

# Circular and spherical projected Cauchy distributions

Michail Tsagris and Omar Alzeley

Department of Economics, University of Crete,  
Gallos Campus, Rethymnon, Greece

[mtsagris@uoc.gr](mailto:mtsagris@uoc.gr)

Department of Mathematics, Umm Al-Qura University,  
Al-Qunfudah University College, Saudi Arabia

[oazeley@uqu.edu.sa](mailto:oazeley@uqu.edu.sa)

February 5, 2023

## Abstract

Two new distributions are proposed: the circular projected and the spherical projected Cauchy distributions. A special case of the circular projected Cauchy coincides with the wrapped Cauchy distribution, and for this, a generalization is suggested that offers better fit via the inclusion of an extra parameter. For the spherical case, by imposing two conditions on the scatter matrix we end up with an elliptically symmetric distribution. All distributions allow for a closed-form normalizing constant and straightforward random values generation, while their parameters can be estimated via maximum likelihood. The bias of the estimated parameters is assessed via numerical studies, while exhibitions using real data compare them further to some existing models indicating better fits.

**Keywords:** Directional data, Cauchy distribution, projected distribution

## 1 Introduction

Multivariate data whose norm equals unity is termed directional data, and whose sample space can be expressed as

$$\mathbb{S}^{d-1} = \left\{ \mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\| = 1 \right\},$$

where  $\|\cdot\|$  denotes the Euclidean norm. When  $d = 2$  the data lie on a circle and thus are referred to as circular, while when  $d = 3$ , the data lie on a sphere, and are hence referred to as spherical.

The literature describes plenty of circular distributions, with the oldest being the von Mises [von Mises \(1918\)](#) and studied by [\(Mardia, 1972, Mardia and Jupp, 2000\)](#). This distribution arises as the conditional distribution of a bivariate normal random vector with some mean vector and identity covariance matrix, given that the vector lies on the unit sphere. Various generalizations of this distribution have been proposed over the years [\(Dietrich and Richter, 2017, Gatto and Jammalamadaka, 2007, Kim and SenGupta, 2013\)](#), and more circular distributions have been proposed by [Pewsey \(2000\)](#), [Jones and Pewsey \(2005\)](#), [Abe and Pewsey \(2011\)](#), and [Jones and Pewsey \(2012\)](#). On the contrary, the wrapped distributions are another type of distributions that arises from wrapping a univariate random vector on the circle. These include the wrapped  $t$  family of distributions [\(Pewsey et al., 2007\)](#), the wrapped stable family [\(Pewsey, 2008\)](#), the wrapped normal, and wrapped Cauchy (WC) distributions

(Mardia and Jupp, 2000) and extensions (Kato and Jones, 2010, 2013). Finally, another type of distribution somewhat less studied is the so-called projected distributions that arise as the distribution of a multivariate random vector projected onto a circle, where the projected normal (Mardia and Jupp, 2000, Presnell et al., 1998, Watson, 1983) is probably the unique distribution of this type.

In an attempt to offer increased flexibility or better capture of skewed data, the list of the aforementioned distributions may include distributions that entail more than two parameters. However, the majority suffer from the problem induced by assuming a diagonal scatter matrix. The projected normal is an exception, which has addressed this issue by adopting the Bayesian stance (Nuñez-Antonio and Gutiérrez-Peña, 2005, Wang and Gelfand, 2013).

Spherical (and hyper-spherical) distributions have been proposed over the years, with the oldest being the von Mises-Fisher distribution (Fisher, 1953), followed by the projected normal, also called offset normal (Mardia, 1972) or displaced normal (Kendall, 1974). This distribution was studied by Watson (1983) who showed its close proximity to the von Mises-Fisher distribution. The drawback of these two distributions is that they assume rotational symmetry, that is they are obtained via a multivariate normal (via the aforementioned manners) with an identity covariance matrix. It was this drawback that motivated the work of Kent (1982), who proposed the first elliptical symmetric distribution, a special case of the Fisher-Bingham distribution (Mardia, 1972, 1975). Paine et al. (2018) proposed a special case of the projected normal distribution, yielding the second elliptical symmetric distribution. Finally, Scealy and Wood (2019) proposed the Scaled von Mises-Fisher family of distributions that entail the same symmetry properties as the two previous distributions.

Moving along the lines of the projected types of distributions, we first propose the circular projected Cauchy distribution. To the best of our knowledge, this is the second projection type of distribution, apart from the normal distribution. We first derive the general circular projected Cauchy distribution that arises by projecting the bivariate Cauchy distribution (with some scatter matrix) onto the circle. To avoid over-identifiability issues we take two special cases; first, we consider the case of independent components, i.e., a diagonal scatter matrix, and we show its equivalence to the WC distribution. Since the diagonal scatter matrix is a rather strict assumption we relax this assumption by imposing one restriction, moving along the lines of Paine et al. (2018), a mean-constrained scatter matrix.

We then move on to the spherical case, and propose the spherical projected Cauchy distribution. As in the circular case, the scatter matrix may be equal to the identity matrix, yielding rotational symmetry. Imposing the same conditions as in Paine et al. (2018), yields an elliptically symmetric distribution. This feature is highly important with regard to spherical data, given that the literature describes only a few distributions with this property.

A remark that can be made about either circular or spherical projected Cauchy distributions is that they have a closed form normalizing constant, a feature that is not the norm with directional distributions. Another advantage is that simulation from the projected Cauchy family is straightforward.

The proposed circular and spherical projected Cauchy distributions are presented in Sections 2 and 3, respectively. In order to compare the performance of the different forms of the proposed projected Cauchy distributions, with both each other and with other distributions, we perform simulation studies that focus on the bias of the estimated parameters in Section 4. We illustrate the performance of the proposed distributions and of some competing distributions using examples with real data in Section 5. Finally, Section 6 concludes the paper.

## 2 Circular projected Cauchy distribution and its special cases

The probability density function of the random variable  $X$  projected onto a circle/sphere/hypersphere is given by  $\mathbf{Y} = \frac{\mathbf{X}}{r}$ , where  $r = \|\mathbf{X}\|$ . The marginal distribution of  $\mathbf{Y}$  is obtained via integrating out the  $r$  over the

positive line, thus ending up with

$$f(\mathbf{y}) = \int_0^\infty r^{d-1} f(r\mathbf{y}) dr. \quad (1)$$

The probability density function of the bivariate Cauchy distribution is given by

$$f(\mathbf{x}) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \left[ 1 + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^{-3/2}. \quad (2)$$

By substituting (2) into (1) and calculating the integral we end up with a new distribution on the circle, termed the circular projected Cauchy (CPC) distribution.

$$\begin{aligned} f(\mathbf{y}) &= \int_0^\infty \frac{r}{2\pi|\boldsymbol{\Sigma}|^{1/2}} (1 + r^2 \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} - 2r \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})^{-3/2} dr \\ &= \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2} (B\sqrt{\Gamma^2 + 1} - A\sqrt{B})}, \end{aligned} \quad (3)$$

where  $A = \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ ,  $B = \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}$  and  $\Gamma^2 = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ . We highlight that  $\mathbf{y} \in \mathbb{S}^1$ , whereas  $\boldsymbol{\mu} \in \mathbb{R}^2$ .

## 2.1 The WC as a special case of the CPC distribution

The difficulty with CPC is the excessive number of parameters leading to over-identifiability issues during the estimation process. To avoid such issues, we assume that  $\boldsymbol{\Sigma} = \mathbf{I}_2$  leading to the following representation

$$f(\mathbf{y}) = \frac{1}{2\pi (\sqrt{\gamma^2 + 1} - \alpha)}, \quad (4)$$

where  $\alpha = \mathbf{y}^\top \boldsymbol{\mu}$  and  $\gamma = \|\boldsymbol{\mu}\|$ . The density in (4) can also be written as

$$f(\theta) = \frac{1}{2\pi (\sqrt{\gamma^2 + 1} - \gamma \cos(\theta - \omega))} \quad (5)$$

since  $\alpha = \gamma \cos(\theta - \omega)$ , where  $\omega$  denotes the mean. When  $\gamma = 0$ , the distribution reduces to the circular uniform, whose density function is given by  $f(\theta) = (2\pi)^{-1}$ . We will denote the distribution whose density is given in 4 by circular independent projected Cauchy (CIPC) distribution.

