\theta$ . This model has many important applications, as described by Mardia & Jupp (2000) and Jammalamadaka & Sengupta (2001). In particular,



when  $\kappa \rightarrow 0$ , the distribution converges to a uniform distribution on the sphere, whereas the larger the value of  $\kappa$ , the greater is the clustering around the mean parameter  $\mu$ , which is also the mode. It provides an adequate model for phenomena that are rotationally symmetric around  $\mu$ . In particular, Fisher (1953) described this distribution in detail when  $d = 2$ , in the context of statistical mechanics and palaeomagnetism problems. The von Mises distribution has served as a probability model for directions in the plane and is the natural analogue of the normal distribution on  $\mathbb{R}$  (see, for instance, Fisher, 1995, for applications on real circular data sets).

Under the null hypothesis, we generated  $\mathbf{x}_1, \dots, \mathbf{x}_n$  random variables from a von Mises density with mean equal to  $\pi$  and concentration parameter 5, that is,  $f_o = f_{\pi,5}$ . More precisely, the simple null hypothesis corresponds to testing  $H_o : f = f_o$ . We have performed NR = 1000 replications, whereas for the bootstrap procedure, we considered  $B = 5000$  replications.

In the smoothing procedure, we have used Epanechnikov's kernel with bandwidth parameter  $h$ . When performing the bootstrap approximation, we selected several bandwidths to investigate the sensitivity of the tests, in both level and power, with respect to bandwidth choice. In particular, the effects of undersmoothing and oversmoothing on the test power are investigated.

To analyse the performance of bootstrap tests under the null and alternative hypotheses, the sample size is taken as  $n = 100$ , whereas the bandwidths considered are  $h = 0.05, 0.1, 0.25, 0.5, 0.7, 0.9, 1.1$ , and  $1.2$ . We select four particular alternatives denoted by  $H_\delta$  for  $\delta = 0.1, 0.2, 0.35$ , and  $0.5$ . They correspond to generated observations  $\mathbf{x}_1, \dots, \mathbf{x}_n$  with density  $f_\delta = (1 - \delta)f_o + \delta f_1$ , where  $f_o = f_{\pi,5}$  and  $f_1 = f_{\pi/2,5}$ . Figure S1 in the online Supporting Information on the journal website shows the plot of the densities  $f_\delta$  chosen as alternatives.

This set of alternatives can be written as in '(6)', taking  $\Delta(x) = b(\delta) (f_1(x) - f_o(x))$ , where  $b(\delta) = \sqrt{nh^{1/2}}\delta$  depends on the alternative index  $\delta$  and also on the selected bandwidth. The values of  $b(\delta)$  are given in Table S3 in the online Supporting Information on the journal website to make fair comparisons of the obtained results. For instance, the results for  $\delta = 0.2$  and  $h = 0.05$  need to be compared with those related to  $\delta = 0.1$  if the bandwidth lies between 0.5 and 0.7.

For the simple hypothesis  $H_o : f = f_o = f_{\pi,5}$ , the observed frequencies of rejection of the bootstrap tests based on the  $L^2$  and  $L^1$  distances, at the 5% level, are reported in Fig. 1 and Table 1. Table 1 shows the improvement attained in level when the finite sample approximation provided by the bootstrap test is used. For almost all the range of bandwidths, the observed level of the bootstrap test is quite close to the nominal one, when the sample size is  $n = 100$ .

The results of the power study reported in Table 1 and Fig. 1 indicate that both the  $L^2$  and  $L^1$  bootstrap test statistics have a good performance under the null hypothesis. As mentioned before, both tests attain the significance level for all the bandwidths considered. Besides, the test based on distance  $L^1$  detects more easily the alternatives considered for different values of bandwidth. Note that when  $\delta = 0.2$ , the observed frequency of rejection is larger than 0.8 for all the bandwidths. On the other hand, the  $L^2$  test shows a loss of power for small bandwidths. This fact may be related to the rate of convergence of the  $L^2$  test, which depends on the bandwidth parameter.

