

Minimum distance method for directional data and outlier detection*

Abstract

In this paper, we propose estimators based on the minimum distance for the unknown parameters of a parametric density on the unit sphere. We show that these estimators are consistent and asymptotically normally distributed. Also, we apply our proposal to develop a method that allows us to detect potential atypical values. The behavior under small samples of the proposed estimators is studied using Monte Carlo simulations. Two applications of our procedure are illustrated with real data sets.

Key words and phrases: Directional Data, Robust Estimation, Outlier Detection, Asymptotic Properties.

1 Introduction

Statistics of directional data refer to methodologies and procedures designed to deal with random variables taking values over a d -dimensional space. For instance, in a two-dimensional space, directional observations can be regarded as points on the unit circle (circular data) and in a three-dimensional space as points on the unitary sphere of dimension 2 (spherical data). These types of data arise naturally in a wide range of natural science problems such as Climatology (see Mardia and Jupp (2000)), Neuroscience (see Leong and Carlile (1998)), Biology (see Ferguson (1967)) and Geology (see Embleton and Giddings (1974)), among others. However, in order to handle these problems appropriately, special methods of estimation and inference are necessary. The reason for this is the structure and the complexity of these sample spaces.

Several authors (see, for instance, Fisher et al. (1993), Mardia and Jupp (2000) and the references therein) have discussed estimation methods for spherical and circular data. Exponential models in the context of directional data were considered by Beran (1979). Hering and Genton (2010) worked on the problem of predicting electricity production and showed the advantage of considering the wind direction as a circular variable instead of a linear variable. Cox (1974) studied maximum likelihood estimators for the von Mises distribution and Beran (1979) extended these results to exponential families, which include the von Mises and Bingham distributions. Non-parametric kernel methods for estimating densities of spherical data have been studied in Hall et al. (1987) and Bai et al. (1988).

* *AMS Subject Classification 1990:* Primary 62F35, Secondary 62G05.

Robustness is a key notion in statistics. Maximum likelihood estimators work well if the assumed model is accurate. However, these estimators are badly affected by outliers. The goal of robust statistical methods is to work under the presence of departures from the assumed models. The development of robust procedures for directional data has gained a lot of attention over the last years. The fact that the variables belong to a compact set requires that the problem be addressed in a different way than in the Euclidean case. A survey on robust methods for circular data can be found in He (1992). Ko and Chang (1993) extended M-estimators for location and concentration parameters to the case of the von Mises distribution, while Ko and Guttorp (1988) studied some properties in order to quantify robustness introducing the concept of SB-robustness. The SB-robustness considers a new measure using the influence function and introduces this concept of standardized bias robustness. Recently, Agostinelli (2007) extended the minimum disparity and weighted likelihood methods to circular data, i.e., the special case of $d = 1$.

The aim of this paper is to study a natural extension of the minimum distance estimators introduced in Cao et al. (1995). More precisely, under the assumption that directional variables follow a parametric model, we develop estimators for the unknown parameters. For this purpose, we assume that directional variables have a common density function f that belongs to a parametric class of densities. Then, we measure the distance between a nonparametric density estimator of f adapted to circular data and the parametric family. We define the estimators for the unknown parameters as the parameters that minimize this distance.

A large number of families of parametric densities exist in the context of directional data; some of them are described in Mardia and Jupp (2000). In particular, an important family corresponds to the model with rotational symmetry that generalizes the von Mises distribution. We would like to remark that our proposal can be applied to any parametric family.

As pointed out in Basu and Lindsay (1994), the minimum distance procedures allow a simple parametric adjustment with a trade-off between efficiency and robustness, just as one can do with tuning constants in M-estimation. However as we remark below, unlike M-estimation, the minimum distance methods are applicable to a wide range of models, not just to location or scale models. Therefore, the minimum distance methods provide a robust alternative to take into account when we consider parametric models.

As summary, in this work we propose an extension of the minimum distance estimator in the context of directional data. This estimators require the use of a proper kernel density estimator adapted to this type of data. The proposed estimators allow us to develop a method to outliers detection. We believe that the progress presented in this work constitute an advance when dealing with directional data, extended the technics and results known in the Euclidean case.

The paper is organized as follows. In Section 2, we propose the estimators and analyze their asymptotic properties. In Section 3, we present a procedure based on the proposed estimators to detect anomalous data. In Section 4, we report the results of a Monte Carlo study conducted to evaluate the performance of the proposed estimators. In Section 5, we show how the method developed to detect outliers can be applied to analyze two real data sets. Finally, proofs are relegated to the Appendix.

2 Minimum distance estimators

Let $\mathcal{S}_d \subset \mathbb{R}^{d+1}$ be the d -dimensional unit sphere of radius one centered at 0. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be independent observations of a random variable, \mathbf{x} taking values in \mathcal{S}_d . We denote by f the density function of \mathbf{x} . Notice that $\int_{\mathcal{S}_d} f(\mathbf{x}) \omega_d(d\mathbf{x}) = 1$ where ω_d is the Lebesgue measure on the sphere. Let $f_\theta(\mathbf{x})$ be a family of density functions parametrized with a vector of parameters $\theta \in \mathbb{R}^q$. We are interested in estimating the unknown parameter θ .

Given a distance function in the space of density functions, we define a *minimum distance density-based estimator* θ_n as any value satisfying

$$D(\hat{f}_n, f_{\theta_n}) \leq \inf_{\theta} D(\hat{f}_n, f_{\theta}) + \delta_n, \quad (1)$$

where $\delta_n \rightarrow 0$ as $n \rightarrow \infty$ and \hat{f}_n is a nonparametric density estimator. Notice that if the infimum of $D(\hat{f}_n, f_{\theta})$ is achieved, then the estimator can be defined as follows:

$$\hat{\theta} = \arg \min_{\theta} D(\hat{f}_n, f_{\theta}). \quad (2)$$

For instance, this occurs when θ belongs to a compact set and the family of densities is continuous with respect to the parameter θ . The family of rotationally symmetric densities has this property as can be observed below.

In the context of spherical data, the kernel density estimator needs to be adapted to the structure of the data. In the literature there are several contributions in this direction. Let us mention, for example, the proposals suggested in Bai et al. (1988) and Hall et al. (1987). These proposals are based on the inner products between the sample $\mathbf{x}_1 \dots \mathbf{x}_n$ and \mathbf{x} , but have a slightly different notation. We consider the following estimator,

$$\hat{f}_{h_n}(\mathbf{x}) = \frac{c(h_n)}{n} \sum_{i=1}^n K\left(\frac{1 - \mathbf{x}^T \mathbf{x}_i}{h_n^2}\right), \quad (3)$$

where h_n is the smoothing parameter and $c(h_n)$ is a normalizing constant given by $c(h_n)^{-1} = \int_{\mathcal{S}_d} K((1 - \mathbf{x}^T \mathbf{y})/h_n^2) \omega_d(d\mathbf{y})$.