Consider now a random variable  $X$  on the real line which we wrap around the circumference of a circle of unit radius by  $\theta = X \pmod{2\pi}$ . If  $X$  follows the Cauchy distribution, then  $\theta$  follows the WC distribution on the circle, and its density function is given by (Mardia and Jupp, 2000)

$$f(\theta) = \frac{1 - \rho^2}{2\pi [1 + \rho^2 - 2\rho \cos(\theta - \omega)]}, \quad (6)$$

where  $\omega$  denotes the mean and  $\rho \in [0, 1)$ . The density in (4) is another parameterization of the WC distribution<sup>1</sup> in (6), where  $\sqrt{\gamma^2 + 1} = \frac{1+\rho^2}{1-\rho^2}$  and  $\gamma = \frac{2\rho}{1-\rho^2}$ , or reverse  $\rho = (\sqrt{\gamma^2 + 1} - 1)/\gamma$ .

Jones and Pewsey (2005) mentioned that the wrapped Cauchy is the only member of their suggested family of symmetric unimodal distributions that arise by projection onto, rather than conditioning on, the unit circle. Specifically, as pointed out in (Mardia and Jupp, 2000), if  $\theta$  follows the angular central Gaussian distribution, then  $2\theta$  follows the WC. They also stated that "... the WC distribution does not arise from projecting a bivariate spherically symmetric distribution with non-zero mean onto the circle.". According to our previous results, this statement is not entirely accurate. Further, the fact that we derived the WC via a different route enables us to generalize the WC distribution in a straightforward manner, as described in the section below.

<sup>1</sup>This equivalence is not observed for the wrapped normal and the projected normal (Presnell et al., 1998) whatsoever.

## 2.2 The generalized circular projected Cauchy distribution

The usual problem with circular distributions is that in the two-dimensional space, they impose no correlation between the components. However, we will employ one of the conditions imposed in [Paine et al. \(2018\)](#), that is,  $\Sigma\boldsymbol{\mu} = \boldsymbol{\mu}$ , but not  $|\Sigma| = 1$ . This implies that the leading eigenvector  $\boldsymbol{\xi}_1$  of  $\Sigma$  is the normalized mean vector  $\boldsymbol{\xi}_1 = (\mu_1, \mu_2)^\top / \gamma$ , while the second eigenvector can be defined up to the sign as  $\boldsymbol{\xi}_2 = (-\mu_2, \mu_1)^\top / \gamma$  or  $\boldsymbol{\xi}_2 = (\mu_2, -\mu_1)^\top / \gamma$ . The leading eigenvalue is equal to 1, while the second eigenvalue is equal to  $\lambda \in (0, 1]$ , hence  $|\Sigma| = \lambda$  and (3) becomes

$$f(\mathbf{y}) = \frac{1}{2\pi\lambda^{1/2} \left( B\sqrt{\gamma^2 + 1} - \alpha\sqrt{B} \right)}, \quad (7)$$

where the inverse of the scatter matrix is given by

$$\Sigma^{-1} = \frac{1}{\gamma^2} \begin{pmatrix} \mu_1^2 + \mu_2^2/\lambda & \mu_1\mu_2(1 - 1/\lambda) \\ \mu_1\mu_2(1 - 1/\lambda) & \mu_2^2 + \mu_1^2/\lambda \end{pmatrix} = \boldsymbol{\xi}_1\boldsymbol{\xi}_1^\top + \boldsymbol{\xi}_2\boldsymbol{\xi}_2^\top/\lambda. \quad (8)$$

The density in (7) may also be expressed in polar coordinates by

$$\begin{aligned} f(\theta) &= \frac{1}{2\pi\lambda^{1/2} \left[ (\cos^2(\theta - \omega) + \frac{1}{\lambda} \sin^2(\theta - \omega)) \sqrt{\gamma^2 + 1} - \gamma \cos(\theta - \omega) \sqrt{\cos^2(\theta - \omega) + \frac{1}{\lambda} \sin^2(\theta - \omega)} \right]} \\ f(\theta) &= \frac{1}{2\pi\lambda^{1/2} \left( b\sqrt{\gamma^2 + 1} - a\sqrt{b} \right)}, \end{aligned} \quad (9)$$

where  $b = \cos^2(\theta - \omega) + \frac{1}{\lambda} \sin^2(\theta - \omega)$ .

We will denote the distribution whose density is given by (7) or (9) by generalized circular projected Cauchy (GCPC) distribution. It is straightforward to note that the distribution is symmetrical about  $\theta$ , that is,  $f(\theta - \omega) = f(\omega - \theta)$ , and hence it is symmetrical about  $\mu + \pi$ . This fact, combined with unimodality, implies that the mean, median, and mode are all equal. straightforward to see that if  $\lambda = 1$ , the GCPC reduces to the WC distribution. Unfortunately, there is no closed form for the cumulative probability function, the trigonometric characteristic function, the mean resultant, and its length. This mandates that numerical integration techniques are applied to compute these quantities.

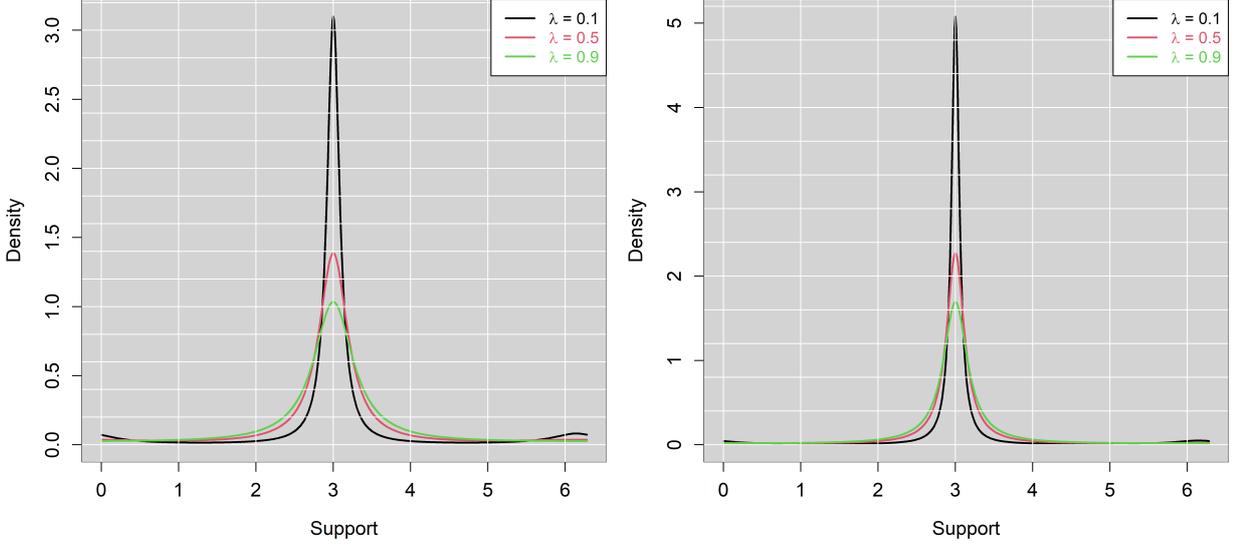
**Proposition 1** *The GCPC has a unique, global maximum at  $\mathbf{y} = \boldsymbol{\xi}_1 = \boldsymbol{\mu}/\gamma$ .*

A rigorous proof appears difficult, but the proposition is strongly supported by extensive numerical investigations. Assume without loss of generality that  $\boldsymbol{\mu} = \gamma(0, 1)^\top$ , and hence that  $\boldsymbol{\xi}_1 = (0, 1)^\top$ . Following [Paine et al. \(2018\)](#), let us write  $\mathbf{y} = \left( u, (1 - u^2)^{1/2} \right)^\top$  and restrict our attention to  $u$ , which lies in a neighborhood of 0. By substituting  $\mathbf{y}$  into the logarithm of (7) and by differentiating this and equating it to zero, we end up with  $\left. \frac{\partial \log f(u)}{\partial u} \right|_0 = 0$ . Then, the second derivative must be negative  $\left. \frac{\partial^2 \log f(u)}{\partial u^2} \right|_0 < 0$ . For a range of values of  $\gamma$  from 1 to 10, and a range of values for  $\lambda$  from 0.01 to 0.99, both considered in increments of 0.01, we computed the second derivative. The second derivative was always negative for all combinations of  $\gamma$  and  $\lambda$ .

As for the uniqueness of the global maximum, i.e., unimodality, we followed a similar approach to [Paine et al. \(2018\)](#). We computed the density of the GCPC for a grid of values of the parameters and counted the associated number of global maxima. The results showed that there was always a unique maximum.

Figure 1 visualizes the GCPC distribution for different values of  $\lambda$  and  $\gamma$ . Note that alternative restrictions/parameterizations that include one extra parameter (isotropic scatter matrix or a non-zero correlation for instance) could be employed as well.