## 7. Real data example

In this section, we applied the proposed bootstrap test procedure to a real data set. The original data are measurements of the wind directions recorded each minute daily in two meteorological stations at Galicia, in the northwest of Spain. The stations will be referred to as B1 and C9, and their locations are plotted in Fig. 2. The observations studied here correspond to August 2009 with a sample size of 26,426. In the figures and tables, we denote as zero the north direction,

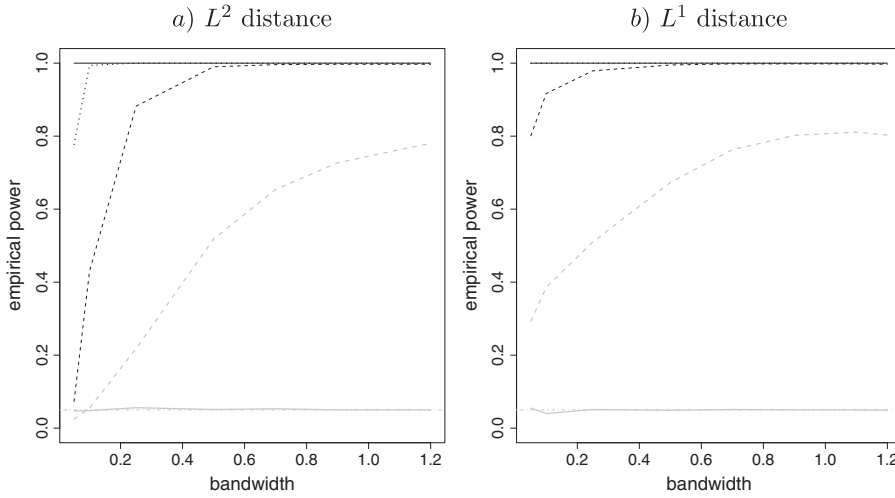


Fig. 1. Frequencies of rejections of the bootstrap test at the 5% level for the simple hypothesis  $H_0 : f = f_0 = f_{\pi,5}$  corresponding to several bandwidths  $h$ . The line in grey corresponds to the null hypothesis ( $\delta = 0$ ), whereas the solid, dotted, dashed black, and dashed grey lines correspond to the alternatives  $H_\delta$  with  $\delta = 0.5, \delta = 0.35, \delta = 0.2$ , and  $\delta = 0.1$ , respectively.

Table 1. Observed frequencies of rejection of the bootstrap test at the 5% level for the simple hypothesis  $H_0 : f = f_{\pi,5}$  and different values of the smoothing parameter  $h$ .  $H_\delta$  stands for the alternative hypothesis when we consider  $f = f_\delta = (1 - \delta)f_0 + \delta f_1$  with  $f_0 = f_{\pi,5}$  and  $f_1 = f_{\pi/2,5}$ , for  $\delta = 0.1, 0.2, 0.35$ , and  $0.5$

| Distance | Hypothesis | $h$   |       |       |       |       |       |       |       |
|----------|------------|-------|-------|-------|-------|-------|-------|-------|-------|
|          |            | 0.05  | 0.1   | 0.25  | 0.5   | 0.7   | 0.9   | 1.1   | 1.2   |
| $L^2$    | $H_0$      | 0.047 | 0.048 | 0.056 | 0.051 | 0.053 | 0.050 | 0.050 | 0.050 |
| $L^1$    | $H_0$      | 0.055 | 0.040 | 0.051 | 0.049 | 0.051 | 0.050 | 0.050 | 0.049 |
| $L^2$    | $H_{0.1}$  | 0.025 | 0.055 | 0.218 | 0.518 | 0.654 | 0.727 | 0.763 | 0.780 |
| $L^1$    | $H_{0.1}$  | 0.293 | 0.386 | 0.51  | 0.673 | 0.763 | 0.802 | 0.811 | 0.803 |
| $L^2$    | $H_{0.2}$  | 0.073 | 0.430 | 0.882 | 0.990 | 0.996 | 0.997 | 0.997 | 0.997 |
| $L^1$    | $H_{0.2}$  | 0.801 | 0.917 | 0.979 | 0.995 | 0.998 | 0.998 | 0.997 | 0.997 |
| $L^2$    | $H_{0.35}$ | 0.777 | 0.994 | 1     | 1     | 1     | 1     | 1     | 1     |
| $L^1$    | $H_{0.35}$ | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     |
| $L^2$    | $H_{0.5}$  | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     |
| $L^1$    | $H_{0.5}$  | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     |

and we measure the data in radians clockwise. The rose diagrams given in Fig. 3 show that the wind directions in both stations have a unimodal distribution with mean around the north.