Some standard choices for D are the distance induced by the L^p norm or the L^∞ norm, i.e., $D_p(\hat{f}_{h_n}, f_{\theta}) = \int |\hat{f}_{h_n}(\mathbf{x}) - f_{\theta}(\mathbf{x})|^p \omega_d(d\mathbf{x})$ and $D_\infty(\hat{f}_{h_n}, f_{\theta}) = \sup_{\mathbf{x}} |\hat{f}_{h_n}(\mathbf{x}) - f_{\theta}(\mathbf{x})|$, respectively. The Hellinger distance $D_H(\hat{f}_{h_n}, f_{\theta}) = \int (\sqrt{\hat{f}_{h_n}(\mathbf{x})} - \sqrt{f_{\theta}(\mathbf{x})})^2 \omega_d(d\mathbf{x})$ could be another choice. This distance was studied in Beran (1979), where the author showed that robustness properties together with efficiency could be obtained using this distance. Another alternative could be to consider D as discrepancy measures or divergences. This approach was considered in Agostinelli (2007) for circular data and in Basu and Lindsay (1994) for Euclidean data, among others.

To compute the estimator, it is necessary to select the smoothing parameter h_n . The choice of the smoothing parameter is a major problems in density estimation. In the linear case, several authors dealt with the problem of providing an automatic procedure to select the bandwidth. For directional data, a cross validation criterion was considered in Hall et al. (1987), while a plug-in bandwidth selector was considered in Taylor (2008). García-Portugués (2013) introduced

three alternatives. One of them assumes that the underlying distribution is von Mises and the other considers a mixture of von Mises densities. However, as Cao et al. (1995) remarked, it seems reasonable to take into account the particular features of the problem and consider another alternative. We consider an automatic bandwidth as in Cao et al. (1995). The idea is to incorporate the smoothing parameter h_n as an additional component in the vector θ . Therefore, we can obtain simultaneously an automatic bandwidth and the estimator of θ as follows:

$$(\tilde{\theta}, \tilde{h}_n) = \arg \min_{\theta, h} D(\hat{f}_h, f_\theta) \quad (4)$$

The proposed estimators, defined in (2), are strongly consistent and asymptotically normally distributed. These properties follow easily using analogous arguments to those considered in Cao et al. (1995). The following assumptions are needed in order to obtain the desired results.

Assumptions:

- A1. The nonparametric estimator satisfies that $\lim_{n \rightarrow \infty} D(\hat{f}_{h_n}, f_\theta) = 0$ a.s. under f_θ .
- A2. For all $\theta_0 \in \Theta$ and a sequence $\theta_m \subset \Theta$ such that $\lim_{m \rightarrow \infty} D(f_\theta, f_{\theta_m}) = D(f_\theta, f_{\theta_0})$, we have that $\lim_{m \rightarrow \infty} \theta_m = \theta_0$.
- A3. The kernel $K : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is a bounded and integrable function with compact support.

For the following hypotheses, we are considering the extension of f to $\mathbb{R}^{d+1} \setminus \{0\}$ given by $f(\mathbf{x}) = f(\mathbf{x}/\|\mathbf{x}\|)$ for all $\mathbf{x} \neq 0$, where $\|\mathbf{x}\|$ denotes the Euclidean norm of \mathbf{x} .

- A4. The density function f_θ is such that (for each $\theta \in \Theta$)
 - (a) $\nabla f_\theta(\mathbf{x}) = (\partial f_\theta(\mathbf{x})/\partial x_1, \dots, \partial f_\theta(\mathbf{x})/\partial x_{d+1})^T$ and $H f_\theta(\mathbf{x}) = (\partial^2 f_\theta(\mathbf{x})/(\partial x_i \partial x_j))_{1 \leq i, j \leq d+1}$ exist and are continuous in $\mathbb{R}^{d+1} \setminus \{0\}$.
 - (b) $\frac{\partial}{\partial \theta} f_\theta(\mathbf{x}) = (\partial f_\theta(\mathbf{x})/\partial \theta_1, \dots, \partial f_\theta(\mathbf{x})/\partial \theta_q)^T$ is integrable with respect to the measure generated by K and f_θ .
- A5. $\Psi(\mathbf{x}, \theta) = \frac{\partial}{\partial \theta} f_\theta(\mathbf{x}) - E_\theta(\frac{\partial}{\partial \theta} f_\theta(\mathbf{x}))$ is a differentiable function with respect to θ in an open neighbourhood of θ_0 .
- A6. If we denote by θ_0 the true value of the parameter, the matrix $A = E_{\theta_0}(\frac{\partial}{\partial \theta} \Psi(\mathbf{x}, \theta)|_{\theta=\theta_0})$ is nonsingular.
- A7. The sequence h_n satisfies $nh_n^4 \rightarrow 0$ as $n \rightarrow \infty$ and $h_n^d c(h_n) \rightarrow \lambda$ as $n \rightarrow \infty$ with $\lambda^{-1} = 2^{d/2-1} \tau_{d-1} \int_0^\infty K(r) r^{d/2-1} dr$, where τ_d is the area of \mathcal{S}_d and $\tau_{d-1} = 2\pi^{d/2}/\Gamma(d/2)$, for $d \geq 1$.

Remark 1.

- a) Assumption A1 was studied in Bai et al. (1988) for some metrics. In particular, if $\int_0^\infty s^{(d-2)/2} K(s) ds < \infty$, $h_n \rightarrow 0$ and $nh_n^d \rightarrow \infty$ as $n \rightarrow \infty$, Bai et al. (1988) showed the strong L^1 -norm consistency, i.e., $D_1(\hat{f}_{h_n}, f_\theta) \xrightarrow{a.s.} 0$. Also, they derived the uniform strong

consistency. More precisely, under the assumptions $h \rightarrow 0$, $\frac{nh^d}{\log n} \rightarrow \infty$ and if the kernel function is bounded on $\mathbb{R}_{>0}$ and satisfies $\int_0^\infty \sup\{K(u) : |\sqrt{u} - \sqrt{v}| < 1\} v^{(k-2)/2} dv < \infty$ they proved that $D_\infty(\hat{f}_{h_n}, f_\theta) \xrightarrow{a.s.} 0$. Besides, using the convergence in L^∞ and L^1 , it is easy to see that $D_2(\hat{f}_{h_n}, f_\theta) \xrightarrow{a.s.} 0$ and $D_H(\hat{f}_{h_n}, f_\theta) \xrightarrow{a.s.} 0$, respectively.

