Bernoulli **26**(4), 2020, 3224–3248
https://doi.org/10.3150/20-BEJ1222

Some properties of a Cauchy family on the sphere derived from the Möbius transformations

SHOGO KATO[1] and PETER MCCULLAGH[2]

[1] *Institute of Statistical Mathematics, 10-3 Midori-Cho, Tachikawa, Tokyo 190-8562, Japan. E-mail: skato@ism.ac.jp*
[2] *Department of Statistics, University of Chicago, George Herbert Jones Laboratory, 5747 South Ellis Avenue, Chicago, IL 60637, USA. E-mail: pmcc@galton.uchicago.edu*

We present some properties of a Cauchy family of distributions on the sphere, which is a spherical extension of the wrapped Cauchy family on the circle. The spherical Cauchy family is closed under the Möbius transformations on the sphere and the parameter of the transformed family is expressed using extended Möbius transformations on the compactified Euclidean space. Stereographic projection transforms the spherical Cauchy family into a multivariate t-family with a certain degree of freedom on Euclidean space. The Möbius transformations and stereographic projection enable us to obtain some results related to the spherical Cauchy family such as an efficient algorithm for random variate generation, a simple form of pivotal statistic and straightforward calculation of probabilities of a region. A method of moments estimator and an asymptotically efficient estimator are expressed in closed form. Maximum likelihood estimation is also straightforward.

Keywords: directional statistics; high dimensional data; stereographic projection; von Mises–Fisher distribution; wrapped Cauchy distribution

1. Introduction

This paper discusses a family of distributions on the unit sphere with probability density function

$$f(y;\mu,\rho)=\frac{\Gamma\{(d+1)/2\}}{2\pi^{(d+1)/2}}\left(\frac{1-\rho^2}{1+\rho^2-2\rho\mu^T y}\right)^d,\quad y\in S^d, \tag{1.1}$$

with respect to surface area, where $\mu\in S^d$ is the location parameter, $\rho\in[0,1)$ is the concentration parameter, and $S^d=\{y\in\mathbb{R}^{d+1};\|y\|=1\}$ denotes the unit sphere in $\mathbb{R}^{d+1}$. The circular case ($d=1$) is well-known as the wrapped Cauchy or circular Cauchy family (see, for example, [26,35]). In this paper, the distribution (1.1) is called the Cauchy distribution on the sphere or the spherical Cauchy distribution.

McCullagh [35] showed that the wrapped Cauchy family is closed under conformal maps preserving the unit circle which are called the Möbius transformations on the unit circle, and that there are similar induced transformations on the parameter space. Related results about the Cauchy family on the real line and on the Euclidean space have been given by [34] and [28], respectively. A power of a real Cauchy density is transformed into an invariant hyperbolic Laplace

1350-7265

density through the inverse Helgason–Fourier transform which is related in form to the Möbius transformation preserving the upper half plane [16]. To our knowledge, however, there has been no literature about the association between the Möbius transformations and the spherical Cauchy family (1.1). Since there have been various statistical applications of the wrapped Cauchy family and/or the Möbius transformations in directional statistics [6,7,18–20,22,35,41], it is potentially useful to consider the Cauchy family on the sphere and its relationship with the Möbius transformations.

This paper presents some properties of the Cauchy family on the sphere, especially, those related to the Möbius transformations. The spherical Cauchy family is closed under the Möbius transformations on the sphere, and the transformed parameter is given by the extended Möbius transformation on the compactified Euclidean space. The statistical benefits of this property include: (i) an efficient algorithm for random variate generation; (ii) a simple pivotal statistic for parametric inference; (iii) straightforward calculation of probabilities of a region; (iv) closed form expression for the maximum likelihood estimator for $n \leq 3$; and (v) straightforward calculation of the Fisher information matrix. The spherical Cauchy family is a transformation model [2], Sections 1.4 and 2.8, under the action of the Möbius transformations. A method of moments estimator can be expressed in simple form. A simple algorithm for maximum likelihood estimation is available. The likelihood for the spherical Cauchy is equivalent to that for the t-family with a certain degree of freedom which is related to the spherical Cauchy via stereographic projection. An asymptotically efficient estimator is presented which our simulation study suggests outperforms the method of moments estimator and the maximum likelihood estimator in certain settings.

Comparing the densities of the spherical Cauchy and von Mises–Fisher, the spherical Cauchy density takes greater values around the mode and antimode and smaller values in the other area of the sphere. The advantages of the spherical Cauchy over the von Mises–Fisher in terms of properties include the closure under the Möbius transformations and the related properties, while the von Mises–Fisher compares favourably with the spherical Cauchy in terms of its membership in the exponential family, straightforward maximum likelihood estimation and a well-developed theory of hypothesis testing.

Apart from the well-known von Mises–Fisher distribution, there exist other probability distributions on the sphere (see, e.g., Section 2.3 of [29]). Many of these existing models are members of the exponential family including extensions of the von Mises–Fisher distribution. Among non-members of the exponential family, the directional Cauchy distribution has been proposed in [12], Section 5.1. Both the spherical Cauchy (1.1) and the directional Cauchy are multi-dimensional extensions of the wrapped Cauchy distribution for $d = 1$. Also, the densities of both models have higher concentration around a peaked mode than the density of the von Mises–Fisher distribution. A difference between the two Cauchy models is that the directional Cauchy is not closed under the Möbius transformations for $d \geq 2$ and most of the discussion given in this paper is not directly applicable to the directional Cauchy for $d \geq 2$.

Throughout this paper, d is a positive integer. We let $\mathbb{R}^d$ and $\overline{\mathbb{R}}^d$ denote the d-dimensional Euclidean space and compactified Euclidean space $\mathbb{R}^d \cup \{\infty\}$, respectively. Suppose that $\| \cdot \|$ is the Euclidean norm and that S^d is the d-dimensional unit sphere in $\mathbb{R}^{d+1}$, namely, $S^d = \{y \in \mathbb{R}^{d+1}; \|y\| = 1\}$. Let D^{d+1} and $\overline{D}^{d+1}$ denote the open and closed unit balls in $\mathbb{R}^{d+1}$, so that $S^d =$

$\overline{D}^{d+1} \setminus D^{d+1}$. We let $\mathbb{C}$ and $\overline{\mathbb{C}}$ denote the complex plane and the compactified complex plane $\mathbb{C} \cup \{\infty\}$, respectively. The set of all $(d+1) \times (d+1)$ rotation matrices is denoted by $\mathrm{SO}(d+1)$. The $(d+1)$-dimensional unit vector whose jth element equals one is e_j. The $(d+1) \times (d+1)$ identity matrix is denoted by I.

The paper is organized as follows. In Section 2, we introduce the Möbius transformations and a Cauchy family on the sphere. Some statistical properties of the spherical Cauchy family are derived from the Möbius transformations. In Section 3, an extension of the stereographic projection is defined and it is used to discuss the relationship between the spherical Cauchy family and a multivariate t-family with a certain degree of freedom. Three estimators of the parameter for the spherical Cauchy family are proposed in Section 4. In Section 5, a simulation study is conducted to compare the performance of the proposed estimators. Finally, the spherical Cauchy family is compared with the von Mises–Fisher family in Section 6. A marginal distribution of the spherical Cauchy family is discussed in Appendix A.

Proofs can be found in the online Supplementary Material [21].

2. Möbius transformations and a Cauchy family on the sphere

2.1. Möbius transformations on the sphere

The goal of this section is to discuss the Möbius transformations preserving the unit sphere S^d, and to investigate their association with the Cauchy family (1.1). The first step to achieve this is to consider the following function

$$\tilde{\mathcal{M}}_{R,\psi}(y) = R\left\{\frac{1 - \|\psi\|^2}{\|y+\psi\|^2}(y+\psi) + \psi\right\}, \quad y \in S^d, \tag{2.1}$$

where $\psi \in \mathbb{R}^{d+1} \setminus S^d$ and $R \in \mathrm{SO}(d+1)$. The transformation (2.1) maps the unit sphere onto itself: it is called the Möbius transformation on the sphere. This transformation is a special case of the Möbius transformation on $\overline{\mathbb{R}}^{d+1}$ with a restricted domain. (For the Möbius transformations on $\overline{\mathbb{R}}^{d+1}$, see, for example, Section 2 of [17] and equation (2.7) below.)

The Möbius transformation on the sphere (2.1) is an extension of the function given in Section 10 of [33] which has the restriction $R = I$, $\psi = \psi_1 e_1$ and $-1 < \psi_1 < 1$. That function was derived to obtain the reflected point $\tilde{M}_{I,\psi_1 e_1}(y)$ by extending a chord from $-y$ through the point $\psi_1 e_1$ to intersect the unit sphere. Similarly, the Möbius transformation on the sphere (2.1) first sends $y \in S^d$ to the point $\tilde{M}_{I,\psi}(y)$ which intersects the unit sphere and the chord extended from $-y$ through ψ. Then the transformed point $\tilde{M}_{R,\psi}(y)$ is obtained by rotating $\tilde{M}_{I,\psi}(y)$ using R.

Remark 2.1. An alternative derivation of the Möbius transformation on the sphere (2.1) with $\psi \neq 0$ is given as follows:

(i) First, rotate y to obtain $u = Py$, where $P \in \mathrm{SO}(d+1)$ is defined to satisfy $P\psi = (\|\psi\|, 0, \ldots, 0)^T$ and $Py = (y^T\psi/\|\psi\|, \{\|y\|^2 - (y^T\psi/\|\psi\|)^2\}^{1/2}, 0, \ldots, 0)^T$. Note that the first row of P is $\psi^T/\|\psi\|$ and, for $y \neq \pm\psi/\|\psi\|$, the second row of P is $\{y - (y^T\psi/\|\psi\|^2)\psi\}^T/\|y - (y^T\psi/\|\psi\|^2)\psi\|$.

(ii) Second, define the complex number $z = u_1 + iu_2$, where u_1 and u_2 are the first and second components of u, respectively, and transform z via the Möbius transformation preserving the unit circle

$$\breve{\mathcal{M}}_c(z) = \frac{z + \|\psi\|}{\|\psi\| z + 1}. \tag{2.2}$$

Then put $\breve{u} = (\mathrm{Re}[\breve{\mathcal{M}}_c(z)], \mathrm{Im}[\breve{\mathcal{M}}_c(z)], 0, \ldots, 0)^T$.

(iii) Finally, the Möbius transformation on the unit sphere (2.1) is obtained by rotating $\breve{u}$ using RP^T, namely, $\tilde{\mathcal{M}}_{R,\psi}(y) = RP^T \breve{u}$.

As will be seen in the next subsection, we will extend the domain of the transformation (2.2) from the unit circle to the compactified complex plane $\overline{\mathbb{C}}$ in order to derive an extension of $\tilde{\mathcal{M}}_{R,\psi}$.

