MATHEMATICA APPLICANDA
Vol. 50(1) 2022, p. 43–63
doi: 10.14708/ma.v50i1.7119

Elena Castilla (Madrid)

Robust Estimation of the Spherical Normal Distribution

Abstract This paper develops a new family of estimators, the minimum density power divergence estimators, for the parameters of the Spherical Normal Distribution. This family contains the maximum likelihood estimator as a particular case. The robustness is empirically illustrated through a Monte Carlo simulation study and two biological numerical examples. Tools needed to implement these methods are also provided.

2010 Mathematics Subject Classification: Primary: 94A15; Secondary: 62F35; 65C05; 62H11.

Key words and phrases: Directional data; Divergence; Robustness; Spherical Normal distribution.

Abbreviations

The following abbreviations are used in this article:

vMF von Misses-Fisher (p. 44)
SN spherical normal (p. 44)
DPD density power divergence (p. 44)
MLE maximum likelihood estimator (p. 47)
RS random search (p. 49)
MAE mean absolute error (p. 50)
MEV mean estaimated value (p. 54)
HL Hodges-Lehmann (p. 57)

1. Introduction. Directional statistics is a major area of interest within the field of statistics on Riemannian manifolds (Bhattacharya and Bhattacharya (2012)). Maybe the most representative example is circular data, which can be represented on the circunference of a unit circle as a point of $\mathbb{S}^1$, as an angle α (measured in radians or degrees), as a unit vector $(\cos\alpha, \sin\alpha)$ of $\mathbb{R}^2$, or even as a complex number with unit modulus $e^{i\alpha}$. Circular data arise in many diverse scientific fields such as medicine (Demir and Bilgin (2019)), biology (Landler et al. (2018)) or political science (Gill and Hangartner (2010)). More generally, many examples in directional statistics involve observations on the unit hypersphere $\mathbb{S}^p$ (Fisher (1993), Agostinelli (2007)).

Note that classical statistics applied to linear data are not valid for directional data because of their geometrical properties. For example, if we have points in circular data, they can be expressed by their angle (in degrees) from 0° to 360°, taking into account that 0° is identical to 360^0. Moreover, we have to determine the starting direction and the direction of rotation. In this paper, we consider "East" and "counterclockwise", respectively. On the other hand, the arithmetic mean between two circular data points according to the linear measure does not coincide with the geometrical mean. All these difficulties make this problem challenging.

Let us consider data distributed on the unit sphere $\boldsymbol{x} \in \mathbb{S}^p$. Note that in the case $p = 1$, we have circular data. The two most important distributions used in literature to describe this type of data are the so called uniform and von Misses-Fisher (vMF) distributions. Uniform distribution assigns equal probabilities to all the points of the sphere

$$f_{\text{Uniform}_{\mathbb{S}^p}}(\boldsymbol{x}) = A_{p-1}^{-1} = \frac{\Gamma(\frac{p}{2})}{2\pi^{p/2}}, \tag{1}$$

where A_{p-1} is the hypervolume or surface area of $\mathbb{S}^{p-1}$ and $\Gamma(\cdot)$ is the standard gamma function. The density function of the vMF distribution (Khatri and Mardia (1977)) is given by

$$f_{\text{vMF}}(\boldsymbol{x}; \boldsymbol{\mu}, \kappa) = C_{p+1}(\kappa) \exp(\kappa \boldsymbol{\mu}^t \boldsymbol{x}), \tag{2}$$

where $\boldsymbol{\mu} \in \mathbb{S}^p$ and $\kappa \in \mathbb{R}^+$ are called the mean direction and concentration parameter, respectively, and $C_p(\kappa)$ is the normalizing constant given by

$$C_p(\kappa) = \frac{\kappa^{p/2-1}}{(2\pi)^{p/2} I_{p/2-1}(\kappa)},$$

with I_ν denoting the modified Bessel function of the first kind at order ν. If $p = 1$ we have the von Misses distribution for circular data (Mardia and Zemroch (1975)). A considerable amount of literature has been published on vMF distribution. See Bangert et al. (2010), among others. In the last years, some extensions or alternatives to vMF distributions have been also developed. See, for example Moghimbeygi and Golalizadeh (2021) for a generalization of the vMF distribution, Kent (1982) for the Fisher-Bingham distribution or, more recently, the Power Spherical distribution (De Cao and Aziz (2020)). In this paper, we consider the Spherical Normal (SN) distribution presented in Hauberg (2018).

Recent developments in the field of directional statistics have led to an increasing interest in robustness of spherical data. See, for example, Ko and Guttorp (1988), Collett (1980), Arellano-Valle et al. (2006), Agostinelli (2007) and Laha and Mahesh (2011). In particular, Kato and Eguchi (2016) used distance-based methods to develop robust estimators for the vMF distribution. Following this idea, we develop here a new family of robust estimators, the minimum density power divergence (DPD) estimators, for the SN distribution. The minimum DPD method for the estimation of the parameters of a distribution is well-known for its robustness (good behaviour in presence of outliers), and it has been successfully applied in several distributions. In this paper, we apply it to the SN distribution, emprically proving its robustness and applicability through an extensive simulation study and two numerical examples.

This paper is organized as follows: Section 2 introduces the SN distribution. In Section 3, we present the minimum DPD method as a generalization of the maximum likelihood method for the estimation of the parameters. Section 4 and Section 5 deal with the numerical results related to this research. Finally, we conclude this paper in Section 6.

2. The SN distribution. Let us consider data distributed on the unit sphere $\boldsymbol{x} \in \mathbb{S}^p$ (i.e., $\boldsymbol{x} \in \mathbb{R}^{p+1}$ and $\|\boldsymbol{x}\|_2 = 1$), and let us consider the geodesic distance:

$$d^2(\boldsymbol{x}, \boldsymbol{y}) = arcos^2(\boldsymbol{x}, \boldsymbol{y}).$$

Definition 2.1 The density function of the SN distribution is given by

$$f_{SN}(\boldsymbol{x}; \boldsymbol{\mu}, \lambda) = \frac{1}{Z_p(\lambda)} \exp\left(-\frac{\lambda}{2} d^2(\boldsymbol{x}, \boldsymbol{\mu})\right),$$

where $\boldsymbol{\mu} \in \mathbb{S}^p$ and $\lambda \in \mathbb{R}^+$ are called the location and concentration parameters, respectively, and $Z_p(\lambda)$ is the normalizing constant given by

$$Z_p(\lambda) = \int_{\mathbb{S}^p} \exp\left(-\frac{\lambda}{2} d^2(\boldsymbol{x}, \boldsymbol{\mu})\right) dx = A_{p-1} \int_0^\pi \exp\left(-\frac{\lambda r^2}{2}\right) (\sin(r))^{p-1} dr,$$

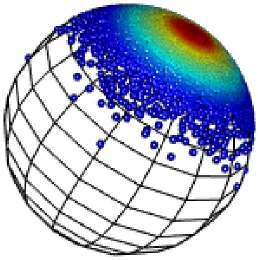
where $A_{p-1} = \frac{2\pi^{p/2}}{\Gamma(\frac{p}{2})}$ was the hypervolume or the surface area of $\mathbb{S}^{p-1}$.

Note that the SN distribution is an instance of the Riemannian normal distribution (Pennec (2006)) that uses as the distance measure between two points the arc-length of the shortest connecting curve on the sphere. This distribution was formally presented in Hauberg (2018) and deeply studied in You (2021) and You and Suh (2022). This last author developed the `Riemann` package in the `R` statistical software. Through the function `rspnorm()`, we can generate data from the SN distribution.