Let  $f$  be the density function of the wind direction. For both stations, the null hypothesis considered was  $H_0 : f \in \{f_{\mu,\kappa}; (\mu, \kappa) \in [0, 2\pi) \times (0, +\infty)\}$ , where  $f_{\mu,\kappa}$  is the density function of a von Mises distribution with mean  $\mu$  and concentration parameter  $\kappa$  as described in Section 6. Table 2 reports the values of the maximum likelihood estimators of the parameters  $(\mu, \kappa)$ . For the null hypothesis of interest, we compute the  $p$ -values for the two bootstrap tests taking  $B = 5000$  bootstrap replications. The  $p$ -values were computed for different bandwidths to analyse the sensitivity of the results. For each station, the set of bandwidths chosen includes

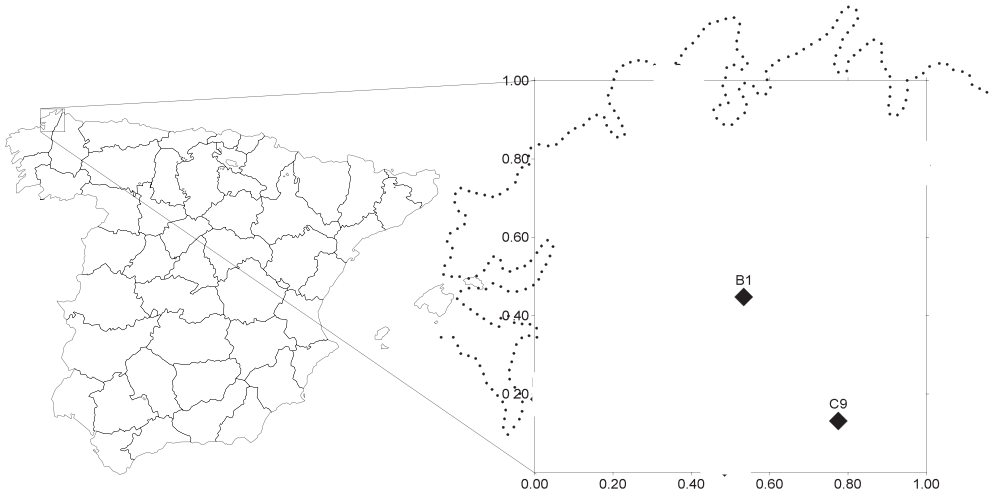


Fig. 2. Location of the meteorological stations B1 and C9 considered.

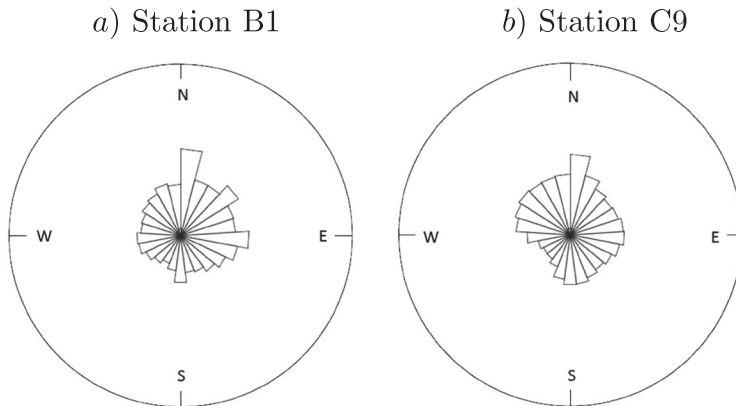


Fig. 3. Rose diagram for stations B1 and C9.

the selector obtained via a cross-validation procedure using the Kullback–Leibler distance (for example, Marron, 1987). To perform the cross-validation, a grid with a length of 46 equally spaced bandwidths between 0.001 and  $2\pi$  was taken. For both stations, the data-driven cross-validation bandwidth was 1.727154. Figure S3 in the online Supporting Information on the journal website shows the plots of the parametric density estimator and the non-parametric estimators for different values of bandwidths.

Note that, for both B1 and C9, a bandwidth of 0.8 shows a good behaviour that is quite similar to that provided by the parametric estimation. Besides, the bandwidth obtained using the cross-validation procedure produces oversmoothed estimations. The obtained  $p$ -values for nine different bandwidths are reported in Table 3 and Fig. 4.

From the obtained results, for both meteorological stations, the wind directions can be assumed to have a von Mises distribution. In both cases, for all bandwidths considered, the empirical  $p$ -values imply that the null hypothesis cannot be rejected.

Table 2. Maximum likelihood estimators ( $\hat{\mu}, \hat{\kappa}$ ) of the parameters ( $\mu, \kappa$ ) under a von Mises parametric model for stations B1 and C9

| Stations | $\hat{\mu}$ | $\hat{\kappa}$ |
|----------|-------------|----------------|
| B1       | 0.6086      | 0.645          |
| C9       | 0.181       | 0.4697         |