Maximum likelihood estimation of the parameters is possible via the Newton-Raphson algorithm, but the relevant derivatives are highly complicated and lengthy, and we rely on numerical optimizers. The regression setting is straightforward to implement. In a similar manner to the spherically projected multivariate linear model ([Presnell et al., 1998](#)), the response angular data are transformed into their polar coordinates, and the mean direction is linked to the covariates. Finally, random values are straightforward to generate from the GCPC using the bivariate Cauchy distribution.



(a) GCPC distribution with  $\gamma = 3$

(b) GCPC distribution with  $\gamma = 5$

Figure 1: Densities of the GCPC distribution with (a)  $\omega = 3$  and (b)  $\gamma = (3, 5)$  and  $\lambda = (0.1, 0.5, 0.9)$ .

### 3 Spherical projected Cauchy distributions

The density of the Cauchy distribution in  $\mathbb{R}^3$  with some scatter matrix and a mean vector  $\boldsymbol{\mu} \in \mathbb{R}^3$  is given by

$$f(\mathbf{x}) = \frac{\Gamma(2)}{\Gamma(0.5)\pi^{3/2}|\boldsymbol{\Sigma}|^{1/2}} \left[1 + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]^{-2} = \frac{1}{\pi^2|\boldsymbol{\Sigma}|^{1/2}} \left[1 + (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]^{-2}.$$

Following the same notation as before, the probability density function of the projected Cauchy variable  $\mathbf{Y} \in \mathbf{S}^2$  is given by

$$\begin{aligned} f(\mathbf{y}) &= \int_0^\infty \frac{r^2}{\pi^2|\boldsymbol{\Sigma}|^{1/2} \left[1 + (r\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (r\mathbf{y} - \boldsymbol{\mu})\right]^2} dr \\ &= \frac{B(\gamma^2 + 1)\sqrt{\Delta} \left[ \arctan 2(\sqrt{\Delta}, -A) - \arctan 2(\sqrt{\Delta}, A) + \pi \right] + 2A\Delta}{4\pi^2|\boldsymbol{\Sigma}|^{1/2}B\Delta^2}, \end{aligned} \quad (10)$$

where  $\gamma = \|\boldsymbol{\mu}\|$  and  $\Delta = B\gamma^2 + B - A^2$ .  $\arctan 2(y, x)$  is the two-argument arc-tangent<sup>2</sup>.

#### 3.1 The spherical independent projected Cauchy distribution

Evidently, when  $\boldsymbol{\Sigma} = \mathbf{I}_3$ , (10) becomes<sup>3</sup>

$$f(\mathbf{y}) = \frac{(\gamma^2 + 1)\sqrt{\delta} \left[ \arctan 2(\sqrt{\delta}, -\alpha) - \arctan 2(\sqrt{\delta}, \alpha) + \pi \right] + 2\alpha\delta}{4\pi^2\delta^2}, \quad (11)$$

where  $\delta = \gamma^2 + 1 - \alpha^2$ . This is the density function of the spherical independent symmetric projected Cauchy (SIPC) distribution.

<sup>2</sup>This returns the angle between the  $x$ -axis and the vector from the origin to  $(x, y)$ , i.e., for positive arguments  $\arctan 2$  equals the tangent of  $y/x$ ,  $\arctan 2(y, x) = \tan^{-1}(y/x)$

<sup>3</sup>From (11) we can verify that when  $\gamma = 0$ , which implies that  $\boldsymbol{\mu} = (0, 0, 0)^\top$ , then  $\alpha = 0$  and  $\delta = 1$ , and  $\arctan 2(\sqrt{E}, 0) = \pi/2$ , hence the numerator is left with  $\pi$ , whereas the denominator is left with  $4\pi^2$ . Hence the density function reduces to  $(4\pi)^{-1}$ , which is the density function of the spherical uniform distribution.

### 3.1.1 The spherical Cauchy distribution

A model that appears to be closely related to the SIPC model is the spherical Cauchy (SC) distribution (Kato and McCullagh, 2020), which can be seen as the generalization of the WC distribution to the sphere (and hyper-sphere). Its density on  $\mathbb{S}^2$  is given by

$$f(\mathbf{y}) = \frac{\Gamma(1.5)}{2\pi^{1.5}} \left( \frac{1 - \rho^2}{1 + \rho^2 - 2\mathbf{y}^\top \boldsymbol{\mu}} \right)^2,$$

where  $\boldsymbol{\mu} \in \mathbb{S}^2$  and  $\rho \in [0, 1)$ . We stress that, unlike the circular projected Cauchy with an identity covariance matrix, the SIPC is not identical to the SC distribution.

## 3.2 The spherical elliptically symmetric projected Cauchy distribution

We impose the same conditions as in Paine et al. (2018), that is,  $\boldsymbol{\Sigma}\boldsymbol{\mu} = \boldsymbol{\mu}$  and  $|\boldsymbol{\Sigma}| = 1$ , and hence (10) becomes

$$f(\mathbf{y}) = \frac{B(\gamma^2 + 1)\sqrt{E} \left[ \arctan 2 \left( \sqrt{E}, -\alpha \right) - \arctan 2 \left( \sqrt{E}, \alpha \right) + \pi \right] + 2\alpha E}{4\pi^2 B E^2}, \quad (12)$$

where  $E = B\gamma^2 + B - \alpha^2$ . Eq. (12) defines the density function of the spherical elliptically symmetric projected Cauchy (SESPC) distribution. The remark here is that the parameters  $\boldsymbol{\mu}$  and  $\boldsymbol{\theta}$  are orthogonal.

## 3.3 A detailed explanation of the scatter matrix

We remind the reader that the scatter matrix  $\boldsymbol{\Sigma}$  is embedded in (12) via  $A = \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ ,  $B = \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}$ , which are also included in the term  $E$ . The largest eigenvalue of the positive definite matrix  $\boldsymbol{\Sigma}$   $\lambda_3$  is equal to 1, whilst the other two eigenvalues are  $0 < \lambda_1 < \lambda_2$ , such that  $\lambda_1 \lambda_2 = 1$  and the inverse of  $\boldsymbol{\Sigma}$  can be written as  $\boldsymbol{\Sigma}^{-1} = \sum_{j=1}^3 \boldsymbol{\xi}_j \boldsymbol{\xi}_j^\top / \lambda_j$ , where  $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$  and  $\boldsymbol{\xi}_3 = \boldsymbol{\mu} / \gamma$ , is a set of mutually orthogonal unit vectors. Note that the third axis, or third eigenvector, is the mean direction, the impact of which is discussed in § 3.4.

As in Paine et al. (2018) we define a pair of unit vectors,  $\tilde{\boldsymbol{\xi}}_1$  and  $\tilde{\boldsymbol{\xi}}_2$ , which are orthogonal to each other and to the mean direction  $\boldsymbol{\xi}_3$ :  $\tilde{\boldsymbol{\xi}}_1 = (-\mu_0^2, \mu_1 \mu_2, \mu_1 \mu_3)^\top / (\gamma \mu_0)$  and  $\tilde{\boldsymbol{\xi}}_2 = (0, -\mu_3, \mu_2)^\top / \mu_0$ , where  $\mu_0 = (\mu_2^2 + \mu_3^2)^{1/2}$ . Thus  $\boldsymbol{\Sigma}^{-1}$  can be written as

$$\boldsymbol{\Sigma}^{-1} = \mathbf{I}_3 + \theta_1 \left( \tilde{\boldsymbol{\xi}}_1 \tilde{\boldsymbol{\xi}}_1^\top - \tilde{\boldsymbol{\xi}}_2 \tilde{\boldsymbol{\xi}}_2^\top \right) + \theta_2 \left( \tilde{\boldsymbol{\xi}}_1 \tilde{\boldsymbol{\xi}}_2^\top + \tilde{\boldsymbol{\xi}}_2 \tilde{\boldsymbol{\xi}}_1^\top \right) + \left[ (\theta_1^2 + \theta_2^2 + 1)^{1/2} - 1 \right] \left( \tilde{\boldsymbol{\xi}}_1 \tilde{\boldsymbol{\xi}}_1^\top + \tilde{\boldsymbol{\xi}}_2 \tilde{\boldsymbol{\xi}}_2^\top \right).$$

The relationship between the  $\boldsymbol{\xi}_i$ s and the  $\tilde{\boldsymbol{\xi}}_i$ s that allows the axes of symmetry,  $\boldsymbol{\xi}_1$  and  $\boldsymbol{\xi}_2$ , to be an arbitrary rotation of  $\tilde{\boldsymbol{\xi}}_1$  and  $\tilde{\boldsymbol{\xi}}_2$  is detailed in Paine et al. (2018). The use of the  $\tilde{\boldsymbol{\xi}}_i$ s axes allows for unconstrained parameter estimation, since, unlike the eigenvalues of  $\boldsymbol{\Sigma}$ ,  $\theta_1$  and  $\theta_2$  lie on the real line. Further, note that the total number of free parameters is five, the same as for the multivariate Cauchy distribution in a tangent space  $\mathbb{R}^2$  to the sphere.