The two parameters R and ψ have a clear interpretation. The matrix R works as a rotation parameter. In order to discuss the interpretation of ψ, assume, without loss of generality, that $R = I$. If $\|\psi\| < 1$, ψ can be interpreted as a parameter vector that attracts the points on the sphere towards $\psi/\|\psi\|$, with the concentration of the points around $\psi/\|\psi\|$ increasing as $\|\psi\|$ increases. In particular, if $\psi = 0$, then $\tilde{\mathcal{M}}_{I,\psi}$ reduces to the identify mapping. As $\|\psi\| \to 1$, $\tilde{\mathcal{M}}_{I,\psi}(y) \to \psi/\|\psi\|$ for any $y \neq -\psi/\|\psi\|$. The points $y = \psi/\|\psi\|$ and $y = -\psi/\|\psi\|$ are invariant under $\tilde{\mathcal{M}}_{I,\psi}$, i.e., $\tilde{\mathcal{M}}_{I,\psi}(\psi/\|\psi\|) = \psi/\|\psi\|$ and $\tilde{\mathcal{M}}_{I,\psi}(-\psi/\|\psi\|) = -\psi/\|\psi\|$ for any $\psi \neq 0$. For the case of $\|\psi\| > 1$, the transformation $\tilde{\mathcal{M}}_{I,\psi}$ consists of the two steps of transformations, namely, the transformation $\tilde{\mathcal{M}}_{I,\psi/\|\psi\|^2}(y)(= y')$ and the reflection of $-y'$ along ψ.

2.2. An extension of the Möbius transformations on the sphere

In this subsection, we discuss a set of functions which is an extension of the Möbius transformations on the sphere (2.1) and is related to the Cauchy family on the sphere. The function is defined by

$$\mathcal{M}_{R,\psi}(x) = R\left\{\frac{1 - \|\psi\|^2}{\|\check{x} + \psi\|^2}(\check{x} + \psi) + \psi\right\}, \quad x \in \mathbb{R}^{d+1} \setminus \{0, -\psi/\|\psi\|^2\}. \tag{2.3}$$

where $\check{x} = x/\|x\|^2$, $\psi \in \mathbb{R}^{d+1} \setminus S^d$, and $R \in \mathrm{SO}(d+1)$. Also, we define $\mathcal{M}_{R,\psi}(0) = R\psi$, $\mathcal{M}_{R,\psi}(-\psi/\|\psi\|^2) = \infty$ and $\mathcal{M}_{R,\psi}(\infty) = R\psi/\|\psi\|^2$.

If we restrict the domain of x to be S^d, then $\mathcal{M}_{R,\psi}$ reduces to the Möbius transformation $\tilde{\mathcal{M}}_{R,\psi}$ on the sphere. For $\psi \neq 0$, the transformation $\mathcal{M}_{R,\psi}$ can also be expressed as

$$\mathcal{M}_{R,\psi}(x) = RT_{\check{\psi}}\left\{\frac{1 - \|\check{\psi}\|^2}{\|x + \check{\psi}\|^2}(x + \check{\psi}) + \check{\psi}\right\}, \quad x \in \mathbb{R}^{d+1} \setminus \{-\check{\psi}\}, \tag{2.4}$$

where $\check{\psi} = \psi/\|\psi\|^2$ and $T_{\check{\psi}} = 2\check{\psi}\check{\psi}^T/\|\check{\psi}\|^2 - I$. Throughout the paper, the transformation (2.3) is denoted by $\mathcal{M}_{R,\psi}$.

The function $\mathcal{M}_{R,\psi}$ with $\psi \neq 0$ can be derived by replacing $y \in S^d$ with $x \in \mathbb{R}^{d+1} \setminus \{-\check{\psi}\}$ in the three steps (i)–(iii) given in Remark 2.1. Notice that, in the step (ii), the Möbius transformation on the complex plane (2.2) can be expressed as

$$\check{\mathcal{M}}_c(z) = \frac{1-\|\psi\|^2}{|\check{z}+\|\psi\||^2}\big(\check{z}+\|\psi\|\big)+\|\psi\| = \overline{\frac{1-\|\psi\|^{-2}}{|z+\|\psi\|^{-1}|^2}\big(z+\|\psi\|^{-1}\big)+\|\psi\|^{-1}},$$

where $\check{z} = z/|z|^2$. The middle and last expressions above lead to the expressions for $\mathcal{M}_{R,\psi}$ given by (2.3) and (2.4), respectively. If $\psi = 0$, then $\mathcal{M}_{R,0}(x) = Rx$.

In a more intuitively accessible manner, the transformation $\mathcal{M}_{R,\psi}$ can be induced as follows. The function

$$\mathcal{M}_c(z) = \alpha_R \frac{z+\alpha_\psi}{\overline{\alpha}_\psi z + 1}, \quad z \in \overline{\mathbb{C}},$$

is a Möbius transformation $\overline{\mathbb{C}} \to \overline{\mathbb{C}}$ that preserves the unit circle, where α_R and α_ψ are complex numbers such that $|\alpha_R| = 1$ and $|\alpha_\psi| \neq 1$. (See, e.g., Section 3.3 of [1] and Chapter 12 of [39] for the the Möbius transformation on $\overline{\mathbb{C}}$, especially, the latter for the Möbius transformation preserving the unit circle.) The inside of the unit circle is preserved for $|\alpha_\psi| < 1$ and the inside is mapped to the outside for $|\alpha_\psi| > 1$. Note that $\mathcal{M}_c$ can also be expressed as

$$\mathcal{M}_c(z) = \alpha_R \left\{ \frac{1-|\alpha_\psi|^2}{|\check{z}+\alpha_\psi|^2}(\check{z}+\alpha_\psi)+\alpha_\psi \right\}, \quad z \in \overline{\mathbb{C}} \setminus \big\{0, -\alpha_\psi/|\alpha_\psi|^2\big\}. \tag{2.5}$$

If follows from this expression that the transformation (2.5) is essentially the same as $\mathcal{M}_{R,\psi}$ with $d = 1$ apart from the parametrization if the real and imaginary parts of $\mathcal{M}_c(z)$ are identified as the first and second components of $\mathcal{M}_{R,\psi}(x)$, respectively. Then it can be readily induced from this expression that $\mathcal{M}_c$ extends to the mapping $\mathcal{M}_{R,\psi}$ expressed in form of (2.3).

Theorem 2.1. *The following hold for the transformation* $\mathcal{M}_{R,\psi}$:

(a) *The transformation* $\mathcal{M}_{R,\psi}$ *is a bijective conformal map which maps* $\overline{\mathbb{R}}^{d+1}$ *onto itself.*
(b) *For any* $\psi \in \overline{\mathbb{R}}^{d+1} \setminus S^d$, *the transformation* $\mathcal{M}_{R,\psi}$ *maps the unit sphere* S^d *onto itself.*
(c) *If* $\|\psi\| < 1$, *then* $\mathcal{M}_{R,\psi}(D^{d+1}) = D^{d+1}$ *and* $\mathcal{M}_{R,\psi}(\overline{\mathbb{R}}^{d+1} \setminus \overline{D}^{d+1}) = \overline{\mathbb{R}}^{d+1} \setminus \overline{D}^{d+1}$.
(d) *If* $\|\psi\| > 1$, *then* $\mathcal{M}_{R,\psi}(D^{d+1}) = \overline{\mathbb{R}}^{d+1} \setminus \overline{D}^{d+1}$ *and* $\mathcal{M}_{R,\psi}(\overline{\mathbb{R}}^{d+1} \setminus \overline{D}^{d+1}) = D^{d+1}$.

It is remarked that, if $\tilde{\mathcal{M}}_{R,\psi}$ in (2.1) is directly extended to be defined on $\overline{\mathbb{R}}^{d+1}$, then the properties (c) and (d) of Theorem 2.1 do not hold for $\tilde{\mathcal{M}}_{R,\psi}$ in general. This extended function is an anti-Möbius transformation which maps D^{d+1} to the outside for $\|\psi\| < 1$ and preserves D^{d+1} for $\|\psi\| > 1$.

The following holds for the set of the Möbius transformations $\{\mathcal{M}_{R,\psi}\}$ with $\psi \neq 0$, whose elements can be expressed as (2.4).

Lemma 2.1. *Let $\psi_1, \psi_2 \neq 0$. Then, for $\psi_2 \neq -R_1\psi_1$,*

$$\mathcal{M}_{R_2,\psi_2}\{\mathcal{M}_{R_1,\psi_1}(x)\} = \mathcal{M}_{R_{12},\psi_{12}}(x), \quad x \in \overline{\mathbb{R}}^{d+1}, \tag{2.6}$$

where $R_{12} = R_2 T_{\psi_2} T_{R_1\check{\psi}_1+\check{\psi}_2} R_1 T_{\psi_1} T_{\psi_{12}}$, $\psi_{12} = \mathcal{M}_{I,\psi_1}(R_1^T\check{\psi}_2)/\|\mathcal{M}_{I,\psi_1}(R_1^T\check{\psi}_2)\|^2 = T_{\psi_1} R_1^T \times T_{R_1\check{\psi}_1+\check{\psi}_2} T_{\psi_2}\mathcal{M}_{I,\psi_2}(R_1\psi_1)$, $\check{\psi}_j = \psi_j/\|\psi_j\|^2$, $j = 1,2$, and T_ψ is defined as in (2.4). For $\psi_2 = -R_1\psi_1$, $\mathcal{M}_{R_2,\psi_2}\{\mathcal{M}_{R_1,\psi_1}(x)\} = \mathcal{M}_{R_2R_1,0}(x)$ for any $x \in \overline{\mathbb{R}}^{d+1}$.

In addition the expression (2.3) implies that $\mathcal{M}_{R_2,\psi_2}\{\mathcal{M}_{R_1,\psi_1}(x)\} = \mathcal{M}_{R_2R_1,R_1^T\psi_2}(x)$ for $\psi_1 = 0$ and $\mathcal{M}_{R_2,\psi_2}\{\mathcal{M}_{R_1,\psi_1}(x)\} = \mathcal{M}_{R_2R_1,\psi_1}(x)$ for $\psi_2 = 0$. Thus, we have the following theorem.

Theorem 2.2. *Let $\mathcal{G}$ be the set of the transformations $\{\mathcal{M}_{R,\psi}\}$ with all possible combinations of $R \in \mathrm{SO}(d+1)$ and $\psi \in \mathbb{R}^{d+1} \setminus S^d$, namely, $\mathcal{G} = \{\mathcal{M}_{R,\psi};\, R \in \mathrm{SO}(d+1),\, \psi \in \mathbb{R}^{d+1} \setminus S^d\}$. Then $\mathcal{G}$ forms a group under composition. The identity element of $\mathcal{G}$ is $\mathcal{M}_{I,0}$ and the inverse of $\mathcal{M}_{R,\psi}$ is $\mathcal{M}_{R^T,-R\psi}$.*

This theorem implies that the Möbius transformations $\mathcal{G}$ are tractable transformations for spherical variables in terms of group operation.

It follows from Theorem 2.2 that $\mathcal{G}$ is a subgroup of the Möbius group on $\overline{\mathbb{R}}^{d+1}$. Here the Möbius group on $\overline{\mathbb{R}}^{d+1}$ is defined by $\mathcal{G}' = \{\mathcal{M}';\, a, b \in \mathbb{R}^{d+1},\ \gamma \in \mathbb{R} \setminus \{0\},\ A \in O(d+1),\ \varepsilon \in \{0,2\}\}$, where

$$\mathcal{M}'(x) = A\left(\gamma \frac{x+a}{\|x+a\|^\varepsilon} + b\right), \quad x \in \mathbb{R}^{d+1} \setminus \{-a\}, \tag{2.7}$$

and $O(d+1)$ is the set of all $(d+1) \times (d+1)$ orthogonal matrices. If $x \in \{-a, \infty\}$, we define $\mathcal{M}'(-a) = Ab$ and $\mathcal{M}'(\infty) = \infty$ for $\varepsilon = 0$ and $\mathcal{M}'(-a) = \infty$ and $\mathcal{M}'(\infty) = b$ for $\varepsilon = 2$. See, for example, Section 2 of [17] for more details of the Möbius group on $\overline{\mathbb{R}}^{d+1}$.