- b) If the parametric density family f_θ has a reasonable parameterization, the assumed A2 is guaranteed. This assumption is related to the purpose of ensuring the identifiability of the parameters. For example, if we consider the distance L^1 , L^2 or L^∞ and the von Mises family, it is easy to see that the assumption is fulfilled.
- c) Assumptions A3 and A7 are standard when we consider kernel estimators in directional data. Note that assumption A3 implies that the kernel must decrease fast unlike the kernel involved in the Euclidean case. Notice that the assumption $nh_n^4 \rightarrow 0$ is related to the bias of the non-parametric estimation (Zhao and Wu (2001)). When $d > 4$, an alternative is to take a kernel with a higher order as in the Euclidean case, obtaining a different bias rate of convergence. However, this interesting topic may be the subject of future research.

Under assumptions A1 and A2, the estimator defined in (2) is strongly consistent, i.e., $\theta_n \xrightarrow{a.s.} \theta$. The proof is the same as the one given in Cao et al. (1995), which is a direct consequence of a general result given in Parr and Schucany (1982). Assumptions A3 to A7 are needed to obtain the asymptotic distribution. We will establish the asymptotic normality for the case that D is the distance induced by the L^2 norm.

Theorem 1. *Under assumptions A3 to A7, the estimator defined in (2) satisfies*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, A^{-1}\Sigma A^{-1})$$

where Ψ and A were defined in A5 and A6, respectively and $\Sigma = E_{\theta_0}(\Psi(\mathbf{x}, \theta_0)\Psi(\mathbf{x}, \theta_0)^T)$.

Remark 2. *It is worth noting that consistency and asymptotic normality results for $\hat{\theta}$ can be derived when considering an automatic bandwidth parameter instead of a fixed one. More precisely, if \hat{h}_n stands for a data-driven bandwidth, for instance, the cross-validation bandwidth defined in Hall et al. (1987), and h_n represents a sequence of real numbers satisfying A7 and $\hat{h}_n/h_n \xrightarrow{P} 1$, then a careful inspection of the proof of Theorem 1 could show that the estimator defined in (2), using \hat{h}_n has asymptotic distribution stated in Theorem 1. Nevertheless, this situation does not include the estimator $\tilde{\theta}$ defined in (4). This interesting situation is beyond the scope of the paper and may be object of future research. A possible strategy in this case could be adapt the argument given in Delecroix et al. (2006) to the directional setting.*

Another property is that the estimator defined in (2) satisfies qualitative robustness. Hampel (1971) introduced this concept with the idea to complement the notion of an influence function with continuity conditions with respect to the Prokhorov distance. A simplified version of his definition is that a sequence of estimators T_n is qualitatively robust at F if for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\pi(\mathcal{L}_F(T_n), \mathcal{L}_G(T_n)) < \varepsilon$ if $\pi(F, G) < \delta$, where π and $\mathcal{L}_F(T_n)$ denote the Prokhorov distance and the distribution of T_n under the distribution F , respectively. The proof that our proposal satisfies this property is the same as that for the linear case studied in Cao et al. (1995).

3 Outlier Detection

One of the first steps towards obtaining a coherent data analysis is the detection of outlying observations. Outliers are often considered as an error or noise and may adversely lead to the parameter estimation introducing biased and incorrect results. The estimators defined in the previous section allow us to introduce a tool to detect when an observation can be considered an atypical point.

A high percentage of the methods to detect outliers rely on the assumption that the data follow an underlying known distribution. We will consider a central assumption that a small number of observations follow a different distribution instead of the target distribution. In the Euclidean case, the target distribution is often taken as a normal distribution. However, when dealing with directional data, we must consider other symmetric and unimodal distributions. The outlier detection problem consists in identifying those observations that lie outside the center of the distribution. The most popular and used distribution on the d -dimensional sphere is the von Mises distribution, and we are going to consider this distribution as the target distribution. Notice that the method can be easily extended to any unimodal symmetric distribution.

We denote the von Mises distribution with parameters μ and κ by $VM(\mu, \kappa)$. Its density f is defined as follows:

$$f(\mathbf{x}) = \left(\frac{\kappa}{2}\right)^{\frac{d-1}{2}} \frac{1}{\Gamma((d+1)/2)I_{(d-1)/2}(\kappa)} e^{\kappa \mu^T \mathbf{x}}$$

where the function Γ represents the Gamma function and I_p is the modified Bessel function of order p , i.e.,

$$I_p(z) = \left(\frac{z}{2}\right)^p \frac{1}{\pi^{1/2} \Gamma(p + \frac{1}{2})} \int_{-1}^1 (1-t^2)^{p-\frac{1}{2}} e^{zt} dt.$$

We wish to explore the distribution of $\|\mathbf{x} - \mu\|^2$. For that purpose, we consider the change of variables given by

$$\begin{aligned} \mathbf{x} &= t\mu + (1-t^2)^{1/2} B_d \xi, \\ \omega_d(d\mathbf{x}) &= (1-t^2)^{d/2-1} dt \omega_{d-1}(d\xi) \end{aligned} \tag{5}$$

where ξ belongs to \mathcal{S}_{d-1} and B_d is a $(d+1) \times d$ semi-orthonormal matrix whose columns join μ to form an orthonormal basis. It is easy to see that the density function of $U = \mu^T \mathbf{x}$ is given by

$$f_U(u) = \frac{\left(\frac{\kappa}{2}\right)^{\frac{d-1}{2}}}{\sqrt{\pi} \Gamma(d/2) I_{\frac{d-1}{2}}(\kappa)} e^{\kappa u} (1-u^2)^{d/2-1}.$$

Therefore, the density of $W = \|\mathbf{x} - \mu\|^2 = 2(1-U)$ is $f_W(w) = f_U((2-w)/2)/2$.

As in Davies and Gather (1993), we can define the α -outlier region as follows. For any confidence coefficient α , $0 \leq \alpha \leq 1$, the α -outlier region with respect to the von Mises distribution is defined by:

$$out(\mu, \alpha) = \{\mathbf{x} : \|\mathbf{x} - \mu\|^2 > v_{1-\alpha}\}, \tag{6}$$

where v_q is the q -quantile corresponding to the density of W . Hence, we consider that an observation \mathbf{x} is an α -outlier with respect to the von Mises distribution if $\mathbf{x} \in out(\mu, \alpha)$.