If  $\theta_1 = \theta_2 = 0$  then  $\boldsymbol{\Sigma} = \mathbf{I}_3$  and hence Eq. (12) reduces to Eq. (11). The rotational symmetry can hence be tested using the log-likelihood ratio test which, asymptotically, follows a  $\chi^2$  distribution with two degrees of freedom. Rejection of the rotational symmetry favors the SESPC model (12) over the SIPC model (11).

## 3.4 The impact of the mean direction being an eigenvector

The first condition imposed was that  $\boldsymbol{\Sigma}\boldsymbol{\mu} = \boldsymbol{\mu}$ , that is, the second eigenvector (i.e., the eigenvector corresponding to the second highest eigenvalue  $\lambda_2 = 1$ ) is equal to the mean direction  $\boldsymbol{\mu}$ . As Paine et al. (2018) state, the condition imposes symmetry about the eigenvectors of  $\boldsymbol{\Sigma}$ . Without loss of generality, suppose that the eigenvectors are parallel to the coordinate axes; that is, each element of the vector  $\boldsymbol{\xi}_j$  equals 0 except the  $j$ -th element, which equals 1. Then, if  $\mathbf{y} = (y_1, y_2, y_3)^\top$ ,

$$\mathbf{y}^\top \boldsymbol{\xi}_3 = y_3 \quad \text{and} \quad \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} = y_3^2 + \sum_{j=1}^2 y_j^2 / \lambda_j. \quad (13)$$

In this case, the density (7) depends only on  $y_j$  through  $y_j^2$  for  $j = 1, 2$ . Consequently, the density is invariant with respect to sign changes of the  $y_1, y_2$ , that is,  $f_{SESPC}(\pm y_1, \pm y_2, y_3) = f_{SESPC}(y_1, y_2, y_3)$ , which implies reflective symmetry about 0 along the axes defined by  $\xi_1$  and  $\xi_2$ . This type of symmetry is implied by ellipse-like contours of constant density inscribed on the sphere, and such contours arise when the density (7) is unimodal. Whether the density is unimodal depends on the nature of the stationary point at  $\mathbf{y} = \boldsymbol{\mu}/\gamma$ .

**Proposition 2** *The SESPC has a global maximum at  $\mathbf{y} = \xi_3 = \boldsymbol{\mu}/\gamma$ .*

Again, rigorous proof appears difficult, and our conjecture is that if the stationary point is a local maximum, then it is a global maximum and the distribution is unimodal. The conjecture is strongly supported by the extensive numerical investigations we have performed. Assume without loss of generality that  $\boldsymbol{\mu} = \gamma(0, 0, 1)^\top$ , and hence  $\xi_3 = (0, 0, 1)^\top$ . This means that  $\xi_1 = (1, 0, 0)^\top$  and  $\xi_2 = (0, 1, 0)^\top$ . Following Paine et al. (2018), let us write  $\mathbf{y} = (0, u, (1 - u^2)^{1/2})^\top$  and restrict our attention to  $u$ , which lies in a neighborhood of 0. By substituting  $\mathbf{y}$  into the logarithm of (12) and by differentiating this and equating it to zero, we end up with  $\left. \frac{\partial \log f(u)}{\partial u} \right|_0 = 0$ . The second derivative must then be negative  $\left. \frac{\partial^2 \log f(u)}{\partial u^2} \right|_0 < 0$ . For a range of values of  $\gamma$  from 1 up to 10, and a range of values for  $\lambda_2$  from 0.01 up to 0.99, both in increments of 0.01, we computed the second derivative. The second derivative was always negative for all combinations of  $\gamma$  and  $\lambda_2$  values.

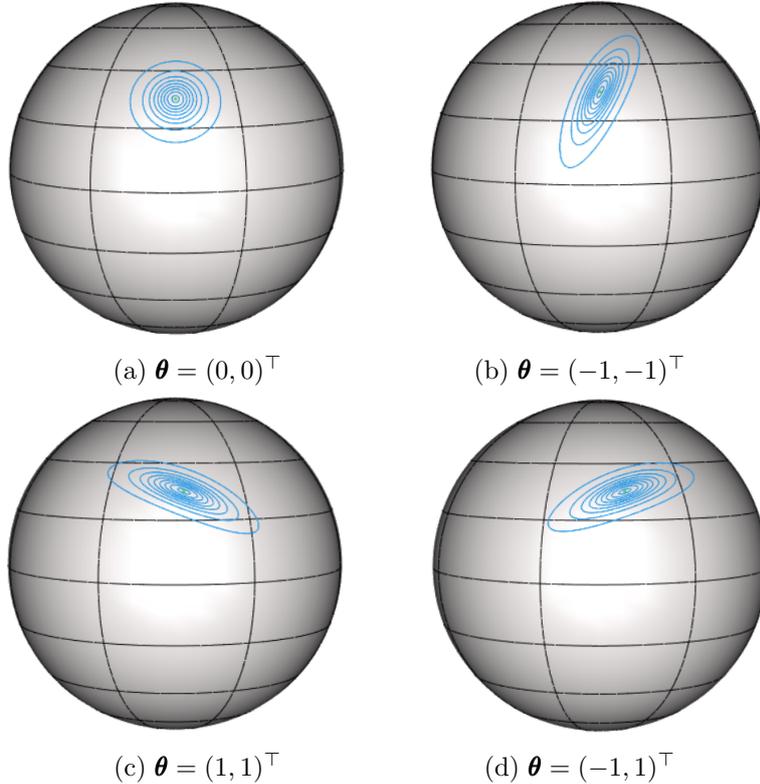


Figure 2: Contour plots of the SESPC distribution with  $\boldsymbol{\mu} = (5.843, 3.057, 3.758)^\top$  and various  $\boldsymbol{\theta}$  parameters.

## 4 Simulation studies

The simulation studies cover all aspects discussed earlier, that is, the examination of certain properties of the maximum likelihood estimation of the circular and spherical models with and without covariates. Further, we compare the models to some already-known distributions.

## 4.1 WC versus GCPC

At first, we compare the bias in the estimated mean vector when the data are generated from GCPC with a mean vector equal to  $\mu = c(3, 10)^\top$  and  $\lambda = (0.1, 0.3, 0.5, 0.7, 0.7, 1)$  for various sample sizes,  $n = (50, 100, 300, 500, 1000)$ . The purposes of this study are: a) to assess the bias of the estimated mean vector between the two models, the WC and GCPC, and b) to assess the power of the log-likelihood ratio in discriminating between the two. The log-likelihood ratio test statistic in this case follows a mixture of Dirac distribution at 1 and a  $\chi_1^2$  distribution with mixing probability equal to 0.5.

Table 1 contains the Euclidean distances of the estimated mean vectors under the WC and GCPC models, averaged over 1,000 repetitions. It also contains the estimated power for testing whether  $\lambda = 1$ , i.e., discriminating between the two models. The results evidently show that when  $\lambda \ll 1$ , the accuracy of the GCPC is substantially higher than that of WC. Further, the discrepancy between them increases with the sample size. As  $\lambda$  tends to 1 the differences diminish. When  $\lambda = 0.9$  and  $\lambda = 1$  the Euclidean distance of the estimated mean vector under the GCPC is greater than that under the WC model. Our explanation for this phenomenon is that the causal factor is the small proportion of times that the log-likelihood ratio rejects the WC in favor of the GCPC. To validate this, we computed the averages excluding these cases and observed that the average Euclidean distances were nearly identical.

Table 1: Average Euclidean distances between the true mean vector  $\mu = (3, 10)^\top$  and the estimated mean vector using WC and GCPC distributions for various  $\lambda$  values. The row termed Power contains the estimated power for testing whether  $\lambda = 1$  (when  $\lambda = 1$ , the GCPC reduces to the WC model).