Table 3. Empirical  $p$ -values for stations B1 and C9

| Test  | $h$        |        |        |        |        |        |        |         |        |  |
|-------|------------|--------|--------|--------|--------|--------|--------|---------|--------|--|
|       | 0.5        | 0.6    | 0.7    | 0.8    | 1.2    | 1.4    | 1.6    | 1.8     | 2      |  |
|       | Station B1 |        |        |        |        |        |        |         |        |  |
| $L^2$ | 0.6212     | 0.6138 | 0.6006 | 0.5954 | 0.5808 | 0.6008 | 0.6726 | 0.8674  | 1.0000 |  |
| $L^1$ | 0.8548     | 0.9004 | 0.9298 | 0.9714 | 0.9994 | 1.0000 | 1.0000 | 1.0000  | 0.9998 |  |
|       | Station C9 |        |        |        |        |        |        |         |        |  |
| $L^2$ | 0.5182     | 0.5134 | 0.5122 | 0.5340 | 0.5838 | 0.5448 | 0.5890 | 0.6192  | 0.8610 |  |
| $L^1$ | 0.2856     | 0.3430 | 0.4054 | 0.4500 | 0.5214 | 0.8186 | 0.9314 | 0.75787 | 0.6588 |  |

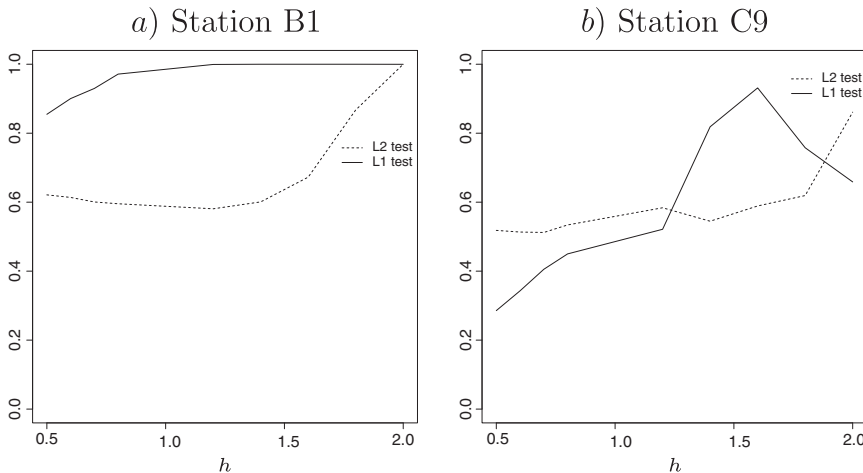


Fig. 4. Empirical  $p$ -values for stations B1 and C9.

**8. Concluding remarks**

Two test statistics were introduced to test if a sample from directional data has a density that belongs to a given parametric family. The asymptotic distribution of the test based on the  $L^2$  distance was given both under the null hypothesis and under a set of contiguous alternatives. The asymptotic performance of  $L^1$  deserves further study, but we conjecture from the results obtained in the Euclidean case that the order of convergence of the  $L^2$  test may be improved using instead the  $L^1$  distance statistic.

Our simulation study illustrates the well-known disadvantage of tests based on non-parametric estimators because, even if they have a normal distribution, their convergence towards the normal distribution is very slow. For this reason, approximations of critical values

are needed. The bootstrap procedure proposed overcomes this problem, as the nominal level is attained for moderate sample sizes.

Finally, the results in Sections 6 and 7 suggest that a deeper study to select the smoothing parameter for testing problems is needed. This topic requires further careful investigation and is still an open problem, even in the setting of Euclidean data.

### Acknowledgements

This work was carried out while Daniela Rodriguez was visiting the Department of Statistics and Operation Research of the Universidad de Santiago de Compostela, supported by a post-doctoral fellowship from CONICET. She is very grateful to all the members of the Group of Statistics for their kind hospitality. This research was partially supported by grants from the Universidad de Buenos Aires, the CONICET, and the ANPCYT, Argentina, and also by a Spanish grant from the Ministerio Español de Ciencia e Innovación and a XUNTA grant from Galicia, Spain.

We wish to thank the Editor, the Associate Editor, and two anonymous referees for valuable comments that led to an improved version of the original paper.

### Supporting Information

Additional supporting information may be found in the online version of this article including the simulation results when using the asymptotic approximation for simple and composite null hypotheses and the results for the bootstrap procedure for a composite null hypothesis.

**Tables S1 and S2:** Observed frequencies of rejection of the asymptotic test based on  $T_{o,n}$  for different values of the bandwidth.

**Table S3:** Values for  $b(\delta)$  for different values of the bandwidth  $h$  and alternative index  $\delta$ .

**Table S4 and Fig. S2:** Observed frequencies of rejection of the bootstrap test at the 5% level for the composite null hypothesis and different values of the smoothing parameter.

**Fig. S1:** Plots of the densities  $f_\delta = (1 - \delta)f_o + \delta f_1$  ( $f_o = f_{\pi,5}$  and  $f_1 = f_{\pi/2,5}$ ) considered in the alternative hypothesis  $H_\delta$ .