In order to apply the outlier detection procedure, note that the α -outlier region depends on μ and therefore, we will need to estimate it. It is well known that when we use classical estimators to detect outlier, their effects may interact in such a way that some or all of them remain unnoticed. This is called in the literature masking effect. Therefore, we consider the minimum distance estimator defined in the previous section as a robust estimate in order to provide reliable measures for the recognition of outliers. It is important to observe that the exact distribution of $\|\mathbf{x} - \hat{\mu}\|^2$ differs from the distribution of $\|\mathbf{x} - \mu\|^2$ and it can be difficult to compute. However, we note that $\|\mathbf{x} - \hat{\mu}\|^2 = \|\mathbf{x} - \mu\|^2 + R_n$ where $R_n \xrightarrow{p} 0$. Therefore, the asymptotic distribution of $\|\mathbf{x} - \hat{\mu}\|^2$ converges to the distribution of $\|\mathbf{x} - \mu\|^2$. Thus, the tails of both distributions can be compared.

As we mentioned before, the definition of an α -outlier can be easily extended to any unimodal symmetric distribution, in particular, to the densities that are rotationally symmetric, i.e., their contour lines are $(d-1)$ -spheres orthogonal to a fixed particular direction. This characteristic can be exploited by means of the so called tangent-normal decomposition that leads to the change of variables considered in (5). The rotationally symmetric densities can be parametrized as follows

$$f(\mathbf{x}) = \frac{1}{\omega_{d-1}(1 - (\mu^T \mathbf{x})^2)^{\frac{d}{2}-1}} g_\nu(\mu^T \mathbf{x}) \quad (7)$$

where ω_q is the area of the q -dimensional sphere and g_ν is a density in $(-1, 1)$ with $\nu \in \mathbb{R}^p$. Also, for this family of distributions we obtain that the density function of U reduces to $f_U(u) = g_\nu(u)$. In the case that the target distribution f is a rotationally symmetric density, we will say that \mathbf{x} is an outlier if $\mathbf{x} \in \text{out}(\mu, \alpha)$. Here, $\text{out}(\mu, \alpha)$ is the same region defined in (6) but v_q is now the q -quantile computed with density $g_\nu((2-w)/2)/2$.

4 Simulation study

This section contains the results of a simulation study designed to evaluate the performance of the estimator defined in Section 2. The simulation study was carried out in the R software package. The code is available at the webpage of the second author. We considered three models in dimensions 2 and 3. For the circle (Models 2 and 3) and the sphere (Model 1), we performed $N = 1000$ replications of independent samples of sizes $n = 100$ and $n = 200$, respectively. We denote by C_0 the uncontaminated case and we use C_1, \dots, C_4 to denote different contaminated cases. The models considered are:

Model 1 (on the sphere): We generated a random sample $\mathbf{x}_i \in \mathcal{S}_2$ for $1 \leq i \leq n$ following a von Mises distribution $VM(\mu, \kappa)$ with mean $\mu = (0, 1, 0)$ and concentration parameter $\kappa = 1$ and we denoted it by C_0 . This model had many important applications (see Jammalamadaka and SenGupta (2001) and Mardia and Jupp (2000)). We also considered a contamination setting taking 10% of the data following a $VM((0, -1, 1), 15)$. More precisely, under C_1 the variables $\mathbf{x}_1 \dots \mathbf{x}_n$ have density function given by $0.9f_{(0,1,0),1}(\mathbf{x}) + 0.1f_{(0,-1,1),15}(\mathbf{x})$.

Model 2 (on the circle): We simulated a random sample according to a von Mises distribution $VM((0, 1), 5)$ for the non contaminated case (C_0). For the first contaminated setting (C_1), we considered a 10% of contamination from a $VM((-1, 0), 20)$. Also, we studied the

behavior of our proposal even when the underlying distribution was wrong. For this purpose, we considered the contaminations C_2 , C_3 and C_4 where the data are generated according to the wrapped Cauchy distribution with $\mu = (0, 1)$ and $\rho = 0.6, 0.7$ and 0.9 , respectively.

Model 3 (on the circle): We considered random variables \mathbf{x}_i for $1 \leq i \leq n$ with a rotationally symmetric density function defined in (7) where $g_\nu(t) = g_{a,b}(t) = \frac{1}{2}\tilde{g}_{a,b}\left(\frac{t+1}{2}\right)$ with $\tilde{g}_{a,b}$ the density function of a Beta distribution with parameters a and b . More precisely, $\tilde{g}_{a,b}(t) = \Gamma(a+b)/(\Gamma(a)\Gamma(b))t^{a-1}(1-t)^{b-1}$. We denoted by $f_{a,b,\mu}$ the rotationally symmetric density obtained. The uncontaminated case (C_0) was generated from $f_{0.5,3,(0,1)}$. For the contaminated case, we considered two schemes of contamination. In this case, C_1 corresponds to a 10% of contamination from a $f_{0.5,3,(0,-1)}$ and C_2 is generated with a 10% of contamination from a $VM((0, 1), 5)$.

For the three models, we focused on the parameter μ . For Models 1 and 3, we considered the other parameters κ , as a and b are fixed and known. In Model 2 for the case of the minimum distance estimators, we imputed an estimator of the concentration parameter defined in Ko (1992) for the von Mises distribution, while the competitor estimators were computed simultaneously for the concentration parameter κ and position parameter μ .

In all the cases, for the smoothing procedure, the kernel was taken as the positive part of the Epanechnikov kernel, i.e., $K(t) = 1.5(1-t^2)I(0 < t < 1)$ and we considered an equidistant grid of 10 values of h_n on the interval $[0.2, 1.4]$. Note that $(1 - \mathbf{x}^T \mu)$ is bounded by 2 and therefore, only the bandwidths between 0 and $\sqrt{2}$ are considered. In Models 2 and 3, we also applied the estimator proposed in (4), hence computing automatic bandwidths.

To study the performance of the estimators, we have considered the mean squared error (MSE) and the proportion of outlier classification (PC), i.e.,

$$\text{MSE}(\hat{\mu}) = \frac{1}{N} \sum_{i=1}^N \|\mu - \hat{\mu}_i\|^2 \quad \text{and} \quad \text{PC}(\hat{\mu}) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^n I_{\{\|\mathbf{x}_j - \hat{\mu}_i\|^2 > v_{1-\alpha}\}}$$

where $\hat{\mu}_i$ for $1 \leq i \leq N$ corresponds to the estimator computed in replication i . For all the cases, we consider $\alpha = 0.05$. For Model 1, v_q has a closed expression $v_q = -2 \ln(1 - q(e^k - e^{-k})/e^k) / \kappa$. In the case of the circle ($d = 1$), we performed a numerical integration to obtain the quantiles v_q for each model. In all the Models, we computed the estimator induced by the L^2 distance, denoted in the tables by minimum distance (MD), and the maximum likelihood estimators denoted by ML . In the cases of Model 2 and Model 3, we considered our estimator with distance L^2 (MD), L^1 (MD_{l^1}), and the Hellinger distance (MD_{he}) and we compared our procedure to the maximum likelihood estimator (ML). Moreover, under Model 2, we considered Agostinelli's estimator, which we denoted with MD_{ag} .