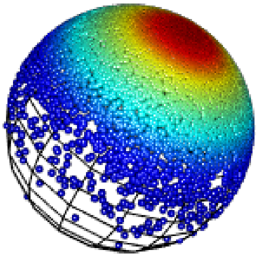
The parameter λ measures somehow the inverse of the variance of the data, so a large λ leads to concentrated mass near $\boldsymbol{\mu}$ and a small λ shows greater dispersion of the mass. In particular, it can be seen that when the concentration goes towards zero, the SN leads to the uniform distribution in (1), i.e.:

$$\lim_{\lambda \to 0^+} \text{SN}(\boldsymbol{\mu}, \lambda) = \text{Uniform}_{\mathbb{S}^p}.$$

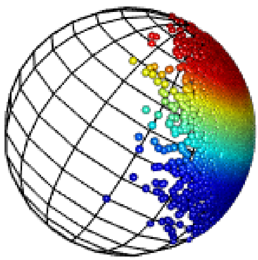
In Figure 1, we present the SN distribution over a sphere of dimension $\mathbb{S}^2$ for different parameter values. Note that the spheres are slightly rotated for the sake of illustration.



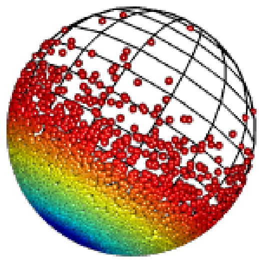
(a) $\boldsymbol{\mu} = (0, 0, 1)^T$, $\lambda = 10$.



(b) $\boldsymbol{\mu} = (0, 0, 1)^T$, $\lambda = 3$.



(c) $\boldsymbol{\mu} = (0, 0.8, 0.6)^T$, $\lambda = 10$.



(d) $\boldsymbol{\mu} = (0, 0, -1)^T$, $\lambda = 3$.

Figure 1: SN distribution in $\mathbb{S}^2$ for different parameter values.

One must not confound the term "spherical normal distribution" with that used in literature to refer to a multivariate normal distribution which components are mutually independent, unit normal random variables. A very nice illustrative explanation of the use of the word "spherical" here is given in Pratt et al. (1995) (Section 22.1). On the other hand, the determination of the probability content of geometrically well-defined regions in Euclidean N-space with this underlying distribution was discussed by Ruben (1960a, 1960b, 1961, 1962) and by Guenther and Terragno (1964), among others.

3. Estimation of parameters

3.1. Maximum Likelihood estimation. The log-likelihood function is given by

$$\ell(\boldsymbol{\mu}, \lambda; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \log f_{SN}(\boldsymbol{x}_i; \boldsymbol{\mu}, \lambda) \qquad (3)$$
$$= -\frac{\lambda}{2} \sum_{i=1}^{n} d^2(\boldsymbol{x}_i, \boldsymbol{y}) - n \log Z_p(\lambda).$$

Note that the second part of (3) does not depend on the concentration parameter λ. Therefore, we can give the following definition:

Definition 3.1 Let us consider the SN distribution. The maximum likelihood estimator (MLE) of $\boldsymbol{\mu}$, $\widehat{\boldsymbol{\mu}}_{\text{MLE}}$, is given by

$$\widehat{\boldsymbol{\mu}}_{\text{MLE}} = \underset{\boldsymbol{\mu} \in \mathbb{S}^p}{\text{argmax}} - \frac{\lambda}{2} \sum_{i=1}^{n} d^2(\boldsymbol{x}_i, \boldsymbol{y}) = \underset{\boldsymbol{\mu} \in \mathbb{S}^p}{\text{argmin}} \sum_{i=1}^{n} d^2(\boldsymbol{x}_i, \boldsymbol{y}).$$

Once $\widehat{\boldsymbol{\mu}}_{\text{MLE}}$ is computed, solving $\widehat{\lambda}_{\text{MLE}}$ reduces to

$$\widehat{\lambda}_{\text{MLE}} = \underset{\lambda \in \mathbb{R}^+}{\text{argmin}} \frac{\lambda}{2} \sum_{i=1}^{n} d^2(\boldsymbol{x}_i, \boldsymbol{y}) + n \log Z_p(\lambda).$$

Hauberg (2018) proposed the steepest descent algorithm to estimate $\boldsymbol{\mu}$. As noted there, it can also be computed in an online fashion by repeated geodesic interpolation (Salehian et al. (2015)). On the other hand, You (2021) and You and Suh (2022) made a comparison between Newton's and Halley's method to estimate λ. Function `mle.spnorm()` in the cited `Riemann` package on `R` allows us to compute these estimators.

3.2. Minimum DPD estimation. Let $f_{\boldsymbol{\theta}}$ be a parametric density with $\boldsymbol{\theta} \in \Theta$, and g the density underlying the data. The DPD between g and $f_{\boldsymbol{\theta}}$ is given by

$$d_\beta(g, f_{\boldsymbol{\theta}}) = \int \left\{ \frac{1}{\beta(1+\beta)} g^{1+\beta} - \frac{1}{\beta} g f_{\boldsymbol{\theta}}^{\beta}(x) + \frac{1}{1+\beta} f_{\boldsymbol{\theta}}^{1+\beta}(x) \right\} dx, \quad \beta > 0,$$
$$d_{\beta=0}(g, f_{\boldsymbol{\theta}}) = \lim_{\beta \to 0} d_\beta(g, f_{\boldsymbol{\theta}}) = \int g \log(g / f_{\boldsymbol{\theta}}(x)) dx.$$

The minimum DPD estimator is given by

$$\widehat{\boldsymbol{\theta}}_\beta = \underset{\boldsymbol{\theta} \in \Theta}{\text{argmin}}\ d_\beta(g, f_{\boldsymbol{\theta}}).$$

In the particular case of $\beta = 0$ it can be shown that it coincides with the MLE, $\widehat{\boldsymbol{\theta}}_{\text{MLE}}$. The minimum DPD estimator was originally presented in the work of Basu et al. (1998) and, since then, it has been applied in many statistical areas to develop robust estimation procedures.

Let consider here the SN distribution $f_{SN}(\boldsymbol{x}; \boldsymbol{\mu}, \lambda)$. Here, our parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\mu}, \lambda)^T$ and, for $\beta > 0$:

$$d_\beta(g, f_{SN}) = \int_{\mathbb{S}^p} \left\{ \frac{1}{\beta(1+\beta)} g^{1+\beta} - \frac{1}{\beta} g f_{SN}^{\beta}(\boldsymbol{x}; \boldsymbol{\mu}, \lambda) + \frac{1}{1+\beta} f_{SN}^{1+\beta}(\boldsymbol{x}; \boldsymbol{\mu}, \lambda) \right\}.$$

But the term $\int_{\mathbb{S}^p} g^{1+\beta}$ has no role in the minimization in $(\boldsymbol{\mu}, \lambda)$ of $d_\beta(g, f_{SN})$. Thus, if we want to obtain the minimum DPD estimator, we must minimize

$$\frac{1}{1+\beta} \int_{\mathbb{S}^p} f_{SN}^{1+\beta}(\boldsymbol{x}; \boldsymbol{\mu}, \lambda) d\boldsymbol{x} - \frac{1}{\beta} \int_{\mathbb{S}^p} f_{SN}^{\beta}(\boldsymbol{x}; \boldsymbol{\mu}, \lambda) dG(\boldsymbol{x}).$$