It is clear from Theorem 2.2 that the set of Möbius transformations $\{\tilde{\mathcal{M}}_{R,\psi}\}$ on the sphere also forms a group under composition. The set of transformations $\{\mathcal{M}_{R,\psi}\}$ is not an Abelian group, implying that that $\mathcal{M}_{R_1,\psi_1}\{\mathcal{M}_{R_2,\psi_2}(x)\} = \mathcal{M}_{R_2,\psi_2}\{\mathcal{M}_{R_1,\psi_1}(x)\}$ does not hold in general. However, for fixed $\mu \in S^d$, the subset of transformations $\{\mathcal{M}_{I,\rho\mu};\, |\rho| \neq 1\}$ forms an Abelian group. Since any subgroup of an Abelian group is Abelian, the group of the Möbius transformations $\{\tilde{\mathcal{M}}_{I,\rho\mu};\, |\rho| \neq 1\}$ on the sphere is also Abelian.

2.3. A Cauchy family on the sphere

The parameters μ and ρ of the spherical Cauchy family (1.1) can be clearly interpreted. The parameter μ controls the mode of the density. The concentration of the distribution is regulated by ρ. The greater the value of ρ, the greater the concentration of the density (1.1) around the mode. In particular, when $\rho = 0$, the distribution (1.1) reduces to the uniform distribution on S^d. On the other hand, as ρ tends to 1, the distribution converges to a point distribution with singularity at $y = \mu$. See Figure 2 given in Section 6 for some plots of the densities of the spherical Cauchy (1.1).

In order to investigate the relationship between the spherical Cauchy family (1.1) and the set of transformations (2.3), it is advantageous to write the parameters of the spherical Cauchy (1.1) as $\phi = \rho\mu$ and extend the parameter space to be $\overline{\mathbb{R}}^{d+1}$. Specifically, we write the density of the spherical Cauchy as

$$f(y;\phi) = \frac{\Gamma\{(d+1)/2\}}{2\pi^{(d+1)/2}} \left(\frac{|1-\|\phi\|^2|}{\|y-\phi\|^2} \right)^d, \quad y \in S^d, \tag{2.8}$$

where $\phi \in \mathbb{R}^{d+1} \setminus S^d$. For $\phi \in S^d$ we assume that the distribution is a point mass at $y = \phi$. Also define that the density is uniform if $\phi = \infty$. It can be seen that $f(y;\phi) = f(y;\phi/\|\phi\|^2)$ for any ϕ. Write $Y \sim C_d^*(\phi)$ if an S^d-valued random vector Y has density (2.8).

The following result can be readily established from Theorem 2.1 and Lemma 2.1.

Theorem 2.3. *The following hold for the spherical Cauchy family* (2.8) *and the transformation* $\mathcal{M}_{R,\psi}$:

$$Y \sim C_d^*(\phi) \quad \Longrightarrow \quad \mathcal{M}_{R,\psi}(Y) \sim C_d^*\{\mathcal{M}_{R,\psi}(\phi)\}.$$

If $d = 1$, Theorem 2.3 is essentially the same as the result for the circular Cauchy or wrapped Cauchy family given in [35]. Regarding the parameter space, Theorem 2.1(c) implies that, if $\|\phi\| < 1$, then $\|\mathcal{M}_{R,\psi}(\phi)\| < 1$ for $\|\psi\| < 1$ and $\|\mathcal{M}_{R,\psi}(\phi)\| > 1$ for $\|\psi\| > 1$. A similar discussion can be made for $\|\phi\| > 1$ by applying Theorem 2.1(d).

Theorems 2.2 and 2.3 imply the following result.

Corollary 2.1. *The spherical Cauchy family* (2.8) *is a transformation model under the action of the Möbius group $\mathcal{G}$ in the sense of Section* 2.8 *of* [2], *where $\mathcal{G}$ is defined as in Theorem* 2.2.

See also Section 1.4 of [2] for the transformation models. The spherical Cauchy family (2.8) is the orbit of the uniform distribution on S^d under the action of the Möbius group $\mathcal{G}$.

There are some other statistical applications of Theorem 2.3. For example, this theorem can be applied to propose an efficient algorithm to generate a random variate following the Cauchy family on S^d.

Corollary 2.2. *If a random vector U follows the uniform distribution on S^d, then $\mathcal{M}_{I,\phi}(U)$ has the Cauchy distribution $C_d^*(\phi)$ on the sphere.*

Note that it immediately follows from this corollary that the integral of the density (2.8) over S^d is equal to one.

Theorem 2.2 guarantees that, for any $\phi \notin S^d$, there always exists a Möbius transformation which transforms the spherical Cauchy $C^*(\phi)$ into the uniform distribution on the sphere. Such a Möbius transformation can be obtained from Theorem 2.3, and it can be applied to constructing a pivotal statistic for ϕ and calculating probabilities of a surface area under the spherical Cauchy density (2.8).

Corollary 2.3. *Suppose $Y \sim C^*(\phi)$. Then $\mathcal{M}_{I,-\phi}(Y)$ is a pivotal statistic for ϕ.*

Corollary 2.4. *Let $f(y;\phi)$ denote the spherical Cauchy density (2.8). Assume $A \subset S^d$. Then*

$$\int_A f(y;\phi)\,dy = \frac{\mathrm{Area}\{\mathcal{M}_{I,-\phi}(A)\}}{\mathrm{Area}(S^d)},$$

where Area(C) *denotes the area of C with respect to the surface measure.*

The proofs of Corollaries 2.2–2.4 are straightforward from Theorem 2.3 and omitted.

Theorem 2.2 implies that the inverse function of a Möbius transformation is unique. However, as can be seen in Corollary 2.3, the Möbius transformation transforming a spherical Cauchy into the spherical uniform is not unique because the spherical uniform is invariant under rotation.

3. Extended stereographic projection

In this section, we consider a transformation of the Cauchy family on the sphere (2.8) via the stereographic projection. The stereographic projection $S^d \to \overline{\mathbb{R}}^d$ is known to be

$$\tilde{\mathcal{P}}(y) = \frac{1}{1-y_{d+1}}(y_1,\ldots,y_d)^T, \quad y \in S^d \setminus \{e_{d+1}\}. \tag{3.1}$$

Also define $\tilde{\mathcal{P}}(e_{d+1}) = \infty$. It is known that the stereographic projection (3.1) maps the unit sphere S^d onto $\overline{\mathbb{R}}^d$. A geometrical interpretation of (3.1) is that $\tilde{\mathcal{P}}(y)$ corresponds to the point at the intersection of the embedded Euclidean space $\overline{\mathbb{R}}^d \times \{0\}$ and the line connecting y and the north pole e_{d+1}.

In order to discuss the transformation of the spherical Cauchy family (2.8) via the stereographic projection (3.1), we define an extension of the complex numbers to higher dimensions and an extended stereographic projection.

Definition 3.1. We define an extension of the complex numbers by

$$\theta = \mu + i\sigma,$$

where $\mu \in \mathbb{R}^d$, $\sigma \in \mathbb{R}$ and i is the square root of -1. We write $\mu + i\sigma = \mu$ if $\sigma = 0$.

Definition 3.2. We define a function on $\overline{\mathbb{R}}^{d+1}$ by

$$\mathcal{P}(x) = 2\frac{(x_1,\ldots,x_d)^T}{\|x-e_{d+1}\|^2} + i\frac{1-\|x\|^2}{\|x-e_{d+1}\|^2}, \quad x \in \mathbb{R}^{d+1} \setminus \{e_{d+1}\}. \tag{3.2}$$

Also, $\mathcal{P}(\infty) = -i$ and $\mathcal{P}(e_{d+1}) = \infty$.

Theorem 3.1. *The following hold for the function $\mathcal{P}$.*

(i) *The function $\mathcal{P}$ is a bijective function which maps $\overline{\mathbb{R}}^{d+1}$ onto $(\mathbb{R}^d + i\mathbb{R}) \cup \{\infty\}$.*
(ii) *The function $\mathcal{P}$ reduces to $\tilde{\mathcal{P}}$ if $x \in S^d$.*
(iii) *If $\|x\| < 1$ ($\|x\| > 1$), then the imaginary part of $\mathcal{P}(x)$ is positive (negative).*

A Möbius transformation, which maps $\mathbb{R}^d \times \mathbb{R}^-$ onto D^{d+1} and is related to the function (3.2), is given in equation (4.4.2) of [38]. Theorem 3.1 can also be proved by transforming the function (3.2) and utilizing the results for the Möbius transformation of [38].

Theorem 3.1 implies that there exists the inverse function of (3.2) which is

$$\mathcal{P}^{-1}(\theta) = \frac{2}{\|\mu\|^2 + (1+\sigma)^2} \left(\mu^T, \frac{\|\mu\|^2 + \sigma^2 - 1}{2} \right)^T,$$

where $\theta = \mu + i\sigma \in (\mathbb{R}^d + i\mathbb{R}) \setminus \{-i\}$. Also, define $\mathcal{P}^{-1}(-i) = \infty$ and $\mathcal{P}^{-1}(\infty) = e_{d+1}$. Then the following result is established.

Theorem 3.2. *The following hold for the spherical Cauchy family (2.8) and the extended stereographic projection $\mathcal{P}$:*

$$Y \sim C_d^*(\phi) \quad \Longrightarrow \quad \mathcal{P}(Y) \sim C_d\big\{\mathcal{P}(\phi)\big\}.$$

Equivalently,

$$X \sim C_d(\theta) \quad \Longrightarrow \quad \mathcal{P}^{-1}(X) \sim C_d^*\big\{\mathcal{P}^{-1}(\theta)\big\}.$$

Here $C_d(\theta)$ denotes a d-variate t-distribution with d degrees of freedom with location parameter μ and scale matrix $d^{-1}\sigma^2 I$, which has density

$$f(x;\theta) = \frac{\Gamma(d)}{(\pi\sigma^2)^{d/2}\Gamma(d/2)} \left(1 + \frac{\|x-\mu\|^2}{\sigma^2} \right)^{-d}, \quad x \in \overline{\mathbb{R}}^d, \tag{3.3}$$

where $\theta = \mu + i\sigma$, $\mu \in \mathbb{R}^d$ and $\sigma \neq 0$. For $\sigma = 0$, we interpret the distribution with density (3.3) as a point mass at $x = \mu$. If $\theta = \infty$, then the distribution is assumed to be the point distribution with singularity at $x = \infty$.

This is an extended result of [8] that the uniform distribution on S^d is transformed into the standard multivariate t-distribution with d degrees of freedom. This theorem and Theorem 2.3 imply that a random variate following the t-distribution with d degrees of freedom $C_d(\theta)$ can be generated from the uniform distribution on S^d.