The maximum likelihood estimators were computed using the libraries `circular` and `movMF` of R package. The library `circular` for dimension two and the library `movMF` in the case of dimension three, see [3] and [21] for more details. Agostinelli's estimator was calculated using the weighted likelihood estimation (WLE) library, see ([1]).

		<i>MD</i>										<i>ML</i>
		bandwidth										
		0.200	0.333	0.467	0.600	0.733	0.867	1.000	1.133	1.267	1.400	
MSE($\hat{\mu}$)	C_0	0.046	0.039	0.035	0.033	0.031	0.029	0.027	0.025	0.022	0.018	0.031
	C_1	0.052	0.045	0.042	0.040	0.039	0.039	0.041	0.042	0.042	0.038	0.155
PC($\hat{\mu}$)	C_0	0.050	0.050	0.050	0.050	0.049	0.049	0.049	0.049	0.049	0.048	0.052
	C_1	0.141	0.141	0.141	0.141	0.141	0.140	0.140	0.140	0.139	0.140	0.061

Table 4.1: The mean squared error and percentage of outlier classification for the estimators of the mean direction in Model 1. *MD* corresponds to the results for the minimum distance estimator with L^2 distance using different values of bandwidth and *ML* corresponds to the results for the maximum likelihood estimators.

		<i>ML</i>	<i>MD</i>	MD_{l^1}	MD_{he}	MD_{ag}
MSE($\hat{\mu}$)	C_0	0.0024	0.0033	0.0034	0.0024	0.0024
	C_1	0.0193	0.0039	0.0041	0.0108	0.0097
	C_2	0.0082	0.0057	0.0094	0.0072	0.0053
	C_3	0.0053	0.0027	0.0058	0.0036	0.0053
	C_4	0.0011	0.0002	0.0006	0.0003	0.0014
PC($\hat{\mu}$)	C_0	0.0489	0.0616	0.0618	0.0600	0.0656
	C_1	0.0009	0.1232	0.1227	0.1198	0.1188
	C_2	0.0785	0.1460	0.1446	0.1448	0.1158
	C_3	0.0874	0.1857	0.1839	0.1849	0.1354
	C_4	0.0694	0.1857	0.184	0.1849	0.1582
$\widehat{h_n}$	C_0	-	0.2427	0.2368	0.2745	-
	C_1	-	0.3209	0.3050	0.3765	-
	C_2	-	0.4494	0.426	0.4731	-
	C_3	-	0.3229	0.2943	0.3485	-
	C_4	-	0.2789	0.2667	0.2871	-

Table 4.2: The mean squared error and percentage of outlier classification for the estimators of the mean direction in Model 2. *MD*, MD_{l^1} and MD_{he} correspond to the results for the minimum distance estimator using an automatic bandwidth with respect to L^2 , L^1 and the Hellinger distance, respectively. *ML* corresponds to the results for the maximum likelihood estimators and MD_{ag} for Agostinelli's estimator.

		<i>MD</i>									
		bandwidth									
		0.200	0.333	0.467	0.600	0.733	0.867	1.000	1.133	1.267	1.400
MSE($\hat{\mu}_{MD}$)	C_0	0.0034	0.0032	0.0029	0.0026	0.0024	0.0024	0.0024	0.0025	0.0026	0.0026
	C_1	0.0039	0.0039	0.0040	0.0045	0.0060	0.0100	0.0185	0.0304	0.0404	0.0446
PC($\hat{\mu}_{MD}$)	C_0	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602
	C_1	0.1220	0.1220	0.1220	0.1220	0.1220	0.1220	0.1220	0.1220	0.1220	0.1220
MSE($\hat{\mu}_{MD_{l^1}}$)	C_0	0.0033	0.0035	0.0032	0.0027	0.0025	0.0025	0.0025	0.0027	0.0027	0.0027
	C_1	0.0041	0.0040	0.0040	0.0039	0.0053	0.0104	0.0217	0.0397	0.0525	0.0525
PC($\hat{\mu}_{MD_{l^1}}$)	C_0	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602
	C_1	0.1219	0.1219	0.1219	0.1219	0.1219	0.1219	0.1219	0.1219	0.1219	0.1219
MSE($\hat{\mu}_{MD_{he}}$)	C_0	0.0025	0.0025	0.0024	0.0024	0.0024	0.0024	0.0024	0.0024	0.0024	0.0025
	C_1	0.0104	0.0107	0.0110	0.0114	0.0122	0.0138	0.0166	0.0207	0.0254	0.0291
PC($\hat{\mu}_{MD_{he}}$)	C_0	0.0598	0.0598	0.0598	0.0598	0.0598	0.0598	0.0598	0.0598	0.0598	0.0598
	C_1	0.1202	0.1202	0.1202	0.1202	0.1202	0.1202	0.1202	0.1202	0.1202	0.1202

Table 4.3: The mean squared error and percentage of outlier classification for the estimators of the mean direction in Model 2. MD , MD_{l^1} and MD_{he} correspond to the results for the minimum distance estimator using different values of bandwidth with respect to L^2 , L^1 and the Hellinger distance, respectively.

		<i>ML</i>	<i>MD</i>	<i>MD_{l¹}</i>	<i>MD_{he}</i>
MSE($\hat{\mu}$)	C_0	0.0066	0.0097	0.0103	0.0066
	C_1	0.0874	0.0123	0.0128	0.0122
	C_2	0.3882	0.0119	0.0110	0.0085
PC($\hat{\mu}$)	C_0	0.0500	0.0512	0.0505	0.0503
	C_1	0.0426	0.1458	0.1445	0.1447
	C_2	0.0353	0.1456	0.1450	0.1449
\hat{h}_n	C_0	-	0.5146	0.5004	0.4571
	C_1	-	0.4562	0.4157	0.4130
	C_2	-	0.4396	0.4032	0.3654

Table 4.4: The mean squared error and percentage of the outlier classification for the estimators of the mean direction in Model 3. MD , MD_{l^1} and MD_{he} correspond to the results for the minimum distance estimator using an automatic bandwidth with respect to L^2 , L^1 and the Hellinger distance, respectively. ML corresponds to the results for the maximum likelihood estimators.