**Fig. S3:** Plots of the parametric and non-parametric density estimators.

### References

- Ahmad, I. A. & Cerrito, P. B. (1993). Goodness of fit tests based on the  $L^2$ -norm of multivariate probability density functions. *J. Nonparametr. Stat.* **2**, 169–181.
- Bai, Z., Rao, C. & Zhao, L. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.* **27**, 24–39.
- Beran, R. (1979). Exponential models for directional data. *Ann. Stat.* **7**, 1162–1178.
- Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *Eur. J. Financ.* **15**, 675–701.
- Bickel, P. J. & Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Stat.* **1**, 1071–1095.
- Bücher, A. & Dette, H. (2010). Some comments on goodness-of-fit tests for the parametric form of the copula based on  $L^2$ -distances. *J. Multivariate Anal.* **101**, 749–763.
- Cao, R. & Lugosi, G. (2005). Goodness-of-fit tests based on the kernel density estimator. *Scand. J. Stat.* **32**, 599–616.
- Cox, D. (1974). *Theoretical statistics*, Chapman & Hall, London.
- Csörgö, M. & Horváth, L. (1987). Asymptotics for  $L_p$ -norms of kernel estimators of densities. *Comput. Stat. Data Anal.* **6**, 241–250.

- Csörgő, M. & Horváth, L. (1988). Central limit theorems for  $L_p$ -norms of density estimators. *Probab. Theory Related Fields* **80**, 269–291.
- Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Economet. Theor.* **10**, 316–356.
- Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Economet. Theor.* **14**, 604–621.
- Fisher, N. I. (1995). *Statistical analysis of circular data*, University Press, Cambridge.
- Fisher, N. I., Lewis, T. & Embleton, B. J. J. (1993). *Statistical Analysis of Spherical Data*, University Press, Cambridge.
- Fisher, R. A. (1953). Dispersion on a sphere. *Proc. R. Soc. London, Ser. A* **217**, 295–305.
- Genest, Ch. & Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. H. Poincaré Probab. Statist.* **44**, 1096–1127.
- Genest, Ch., Rémillard, B. & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: a review and a power study. *Insur. Math. Econ.* **44**, 199–213.
- Genton, M. G. & Hering, A. S. (2007). Blowing in the wind. *Significance* **4**, 11–14.
- 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. & Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74**, 751–762.
- Härdle, W. & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Stat.* **21**, 1926–1947.
- Hering, A. S. & Genton, M. G. (2010). Powering up with space–time wind forecasting. *J. Am. Stat. Assoc.* **105**, 92–104.
- Horváth, L. (1991). On  $L_p$ -norms of multivariate density estimators. *Ann. Stat.* **19**, 1933–1949.
- Jammalamadaka, S. & Sengupta, A. (2001). *Topics in Circular Statistics*, Multivariate Analysis, vol. 5, World Scientific, Singapore.
- Mardia, K. & Jupp, P. (2000). *Directional data*, Wiley, New York.
- Marron, J. S. (1987). A comparison of cross validation techniques in density estimation. *Ann. Stat.* **13**, 152–162.
- Neumann, M. & Paparoditis, E. (2000). On bootstrapping  $L^2$ -type statistics in density testing. *Stat. Probabil. Lett.* **50**, 137–147.
- Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Stat.* **3**, 1–14.
- Rosenblatt, M. (1991). *Stochastic curve estimation*, NSF-CMBS Regional Conference Series in Probability and Statistics, vol. 3, Institute of Mathematical Statistics, Hayward, California.
- Stute, W., González-Manteiga, W. & Presedo-Quidimil, M. (1993). Bootstrap based goodness-of-fit tests. *Metrika* **40**, 243–256.
- Zhao, L. & Wu, C. (2001). Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China, Ser. A* **44**, 474–483.

Received August 2011, in final form February 2013

Graciela Boente, Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón 1, Buenos Aires, C1428EHA, Argentina.  
E-mail: gboente@dm.uba.ar