$\lambda$	Model and Power	Sample size				
		n=50	n=100	n=300	n=500	n=1000
$\lambda = 0.1$	WC	23.467	23.043	22.824	22.505	22.470
	GCPC	9.856	6.343	2.924	2.226	1.688
	Power	0.438	0.633	0.942	0.983	0.985
$\lambda = 0.3$	WC	9.337	8.874	8.646	8.612	8.589
	GCPC	5.762	4.475	2.655	2.026	1.355
	Power	0.231	0.345	0.622	0.819	0.960
$\lambda = 0.5$	WC	5.028	4.485	4.363	4.392	4.299
	GCPC	3.995	3.134	2.181	1.778	1.309
	Power	0.125	0.180	0.367	0.503	0.732
$\lambda = 0.7$	WC	3.014	2.388	2.096	2.117	2.054
	GCPC	2.983	2.306	1.725	1.491	1.171
	Power	0.091	0.100	0.197	0.236	0.355
$\lambda = 0.9$	WC	1.990	1.373	0.884	0.743	0.615
	GCPC	2.587	1.980	1.272	1.039	0.826
	Power	0.051	0.061	0.082	0.094	0.114
$\lambda = 1$	WC	1.761	1.207	0.682	0.543	0.378
	GCPC	2.372	1.923	1.095	0.964	0.715
	Power	0.049	0.057	0.049	0.057	0.058

## 4.2 SIPC versus SESPC

As before, we compare the bias in the estimated mean vector when the data are generated from SESPC with a mean vector equal to  $\mu = c(5.843, 3.057, 3.758)^\top$  and  $\theta_1 = \theta_2 = (0.1, 0.3, 0.5, 0.7, 0.7, 1)$  for various sample sizes,  $n = (50, 100, 300, 500, 1000)$ . The purposes of this study are: a) to assess the bias of the estimated mean vector between the two models, SIPC and SESPC, and b) to assess the power of the log-likelihood ratio in

discriminating between the two.

Table 2 contains the Euclidean distances of the estimated mean vectors under the SIPC and SESPC models, averaged over 1,000 repetitions. It also contains the estimated power for testing whether  $\theta_1 = \theta_2 = 0$ , i.e., discriminating between the two models. The results evidently show that as  $\theta$  increases, the accuracy of the SESPC becomes substantially higher than that of the SIPC. Further, the discrepancy between the two increases with the sample size. Naturally, as the  $\theta$  values tend to 0, the differences diminish.

Table 2: Average Euclidean distances between the true mean vector is  $\boldsymbol{\mu} = (5.843, 3.057, 3.758)^\top$  and the estimated mean vector  $\hat{\boldsymbol{\mu}}$  using SIPC and SESPC distributions for various values of common  $\theta$  parameters. The row termed Power contains the estimated power for testing rotational symmetry ( $\theta_1 = \theta_2 = 0$ ), i.e., the SESPC distribution versus the SIPC distribution.

$(\theta_1, \theta_2)$	Model and Power	Sample size				
		n=50	n=100	n=300	n=500	n=1000
$\theta_1 = \theta_2 = 0$	SIPC	1.106	0.739	0.422	0.332	0.221
	SESPC	1.123	0.747	0.424	0.334	0.222
	Power	0.057	0.060	0.065	0.049	0.050
$\theta_1 = \theta_2 = 0.2$	SIPC	1.082	0.740	0.426	0.331	0.243
	SESPC	1.111	0.755	0.429	0.329	0.237
	Power	0.249	0.477	0.930	0.991	1.000
$\theta_1 = \theta_2 = 0.4$	SIPC	1.057	0.765	0.450	0.373	0.309
	SESPC	1.151	0.778	0.430	0.327	0.234
	Power	0.764	0.968	1.000	1.000	1.000
$\theta_1 = \theta_2 = 0.6$	SIPC	1.071	0.819	0.552	0.498	0.465
	SESPC	1.100	0.776	0.429	0.336	0.233
	Power	0.959	1.000	1.000	1.000	1.000
$\theta_1 = \theta_2 = 0.8$	SIPC	1.133	0.886	0.736	0.730	0.726
	SESPC	1.206	0.765	0.425	0.339	0.242
	Power	0.988	0.999	1.000	1.000	1.000
$\theta_1 = \theta_2 = 1$	SIPC	1.226	1.022	0.997	0.970	0.985
	SESPC	1.258	0.843	0.433	0.327	0.241
	Power	0.981	0.993	1.000	1.000	0.999

### 4.3 SESPC versus SIPC and SC

We then examined the accuracy of SESPC, SIPC, and SC in estimating the mean direction of the fitted model,  $\hat{\mathbf{m}} = \boldsymbol{\mu}/\gamma$ , when the data are generated using the SESPC model for a range of  $\theta_1$  and  $\theta_2$  values. The measure we used for this is  $\sqrt{2[1 - \mathbf{E}(\hat{\mathbf{m}}^\top \hat{\mathbf{m}})]}$ , where the expectation was approximated by Monte Carlo.

Hence, Table 3 reports the following quantity

$$\text{Error}(\hat{\mathbf{m}}) = \sqrt{2 \left[ 1 - B^{-1} \sum_{b=1}^B (\hat{\mathbf{m}}_i^\top \mathbf{m}) \right]},$$

where  $(\hat{\mathbf{m}}_i)$  is the estimated mean direction of the fitted model for the  $i$ -run out of  $B$  Monte Carlo runs. Regardless of the values of the  $\theta_s$ , the errors for SESPC and the SIPC are nearly identical for large sample sizes ( $n \geq 300$ ), indicating that even when elliptical symmetry exists, SIPC does a good job at estimating the mean direction. However, we highlight that this does not hold for the estimated concentration parameters (not shown here).

Table 3: The Error ( $\widehat{\mathbf{m}}$ ) of the fitted model when the true mean vector is  $\boldsymbol{\mu} = (5.843, 3.057, 3.758)^\top$  and hence the true mean direction is  $\mathbf{m} = (0.770, 0.403, 0.495)^\top$ .

$(\theta_1, \theta_2)$	Model and Power	Sample size				
		n=50	n=100	n=300	n=500	n=1000
$\theta_1 = \theta_2 = 0$	SESPC	0.035	0.024	0.014	0.011	0.008
	SIPC	0.035	0.024	0.014	0.011	0.008
	SC	0.151	0.139	0.151	0.162	0.147
$\theta_1 = \theta_2 = 0.2$	SESPC	0.036	0.026	0.014	0.011	0.008
	SIPC	0.036	0.026	0.014	0.011	0.008
	SC	0.136	0.158	0.128	0.162	0.159
$\theta_1 = \theta_2 = 0.4$	SESPC	0.095	0.051	0.015	0.012	0.046
	SIPC	0.038	0.027	0.015	0.012	0.008
	SC	0.156	0.137	0.143	0.159	0.150
$\theta_1 = \theta_2 = 0.6$	SESPC	0.109	0.029	0.017	0.012	0.063
	SIPC	0.043	0.030	0.017	0.012	0.009
	SC	0.149	0.152	0.147	0.104	0.138
$\theta_1 = \theta_2 = 0.8$	SESPC	0.193	0.067	0.017	0.013	0.046
	SIPC	0.045	0.031	0.018	0.014	0.010
	SC	0.160	0.161	0.160	0.153	0.151
$\theta_1 = \theta_2 = 1$	SESPC	0.241	0.111	0.019	0.014	0.010
	SIPC	0.053	0.034	0.020	0.015	0.011
	SC	0.177	0.151	0.141	0.135	0.147

## 5 Demonstrations using real data

Examples using real data illustrate the superior performance of the suggested models over some known models.

### 5.1 Circular data

The first example refers to measurements of the directions taken by 76 turtles after some treatment<sup>4</sup> (Fisher, 1995, pg. 241). The estimated parameters of the WC and the GCPC models are presented in Table 4. The parameters in general appear to be close, with the exception of the  $\gamma$  parameter for the WC, which is higher. The p-value for the hypothesis that  $\lambda = 1$  is equal to 0.0002, hence the GCPC evidently is preferred to the WC model. Figure 3 contains the kernel density estimate of the data along with the fitted densities of the WC and the GCPC models. The WC and GCPC models produce a higher density for the first mode, but only the GCPC model captures the second.

Table 4: Estimated parameters for the CPC and GCPC models for the Turtles dataset.

Model	$\hat{\mu}$	$\hat{\boldsymbol{\mu}}$ in $\mathbb{S}^1$	$\hat{\boldsymbol{\mu}}$ in $\mathbb{R}^2$	$\hat{\gamma}$	$\hat{\lambda}$	Log-likelihood
CPC	1.107	$(0.448, 0.894)^\top$	$(0.730, 1.458)^\top$	1.630		-113.248
GCPC	1.094	$(0.459, 0.889)^\top$	$(0.458, 0.888)^\top$	0.999	0.341	-109.857

### 5.2 Spherical data

To visually compare the density values of the SIPC, SESPC, and spherical Cauchy distributions, we will use the Paleomagnetic pole dataset (Schmidt, 1976). This dataset contains estimates of the position of the Earth's

<sup>4</sup>The dataset is available to download from the *R* package *circular* (Agostinelli and Lund, 2017).