		<i>MD</i>									
		bandwidth									
		0.200	0.333	0.467	0.600	0.733	0.867	1.000	1.133	1.267	1.400
MSE($\widehat{\mu}$)	C_0	0.0103	0.0101	0.0098	0.0095	0.0091	0.0086	0.0082	0.0077	0.0073	0.0071
	C_1	0.0127	0.0125	0.0123	0.0120	0.0117	0.0114	0.0113	0.0113	0.0117	0.0126
	C_2	0.0117	0.0118	0.0119	0.0121	0.0125	0.0134	0.0151	0.0184	0.0248	0.0394
PC($\widehat{\mu}$)	C_0	0.0507	0.0507	0.0507	0.0507	0.0507	0.0507	0.0507	0.0507	0.0507	0.0507
	C_1	0.1454	0.1454	0.1454	0.1454	0.1454	0.1454	0.1454	0.1454	0.1454	0.1454
	C_2	0.1459	0.1459	0.1459	0.1459	0.1459	0.1459	0.1459	0.1459	0.1459	0.1459
MSE($\widehat{\mu}_{l^1}$)	C_0	0.0093	0.0096	0.0100	0.0108	0.0112	0.0099	0.0087	0.0078	0.0073	0.0070
	C_1	0.0130	0.0132	0.0124	0.0120	0.0122	0.0126	0.0116	0.0113	0.0122	0.0141
	C_2	0.0108	0.0111	0.0117	0.0143	0.0200	0.0248	0.0267	0.0303	0.0389	0.0585
PC($\widehat{\mu}_{l^1}$)	C_0	0.0504	0.0504	0.0504	0.0504	0.0504	0.0504	0.0504	0.0504	0.0504	0.0504
	C_1	0.1450	0.1450	0.1450	0.1450	0.1450	0.1450	0.1450	0.1450	0.1450	0.1450
	C_2	0.1457	0.1457	0.1457	0.1457	0.1457	0.1457	0.1457	0.1457	0.1457	0.1457
MSE($\widehat{\mu}_{he}$)	C_0	0.0067	0.0066	0.0066	0.0068	0.0069	0.0071	0.0073	0.0074	0.0075	0.0075
	C_1	0.0132	0.0125	0.0121	0.0120	0.0118	0.0117	0.0116	0.0115	0.0115	0.0115
	C_2	0.0080	0.0084	0.0089	0.0095	0.0103	0.0113	0.0127	0.0145	0.0172	0.0230
PC($\widehat{\mu}_{he}$)	C_0	0.0501	0.0501	0.0501	0.0501	0.0501	0.0501	0.0501	0.0501	0.0501	0.0501
	C_1	0.1450	0.1450	0.1450	0.1450	0.1450	0.1450	0.1450	0.1450	0.1450	0.1450
	C_2	0.1454	0.1454	0.1454	0.1454	0.1454	0.1454	0.1454	0.1454	0.1454	0.1454

Table 4.5: The mean squared error and percentage of outlier classification for the estimators of the mean direction in Model 3. MD , MD_{l^1} and MD_{he} correspond to the results for the minimum distance estimator using different values of bandwidth with respect to L^2 , L^1 and the Hellinger distance, respectively.

In all the considered models, we can observe that under noncontamination, the minimum distance estimators have a good behavior, similar to that of the maximum likelihood estimator. With the Hellinger distance, the minimum distance estimator has the same behavior as the maximum likelihood estimator. This shows that efficiency is achieved with the Hellinger distance. Note that in Model 1 for the noncontaminated case and large bandwidths, the minimum distance estimators performed better than the maximum likelihood estimators. As we mentioned above, the ML estimators were computed using the R library movMV. This library uses the expectation maximization algorithm to compute the ML estimators and this algorithm does not guarantee that the global maximum will be found. We do not consider that the proposed estimators are more efficient than the ML estimators. We believe that these results are only due to computational limitations.

In the contaminated cases, the performance of the minimum distance estimators is slightly affected; however, the ML estimator increases its MSE considerably. The same behavior can be observed for the classification proportion. For the noncontaminated cases, all the estimators considered around 5% of outliers that correspond to the cut-off point $\alpha = 0.05$. However, under contamination, the MD estimators increase the percentage of outlier classification around 14%. This is attributed to the percentage of contamination introduced in C_1 with $\epsilon = 10\%$. However, the ML estimator detects fewer outliers than the 15% expected.

Finally, note that for the case of contaminations C_2 , C_3 and C_4 in Model 2, the proposed estimators have an acceptable behavior even when all the data are generated from a different

distribution coming from the central model.

5 Real Examples

In this Section, we discuss two examples based on real data sets. In both examples, we consider the Epanechnikov kernel as in the simulation study, an automatic bandwidth computation and the estimators as in (4).

5.1 Frogs' dataset

The first example was studied by Collett (1980) and corresponds to frogs' directions. Ferguson (1967) conducted an experiment to investigate the homing ability of a species of frogs. A total of 14 frogs were collected from the mud flats of an abandoned stream meandering near Indianola, Mississippi, in the United States. After 30 hours, the frogs were released and the directions taken by the frogs were recorded. The results are expressed in radians, where the point 0 corresponds to the north and the circle is oriented clockwise. We analyzed this data set with the von Mises model.

Table 5.1 reports the results for the estimation of the mean direction $\mu = (\cos(\theta), \sin(\theta))$ and the bandwidth obtained in (4). In order to simplify the notation, we summarized the estimators of μ with the corresponding angle θ . We denoted with OC_α the set of the number of observations classified as outliers using an α -outlier region. In order to compute the estimators of μ , we fixed several values of concentration parameters. These results are given in Table 5.1. We also computed the estimator of the concentration parameter defined in Ko (1992). When we estimated the concentration parameter with the proposal by Ko, we obtained $\hat{\kappa} = 2.399$ and with this value we computed the estimator of θ . The result with this procedure was $\hat{\theta} = 2.474752$ and the 14th observation was detected as atypical considering levels $\alpha = 0.01, 0.05$, i.e., $OC_{0.01} = OC_{0.05} = \{14\}$.

In Table 5.1, we observe that for all values of the concentration parameter, the 14th observation was classified as an outlier. This observation corresponds to the point located in the anti-modal direction. Besides, when we increase the concentration parameter, other observations located around the south direction were classified as atypical values. This relation is reasonable and expected, because large values of the concentration parameters imply a distribution with less variability around the mean. The same conclusion regarding the 14th observation was obtained in Collett (1980) and Agostinelli (2007).

Furthermore in the example, we can appreciate the relationship between the concentration parameter and the bandwidth selector. Large values of concentration parameters produce small bandwidths and large bandwidths are related to small values of κ . This trade-off between the concentration parameter and the bandwidth does not allow estimating the parameter κ simultaneously with the position parameter μ .