We can estimate the second integral using the empirical distribution function based on a random sample of size n; $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$; i.e, we must minimize, for $\beta > 0$

$$\frac{1}{1+\beta} \int_{\mathbb{S}^p} f_{SN}^{1+\beta}(\boldsymbol{x}; \boldsymbol{\mu}, \lambda) dx - \frac{1}{\beta} \frac{1}{n} \sum_{i=1}^{n} f_{SN}^{\beta}(\boldsymbol{x}_i; \boldsymbol{\mu}, \lambda).$$

But

$$\begin{aligned}
f_{SN}^{1+\beta}(\boldsymbol{x}; \boldsymbol{\mu}, \lambda) &= \left[\frac{1}{Z_p(\lambda)} \exp\left(-\frac{\lambda}{2} d^2(\boldsymbol{x}, \boldsymbol{\mu}) \right) \right]^{1+\beta} \\
&= \frac{1}{[Z_p(\lambda)]^{1+\beta}} \exp\left(-\frac{\lambda(1+\beta)}{2} d^2(\boldsymbol{x}, \boldsymbol{\mu}) \right) \\
&= \frac{Z_p(\lambda(1+\beta))}{[Z_p(\lambda)]^{1+\beta}} \frac{1}{Z_p(\lambda(1+\beta))} \exp\left(-\frac{\lambda(1+\beta)}{2} d^2(\boldsymbol{x}, \boldsymbol{\mu}) \right) \\
&= \frac{Z_p(\lambda(1+\beta))}{[Z_p(\lambda)]^{1+\beta}} f_{SN}(\boldsymbol{x}; \boldsymbol{\mu}, (1+\beta)\lambda),
\end{aligned}$$

so

$$\int_{\mathbb{S}^p} f_{SN}^{1+\beta}(\boldsymbol{x}; \boldsymbol{\mu}, \lambda) d\boldsymbol{x} = \frac{Z_p(\lambda(1+\beta))}{[Z_p(\lambda)]^{1+\beta}} \int_{\mathbb{S}^p} f_{SN}(\boldsymbol{x}; \boldsymbol{\mu}, (1+\beta)\lambda) d\boldsymbol{x} = \frac{Z_p(\lambda(1+\beta))}{[Z_p(\lambda)]^{1+\beta}}.$$

Therefore, we can give the following definition

DEFINITION 3.2 Let us consider the SN distribution. The minimum DPD estimator of $(\boldsymbol{\mu}, \lambda)$, $(\widehat{\boldsymbol{\mu}}_\beta, \widehat{\lambda}_\beta)$ is given by

$$(\widehat{\boldsymbol{\mu}}_\beta, \widehat{\lambda}_\beta) = \operatorname*{argmin}_{\substack{\boldsymbol{\mu}\in\mathbb{S}^p \\ \lambda\in\mathbb{R}^+}} \left\{ \frac{1}{1+\beta} \frac{Z_p(\lambda(1+\beta))}{[Z_p(\lambda)]^{1+\beta}} - \frac{1}{\beta}\frac{1}{n}\sum_{i=1}^{n} f_{SN}^{\beta}(\boldsymbol{x}_i; \boldsymbol{\mu}, \lambda) \right\}. \quad (4)$$

If $\boldsymbol{\mu}$ is known, solving $\widehat{\lambda}_\beta$ reduces to

$$\widehat{\lambda}_\beta = \operatorname*{argmin}_{\lambda\in\mathbb{R}^+} \left\{ \frac{1}{1+\beta} \frac{Z_p(\lambda(1+\beta))}{[Z_p(\lambda)]^{1+\beta}} - \frac{1}{\beta}\frac{1}{n}\sum_{i=1}^{n} f_{SN}^{\beta}(\boldsymbol{x}_i; \boldsymbol{\mu}, \lambda) \right\}. \quad (5)$$

While estimating the concentration parameter (5) reduces to the classical optimization problem on $\mathbb{R}^+$ (we can use, for example, `optimize()` function in R), estimating $\boldsymbol{\mu}$ presents the additional difficult of restricting the parametric space to $\mathbb{S}^p$. This could done through a penalty method. The idea is to approximate the constrained problem ($\|\boldsymbol{\mu}\|_2 = 1$) with an unconstrained one and then apply standard search techniques to obtain solution. This approximation is done by adding a term to the objective function that prescribes a high cost for violation of the constraints, see Bertsekas (1976).

Another option is to use a Random Search (RS) method. RS algorithms are very common in Machine Learning and are particularly useful when the objective function is not continuous or differentiable (see, for example, Bergstra and Bengio (2012)). Starting in a random initial point on the parametric space, RS methods may iteratively move to better positions in the search-space. A simple version applied to our problem in (4) can be the following:

`STEP 1:` Initialize $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_{\text{MLE}}$ and $\lambda_0 = \lambda_{\text{MLE}}$.

`STEP 2:` For i from 1 to N do:

Sample a new position $\boldsymbol{\mu}_1$ in $\mathbb{S}^p$. If

$$\sum_{i=1}^{n} f_{SN}^{\beta}(\boldsymbol{x}_i; \boldsymbol{\mu}_0, \lambda_0) < \sum_{i=1}^{n} f_{SN}^{\beta}(\boldsymbol{x}_i; \boldsymbol{\mu}_1, \lambda_0)$$

then $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$ and update λ_0 by (5).

`STEP 3:` Return $(\widehat{\boldsymbol{\mu}}_\beta, \widehat{\lambda}_\beta) = (\boldsymbol{\mu}_0, \lambda_0)$.

One of the main drawbacks of this approximation method is that we may need a very large value of N to guarantee a precise solution, overall when p is large.

The asymptotic normality of the estimator could be shown from the M-estimation theory. See Kato and Eguchi (2016) for an equivalent argument in the vMF distribution.

4. Monte Carlo simulation study. We develop here a simulation study to illustrate the robustness of the proposed methods. This study is developed in the `R` statistical software. The following packages are used: `Riemann`, `circular` (Lund et al. (2017)) and `Rfast` (Papadakis et al. (2021)).

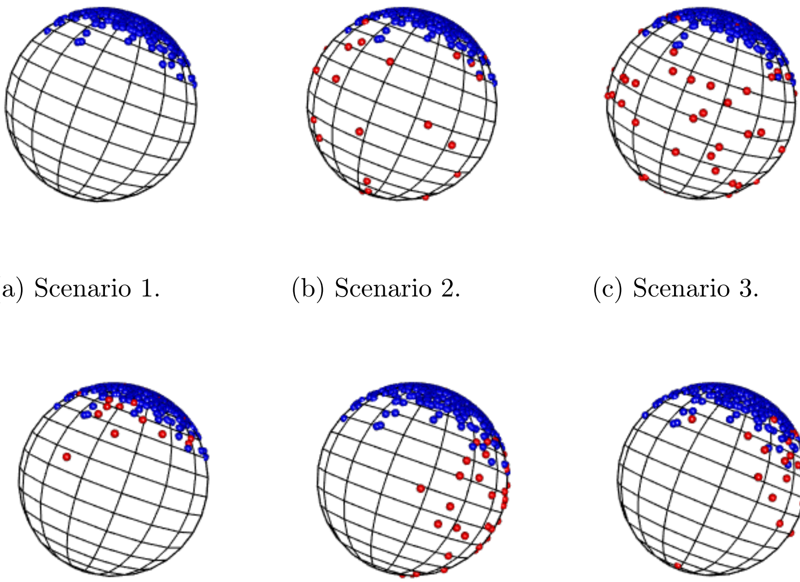
4.1. Known $\boldsymbol{\mu}$, unknown λ. Let us first assume that the location parameter $\boldsymbol{\mu}$ is known in a sphere of dimension 2, $\mathbb{S}^2$. For different sample sizes, $n \in \{20, 60, 100, \ldots, 500\}$, we simulate data from the following six scenarios (see Figure 2):