## Appendix: Proofs

*Proof of theorem 1.* Denoting  $Y_{n,i}(\mathbf{x}) = K((1 - \mathbf{x}^T \mathbf{x}_i)/h^2) - \mathbb{E}(K((1 - \mathbf{x}^T \mathbf{x}_i)/h^2))$ , we have the expansion  $\int (f_n(\mathbf{x}) - K_h f_{\hat{\beta}}(\mathbf{x}))^2 \omega_d(\mathbf{x}) = T_{n1} + T_{n2} + T_{n3} + 2T_{n4}$  with  $T_{n3} = \int [K_h(f(\mathbf{x}) - f_{\hat{\beta}}(\mathbf{x}))]^2 \omega_d(\mathbf{x})$ ,  $T_{n4} = \int (f_n(\mathbf{x}) - K_h f(\mathbf{x})) K_h(f(\mathbf{x}) - f_{\hat{\beta}}(\mathbf{x})) \omega_d(\mathbf{x})$ , and

$$T_{n1} = \frac{c(h)^2}{n^2} \sum_{i=1}^n \int Y_{n,i}(\mathbf{x})^2 \omega_d(\mathbf{x}),$$

$$T_{n2} = 2 \frac{c(h)^2}{n^2} \sum_{1 \leq i < j \leq n} \int Y_{n,i}(\mathbf{x}) Y_{n,j}(\mathbf{x}) \omega_d(d\mathbf{x}).$$

Note that, under  $H_o$ ,  $f = f_{\beta_o}$ . Hence, assumptions **A1**, **A2**, and **A4** together with lemma 4 in Zhao & Wu (2001) entail that  $nh^{d/2}(T_{n1} - \mathbb{E}(T_{n1})) \xrightarrow{p} 0$ , whereas lemma 6 in Zhao & Wu (2001) leads us to  $nh^{d/2}T_{n2} \xrightarrow{D} 2^{1/2}\lambda^{-2}\sigma Z$  with  $Z \sim N(0, 1)$ . As in the proof of lemma 4 in Zhao & Wu (2001), lemma 1(i) in Zhao & Wu (2001) together with assumptions **A1**, **A2**, and **A4** imply that

$$\begin{aligned} \mathbb{E}(T_{n1}) &= \frac{c(h)^2}{n} \int_{S_d} \text{var}(K((1 - \mathbf{x}^T \mathbf{x})/h^2)) \omega_d(d\mathbf{x}) \\ &= \frac{1}{nh^d 2^{d/2-1} \tau_{d-1}} \frac{\int_0^\infty K^2(r)r^{d/2-1} dr}{\left[ \int_0^\infty K(r)r^{d/2-1} dr \right]^2} - \frac{1}{n} \int_{S_d} f^2(\mathbf{x}) \omega_d(d\mathbf{x}) + o\left(\{nh^{d/2}\}^{-1}\right) \end{aligned}$$

Therefore, it only remains to be shown that

$$nh^{d/2}T_{n3} \xrightarrow{p} 0, \tag{8}$$

$$nh^{d/2}T_{n4} \xrightarrow{p} 0. \tag{9}$$

Using a Taylor expansion of order 1 and the fact that, under  $H_o$ ,  $f = f_{\beta_o}$ , we obtain

$$T_{n3} = (\widehat{\beta} - \beta_o)^T \int K_h \left. \frac{\partial f_{\beta}(\mathbf{x})}{\partial \beta} \right|_{\beta=\xi_n} \left( K_h \left. \frac{\partial f_{\beta}(\mathbf{x})}{\partial \beta} \right|_{\beta=\xi_n} \right)^T \omega_d(d\mathbf{x}) (\widehat{\beta} - \beta_o),$$

where  $\xi_n$  is an intermediate point between  $\widehat{\beta}$  and  $\beta_o$ . Then, lemma 1 in Zhao & Wu (2001), the dominated convergence theorem, and the fact that from **A5(a)**  $\widehat{\beta}$  has a root- $n$  order of convergence entail that  $nT_{n3} = O_{\mathbb{P}}(1)$ , concluding the proof of ‘(8)’.

Likewise, under  $H_o$ , a second-order Taylor expansion gives  $T_{n4} = T_{n41}(\widehat{\beta} - \beta_o) + (\widehat{\beta} - \beta_o)^T T_{n42}(\widehat{\beta} - \beta_o)$ , where

$$T_{n41} = \frac{1}{n} \sum_{i=1}^n c(h) \int \left[ K \left( \frac{1 - \mathbf{x}^T \mathbf{x}_i}{h^2} \right) - \mathbb{E}K \left( \frac{1 - \mathbf{x}^T \mathbf{x}_1}{h^2} \right) \right] K_h \left( \left. \frac{\partial f_{\beta}(\mathbf{x})}{\partial \beta} \right|_{\beta=\beta_o} \right) \omega_d(d\mathbf{x}),$$

$$T_{n42} = \frac{1}{2n} \sum_{i=1}^n c(h) \int \left[ K \left( \frac{1 - \mathbf{x}^T \mathbf{x}_i}{h^2} \right) - \mathbb{E}K \left( \frac{1 - \mathbf{x}^T \mathbf{x}_1}{h^2} \right) \right] K_h \left( \left. \frac{\partial^2 f_{\beta}(\mathbf{x})}{\partial \beta^T \partial \beta} \right|_{\beta=\xi_n} \right) \omega_d(d\mathbf{x}).$$