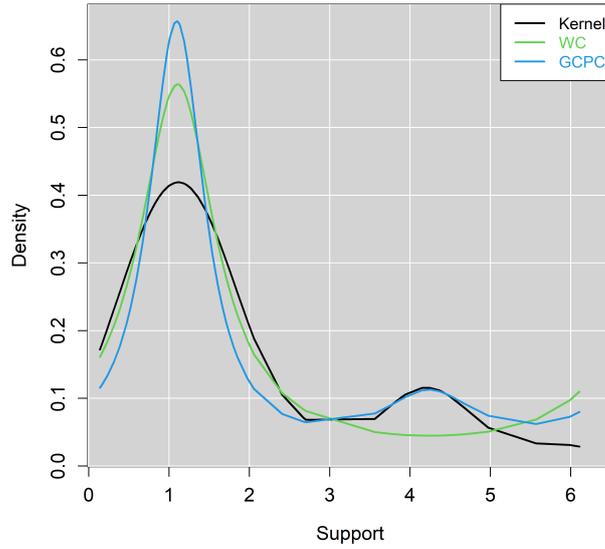


Figure 3: (a) The kernel density estimate, and the WC & GCPC densities of the turtles dataset.

historic magnetic pole calculated from 33 different sites in Tasmania. The data are visualized in Figure 4. Evidently, the rotational symmetry is rejected ( $p$ -value = 0.0026), corroborating the findings of [Paine et al. \(2018\)](#).

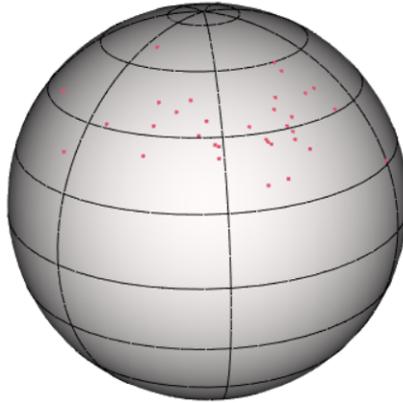


Figure 4: Sphere plot of the Paleomagnetic dataset.

Table 5 reports the estimated parameters for each of the three distributions. Evidently, the mean directions are not far from each other. Figure 5 shows the spherical contour plots of the fitted densities. It is evident that SESPC has managed to capture the shape of the data more accurately.

Figure 6 presents the transects of the densities of the three distributions. We computed the density values for a range of 1,000 values of latitude and longitude along the observed range of the data. We then matched the latitude to the latitude of the mean direction of the SESPC distribution and left the longitude varying. Hence, Figure 6 shows a slice of the multivariate density as a function of the longitude, when the latitude is  $134.71^\circ$ .

The assumption of rotational symmetry is rejected ( $p$ -value=0.0026) and hence the densities of the SIPC and SESPC are very similar, as one might expect and the densities differ. Their shape is similar, but the locations of the modes are a little different, showing the effect of the elliptical symmetry present in the data.

Table 5: Estimated parameters for the SIPC, SESPC, and spherical Cauchy distributions. The columns present the estimated mean direction in Euclidean and polar coordinates, the estimated concentration parameters, and the estimated  $\theta$  parameters for the SESPC.

Model	Estimated parameters			
Spherical Cauchy	$\hat{\boldsymbol{\mu}} = (-0.699, 0.224, 0.679)^\top$	$\hat{\boldsymbol{\mu}} = (134.36^\circ, 71.75^\circ)^\top$	$\hat{\rho} = 0.752$	
SIPC	$\hat{\boldsymbol{\mu}} = (-4.155, 1.130, 3.834)^\top$	$\hat{\boldsymbol{\mu}} = (136.11^\circ, 73.58^\circ)^\top$	$\hat{\gamma} = 5.766$	
SESPC	$\hat{\boldsymbol{\mu}} = (-4.207, 1.433, 3.991)^\top$	$\hat{\boldsymbol{\mu}} = (134.77^\circ, 70.25^\circ)^\top$	$\hat{\gamma} = 5.973$	$\hat{\boldsymbol{\theta}} = (0.219, -0.846)^\top$

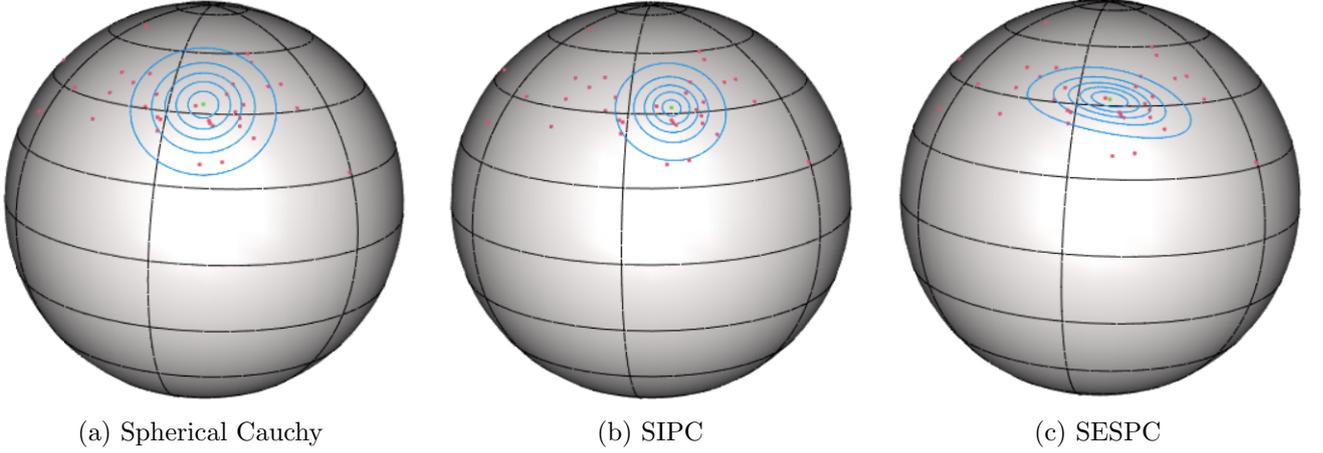


Figure 5: Comparison of the three distributions fitted to the Paleomagnetic dataset: spherical contour plots of each fitted distribution.

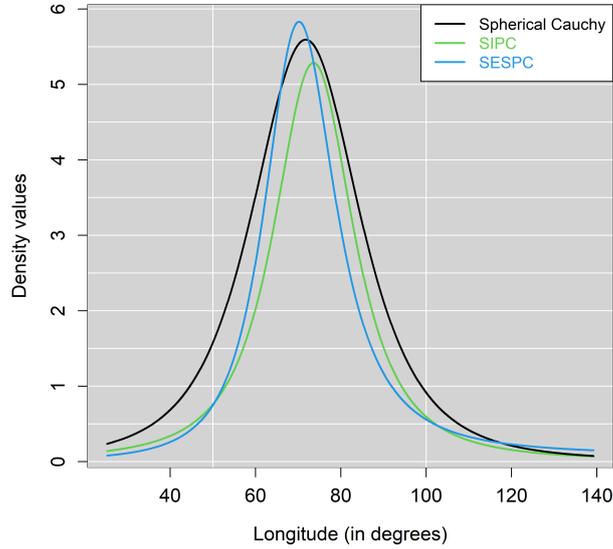


Figure 6: Comparison of the three distributions fitted to the Paleomagnetic dataset: transects of the densities when Latitude=134.71°.

## 6 Conclusions

We introduced the projected Cauchy distribution on the circle and the sphere. For the circular case, the projection of the bivariate Cauchy distribution resulted in the wrapped Cauchy distribution, but with a different

parameterization. We then induced a restriction in the scatter matrix that led to the generalized projected Cauchy distribution that includes one extra parameter. We highlight that alternative scatter matrix structures could be imposed as well. The comparison of the proposed generalized Cauchy distribution, with other variants stemming from the projected Cauchy and with other three-parameter circular distributions, might well be of interest to future research.

Moving to the spherical case, we employed the same strategy and projected the trivariate Cauchy distribution on the sphere and then imposed two conditions on the scatter matrix yielding elliptical symmetry. However, two drawbacks are a) the density does not have a convenient formula, and b) generalization of the density formula to higher dimensions is not straightforward and we could not find a general form. However, the advantages of the projected Cauchy family of distributions include a closed form for the normalizing constant and an efficient way to simulate values.