4. Statistical inference

4.1. Method of moments estimation

In statistical inference for rotationally symmetric distributions, it is common to express their parameters in terms of one unit vector (as the mean direction) and one positive scalar (as the concentration parameter). However, for the spherical Cauchy family, it is more convenient and often more interpretable to express its parameter in terms of ϕ as in (2.8) rather than μ and ρ as in (1.1). Nonetheless, it is possible to discuss the statistical inference for the parameters μ and ρ in a similar manner as will be seen in this section.

Throughout this section, we assume that $Y_1, \dots, Y_n$ is a random sample from the multivariate Cauchy distribution $C_d^*(\phi)$ on the sphere with $\|\phi\| < 1$.

Theorem 4.1. *Let Y have the spherical Cauchy distribution $C_d^*(\phi)$ with $\|\phi\| < 1$. Then, for $\phi \neq 0$,*

$$\mathbb{E}(Y) = \eta_{1,d}\big(\|\phi\|\big)\frac{\phi}{\|\phi\|},$$

$$\mathbb{E}\big(YY^T\big) = \frac{1}{d}\left[\big\{1 - \eta_{2,d}\big(\|\phi\|\big)\big\}I + \big\{(d+1)\eta_{2,d}\big(\|\phi\|\big) - 1\big\}\frac{\phi\phi^T}{\|\phi\|^2}\right],$$

where

$$\eta_{1,d}(\rho) = \frac{1+\rho^2}{2\rho}\left[1 - \frac{(1+\rho)^2}{1+\rho^2}F\left\{1, \frac{d}{2}; d; \frac{-4\rho}{(1-\rho)^2}\right\}\right],$$

$$\begin{aligned}\eta_{2,d}(\rho) = \frac{(1+\rho^2)^2}{4\rho^2}&\left[1 - 2\frac{(1+\rho)^2}{1+\rho^2}F\left\{1, \frac{d}{2}; d; \frac{-4\rho}{(1-\rho)^2}\right\}\right.\\ &\left.+ \frac{(1+\rho)^4}{(1+\rho^2)^2}F\left\{2, \frac{d}{2}; d; \frac{-4\rho}{(1-\rho)^2}\right\}\right],\end{aligned}$$

and F denotes the hypergeometric series [[13], *equation* (9.111)]. *If $\phi = 0$, $\mathbb{E}(Y) = 0$ and $\mathbb{E}(YY^T) = (d+1)^{-1}I$.*

As can be seen in the Supplementary Material [21], the proof of this theorem can be partly simplified using the general result for rotationally symmetric distributions given in Section 3.4 of [42]. This theorem and Theorems A.3 and A.4 of Appendix A imply that $\mathbb{E}(Y)$ and $\mathbb{E}(YY^T)$ can be expressed in closed form without hypergeometric functions for any d.

A method of moments estimator of ϕ is obtained by equating the expectation of Y and its sample analogue. In other words the method of moments estimator is the solution of the equation

$$\eta_{1,d}\big(\|\phi\|\big)\frac{\phi}{\|\phi\|} = \overline{Y}, \tag{4.1}$$

where $\eta_{1,d}(\|\phi\|)$ is defined as in Theorem 4.1 and $\overline{Y} = n^{-1}\sum_{j=1}^n Y_j$. As is clear from Lemma A.1 of Appendix A, it holds that $\eta_{1,d}(0) = 0$, $\lim_{\|\phi\|\to 1} \eta_{1,d}(\|\phi\|) = 1$, and $\eta_{1,d}(\|\phi\|)$ is monotonically increasing with respect to $\|\phi\|$. This immediately leads to the following theorem.

Theorem 4.2. *The equation* (4.1) *has the unique solution* $\hat{\phi}_{\mathrm{MM}}$ *on the* $(d+1)$*-dimensional unit disc*

$$\hat{\phi}_{\mathrm{MM}} = \eta_{1,d}^{-1}\big(\|\overline{Y}\|\big)\frac{\overline{Y}}{\|\overline{Y}\|}, \tag{4.2}$$

where $\eta_{1,d}^{-1}(\rho)$ *is the inverse of* $\eta_{1,d}(\rho)$ *for* $0 \le \rho < 1$.

Since $\eta_{1,d}$ is monotonically increasing, the method of moments estimate $\hat{\phi}_{\mathrm{MM}}$ can be estimated numerically via usual optimization algorithms.

Theorem 4.3. *Let* $\hat{\phi}_{\mathrm{MM}}$ *be the method of moments estimator* (4.2)*. Then* $\sqrt{n}(\hat{\phi}_{\mathrm{MM}} - \phi)$ *tends in distribution to* $N(0, \Lambda\Sigma\Lambda)$ *as* $n \to \infty$, *where*

$$\Sigma = d^{-1}\left[\big\{1 - \eta_{2,d}\big(\|\phi\|\big)\big\}I + \big\{(d+1)\eta_{2,d}\big(\|\phi\|\big) - 1 - d\eta_{1,d}^2\big(\|\phi\|\big)\big\}\frac{\phi\phi^T}{\|\phi\|^2}\right],$$

$$\Lambda = \eta_{1,d}^{-1\,\prime}\big\{\eta_{1,d}\big(\|\phi\|\big)\big\}\frac{\phi^T\phi}{\|\phi\|^2} + \frac{\|\phi\|}{|\eta_{1,d}(\|\phi\|)|}\left(I - \frac{\phi^T\phi}{\|\phi\|^2}\right),$$

$$\eta_{1,d}^{-1\,\prime}\big\{\eta_{1,d}\big(\|\phi\|\big)\big\} = \frac{d+1}{2d}\frac{(1-\|\phi\|)^3}{1+\|\phi\|}\left\{F\left(2, \frac{d}{2}+1; d+2; -\frac{4\|\phi\|}{(1-\|\phi\|)^2}\right)\right\}^{-1}.$$

Using the expression of the density (1.1), the method of moments estimators of μ and ρ are given by

$$\hat{\mu}_{\mathrm{MM}} = \frac{\overline{Y}}{\|\overline{Y}\|}, \qquad \hat{\rho}_{\mathrm{MM}} = \eta_{1,d}^{-1}\big(\|\overline{Y}\|\big),$$

respectively. The asymptotic normality of these estimators can be proved in a similar manner as in Theorem 4.3. Note that $\hat{\mu}_{\mathrm{MM}}$ is known as the extrinsic sample mean of spherical random vectors. The asymptotic normality of the extrinsic sample mean holds not only for the random vectors following the spherical Cauchy but also for random vectors taking values in rather general manifolds under certain conditions. (Any bounded subset would suffice.) See Theorem 3.6 of [3] and Theorem 2 of [14] for details.

4.2. Maximum likelihood estimation

As the following theorem shows, maximum likelihood estimation for the Cauchy distribution on the sphere and that for the multivariate t-distribution with a certain degree of freedom are equivalent.

Theorem 4.4. *Let* $Y_1, \ldots, Y_n$ *be an i.i.d. sample from the spherical Cauchy distribution* $C_d^*(\phi)$. *Suppose that* $\mathcal{P}$ *is the function* (3.2)*. Then the maximum likelihood estimator of* ϕ *is equal to* $\mathcal{P}^{-1}(\hat{\theta})$, *where* $\hat{\theta}$ *is the maximum likelihood estimator of the* d*-variate* t*-distribution with* d *degrees of freedom* $C_d(\theta)$, *given in* (3.3), *for the sample* $\mathcal{P}(Y_1), \ldots, \mathcal{P}(Y_n)$.

The proof is clear from Theorems 3.1 and 3.2 and omitted. This theorem implies that, in order to estimate the parameter of the spherical Cauchy, it suffices to estimate the parameter of the d-variate t-distribution with d degrees of freedom.

Although Theorem 4.4 is helpful for computing the maximum likelihood estimates of the parameter, there remain various properties of the maximum likelihood estimator which are not clear from this theorem. The rest of this subsection is devoted to investigating properties of the maximum likelihood estimator which are not clear from Theorem 4.4. The loglikelihood function is

$$\ell(\phi) = \sum_{j=1}^{n} \log f(y_j;\phi) = C + d\left\{ n \log\bigl(1 - \|\phi\|^2\bigr) - \sum_{j=1}^{n} \log\bigl(1 + \|\phi\|^2 - 2\phi^T y_j\bigr) \right\}, \quad (4.3)$$

where $C = n\log\Gamma\{(d+1)/2\} - n\log\{2\pi^{(d+1)/2}\}$. The first derivative of the loglikelihood function with respect to ϕ is

$$\frac{\partial \ell}{\partial \phi} = -2d\left(\frac{n\phi}{1-\|\phi\|^2} + \sum_{j=1}^{n} \frac{\phi - y_j}{1 + \|\phi\|^2 - 2\phi^T y_j} \right) = \frac{2d}{1-\|\phi\|^2} \sum_{j=1}^{n} \mathcal{M}_{I,-\phi}(y_j), \quad (4.4)$$

where $\mathcal{M}_{R,\psi}$ is as in (2.3). Therefore the estimating equation for ϕ has a simple form

$$\sum_{j=1}^{n} \mathcal{M}_{I,-\phi}(y_j) = 0.$$

Theorem 4.5. *For $n \le 3$, the maximum likelihood estimator of ϕ, $\hat{\phi}_{\mathrm{ML}}$, can be expressed as follows.*

(i) *For $n = 1$, the maximum likelihood estimator of ϕ is $\hat{\phi}_{\mathrm{ML}} = y_1$.*

(ii) *Suppose $n = 2$. If $y_1 \ne \pm y_2$, the contour of maximum likelihood of ϕ is the circle perpendicular to the unit sphere with chord (y_1, y_2) in the two-dimensional plane spanned by y_1 and y_2. When $y_1 = -y_2$, the contour of maximum likelihood of ϕ is the line connecting y_1 and y_2. If $y_1 = y_2$, then $\hat{\phi}_{\mathrm{ML}} = y_1$.*

(iii) *Assume $n = 3$ and $y_j \ne y_k$ $(j \ne k)$. Then the maximum likelihood estimator of ϕ is*

$$\hat{\phi}_{\mathrm{ML}} = \mathcal{P}^{-1}(\hat{\mu} + i\hat{\sigma}),$$

where $\mathcal{P}$ is defined as in (3.2) and

$$\hat{\mu} = \frac{\|\mathcal{P}(y_1) - \mathcal{P}(y_2)\|^2 \mathcal{P}(y_3) + \|\mathcal{P}(y_2) - \mathcal{P}(y_3)\|^2 \mathcal{P}(y_1) + \|\mathcal{P}(y_3) - \mathcal{P}(y_1)\|^2 \mathcal{P}(y_2)}{\|\mathcal{P}(y_1) - \mathcal{P}(y_2)\|^2 + \|\mathcal{P}(y_2) - \mathcal{P}(y_3)\|^2 + \|\mathcal{P}(y_3) - \mathcal{P}(y_1)\|^2},$$

$$\hat{\sigma} = \sqrt{3}\,\frac{\|\mathcal{P}(y_1) - \mathcal{P}(y_2)\|\,\|\mathcal{P}(y_2) - \mathcal{P}(y_3)\|\,\|\mathcal{P}(y_3) - \mathcal{P}(y_1)\|}{\|\mathcal{P}(y_1) - \mathcal{P}(y_2)\|^2 + \|\mathcal{P}(y_2) - \mathcal{P}(y_3)\|^2 + \|\mathcal{P}(y_3) - \mathcal{P}(y_1)\|^2}.$$

For $d = 1$ and $n = 4$, Ferguson [9] and McCullagh [35] showed the maximum likelihood estimator of $\hat{\phi}_{\mathrm{ML}}$ can be expressed in closed form. However it does not appear clear that there are closed form expressions for the maximum likelihood estimators for $n \geq 4$ for general d.