Using lemma 1 from Zhao & Wu (2001) and assumptions **A1** and **A3**, we obtain  $\text{var}(\sqrt{nh^{d/2}}T_{n41}) \rightarrow 0$ . Therefore, as  $\mathbb{E}(T_{n41}) = 0$ , we have  $\sqrt{nh^{d/2}}T_{n41} \xrightarrow{p} 0$ , which together with the fact that  $\sqrt{n}(\widehat{\beta} - \beta_o) = O_{\mathbb{P}}(1)$  implies that the first term of  $nh^{d/2}T_{n4}$  converges to 0 in probability. Straightforward calculations allow us to show that the second term of  $T_{n4}$  is negligible compared with the first one, as from **A5(a)**,  $\widehat{\beta} - \beta_o = O_{\mathbb{P}}(n^{-1/2})$ , concluding the proof of ‘(9)’.  $\square$

*Proof of theorem 2.* We have the expansion  $\int (f_n(\mathbf{x}) - K_h f_{\widehat{\beta}}(\mathbf{x}))^2 \omega_d(d\mathbf{x}) = T_{n1} + T_{n2} + T_{n3} + 2T_{n4}$ , where  $T_{nj}$  are defined as in the proof of theorem 1, with  $f = f_{\beta_o} + (1/\sqrt{nh^{d/2}})\Delta$ . Note that  $T_{n4} = T_{n41} + 2T_{n42}$ , where

$$T_{n41} = \int (f_n(\mathbf{x}) - K_h f(\mathbf{x})) K_h (f_{\beta_o}(\mathbf{x}) - f_{\hat{\beta}}(\mathbf{x})) \omega_d(\mathbf{d}\mathbf{x}),$$

$$T_{n42} = \frac{1}{\sqrt{nh^{d/2}}} \int (f_n(\mathbf{x}) - K_h f(\mathbf{x})) K_h \Delta(\mathbf{x}) \omega_d(\mathbf{d}\mathbf{x}),$$

and

$$T_{n3} = \int [K_h (f(\mathbf{x}) - f_{\hat{\beta}}(\mathbf{x}))]^2 \omega_d(\mathbf{d}\mathbf{x}) = \int [K_h (f_{\beta_o}(\mathbf{x}) - f_{\hat{\beta}}(\mathbf{x}))]^2 \omega_d(\mathbf{d}\mathbf{x})$$

$$+ \frac{2}{\sqrt{nh^{d/2}}} \int K_h (f_{\beta_o}(\mathbf{x}) - f_{\hat{\beta}}(\mathbf{x})) K_h \Delta(\mathbf{x}) \omega_d(\mathbf{d}\mathbf{x}) + \frac{1}{nh^{d/2}} \int [K_h \Delta(\mathbf{x})]^2 \omega_d(\mathbf{d}\mathbf{x})$$

$$= T_{n31} + 2T_{n32} + T_{n33}.$$

As in the proof of theorem 1, we have  $\mathbb{E}(T_{n42}) = 0$ . Besides,  $\text{var}(nh^{d/2}T_{n42}) = h^{d/2}\mathbb{E}(Z_{1n}^2)$ , where

$$Z_{1n} = c(h) \int \left[ K \left( \frac{1 - \mathbf{x}^T \mathbf{x}_1}{h^2} \right) - \mathbb{E}K \left( \frac{1 - \mathbf{x}^T \mathbf{x}_1}{h^2} \right) \right] K_h \Delta(\mathbf{x}) \omega_d(\mathbf{d}\mathbf{x}).$$

Using lemma 1 in Zhao & Wu (2001) and the dominated convergence theorem, we can easily derive that  $\mathbb{E}(Z_{1n}^2) = O(1)$ , as **A1** and **A2** hold. Therefore,  $nh^{d/2}T_{n42} = o_{\mathbb{P}}(1)$ . Arguing as in theorem 1, we obtain  $nh^{d/2}T_{n41} = o_{\mathbb{P}}(1)$ , which entails that  $nh^{d/2}T_{n4} = o_{\mathbb{P}}(1)$ .