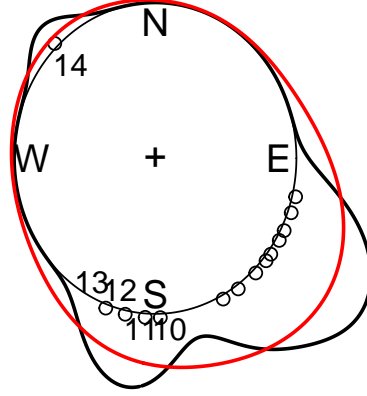


Figure 5.1: Frogs' data set and the nonparametric kernel density estimation using two different smoothing parameters.

κ	1.00	1.74	2.47	3.21	3.95	4.68	5.42	6.16	6.89	7.63
$\hat{\theta}$	2.544	2.514	2.471	2.439	2.408	2.374	2.338	2.304	2.278	2.258
\hat{h}_n	1.113	0.809	0.588	0.522	0.484	0.447	0.409	0.370	0.339	0.313
1% obs. n°		14	14	14	14	14	14	13-14	12-14	12-14
5% obs. n°	14	14	14	14	14	14	12-14	12-14	11-14	10-14
κ	8.37	9.11	9.84	10.58	11.32	12.05	12.79	13.53	14.26	15.00
$\hat{\theta}$	2.244	2.234	2.223	2.218	2.212	2.20	2.205	2.203	2.201	2.196
\hat{h}_n	0.292	0.274	0.252	0.239	0.226	0.217	0.208	0.204	0.199	0.182
$OC_{0.01}$	11-14	11-14	10-14	10-14	10-14	10-14	10-14	10-14	10-14	10-14
$OC_{0.05}$	10-14	10-14	10-14	10-14	10-14	10-14	10-14	10-14	10-14	10-14

Table 5.1: The results for the mean direction and the bandwidth obtained using the minimum distance estimator based on the L^2 distance for different values of the concentration parameters κ . $OC_{0.01}$ and $OC_{0.05}$ denote the set of the number of observations classified as outliers considering confidence coefficients $\alpha = 0.01$ and 0.05 , respectively. The notation $l - m$ corresponds to numbers from l to m .

5.2 Paleozoic data

The second example corresponds to a data set of 17 measurements of magnetic remanence from specimens collected from the lower Paleozoic Tumblagooda Sandstone, Western Australia. These data had been analyzed by Embleton and Giddings (1974) and Ko and Chang (1993). The aim of the study was to determine the origin of natural remanent magnetization in red beds. Fisher et al.

(1993) considered a von Mises distribution and detected two observations that can be considered as atypical values.

If we apply our procedure using a previous estimator of the concentration parameter computed as in Ko (1992), we obtain the following results $\hat{\mu} = (0.3580, 0.7875, -0.5015)$, $\hat{\kappa} = 34.6574$ and $OC_{0.01} = OC_{0.05} = \{5, 6, 7\}$. More precisely, the method applied with confidence coefficients $\alpha = 0.01, 0.05$ detects three potential outliers: observations 5, 6 and 7.

Table 5.2 shows the results obtained with estimators that were also computed considering the von Mises distribution. It is important to note that while a unimodal distribution is a reasonable assumption, Fisher et al. (1993) concluded that the Fisher distribution may not be an appropriate model for the remaining 15 observations. However, under the same assumption, we obtained analogous results to those obtained in Fisher et al. (1993). We also detected another point that could be an atypical observation. This may reinforce the idea that the von Mises model was not appropriate. Figure 5.2 shows the data set and the observations classified as atypical points. Observation numbers 6 and 7 correspond to those also detected also in Fisher et al. (1993).

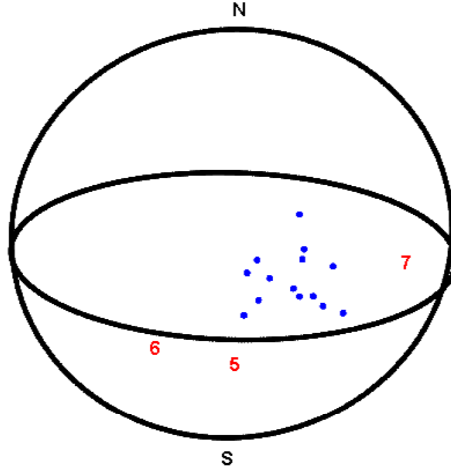


Figure 5.2: The Paleozoic data (blue dots) and the potential outlier (red numbers).

κ	$\hat{\mu}$	1% obs. n°	5%obs. n°
11.0000	(0.3539, 0.7696, -0.5315)		6
12.7143	(0.3517, 0.7748, -0.5253)		6-7
14.4286	(0.3526, 0.7749, -0.5245)		6-7
16.1429	(0.3533, 0.7753, -0.5235)	6	6-7
17.8571	(0.3538, 0.7759, -0.5222)	6	6-7
19.5714	(0.3665, 0.7796, -0.5079)	6	6-7
21.2857	(0.3660, 0.7801, -0.5074)	6	6-7
23.0000	(0.3653, 0.7809, -0.5068)	6-7	5-7
24.7143	(0.3643, 0.7817, -0.5061)	6-7	5-7
26.4286	(0.3632, 0.7827, -0.5055)	6-7	5-7
28.1429	(0.3620, 0.7837, -0.5047)	6-7	5-7
29.8571	(0.3608, 0.7847, -0.5040)	6-7	5-7
31.5714	(0.3597, 0.7858, -0.5032)	6-7	5-7
33.2857	(0.3588, 0.7868, -0.5023)	6-7	5-7
35.0000	(0.3579, 0.7877, -0.5014)	5-7	5-7

Table 5.2: The results for the mean direction using the minimum distance estimator based on the L^2 distance for each value of the concentration parameter κ . $OC_{0.01}$ and $OC_{0.05}$ denote the set of the number of the observations classified as outliers considering a confidence coefficient $\alpha = 0.01$ and 0.05 , respectively. The notation $l - m$ corresponds to numbers l to m .