# Appendix

## A1 Derivation of (3)

We recall the terms used,  $A = \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ ,  $B = \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}$  and  $\Gamma^2 = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ . Thus, the indefinite integral above Eq. (3), which we will solve in the first place, can be written as

$$I = \int \frac{r}{(1 + Br^2 - 2Ar + \Gamma^2)^{3/2}} dr \quad (\text{A1.1})$$

$$= \int \left( \frac{Br - A}{B(Br^2 - 2Ar + \Gamma^2 + 1)^{3/2}} + \frac{A}{B(Br^2 - 2Ar + \Gamma^2 + 1)^{3/2}} \right) dr \quad (\text{A1.2})$$

$$= \frac{1}{B} \int \frac{Br - A}{(Br^2 - 2Ar + \Gamma^2 + 1)^{3/2}} dr + \frac{A}{B} \int \frac{1}{(Br^2 - 2Ar + \Gamma^2 + 1)^{3/2}} dr = \frac{1}{B} I_1 + \frac{A}{B} I_2 \quad (\text{A1.3})$$

$$(\text{A1.4})$$

Let us now solve the first integral ( $I_1$ ). Substitute  $u = Br^2 - 2Ar + \Gamma^2 + 1$  and thus  $du = 2(Br - A)dr$ , hence the first integral becomes

$$I_1 = \frac{1}{2} \int \frac{1}{u^{3/2}} du = -\frac{1}{\sqrt{u}} = -\frac{1}{\sqrt{Br^2 - 2Ar + \Gamma^2 + 1}}. \quad (\text{A1.5})$$

Let us now solve the second integral ( $I_2$ ).

$$I_2 = B^{3/2} \int \frac{1}{\left[ (Br - A)^2 + B(\Gamma^2 + 1) - A^2 \right]^{3/2}} dr = B^{3/2} I_3. \quad (\text{A1.6})$$

Substitute  $u = Br - A$  and hence  $du = Bdr$ , thus  $I_3$  can be written as

$$I_3 = \frac{1}{B} \int \frac{1}{[u^2 + B(\Gamma^2 + 1) - A^2]^{3/2}} du = \frac{1}{B} I_4. \quad (\text{A1.7})$$

Again, using substitution,  $u = \sqrt{B(\Gamma^2 + 1) - A^2} \tan(v)$  and  $v = \arctan\left(\frac{u}{\sqrt{B(\Gamma^2 + 1) - A^2}}\right)$ . Then  $du = \sqrt{B(\Gamma^2 + 1) - A^2} \sec^2(v) dv$ . Thus,  $I_4$  becomes

$$I_4 = \int \frac{\sqrt{B(\Gamma^2 + 1) - A^2} \sec^2(v)}{[(B(\Gamma^2 + 1) - A^2) \tan^2(v) + B(\Gamma^2 + 1) - A^2]^{3/2}} dv = \frac{1}{B\Gamma^2 + B - A^2} \int \frac{1}{\sec(v)} dv \quad (\text{A1.8})$$

$$= \frac{\sin(v)}{B\Gamma^2 + B - A^2}. \quad (\text{A1.9})$$

We undo the last substitution, and hence  $\sin(v) = \sin\left[\arctan\left(\frac{u}{\sqrt{B(\Gamma^2 + 1) - A^2}}\right)\right] = \frac{u}{\sqrt{B(\Gamma^2 + 1) - A^2} \sqrt{\frac{u^2}{B(\Gamma^2 + 1) - A^2} + 1}}$ .

We plug this last result into  $I_4$  to obtain

$$I_4 = \frac{u}{(B\Gamma^2 + B - A^2) \sqrt{B(\Gamma^2 + 1) - A^2} \sqrt{\frac{u^2}{B(\Gamma^2 + 1) - A^2} + 1}} \quad (\text{A1.10})$$

and hence  $I_3$  becomes

$$I_3 = \frac{u}{B(B\Gamma^2 + B - A^2) \sqrt{B(\Gamma^2 + 1) - A^2} \sqrt{\frac{u^2}{B(\Gamma^2 + 1) - A^2} + 1}}. \quad (\text{A1.11})$$

We now undo the substitution that took us from  $I_2$  to  $I_3$  and obtain

$$I_2 = \frac{B^{3/2} (Br - A)}{B(B\Gamma^2 + B - A^2) \sqrt{B(\Gamma^2 + 1) - A^2} \sqrt{\frac{(Br - A)^2}{B(\Gamma^2 + 1) - A^2} + 1}}. \quad (\text{A1.12})$$

Finally,  $I$  is written as

$$I = -\frac{1}{B} \frac{1}{\sqrt{Br^2 - 2Ar + \Gamma^2 + 1}} + \frac{A}{B} \frac{B^{3/2} (Br - A)}{B (B\Gamma^2 + B - A^2) \sqrt{B(\Gamma^2 + 1) - A^2} \sqrt{\frac{(Br-A)^2}{\sqrt{B(\Gamma^2+1)-A^2}} + 1}} + c, \quad (\text{A1.13})$$

where  $c$  is a constant. After some rearrangement, the above integral becomes

$$I = \frac{Ar - \Gamma^2 - 1}{(B\Gamma^2 + B - A^2) \sqrt{Br^2 - 2Ar + \Gamma^2 + 1}} + c. \quad (\text{A1.14})$$

Hence the definite integral is equal to

$$I = \int_0^\infty \frac{r}{(1 + Br^2 - 2Ar + \Gamma^2)^{3/2}} dr = \frac{B\sqrt{\Gamma^2 + 1} + A\sqrt{B}}{B(B\Gamma^2 + B - A^2)}. \quad (\text{A1.15})$$

By simplifying the above expression using identities and by adding the ignored constant terms, we end up with the expression in (3).

## A2 Derivatives of the log-likelihood of the WC with the parameterization of (4)

$$\begin{aligned} \ell &= -n \log(2\pi) - \sum_{i=1}^n \log(\sqrt{\gamma^2 + 1} - \alpha_i) \\ \frac{\partial \ell}{\partial \boldsymbol{\mu}} &= - \sum_{i=1}^n \frac{\frac{\boldsymbol{\mu}}{\sqrt{\gamma^2 + 1}} - \mathbf{y}_i}{\sqrt{\gamma^2 + 1} - \alpha_i} \\ \frac{\partial^2 \ell}{\partial \boldsymbol{\mu} \boldsymbol{\mu}^\top} &= - \sum_{i=1}^n \frac{\frac{\mathbf{I}_2 \sqrt{\gamma^2 + 1} - \frac{\boldsymbol{\mu} \boldsymbol{\mu}^\top}{\sqrt{\gamma^2 + 1}}}{\gamma^2 + 1} (\sqrt{\gamma^2 + 1} - \alpha_i) - \left( \frac{\boldsymbol{\mu}}{\sqrt{\gamma^2 + 1}} - \mathbf{y}_i \right) \left( \frac{\boldsymbol{\mu}}{\sqrt{\gamma^2 + 1}} - \mathbf{y}_i \right)^\top}{(\sqrt{\gamma^2 + 1} - \alpha_i)^2}. \end{aligned}$$

## A3 Derivatives of the log-likelihood of the WC with the parameterization of (5)

$$\begin{aligned} \ell &= -n \log(2\pi) - \sum_{i=1}^n \log(\sqrt{\gamma^2 + 1} - \gamma \cos(\theta_i - \omega)) \\ \frac{\partial \ell}{\partial \omega} &= \sum_{i=1}^n \frac{\gamma \sin(\theta_i - \omega)}{\sqrt{\gamma^2 + 1} - \gamma \cos(\theta_i - \omega)} \\ \frac{\partial \ell}{\partial \gamma} &= - \sum_{i=1}^n \frac{\frac{\gamma}{\sqrt{\gamma^2 + 1}} - \cos(\theta_i - \omega)}{\sqrt{\gamma^2 + 1} - \gamma \cos(\theta_i - \omega)} \\ \frac{\partial^2 \ell}{\partial \omega^2} &= \sum_{i=1}^n \frac{\gamma^2 \sin^2(\theta_i - \omega)}{(\sqrt{\gamma^2 + 1} - \gamma \cos(\theta_i - \omega))^2} - \frac{\gamma \cos(\theta_i - \omega)}{\sqrt{\gamma^2 + 1} - \gamma \cos(\theta_i - \omega)} \\ \frac{\partial^2 \ell}{\partial \gamma^2} &= \sum_{i=1}^n \frac{\left( \frac{\gamma}{\sqrt{\gamma^2 + 1}} - \cos(\theta_i - \omega) \right)^2}{(\sqrt{\gamma^2 + 1} - \gamma \cos(\theta_i - \omega))^2} - \frac{\frac{1}{\sqrt{\gamma^2 + 1}} - \frac{\gamma^2}{(\gamma^2 + 1)^{3/2}}}{\sqrt{\gamma^2 + 1} - \gamma \cos(\theta_i - \omega)} \\ \frac{\partial^2 \ell}{\partial \omega \partial \gamma} &= \sum_{i=1}^n \frac{\sin(\theta_i - \omega)}{\sqrt{\gamma^2 + 1} (\sqrt{\gamma^2 + 1} - \gamma \cos(\theta_i - \omega))^2}. \end{aligned}$$