Lemma 4.1. *Let $f(y)$ be the density (2.8) with $\|\phi\| < 1$. Then the Fisher information matrix is*

$$\mathcal{I} = -\mathbb{E}\left\{ \frac{\partial}{\partial\phi\partial\phi^T} \log f(Y) \right\} = \frac{4}{(1 - \|\phi\|^2)^2} \frac{d^2}{d+1} I. \tag{4.5}$$

Thus the asymptotic variance of the maximum likelihood estimator of ϕ can be expressed in simple form. The proof is clear from the general theory and therefore omitted.

Theorem 4.6. *Let $Y_1, \dots, Y_n$ be a random sample from $C_d^*(\phi)$ with $\|\phi\| < 1$. Assume $\hat{\phi}_{\mathrm{ML}}$ is the maximum likelihood estimator of ϕ. Then $\sqrt{n}(\hat{\phi}_{\mathrm{ML}} - \phi)$ tends in distribution to $N(0, \mathcal{I}^{-1})$ as $n \to \infty$, where $\mathcal{I}^{-1} = (1 - \|\phi\|^2)^2(d+1)/(4d^2)I$.*

As seen in Theorem 4.4, the maximum likelihood estimates for the sample from the spherical Cauchy (2.8) for general sample size can be estimated via the transformation of the spherical Cauchy into the d-variate t with d degrees of freedom. For the maximum likelihood estimation for the multivariate t-distribution with known degrees of freedom, one can apply the EM algorithm (see, e.g., Section 2.6.1 of [36]).

However it would be more efficient if the parameter estimates are obtained directly from the sample without transformation. For $d = 1$, the algorithm of [26] is available to estimate the parameter ϕ. Using the fact that the Fisher information (4.5) and the score function (4.4) for the spherical Cauchy are expressed in simple and closed form, here we present a simple algorithm based on the method of scoring [e.g., [10], Section 20].

Algorithm 4.1.

Step 1. Take an initial value ϕ_0.

Step 2. Compute $\phi_1, \dots, \phi_N$ as follows until the value ϕ_N remains virtually unchanged from the previous value ϕ_{N-1}.

$$\phi_t = \phi_{t-1} + \frac{(d+1)(1 - \|\phi_{t-1}\|^2)}{2dn} \sum_{j=1}^{n} \mathcal{M}_{I,-\phi_{t-1}}(y_j), \quad t = 1, \dots, N.$$

Step 3. Record ϕ_N as the maximum likelihood estimate of ϕ.

The convergence of this algorithm is not proved mathematically. However our simulation study implies that the algorithm converges fast when the method of moments estimate (4.2) is used as the initial value ϕ_0. In addition, for $d = 1$, it seems that the parameter estimates based on Algorithm 4.1 numerically coincide with those based on the algorithm of [26].

The following tractable property holds for stationary points of the loglikelihood function.

Theorem 4.7. *Let $Y_1, \dots, Y_n$ be a random sample from the spherical Cauchy distribution $C^*(\phi)$. Assume that $Y_j \neq Y_k$ for some (j,k). Then any stationary point of the loglikelihood function* (4.3) *is a local maximum.*

In this theorem, if the loglikelihood function (4.3) satisfies certain conditions, it can be seen that there exists only one stationary point in (4.3) and therefore the loglikelihood function (4.3) is unimodal (see [11] for details).

Maximum likelihood estimation for the reparametrized parameters μ and ρ in (1.1) can be discussed in a somewhat similar manner. The maximum likelihood estimators of μ and ρ are essentially the same as $\hat{\phi}_{\mathrm{ML}}$, namely, $\hat{\mu}_{\mathrm{ML}} = \hat{\phi}_{\mathrm{ML}}/\|\hat{\phi}_{\mathrm{ML}}\|$ and $\hat{\rho}_{\mathrm{ML}} = \|\hat{\phi}_{\mathrm{ML}}\|$. The score function and the Fisher information can be readily calculated using the chain rule.

4.3. Asymptotically efficient estimation

Consider the estimator

$$\hat{\phi}_{\mathrm{AE}} = \eta_{1,d}^{-1}(\|\overline{Y}\|)\frac{\overline{Y}}{\|\overline{Y}\|} + \frac{d+1}{2dn}\sum_{j=1}^{n}\mathcal{M}_{I,-\hat{\phi}_{\mathrm{MM}}}(Y_j). \tag{4.6}$$

This estimator is derived as $\hat{\phi}_{\mathrm{AE}} = \hat{\phi}_{\mathrm{MM}} + (n\mathcal{I})^{-1}\partial\ell/\partial\phi$, where $\hat{\phi}_{\mathrm{MM}}$ is the method of moments estimator (4.2), a consistent estimator of ϕ, and $\mathcal{I}$ denotes the Fisher information matrix (4.5) evaluated at $\hat{\phi}_{\mathrm{MM}}$. It follows, for example, from Theorem 20 of [10] that the estimator (4.6) is an asymptotically efficient estimator of ϕ with asymptotic variance $\mathcal{I}^{-1} = (1-\|\phi\|^2)^2(d+1)/(4d^2)I$. The estimator (4.6) also appears as ϕ_1 in Algorithm 4.1 when the method of moments estimator (4.2) is taken as the initial value ϕ_0.

An advantage of the estimator (4.6) is that it achieves both closed-form expression and asymptotic efficiency, whereas the method of moments estimator (4.2) has only closed-form expression and the maximum likelihood estimator only asymptotic efficiency.

5. Simulation study

We compare the method of moments estimator (4.2), the maximum likelihood estimator and the asymptotically efficient estimator (4.6) in terms of their performance for finite sample sizes and their asymptotic behaviour. In order to compare the performance of the three estimators, the mean squared error $\mathrm{MSE} = \mathbb{E}(\|\hat{\phi} - \phi\|^2)$ is considered, where $\hat{\phi}$ is an estimator of ϕ of the spherical Cauchy $C_d^*(\phi)$. We consider the relative mean squared error defined by

$$\mathrm{RMSE}_{\mathrm{E/ML}} = \mathrm{MSE}_{\mathrm{E}}/\mathrm{MSE}_{\mathrm{ML}},$$

where $\mathrm{MSE}_{\mathrm{ML}}$ denotes the MSE of the maximum likelihood estimator and $\mathrm{MSE}_{\mathrm{E}}$ is the MSE of the method of moments estimator (4.2) or the asymptotically efficient estimator (4.6). The standard errors of the MSE of the maximum likelihood estimator, the method of moments estimator (4.2) and the asymptotically efficient estimator (4.6) were also computed.

First, we consider the performance of the three estimators for finite sample sizes via a Monte Carlo simulation study. Random samples of sizes $n = 10,\ 25$ and 100 were generated from the spherical Cauchy $C_d^*(\phi)$ with $\phi/\|\phi\| = e_1,\ \|\phi\| = \eta_{1,d}^{-1}(0.1),\ \eta_{1,d}^{-1}(0.5)$ and $\eta_{1,d}^{-1}(0.9)$ and $d = 1,\ 2,\ 10$ and 100. The values of $\|\phi\|$ are defined such that the mean resultant lengths of the underlying distributions are 0.1, 0.5 and 0.9. For each combination of d, n and $\|\phi\|$, $r = 2000$ random samples were generated using Corollary 2.2. Then the three estimators were computed for each random sample. We used Algorithm 4.1 to estimate the maximum likelihood estimator and the method of moments estimator (4.2) was adopted as the initial value of the algorithm.

An estimate of MSE based on r random samples is defined by $\widehat{\mathrm{MSE}} = r^{-1} \sum_{j=1}^{r} \|\hat{\phi}_j - \phi\|^2$, where $\hat{\phi}_j$ is an estimator of ϕ estimated from the jth random sample $(j = 1, \dots, r)$. We then discuss an estimate of relative mean squared error defined by

$$\widehat{\mathrm{RMSE}}_{\mathrm{E/ML}} = \widehat{\mathrm{MSE}}_{\mathrm{E}} / \widehat{\mathrm{MSE}}_{\mathrm{ML}}, \tag{5.1}$$

where $\widehat{\mathrm{MSE}}_{\mathrm{ML}}$ denotes $\widehat{\mathrm{MSE}}$ of the maximum likelihood estimator and $\widehat{\mathrm{MSE}}_{\mathrm{E}}$ is $\widehat{\mathrm{MSE}}$ of the method of moments estimator (4.2) or the asymptotically efficient estimator (4.6).

Table 1 shows estimates of relative mean squared error (5.1) for some selected combinations of d, n and $\|\phi\|$. The values of the relative mean squared error (5.1) for $n = \infty$ given in the table are those of the asymptotic relative mean squared error, namely, $\lim_{n\to\infty} \mathrm{RMSE}_{\mathrm{E/ML}}$, which can be calculated using Theorems 4.3 and 4.6. Also, Table 1 provides estimates of standard error of the relative mean squared error (5.1). Since the number of simulation samples r is large, we apply the central limit theorem and delta method to obtain estimates of standard error of the relative mean squared error (5.1) as

$$\widehat{\mathrm{SE}}_{\mathrm{E/ML}} = r^{-1/2} \left(\frac{1}{\widehat{\mathrm{MSE}}_{\mathrm{ML}}^{2}} \hat{\sigma}_{\mathrm{E}}^2 - \frac{2\widehat{\mathrm{MSE}}_{\mathrm{E}}}{\widehat{\mathrm{MSE}}_{\mathrm{ML}}^{3}} \hat{\sigma}_{\mathrm{E,ML}} + \frac{\widehat{\mathrm{MSE}}_{\mathrm{E}}^{2}}{\widehat{\mathrm{MSE}}_{\mathrm{ML}}^{4}} \hat{\sigma}_{\mathrm{ML}}^2 \right)^{1/2},$$

where $\hat{\sigma}_{\mathrm{E}}^2$ and $\hat{\sigma}_{\mathrm{ML}}^2$ are the sample variance of $\{\|\hat{\phi}_{\mathrm{E}j} - \phi\|^2\}_{j=1}^{r}$ and $\{\|\hat{\phi}_{\mathrm{ML}j} - \phi\|^2\}_{j=1}^{r}$, respectively, $\hat{\sigma}_{\mathrm{E,ML}}$ is the sample covariance of $\{(\|\hat{\phi}_{\mathrm{E}j} - \phi\|^2, \|\hat{\phi}_{\mathrm{ML}j} - \phi\|^2)\}_{j=1}^{r}$, $\hat{\phi}_{\mathrm{E}j}$ is the method of moments estimate (4.2) or the asymptotically efficient estimate (4.6) estimated from the jth random sample, and $\hat{\phi}_{\mathrm{ML}j}$ is the maximum likelihood estimate estimated from the jth random sample $(j = 1, \dots, r)$.