Appendix

Proof of Theorem 1. Since $\hat{\theta}$ is the minimum of $D(\hat{f}_{h_n}, f_{\hat{\theta}})$, we have that

$$0 = \frac{\partial}{\partial \theta} D(\hat{f}_{h_n}, f_{\theta})|_{\hat{\theta}} = 2 \int (\hat{f}_{h_n}(\mathbf{x}) - f_{\theta}(\mathbf{x})) \frac{\partial}{\partial \theta} f_{\theta}(\mathbf{x}) \omega_d(d\mathbf{x}) \quad (8)$$

$$= 2 \int \hat{f}_{h_n}(\mathbf{x}) \frac{\partial}{\partial \theta} f_{\theta}(\mathbf{x}) \omega_d(d\mathbf{x}) - 2E_{\theta} \left(\frac{\partial}{\partial \theta} f_{\theta}(\mathbf{x}) \right) \quad (9)$$

$$= 2 \frac{1}{n} \sum_{i=1}^n \Psi_h(\mathbf{x}_i, \hat{\theta}) \quad (10)$$

$$(11)$$

where $\Psi_h(\mathbf{t}, \theta) = c(h_n) \int K \left(\frac{1 - \mathbf{x}^T \mathbf{t}}{h_n^2} \right) \frac{\partial}{\partial \theta} f_{\theta}(\mathbf{x}) \omega_d(d\mathbf{x}) - E_{\theta} \left(\frac{\partial}{\partial \theta} f_{\theta}(\mathbf{x}) \right)$. On the other hand,

$$\begin{aligned} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\Psi_h(\mathbf{x}_i, \hat{\theta}) - \Psi(\mathbf{x}_i, \hat{\theta}) \right) \right| &= \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(c(h_n) \int K \left(\frac{1 - \mathbf{x}^T \mathbf{x}_i}{h_n^2} \right) \frac{\partial}{\partial \theta} f_{\theta}(\mathbf{x}) \omega_d(d\mathbf{x}) - \frac{\partial}{\partial \theta} f_{\theta}(\mathbf{x}_i) \right) \right| \\ &\leq \frac{1}{\sqrt{n}} \sum_{i=1}^n \left| c(h_n) \int K \left(\frac{1 - \mathbf{x}^T \mathbf{x}_i}{h_n^2} \right) \frac{\partial}{\partial \theta} f_{\theta}(\mathbf{x}) \omega_d(d\mathbf{x}) - \frac{\partial}{\partial \theta} f_{\theta}(\mathbf{x}_i) \right| \\ &\leq \sqrt{h_n^4 n} \frac{1}{n} \sum_{i=1}^n \left| b\phi \left(\frac{\partial}{\partial \theta} f_{\theta}, \mathbf{x}_i \right) \right| \end{aligned}$$

the last inequality follows from Lemma 1 [30], where

$$b = \frac{\int_0^\infty K(r)r^{d/2}dr}{\int_0^\infty K(r)r^{d/2-1}dr} \quad \text{and} \quad \phi(g, x) = -x^T \nabla g(x) + d^{-1}[\nabla^2 g(x) - x^T (Hg(x))x]$$

with ∇ , ∇^2 and H the gradient, Laplacian and Hessian, respectively are defined in A4. From A7 and (10), we obtain that $\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi(\mathbf{x}_i, \hat{\theta}) \xrightarrow{a.s.} 0$ and the proof follows as in Cao, et al. (1995). \square

References

- [1] Agostinelli, C. (2006). R package ‘WLE’: Weighted Likelihood Estimation. <https://cran.r-project.org/web/packages/wle/>
- [2] Agostinelli, C. (2007). Robust estimation for circular data. *Comput. Stat. Data Anal.*, **51**, 12, 5867–5875.
- [3] Agostinelli, C. and Lund, U. (2013). R package ‘circular’: Circular Statistics. <https://r-forge.r-project.org/projects/circular/>
- [4] Bai, Z.; Rao, C. and Zhao, L. (1988). Kernel estimators of density function of directional data. *J. Multivariate Anal.*, **27**, 24–39.
- [5] Basu, A. and Lindsay, B. (1994). Minimum disparity estimation for continuous models: efficiency, distributions and robustness. *Ann. Inst. Statist. Math.*, **46**, 4, 683–705.
- [6] Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.*, **5**, 445–463.
- [7] Beran, R. (1979). Exponential models for directional data. *Ann. Statist.*, **7**, 1162–1178.
- [8] Cao, R.; Cuevas, A. and Fraiman, R. (1995). Minimum distance density-based estimators. *Comput. Stat. Data Anal.*, **20**, 611–631.
- [9] Collett, D. (1980). Outliers in circular data. *Appl. Stat.*, **29**, 50–57.
- [10] Cox, D. (1974). *Theoretical Statistics*. Chapman & Hall, London.
- [11] Davies, L. and Gather, U. (1993). The identification of multiple outliers. *J. Amer. Statist. Assoc.*, **88**, 782–792.
- [12] Delecroix, M.; Hristache, M. and Patilea, V. (2006). On semiparametric M -estimation in single index regression. *J. Statist. Plann. Inference*, **136**, 730–769.
- [13] Embleton, B. and Giddings, J. (1974). Late Precambrian and lower Palaeozoic palaeomagnetic results from South Australia and Western Australia. *Earth Planet. Sci. Lett.*, **22**, 355–365.
- [14] Ferguson, D. (1967). Sun compass orientation of the northern cricket frog *Acris crepitans*. *Anim. Behav.*, **14**, 45–53.

- [15] Fisher, N.; Lewis, T. and Embleton, B. (1993). *Statistical Analysis of Spherical Data*. Cambridge: University Press.
- [16] García-Portugués, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat.*, **7**, 1655–1685.
- [17] Hall, P.; Watson, G. and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, **74**, 751–762.
- [18] Hampel, F. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.*, **42**, 1887–1896.
- [19] He, X. (1992). *Robust statistics of directional data: a survey*. Nonparametric statistics and related topics. North Holland, New York, 87–95.
- [20] Hering, A. and Genton, M. (2010). Powering up with space-time wind forecasting. *J. Amer. Statist. Assoc.*, **105**, 92–104.
- [21] Hornik, K. and Grün, B. (2014). movMF: An R Package for Fitting Mixtures of von Mises-Fisher Distributions. *J. Stat. Softw.*, **58**, 1–31.
- [22] Jammalamadaka, S. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific, Singapore.
- [23] Ko, D. (1992). Robust estimation of the concentration parameter of the von Mises-Fisher distribution. *Ann. Statist.*, **20**, 917–928.
- [24] Ko, D. and Chang, T. (1993). Robust M -estimators on spheres. *J. Multivariate Anal.*, **45**, 104–136.
- [25] Ko, D. and Guttorp, P. (1988). Robustness of estimators for directional data. *Ann. Statist.*, **16**, 609–618.
- [26] Leong, P. and Carlile, S. (1998). Methods for spherical data analysis and visualization. *J. Neurosci. Meth.*, **80**, 191–200.
- [27] Mardia, K. and Jupp, P. (2000). *Directional Data*, New York: Wiley.
- [28] Parr, W. and Schucany, W. (1982). Minimum distance estimation and components-of-fit statistics. *J. Roy. Statist. Soc. B.*, 178–189.
- [29] Taylor, C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.*, **52**, 3493–3500.
- [30] Zhao, L. and Wu, C. (2001). Central limit theorem for integrated square error of kernel estimators of spherical density. *Sci. China Serie A*, **44**, 474–483.