## The log density and its derivatives used in Proposition 1

$$\begin{aligned}
\log f(u) &= -\log \left[ \left( \frac{u^2}{\lambda} + 1 - u^2 \right) \sqrt{\gamma^2 + 1} - \sqrt{1 - u^2} \mu_2 \sqrt{\frac{u^2}{\lambda} + 1 - u^2} \right] \\
\frac{\partial \log f(u)}{\partial u} &= - \frac{\frac{\mu_2 u \sqrt{\frac{u^2}{\lambda} - u^2 + 1}}{\sqrt{1 - u^2}} - \frac{\mu_2 \cdot (\frac{2u}{\lambda} - 2u) \sqrt{1 - u^2}}{2\sqrt{\frac{u^2}{\lambda} - u^2 + 1}} + \sqrt{\gamma^2 + 1} \left( \frac{2u}{\lambda} - 2u \right)}{\sqrt{\gamma^2 + 1} \left( \frac{u^2}{\lambda} - u^2 + 1 \right) - \mu_2 \sqrt{1 - u^2} \sqrt{\frac{u^2}{\lambda} - u^2 + 1}} \\
\frac{\partial^2 \log f(u)}{\partial u^2} &= \frac{\left( \frac{\mu_2 u \sqrt{\frac{u^2}{\lambda} - u^2 + 1}}{\sqrt{1 - u^2}} - \frac{\mu_2 \cdot (\frac{2u}{\lambda} - 2u) \sqrt{1 - u^2}}{2\sqrt{\frac{u^2}{\lambda} - u^2 + 1}} + \sqrt{\gamma^2 + 1} \left( \frac{2u}{\lambda} - 2u \right) \right)^2}{\left( \sqrt{\gamma^2 + 1} \left( \frac{u^2}{\lambda} - u^2 + 1 \right) - \mu_2 \sqrt{1 - u^2} \sqrt{\frac{u^2}{\lambda} - u^2 + 1} \right)^2} \\
&\quad - \frac{\frac{\mu_2 \sqrt{\frac{u^2}{\lambda} - u^2 + 1}}{\sqrt{1 - u^2}} + \frac{\mu_2 u^2 \sqrt{\frac{u^2}{\lambda} - u^2 + 1}}{(1 - u^2)^{\frac{3}{2}}} - \frac{\mu_2 \cdot (\frac{2}{\lambda} - 2) \sqrt{1 - u^2}}{2\sqrt{\frac{u^2}{\lambda} - u^2 + 1}} + \frac{\mu_2 u \cdot (\frac{2u}{\lambda} - 2u)}{\sqrt{1 - u^2} \sqrt{\frac{u^2}{\lambda} - u^2 + 1}}}{\sqrt{\gamma^2 + 1} \left( \frac{u^2}{\lambda} - u^2 + 1 \right) - \mu_2 \sqrt{1 - u^2} \sqrt{\frac{u^2}{\lambda} - u^2 + 1}} \\
&\quad - \frac{\frac{\mu_2 \cdot (\frac{2u}{\lambda} - 2u)^2 \sqrt{1 - u^2}}{4 \left( \frac{u^2}{\lambda} - u^2 + 1 \right)^{\frac{3}{2}}} + \sqrt{\gamma^2 + 1} \left( \frac{2}{\lambda} - 2 \right)}{\sqrt{\gamma^2 + 1} \left( \frac{u^2}{\lambda} - u^2 + 1 \right) - \mu_2 \sqrt{1 - u^2} \sqrt{\frac{u^2}{\lambda} - u^2 + 1}}
\end{aligned}$$

## The log density used in Proposition 2

$$\begin{aligned}
\log f(u) &= \log \left[ (\gamma^2 + 1) \left( \frac{u^2}{\lambda_2} - u^2 + 1 \right) \sqrt{\frac{(\gamma^2 + 1) u^2}{\lambda_2} - u^2 + 1} \left( -\arctan 2 \left( \sqrt{\frac{(\gamma^2 + 1) u^2}{\lambda_2} - u^2 + 1}, \gamma \sqrt{1 - u^2} \right) \right. \right. \\
&\quad \left. \left. + \arctan 2 \left( \sqrt{\frac{(\gamma^2 + 1) u^2}{\lambda_2} - u^2 + 1}, -\gamma \sqrt{1 - u^2} \right) + \pi \right) + 2\gamma \sqrt{1 - u^2} \right] \\
&\quad - 2 \log \left( \frac{(\gamma^2 + 1) u^2}{\lambda_2} - u^2 + 1 \right) - \log \left( \frac{u^2}{\lambda_2} - u^2 + 1 \right)
\end{aligned}$$

## References

- Abe, T. and A. Pewsey (2011). Sine-skewed circular distributions. *Statistical Papers* 52(3), 683–707.
- Agostinelli, C. and U. Lund (2017). *R package circular: Circular Statistics (version 0.4-93)*. CA: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA.
- Dietrich, T. and W.-D. Richter (2017). Classes of geometrically generalized von Mises distributions. *Sankhya B* 79(1), 21–59.
- Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge University Press.
- Fisher, R. A. (1953). Dispersion on a sphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 217(1130), 295–305.
- Gatto, R. and S. R. Jammalamadaka (2007). The generalized von Mises distribution. *Statistical Methodology* 4(3), 341–353.
- Jones, M. and A. Pewsey (2005). A family of symmetric distributions on the circle. *Journal of the American Statistical Association* 100(472), 1422–1428.

- Jones, M. and A. Pewsey (2012). Inverse Batschelet distributions for circular data. *Biometrics* 68(1), 183–193.
- Kato, S. and M. Jones (2010). A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *Journal of the American Statistical Association* 105(489), 249–262.
- Kato, S. and M. Jones (2013). An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. *Bernoulli* 19(1), 154–171.
- Kato, S. and P. McCullagh (2020). Some properties of a Cauchy family on the sphere derived from the Möbius transformations. *Bernoulli* 26(4), 3224–3248.
- Kendall, D. G. (1974). Pole-seeking Brownian motion and bird navigation. *Journal of the Royal Statistical Society: Series B (Methodological)* 36(3), 365–402.
- Kent, J. T. (1982). The Fisher-Bingham distribution on the sphere. *Journal of the Royal Statistical Society: Series B (Methodological)* 44(1), 71–80.
- Kim, S. and A. SenGupta (2013). A three-parameter generalized von Mises distribution. *Statistical Papers* 54(3), 685–693.
- Mardia, K. (1972). *Statistics of directional data*. Academic Press, NY.
- Mardia, K. and P. Jupp (2000). *Directional Statistics*. John Wiley & Sons.
- Mardia, K. V. (1975). Statistics of directional data. *Journal of the Royal Statistical Society: Series B (Methodological)* 37(3), 349–371.
- Nuñez-Antonio, G. and E. Gutiérrez-Peña (2005). A Bayesian analysis of directional data using the projected normal distribution. *Journal of Applied Statistics* 32(10), 995–1001.
- Paine, P., S. P. Preston, M. Tsagris, and A. T. Wood (2018). An elliptically symmetric angular Gaussian distribution. *Statistics and Computing* 28(3), 689–697.
- Pewsey, A. (2000). The wrapped skew-normal distribution on the circle. *Communications in Statistics-Theory and Methods* 29(11), 2459–2472.
- Pewsey, A. (2008). The wrapped stable family of distributions as a flexible model for circular data. *Computational Statistics & Data Analysis* 52(3), 1516–1523.
- Pewsey, A., T. Lewis, and M. Jones (2007). The wrapped t family of circular distributions. *Australian & New Zealand Journal of Statistics* 49(1), 79–91.
- Presnell, B., S. P. Morrison, and R. C. Littell (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association* 93(443), 1068–1077.
- Scealy, J. and A. T. Wood (2019). Scaled von Mises–Fisher distributions and regression models for paleomagnetic directional data. *Journal of the American Statistical Association*.
- Schmidt, P. (1976). The non-uniqueness of the Australian Mesozoic palaeomagnetic pole position. *Geophysical Journal International* 47(2), 285–300.
- von Mises, R. (1918). Über die” Ganzzahligkeit” der Atomgewicht und verwandte Fragen. *Physikal. Z.* 19, 490–500.
- Wang, F. and A. E. Gelfand (2013). Directional data analysis under the general projected normal distribution. *Statistical methodology* 10(1), 113–127.
- Watson, G. S. (1983). *Statistics on spheres*. Wiley New York.