- S.1: Pure data on the form $\mathrm{SN}(\boldsymbol{\mu}, \lambda)$, with $\boldsymbol{\mu} = (0, 0, 1)^T$ and $\lambda = 10$.
- S.2: Contaminated data on the form $(1-\varepsilon)\mathrm{SN}(\boldsymbol{\mu}, \lambda) + \varepsilon \mathrm{Uniform}_{\mathbb{S}^2}$, with $\varepsilon = 0.05$.
- S.3: Contaminated data on the form $(1-\varepsilon)\mathrm{SN}(\boldsymbol{\mu}, \lambda) + \varepsilon \mathrm{Uniform}_{\mathbb{S}^2}$, with $\varepsilon = 0.10$.
- S.4: Contaminated data on the form $(1-\varepsilon)\mathrm{SN}(\boldsymbol{\mu}, \lambda) + \varepsilon \mathrm{SN}(\boldsymbol{\mu}, \tilde{\lambda})$, with $\tilde{\lambda} = 3$ and $\varepsilon = 0.05$.
- S.5: Contaminated data on the form $(1-\varepsilon)\mathrm{SN}(\boldsymbol{\mu}, \lambda) + \varepsilon \mathrm{SN}(\tilde{\boldsymbol{\mu}}, \tilde{\lambda})$, with $\tilde{\boldsymbol{\mu}} = (0, 1, 0)^T$, $\tilde{\lambda} = 3$ and $\varepsilon = 0.05$.
- S.6: Contaminated data on the form $(1-\varepsilon)\mathrm{SN}(\boldsymbol{\mu}, \lambda) + \varepsilon \mathrm{vMF}(\tilde{\boldsymbol{\mu}}, k)$, with $\tilde{\boldsymbol{\mu}} = (0, 0.8, 0.6)^T$, $k = 3$ and $\varepsilon = 0.05$.

For each scenario, we compute the minimum DPD estimator of λ for $\beta \in \{0, 0.2, 0.4, 0.6\}$ (note that $\beta = 0$ corresponds to the MLE). The mean absolute error (MAE) is then computed by

$$\mathrm{MAE}(\boldsymbol{\beta}) = \frac{1}{S} \sum_{s=1}^{S} \left| \widehat{\lambda}_{\beta}^{(s)} - \lambda \right|,$$

where $S = 1,000$ is the number of samples used in the simulation. Results are presented in Figure 3. When considering a pure scenario, MLE slightly outperforms minimum DPD estimators. On the other hand, in all the alternative contamination scenarios, minimum DPD scenarios present a much more robust behaviour than classical MLE. Only in Scenario 4, this difference is not so extreme, as the contaminating distribution only differs in the dispersion parameter. However, minimum DPD estimators also outperform MLE in this case.



(a) Scenario 1. (b) Scenario 2. (c) Scenario 3.

(d) Scenario 4. (e) Scenario 5. (f) Scenario 6.

Figure 2: Scenarios in the simulation study for known $\boldsymbol{\mu}$, unknown λ. Blue points represent pure data and red points represent outliers. Data simulated with $n = 500$.

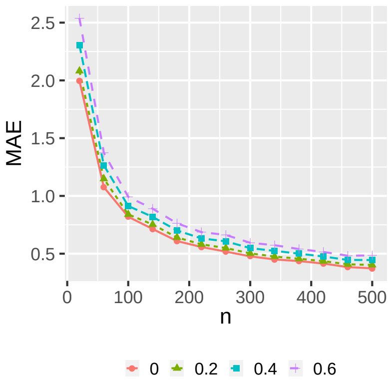
We next study how the degree of contamination on the parameter λ affects to its estimation. Let us consider the contamination scenario $(1-\varepsilon)\text{SN}(\boldsymbol{\mu}, \lambda)+\varepsilon\text{SN}(\boldsymbol{\mu}, \tilde{\lambda})$, for $\varepsilon \in \{0.05, 0.1\}$ and $n \in \{60, 100, 200\}$. The degree of contamination of λ is measured as

$$\tau = \frac{\lambda - \tilde{\lambda}}{\lambda},$$

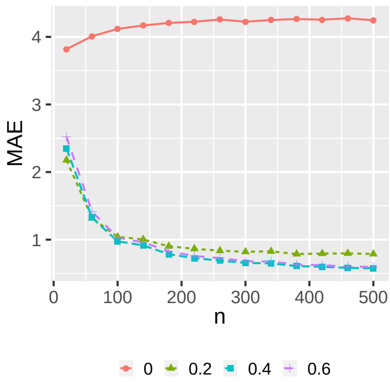
for $\tilde{\lambda} \in \{10, 9, \ldots, 1\}$. This way, when $\tilde{\lambda} = \lambda = 10$, $\tau = 0$ and we are considering a pure scheme. As greater is the value of τ, stronger is this contamination. Results are presented in Figure 4, illustrating again the robustness for the proposed estimators.

4.2. Unknown $\boldsymbol{\mu}$, unknown λ. Let us now consider the case in which both location and dispersion parameters are unknown. Let us consider $\boldsymbol{\mu} = (1, 0)^T$ and $\lambda = 10$. To illustrate the behaviour of the proposed estimators we consider two contamination scenarios in $\mathbb{S}^1$ (see Figure 2).

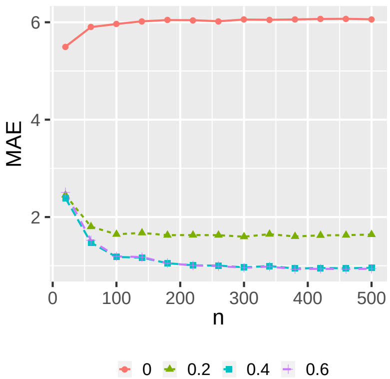
S.1: Contaminated data on the form $(1-\varepsilon)\text{SN}(\boldsymbol{\mu}, \lambda) + \varepsilon\text{SN}(\tilde{\boldsymbol{\mu}}, \tilde{\lambda})$, with $\tilde{\boldsymbol{\mu}} = (-0.6, 0.8)^T$, $\tilde{\lambda} = 12$ and $\varepsilon = 0.05$.



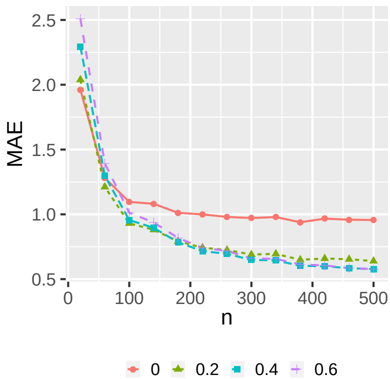
(a) Scenario 1.