On the other hand, **A1** and **A2** and lemma 1 in Zhao & Wu (2001) imply that  $nh^{d/2}T_{n33} \rightarrow \int \Delta^2(\mathbf{x})\omega_d(\mathbf{d}\mathbf{x})$ . Similarly, using  $\hat{\beta} - \beta_o = O_{\mathbb{P}}(n^{-1/2})$  from **A5(b)** and the Cauchy–Schwartz inequality, it is easy to see that  $nh^{d/2}T_{n32} = o_{\mathbb{P}}(1)$ . Besides, as in theorem 1,  $nh^{d/2}T_{n31} = o_{\mathbb{P}}(1)$ , and so  $nh^{d/2}T_{n3} \xrightarrow{p} \int \Delta^2(\mathbf{x})\omega_d(\mathbf{d}\mathbf{x})$ .

Then, as in the proof of theorem 1, we obtain

$$nh^{d/2} \left( T_n - \frac{b}{nh^d} \right) = nh^{d/2}[T_{n1} - \mathbb{E}(T_{n1})] + nh^{d/2}T_{n2} + nh^{d/2} \left[ \mathbb{E}(T_{n1}) - \frac{b}{nh^d} \right]$$

$$+ nh^{d/2}T_{n3} + 2nh^{d/2}T_{n4}$$

$$= nh^{d/2}T_{n2} + \int \Delta^2(\mathbf{x})\omega_d(\mathbf{d}\mathbf{x}) + o_{\mathbb{P}}(1),$$

which concludes the proof. □

*Proof of theorem 3.* Let  $T_{n,\hat{\beta}}^* = \int (f_n^*(\mathbf{x}) - K_h f_{\hat{\beta}}(\mathbf{x}))^2 \omega_d(\mathbf{d}\mathbf{x})$ , and note that

$$T_n^* - T_{n,\hat{\beta}}^* = \int [K_h (f_{\hat{\beta}}(\mathbf{x}) - f_{\hat{\beta}}^*(\mathbf{x}))]^2 \omega_d(\mathbf{d}\mathbf{x})$$

$$+ 2 \int (f_n^*(\mathbf{x}) - K_h f_{\hat{\beta}}(\mathbf{x})) K_h (f_{\hat{\beta}}(\mathbf{x}) - f_{\hat{\beta}}^*(\mathbf{x})) \omega_d(\mathbf{d}\mathbf{x}).$$

Using arguments analogous to those considered in the proof of theorem 1, when deriving the asymptotic behaviour of  $T_{n3}$  and  $T_{n4}$  and using **A6**, we obtain  $nh^{d/2}[T_n^* - T_{n,\hat{\beta}}^*] = o_{\mathbb{P}^*}(1)$ . Then, it is enough to obtain the asymptotic behaviour of  $T_{n,\hat{\beta}}^*$ . Let

$$H(\mathbf{x}, \mathbf{y}) = c^2(h) \int \left[ K \left( \frac{1 - \mathbf{x}^T \mathbf{z}}{h^2} \right) - m(\mathbf{x}, h) \right] \left[ K \left( \frac{1 - \mathbf{y}^T \mathbf{z}}{h^2} \right) - m(\mathbf{y}, h) \right] \omega_d(\mathbf{d}\mathbf{z}),$$

where  $m(\mathbf{y}, h) = \mathbb{E}^* K((1 - \mathbf{y}^T \mathbf{x}_1^*)/h^2)$  and  $\mathbb{E}^*$  denotes expectation with respect to the bootstrap distribution. Then,  $T_{n,\hat{\beta}}^* = (1/n^2) \sum_{1 \leq i \neq j \leq n} H(\mathbf{x}_i^*, \mathbf{x}_j^*) + (1/n^2) \sum_{i=1}^n H(\mathbf{x}_i^*, \mathbf{x}_i^*)$ .



The term  $T_{n, \hat{\beta}_1}^* = (1/n^2) \sum_{1 \leq i \neq j \leq n} H(\mathbf{x}_i^*, \mathbf{x}_j^*)$  is a quadratic form with  $\mathbb{E}^*(T_{n, \hat{\beta}_1}^*) = 0$ . Then, as in lemma 6 in Zhao & Wu (2001), from **A1** and **A2**, the fact that  $\hat{\beta}$  is a consistent estimator of  $\beta_1$ , and the dominated convergence theorem, we obtain  $\text{var}^*(nh^{d/2}T_{n, \hat{\beta}_1}^*) \xrightarrow{p} 2\sigma_{\beta_1}^2$ . Arguing as in lemma 6 in Zhao & Wu (2001), but conditional on  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , we can see that the conditions of theorem 1 in Hall (1984) are satisfied. Finally, arguments analogous to those considered in lemma 4 in Zhao & Wu (2001) allow us to show that **A1**, **A2**, and **A4** imply that  $nh^{d/2} [\sum_{i=1}^n H(\mathbf{x}_i^*, \mathbf{x}_i^*) / n^2 - (b/(nh^d))] \xrightarrow{p} 0$ .  $\square$