The table suggests that, for high dimensional cases, that is, $d \geq 10$, the asymptotically efficient estimator (4.6) slightly outperforms the method of moments estimator (4.2) and the maximum likelihood estimator in terms of the mean squared error. For low dimensional cases, that is, $d = 1$ or 2, the asymptotically efficient estimator (4.6) outperforms the other two estimators for small values of $\|\phi\|$ (low concentration) and the maximum likelihood estimator is preferable otherwise. The method of moments estimator (4.2) shows worse performance than the asymptotically efficient estimator (4.6) in all the cases, especially, those of small d and large $\|\phi\|$ (high concentration). As n increases, the value of $\widehat{\mathrm{RMSE}}_{\mathrm{E/ML}}$ for the method of moments estimator (4.2) increases and the value of $\widehat{\mathrm{RMSE}}_{\mathrm{E/ML}}$ for the asymptotically efficient estimator (4.6) approaches one. Estimated values of standard error of $\widehat{\mathrm{RMSE}}_{\mathrm{E/ML}}$ are small, especially, for high dimensional cases, and hence the observations given above seem reasonable ones.

Table 1. Relative mean squared error (RMSE) of the method of moments estimator (4.2) (MM) and that of the asymptotically efficient estimator (4.6) (AE) with respect to the maximum likelihood estimator and, in brackets, standard error (SE) of the RMSE of the MM and that of the AE with respect to the maximum likelihood estimator estimated from 2000 simulation samples of size n from the spherical Cauchy $C_d^*(\phi)$ with $\phi/\|\phi\| = e_1$ and (a) $d = 1$, (b) $d = 2$, (c) $d = 10$ and (d) $d = 100$

$\|\phi\|$	Est.	$n = 10$ RMSE (SE)	$n = 25$ RMSE (SE)	$n = 100$ RMSE (SE)	$n = \infty$ RMSE
			(a)		
$\eta_{1,1}^{-1}(0.1)$	MM	0.920 (0.011)	0.964 (0.008)	0.996 (0.005)	1.010
	AE	0.844 (0.005)	0.932 (0.002)	0.980 (0.001)	1.000
$\eta_{1,1}^{-1}(0.5)$	MM	1.123 (0.017)	1.250 (0.019)	1.304 (0.020)	1.333
	AE	0.936 (0.006)	0.981 (0.004)	0.994 (0.003)	1.000
$\eta_{1,1}^{-1}(0.9)$	MM	3.814 (0.153)	4.623 (0.145)	5.015 (0.145)	5.263
	AE	2.025 (0.065)	1.812 (0.044)	1.291 (0.023)	1.000
			(b)		
$\eta_{1,2}^{-1}(0.1)$	MM	0.940 (0.008)	0.980 (0.006)	0.997 (0.003)	1.005
	AE	0.898 (0.003)	0.959 (0.001)	0.990 (0.000)	1.000
$\eta_{1,2}^{-1}(0.5)$	MM	1.062 (0.011)	1.120 (0.011)	1.132 (0.011)	1.153
	AE	0.922 (0.004)	0.971 (0.002)	0.993 (0.001)	1.000
$\eta_{1,2}^{-1}(0.9)$	MM	1.902 (0.037)	2.149 (0.040)	2.154 (0.041)	2.234
	AE	1.059 (0.011)	1.044 (0.007)	1.012 (0.003)	1.000
			(c)		
$\eta_{1,10}^{-1}(0.1)$	MM	0.990 (0.002)	0.998 (0.001)	1.000 (0.001)	1.001
	AE	0.977 (0.000)	0.992 (0.000)	0.998 (0.000)	1.000
$\eta_{1,10}^{-1}(0.5)$	MM	1.014 (0.003)	1.026 (0.002)	1.028 (0.002)	1.027
	AE	0.978 (0.001)	0.992 (0.000)	0.998 (0.000)	1.000
$\eta_{1,10}^{-1}(0.9)$	MM	1.086 (0.005)	1.111 (0.005)	1.113 (0.005)	1.111
	AE	0.984 (0.001)	0.994 (0.000)	0.999 (0.000)	1.000
			(d)		
$\eta_{1,100}^{-1}(0.1)$	MM	0.999 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000
	AE	0.998 (0.000)	0.999 (0.000)	1.000 (0.000)	1.000
$\eta_{1,100}^{-1}(0.5)$	MM	1.002 (0.000)	1.002 (0.000)	1.002 (0.000)	1.003
	AE	0.998 (0.000)	0.999 (0.000)	1.000 (0.000)	1.000
$\eta_{1,100}^{-1}(0.9)$	MM	1.007 (0.000)	1.008 (0.000)	1.007 (0.000)	1.008
	AE	0.998 (0.000)	0.999 (0.000)	1.000 (0.000)	1.000

Next, we discuss the limits of the values of the relative mean squared error of the method of moments estimator (4.2) with respect to the maximum likelihood estimator as $n \to \infty$. Figure 1 displays the asymptotic relative mean squared error of the method of moments estimator (4.2) with respect to the maximum likelihood estimator as a function of $\|\phi\|$ or d. This figure implies

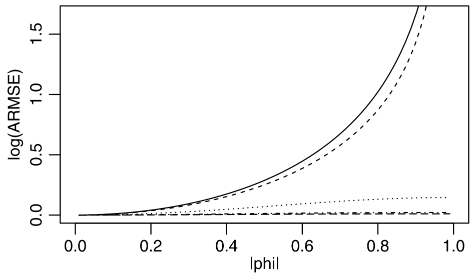

Figure 1. Asymptotic relative mean squared error of the method of moments estimator (4.2) with respect to the maximum likelihood estimator for the spherical Cauchy $C_d^*(\phi)$ as a function of $\|\phi\|$ for $d=1$ (solid), $d=2$ (dashed), $d=10$ (dotted), $d=50$ (dotdashed), and $d=100$ (longdashed). The vertical axis is plotted in logarithmic scale.

that, when $\|\phi\|$ is small, the asymptotic relative mean squared error is close to one for any d. The figure also suggests that the asymptotic relative mean squared error is monotonically increasing with respect to $\|\phi\|$. In particular, when d is small and $\|\phi\|$ is large, the asymptotic relative mean squared error is very large. As d increases, The asymptotic relative mean squared error approaches one for any $\|\phi\|$. The same discussion can be given to the relative mean squared error of the method of moments estimator (4.2) with respect to the asymptotically efficient estimator (4.6).

Given these observations, the following conclusions can be made as to the choice of the three estimators of the parameter of the spherical Cauchy in terms of mean squared error. If the dimension of the data is large, then the asymptotically efficient estimator (4.6) is preferred. This estimator outperforms both the maximum likelihood estimator and method of moments estimator (4.2) in terms of mean squared error for large d. When the dimension of the data is small, the asymptotically efficient estimator (4.6) is preferred for dispersed data and the maximum likelihood estimator is recommended otherwise.

The calculation of the asymptotically efficient estimator (4.6) is as efficient as that of the method of moments estimator (4.2) and is more efficient than that of the maximum likelihood estimator. However, our simulation study suggests that the convergence of the maximum likelihood estimation based on Algorithm 4.1 is very fast when n is not very small and d is greater than one. Actually, our computation for producing Table 1 implies that Algorithm 4.1 converges in almost all the combinations of $(d, n, \|\phi\|)$ when the method of moments estimator (4.2) is adopted as the initial value. To be more precise, using the stopping rule $\|\phi_t - \phi_{t-1}\| < 1 \times 10^{-7}$ and $t \leq 100$, Algorithm 4.1 failed to converge only once for $(d, n, \|\phi\|) = (1, 10, \eta_{1,1}^{-1}(0.1))$, $(d, n, \|\phi\|) = (1, 10, \eta_{1,1}^{-1}(0.5))$ and $(d, n, \|\phi\|) = (1, 10, \eta_{1,1}^{-1}(0.9))$ among 2000 simulation samples for each combination of $(d, n, \|\phi\|)$. When the stopping rule is relaxed to be $\|\phi_t - \phi_{t-1}\| < 1 \times 10^{-5}$ and $t \leq 100$, then Algorithm 4.1 converged in all the cases. Also, when $d = 1$, our simulation study implies that the maximum likelihood estimates estimated via Algorithm 4.1 numerically coincide with those estimated via the algorithm of [26] in the sense that the sum of squared differences between these two estimates is very small.

6. Comparison with von Mises–Fisher family

We compare the spherical Cauchy family with the von Mises–Fisher family or the Fisher–von Mises–Langevin family, which is a well-known family of distributions on the sphere. The von Mises–Fisher family on S^d has density

$$f(y) = \frac{\kappa^{(d-1)/2}}{(2\pi)^{(d+1)/2} I_{(d-1)/2}(\kappa)} \exp\bigl(\kappa\mu^T y\bigr), \quad y \in S^d, \tag{6.1}$$

where $\mu \in S^d$ controls the mode of the density, $\kappa \geq 0$ regulates the concentration of the distribution, and I_ν denotes the modified Bessel function of the first kind and order ν. For an S^d-valued random vector Y, its mean direction is defined by $E(Y)/\|E(Y)\|$ provided $E(Y) \neq 0$ and its mean resultant length by $\|E(Y)\|$. If Y has the von Mises–Fisher distribution (6.1), the mean direction and mean resultant length of Y are given by μ and $A_d(\kappa)$, respectively, where $A_d(\kappa) = I_{(d-1)/2}(\kappa)/I_{(d+1)/2}(\kappa)$. See, for example, for Section 9.3.2 of [32] and Section 2.3.1 of [29] for properties of the von Mises–Fisher family.

First, we discuss similarities and differences between the densities of the spherical Cauchy family (2.8) and von Mises–Fisher family (6.1). The densities of both families are unimodal and rotationally symmetric around their modes. If the mean resultant lengths are small, the densities of both models have similar shapes. However, when the mean resultant lengths are not small, the densities of the spherical Cauchy and von Mises–Fisher show different behaviour. Figure 2 displays densities and their log ratios of the spherical Cauchy distributions (2.8) and the von Mises–Fisher distributions (6.1) for some selected values of d, ϕ and $\kappa\mu$. The values of the concentration parameters are selected such that the mean resultant lengths of both models are 0.5 in Figure 2(a)–(c) and 0.9 in Figure 2(d). The figure suggests that, when the mean resultant lengths are not small, the spherical Cauchy density takes greater values than the von Mises–Fisher density around the mode and antimode and smaller values than the von Mises–Fisher density in the other area of the sphere. The comparison between Figure 2(a) and (b) implies that,

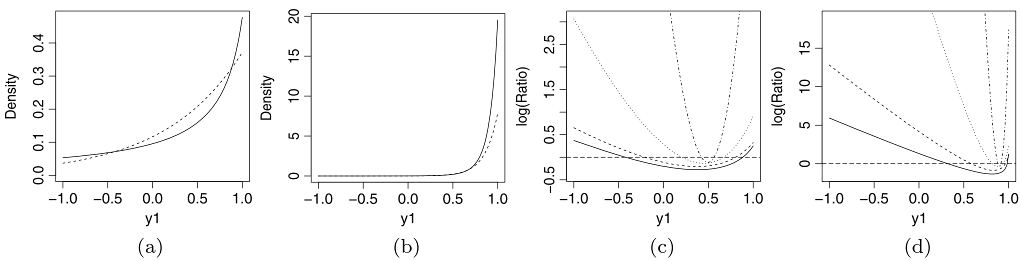

Figure 2. (a),(b): Density of the spherical Cauchy (solid) and von Mises–Fisher (dashed) as a function of y_1, for $\phi = \eta_{1,d}^{-1}(0.5)e_1$, $\kappa\mu = A_d^{-1}(0.5)e_1$ and (a) $d = 1$, (b) $d = 10$. (c),(d): The spherical Cauchy to von Mises–Fisher log density ratio as a function of y_1 for $d = 1$ (solid), $d = 2$ (dashed), $d = 10$ (dotted), and $d = 100$ (dotdashed). The parameters in (c) are $\phi = \eta_{1,d}^{-1}(0.5)e_1$, $\kappa\mu = A_d^{-1}(0.5)e_1$, and $\phi = \eta_{1,d}^{-1}(0.9)e_1$, $\kappa\mu = A_d^{-1}(0.9)e_1$ in (d). In (c) and (d), the longdashed line represents the horizontal line whose intercept is $\log(1)(= 0)$.

compared with the densities with $d = 2$, the densities with $d = 10$ take greater values around the mode. Figure 2(c) and (d) suggests that, the greater the value of d, the smaller the range of y_1 in which the von Mises–Fisher density takes greater values than the spherical Cauchy density. When the mean resultant lengths are large, the von Mises–Fisher density takes greater values than the spherical Cauchy density in a small range of y_1.