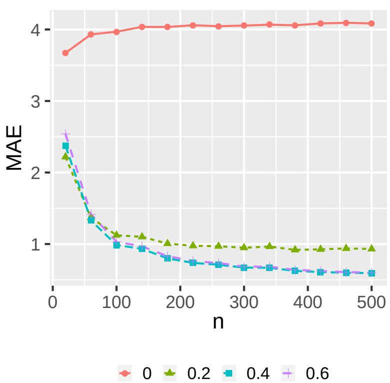
(b) Scenario 2.



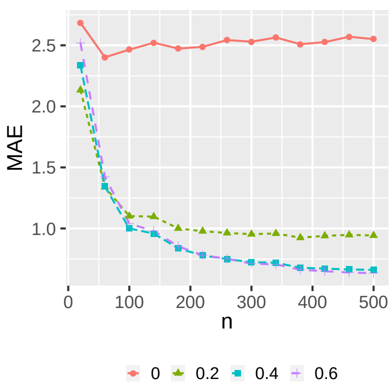
(c) Scenario 3.



(d) Scenario 4.

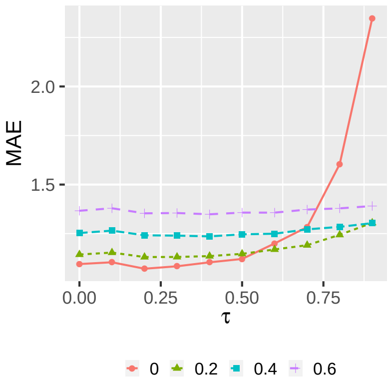


(e) Scenario 5.

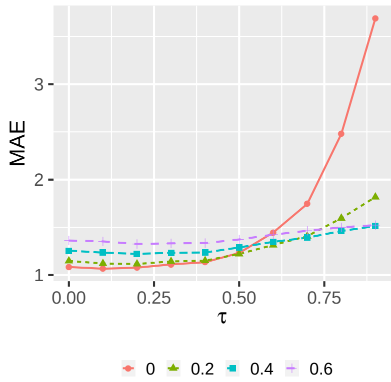


(f) Scenario 6.

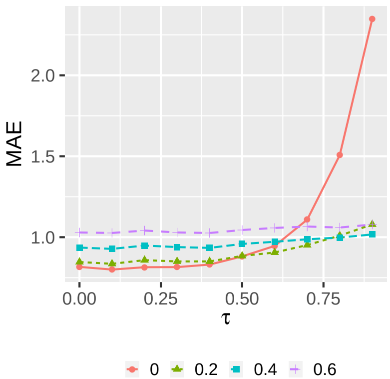
Figure 3: Scenarios in the simulation study for known $\boldsymbol{\mu}$, unknown λ. Mean absolute errors (MAEs).



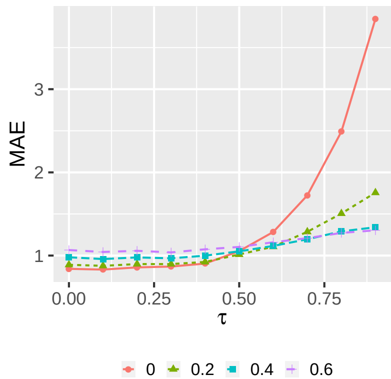
(a) $n = 60$, $\varepsilon = 0.05$.



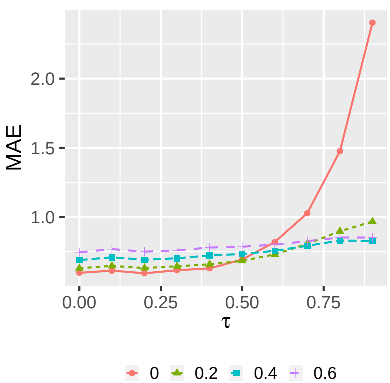
(b) $n = 60$, $\varepsilon = 0.10$.



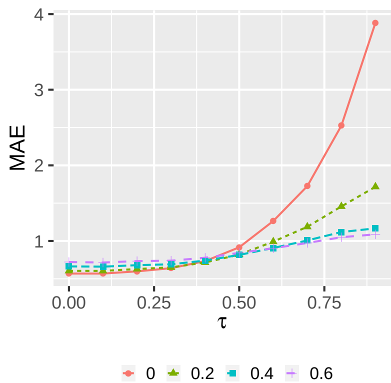
(c) $n = 100$, $\varepsilon = 0.05$.



(d) $n = 100$, $\varepsilon = 0.10$.



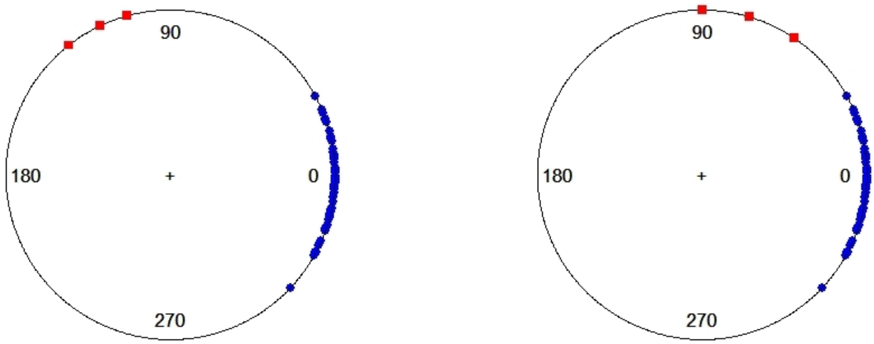
(e) $n = 200$, $\varepsilon = 0.05$.



(f) $n = 200$, $\varepsilon = 0.10$.

Figure 4: Scenarios in the simulation study for known $\boldsymbol{\mu}$, unknown λ. Mean absolute errors (MAEs).

S.2: Contaminated data on the form $(1-\varepsilon)\text{SN}(\boldsymbol{\mu}, \lambda) + \varepsilon\text{SN}(\tilde{\boldsymbol{\mu}}, \tilde{\lambda})$, with $\tilde{\boldsymbol{\mu}} = (1, 0)^T$, $\tilde{\lambda} = 3$ and $\varepsilon = 0.05$.



(a) Scenario 1. (b) Scenario 2.

Figure 5: Scenarios in the simulation study for unknown $\boldsymbol{\mu}$, unknown λ. Blue points represent pure data and red squares represent outliers. Data simulated with $n = 60$.

For each scenario, we compute the minimum DPD estimators of $\boldsymbol{\mu}$ and λ for $\beta \in \{0, 0.2, 0.4, 0.6\}$ and $n \in \{60, 80, 100\}$ with the RS algorithm based on $N = 10^4$ points. Mean estimated values (MEVs) and MAEs are then computed and results are presented in Table 1 and Table 2, respectively. The effect of the contamination on λ is again very clear. However, the effect on the estimation of $\boldsymbol{\mu}$ is not so clear. In fact, although this estimation is affected with the contamination, it remains very near to the true value. These results are in concordance with previous literature. See Kato and Eguchi (2016).

5. Numerical examples. Let us illustrate the robust behaviour of the proposed estimators through the study of two classical data sets in $\mathbb{S}^1$.

Example 5.1 (**Sardinian sea stars**) These data, presented in Fisher (1993) represent the directions of 22 Sardinian sea stars 11 days after being displaced from their natural habitat. These directions are presented (in degrees) in the left of Figure 6 and they can be treated as points in $\mathbb{S}^1$. As it is seen in the plot, there is one observation (147°) surprisingly remote from the others. This point has been shown to be a real outlier in Fisher (1993) and Kato and Eguchi (2016), among others.

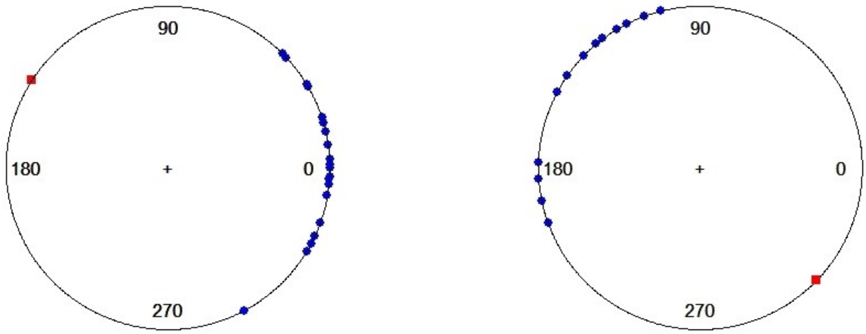
Table 1: Scenarios in the simulation study for unknown $\boldsymbol{\mu}$, unknown λ. MEVs.

		Scenario 1				Scenario 2			
Param.	Value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
$n = 60$									
λ	10	3.27371	11.75964	11.40701	10.89353	3.78810	11.29220	11.37552	10.88994
μ_1	1	0.99949	0.99880	0.99728	0.99728	0.99960	0.99880	0.99728	0.99728
μ_2	0	-0.03127	-0.04862	-0.07333	-0.07331	-0.02749	-0.04862	-0.07333	-0.07331
$n = 80$									
λ	10	3.01664	10.02490	10.32019	10.08790	5.45611	8.03600	9.66097	9.90902
μ_1	1	0.99999	0.99990	0.99853	0.99853	0.99996	0.99990	0.99880	0.99853
μ_2	0	0.00149	-0.01404	-0.05411	-0.05411	0.00822	-0.01404	-0.04872	-0.05411
$n = 100$									
λ	10	2.83623	10.06420	10.22507	10.01820	4.29350	8.75831	9.87675	9.91226
μ_1	1	0.99996	0.99999	0.99933	0.99880	0.99983	0.99997	0.99933	0.99880
μ_2	0	0.00873	-0.00334	-0.03650	-0.04872	0.01810	-0.00672	-0.03650	-0.04872

Table 2: Scenarios in the simulation study for unknown $\boldsymbol{\mu}$, unknown λ. MAEs.

		Scenario 1				Scenario 2			
Param.	Value	0	0.2	0.4	0.6	0	0.2	0.4	0.6
$n = 60$									
λ	10	6.72629	1.75964	1.40701	0.89353	6.21190	1.29820	1.37552	0.88994
μ_1	1	0.00051	0.00120	0.00272	0.00272	0.00040	0.00120	0.00272	0.00272
μ_2	0	0.03167	0.04899	0.07366	0.07368	0.02793	0.04899	0.07366	0.07368
$n = 80$									
λ	10	6.98336	0.02988	0.32356	0.09125	4.54389	1.96400	0.33903	0.09098
μ_1	1	0.00001	0.00010	0.00147	0.00147	0.00004	0.00010	0.00120	0.00147
μ_2	0	0.00149	0.01429	0.05427	0.05427	0.00822	0.01429	0.04889	0.05427
$n = 100$									
λ	10	7.16377	0.06690	0.22630	0.01961	5.70650	1.24169	0.12325	0.08774
μ_1	1	0.00004	0.00001	0.00067	0.00120	0.00017	0.00003	0.00067	0.00120
μ_2	0	0.00873	0.00359	0.03667	0.04889	0.01810	0.00697	0.03667	0.04889

Example 5.2 (**Northern cricket frogs**) These data contain the orientation of 14 northen cricket frogs after 30 hours enclosure within a dark environmental chamber. Originally appeared in Ferguson et al. (1967) to investigate the homing ability of this type of frog, these data were presented in Collett (1980) to illustrate outliers in circular data. Taking the north as 0°, the orientations are presented in the right of Figure 6. As noted by Collett (1980), there is one observation (316°) which clearly represents an outlier.



(a) Sardinian sea stars dataset. (b) Northern cricket frogs dataset.

Figure 6: Examples over $\mathbb{S}^1$ (in degrees). Blue points represent regular observations and red squares represent possible outliers.

MLEs for the parameters in the full data and for the data excluding the potential outlier, are obtained jointly with minimum DPD estimators for different values of the tuning parameter β. Results are presented in Table 3. When comparing MLE for the full datasets and for the datasets excluding the one sample, the value of the location parameter $\boldsymbol{\mu}$ is not highly affected, while the estimation of the concentration parameter λ clearly differs. These results are in concordance with these discussed in Section 4.2 and illustrate how outliers do not influence the estimation of $\boldsymbol{\mu}$. The minimum DPD estimates for moderate tuning parameter are more similar to the MLE for datasets without outliers, implying their robustness, overall for the estimation of the concentration parameter λ.

5.1. Choice of the optimal tuning parameter As derived from the simulation study, the robustness of the minimum DPD estimator directly depends on the chosen tuning parameter β. As a general advice, it seems that a moderate choice of the tuning parameter would lead to a suitable balance between robustness and efficiency. However, a data-driven choice of β would be more helpful in practice. Several methods for that purpose have been discussed in literature. The method proposed by Warwick and Jones (2005) consists of minimizing the estimated mean squared error. However, this method needs of the asymptotic variance-covariance matrix of the corresponding estimators, which would require heavy computations. Further discussions can be found in Sugasawa and Yonekura (2021) and in Castilla and Chocano (2023), among others.

A possible approach may be that of choosing the tuning parameter which minimizes the difference between the estimated mean angle and an empirical estimate of the angle. For definition and discussion of circular median see Fisher (1993) . It can be computed in `R` through the function `median.circular()` in the package `circular`. Another interesting measure to consider is the Hodges-Lehmann (HL) median for circular data, proposed by Otieno (2002). It can be implemented in `R` through the function `medianHL.circular()`. The obtained medians and HL-medians are, respectively, −1.14 and 2.59 (sea stars data) and 137.00 and 137.00 (frogs data). Therefore, for the frogs data, $\beta = 1$ would be chosen as the optimal tuning parameter, while for the sea stars data, $\beta = 1$ or $\beta = 0.2$ would be chosen as the optimal tuning parameter, depending the median statistic considered. This is in concordance with the presence of outliers previously noted.

Table 3: Sea Stars and Frogs data. MLE estimator for the full data (MLE) for the data without the potential outlier observation (MLE$^{(1)}$) and minimum DPD estimators for the full data and different tuning parameter β. Here $\widehat{\alpha}$ represents the estimated angle (in degrees).