Next, we compare other statistical aspects of the spherical Cauchy family and von Mises–Fisher family. The von Mises–Fisher has a well-developed theory of statistical inference. Some tractable results about statistical inference for the von Mises–Fisher partly follow from the fact that, unlike the spherical Cauchy, the von Mises–Fisher is a member of the exponential family. The maximum likelihood estimator of the parameter for the von Mises–Fisher distribution can be expressed in closed form. On the other hand, a closed form expression for the maximum likelihood estimator of the parameter for the spherical Cauchy has not been found apart from $n \leq 4$ for $d = 1$ and $n \leq 3$ for $d \geq 2$. As for hypothesis testing, many tractable test statistics have been proposed in the literature for testing the location parameter and/or the concentration parameter of the von Mises–Fisher family in various settings. Apart from the use of pivotal statistics, methods of hypothesis testing for the spherical Cauchy do not seem immediately clear. However some general methods of hypothesis tests related to rotationally symmetric models are available, including tests of mode [37], tests of uniformity [4] and tests of concentration [5]. Also various extensions are available for the von Mises–Fisher distribution, including the Fisher–Bingham distribution [31] as a general and flexible model for any d and Kent distributions [24,25] as tractable and interpretable models for $d = 2$. There have not been extensions of the spherical Cauchy distribution at the moment apart from general constructions given, for example, in [30]. A flexible extension of the spherical Cauchy family is a potential future research topic.

The spherical Cauchy family has the tractable property that it forms a transformation model: the family is closed under the Möbius transformations on the sphere and there are similar induced transformations on the parameter space. This result can be applied to derive tractable properties of the spherical Cauchy family such as an efficient algorithm for random variate generation, a simple form of pivotal statistic, a closed form expression for probabilities of a surface area under the spherical Cauchy density. These properties do not hold for the von Mises–Fisher family in general. Furthermore, the spherical Cauchy family is related to the t-family with d degrees of freedom via the stereographic projection. A simple algorithm for maximum likelihood estimation and the asymptotically efficient estimator (4.6) enable us to use the spherical Cauchy, which has a different shape of the density from the von Mises–Fisher in general, as a practical statistical model. Since the Möbius transformations and/or the wrapped Cauchy family are applied to propose statistical models for circular data including regression models [7,23] and time series models [15,19], the theory of the Möbius transformations and/or the spherical Cauchy presented in this paper can be potentially useful for the development of statistical models for spherical data.

Appendix A: A marginal distribution of the Cauchy family on the sphere

A.1. A marginal distribution and real Möbius group

Here we discuss a marginal distribution of the spherical Cauchy family.

Theorem A.1. *Suppose $Y = (Y_1, \ldots, Y_{\nu+1})^T \sim C^*_\nu(\phi)$, where $\phi = (\rho, 0, \ldots, 0)^T$ and $\rho \in \mathbb{R} \setminus \{-1, 1\}$. Then the marginal density of Y_1 is of the form*

$$f(y_1; \rho, \nu) = \frac{1}{B(\nu/2, 1/2)} \left(\frac{|1-\rho^2|}{1+\rho^2 - 2\rho y_1} \right)^\nu \left(1 - y_1^2\right)^{(\nu-2)/2}, \quad -1 < y_1 < 1, \tag{A.1}$$

where $B(\cdot, \cdot)$ denotes the beta function.

The proof is straightforward and therefore omitted. It is important to discuss the marginal distribution (A.1) because this marginal is essentially the distribution of the inner product of a spherical Cauchy variable and its mean direction; if $\tilde{Y} \sim C^*(\tilde{\phi})$ with $\|\tilde{\phi}\| \neq 1$, then the distribution of $\mu^T \tilde{Y}$ has the density (A.1) with ρ replaced by $\|\tilde{\phi}\|$.

In a similar manner as in [33], if we view ν as a continuous-valued parameter with $\nu \geq 0$, then (A.1) can be considered a two-parameter family. Clearly, $f(y_1; \rho, \nu) = f(y_1; \rho^{-1}, \nu)$. If $\rho = 0$, then the distribution (A.1) reduces to the symmetric beta distribution with density

$$f(y_1; \nu) = \frac{(1 - y_1^2)^{(\nu-2)/2}}{B(\nu/2, 1/2)}, \quad -1 < y_1 < 1. \tag{A.2}$$

It can be readily seen from equation (8.384.5) of [13] that the family (A.1) with $-1 < \rho < 1$ is equivalent to Seshadri's family [40] with the parameterization given in Example 1 of his paper. As discussed there, if $\nu = 1$, then the family (A.1) reduces to the family discussed in [27,33] whose density is given by equation (2) of the latter paper.

Theorem A.2. *Let $\mathcal{R}$ be the real Möbius transformation*

$$\mathcal{R}(y_1) = \frac{y_1 + b}{b y_1 + 1}, \quad -1 < y_1 < 1; -1 < b < 1. \tag{A.3}$$

If Y_1 has the density (A.1), then $\mathcal{R}(Y_1)$ belongs to the same family with the parameter ρ replaced by $(\rho + \rho')/(\rho\rho' + 1)$, where $\rho' = (1 - \sqrt{1 - b^2})/b$.

The proof is clear from straightforward calculation and therefore omitted. Another approach to proving this result is to remember the derivation of the model given in Theorem A.1 and apply Theorem 2.3 with $R_1 = R_2 = I$ and $\phi_1 = (\rho, 0, \ldots, 0)^T$ and $\phi_2 = (\rho', 0, \ldots, 0)^T$.

This is an extension of the result given in [40] that the family (A.1) is transformed into the symmetric beta density (A.2) via a special case of the Möbius transformation (A.3) with $b = -2\rho/(1 + \rho^2)$.

A.2. Moments

We discuss some moments of the marginal family (A.1) which can be applied to obtain moments for the spherical Cauchy family. Define

$$\eta_{k,\nu}(\rho) = \mathbb{E}\left(Y_1^k\right) = \int_{-1}^{1} \frac{y_1^k}{B(\nu/2, 1/2)} \left(\frac{|1-\rho^2|}{1+\rho^2 - 2\rho y_1} \right)^\nu \left(1 - y_1^2\right)^{(\nu-2)/2} dy_1, \quad -1 < \rho < 1,$$

where Y_1 has the density (A.1). As the following lemma shows, the monotonicity holds for $\eta_{k,\nu}$ for an odd integer of k.

Lemma A.1. *Suppose that k is an odd integer. Then $\eta_{k,\nu}(0) = 0$, $\lim_{\rho\to 1} \eta_{k,\nu}(\rho) = 1$, and $\partial\eta_{k,\nu}(\rho)/\partial\rho > 0$ for $0 < \rho < 1$.*

Next the first and second moments of the marginal family (A.1) are discussed. Seshadri [40] obtained closed-form expressions for the mean and variance of the family (A.1) with $\nu = 1$ and approximated values of these statistics with general ν. Here we provide exact expressions for the moments for general ν.

Theorem A.3. *The following hold for $\eta_{1,\nu}(\rho)$:*

(i) *for any $\nu \geq 0$,*

$$\begin{aligned}\eta_{1,\nu}(\rho) &= \frac{1+\rho^2}{2\rho}\left[1 - \frac{(1+\rho)^2}{1+\rho^2} F\left\{1, \frac{\nu}{2}; \nu; -\frac{4\rho}{(1-\rho)^2}\right\}\right]\\ &= \frac{1+\rho^2}{2\rho}\left[1 - \frac{1-\rho^2}{1+\rho^2} F\left\{\frac{1}{2}, \frac{\nu-1}{2}; \frac{\nu+1}{2}; -\frac{4\rho^2}{(1-\rho^2)^2}\right\}\right],\end{aligned}$$

where F denotes the hypergeometric series [[13], *equation* (9.111)],

(ii) *for $\nu = 1, \ldots, 4$,*

$$\begin{aligned}&\eta_{1,1}(\rho) = \rho, \qquad \eta_{1,2}(\rho) = \frac{1+\rho^2}{2\rho}\left\{1 - \frac{(1-\rho^2)^2}{2\rho(1+\rho^2)} \log\left(\frac{1+\rho}{1-\rho}\right)\right\},\\ &\eta_{1,3}(\rho) = \frac{\rho(3-\rho^2)}{2},\\ &\eta_{1,4}(\rho) = \frac{1+\rho^2}{2\rho}\left\{1 - \frac{3(1-\rho^2)^2}{8\rho^2} + \frac{3}{16\rho^3}\frac{(1-\rho^4)^4}{1+\rho^2}\log\left(\frac{1+\rho}{1-\rho}\right)\right\},\end{aligned}$$

(iii) *for $\nu \geq 4$,*

$$\begin{aligned}\eta_{1,\nu}(\rho) = \frac{\nu-1}{(\nu-2)(\nu-3)}\Bigg[&\left\{\nu - 2 + \frac{(\nu-3)(1-\rho^2)^2}{4\rho^2}\right\}\mu_1(\nu-2)\\ &- \frac{(\nu-3)(1-\rho^2)^2}{4\rho^2}\mu_1(\nu-4) - \frac{\nu-2}{\nu-1}\frac{1+\rho^2}{\rho}\Bigg].\end{aligned}$$

It follows from these results that, for any positive integer ν, the mean of $\eta_{1,\nu}(\rho)$ can be expressed in closed form.

Theorem A.4. *The following results hold for $\eta_{2,\nu}(\rho)$:*

(i) *for* $\nu \geq 0$,

$$\eta_{2,\nu}(\rho) = \frac{(1+\rho^2)^2}{4\rho^2}\left[1 - 2\frac{(1+\rho)^2}{1+\rho^2}F\left\{1, \frac{\nu}{2}; \nu; -\frac{4\rho}{(1-\rho)^2}\right\}\right.$$
$$\left. + \frac{(1+\rho)^4}{(1+\rho^2)^2}F\left\{2, \frac{\nu}{2}; \nu; -\frac{4\rho}{(1-\rho)^2}\right\}\right], \tag{A.4}$$

(ii) *for* $\nu = 1, \ldots, 4$,

$$\eta_{2,1}(\rho) = \frac{1+\rho^2}{2}, \qquad \eta_{2,2}(\rho) = \frac{1+\rho^2}{4\rho^2}\left\{\frac{2(1+\rho^4)}{1+\rho^2} - \frac{(1-\rho^2)^2}{\rho}\log\left(\frac{1+\rho}{1-\rho}\right)\right\},$$
$$\eta_{2,3}(\rho) = \frac{1+6\rho^2-3\rho^4}{4},$$
$$\eta_{2,4}(\rho) = \frac{1+\rho^2}{16\rho^4}\left\{\frac{-2(3-8\rho^2+2\rho^4-8\rho^6+3\rho^8)}{1+\rho^2} + \frac{3(1-\rho^2)^4}{\rho}\log\left(\frac{1+\rho}{1-\rho}\right)\right\},$$

(iii) *for* $\nu > 4$,

$$\eta_{2,\nu}(\rho) = \frac{1}{(\nu-3)(\nu-4)}\left[-\frac{3(1+\rho^2)^2}{2\rho^2} + \frac{1+\rho^2}{\rho}\{(\nu-3)(\nu-4)\eta_{1,\nu}(\rho)\right.$$
$$\left. - c_1\eta_{1,\nu-2}(\rho) + c_2\eta_{1,\nu-4}(\rho)\} + c_1\eta_{2,\nu-2}(\rho) - c_2\eta_{2,\nu-4}(\rho)\right],$$

where $c_1 = (\nu-1)(\nu-6) - (\nu-1)(\nu-3)(1-\rho^2)^2/(4\rho^2)$ *and* $c_2 = -(\nu-1)(\nu-3)(1-\rho^2)^2/(4\rho^2)$.

It follows from these results and equation (9.134.3) of [13] that $\eta_{2,\nu}(\rho)$ has a closed-form expression for any $\nu \in \mathbb{N}$. Thus the variance of Y_1 can also be expressed in closed from for any positive integer ν.

Theorems A.3 and A.4 can be applied to express certain moments of the spherical Cauchy family (see Section 4.1).

Acknowledgments

Kato is grateful to Department of Statistics at the University of Chicago for its hospitality during the research visit that led to this paper. We thank Tomonari Sei for informing us of the references [11] and [38] and the Associate Editor and referees for their helpful comments. The work of Kato was supported by JSPS KAKENHI Grant Number 17K05379.

Supplementary Material

Supplement to "Some properties of a Cauchy family on the sphere derived from the Möbius transformations" (DOI: 10.3150/20-BEJ1222SUPP; .pdf). The supplement [21] provides proofs for claims made in Sections 2, 3 and 4 and Appendix A. Specifically, we prove Lemma 2.1 and Theorems 2.1 and 2.3 of Section 2, Theorems 3.1 and 3.2 of Section 3, Lemma 4.1 and Theorems 4.1, 4.3, 4.5 and 4.7 of Section 4, and Lemma A.1 and Theorems A.3 and A.4 of Appendix A.

References

[1] Ahlfors, L.V. (1978). *Complex Analysis*: *An Introduction to the Theory of Analytic Functions of One Complex Variable*, 3rd ed. New York: McGraw-Hill. MR0510197

[2] Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*. *Monographs on Statistics and Applied Probability* **52**. London: CRC Press. MR1317097

[3] Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann*. *Statist*. **31** 1–29. MR1962498 https://doi.org/10.1214/aos/1046294456

[4] Cutting, C., Paindaveine, D. and Verdebout, T. (2017). Testing uniformity on high-dimensional spheres against monotone rotationally symmetric alternatives. *Ann*. *Statist*. **45** 1024–1058. MR3662447 https://doi.org/10.1214/16-AOS1473

[5] Cutting, C., Paindaveine, D. and Verdebout, T. (2017). Tests of concentration for low-dimensional and high-dimensional directional data. In *Big and Complex Data Analysis*. *Contrib*. *Stat*. 209–227. Cham: Springer. MR3674718

[6] Downs, T.D. (2003). Spherical regression. *Biometrika* **90** 655–668. MR2006842 https://doi.org/10.1093/biomet/90.3.655

[7] Downs, T.D. and Mardia, K.V. (2002). Circular regression. *Biometrika* **89** 683–697. MR1929172 https://doi.org/10.1093/biomet/89.3.683

[8] Dunau, J.-L. and Sénateur, H. (1988). Une caracterisation du type de la loi de Cauchy-conforme sur $\mathbf{R}^n$. *Probab*. *Theory Related Fields* **77** 129–135. MR0921823 https://doi.org/10.1007/BF01848135

[9] Ferguson, T. (1978). Maximum likelihood estimates of the parameters of the Cauchy distribution for samples of size 3 and 4. *J*. *Amer*. *Statist*. *Assoc*. **73** 211–213.

[10] Ferguson, T.S. (1996). *A Course in Large Sample Theory*. *Texts in Statistical Science Series*. London: CRC Press. MR1699953

[11] Gabrielsen, G. (1986). Global maxima of real-valued functions. *J*. *Optim*. *Theory Appl*. **50** 257–266. MR0857236 https://doi.org/10.1007/BF00939272

[12] García-Portugués, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron*. *J*. *Stat*. **7** 1655–1685. MR3070874 https://doi.org/10.1214/13-EJS821

[13] Gradshteyn, I.S. and Ryzhik, I.M. (2007). *Table of Integrals*, *Series*, *and Products*, 7th ed. San Diego, CA: Academic Press.

[14] Hendriks, H. and Landsman, Z. (1996). Asymptotic behavior of sample mean location for manifolds. *Statist*. *Probab*. *Lett*. **26** 169–178. MR1381468 https://doi.org/10.1016/0167-7152(95)00007-0

[15] Holzmann, H., Munk, A., Suster, M. and Zucchini, W. (2006). Hidden Markov models for circular and linear-circular time series. *Environ*. *Ecol*. *Stat*. **13** 325–347. MR2242193 https://doi.org/10.1007/s10651-006-0015-7

[16] Huckemann, S.F., Kim, P.T., Koo, J.-Y. and Munk, A. (2010). Möbius deconvolution on the hyperbolic plane with application to impedance density estimation. *Ann. Statist.* **38** 2465–2498. MR2676895 https://doi.org/10.1214/09-AOS783

[17] Iwaniec, T. and Martin, G. (2001). *Geometric Function Theory and Non-linear Analysis*. *Oxford Mathematical Monographs*. New York: Oxford University Press. MR1859913

[18] Jones, M.C. (2004). The Möbius distribution on the disc. *Ann. Inst. Statist. Math.* **56** 733–742. MR2126808 https://doi.org/10.1007/BF02506486

[19] Kato, S. (2010). A Markov process for circular data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 655–672. MR2758240 https://doi.org/10.1111/j.1467-9868.2010.00748.x

[20] Kato, S. and Jones, M.C. (2010). A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *J. Amer. Statist. Assoc.* **105** 249–262. MR2656051 https://doi.org/10.1198/jasa.2009.tm08313

[21] Kato, S. and McCullagh P. (2020). Supplement to "Some properties of a Cauchy family on the sphere derived from the Möbius transformations." https://doi.org/10.3150/20-BEJ1222SUPP

[22] Kato, S. and Pewsey, A. (2015). A Möbius transformation-induced distribution on the torus. *Biometrika* **102** 359–370. MR3371009 https://doi.org/10.1093/biomet/asv003

[23] Kato, S., Shimizu, K. and Shieh, G.S. (2008). A circular-circular regression model. *Statist. Sinica* **18** 633–645. MR2432283

[24] Kent, J.T. (1982). The Fisher-Bingham distribution on the sphere. *J. Roy. Statist. Soc. Ser. B* **44** 71–80. MR0655376

[25] Kent, J.T., Hussein, I. and Jah, M.K. (2016). Directional distributions in tracking of space debris. In Proceedings of the 19th International Conference on Information Fusion (FUSION) 2081–2086.

[26] Kent, J.T. and Tyler, D.E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *J. Appl. Stat.* **15** 247–254.

[27] Leipnik, R.B. (1947). Distribution of the serial correlation coefficient in a circularly correlated universe. *Ann. Math. Stat.* **18** 80–87. MR0019892 https://doi.org/10.1214/aoms/1177730494

[28] Letac, G. (1986). Seul le groupe des similitudes-inversions préserve le type de la loi de Cauchy-conforme de $\mathbf{R}^n$ pour $n > 1$. *J. Funct. Anal.* **68** 43–54. MR0850153 https://doi.org/10.1016/0022-1236(86)90056-X

[29] Ley, C. and Verdebout, T. (2017). *Modern Directional Statistics*. *Chapman & Hall/CRC Interdisciplinary Statistics Series*. Boca Raton, FL: CRC Press. MR3752655

[30] Ley, C. and Verdebout, T. (2017). Skew-rotationally-symmetric distributions and related efficient inferential procedures. *J. Multivariate Anal.* **159** 67–81. MR3668548 https://doi.org/10.1016/j.jmva.2017.02.010

[31] Mardia, K.V. (1975). Statistics of directional data. *J. Roy. Statist. Soc. Ser. B* **37** 349–393. MR0402998

[32] Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR1828667

[33] McCullagh, P. (1989). Some statistical properties of a family of continuous univariate distributions. *J. Amer. Statist. Assoc.* **84** 125–129. MR0999670

[34] McCullagh, P. (1992). Conditional inference and Cauchy models. *Biometrika* **79** 247–259. MR1185127 https://doi.org/10.1093/biomet/79.2.247

[35] McCullagh, P. (1996). Möbius transformation and Cauchy parameter estimation. *Ann. Statist.* **24** 787–808. MR1394988 https://doi.org/10.1214/aos/1032894465

[36] McLachlan, G.J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*, 2nd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley Interscience. MR2392878 https://doi.org/10.1002/9780470191613

[37] Paindaveine, D. and Verdebout, T. (2017). Inference on the mode of weak directional signals: A Le Cam perspective on hypothesis testing near singularities. *Ann. Statist.* **45** 800–832. MR3650401 https://doi.org/10.1214/16-AOS1468

[38] Ratcliffe, J.G. (2019). *Foundations of Hyperbolic Manifolds*, 3rd ed. *Graduate Texts in Mathematics* **149**. New York: Springer.

[39] Rudin, W. (1987). *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill.

[40] Seshadri, V. (1991). A family of distributions related to the McCullagh family. *Statist*. *Probab*. *Lett*. **12** 373–378. MR1142090 https://doi.org/10.1016/0167-7152(91)90025-M

[41] Uesu, K., Shimizu, K. and SenGupta, A. (2015). A possibly asymmetric multivariate generalization of the Möbius distribution for directional data. *J*. *Multivariate Anal*. **134** 146–162. MR3296040 https://doi.org/10.1016/j.jmva.2014.11.004

[42] Watson, G.S. (1983). *Statistics on Spheres*. *University of Arkansas Lecture Notes in the Mathematical Sciences* **6**. New York: Wiley. MR0709262

Received February 2019 and revised March 2020