Sardinian sea stars dataset					Northern cricket frogs dataset				
Estimator	$\widehat{\lambda}$	$\widehat{\mu}_1$	$\widehat{\mu}_2$	$\widehat{\alpha}$	Estimator	$\widehat{\lambda}$	$\widehat{\mu}_1$	$\widehat{\mu}_2$	$\widehat{\alpha}$
MLE	2.117	0.9985	0.0541	3.10	MLE	1.089	-0.8288	0.5596	145.97
MLE$^{(1)}$	5.175	0.9997	0.0232	1.33	MLE$^{(1)}$	3.337	-0.8200	0.5724	145.08
β =0.1	3.064	0.9962	0.0867	4.97	β =0.1	1.405	-0.8995	0.4370	154.09
β =0.2	4.689	0.9989	0.0462	2.65	β =0.2	2.002	-0.8631	0.5051	149.66
β =0.3	5.224	0.9994	0.0357	2.04	β =0.3	2.533	-0.8256	0.5642	145.65
β =0.4	5.381	0.9993	0.0373	2.14	β =0.4	2.615	-0.8036	0.5952	143.47
β =0.5	5.467	0.9993	0.0386	2.21	β =0.5	2.580	-0.7921	0.6104	142.38
β =0.6	5.536	0.9992	0.0389	2.23	β =0.6	2.531	-0.7825	0.6226	141.49
β =0.7	5.599	0.9993	0.0386	2.21	β =0.7	2.491	-0.7733	0.6340	140.65
β =0.8	5.659	0.9993	0.0373	2.14	β =0.8	2.468	-0.7634	0.6459	139.77
β =0.9	5.719	0.9994	0.0358	2.05	β =0.9	2.466	-0.7527	0.6583	138.83
β =1	5.782	0.9994	0.0338	1.94	β =1	2.490	-0.7410	0.6715	137.82

6. Concluding remarks In this paper we have developed a new family of distance-based estimators, the minimum DPD estimators, for the parameters of the SN distribution, introduced by Hauberg (2018). This way, we apply a very well-known robust method for the estimation of the parameters to a distribution on the sphere. This family of estimators depends on a tuning parameter $\beta \geq 0$. When $\beta = 0$, we have the classical MLE. It is empirically shown that increasing β leads to more robust estimation, with an unavoidable loss of efficiency. This effect is seen on the estimation of the concentration parameter λ, which is more affected by the presence of outliers.

This study is limited to the definition and application of the new family of estimators. Further study should show the asymptotical behaviour of the proposed estimators. On the other hand, further research should be undertaken to explore how to choose the optimal tuning parameter. Additionally, we hope to develop an appropriate `R` package in future for a wider range of practitioners to enhance the practical applications of the proposed methodologies.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgments: The author would like to thank Professor Szajowski for his continuous help and support during the publication process. The comments made by an anonymous reviewer are also greatly appreciated. Finally, the author thanks the invaluable support recieved by Professor P.J. Chocano.

7. References

[1] C. Agostinelli. Robust estimation for circular data. *Comput. Statist. Data Anal.*, 51(12):5867–5875, 2007. ISSN 0167-9473. doi: 10.1016/j.csda.2006.11.002. MR 2407683. Cited on pp. 43 and 44.

[2] R. B. Arellano-Valle, G. del Pino, and P. Iglesias. Bayesian inference in spherical linear models: robustness and conjugate analysis. *J. Multivariate Anal.*, 97(1):179–197, 2006. ISSN 0047-259X. doi: 10.1016/j.jmva.2004.12.002. MR 2208848. Cited on p. 44.

[3] M. Bangert, P. Hennig, and U. Oelfke. Using an infinite von mises-fisher mixture model to cluster treatment beam directions in external radiation therapy. In S. Draghici, T. M. Khoshgoftaar, V. Palade, W. Pedrycz, M. A. Wani, and X. Zhu, editors, *ICMLA*, pages 746–751. IEEE Computer Society, 2010. ISBN 978-0-7695-4300-0. doi: 10.1109/ICMLA.2010.114. Cited on p. 44.

[4] A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85 (3):549–559, 1998. ISSN 0006-3444. doi: 10.1093/biomet/85.3.549. MR 1665873. Cited on p. 48.

[5] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, 13:281–305, 2012. ISSN 1532-4435. MR 2913701. Cited on p. 49.

[6] D. P. Bertsekas. On penalty and multiplier methods for constrained minimization. *SIAM J. Control Optim.*, 14(2):216–235, 1976. ISSN 0363-0129. doi: 10.1137/0314017. MR 408830. Cited on p. 49.

[7] A. Bhattacharya and R. Bhattacharya. *Nonparametric inference on manifolds with applications to shape spaces.*, volume 2 of *Institute of Mathematical Statistics (IMS) Monographs*. Cambridge University Press, Cambridge, 2012. ISBN 978-1-107-01958-4. doi: 10.1017/CBO9781139094764. MR 2934285. Cited on p. 43.

[8] E. Castilla and P. Chocano. *On the choice of the optimal tuning parameter in robust one-shot device testing analysis.*, volume 445 of *Trends in Mathematical, Information and Data Sciences. Studies in Systems, Decision and Control*, pages 169–180. Springer, Cham, 2023. ISBN 978-3-031-04136-5. doi: 10.1007/978-3-031-04137-2_16. Cited on p. 57.

[9] D. Collett. Outliers in circular data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(1):50–57, 1980. doi: 10.2307/2346410. Cited on pp. 44 and 56.

[10] N. De Cao and W. Aziz. The power spherical distribution. *arXiv preprint arXiv:2006.04437*, 2020. Cited on p. 44.

[11] Y. Demir and Ö. C. Bilgin. Application of circular statistics to life science data. *Medical Science and Discovery*, 6(3):63–72, 2019. ISSN 2148-6832. doi: 10.17546/msd. Cited on p. 43.

[12] D. E. Ferguson, H. F. Landreth, and J. P. Mckeown. Sun compass orientation of the northern cricket frog, acris crepitans. *Animal Behaviour*, 15(1):45–53, 1967. doi: 0.1016/S0003-3472(67)80009-5. Cited on p. 56.

[13] N. I. Fisher. *Statistical analysis of circular data.* Cambridge University Press, Cambridge, 1993. ISBN 0-521-35018-2; 0-521-56890-0. doi: 10.1017/CBO9780511564345. MR 1251957. Cited on pp. 43, 54, and 57.

[14] J. Gill and D. Hangartner. Circular data in political science and how to handle it. *Political Analysis*, 18(3):316–336, 2010. doi: 10.1093/pan/mpq00. Cited on p. 43.

[15] W. C. Guenther and P. J. Terragno. A review of the literature on a class of coverage problems. *Ann. Math. Statist.*, 35:232–260, 1964. ISSN 0003-4851. doi: 10.1214/aoms/1177703747. MR 0159406. Cited on p. 46.

[16] S. Hauberg. Directional statistics with the spherical normal distribution. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 704–711. IEEE, 2018. doi: 10.23919/ICIF.2018.8455242. Cited on pp. 44, 45, 47, and 58.

[17] S. Kato and S. Eguchi. Robust estimation of location and concentration parameters for the von Mises–Fisher distribution. *Statist. Papers*, 57(1): 205–234, 2016. ISSN 0932-5026. doi: 10.1007/s00362-014-0648-9. MR 3461956. Cited on pp. 44, 50, and 54.

[18] J. T. Kent. The Fisher-Bingham distribution on the sphere. *J. Roy. Statist. Soc. Ser. B*, 44(1):71–80, 1982. ISSN 0035-9246. URL `http://links.jstor.org/sici?sici=0035-9246(1982)44:1<71:TFDOTS>2.0.CO;2-W&origin=MSN`. MR 0655376. Cited on p. 44.

[19] C. G. Khatri and K. V. Mardia. The von Mises-Fisher matrix distribution in orientation statistics. *J. Roy. Statist. Soc. Ser. B*, 39(1):95–106, 1977. ISSN 0035-9246. doi: 10.1111/j.2517-6161.1977.tb01610.x. MR 0494687. Cited on p. 44.

[20] D. Ko and P. Guttorp. Robustness of estimators for directional data. *Ann. Statist.*, 16(2):609–618, 1988. ISSN 0090-5364. doi: 10.1214/aos/1176350822. MR 0947564. Cited on p. 44.

[21] A. K. Laha and K. C. Mahesh. SB-robustness of directional mean for circular distributions. *J. Statist. Plann. Inference*, 141(3):1269–1276, 2011. ISSN 0378-3758. doi: 10.1016/j.jspi.2010.09.027. MR 2739166. Cited on p. 44.

[22] L. Landler, G. D. Ruxton, and E. P. Malkemper. Circular data in biology: advice for effectively implementing statistical procedures. *Behavioral ecology and sociobiology*, 72(8):1–10, 2018. doi: 10.1007/s00265-018-2538-y. Cited on p. 43.

[23] U. Lund, C. Agostinelli, and M. C. Agostinelli. Package 'circular'. *Repository CRAN*, pages 1–142, 2017. URL `https://cran.r-project.org/web/packages/circular/index.html`. Cited on p. 50.

[24] K. Mardia and P. Zemroch. Algorithm as 86: The von mises distribution function. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 24(2):268–272, 1975. doi: 10.2307/2346578. Cited on p. 44.

[25] M. Moghimbeygi and M. Golalizadeh. A new extension of von Mises–Fisher distribution. *Hacet. J. Math. Stat.*, 50(6):1838–1854, 2021. ISSN 1303-5010. doi: 10.15672/hujms.788296. MR 4351505. Cited on p. 44.

[26] B. S. Otieno. *An alternative estimate of preferred direction for circular data*. ProQuest LLC, Ann Arbor, MI, 2002. ISBN 978-0496-90178-4. URL http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqdiss&rft_dat=xri:pqdiss:3157778. Thesis (Ph.D.)–Virginia Polytechnic Institute and State University. Cited on p. 57.

[27] M. Papadakis, M. Tsagris, M. Dimitriadis, S. Fafalios, I. Tsamardinos, M. Fasiolo, G. Borboudakis, J. Burkardt, C. Zou, K. Lakiotaki, and C. Chatzipantsiou. A Collection of Efficient and Extremely Fast R Functions, 2021. URL https://cran.r-project.org/web/packages/Rfast/index.html. Package 'Rfast' Version 2.0.6. Manos Papadakis<rfastofficial@gmail.com>–Maintainer. Date 2022-02-15. Cited on p. 50.

[28] X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006. doi: 10.1007/s10851-006-6228-4. Cited on p. 45.

[29] J. W. Pratt, H. Raiffa, and R. Schlaifer. *Introduction to statistical decision theory*. MIT Press, Cambridge, MA, 1995. ISBN 0-262-16144-3. Cited on p. 46.

[30] H. Ruben. Probability content of regions under spherical normal distributions. I. *Ann. Math. Statist.*, 31:598–618, 1960a. ISSN 0003-4851. doi: 10.1214/aoms/1177705788. MR 0117828. Cited on p. 46.

[31] H. Ruben. Probability content of regions under spherical normal distributions. II. The distribution of the range in normal samples. *Ann. Math. Statist.*, 31:1113–1121, 1960b. ISSN 0003-4851. doi: 10.1214/aoms/1177705683. MR 0125670. Cited on p. 46.

[32] H. Ruben. Probability content of regions under spherical normal distributions. III. The bivariate normal integral. *Ann. Math. Statist.*, 32: 171–186, 1961. ISSN 0003-4851. doi: 10.1214/aoms/1177705149. MR 0117830. Cited on p. 46.

[33] H. Ruben. Probability content of regions under spherical normal distributions. IV. The distribution of homogeneous and non-homogeneous quadratic functions of normal variables. *Ann. Math. Statist.*, 33:542–570, 1962. ISSN 0003-4851. doi: 10.1214/aoms/1177704580. MR 0137189. Cited on p. 46.

[34] H. Salehian, R. Chakraborty, E. Ofori, D. Vaillancourt, and B. C. Vemuri. An efficient recursive estimator of the fréchet mean on a hypersphere

with applications to medical image analysis. *Mathematical Foundations of Computational Anatomy*, 3:143–154, 2015. Cited on p. 47.

[35] S. Sugasawa and S. Yonekura. On selection criteria for the tuning parameter in robust divergence. *Entropy*, 23(9):Paper No. 1147, 10, 2021. doi: 10.3390/e23091147. MR 4319509. Cited on p. 57.

[36] J. Warwick and M. C. Jones. Choosing a robustness tuning parameter. *J. Stat. Comput. Simul.*, 75(7):581–588, 2005. ISSN 0094-9655. doi: 10.1080/00949650412331299120. MR 2162547. Cited on p. 57.

[37] K. You. *Topics in Geometric and Topological Data Analysis*. ProQuest LLC, Ann Arbor, MI, 2021. ISBN 979-8534-69161-0. Thesis (Ph.D.)–University of Notre Dame. MR 4297631. Cited on pp. 45 and 47.

[38] K. You and C. Suh. Parameter estimation and model-based clustering with spherical normal distribution on the unit hypersphere. *Comput. Statist. Data Anal.*, 171:Paper No. 107457, 2022. ISSN 0167-9473. doi: 10.1016/j.csda.2022.107457. MR 4393851. Cited on pp. 45 and 47.

Odporna estymacja parametrów sferycznego rozkładu normalnego.

Elena Castilla

Streszczenie Artykuł przedstawia nową rodzinę estymatorów parametrów sferycznego rozkładu normalnego minimalnej dywergencji. Ta rodzina obejmuje estymator największej wiarygodności jako przypadek szczególny. Odporność tych estymatorów jest zilustrowana empirycznie przez badanie symulacyjne Monte Carlo. Zamieszczone przykłady dla danych rzeczywistych dotyczą zagadnień z biologii. Pokazano również narzędzia potrzebne do wdrożenia tych metod.

Klasyfikacja tematyczna AMS (2010): 94A15; 62F35; 65C05; 62H11.

Słowa kluczowe: Dane kierunkowe; Rozbieżność; odporność; sferyczny rozkład normalny.

Elena Castilla was born in 1993 in Madrid (Spain). She got her M.Sc. and Bachelor Degree in Mathematics and Statistics at Complutense University of Madrid. In 2021, she received Ph.D. degree for the work *Robust Statistical Inference for One-shot devices based on Divergences*, also at Complutense University of Madrid. Currently, she is an Assistant Professor at Rey Juan Carlos University. She is also a member of the Instituto de Matemática Interdisciplinar (IMI). References to her research papers are listed in the European Mathematical Society, FIZ Karlsruhe, and the Heidelberg Academy of Sciences bibliography database known as zbMath under ZBL ai:castilla.elena, in MathSciNet under ID:1256155, Scopus under AU-ID:57200608540 and Web of Science WOS ResearcherID:AAA-7913-2020.

Elena Castilla
Universidad Rey Juan Carlos
Department of Applied Mathematics
Calle Tulipán. 28933 Móstoles, Madrid, Spain
E-mail: elena.castilla@urjc.es
URL: https://sites.google.com/view/elenacastillagonzlez

Communicated by: Roman Zmyślony

(Received: 9th of February 2022; revised: 1st of June 2022)